nature genetics

Article

https://doi.org/10.1038/s41588-025-02184-4

Dominant variants in major spliceosome U4 and U5 small nuclear RNA genes cause neurodevelopmental disorders through splicing disruption

Received: 2 October 2024

Accepted: 28 March 2025

Published online: 16 May 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

The major spliceosome contains five small nuclear RNAs (snRNAs; U1, U2, U4, U5 and U6) essential for splicing. Variants in *RNU4-2*, encoding U4, cause a neurodevelopmental disorder called ReNU syndrome. We investigated de novo variants in 50 snRNA-encoding genes in a French cohort of 23,649 individuals with rare disorders and gathered additional cases through international collaborations. Altogether, we identified 145 previously unreported probands with (likely) pathogenic variants in *RNU4-2* and 21 individuals with de novo and/or recurrent variants in *RNU5B-1* and *RNU5A-1*, encoding U5. Pathogenic variants typically arose de novo on the maternal allele and cluster in regions critical for splicing. *RNU4-2* variants mainly localize to two structures, the stem III and T-loop/quasi-pseudoknot, which position the U6 ACAGAGA box for 5′ splice site recognition and associate with different phenotypic severity. *RNU4-2* variants result in specific defects in alternative 5′ splice site usage and methylation patterns (episignatures) that correlate with variant location and clinical severity. This study establishes *RNU5B-1* as a neurodevelopmental disorder gene, suggests *RNU5A-1* as a strong candidate and highlights the role of de novo variants in snRNAs.

The splicing of pre-mRNA into mature mRNA in eukaryotic cells consists of excising introns and ligating exons through two transesterification reactions catalyzed by the spliceosome[1,2]. This large ribonucleoprotein complex is composed of five uridyl-rich small nuclear RNAs (snRNAs) that are essential for spliceosome assembly and function and differ according to the type of excised intron. The major spliceosome processes the majority (>99%) of introns containing GU–AG splice sites (U2 type) and is composed of snRNAs U1, U2, U4, U5 and U6 (ref. 3). Each snRNA has unique sequence motifs and secondary structures that allow it to interact precisely with its target sites. U1 and U2, respectively, bind to the 5′ splice sites (5′SS) and branch points, while U4, U5 and U6 form the tri-small nuclear ribonucleoprotein particle (snRNP) complex that is recruited to assemble a precatalytic spliceosome complex. U6, initially paired with U4 in an inactive conformation, activates upon dissociation to interact with U2 and form the catalytic site[4]. U5 aligns exons by binding the 5′ and 3′ splice sites, ensuring precise ligation[5].

Spliceosomal snRNAs are encoded by distinct single-exon genes and ubiquitously transcribed by RNA polymerase II or III[6]. Human genomes contain multiple gene copies of snRNAs U1, U2, U4, U5 and U6, some of which are functional and others are pseudogenes[7,8]. After transcription, snRNAs undergo essential processing steps, including 5′-capping, 3′-end processing, nuclear export, Sm protein binding, nuclear re-import and nucleotide modifications (2′-O-methylation, pseudouridylation) guided by small Cajal body-specific RNAs[9–12].

A recent landmark discovery has implicated de novo variants in *RNU4-2*, one of two functional genes encoding U4, as the cause of

e-mail: caroline.nava@aphp.fr; christel.depienne@uk-essen.de

ReNU syndrome (OMIM 620851), an unexpectedly frequent neurodevelopmental disorder (NDD)[13,14]. This discovery was facilitated by the recurrence of a single base insertion (n.64_65insT) representing 78% of pathogenic variants, enriched in the Genomics England (GEL) NDD cohort[15] but absent from gnomAD[16] and highly depleted in UK Biobank[17]. Genome sequencing is necessary to detect these variants, as they are typically not yet captured by exome sequencing. No similar enrichment was found in 28 other brain-expressed snRNA genes, although 14 regions of 13 genes appear more evolutionary constrained[13]. This raises questions of whether variants in other snRNA genes may underlie other rare diseases and how to accurately classify variants in these genes.

In this study, we investigated 50 Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC)-approved snRNA genes in a French cohort of 23,649 patients with rare disorders and collected data for additional patients via international collaborations. Using these data, we implicate two further snRNAs in NDDs and more comprehensively define ReNU syndrome. We also identify ReNU syndrome-associated transcriptional and epigenetic signatures through RNA-sequencing (RNA-seq) and DNA methylation studies.

Results

RNU4-2 variants in cohorts of patients with rare diseases

We investigated de novo variants in *RNU4-2* (NR_003137.2) and/or rare variants (<10 alleles in gnomAD v4.1.0) located in the 18-bp critical region defined in ref. 13 in the Plan France Médecine Génomique 2025 (PFMG2025) cohort comprising 23,649 patients with rare disorders (15,073 with NDD)[18]. This analysis revealed 75 patients with de novo *RNU4-2* variants. Among the patients for whom parental analysis was not possible, four had variants previously reported as de novo in another unrelated individual, and one patient had a single-nucleotide deletion (n.76del) within the critical region.

In parallel, we collected data for 70 previously unreported patients with *RNU4-2* variants identified through genome sequencing data reanalysis (30 patients) or targeted sequencing (40 patients, including one monozygotic twin pair). Variants occurred de novo in 55/56 cases for whom both parents were available. One patient had a variant (n.72_73del) inherited from an affected father, which had occurred de novo in another unrelated patient.

Altogether, 150 individuals (73 males and 77 females, including the twin) had 22 distinct *RNU4-2* variants, of which 106 patients (71%) had the recurrent n.64_65insT insertion (Fig. 1 and Supplementary Table 1). Seven other variants were recurrent—n.76C>T ($n = 10$), n.66A>G ($n = 5$), n.67A>G ($n = 5$), n.65A>G ($n = 3$), n.77_78insT ($n = 3$), n.70T>C and n.72_73del ($n = 2$ patients each). Fourteen de novo variants were identified in a single patient. All but three variants clustered within the highly conserved 18-bp critical region spanning nucleotides 62–79 (chromosome 12 (hg38) (chr12(hg38)): 120,291,825–120,291,842)[13] (Extended Data Fig. 1a). This region overlaps four distinct domains in the U4/U6 structure[4]: stem I (U4 n.62), T-loop (also known as quasi-pseudoknot; n.63–67), RBM42 interaction region (n.68–70) and stem III (n.72–79). We classified 18 variants (in 146 individuals; 145 probands) as pathogenic (P) or likely pathogenic (LP) and four as variants of uncertain significance (VUS) using American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) criteria[19,20] (Methods; Supplementary Note). We observed no difference in CADD PHRED scores or nucleotide conservation (verPhyloP) between LP/P variants and *RNU4-2* variants present in gnomAD v4.1.0. LP/P variants had a greater predicted effect on U4/U6 interaction compared to variants observed ≥10 times in gnomAD (Supplementary Fig. 1 and Supplementary Table 2), but due to the large overlap, these cannot be used to predict variant pathogenicity.

