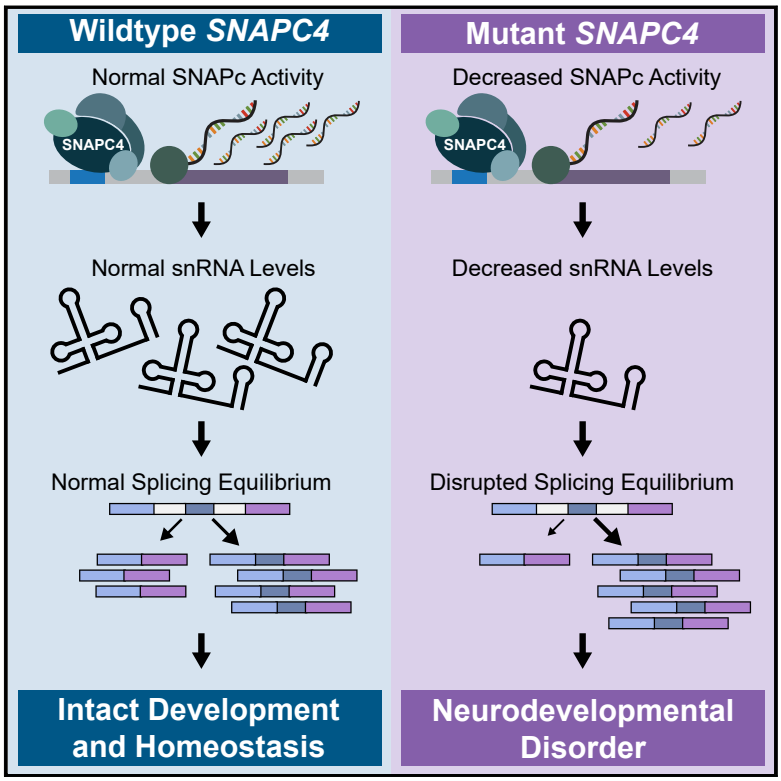
Bi-allelic *SNAPC4* variants dysregulate global alternative splicing and lead to neuroregression and progressive spastic paraparesis

Graphical abstract



Authors

F. Graeme Frost, Marie Morimoto, Prashant Sharma, ..., Cynthia J. Tifft, William A. Gahl, May Christine V. Malicdan

Correspondence

maychristine.malicdan@nih.gov

Frost et al. demonstrate that a neurodevelopmental disorder is associated with deleterious bi-allelic variants in *SNAPC4*, which encodes a protein required for the transcription of spliceosomal small nuclear RNAs. These loss-of-function *SNAPC4* variants lead to global splicing dysregulation, implicating a potential mechanism for disease pathology.

Frost et al., 2023, The American Journal of Human Genetics *110*, 663–680
April 6, 2023
https://doi.org/10.1016/j.ajhg.2023.03.001

Bi-allelic *SNAPC4* variants dysregulate global alternative splicing and lead to neuroregression and progressive spastic paraparesis

F. Graeme Frost,[1] Marie Morimoto,[1] Prashant Sharma,[1] Lyse Ruaud,[2,3] Newell Belnap,[4] Daniel G. Calame,[5,6,7] Yuri Uchiyama,[8,9] Naomichi Matsumoto,[9] Machteld M. Oud,[10] Elise A. Ferreira,[11,12] Vinodh Narayanan,[4] Sampath Rangasamy,[4] Matt Huentelman,[4] Lisa T. Emrick,[5,6,7] Ikuko Sato-Shirai,[13,14] Satoko Kumada,[13] Nicole I. Wolf,[15] Peter J. Steinbach,[16] Yan Huang,[1] Undiagnosed Diseases Network, Barbara N. Pusey,[1] Sandrine Passemard,[3,17] Jonathan Levy,[18,19] Séverine Drunat,[18,19,20] Marie Vincent,[21,22] Agnès Guet,[23] Emanuele Agolini,[24] Antonio Novelli,[24] Maria Cristina Digilio,[25] Jill A. Rosenfeld,[6] Jennifer L. Murphy,[1] James R. Lupski,[5,6,7,26,27] Gilbert Vezina,[28] Ellen F. Macnamara,[1] David R. Adams,[1,29] Maria T. Acosta,[1] Cynthia J. Tifft,[1,29] William A. Gahl,[1,30] and May Christine V. Malicdan[1,30,*]

Summary

The vast majority of human genes encode multiple isoforms through alternative splicing, and the temporal and spatial regulation of those isoforms is critical for organismal development and function. The spliceosome, which regulates and executes splicing reactions, is primarily composed of small nuclear ribonucleoproteins (snRNPs) that consist of small nuclear RNAs (snRNAs) and protein subunits. snRNA gene transcription is initiated by the snRNA-activating protein complex (SNAPc). Here, we report ten individuals, from eight families, with bi-allelic, deleterious *SNAPC4* variants. *SNAPC4* encoded one of the five SNAPc subunits that is critical for DNA binding. Most affected individuals presented with delayed motor development and developmental regression after the first year of life, followed by progressive spasticity that led to gait alterations, paraparesis, and oromotor dysfunction. Most individuals had cerebral, cerebellar, or basal ganglia volume loss by brain MRI. In the available cells from affected individuals, SNAPC4 abundance was decreased compared to unaffected controls, suggesting that the bi-allelic variants affect SNAPC4 accumulation. The depletion of SNAPC4 levels in HeLa cell lines via genomic editing led to decreased snRNA expression and global dysregulation of alternative splicing. Analysis of available fibroblasts from affected individuals showed decreased snRNA expression and global dysregulation of alternative splicing compared to unaffected cells. Altogether, these data suggest that these bi-allelic *SNAPC4* variants result in loss of function and underlie the neuroregression and progressive spasticity in these affected individuals.

Introduction

Alternative splicing is a ubiquitous mechanism within eukaryotic genomes, particularly in the human genome, where 95% of exonic genes encode multiple isoforms.[1,2] This ubiquity not only expands transcriptomic complexity but also serves as a post-transcriptional mechanism of modulating gene expression and function that is critically important in both development[3–5] and maintenance of homeostasis in differentiated tissue.[6–9] This process has

[1]National Institutes of Health Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD, USA; [2]APHP.Nord, Robert Debré University Hospital, Department of Genetics, Paris, France; [3]Université Paris Cité, Inserm UMR 1141, NeuroDiderot, 75019 Paris, France; [4]Center for Rare Childhood Disorders, The Translational Genomics Research Institute, Phoenix, AZ, USA; [5]Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; [6]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; [7]Texas Children's Hospital, Houston, TX, USA; [8]Department of Rare Disease Genomics, Yokohama City University Hospital, Yokohama, Japan; [9]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; [10]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; [11]Department of Pediatrics, Emma Children's Hospital, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam University Medical Centers, Amsterdam, the Netherlands; [12]United for Metabolic Diseases, Amsterdam, the Netherlands; [13]Department of Neuropediatrics, Tokyo Metropolitan Neurological Hospital, Tokyo, Japan; [14]Department of Pediatrics, Shimada Ryoiku Medical Center Hachioji for Challenged Children, Tokyo, Japan; [15]Amsterdam Leukodystrophy Center, Department of Child Neurology, Emma Children's Hospital, Amsterdam University Medical Centers, and Amsterdam Neuroscience, Cellular & Molecular Mechanisms, Vrije Universiteit, Amsterdam, the Netherlands; [16]Bioinformatics and Computational Biosciences Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA; [17]Service de Neurologie Pédiatrique, DMU INOV-RDB, APHP, Hôpital Robert Debré, Paris, France; [18]Department of Genetics, APHP-Robert Debré University Hospital, Paris, France; [19]Laboratoire de biologie médicale multisites Seqoia - FMG2025, Paris, France; [20]INSERM UMR1141, Neurodiderot, University of Paris, Paris, France; [21]Service de Génétique Médicale, CHU Nantes, Nantes, France; [22]Inserm, CNRS, University Nantes, l'institut du thorax, Nantes, France; [23]APHP.Nord, Louis Mourier Hospital, Pediatrics Department, Paris, France; [24]Laboratory of Medical Genetics, Translational Cytogenomics Research Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; [25]Medical Genetics Unit, Bambino Gesù Children Hospital, IRCCS, Rome, Italy; [26]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; [27]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; [28]Department of Diagnostic Radiology and Imaging, Children's National Hospital, Washington, DC, USA; [29]Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; [30]Human Biochemical Genetics Section, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

*Correspondence: maychristine.malicdan@nih.gov

https://doi.org/10.1016/j.ajhg.2023.03.001.

been well characterized in the developing nervous system, where differential alternative splicing can either serve as a master switch over larger transcriptional programs[10,11] or be the mechanistic target of a developmental switch.[12] The importance of alternative splicing within the nervous system is underscored by the neurological disorders linked to its dysfunction, including amyotrophic lateral sclerosis (ALS [MIM: 105400]),[13] autism (MIM: 209850),[14] Huntington disease (MIM: 143100),[15] and NSRP1-related disorders (MIM: 620001).[16]

Splicing is facilitated by small nuclear ribonucleoproteins (snRNPs),[17] which are composed of a small nuclear RNA (snRNA) and several protein subunits. The snRNA components of snRNPs mediate the splicing reaction by base pairing with canonical splicing motifs on pre-mRNAs, mediating RNA-RNA interactions between snRNPs, and facilitating broader RNA-protein reactions within the spliceosome.[18] There are five spliceosomal snRNPs in the major spliceosome, i.e., U1, U2, U4, U5, and U6; each plays a distinct role in catalyzing the splicing reaction. The U1 and U2 snRNPs are responsible for binding the 5′ splice site and branch splice site, respectively, while the U4, U5, and U6 snRNPs form a tri-snRNP complex that is required for assembly and function of the catalytic spliceosome.[18]

The biogenesis of snRNPs is a multifaceted process involving post-transcriptional modifications, assembly of RNA and protein subunits, and trafficking between cellular compartments.[19] snRNP biogenesis starts with the transcription of snRNA genes,[20] all of which share a common promoter sequence, the proximal sequence element (PSE), which is recognized by the transcription-initiating snRNA-activating protein complex (SNAPc).[21] SNAPc binding to the PSE is mediated through the SNAPC4 subunit, while other subunits (SNAPC1, 2, 3, and 5) modify PSE-binding affinity or mediate interactions with transcriptional co-activators.[22,23] There is no human disease associated with variants in *SNAPC4* (MIM: 602777).

In this study, we report ten individuals, from eight families, with bi-allelic variants in *SNAPC4* who presented with a neurodevelopmental disorder (NDD). The disease course involved developmental and motor regression after the first year of life, and progressive, ascending spasticity with increased tone and deep tendon reflexes (DTRs), paraparesis, and oromotor dysfunction. Brain imaging revealed atrophy of the cerebellum, cerebrum, and basal ganglia. To investigate the consequence of the SNAPC4 loss-of-function (LOF) variants, we generated a *SNAPC4*-deficient HeLa cell model and demonstrated a reduction in snRNA expression that was associated with widespread alternative splicing dysregulation. More importantly, we observed similar reductions in snRNA expression and dysregulation of alternative splicing in fibroblasts from affected individuals with LOF variants in *SNAPC4*, in keeping with the role of SNAPC4 in modulating snRNA expression and global alternative splicing. These results, paired with the consistent neuroregression and progressive spasticity observed in these affected individuals, make *SNAPC4* a compelling disease-associated gene for the neurodevelopmental disease trait.

Material and methods

Affected individuals and clinical evaluation

Individual 1 was evaluated through research protocols of the National Institutes of Health Undiagnosed Diseases Program[24–26] (76-HG-0238) and Undiagnosed Diseases Network (15-HG-0130) approved by the NHGRI IRB. Individual 2 was evaluated and followed at Texas Children's Hospital (Houston, TX, USA), and she, her parents, and unaffected siblings were consented under Baylor College of Medicine IRB-approved protocol H-29697. Individuals 3 and 4 were clinically assessed and followed at the Center for Rare Childhood Disorders at the Translational Genomics Research Institute (TGen) (Phoenix, AZ, USA) and were enrolled under the Center for Rare Childhood Disorders (C4RCD) research protocol at TGen (WIRB # 20120789) approved by the Western IRB. Individual 5 was evaluated under a study protocol approved by the IRBs of the Faculty of Medicine, Yokohama City University, Japan. Individual 6 was evaluated at Amsterdam University Medical Centers and consented under a research protocol (NL67721.018.19) approved by the Amsterdam University Medical Center IRB. Individuals 7 and 8 were evaluated through research projects approved by the National Ethics Committee (Comité de Protection des Personnes, Ile-de France II, number 2010OA1481-38) and registered at ClinicalTrials.gov (NCT01565005). Individual 9 was clinically assessed and followed at Robert Debré Hospital and gave written informed consent to DNA analysis and associated studies (Paris, France). Individual 10 was evaluated and followed at Bambino Gesù Children's Hospital (Rome, Italy) and was consented under the hospital's diagnostic procedure protocol. Written consent for all the participants under the age of 18 years was obtained from the parents or guardians.

Sequencing

Quartet genome sequencing (GS) was performed on individual 1 at HudsonAlpha Institute for Biotechnology (Huntsville, AL) through the Undiagnosed Diseases Network. Individual 2 underwent clinical trio exome sequencing (ES) at Baylor Genetics (Houston, TX). Individual 3 underwent clinical trio ES at GeneDx (Gaithersburg, MD). Individual 4 underwent clinical trio ES at TGen (Phoenix, AZ). Individual 5 underwent clinical trio ES at Yokohama City University Graduate School of Medicine (Yokohama, Japan). Individual 6 underwent clinical trio ES by BGI (Shenzhen, China). Individuals 7 and 8 underwent clinical trio ES (Intragen, Evry, France), while individual 9 underwent clinical trio GS (LBMS SeqOIA, Paris, France). Individual 10 underwent an expanded custom clinical exome sequencing that included more than 10,000 genes (Twist Bioscience, California, USA). We utilized GeneMatcher[27] to connect the researchers and clinicians.

Cell culture

Primary dermal fibroblasts from affected individuals 1, 2, 3, 4, and 6 were cultured from a forearm skin biopsy.[28] Unaffected primary dermal fibroblasts GM01652, GM07522, and GM09503 (Coriell Institute for Medical Research) were used as controls. Fibroblasts were cultured in high glucose DMEM (11965092, Thermo Fisher Scientific) with 10% FBS (10082, Thermo Fisher Scientific), and

1× antibiotic-antimycotic (15240, Thermo Fisher Scientific) at 37°C with 5% CO_2.

We isolated lymphocytes from whole blood of individual 5 and transformed them with Epstein-Barr virus to generate a lymphoblastoid cell line. Unaffected lymphoblastoid cell lines AG09392 and AG15022 (Coriell Institute for Medical Research) were used as controls. Lymphoblastoid cell lines were cultured in RPMI (61870127, Thermo Fisher Scientific) with 10% FBS (10082, Thermo Fisher Scientific) and 1× antibiotic-antimycotic (15240, Thermo Fisher Scientific) at 37°C with 5% CO_2.

RNA extraction and reverse transcription

Total RNA was extracted from cells with the RNeasy Mini Kit (74104, Qiagen), following the manufacturer's protocol. cDNA was then synthesized with the Omniscript RT Kit (205111, Qiagen) and random nonamer primers (R7647, Millipore Sigma) at a final concentration of 10 μM, following the manufacturer's protocol.

For individual 7, total RNA from dermal fibroblasts was extracted with a NucleoSpin RNA extraction kit (Macherey-Nagel, Germany) according to the manufacturer's protocol. Extracted mRNA was analyzed for quality and concentration by spectrophotometry (Nanodrop2000, Thermo Fisher Scientific, Waltham, MA, USA). Reverse transcription was performed with an iScript cDNA synthesis kit (BioRad, Hercules, CA, USA).

Quantitative PCR

Relative expression of mRNA was analyzed via TaqMan Gene Expression Assays with the TaqMan Gene Expression Assay Master Mix (4370048, Thermo Fisher Scientific) and the 7500 Fast Real-Time PCR System (Applied Biosystems). A full list of probes and normalization is listed in the supplemental information (Table S2 and supplemental methods, respectively).

cDNA sequencing

To determine the consequence of the splice site variants, we performed RT-PCR followed by TOPO cloning to isolate and sequence single alleles by using the TOPO TA Cloning Kit for Sequencing (450030, Invitrogen), following the manufacturer's protocol. Details are provided in the supplemental methods.

Nonsense-mediated mRNA decay (NMD) analysis

Nonsense-mediated mRNA decay in cultured lymphoblastoid cells from individual 5 was examined by using cycloheximide treatment. DMSO with or without 30 μM cycloheximide (Sigma) was added to the culture medium for 5 h, and total RNA was isolated with RNeasy Mini Kit (74104, Qiagen). cDNA was synthesized from 2 μg of total RNA with random hexamers via the PrimeScript 1st Strand cDNA Synthesis Kit (Takara Bio), and 50 ng of cDNA was used for RT-PCR. PCR products were resolved on 1% agarose gels and detected by ethidium bromide staining.

Sanger sequencing

PCR products were enzymatically cleaned up with the ExoSAP-IT Express PCR Product Cleanup Reagent (75001.200.UL, Applied Biosystems), following the manufacturer's protocol. The clean PCR product was then amplified with the BigDye Terminator v3.1 Cycle Sequencing Kit (4337457, Applied Biosystems), and the sequencing reaction was purified with the BigDye XTerminator Purification Kit (4376485, Applied Biosystems), following the manufacturer's protocol. Sequencing reactions were separated by capillary electrophoresis with the SeqStudio Genetic Analyzer System (A35645, Applied Biosystems). Sequences were analyzed with Sequencher (version 5.4.6, build 46289, Gene Codes).

Generation of a HeLa *SNAPC4*-deficient cell line

We used CRISPR-Cas9 technology to generate a *SNAPC4*-deficient HeLa cell line. In brief, we used guide RNAs (gRNAs) targeting exons 2, 3, and 4 of *SNAPC4* (Figure S1B, Table S1), and we isolated single clones by using limiting dilution. We genotyped clonal populations by using primers designed to encompass the targeted region (Figure S1C). We extracted DNA from the clones of interest (control clone #1, KO clone #12, and KO clone #19) by using the DNeasy Blood and Tissue Kit (Qiagen). The region was then amplified via the Multiplex PCR Kit (206143, Qiagen) with the primers listed in Table S1. Deficiency of *SNAPC4* mRNA expression was confirmed by quantitative PCR and immunoblot. TaqMan assay IDs used for quantitative PCR and antibodies for immunoblot are listed in Tables S2 and S3, respectively.

Immunoblot analysis

Cells were harvested from 70% to 90% confluent flasks of dermal fibroblasts, washed with PBS, and lysed with 2× SDS lysis buffer (250 mM Tris-HCl [pH 6.8], 4% SDS, 10% glycerol, Complete Protease Inhibitor Cocktail [Roche], and PhosSTOP phosphatase inhibitor cocktail [Roche]). Lysates were kept at room temperature for 5 min, followed by sonication. After centrifugation, supernatants were transferred to a new tube and protein concentrations were measured with the DC Protein Assay (Bio-Rad Laboratories). Total protein was resolved by SDS-PAGE on 4%–15% Tris-Glycine Gels (Bio-Rad Laboratories) and transferred to nitrocellulose membranes via a Trans-Blot Turbo Transfer System (Bio-Rad Laboratories). Immunoblotting was carried out as described previously.[29] Full, uncropped images of all blots are included in the supplemental information (Figure S2).

RNA-sequencing (RNA-seq) analysis

For bulk RNA-seq, total RNA was extracted from HeLa cells with the Maxwell RSC simplyRNA Cells Kit (AS1390, Promega). Full details of library preparation are described in the supplemental information (supplemental methods). Briefly, we constructed stranded poly(A)-selected mRNA libraries from total RNA and sequenced these on an S4 flow cell on a NovaSeq 6000 by using version 1.5 chemistry to achieve a minimum of 50 million 150 base read pairs. The data were processed with RTA version 3.4.4. Reads were then aligned to GRCh37 with HISAT2,[30] and read counts were generated with DESeq2.[31] Differential expression was determined with the *results()* function in the DESeq2[31] R package. All differentially expressed genes (adjusted $p \leq 0.05$) were then tested for functional enrichment with the WebGestaltR[32] R package. To analyze alternative splicing in the transcriptome, we ran rMATS[33] with the *--variable-read-length* and *–allow-clipping flags* functions by using the BAM files generated from the read alignment of samples.

For small RNA-seq, total RNA was extracted from HeLa cells and fibroblasts with the Maxwell RSC miRNA from Tissue Kit (AS1460, Promega) according to the manufacturer's specifications. Small RNA-seq was carried out by BGI Genomics (Shenzhen, China). Briefly, total RNA was size selected for fragments between 50 bp and 200 bp, which were sequenced with 50 bp single-end reads at an average depth of 20 million reads per sample. Reads were then aligned to GRCh37 by BGI. Next, counts were generated

with DESeq2.[31] Differential expression was determined with the *results()* function in the DESeq2[31] R package.

Protein structure analysis

The structure of SNAPC4 bound to DNA was modeled for residues Lys 345 to Gly 500, using the Prime software tools (Schrodinger), with the crystal structure 1h88.pdb[34] as the template. Of the 156 residues modeled, 51 (33%) were identical to the nearest residue in the template structure. The model was rendered with the programs MolScript[35] and Raster3D.[36]

Statistical analysis

Differences in *SNAPC4* mRNA, SNAPC4 protein, and various snRNAs between the affected individual and control groups were evaluated with either a two-sample t test ($\alpha = 0.05$) if Shapiro's test showed normal distribution of data within each group ($p > 0.05$) or Mann-Whitney U test ($\alpha = 0.05$) if Shapiro's test showed nonnormal distribution of data within each group ($p \leq 0.05$). Differences in expression of *SNAPC4* mRNA, SNAPC4 protein, and snRNAs between specific cell lines were evaluated with either a two-way ANOVA ($\alpha = 0.05$) if Shapiro's test showed normal distribution of data within each cell line ($p > 0.05$) or the Kruskal-Wallis rank-sum test ($\alpha = 0.05$) if Shapiro's test showed nonnormal distribution of data within one or more cell lines ($p \leq 0.05$). If a two-way ANOVA or Kruskal-Wallis rank-sum test showed significance ($p \leq 0.05$), Tukey's test ($\alpha = 0.05$) or Dunn's test ($\alpha = 0.05$) was used for post-hoc analysis, respectively.

Enrichment for genes containing U12-type introns in differentially spliced events from HeLa cells was tested with Fisher's exact test ($\alpha = 0.05$). All statistical analyses were carried out with R.

Results

Bi-allelic *SNAPC4* variants are associated with a progressive neurodevelopmental disorder and lead to SNAPC4 deficiency

We describe ten individuals, from eight families, who presented with an autosomal recessive NDD (Figure 1A and Table 1). Their NDD is characterized by progressive spasticity and developmental and motor regression after the first year of life, initially presenting as progressive spastic paraparesis that led to gait alterations. Further, they presented with generalized motor compromise including oromotor dysfunction; movement disorder associated with dystonic posturing is seen in some affected individuals, while in one individual a possible later lower motor neuron compromise indicated by grouped fiber atrophy on muscle biopsy was seen (individual 5) (Figure S3). Other phenotypes noted included short stature (individuals 1–10), microcephaly (individuals 1, 3–9) (Figure 1B), and increased deep tendon reflexes (individuals 1–9) (Table 1).

Individuals 1, 2, 3, 5, 6, 7, 8, and 9 had MRI studies of the brain (Figure 2), which showed evidence of structural abnormalities on one or more MRIs in individuals 1, 2, 6, 7, 8, and 9 (Figures 2A, 2B, and 2E–2H). Specifically, diffuse cerebral atrophy (individuals 1 and 6) (Figure 2A and 2E) as well as severe volume loss in the cerebellum (individuals 2 and 6–9) (Figures 2B and 2E–2H) and basal ganglia (individuals 2 and 6) (Figures 2B and 2E) were observed. Interestingly, the age of onset was variable across the affected individuals. While individual 1 had no reported findings at 2 years, cortical atrophy was only detected at 14 years (Figure 2A). Additionally, while individuals 2 and 6 have similar findings in their most recent MRIs, their acquisition of those abnormalities differed. Individual 2 initially showed evidence of only basal ganglia atrophy at 18 months (Figure S4B) and of cerebellar atrophy at 5 years (Figure 2B). Conversely, individual 6 initially presented with severe cerebellar atrophy at 18 months (Figure S4E), with note of basal ganglia atrophy only evident at 2 years (Figure 2E). An MRI of individual 5 at 15 years showed moderate microcephaly (Figure 2D). Of note, cerebellar atrophy was observed to be progressive in individuals 2, 6, 7, and 8 (Figure S4). Childhood head circumferences showed that five individuals were microcephalic at birth (individuals 1, 5, 7–9, Figure 1B), and others had normal head circumferences at birth but fell off the normal growth curve later in life (individuals 3, 4, 6, 10, Figure 1B). Images from all available brain MRI studies (Figure S4), clinical histories (supplemental note: case reports), and phenotyping (Table S4) are available in the supplemental information.