De novo analysis reveals variants in brain-expressed U5 genes

We next analyzed de novo variants in 49 additional HGNC-approved genes encoding snRNAs (Supplementary Table 3) in the PFMG cohort. This analysis revealed rare de novo variants in 17 genes in 36 unrelated patients (Supplementary Table 4). Six patients already had P/LP variants. All other patients remained unsolved after genome analysis (Supplementary Note). Fifteen variants (in 18 patients) were located in genes encoding U5 (six in *RNU5A-1*/NR_002756.2, seven in *RNU5B-1*/NR_002757.3, two in *RNU5E-1*/NR_002754.2 and three in *RNU5F-1*/NR_002753.5). As *RNU5A-1* and *RNU5B-1* are the main genes encoding U5 in the brain (Extended Data Fig. 2), we focused our analysis on these two genes.

In the 100,000 Genomes Project (GEL) and National Health Service (NHS) Genomic Medicine Service (GMS) cohorts available within GEL, six NDD probands had five de novo variants in *RNU5B-1*, compared to a single de novo variant in a non-NDD individual (6/12,724 versus 1/30,058; Fisher's $P = 0.0036$; Supplementary Table 5). Furthermore, five patients with de novo *RNU5B-1* variants were identified across additional cohorts, including the Broad Center for Mendelian Genomics (two patients), the BCH Epilepsy Genetics Program, the Australian Undiagnosed Diseases Network (UDN-Aus) and Care4Rare Canada (one patient each).

Altogether, 18 NDD probands had de novo variants in *RNU5B-1*. Seven of these variants (14 patients) clustered within a small region spanning chr15(hg38): 65,304,713–65,304,720, corresponding to the highly conserved U5 5′ loop I (Fig. 2 and Supplementary Table 6), which is depleted in variants in gnomAD v4.1.0 (Extended Data Fig. 1b) and UK Biobank[13]. Extending the analysis of this critical region to patients analyzed as singletons or duos, we identified three additional patients with variants in GEL/NHS, bringing the total to eight patients with NDD with variants in this region, compared to none in the non-NDD cohort (8/12,724 undiagnosed NDD versus 0/30,058 non-NDD in GEL/NHS-GMS; Fisher's $P = 6.1 \times 10^{-5}$). In total, 17 NDD individuals had variants in the critical region of *RNU5B-1*. Three variants (n.39C>G, n.42_43insA and n.44AG) were recurrent, identified in six, three and four patients, respectively. These variants were absent from all databases and classified as LP.

Pathogenic *RNU* variants mainly occur on the maternal allele

We investigated the parental origin of *RNU4-2* variants in available genome data by phasing de novo variants and informative SNPs in the flanking regions (Methods; Supplementary Fig. 2). The parental origin of mutations was reliably determined in 50 trios and one mother–patient duo. Variants were assigned to the maternal allele in 47 cases and to the paternal allele in four instances. Notably, all n.64_65insT insertions ($n = 38$) were phased to the maternal allele, consistent with previous observations[13]. Of the four variants assigned to the paternal allele, two (n.62T>C and n.68A>C) were classified LP/P and two (n.76del and n.92C>G) were VUS.

Among the phaseable *RNU5B-1* variants located in the U5 5′ loop I, five (n.39C>G, n.42_43insA and three n.44A>G) were phased to the maternal allele, while two (n.39C>G and n.37G>C) were on the paternal allele. The two *RNU5B-1* de novo variants located outside of the conserved 5′ loop I (n.24G>C and n.74T>C) were also phased to the paternal allele. Both n.40_41insA variants in *RNU5A-1* occurred de novo on the maternal allele.

RNU4-2 variants in the T-loop and stem III differ in severity

Clinical data were available for 143 patients with P/LP *RNU4-2* variants (69 males and 74 females, excluding the monozygotic twin with an identical phenotype to her sister; Supplementary Tables 7 and 8). The median age at study entry was 9 years (range = 4 months to 45 years). All patients had NDD with variable degrees of intellectual disability (ID), ranging from mild (7.1%), moderate (27.7%) to severe/profound (65.2%).

We investigated genotype–phenotype correlations in *RNU4-2*-related disorders. Unsupervised clustering of clinical features revealed two separate clusters differing in severity (Extended Data Fig. 3a). Most *RNU4-2* variants in stem III (63%, 12/19) were in the mild