Exome or genome sequencing identified bi-allelic variants in *SNAPC4*, the DNA-binding subunit of the SNAP complex (Figures 3A and 3B), in all affected individuals, which were confirmed by Sanger sequencing and segregated with disease. Affected individuals with these variants were identified with the GeneMatcher platform.[27] All the variants were compound heterozygous pairs, except in individuals 7–9, who come from consanguineous families. Of the 13 *SNAPC4* alleles identified, six were missense, one was nonsense, one was frameshift, and five were splice site variants (Table 1). All splice site variants were confirmed to interfere with splicing by cDNA sequencing and ultimately introduce premature stop codons (Figures S5 and S6). All missense variants—c.553A>G (GenBank: NM_003086.3) (p.Lys185Glu [GenBank: NP_003077.2]), c.595G>A (GenBank: NM_003086.3) (p.Asp199Asn [GenBank: NP_003077.2]), c.1157A>G (GenBank: NM_003086.3) (p.Gln386Arg [GenBank: NP_003077.2]), c.1321G>A (GenBank: NM_003086.3) (p.Asp441Asn [GenBank: NP_003077.2]), c.1436T>C (GenBank: NM_003086.3) (p.Ile479Thr [GenBank: NP_003077.2]), c.2900G>T (GenBank: NM_003086.3) (p.Gly967Val [GenBank: NP_003077.2])—were predicted to be pathogenic with *in silico* tools (Table S5). The locations of the three missense variants within the DNA-binding domain are highlighted in the modeled structure of SNAPC4 (Figures 3C and 3D), with Asp 441 interacting directly with the DNA. Asp 441 corresponds to Glu 132 in the homologous template structure (1h88.pdb[34]), while Gln 386 and Ile 479 are unchanged from the corresponding template residues. The variants p.Gln386Arg and p.Asp441Asn alter the charge of the native residue, which could potentially destabilize the interaction of SNAPC4

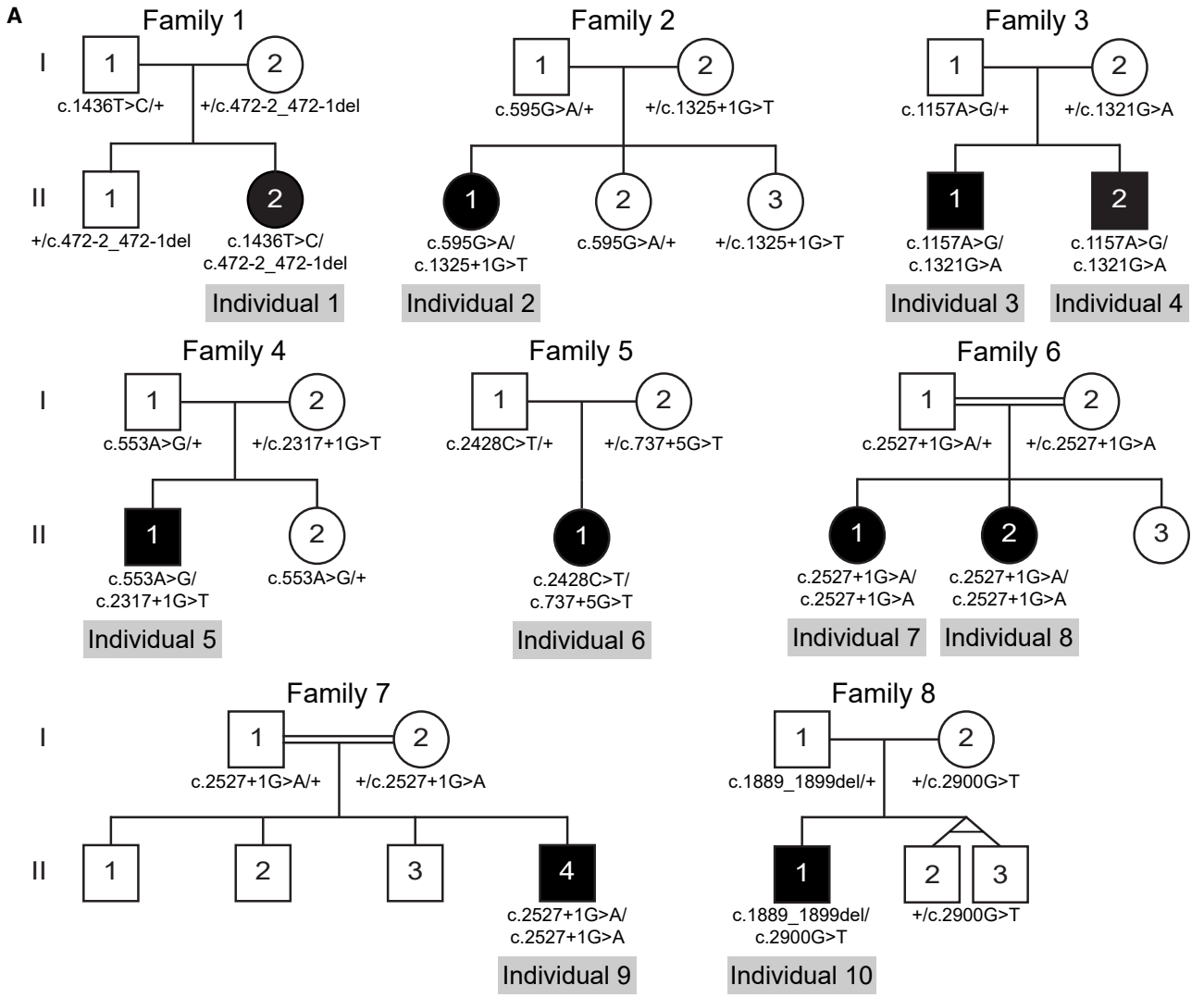


Figure 1. Disease inheritance and head circumference measurements in individuals with bi-allelic *SNAPC4* variants
(A) Pedigrees showing *SNAPC4* genotypes and disease inheritance for families 1–8.
(B) Head circumference measurements for individuals 1–9.

with DNA. The p.Ile479Thr variant changes a hydrophobic side chain that is oriented toward the protein core and packs against Leu 467 and Ile 468 with a hydrophilic side chain less suited to this environment. Consequently, p.Ile479Thr may cause misfolding of this region of SNAPC4, which is proximal to the bound DNA. Therefore,

Table 1. Summary of clinical and variant features of individuals with bi-allelic variants in *SNAPC4*

Variant or phenotype	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8	Individual 9	Individual 10
Genomic changes	9:g.139278469A>G, 9:g.139289331_139289332del	9:g.139288738C>T, 9:139281936:C>A	9:139281941:C>T, 9:139282267:T>C	9:139281941:C>T, 9:139282267:T>C	9:139288780T>C, 9:139276275C>A	9:139275263G>A, 9:139287120C>A	9:g139274288C>T, 9:g139274288C>T	9:g139274288C>T, 9:g139274288C>T	9:g139274288C>T, 9:g139274288C>T	9:g.139277726_139277736del, 9:g.139273379C>A
mRNA changes	c.1436T>C, c.472-1_472-2del	c.595G>A, c.1325+1G>T	c.1321G>A, c.1157A>G	c.1321G>A, c.1157A>G	c.553A>G, c.2317+1G>T	c.2428C>T, c.737+5G>T	c.2527+1G>A, c.2527+1G>A	c.2527+1G>A, c.2527+1G>A	c.2527+1G>A, c.2527+1G>A	c.1889_1899del, c.2900G>T
Amino acid changes	p.Ile479Thr, p.Gly158Valfs*65	p.Asp199Asn, p.Arg434Glyfs*10	p.Asp441Asn, p.Gln386Arg	p.Asp441Asn, p.Gln386Arg	p.Lys185Glu, p.Val752Argfs*155	p.Arg810*, p.Asn245Lysfs*1	p.Ser835Thrfs*86, p.Ser835Thrfs*86	p.Ser835Thrfs*86, p.Ser835Thrfs*86	p.Ser835Thrfs*86, p.Ser835Thrfs*86	p.Val630Glyfs*76, p.Gly967Val
Inheritance pattern	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal	paternal, maternal
Prenatal and perinatal history										
Delivery	NSVD	emergency C/S	NSVD	NSVD	NSVD	NSVD	NSVD	NSVD	NSVD	NSVD
Birth weight	2,490 g (4th centile)	3,345 g (60th centile)	3,288 g (45th centile)	3,773 g (80th centile)	2,138 g (1st centile)	3,190 g (46th centile)	2,980 g (11th centile)	2,790 g (7th centile)	2,640 g (1st centile)	3,400 g (54th centile)
Birth length	45.7 cm (3rd centile)	49.5 cm (58th centile)	unknown	53.3 cm (96th centile)	45.5 cm (1st centile)	51 cm (84th centile)	48 cm (10th centile)	48 cm (16th centile)	47 cm (2nd centile)	unknown
Growth parameters										
Abnormal body weight	+ (95th centile)	+ (<1st centile)	–	–	+ (<1st centile)	+ (1st centile)	+ (<1st centile)	+ (<1st centile)	–	+ (75th–90th centile)
Short stature	+ (12th centile)	+ (<1st centile)	+ (2nd centile)	+ (<1st centile)	+ (<1st centile)	+ (<1st centile)	+ (<1st centile)	+ (<1st centile)	+ (<1st centile)	+ (10th centile)
Microcephaly	+ (−2.17 SD, 1st centile, 9 months)	–	+ (−2.12 SD, 2nd centile, 13 years)	+ (−2.36 SD, 1st centile, 3 years)	+ (−3.4 SD, <1st centile, 8.7 years)	+ (−4.12 SD, <1st centile, 7 years)	+ (−5 SD, <1st centile,12.5 years)	+ (−7 SD, <1st centile, 11 years)	+ (−5.5 SD, <1st centile, 6.5 years)	–
Neurological findings										
Developmental and motor regression	+ (onset: 15 months)	+ (onset: 15 months)	+ (onset: 30 months)	+ (onset: 36 months)	unclear	+ (onset: 6 months)	+ (onset: 16 months)	+ (onset: 18 months)	unclear	+ (onset: 24–36 months)
Spasticity	+ (progressive, ascending)	+ (progressive, ascending)	+ (progressive, ascending)	+ (progressive, ascending)	+	+	+	+	+	+
Spastic quadriparesis	+	+	+	+	–	–	+	+	–	–
Dystonia	+	+	+	+	unknown	+	–	–	–	unknown
Oromotor dysfunction	+ (dysarthria)	+ (drooling, non–verbal)	+ (drooling, dysarthria)	+ (drooling, dysarthria)	+ (drooling, dysarthria)	+ (drooling, dysarthria)	+ (drooling, dysarthria)	+ (drooling, dysarthria)	–	+ (dysarthria)
Gait abnormalities	+	non–ambulatory	+	+	+	non–ambulatory	+	+	+	+
Increased deep tendon reflexes	+	+	+	+	+	+	+	+	+	–
Brain MRI findings	mild cerebral atrophy	severe atrophy of cerebellum, striatum	unknown	none	none	severe atrophy of cerebellum, striatum, cerebrum	severe cerebellar atrophy, mild cerebral atrophy	severe cerebellar atrophy	severe cerebellar atrophy	unknown

All variants are in reference to transcript GenBank: NM_003086.3. Abbreviations: NVSD, normal vaginal spontaneous delivery; C/S, caesarean section.

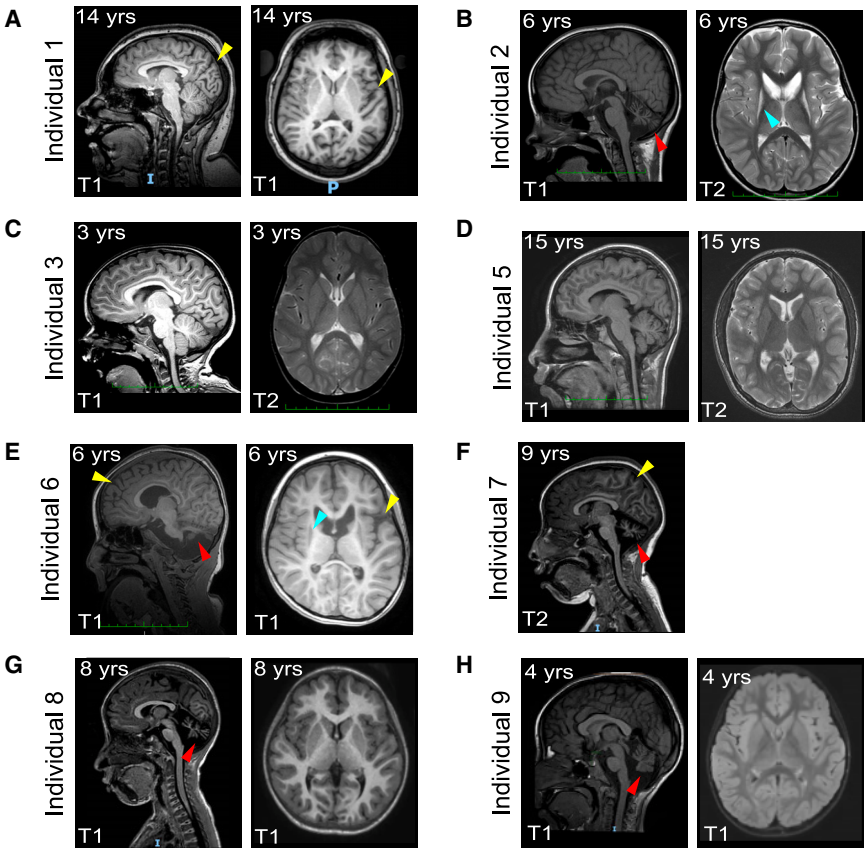


Figure 2. Magnetic resonance imaging (MRI) studies of the brain in individuals with bi-allelic *SNAPC4* variants

To investigate the presence of structural brain abnormalities that may associate with the described neurodevelopmental disorder, MRI studies of the brain from individuals 1 (A), 2 (B), 3 (C), 5 (D), 6 (E), 7 (F), 8 (G), and 9 (H) were reviewed for structural abnormalities. Images demonstrating atrophy to specific structures are indicated with red (cerebellum), yellow (cerebral cortex), or blue (basal ganglia) arrows. Images include T1-weighted (T1), T2-weighted (T2), and FLAIR images.

the structural modeling suggests that each of the three missense mutations has the potential to perturb the affinity of SNAPC4 for DNA.

SNAPC4-deficient cell lines show decreased snRNA expression and dysregulated global alternative splicing

Given that the majority of the *SNAPC4* variants are LOF, we generated *SNAPC4*-deficient HeLa cell lines generated by CRISPR-Cas9 editing (KO12 and KO19) (Figure S1). In these *SNAPC4*-deficient cell lines, relative *SNAPC4* mRNA expression at the exon 2–3 boundary was significantly reduced when compared to controls (Figure 4A, $p \leq 7.09 \times 10^{-5}$), while downstream expression (exon 22–23) was less affected. That difference is consistent with the location of CRISPR guide RNA targets in exons 2 and 3 (Figure S1A). Protein levels reflected this reduction, as immunoblot of whole cell lysates from both cell lines showed significantly less protein compared to a parental control cell line (Figure 4B, $p \leq 1.04 \times 10^{-3}$).

Because SNAPC4 plays a critical role in snRNA transcription, we hypothesized that SNAPC4 deficiency could alter snRNA expression. Via small RNA-seq, we measured snRNA expression in *SNAPC4*-deficient cell lines compared to the parental, wild-type cell line. We observed that the summed expression of all snRNAs was reduced in *SNAPC4*-deficient cell lines, although it did not reach the threshold of significance (Figure 4C, $p = 0.056$). While the expression of the majority of the snRNAs measured were reduced to 45%–50% in both *SNAPC4*-deficient cell lines, only expression of *RNU4-1* and *RNU4ATAC* (Figure 4D) (KO12, $p \leq 0.05$) and *RNU5A-1* (Figure 4D) (KO19, $p \leq 0.05$) were significantly reduced when compared to control.

To determine the effects of *SNAPC4* and snRNA deficiency on alternative splicing throughout the transcriptome, we performed bulk RNA-seq. We applied the Multivariate Analysis of Transcript Splicing (rMATS) software[33] to determine which alternative splicing events occurred significantly more or less frequently in *SNAPC4*-deficient cell lines compared to control. rMATS evaluates five types of alternative splicing events, i.e., skipped exon (SE), retained intron (RI), mutually exclusive exons (MXE), alternative 5′ splice site (A5SS), and alternative 3′ splice site (A3SS). Using rMATS, we detected 7,496 differentially spliced alternative splicing events via a false discovery rate (FDR) of ≤ 0.05 (Table S6) between *SNAPC4*-deficient HeLa cells and control. All samples were hierarchically clustered on the basis of the frequency of differentially spliced alternative splicing events, which showed segregation of all samples (Figure 5A). We also performed PCA on all differentially spliced events to further validate their capacity to distinguish samples. PCA showed that replicates of each sample clustered tightly and distinctly (Figure 5B). To identify functional blocks of genes most likely to be linked to *SNAPC4* deficiency, we used k-means clustering to identify clusters 3 and 4 (Figure 5A), which had average Z scores that positively correlated between KO12 and KO19 but were negatively correlated between control and both KO12 and KO19. We then sought to characterize the biological relevance of the genes containing the alternative splicing events in clusters 3 and 4 by performing Gene Ontology (GO) Biological Process (BP) term enrichment analysis. The top enriched GO BP terms by FDR showed that cluster 3 alternative splicing events were most

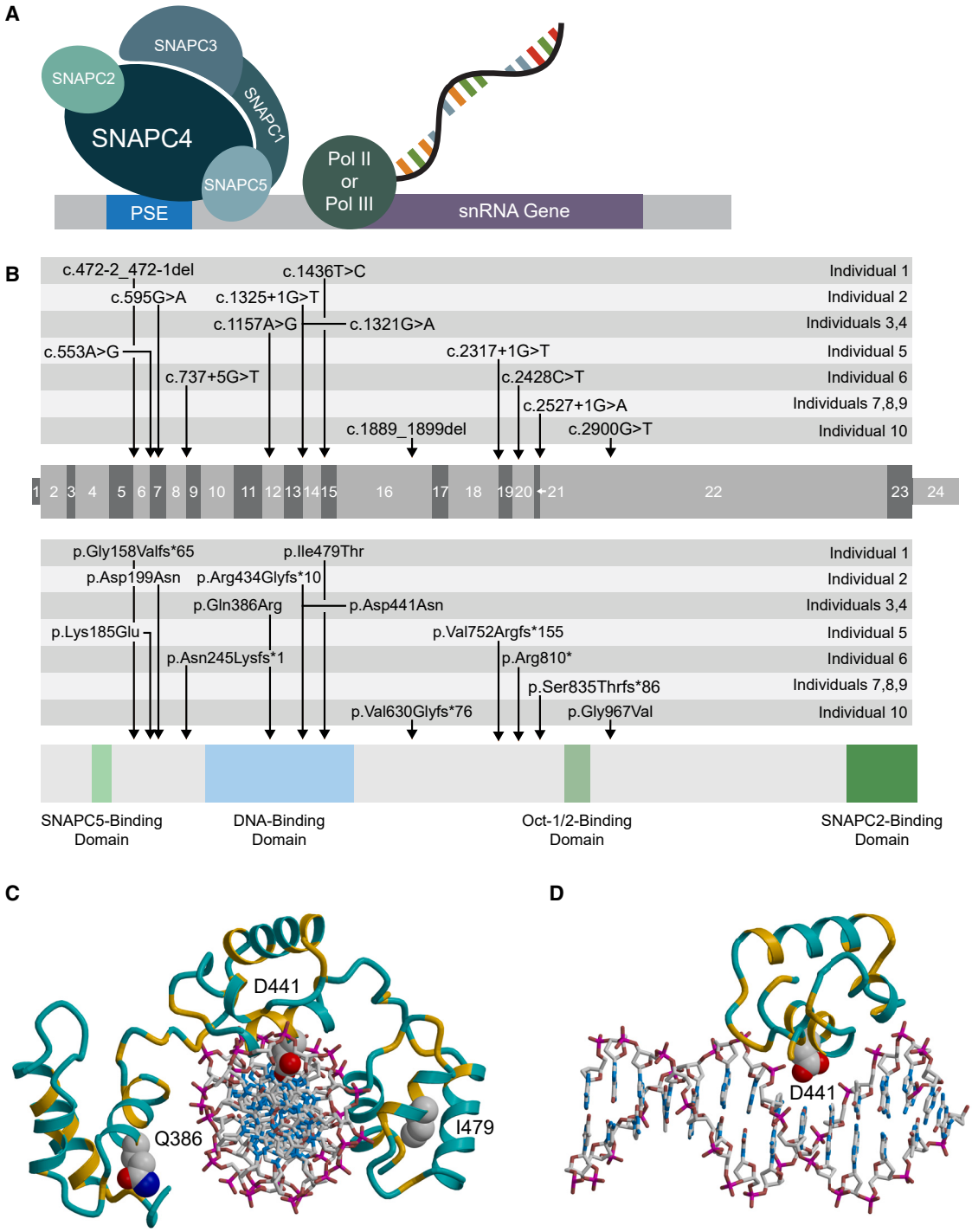


Figure 3. Mapping of *SNAPC4* variants compared to functional elements
(A) Schematic representation of the SNAP complex and snRNA transcription.
(B) Locations of variants from affected individuals relative to functional domains in SNAPC4.
(C) Homology model of human SNAPC4 (amino acid residues 345–500) bound with DNA, built using the crystal structure 1h88.pdb as template. The protein backbone is colored by sequence identity to the template, gold where identical (33%) and turquoise where different. Side chains of residues mutated in affected individuals are shown as space-filling; the DNA is shown as sticks.
(D) Portion of the model containing Asp 441, after a 90-degree rotation.

associated with genes involved in RNA splicing, transport, and processing (Figure 5C) (Table S8). Additional enriched terms in cluster 3 suggest an impact of *SNAPC4* deficiency on microtubule regulation during mitosis (Figure 5C) (Table S8). Cluster 4 had only one enriched GO term, which recapitulated the impact of *SNAPC4* deficiency on microtubule regulation during mitosis (Figure 5C) (Table S8).

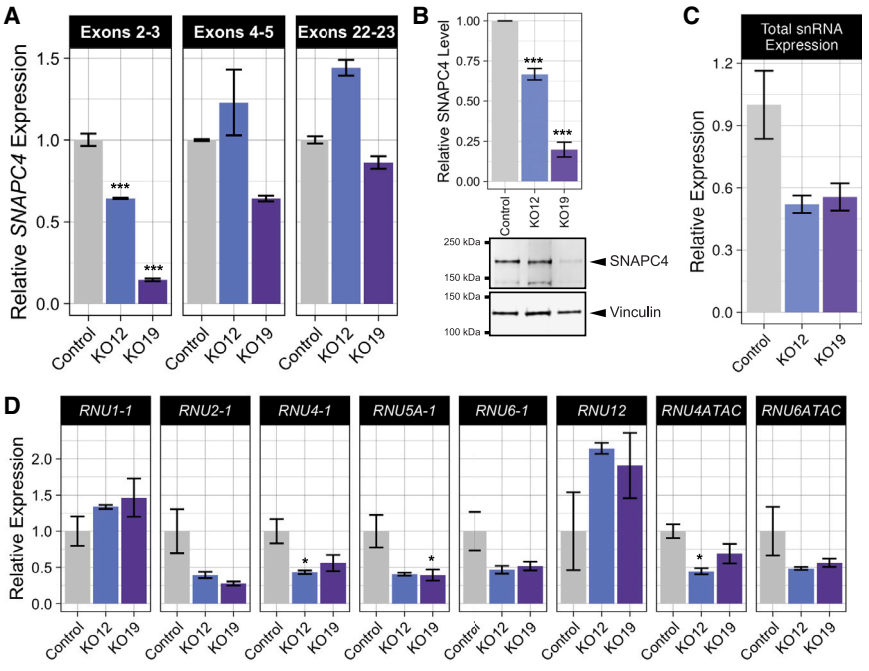


Figure 4. Cellular characterization of *SNAPC4* deficiency in the HeLa *SNAPC4*-deficient cells

(A) To evaluate the effect of CRISPR editing on *SNAPC4* mRNA expression, relative mRNA expression was quantified in *SNAPC4*-deficient (KO12 and KO19) and control HeLa cells at three exon-exon junctions (three technical replicates, error bars represent ±1 SEM); *HPRT1* was used as an endogenous control, and the control cells were used as the reference sample. Expression differences were tested with two-Way ANOVA and Tukey's test for post-hoc analysis.

(B) Immunoblot was used to evaluate the effect of CRISPR editing on SNAPC4 protein levels (n = 3, error bars represent ±1 SEM); vinculin was used as the loading control. Protein differences were tested using two-way ANOVA, and Tukey's test for post-hoc analysis.

(C) Downstream consequences of changes in SNAPC4 levels were evaluated by measuring the summed expression of all major and minor spliceosomal with small RNA-seq; expression was normalized against total counts in that sample (n = 3, error bars represent ±1 SEM).

(D) Expression of specific spliceosomal snRNAs was evaluated with the same normalization procedure (n = 3, error bars represent ±1 SEM). Expression differences across samples were tested with two-way ANOVA and Tukey's test for post-hoc analysis, and expression differences across groups were tested with a two-sample t test. $p < 0.05$, *; $p < 0.01$, **; $p < 0.001$, ***.