https://doi.org/10.1038/s41588-025-02184-4

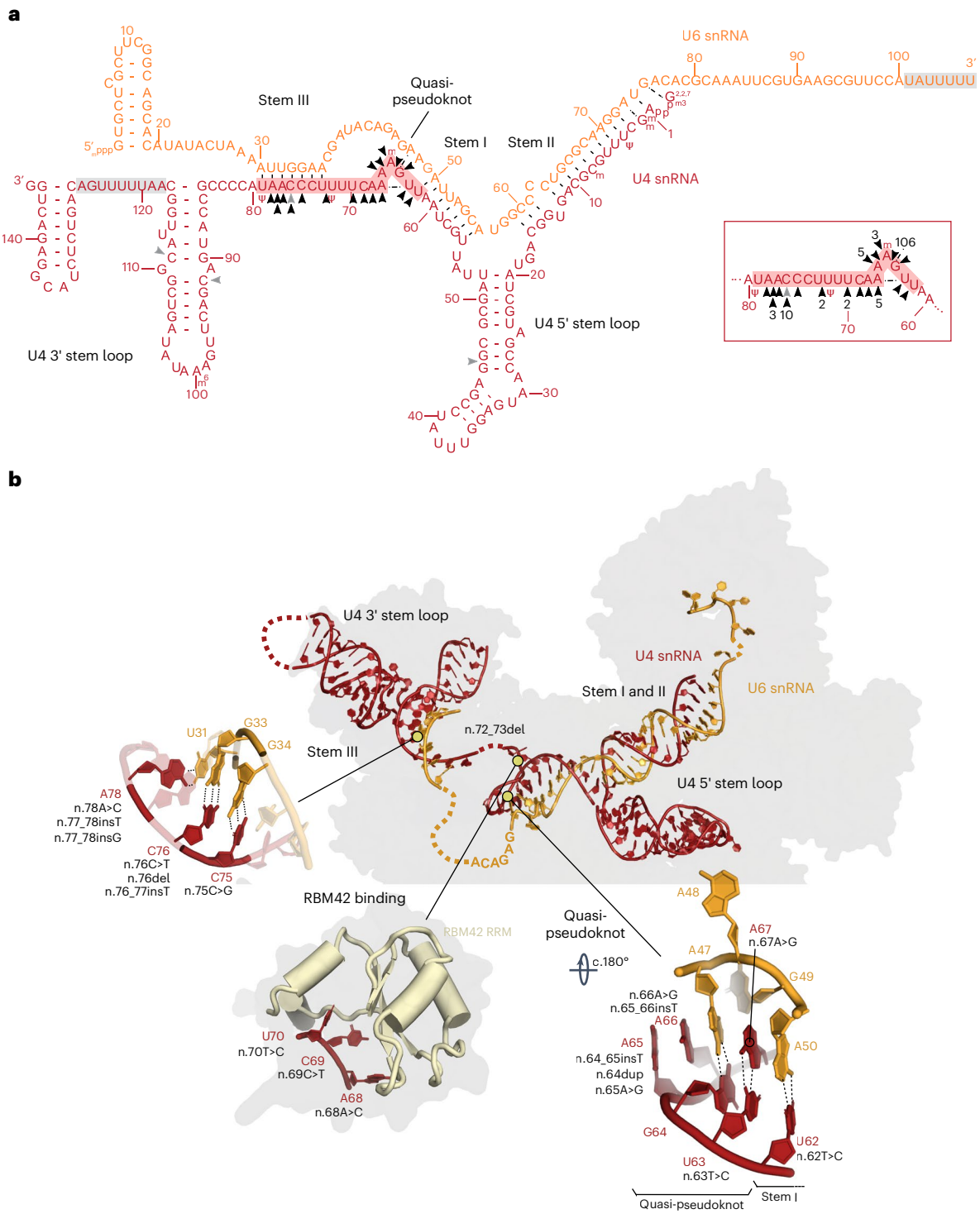


Fig. 1 | Overview of *RNU4-2* variants identified in this study. a, Two-dimensional predicted structure of the interaction between U4 (red) and U6 (orange) snRNAs showing distinct domains. Arrowheads indicate variants identified in this study; P and LP in black, and VUS in gray. The numbers in black within the zoom-in box represent the count of patients with each variant for nucleotide changes that occur more than once. Red and orange numbers refer to the numbering of nucleotides from each snRNA. Red-shaded region, 18-bp critical region; gray-shaded regions, Sm sites. **b**, Organization of the U4–U6 duplex at the tri-snRNP stage (PDB ID: 6QW6) and close-up views of stem III, RBM42 binding and quasi-pseudoknot regions. Interactions stabilizing these structures, as well as LP/P variants potentially affecting their stability, are represented. Ψ, pseudouridine; m, 2′-O-methyl residues; m6, N6-methyladenosine; ^{2,2,7}m3Gppp, 2,2,7-trimethylguanosine cap; $_{m}$pppG, 5′ guanosine triphosphate cap with γ-monomethyl phosphate.

phenotype cluster, whereas most variants in the T-loop and RBM42 interacting region were in the high severity cluster (98%, 121/123). Principal component analysis (PCA) confirmed this result, with variants in stem III and in the T-loop separating on the first principal component (PC) axis, accounting for 12.3% of the variance (Fig. 3a,b and Extended Data Fig. 3b–d). These results suggest that phenotypic variability largely depends on the location of *RNU4-2* variants within U4 functional domains.

https://doi.org/10.1038/s41588-025-02184-4

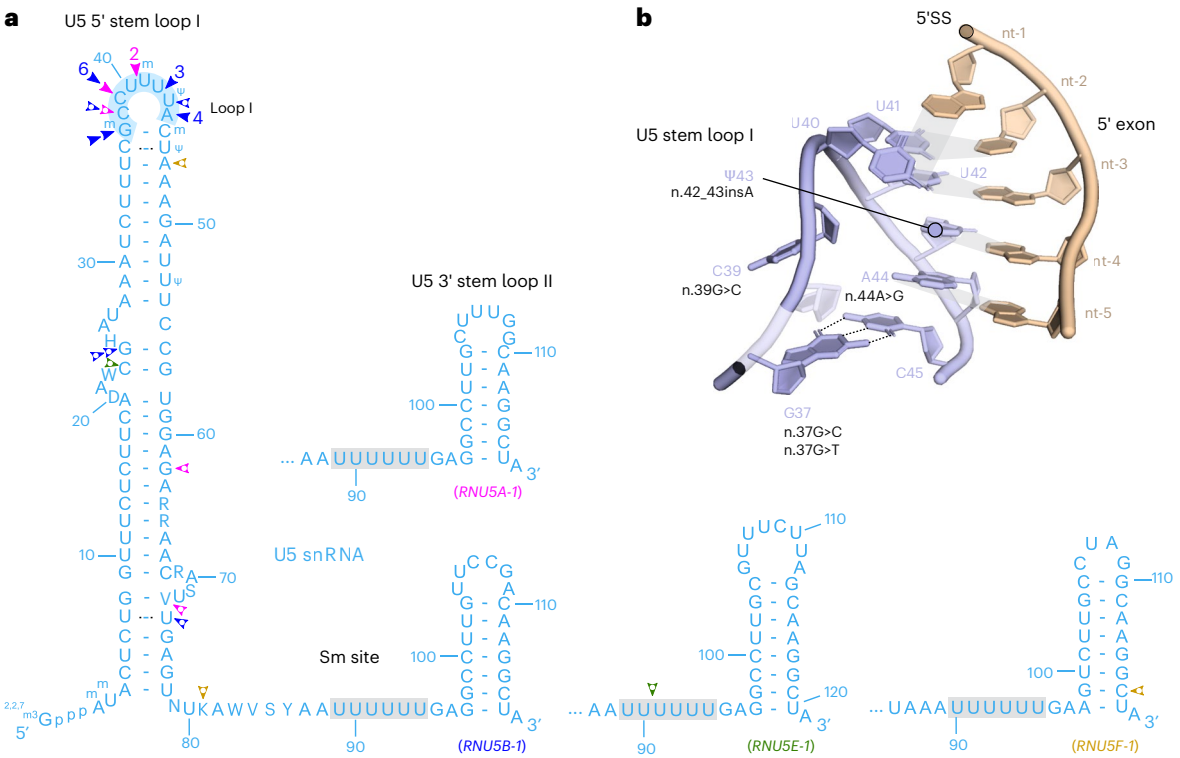


Fig. 2 | Overview of *RNU5A-1*, *RNU5B-1*, *RNU5E-1* and *RNU5F-1* variants identified in this study. a, Two-dimensional predicted structure of U5 (light blue) snRNA showing distinct domains. Arrowheads indicate variants identified in this study—pink, *RNU5A-1*; dark blue, *RNU5B-1*; green, *RNU5E-1* and yellow, *RNU5F-1*. P and LP variants are indicated with a filled color, while VUS are marked with a white dot inside the arrowheads. Numbers near the arrowheads represent the count of patients with each variant for nucleotide changes that occur more than once. Nucleotide differences between *RNU5A-1*, *RNU5B-1*, *RNU5E-1* and *RNU5F-1* are shown using International Union of Pure and Applied Chemistry (IUPAC) codes, except for the highly variable 3′ stem loop II, for which separate loops are displayed. Light blue numbers refer to the numbering of nucleotides from each snRNA. The N at position 79 corresponds to a gap in *RNU5F-1*. Blue-shaded region, critical region. Gray-shaded region, Sm site. **b**, The 5′ exon recognition by the U5 stem loop I at the B-complex stage (PDB ID: 8Q7N). Interactions stabilizing these structures, as well as LP/P variants potentially affecting their stability, are represented.

Among the 103 patients with *RNU4-2* c.64_65insT and clinical data available (Table 1 and Extended Data Fig. 4), prenatal findings were observed in 55 of 92 cases (60%) and predominantly consisted of intrauterine growth restriction (IUGR; 30%) and/or cerebral abnormalities (33%; ventriculomegaly, 19.5%); 38% of fetuses showed isolated abnormalities, while 62% had two or more signs. Neonatal findings (91%), mainly hypotonia (71%) and feeding difficulties (57%), were frequent. Congenital microcephaly was present in 28% of individuals, while microcephaly at the time of last examination was present in 74% of individuals (Supplementary Fig. 3). In total, 60% of individuals had short stature. All 85 patients older than 3 years exhibited developmental delay. Most could walk, with a median walking age of 30 months (range = 13 months to 12 years), but 13% did not reach this milestone. Most patients were nonverbal (61%) or could only speak a few words (34%). The majority had severe/profound ID (78%), with 21% having moderate ID and one patient having mild ID. Behavioral disturbances were common, with autistic features and/or midline stereotypies reminiscent of Rett syndrome in 84% of patients. Epilepsy affected 56%, with an additional 8% experiencing a single seizure. Seizure onset ranged from the neonatal period to 13 years (median = 32 months), and seizures were usually generalized, rare, fever-sensitive and responsive to antiepileptic medications. However, 5 patients were diagnosed with developmental and epileptic encephalopathy, 14 experienced status epilepticus and 7 had drug-resistant epilepsy. Brain magnetic resonance imaging (MRI) abnormalities were prevalent (91%), with the most common findings being enlarged ventricles (84%) and corpus callosum abnormalities (85%). Less common findings included heterotopia ($n = 7$), delayed myelination or hypomyelination ($n = 11$) and abnormal gyration ($n = 5$). In total, 38% of cases had skeletal abnormalities, including osteopenia or fractures ($n = 20$) and hip dysplasia ($n = 10$). Dysmorphic features suggested Pitt–Hopkins syndrome (Fig. 4). Strabismus and drooling were common. Feeding difficulties affected 69%, failure to thrive 55% and constipation 57%. Acrocyanosis or vasomotor disorders (Extended Data Fig. 5) were present in 16 patients, blood count anomalies in 13 patients and hypothyroidism in 8 patients.

The phenotype of the patients with other variants in the T-loop and RBM42 interaction region was indistinguishable from that of patients with n.64_65insT. Individuals with the recurrent n.66A>G ($n = 5$) and n.67A>G ($n = 5$) variants had a similar phenotype, characterized by neonatal hypotonia (5/5 and 3/5), microcephaly (5/5 for both), epilepsy in about half (3/5 and 2/4) and similar dysmorphic features. All patients had severe developmental delay and severe ID, except for one case with moderate ID. Notably, all patients were nonverbal.

Patients with *RNU4-2* n.76C>T variant ($n = 10$) exhibited a distinct clinical profile from patients with n.64_65insT (Table 1 and Fig. 3c). They had less neonatal findings (Fisher's $P = 2.41 \times 10^{-4}$), especially hypotonia ($P = 1.67 \times 10^{-3}$), presented less severe ID ($P = 8.57 \times 10^{-5}$) and developmental delay ($P = 2.05 \times 10^{-5}$), were more proficient in their language abilities ($P = 1.32 \times 10^{-8}$) and rarely showed brain MRI abnormalities ($P = 2.47 \times 10^{-3}$). All patients could walk, with four of them achieving this milestone at a normal age (median walking age = 19 months (12–33 months)), and all could speak, with simple sentences ($n = 6$) or normal language skills ($n = 4$). Microcephaly was noted in three of ten patients, and short stature was noted in only one of ten patients. Two of five patients had autistic features. Six patients had fever-sensitive generalized epilepsy, well-controlled with antiseizure medication, while four