SNAPC4-deficient fibroblasts from affected individuals have decreased snRNA expression and dysregulated global alternative splicing

To assess the consequences of SNAPC4 deficiency in primary cell lines from affected individuals, we measured *SNAPC4* expression. Group-wise relative *SNAPC4* mRNA expression was significantly decreased in fibroblasts derived from affected individuals compared to controls (Figure 6A, $p = 2.15 \times 10^{-4}$) but not in the lymphoblastoid cell lines (Figure 6B, $p = 0.781$). SNAPC4 showed a significant decrease in abundance in fibroblasts (Figure 6C, $p = 2.13 \times 10^{-6}$) and lymphoblastoid cells (Figure 6D, $p = 1.76 \times 10^{-4}$) from affected individuals compared to controls. These results suggest that the bi-allelic variants lead to SNAPC4 deficiency and could alter snRNA transcription. Using small RNA-seq, we quantified snRNA transcription and, although the summed expression of all snRNAs did not vary between cells from affected individuals and controls (Figure 6E, $p = 0.664$), expression of specific snRNAs did vary between the two groups. There was an average decrease in expression of *RNU1-1* (45%), *RNU2-1* (42%), *RNU4-1* (24%), and *RNU5A-1* (72%) in affected individual cells compared to controls and those decreases were statistically significant in *RNU2-1* and *RNU5A-1* (Figure 6F, $p \leq 0.0343$). Within the group-wise patterns of snRNA expression, there was variability across fibroblasts from affected individuals. Expression was reduced in the *RNU1-1* (range: 55%–72%), *RNU4-1* (range: 85%–90%), and *RNU5A-1* (range: 86%–95%) snRNAs in individuals 1, 2, and 6 compared to control, although only *RNU5A-1* was statistically significant in individuals 1 and 6 (Figure 6F, $p \leq 0.00942$). Expression of *RNU2-1* was reduced by at least 43% in individuals 3, 4, and 6, although that difference was not statistically significant (Figure 6F, $p \geq 0.76$). Primary cell lines were not available for individuals 7–10 for functional characterization.

To assess the impact of changes in snRNA expression on alternative splicing throughout the transcriptome, we used rMATS to analyze RNA-seq data. rMATS detected 5,523 differentially spliced alternative splicing events (Table S7, FDR ≤ 0.05) between control and affected individual fibroblasts. All samples were hierarchically clustered on the basis of the frequency of differentially spliced events (Figure 7A), which showed segregation of all sample replicates, as well as the affected individual and control fibroblasts. Next, PCA was performed on all differentially spliced events. PCA showed that replicates of each sample clustered well and affected individual and control fibroblasts clusters were distinct (Figure 7B). Additionally, there was stratification between individuals 1, 2, and 6 and individuals 3 and 4 (Figures 7A and 7B). To characterize functional groups of genes linked to *SNAPC4* deficiency, we used k-means clustering to identify clusters 1 and 2 (Figure 7A). By performing GO BP term enrichment analysis in clusters 1 and 2, we observed that RNA splicing, mRNA processing, and microtubule cytoskeleton organization involved in mitosis were enriched (Figure 7C, Table S8); these pathways were also enriched in at least one cluster in the *SNAPC4*-deficient HeLa cell lines (Figure 5C, Table S8). There are also enriched processes that are distinct to either our HeLa or fibroblast models, including "process utilizing autophagic mechanism" in

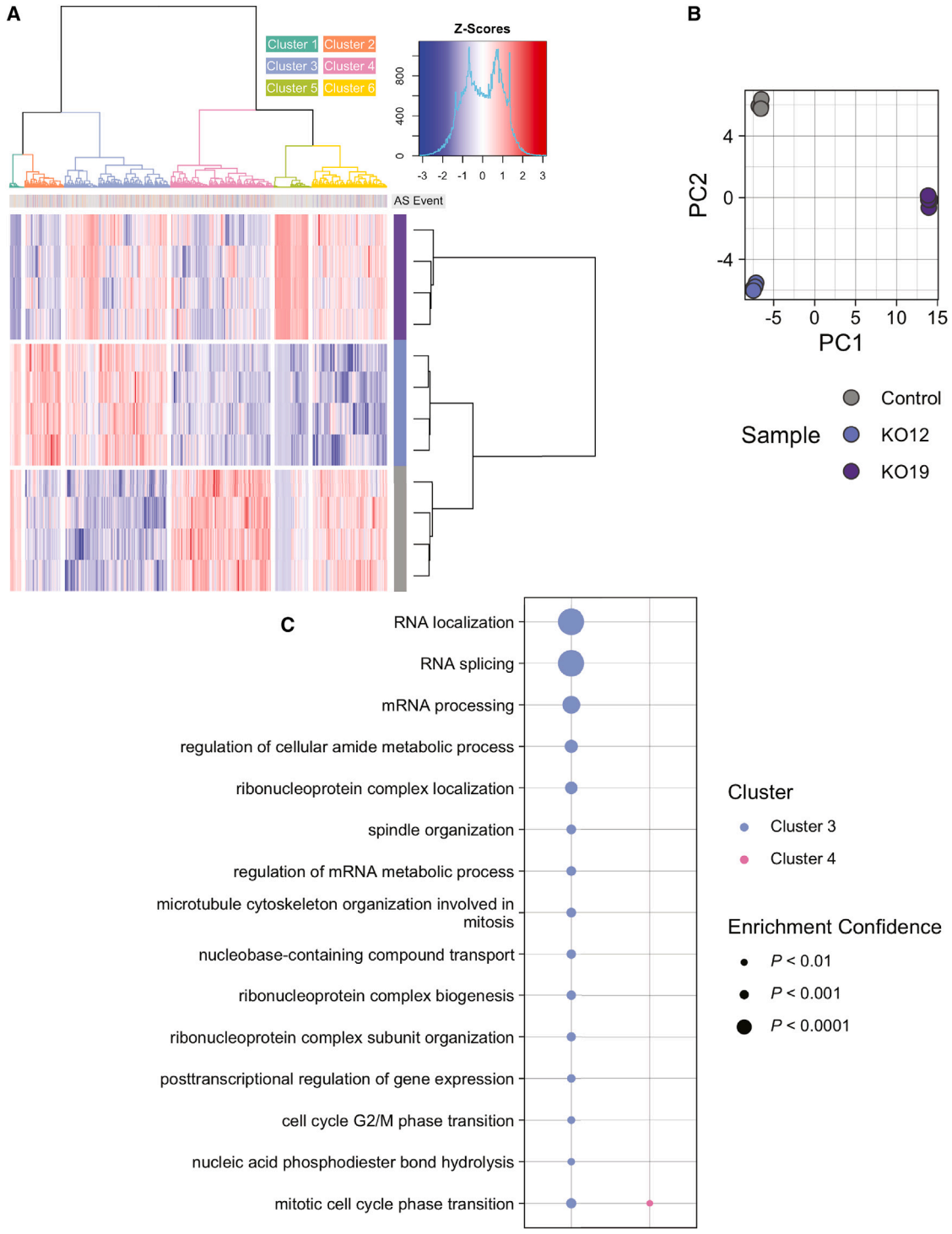


Figure 5. Analysis of global alternative splicing in *SNAPC4*-deficient HeLa cells

In order to assess the downstream functional consequences of SNAPC4 and snRNA deficiency in our HeLa model, we identified significant alternative splicing (differentially spliced) events between *SNAPC4*-deficient and control HeLa cells by using rMATs. The frequencies of all differentially spliced events in each sample were hierarchically clustered (A) to assess sample segregation and to identify blocks of genes with correlated splicing frequency changes. Sample segregation was further assessed by principal-component analysis (PCA) (B) of differentially spliced event frequencies. Splicing events in clusters 3 and 4 were had frequencies that were positively correlated between *SNAPC4*-deficient cells (KO12 and KO19), and negatively correlated with control cells, indicating they are more likely to be linked to SNAPC4 deficiency than splicing events in other clusters. To characterize functional consequences of splicing changes, splicing events in clusters 3 and 4 were then analyzed for Gene Ontology (GO) Biological Process (BP) enrichment and annotated for events belonging to recurrent functional categories (C).

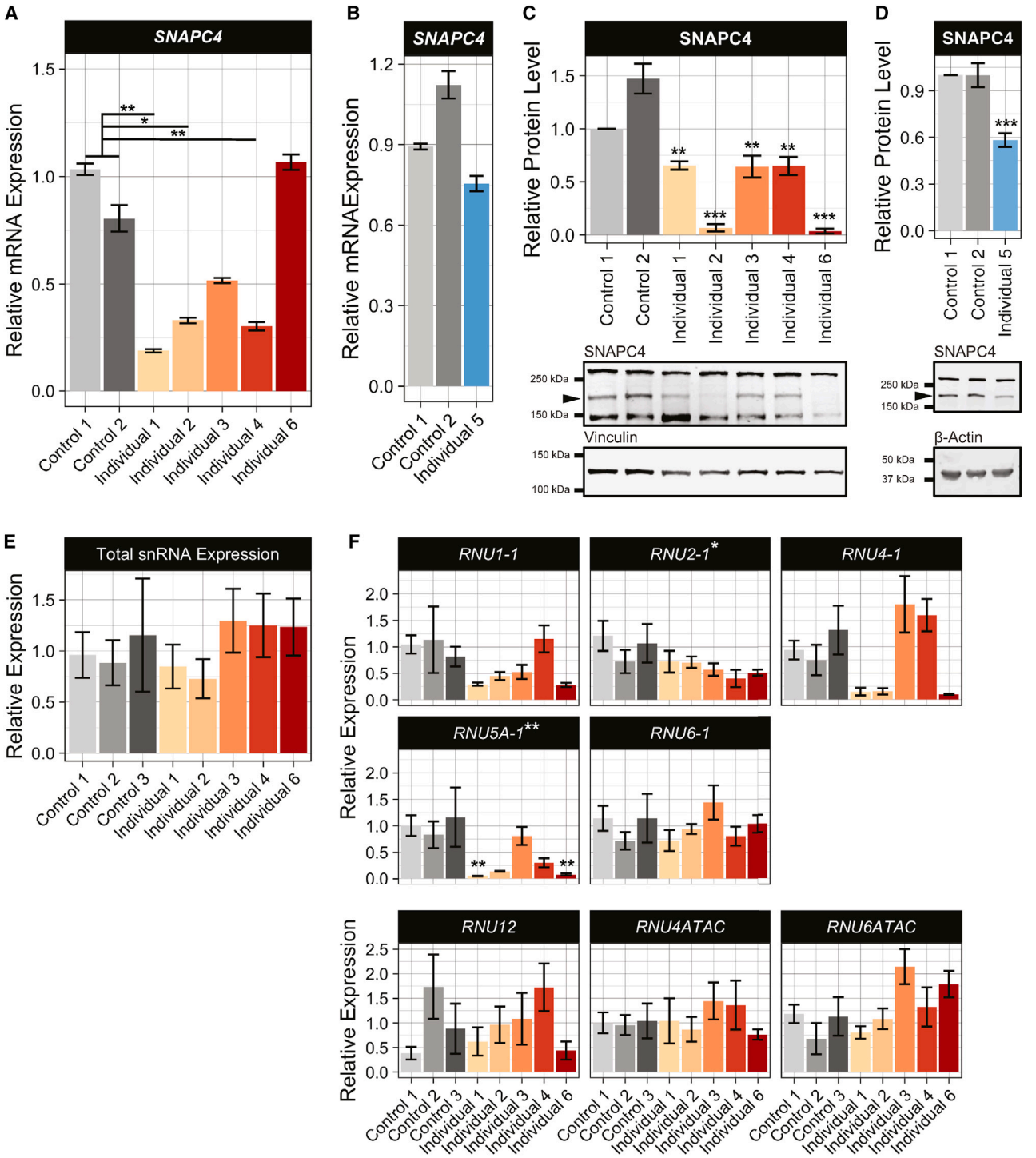


Figure 6. Cellular characterization of fibroblasts derived from individuals with bi-allelic *SNAPC4* variants

To evaluate the impact of bi-allelic *SNAPC4* variants, relative expression of *SNAPC4* mRNA was assessed in affected individual and control fibroblasts (A) and lymphoblastoid cell lines (B) via RT-qPCR (three technical replicates, error bars represent ±1 SEM); *HPRT1*, *TBP*, and *POLR2A* were used as endogenous controls, and control 1 and 2 were used as reference samples. Expression differences were tested with Kruskal-Wallis test and Dunn's test for post-hoc analysis. To evaluate the impact of changes in mRNA expression, SNAPC4 protein levels in fibroblasts (C) and lymphoblastoid cell lines (D) were quantified with immunoblot (n = 3, error bars represent ±1 SEM); vinculin was used as the loading control in fibroblasts, and beta-actin was used in lymphoblastoid cell lines. Expression differences were tested with two-way ANOVA and Tukey's test for post-hoc analysis. Downstream consequences of changes in SNAPC4 abundance were evaluated by measuring the summed expression of all major and minor spliceosomal snRNAs in fibroblasts derived from affected individuals compared to controls via small RNA-seq (E); expression in each sample was normalized against total counts in that sample (n = 4, error bars represent ±1 SEM). (F) The expression of specific major and minor spliceosomal snRNAs were then evaluated with the same normalization procedure (n = 4, error bars represent ±1 SEM). Expression differences across samples were tested using two-way ANOVA and Tukey's Test for post-hoc analysis, and expression differences across groups were tested with a two-sample t test. $p < 0.05$, *; $p < 0.01$, **; $p < 0.001$, ***.