https://doi.org/10.1038/s41588-025-02184-4

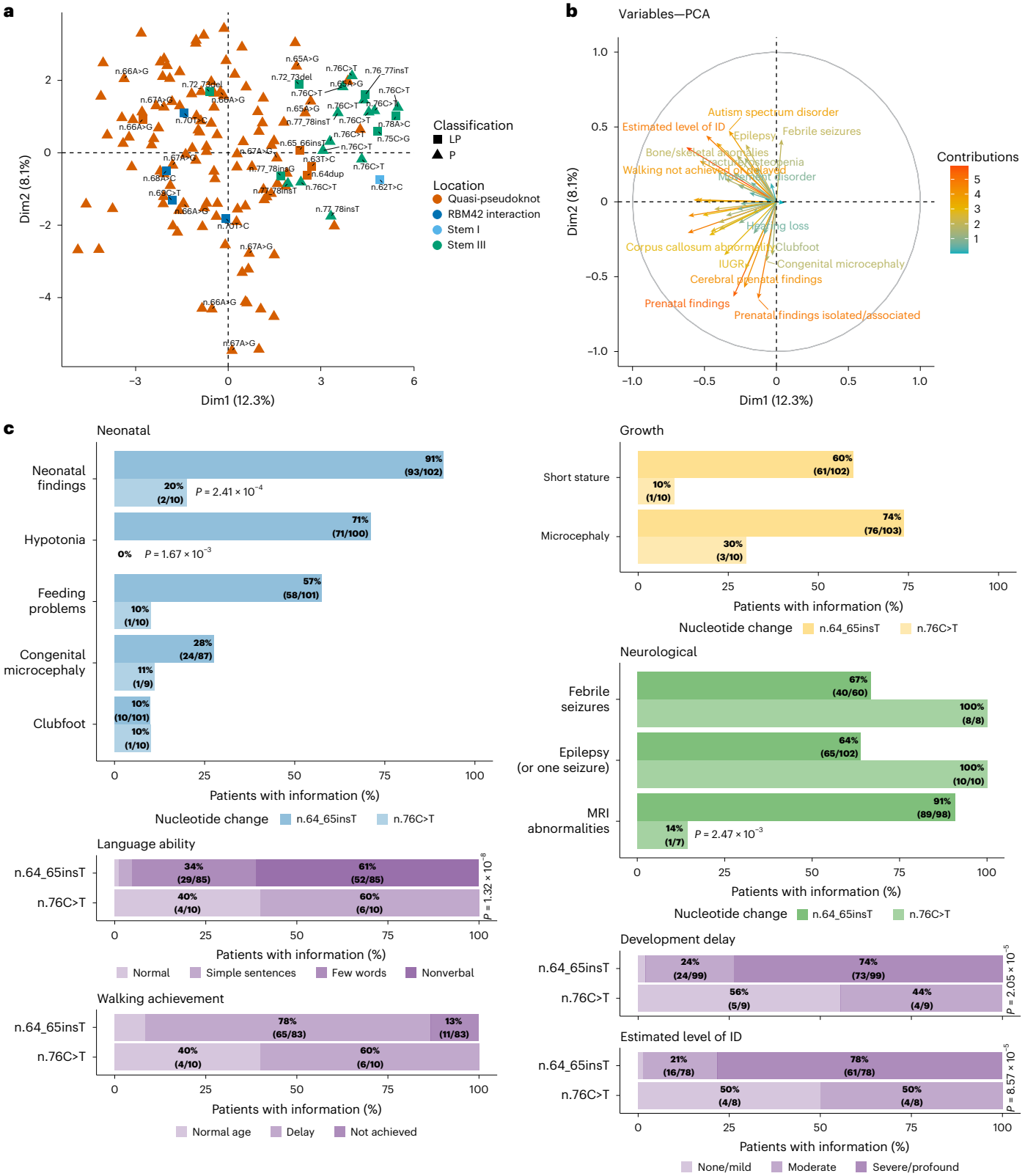


Fig. 3 | *RNU4-2* variants in the T-loop and stem III associate with different phenotype severity. a, PCA of 44 phenotypic features in 143 patients showing the separation of variants with respect to their location within distinct U4:U6 domains. Labels with the nucleotide change appear for variants other than n.64_65insT. *RNU4-2* variants are colored according to their location within the distinct U4:U6 domains; stem I ($n = 1$) in light blue, quasi-pseudoknot ($n = 119$) in orange, RBM42 interaction region ($n = 4$) in blue and stem III ($n = 19$) in green. Triangles, P ($n = 128$) variants; squares, LP ($n = 15$) variants. **b**, Contributions of the clinical features to the PCA. **c**, Comparative analysis of 14 phenotypes related to *RNU4-2* n.64_65insT and n.76C>T variants. The *P* values were calculated using Fisher's exact tests (two-sided; 2 × 2, 2 × 3 or 2 × 4 contingency tables) to compare 41 phenotypes between patients with n.64_65insT variants and those in the other three variant groups. Multiple comparisons were adjusted for using Bonferroni correction. The percentage of patients with the feature, followed by the numerator (number of affected patients) and denominator (total assessed), is shown directly in the bars. Full details of all tests and patient numbers can be found in Supplementary Table 8.

https://doi.org/10.1038/s41588-025-02184-4

Table 1 | Clinical features of individuals with *RNU4-2* variants according to the location of the variants in the different U4 functional domains

Parameters		Total	n.64_65insT	n.76C>T	Other variants in the T-loop or RBM42 interaction region	Other variants in the stem III
Patients		146[a]	106	10	20	9
With clinical data		143[a]	103	10	20	9
Prenatal findings		70/125 (56%)	55/92 (60%)	3/10 (30%)	9/17 (53%)	3/6 (50%)
IUGR		41/131 (31.3%)	28/94 (30%)	1/9 (11%)	9/20 (45%)	3/8 (38%)
Cerebral abnormalities		29/91 (31.87%)	27/82 (33%)	0/1 (0%)	2/6 (33%)	0/2 (0%)
Neonatal findings		115/140 (82.14%)	93/102 (91%)	2/10 (20%)	17/20 (85%)	3/8 (38%)
Neonatal hypotonia		88/138 (63.77%)	71/100 (71%)	0/10 (0%)	15/20 (75%)	2/8 (25%)
Neonatal feeding problems		73/138 (52.9%)	58/101 (57%)	1/10 (10%)	12/19 (63%)	2/8 (25%)
Congenital microcephaly		30/122 (24.59%)	24/87 (28%)	1/9 (11%)	3/19 (16%)	2/7 (29%)
Microcephaly		97/141 (68.79%)	76/103 (74%)	3/10 (30%)	15/19 (79%)	3/9 (33%)
Short stature		72/140 (51.43%)	61/102 (60%)	1/10 (10%)	9/19 (47%)	1/9 (11%)
Walking	Not achieved	16/115 (13.91%)	11/83 (13%)	0/10 (0%)	5/13 (38%)	0/9 (0%)
	Delay	82/115 (71.3%)	65/83 (78%)	6/10 (60%)	7/13 (54%)	4/9 (44%)
	Normal age	17/115 (14.78%)	7/83 (8%)	4/10 (40%)	1/13 (8%)	5/9 (56%)
Language ability	Nonverbal	61/116 (52.59%)	52/85 (61%)	0/10 (0%)	9/12 (75%)	0/9 (0%)
	Few words	31/116 (26.72%)	29/85 (34%)	0/10 (0%)	1/12 (8%)	1/9 (11%)
	Simple sentences	16/116 (13.79%)	3/85 (4%)	6/10 (60%)	1/12 (8%)	6/9 (67%)
	Normal	8/116 (6.9%)	1/85 (1%)	4/10 (40%)	1/12 (8%)	2/9 (22%)
Developmental delay	Severe	87/134 (64.93%)	73/99 (74%)	0/9 (0%)	13/18 (72%)	1/8 (12%)
	Moderate	37/134 (27.61%)	24/99 (24%)	4/9 (44%)	5/18 (28%)	4/8 (50%)
	Mild	9/134 (6.72%)	2/99 (2%)	5/9 (56%)	0/18 (0%)	2/8 (25%)
	No	1/134 (0.75%)	0/99 (0%)	0/9 (0%)	0/18 (0%)	1/8 (12%)
Estimated level of ID	Severe	73/112 (65.18%)	61/78 (78%)	0/8 (0%)	12/17 (71%)	0/9 (0%)
	Moderate	31/112 (27.68%)	16/78 (21%)	4/8 (50%)	5/17 (29%)	6/9 (67%)
	Mild	8/112 (7.14%)	1/78 (1%)	4/8 (50%)	0/17 (0%)	3/9 (33%)
Autism spectrum disorder		52/92 (56.52%)	43/69 (62%)	2/7 (29%)	5/11 (45%)	2/5 (40%)
Epilepsy		81/140 (57.86%)	57/102 (56%)	6/10 (60%)	11/19 (58%)	7/9 (78%)
Febrile seizures		56/81 (69.14%)	40/60 (67%)	8/8 (100%)	4/7 (57%)	4/6 (67%)
Status epilepticus		20/68 (29.41%)	17/51 (33%)	0/5 (0%)	3/9 (33%)	0/3 (0%)
Drug-resistance		12/72 (16.67%)	10/52 (19%)	0/6 (0%)	2/8 (25%)	0/6 (0%)
Abnormal brain MRI		107/130 (82.31%)	89/98 (91%)	1/7 (14%)	15/18 (83%)	2/7 (29%)
Enlarged ventricles		58	54	0	4	0
Corpus callosum abnormality		66	53	1	11	1
Cardiac abnormalities		21/112 (18.75%)	15/85 (18%)	1/8 (12%)	4/14 (29%)	1/5 (20%)
Renal/genitourinary abnormalities		24/110 (21.82%)	17/83 (20%)	1/5 (20%)	5/16 (31%)	1/6 (17%)
Bone/skeletal anomalies		42/123 (34.15%)	35/92 (38%)	1/9 (11%)	4/16 (25%)	2/6 (33%)
Eyes/vision abnormalities		70/124 (56.45%)	59/95 (62%)	0/6 (0%)	7/15 (47%)	4/8 (50%)
Hearing loss		9/125 (7.2%)	6/91 (7%)	0/7 (0%)	3/18 (17%)	0/9 (0%)
Teeth/dental anomalies		19/99 (19.19%)	13/72 (18%)	1/7 (14%)	3/14 (21%)	2/6 (33%)
Skin abnormalities		28/115 (24.35%)	20/84 (24%)	3/7 (43%)	4/17 (24%)	1/7 (14%)
Feeding issues		80/125 (64%)	63/91 (69%)	1/7 (14%)	13/18 (72%)	3/9 (33%)
Failure to thrive		65/126 (51.59%)	51/93 (55%)	2/7 (29%)	10/18 (56%)	2/8 (25%)
Constipation		63/123 (51.22%)	51/90 (57%)	2/8 (25%)	8/18 (44%)	2/7 (29%)
Joint hyperlaxity		46/112 (41.07%)	37/82 (45%)	0/7 (0%)	6/15 (40%)	3/8 (38%)
Acrocyanosis		18/108 (16.67%)	16/79 (20%)	0/6 (0%)	2/16 (12%)	0/7 (0%)
Blood count abnormality		15/93 (16.13%)	13/69 (19%)	0/7 (0%)	1/12 (8%)	1/5 (20%)

A more detailed table with statistical tests is available as Supplementary Table 8. [a]One patient with a variant (n.62T>C) located in stem loop I.

https://doi.org/10.1038/s41588-025-02184-4

Fig. 4 | Facial photographs from 22 patients with the recurrent *RNU4-2* c.64_65insT variant. a–**v**, The main facial features include a large mouth, a short philtrum, downturned corners of the mouth, thick lips, deep-set eyes, sparse eyebrows and strabismus. Older individuals also showed facial asymmetry. **a**–**v** correspond to unrelated patients except **p** and **q** who are monozygotic twins. Consent forms have been obtained for the publication of the facial photographs.

others had a single febrile seizure. None had nystagmus, and only one had ataxia. Brain MRI was normal in six of eight cases. Dysmorphic features were distinct from those seen in patients with the recurrent variant (Fig. 5a).

Similarly, patients with other variants in the stem III ($n = 9$) exhibited a mild/moderate phenotype compared to patients with the n.64_65insT variant, showing less severe developmental delay ($P = 1.90 \times 10^{-2}$) and ID ($P = 7.79 \times 10^{-5}$) and with improved language abilities ($P = 3.20 \times 10^{-6}$). All patients could walk and speak, with varying degrees of language development (normal language, two; simple sentences, six; few words, one). ID was mild in three and moderate in six, with autistic features in two of five cases. Fever-sensitive epilepsy was common (7/9) but well-controlled with antiseizure medication. Brain MRI was normal in five of seven patients.

RNU5 variants lead to NDD with variable malformations

Detailed clinical data were available for nine of 15 patients with NDD with *RNU5B-1* LP variants (Fig. 5b, Supplementary Tables 9 and 10 and Supplementary Fig. 4). Six had severe developmental delay, one had moderate developmental delay and one had normal cognition but attention difficulties. All nine patients showed brain MRI abnormalities, but only one had epilepsy. Three had pectus excavatum, two of whom also had marfanoid habitus. Three had ocular abnormalities, such as congenital glaucoma ($n = 1$), small papillae with retinal vascular tortuosity ($n = 1$) and severe myopia (<−12.25 D). Other malformations included pulmonary issues ($n = 2$), sacrococcygeal abnormalities ($n = 2$), tooth agenesis or fusion ($n = 2$) and cardiac malformation ($n = 2$). Acquired microcephaly was noted in three individuals with n.44A>G, whereas two individuals with n.39C>G had macrocephaly. Human phenotype ontology terms enriched in *RNU5B-1* cases from GEL include seizures, macrocephaly and eye anomalies (Supplementary Table 11).

The three patients with *RNU5A-1* variants for whom clinical data were available also had NDD with variable congenital malformations. One had postaxial polydactyly, dental agenesis and talus feet due to oligohydramnios. Another had anal malposition, sacrococcygeal dimple and caudal appendix, thin and incomplete corpus callosum and septal

https://doi.org/10.1038/s41588-025-02184-4

Fig. 5 | Facial photographs from 13 patients with other variants in *RNU4-2* and two patients with variants in *RNU5B-1*. a, Individuals with other variants in *RNU4-2*. (i), n.65A>G; (ii)–(iv), n.66A>G; (v) and (vi), n.67A>G; (vii), n.68A>C; (viii)–(x), n.76C>T; (xi), n.77_78insG; (xii) and (xiii), n.77_78insT. **b**, Individuals with the *RNU5B-1* n.39C>G variant (i and ii). Consent forms have been obtained for the publication of the facial photographs.

agenesis. The third had cardiac malformations and marfanoid habitus. The two patients with n.40_41insA had seizures. Head circumference (HC) was normal in all.

Pathogenic variants lead to specific splicing defects

We previously reported specific alternative 5′ splice site (5′SS) abnormalities in the blood of individuals with *RNU4-2* variants[13]. To confirm and extend this observation, we conducted RNA-seq on lymphocyte cultures from 19 individuals with *RNU4-2* variants and 21 controls with other NDDs (Supplementary Fig. 5a). Using rMATS-turbo[21], we identified significant aberrant splicing events (Supplementary Tables 12–16). We extracted percent spliced in (PSI) values of significantly altered exons for each splicing category and performed PCA using matrices with samples as columns and PSI values as rows. PCA revealed that the most pronounced effect was for the signal originating from 111 altered 5′SS, with distinct clustering patterns of affected individuals (Fig. 6a and Extended Data Fig. 6a–d). Severe phenotypes (associated with variants n.64_65insT, n.67A>G, n.68A>C and n.70T >C) formed a distinct cluster, while mild phenotypes (n.72_73del, n.75C>G and n.76C>T) appeared intermediate between severe cases and controls. This suggests a common 5′SS usage signature associated with *RNU4-2* pathogenic variants, with distinct profiles correlating with disease severity.

We then applied the 5′SS signature to three individuals with a VUS or atypical clinical presentation. The mildly affected proband with n.64_65insT clustered with more severely affected carriers of the same variant (Fig. 6a). The n.62C>T variant was reclassified as LP due to clustering with mild phenotypes, whereas n.45_46insT showed an intermediate profile between patients and controls and remained a VUS.

We characterized and visually inspected 69 5′SS events using Integrative Genomics Viewer (IGV), including 50 shared by patients with mild and severe phenotypes and 19 unique to severe phenotypes (Fig. 6b–d, Supplementary Fig. 5b and Supplementary Table 17). Decreased 5′SS consistently shows high spliceAI scores (shared sites, median = 0.93; severe-only sites, median = 0.92), indicating that alternative 5′SS usage is not restricted to weak sites (Fig. 6b). In contrast, increased 5′SS events associated with severe phenotypes have significantly lower SpliceAI scores (median = 0.52; Mann–Whitney *U* test, $P = 0.014$) compared to shared sites (median = 0.80). Additionally, only five 5′SS events were absent from controls (mean supporting reads in controls < 3; 4/69, 5.8%), suggesting that the main effect of U4 variants is a shift in existing alternative isoforms rather than the use of new cryptic splice sites.