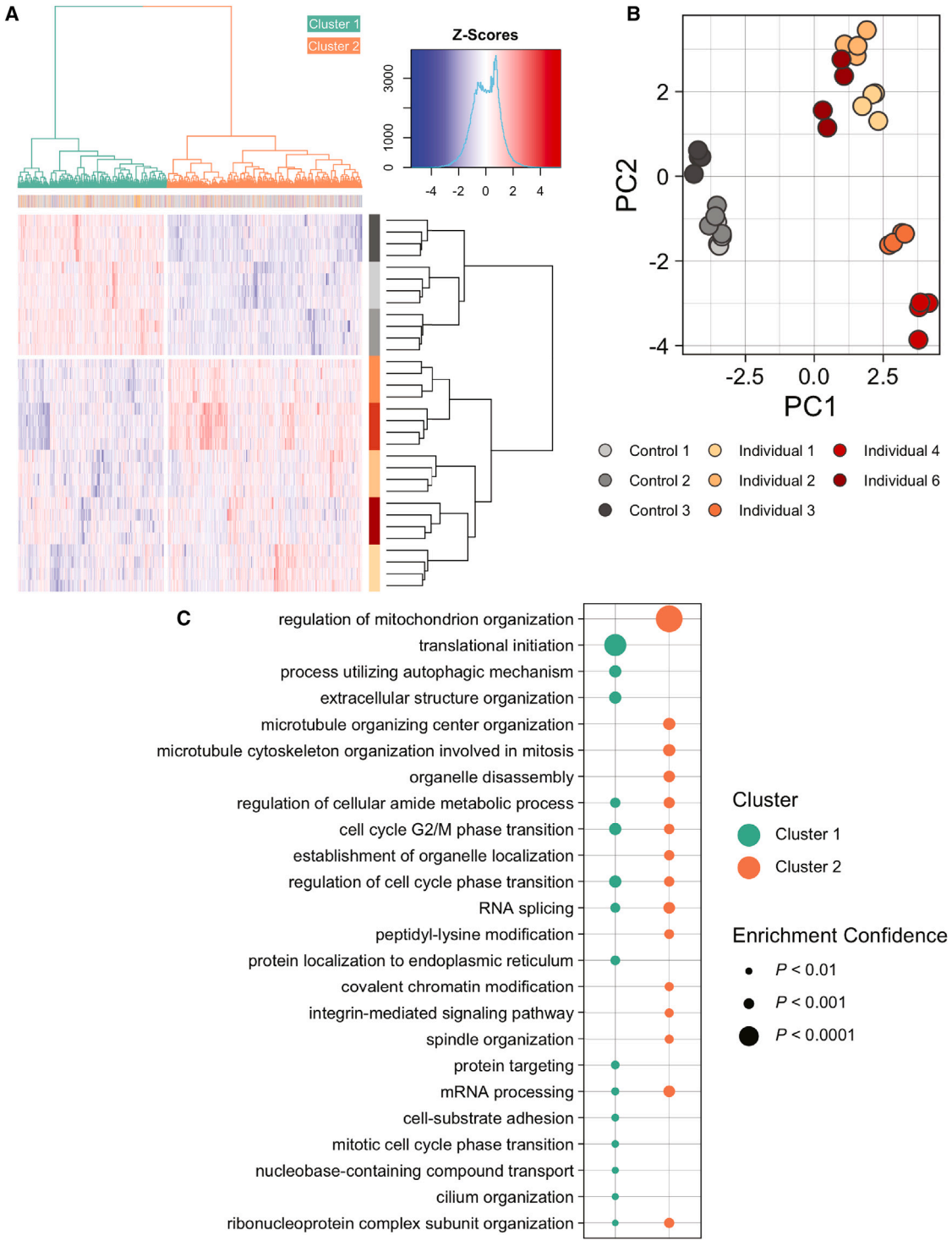


Figure 7. Analysis of global alternative splicing in fibroblasts derived from individuals with bi-allelic *SNAPC4* variants
In order to assess the downstream functional consequences of SNAPC4 and snRNA deficiency in fibroblasts derived from affected individuals, we identified significant alternative splicing (differentially spliced) events compared to unaffected fibroblasts. All differentially spliced events in each sample were hierarchically clustered (A) to assess sample segregation and to identify blocks of genes with correlated splicing frequency changes. Further assessment of sample segregation was done by principal-component analysis (PCA) (B) of differentially spliced event frequencies. To characterize functional consequences of splicing changes, splicing events in clusters 1 and 2 were analyzed for Gene Ontology (GO) Biological Process (BP) enrichment (C).

fibroblasts (Figure 7C, Table S8) and "ribonucleoprotein complex biogenesis" in HeLa cells (Figure 5C, Table S8).

Discussion

In this study, we present ten individuals, from eight unrelated families, with bi-allelic *SNAPC4* variants that are either absent or rare in gnomAD, are predicted to be deleterious by *in silico* tools, segregate with disease (Figure 1A, Table S5), and lead to reduced *SNAPC4* expression in primary fibroblasts. Each individual had at least one null variant, except the siblings (individuals 3 and 4) who had compound heterozygous missense variants. Nonetheless, these missense variants, p.Gln386Arg and p.Asp441Asn, are within the DNA-binding domain and alter the charge of the native residue and could potentially destabilize the interaction of SNAPC4 with DNA. The majority of the affected individuals presented with an NDD characterized by delayed motor development noted around 1 year of age. Delayed developmental milestones were followed by developmental regression with progressive spasticity, initially presenting as progressive spastic paraparesis that led to gait alterations, with further generalized motor compromise including oromotor dysfunction, and movement disorder in some affected individuals. Microcephaly was observed in most affected individuals at birth (individuals 1, 5, and 7–10) (Figure 1B) or later in life (individuals 3, 4, and 6) (Figure 1B).

While the onset and progression of this disorder appears variable, there is an apparent negative correlation between SNAPC4 abundance and severity of MRI findings. Affected individuals with more severe reductions of SNAPC4 abundance (individuals 2 and 6) (Figure 6C) had striking findings of rapidly progressive volume loss in the basal ganglia and rapidly developing severe cerebellar atrophy (Figures 2B and 2E). In addition, individual 6 had progressive cerebral white matter abnormalities and cerebral volume loss by 6 years of age. Individuals 7–9 also showed striking cerebellar atrophy on MRI (Figures 2F–2H). Given that individuals 7–9 are all homozygous for a truncating variant in *SNAPC4*, we hypothesize that their low SNAPC4 protein levels are similar to those of individuals 2 and 6. Further variability in this disorder may be explained by a genotype-phenotype correlation. The presence of two truncating alleles was associated with developmental regression at ages 30–36 months (individuals 3 and 4) (Table 1), and the presence of a single truncating allele was associated with developmental regression at 15 months in both individuals 1 and 2, but 24–36 months in individual 10 (Table 1). Individual 5 also had one truncating allele, but it was not clear his development regressed, although he did show mild spastic paraplegia in early childhood (Table S4). However, it is not clear whether two truncating *SNAPC4* variants are associated with earlier regression when compared to a single truncating variant: individual 6 had developmental regression at 6 months (Table 1), while individuals 7 and 8 regressed at 16–18 months (Table 1). Individual 9 also had two truncating alleles, but it was not clear whether his development regressed (Table 1). Given the more consistent correlation between SNAPC4 levels and MRI findings than truncating alleles and regression onset, it appears that the net effect of *SNAPC4* variants on protein levels are a better predictor of disease severity than genotype alone.

Moreover, in terms of brain MRI findings, affected individuals with reduced, albeit more preserved, SNAPC4 levels (individuals 1, 3, and 5) (Figure 6C) had either normal studies or findings of a nonspecific cerebral volume loss (Figures 2A, 2C, and 2D). In individual 5, lower motor neuron compromise was observed by muscle biopsy and clinical evaluation (Figure S3). These observations suggest that the presence of truncating *SNAPC4* variants, and the resulting decrease in translated SNAPC4 (Figures 6C and 6D), could influence disease progression and severity.

In our *SNAPC4*-deficient cell lines generated by genome editing, there was a decrease in expression of some snRNAs (Figure 4D, $p \leq 0.032$), consistent with the proposed role of SNAPC4 in mammalian snRNA transcription.[23] Comparing snRNA expression in fibroblasts from affected individuals to controls demonstrated statistically significant reduced expression for *RNU2-1* and *RNU5A-1* (Figure 6F, $p \leq 0.0343$), consistent with the selective changes in the snRNA expression we found in our HeLa model and in a *D. rerio* model of *Snapc4* deficiency.[37]

Reduced snRNA expression in these *SNAPC4*-deficient HeLa cell lines was associated with transcriptome-wide changes in alternative splicing for KO12 and KO19 (Figure 5A). Notably, the functional enrichment for mRNA transport and splicing replicate similar findings in an snRNA-knockdown study[38] and are consistent with the known role of spliceosomal proteins in regulating splicing of their own pre-mRNAs.[39] Similarly, changes in snRNA expression in affected individual fibroblasts was associated with a massive disruption of alternative splicing throughout the transcriptome (Figures 7A and 7B). The global alternative splicing changes seen in cells from affected individuals demonstrate that *SNAPC4* variants have important cellular consequences and can be considered pathogenic. Intriguingly, the functional enrichment of clusters of differentially spliced genes in cells from affected individuals showed some similar results to that in our HeLa model. Specifically, there was enrichment for microtubule regulation during mitosis, mRNA transport, and mRNA splicing (Figures 5C and 7C). The overlap in functional enrichment for differentially spliced genes in two independent models of *SNAPC4* deficiency suggests that, despite the high cell-type specificity of splicing programs, there may be some universal consequences of *SNAPC4* deficiency.

Nonetheless, there are distinct pathway enrichments among differentially spliced genes in our two models. Interestingly, these distinct processes are linked to Mendelian neurological disorders.[40,41] In *SNAPC4*-deficient HeLa

cells there was functional enrichment for genes involved in noncoding RNA processing (Figure 5C), including *EXOSC3* (MIM: 606489), which is required for maturation of transcribed snRNAs. Intriguingly, deleterious variants in *EXOSC3* are linked to a neurological disorder characterized in part by cerebellar hypoplasia,[42] a phenotype observed in the affected individuals in this study. In fibroblasts derived from affected individuals, there was enrichment for a variety of pathways, including several related to autophagy (Figure 7C), the dysfunction of which has been linked to neurodegenerative disorders including Parkinson's and Alzheimer's.[43] These distinct processes could represent cell-type-specific consequences of SNAPC4 deficiency. We hypothesize that those differences are the result of discrete transcriptome profiles and therefore distinct populations of molecules regulating splicing. However, evaluating that hypothesis will require additional experiments assessing splicing dysregulation across a wider range of *SNAPC4*-deficient cell types, including neuronal cells.

The critical and ancient role of SNAPc in cell biology, and specifically the SNAPC4 subunit, is validated by mutant models of *SNAPC4* orthologs in animal models. In the *D. rerio* model, a C-terminal truncating variant in *Snapc4* was associated with reduced expression of *U4* and *U5*.[37] This *D. rerio* model displayed embryonic lethality when translation of *Snapc4* was blocked and was reported to have a gut malformation phenotype; neurological phenotypes were not evaluated.[37] Further, completely blocking *Snapc4* translation was embryonic lethal in *D. rerio*, affirming the requirement of Snapc4 for basal cell functions and development. In *A. thaliana*, homozygous deleterious variants in *AtSNAPc4* are most likely lethal,[44] indicating the conserved importance of SNAPC4 function across distant eukaryotic phyla. Unfortunately, no mammalian model deficient in SNAPC4 has been described. In *M. musculus*, phenotyping data are only available for heterozygous null mice.

The mechanism linking *SNAPC4* deficiency to neurological dysfunction remains to be explored. However, there is precedent for a defect in snRNP biogenesis causing a neurological disorder. Spinal muscular atrophy (SMA [MIM: 253300]) is a neurodegenerative disease that results from bi-allelic pathogenic variants in *SMN1* (MIM: 600354), which encodes a protein required for snRNP assembly. The precise etiology of neuronal dysfunction in SMA is still emerging, although several important factors have been identified in murine SMA models. There is a greater reduction of the U11 and U12 snRNPs in brain and spinal cord of a murine SMA model,[45] which may influence the neurological phenotype in SMA as a result of the enrichment for voltage-gated ion channels in genes containing U12-type introns.[45] A similar reduction in expression of minor spliceosomal snRNAs does not appear in our fibroblasts from affected individuals (Figure 6F, $p \geq 0.082$). The expression of the minor spliceosomal snRNA *U4ATAC* is reduced in our HeLa model (Figure 4H, $p = 0.0172$), but that reduction is equivalent to that of major spliceosome snRNAs. Although rMATS analysis in our HeLa model showed 349 significant alternative splicing events in genes with U12-type introns, that does not represent a significant enrichment over events in genes without U12-type introns (Figure 5A, $p = 0.58$). *SMN1* has also been implicated in cellular processes independent of its snRNP assembly function that may contribute to the neuronal pathology of SMA, including DNA repair, translation, macromolecule trafficking, and cytoskeleton maintenance.[46] These differences in affected pathways in SMA compared to those identified in the affected individuals with bi-allelic *SNAPC4* variants may explain, in part, the differences in SMA and SNAPC4-related NDD (Figure 7C, Table S8); while the former is characterized by lower motor neuron degeneration, the latter is currently associated with upper motor neuron dysfunction.