Analysis of the 5′SS usage patterns revealed a consistent trend—A/A/G nucleotides at positions +3/+4/+5 were frequently replaced

https://doi.org/10.1038/s41588-025-02184-4

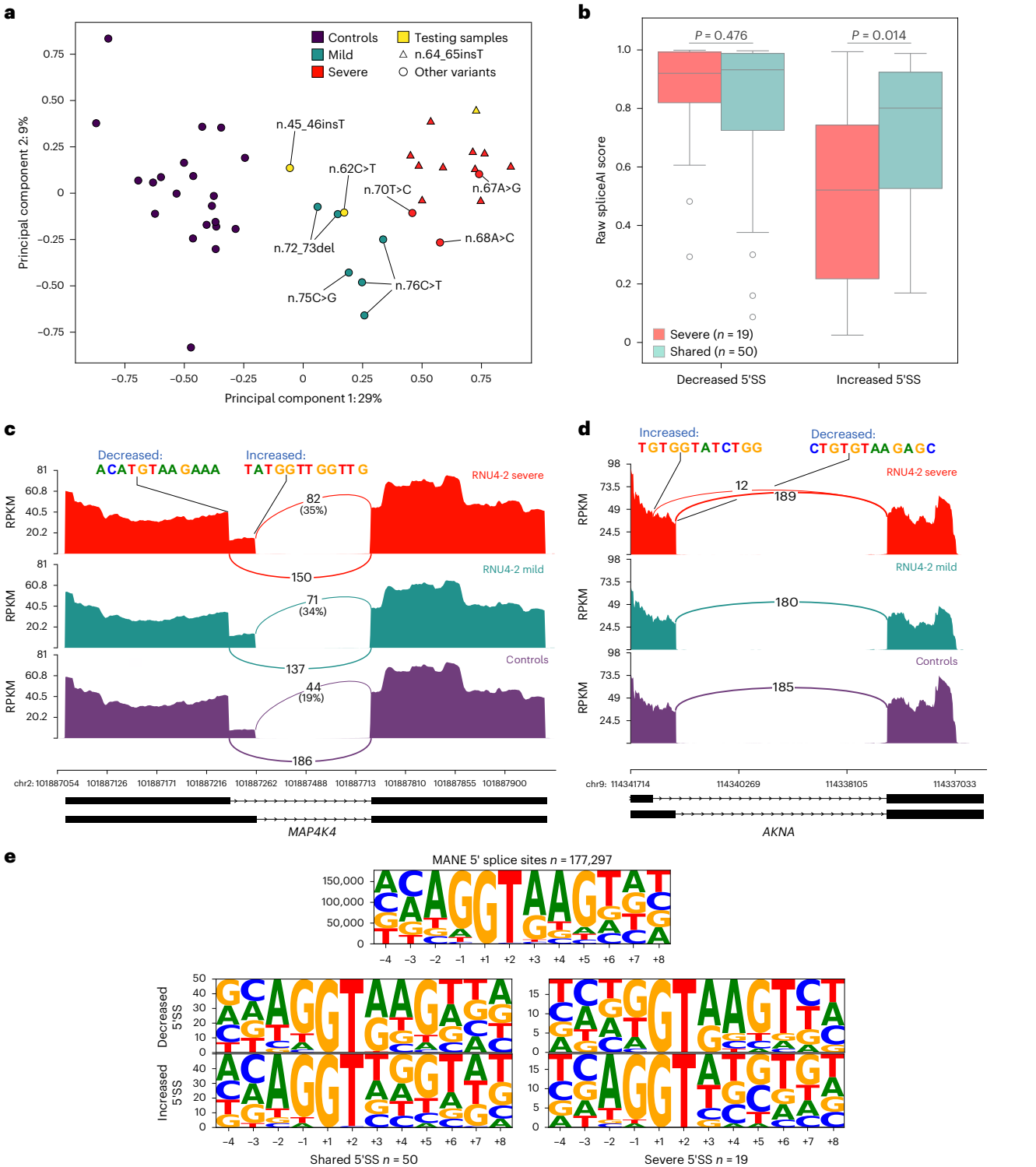


Fig. 6 | RNA-seq identifies an alternative 5′SS signature that differentiates severe from mild *RNU4-2*-related phenotypes. a, PCA based on PSI values from 111 significant 5′SS events detected using rMATS, comparing 19 patients with *RNU4-2* variants (6 mildly affected in teal and 13 severely affected in red) to 21 controls (purple). Triangles, n.64_65insT; circles, other variants. Yellow symbols correspond to three test samples—one variant of uncertain significance (VUS; n.45_46insT), one VUS that could be reclassified as LP and the recurrent variant n.64_65insT from a patient with a milder phenotype. **b**, Box plot showing raw spliceAI scores of the decreased 5′SS site and the increased 5′SS for the 50 events shared between mild ($n = 6$) and severe individuals ($n = 13$) and the 19 events only detected in severe individuals. SpliceAI scores for severe and shared 5′SS were not statistically different for decreased sites ($P = 0.476$) but were significant for increased sites ($P = 0.014$) using the two-sided Mann–Whitney *U* test. Box plot elements are defined as follows: centerline, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers. **c**,**d**, Sashimi plots showing isoform shifts in *MAP4K4* (**c**) and *AKNA* (**d**). Aggregated coverage and splicing-supporting reads from patients with *RNU4-2* variants with mild (n.75C>G, n.76C>T and n.72_73del; $n = 6$) or severe (n.64_65insT, n.67A>G, n.68A>C and n.70T>C; $n = 13$) phenotypes and controls ($n = 21$) are shown. The *MAP4K4* event is present in both patient groups, while the *AKNA* abnormality appears only in severe cases. **e**, Consensus nucleotide sequence of decreased and increased 5′SS for 50 shared events (left) and for the 19 severe-only events (right), in comparison to the consensus sequence of all 5′SS from MANE transcripts (top).

by C/T (Fig. 6e). This change was accompanied by an increased reliance on A/G nucleotides at positions −2/−1, particularly evident in 5′SS events that were reduced in severe phenotypes. Indeed, 2 of 19 severe variants had AG at these positions, compared to 26 of 50 shared variants (two-tailed Fisher's test, $P = 0.0008$). These findings indicate that 5′SSs used exclusively in patients with severe phenotypes tend to depend more on the exonic sequence (end of the exon) and less on the intronic sequence.

Of the 69 5′SS events, 21 (30%) were out-of-frame, indicating that these transcripts were not degraded by nonsense-mediated mRNA decay (NMD). To test whether NMD masked additional out-of-frame 5′SS events, we repeated the splicing signature analysis on puromycin-treated samples. This revealed only nine significant 5′SS defects not present in controls, including one in *KDM6A* (Supplementary Table 18). These results suggest that reduced mRNA expression is not a key factor in *RNU4-2* variant pathogenicity.

To explore the underlying pathophysiology, we examined 5′SS events affecting known NDD-associated genes. We identified 14 such genes: *DPM1*, *KIF2A*, *P4HTM*, *HNRNPH1*, *KMT2A*, *KMT2C*, *KMT2D*, *POMT1*, *MADD*, *TRIO*, *EIF4A2*, *SYNCRIP*, *THOC2* and *PPP1CB*. Notably, three genes belonging to the KMT2 family of H3K4 methyltransferases, *KMT2A* (Wiedemann–Steiner syndrome), *KMT2C* (Kleefstra syndrome) and *KMT2D* (Kabuki syndrome), are related to syndromic NDDs.