Intriguingly, LOF variants in another snRNP assembly factor, *GEMIN5* (MIM: 607005), have been shown to cause a neurodevelopmental disorder characterized by hypotonia and cerebellar ataxia (MIM: 619333), which bears a closer resemblance to the clinical presentation of the affected individuals we report here.[47] This shows that the pathology of variants disrupting snRNP biogenesis is highly gene dependent. The clinical presentation of individuals with LOF *GEMIN5* variants provides corroborating evidence that disruption of snRNP biogenesis can result in a neurodevelopmental disorder that primarily affects upper motor neurons, as we have hypothesized with *SNAPC4*. In addition to variants affecting snRNP protein subunits, variants in a U2 spliceosomal snRNA gene have been associated with neurodegeneration,[48] demonstrating the potential of variants affecting either protein or RNA components of snRNPs to cause neurological disease. Beyond the disruption of snRNP biogenesis, a neurological disorder has also been associated with LOF variants in the regulatory splicing factor *NSRP1*,[16] further demonstrating the connection between spliceosomal disruption and neurological disorders.

There are several limitations to our study. First, while fibroblasts from affected individuals and a *SNAPC4*-deficient HeLa model showed notable *SNAPC4* deficiency and downstream alterations in snRNA expression (Figures 4C, 4D, 6C, 6D, and 6F) and dysregulation of alternative splicing (Figures 5 and 7), we have not assessed cells from the nervous system that seem to be the most affected in this disorder for similar changes. Nevertheless, our results utilizing available cell lines point to an intriguing consequence of *SNAPC4* deficiency: defects in splicing of genes encoding spliceosomal machinery and nucleoporins. Previous work has shown that disrupted nucleoporin activity[49] and dysregulation of spliceosome subunits[50,51] are linked to neurological disease, indicating that misplicing of those genes in the context of *SNAPC4* deficiency may contribute to the neurological phenotype in these affected individuals. Both a *SNAPC4*-deficient HeLa model and skin fibroblasts from affected individuals showed that among differentially spliced genes there was

an enrichment of genes related to the spliceosome and mRNA transport, including nucleoporin genes (Figures 5C and 7C, Table S8).

While alternative splicing is utilized in all human tissues, it is particularly important in the nervous system, which utilizes the highest number of unique isoforms of all tissues. Specifically, alternative splicing plays key roles in regulating gene expression during differentiation of embryonic cells into neurons[10,11,52] and allows for rapid splicing and translation of large transcripts in response to stimuli.[6] The potential of altered splicing of genes encoding spliceosomal machinery to influence neurological disease is not novel, as altered splicing profiles of the U1 snRNP protein subunit gene *SNRNP70* (MIM: 180740) have been linked to a specific ALS subtype.[50] Nucleoporin genes have also been associated with numerous neurological diseases, including Achalasia-Addisonianism-Alacrima syndrome (triple A syndrome [MIM: 231550]) and ALS, and have been shown to be important for nervous system development.[49] We propose that a disruption of alternative splicing associated with *SNAPC4* deficiency impacts neuron differentiation and function; future studies will be required to evaluate this hypothesis in relevant cells, tissues, or animal models.

Second, there are phenotypic characteristics unique to some patients. For example, high arched palate was observed in individuals 1 and 2, variable dysmorphic features were noted in individuals 5, 6, and 9, and other features in other affected individuals (Table S4). We do not know whether these are part of the phenotypic spectrum of *SNAPC4* deficiency or due to other factors, including genotypic background. Identification of additional individuals will most likely help clarify the phenotype associated with *SNAPC4* deficiency. Lastly, while we highlight the function of SNAPC4 in influencing the spliceosome, there remains the possibility there are uncharacterized functions of SNAPC4 beyond its role in snRNA transcription. GO enrichment analysis of differentially expressed genes from fibroblasts of affected individuals and *SNAPC4*-deficient HeLa cells showed enrichment for GO BP terms related to cell cycle phase transition and checkpoints, DNA replication, vesicle localization, and angiogenesis (Table S8), which may contribute to dysfunction within the nervous system.

Conclusion

We present ten individuals with bi-allelic *SNAPC4* variants and an NDD primarily characterized by progressive spasticity. Characterization of *SNAPC4*-deficient HeLa cell lines showed that *SNAPC4* deficiency impairs snRNA transcription and has broad consequences on alternative splicing throughout the transcriptome. Fibroblasts from affected individuals exhibited *SNAPC4* deficiency, with similarly broad consequences as in our HeLa models. Furthermore, there is significant overlap in the functional enrichment of differentially spliced genes in our HeLa model and fibroblasts derived from affected individuals, indicating that there are some universal consequences of *SNAPC4* deficiency, despite the highly tissue-specific nature of splicing programs. These data present compelling evidence for deleterious *SNAPC4* variants as the cause of an NDD with neuroregression and progressive spasticity and implicate a member of the SNAP complex in Mendelian disease.

Consortia

Collaborators of The Undiagnosed Diseases Network (UDN) include Maria T. Acosta, Margaret Adam, David R. Adams, Justin Alvey, Laura Amendola, Ashley Andrews, Euan A. Ashley, Mahshid S. Azamian, Carlos A. Bacino, Guney Bademci, Ashok Balasubramanyam, Dustin Baldridge, Jim Bale, Michael Bamshad, Deborah Barbouth, Pinar Bayrak-Toydemir, Anita Beck, Alan H. Beggs, Edward Behrens, Gill Bejerano, Hugo J. Bellen, Jimmy Bennett, Beverly Berg-Rood, Jonathan A. Bernstein, Gerard T. Berry, Anna Bican, Stephanie Bivona, Elizabeth Blue, John Bohnsack, Devon Bonner, Lorenzo Botto, Brenna Boyd, Lauren C. Briere, Elly Brokamp, Gabrielle Brown, Elizabeth A. Burke, Lindsay C. Burrage, Manish J. Butte, Peter Byers, William E. Byrd, John Carey, Olveen Carrasquillo, Thomas Cassini, Ta Chen Peter Chang, Sirisak Chanprasert, Hsiao-Tuan Chao, Gary D. Clark, Terra R. Coakley, Laurel A. Cobban, Joy D. Cogan, Matthew Coggins, F. Sessions Cole, Heather A. Colley, Cynthia M. Cooper, Heidi Cope, William J. Craigen, Andrew B. Crouse, Michael Cunningham, Precilla D'Souza, Hongzheng Dai, Surendra Dasari, Joie Davis, Jyoti G. Dayal, Esteban C. Dell'Angelica, Katrina Dipple, Daniel Doherty, Naghmeh Dorrani, Argenia L. Doss, Emilie D. Douine, Laura Duncan, Dawn Earl, David J. Eckstein, Lisa T. Emrick, Christine M. Eng, Cecilia Esteves, Marni Falk, Liliana Fernandez, Elizabeth L. Fieg, Paul G. Fisher, Brent L. Fogel, Irman Forghani, William A. Gahl, Ian Glass, Bernadette Gochuico, Rena A. Godfrey, Katie Golden-Grant, Madison P. Goldrich, Alana Grajewski, Irma Gutierrez, Don Hadley, Sihoun Hahn, Rizwan Hamid, Kelly Hassey, Nichole Hayes, Frances High, Anne Hing, Fuki M. Hisama, Ingrid A. Holm, Jason Hom, Martha Horike-Pyne, Yong Huang, Alden Huang, Wendy Introne, Rosario Isasi, Kosuke Izumi, Fariha Jamal, Gail P. Jarvik, Jeffrey Jarvik, Suman Jayadev, Orpa Jean-Marie, Vaidehi Jobanputra, Lefkothea Karaviti, Jennifer Kennedy, Shamika Ketkar, Dana Kiley, Gonench Kilich, Shilpa N. Kobren, Isaac S. Kohane, Jennefer N. Kohler, Susan Korrick, Mary Kozuira, Deborah Krakow, Donna M. Krasnewich, Elijah Kravets, Seema R. Lalani, Byron Lam, Christina Lam, Grace L. LaMoure, Brendan C. Lanpher, Ian R. Lanza, Kimberly LeBlanc, Brendan H. Lee, Roy Levitt, Richard A. Lewis, Pengfei Liu, Xue Zhong Liu, Nicola Longo, Sandra K. Loo, Joseph Loscalzo, Richard L. Maas, Ellen F. Macnamara, Calum A. MacRae, Valerie V. Maduro, Rachel Mahoney, Bryan C. Mak, May Christine V. Malicdan, Laura A. Mamounas, Teri A. Manolio, Rong Mao, Kenneth Maravilla, Ronit Marom, Julian A. Martínez-Agosto, Gabor Marth, Beth A. Martin, Martin G. Martin, Shruti Marwaha, Jacob McCauley, Allyn

McConkie-Rosell, Alexa T. McCray, Elisabeth McGee, Heather Mefford, J. Lawrence Merritt, Matthew Might, Ghayda Mirzaa, Eva Morava, Paolo M. Moretti, Mariko Nakano-Okuno, Stanley F. Nelson, John H. Newman, Sarah K. Nicholas, Deborah Nickerson, Shirley Nieves-Rodriguez, Donna Novacic, Devin Oglesbee, James P. Orengo, Laura Pace, Stephen Pak, J. Carl Pallais, Christina G.S. Palmer, Jeanette C. Papp, Neil H. Parker, John A. Phillips III, Jennifer E. Posey, Lorraine Potocki, Barbara N. Pusey Swerdzewski, Aaron Quinlan, Archana N. Raja, Deepak A. Rao, Anna Raper, Wendy Raskind, Genecee Renteria, Chloe M. Reuter, Lynette Rives, Amy K. Robertson, Lance H. Rodan, Jill A. Rosenfeld, Natalie Rosenwasser, Francis Rossignol, Maura Ruzhnikov, Ralph Sacco, Jacinda B. Sampson, Mario Saporta, Judy Schaechter, Timothy Schedl, Kelly Schoch, Daryl A. Scott, C. Ron Scott, Vandana Shashi, Jimann Shin, Edwin K. Silverman, Janet S. Sinsheimer, Kathy Sisco, Edward C. Smith, Kevin S. Smith, Emily Solem, Lilianna Solnica-Krezel, Ben Solomon, Rebecca C. Spillmann, Joan M. Stoler, Kathleen Sullivan, Jennifer A. Sullivan, Angela Sun, Shirley Sutton, David A. Sweetser, Virginia Sybert, Holly K. Tabor, Queenie K.-G. Tan, Amelia L. M. Tan, Mustafa Tekin, Fred Telischi, Willa Thorson, Cynthia J. Tifft, Camilo Toro, Alyssa A. Tran, Brianna M Tucker, Tiina K. Urv, Adeline Vanderver, Matt Velinder, Dave Viskochil, Tiphanie P. Vogel, Colleen E. Wahl, Melissa Walker, Stephanie Wallace, Nicole M. Walley, Jennifer Wambach, Jijun Wan, Lee-kai Wang, Michael F. Wangler, Patricia A. Ward, Daniel Wegner, Monika Weisz Hubshman, Mark Wener, Tara Wenger, Katherine Wesseling Perry, Monte Westerfield, Matthew T. Wheeler, Jordan Whitlock, Lynne A. Wolfe, Kim Worley, Changrui Xiao, Shinya Yamamoto, John Yang, Diane B. Zastrow, Zhe Zhang, Chunli Zhao, and Stephan Zuchner.

Data and code availability

The data that support the findings of this study are openly available in Gene Expression Omnibus (GEO) at https://www.ncbi.nlm.nih.gov/geo, reference number GEO: GSE211811.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.03.001.

Acknowledgments

The authors thank the patients and their families for participating in this work, as well as Dr. C. Christopher Lau (National Human Genome Research Institute, National Institutes, of Health) for his assistance interpreting genomic data for individual 1. We are grateful to Prof. C. van Karnebeek, Dr. M. Langeveld, Prof. H. Waterham, Dr. Saskia van der Crabben (Amsterdam UMC), and the ZOEMBA team in the Netherlands for their research contributions. This work was supported in part by the Common Fund, Office of the Director, National Institutes of Health; the Intramural Research Program of National Human Genome Research Institute of the National Institutes of Health; the Office of Science Management and Operations (OSMO) of the National Institute of Allergy and Infectious Diseases, NIH; the Japan Agency for Medical Research and Development (AMED) (grant numbers JP22ek0109486, JP22ek0109549, JP22ek0109348); JSPS KAKENHI (grant number JP21k15907); Takeda Science Foundation; United for Metabolic Diseases, The Netherlands; and Stichting Metakids, The Netherlands. J.R.L. was supported by a US National Human Genome Research Institute (NHGRI) and National Heart, Lung, and Blood Institute grant to Baylor-Hopkins Center for Mendelian Genomics (UM1 HG006542), US NHGRI grant to the Baylor College of Medicine Genomics of Research Elucidates the Genetics of Rare disease (GREGoR) Research Center (U01 HG011758), US National Institute of Neurological Disorders and Stroke (R35NS105078), the Spastic Paraplegia Foundation, and the Muscular Dystrophy Association (MDA 512848). D.G.C. was supported by an NIH Brain Disorders and Development Training Grant (T32 NS043124-19) and Muscular Dystrophy Association grant 873841.

Declaration of interests

J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Genetics Center, and is a co-inventor on multiple US and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, genomic disorders, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic and genomic testing conducted at Baylor Genetics (BG); J.R.L. serves on the Scientific Advisory Board (SAB) of BG.

Received: November 2, 2022
Accepted: February 28, 2023
Published: March 24, 2023

References

1. Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593. https://doi.org/10.1126/science.1230612.
2. Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463. https://doi.org/10.1038/nature08909.
3. Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. *18*, 437–451. https://doi.org/10.1038/nrm.2017.27.
4. Wang, E.T., Ward, A.J., Cherone, J.M., Giudice, J., Wang, T.T., Treacy, D.J., Lambert, N.J., Freese, P., Saxena, T., Cooper, T.A., and Burge, C.B. (2015). Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. Genome Res. *25*, 858–871. https://doi.org/10.1101/gr.184390.114.
5. Carreira-Rosario, A., Bhargava, V., Hillebrand, J., Kollipara, R.K., Ramaswami, M., and Buszczak, M. (2016). Repression of pumilio protein expression by Rbfox1 promotes germ cell differentiation. Dev. Cell *36*, 562–571. https://doi.org/10.1016/j.devcel.2016.02.010.
6. Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted intron retention and excision for rapid gene regulation in

response to neuronal activity. Neuron *92*, 1266–1278. https://doi.org/10.1016/j.neuron.2016.11.032.

7. Li, S., Guo, W., Dewey, C.N., and Greaser, M.L. (2013). Rbm20 regulates titin alternative splicing as a splicing repressor. Nucleic Acids Res. *41*, 2659–2672. https://doi.org/10.1093/nar/gks1362.
8. Elizalde, M., Urtasun, R., Azkona, M., Latasa, M.U., Goñi, S., García-Irigoyen, O., Uriarte, I., Segura, V., Collantes, M., Di Scala, M., et al. (2014). Splicing regulator *SLU7* is essential for maintaining liver homeostasis. J. Clin. Invest. *124*, 2909–2920. https://doi.org/10.1172/JCI74382.
9. Lupski, J.R. (2022). Biology in balance: human diploid genome integrity, gene dosage, and genomic medicine. Trends Genet. *38*, 554–571. https://doi.org/10.1016/j.tig.2022.03.001.
10. Yap, K., Lim, Z.Q., Khandelia, P., Friedman, B., and Makeyev, E.V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev. *26*, 1209–1223. https://doi.org/10.1101/gad.188037.112.
11. Yeom, K.-H., Pan, Z., Lin, C.-H., Lim, H.Y., Xiao, W., Xing, Y., and Black, D.L. (2021). Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. Genome Res. *31*, 1106–1119. https://doi.org/10.1101/gr.273904.120.
12. Dillman, A.A., Hauser, D.N., Gibbs, J.R., Nalls, M.A., McCoy, M.K., Rudenko, I.N., Galter, D., and Cookson, M.R. (2013). mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nat. Neurosci. *16*, 499–506. https://doi.org/10.1038/nn.3332.
13. Sun, S., Ling, S.-C., Qiu, J., Albuquerque, C.P., Zhou, Y., Tokunaga, S., Li, H., Qiu, H., Bui, A., Yeo, G.W., et al. (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nat. Commun. *6*, 6171. https://doi.org/10.1038/ncomms7171.
14. Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. Cell Rep. *6*, 1139–1152. https://doi.org/10.1016/j.celrep.2014.02.005.
15. Fernández-Nogales, M., Santos-Galindo, M., Hernández, I.H., Cabrera, J.R., and Lucas, J.J. (2016). Faulty splicing and cytoskeleton abnormalities in Huntington's disease. Brain Pathol. *26*, 772–778. https://doi.org/10.1111/bpa.12430.
16. Calame, D.G., Bakhtiari, S., Logan, R., Coban-Akdemir, Z., Du, H., Mitani, T., Fatih, J.M., Hunter, J.V., Herman, I., Pehlivan, D., et al. (2021). Biallelic loss-of-function variants in the splicing regulator *NSRP1* cause a severe neurodevelopmental disorder with spastic cerebral palsy and epilepsy. Genet. Med. *23*, 2455–2460. https://doi.org/10.1038/s41436-021-01291-x.
17. Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121. https://doi.org/10.1038/nrm3742.
18. Bohnsack, M.T., and Sloan, K.E. (2018). Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. Biol. Chem. *399*, 1265–1276. https://doi.org/10.1515/hsz-2018-0205.
19. Xu, C., Ishikawa, H., Izumikawa, K., Li, L., He, H., Nobe, Y., Yamauchi, Y., Shahjee, H.M., Wu, X.-H., Yu, Y.T., et al. (2016). Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. Genes Dev. *30*, 2376–2390. https://doi.org/10.1101/gad.288340.116.
20. Becker, D., Hirsch, A.G., Bender, L., Lingner, T., Salinas, G., and Krebber, H. (2019). Nuclear Pre-snRNA export is an essential quality assurance mechanism for functional spliceosomes. Cell Rep. *27*, 3199–3214.e3. https://doi.org/10.1016/j.celrep.2019.05.031.
21. Dergai, O., Cousin, P., Gouge, J., Satia, K., Praz, V., Kuhlman, T., Lhôte, P., Vannini, A., and Hernandez, N. (2018). Mechanism of selective recruitment of RNA polymerases II and III to snRNA gene promoters. Genes Dev. *32*, 711–722. https://doi.org/10.1101/gad.314245.118.
22. Ford, E., Strubin, M., and Hernandez, N. (1998). The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. Genes Dev. *12*, 3528–3540. https://doi.org/10.1101/gad.12.22.3528.
23. Egloff, S., O'Reilly, D., and Murphy, S. (2008). Expression of human snRNA genes from beginning to end. Biochem. Soc. Trans. *36*, 590–594. https://doi.org/10.1042/BST0360590.
24. Gahl, W.A., Markello, T.C., Toro, C., Fajardo, K.F., Sincan, M., Gill, F., Carlson-Donohoe, H., Gropman, A., Pierson, T.M., Golas, G., et al. (2012). The National Institutes of Health undiagnosed diseases program: insights into rare diseases. Genet. Med. *14*, 51–59. https://doi.org/10.1038/gim.0b013e318232a005.
25. Gahl, W.A., Mulvihill, J.J., Toro, C., Markello, T.C., Wise, A.L., Ramoni, R.B., Adams, D.R., Tifft, C.J.; and UDN (2016). The NIH undiagnosed diseases program and network: applications to modern medicine. Mol. Genet. Metab. *117*, 393–400. https://doi.org/10.1016/j.ymgme.2016.01.007.
26. Gahl, W.A., and Tifft, C.J. (2011). The NIH undiagnosed diseases program: lessons learned. JAMA *305*, 1904–1905. https://doi.org/10.1001/jama.2011.613.
27. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930. https://doi.org/10.1002/humu.22844.
28. Davids, M., Kane, M.S., Wolfe, L.A., Toro, C., Tifft, C.J., Adams, D., Li, X., Raihan, M.A., He, M., Gahl, W.A., et al. (2019). Glycomics in rare diseases: from diagnosis tomechanism. Transl. Res. *206*, 5–17. https://doi.org/10.1016/j.trsl.2018.10.005.
29. Sharma, P., Reichert, M., Lu, Y., Markello, T.C., Adams, D.R., Steinbach, P.J., Fuqua, B.K., Parisi, X., Kaler, S.G., Vulpe, C.D., et al. (2019). Biallelic *HEPHL1* variants impair ferroxidase activity and cause an abnormal hair phenotype. PLoS Genet. *15*, e1008143. https://doi.org/10.1371/journal.pgen.1008143.
30. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.
31. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
32. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.
33. Shen, S., Park, J.W., Lu, Z.x., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate

RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601. https://doi.org/10.1073/pnas.1419161111.
34. Tahirov, T.H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., et al. (2002). Mechanism of c-Myb–C/EBPβ Cooperation from Separated Sites on a Promoter. Cell *108*, 57–70. https://doi.org/10.1016/S0092-8674(01)00636-5.
35. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950. https://doi.org/10.1107/S0021889891004399.
36. Merritt, E.A., and Bacon, D.J. (1997). Raster3D: Photorealistic molecular graphics. Methods Enzymol. *277*, 505–524. https://doi.org/10.1016/S0076-6879(97)77028-9.
37. Schaub, M., Nussbaum, J., Verkade, H., Ober, E.A., Stainier, D.Y.R., and Sakaguchi, T.F. (2012). Mutation of zebrafish *Snapc4* is associated with loss of the intrahepatic biliary network. Dev. Biol. *363*, 128–137. https://doi.org/10.1016/j.ydbio.2011.12.025.
38. Dvinge, H., Guenthoer, J., Porter, P.L., and Bradley, R.K. (2019). RNA components of the spliceosome regulate tissue- and cancer-specific alternative splicing. Genome Res. *29*, 1591–1604. https://doi.org/10.1101/gr.246678.118.
39. Lareau, L.F., Brooks, A.N., Soergel, D.A.W., Meng, Q., and Brenner, S.E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. Adv. Exp. Med. Biol. *623*, 190–211.
40. Workman, E., Kolb, S.J., and Battle, D.J. (2012). Spliceosomal small nuclear ribonucleoprotein biogenesis defects and motor neuron selectivity in spinal muscular atrophy. Brain Res. *1462*, 93–99. https://doi.org/10.1016/j.brainres.2012.02.051.
41. Stamatakou, E., Wróbel, L., Hill, S.M., Puri, C., Son, S.M., Fujimaki, M., Zhu, Y., Siddiqi, F., Fernandez-Estevez, M., Manni, M.M., et al. (2020). Mendelian neurodegenerative disease genes involved in autophagy. Cell Discov. *6*, 1–13. https://doi.org/10.1038/s41421-020-0158-y.
42. Wan, J., Yourshaw, M., Mamsa, H., Rudnik-Schöneborn, S., Menezes, M.P., Hong, J.E., Leong, D.W., Senderek, J., Salman, M.S., Chitayat, D., et al. (2012). Mutations in the RNA exosome component gene *EXOSC3* cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat. Genet. *44*, 704–708. https://doi.org/10.1038/ng.2254.
43. Nixon, R.A. (2013). The role of autophagy in neurodegenerative disease. Nat. Med. *19*, 983–997. https://doi.org/10.1038/nm.3232.
44. Thiedig, K., Weisshaar, B., and Stracke, R. (2021). Functional and evolutionary analysis of the Arabidopsis 4R-MYB protein SNAPc4 as part of the SNAP complex. Plant Physiol. *185*, 1002–1020. https://doi.org/10.1093/plphys/kiaa067.
45. Coady, T.H., and Lorson, C.L. (2011). SMN in spinal muscular atrophy and snRNP biogenesis. Wiley Interdiscip. Rev. RNA *2*, 546–564. https://doi.org/10.1002/wrna.76.
46. Singh, R.N., Howell, M.D., Ottesen, E.W., and Singh, N.N. (2017). Diverse role of survival motor neuron protein. Biochim. Biophys. Acta. Gene Regul. Mech. *1860*, 299–315. https://doi.org/10.1016/j.bbagrm.2016.12.008.
47. Kour, S., Rajan, D.S., Fortuna, T.R., Anderson, E.N., Ward, C., Lee, Y., Lee, S., Shin, Y.B., Chae, J.-H., Choi, M., et al. (2021). Loss of function mutations in *GEMIN5* cause a neurodevelopmental disorder. Nat. Commun. *12*, 2558. https://doi.org/10.1038/s41467-021-22627-w.
48. Jia, Y., Mu, J.C., and Ackerman, S.L. (2012). Mutation of a U2 snRNA Gene Causes Global Disruption of Alternative Splicing and Neurodegeneration. Cell *148*, 296–308. https://doi.org/10.1016/j.cell.2011.11.057.
49. Nofrini, V., Di Giacomo, D., and Mecucci, C. (2016). Nucleoporin genes in human diseases. Eur. J. Hum. Genet. *24*, 1388–1395. https://doi.org/10.1038/ejhg.2016.25.
50. Nakaya, T. (2022). A specific gene-splicing alteration in the *SNRNP70* gene as a hallmark of an ALS subtype. Gene *818*, 146203. https://doi.org/10.1016/j.gene.2022.146203.
51. Kim, H.K., Pham, M.H.C., Ko, K.S., Rhee, B.D., and Han, J. (2018). Alternative splicing isoforms in health and disease. Pflugers Arch. *470*, 995–1016. https://doi.org/10.1007/s00424-018-2136-x.
52. Su, C.-H., D, D., and Tarn, W.-Y. (2018). Alternative Splicing in Neurogenesis and Brain Development. Front. Mol. Biosci. *5*, 12.