Finally, we analyzed samples from six patients with either *RNU5B-1* (two n.39C>G and two n.44A>G) or *RNU5A-1* (n.40_41insA and n.39del) variants for splicing effects. These variants did not share the *RNU4-2* 5′SS signature (Extended Data Fig. 6e), and unlike *RNU4-2* variants, *RNU5* variants lacked a shared 5′SS or 3′SS signature. An individual analysis against 21 controls, supplemented with 20 additional controls and 19 *RNU4-2* patients, revealed possible variant-specific effects—*RNU5B-1* n.39C>G mainly affected 5′SS, while *RNU5B-1* n.44A>G primarily impacted 3′SS (Extended Data Fig. 7 and Supplementary Tables 19 and 20). *RNU5A-1* n.39del may also alter 3′SS, while *RNU5A-1* n.40_41insA seems to affect both 5′SS and 3′SS (Extended Data Fig. 8 and Supplementary Tables 21 and 22). These results, although obtained from a limited number of patients, may underline the clinical variability observed for patients with *RNU5* variants.

Identification of a specific *RNU4-2* episignature

KMT2A and *KMT2D* variants are associated with specific DNA methylation profiles (episignatures). To determine whether an episignature could also be identified for ReNU syndrome, we compared genome-wide methylation profiles of 35 patients with P/LP *RNU4-2* variants with those of 45 healthy age-matched controls. Adjusting for age, sex and blood cell composition, we identified 147 differentially methylated positions ($P < 10^{-7}$ and $|\Delta\beta| > 5\%$; Supplementary Fig. 6). PCA and heatmap representations clearly separated patients from controls (Fig. 7). The strength of the episignature correlated with disease severity and variant localization. Variants associated with mild phenotypes showed similar levels of hypermethylation as moderate-to-severe phenotypes but exhibited intermediate hypomethylated signals in the lower heatmap cluster (Supplementary Table 23). The first PC axis, separating patients from controls, captured 52% of the methylation dispersion, comparable to *ATRX*, *KMT2D* or *KMT2A* episignatures[22]. After fivefold cross-validation, the overall sensitivity was 0.91 (31/35, 95% binomial confidence interval (CI; 0.77–0.98)), reaching 100% (24/24, 95% binomial CI (0.86–1.00)) for n.64_65insT carriers, with a specificity of 0.98 (44/45, 95% binomial CI (0.88–0.999); Fig. 7b). Two patients with *RNU5A-1* n.40_41insA did not share the *RNU4-2* signature but clustered together at the boundary of the control group, suggesting these variants may have their own distinct episignature (Extended Data Fig. 9). The *RNU4-2* episignature is entirely distinct from the *KMT2A* and *KMT2D* episignatures (Extended Data Fig. 10).

Discussion

Despite extensive genetic testing, 40–60% of NDD cases with suspected genetic origins remain unsolved. The recent discovery of *RNU4-2* variants as a major cause of NDDs, overlooked until 2024, underscores the role of noncoding genes in undiagnosed cases. Here we analyzed 50 HGNC-approved snRNA genes in a large French cohort of patients who underwent genome sequencing as part of routine diagnosis. This led to identifying 76 (likely) pathogenic variants in *RNU4-2* (0.5% of NDD participants) and eight (0.05%) with variants in *RNU5A-1* or *RNU5B-1*. Combining these data with 80 additional patients from other cohorts, we observed that pathogenic variants typically cluster in evolutionarily conserved regions of U4 and U5 critical for splicing. *RNU4-2* variants cluster in the T-loop/quasi-pseudoknot and stem III, while *RNU5B-1* and *RNU5A-1* variants cluster in the conserved 5′ loop I, which pairs with the exon adjacent to the 5′SS[5]. De novo variants in other domains or snRNA genes were also identified, but their clinical relevance remains unclear so far.

This study provides a comprehensive overview of *RNU4-2*-related phenotypes, revealing distinct clinical outcomes based on variant location. Variants in the T-loop, including n.64_65insT, are associated with severe phenotypes, while variants in stem I/stem III, including n.62T>C, n.76C>T and c.72_73del, lead to milder forms. This supports a continuum of *RNU4-2*-related phenotypes with inherited variants also possibly contributing to NDD etiology. At the severe end, prenatal manifestations, mainly cerebral abnormalities (corpus callosum anomalies and enlarged ventricles) and/or IUGR, were observed in 60% of cases, highlighting the importance of genome sequencing or targeted *RNU4-2* analysis in prenatal genetic testing.

A striking observation in line with previous findings[13] is the predominant maternal origin of *RNU4-2* variants, possibly explained by the negative selection of variants severely affecting splicing in the male germline. However, paternal transmission of less severe variants is possible, as evidenced by four cases. The mechanism underlying the high recurrence of *RNU4-2* and its potential link to maternal origin remain unclear. Interestingly, recurrent insertions in *RNU4-2* n.64_65insT and *RNU5A-1* n.40_41insA occur at 2′-O-methylation sites[11,12], although any connection to maternal inheritance or recurrence is yet to be established.

We provide definitive evidence that *RNU4-2* pathogenic variants lead to specific alternative 5′SS anomalies in the blood cells of affected patients, with detected events correlating with phenotype severity. Variants in the T-loop and stem I/stem III indeed show distinct, partially overlapping transcriptional signatures, which could aid in the interpretation of VUS. Furthermore, DNA methylation exhibited a similar pattern, revealing a shared global episignature, albeit with more pronounced and distinct alterations associated with severe NDD phenotypes linked to T-loop variants. Given the widespread use of exome sequencing in routine diagnostics, these transcriptional and epigenetic signatures could help diagnose additional ReNU syndrome cases worldwide. This analysis also revealed that pathogenic variants in snRNAs lead to widespread but mild splicing abnormalities, mainly characterized by a shift in existing isoforms. Although this analysis was performed in lymphocytes rather than neuronal cells, the data suggest that ubiquitously expressed genes, such as *KMT2A*, *KMT2C*, *KMT2D* and *KDM6A*, which encode lysine methyltransferases and a lysine demethylase involved in chromatin remodeling, may also be altered in the brain of affected individuals. Interestingly, a recent study demonstrated that USP39 deficiency disrupts the assembly of the U4/U6.U5 tri-snRNP, resulting in 5′SS abnormalities similar to those observed in patients with *RNU4-2* pathogenic variants. This disruption leads to the accumulation of misfolded proteins in proteotoxic aggregates, triggering endoplasmic reticulum stress and subsequent cell death[23].

Although U4 variants could disrupt spliceosome function at various stages (U4 snRNP biogenesis, U4/U6 di-snRNP, U4/U6.U5 tri-snRNP assembly and spliceosome activation), our results strongly suggest

https://doi.org/10.1038/s41588-025-02184-4

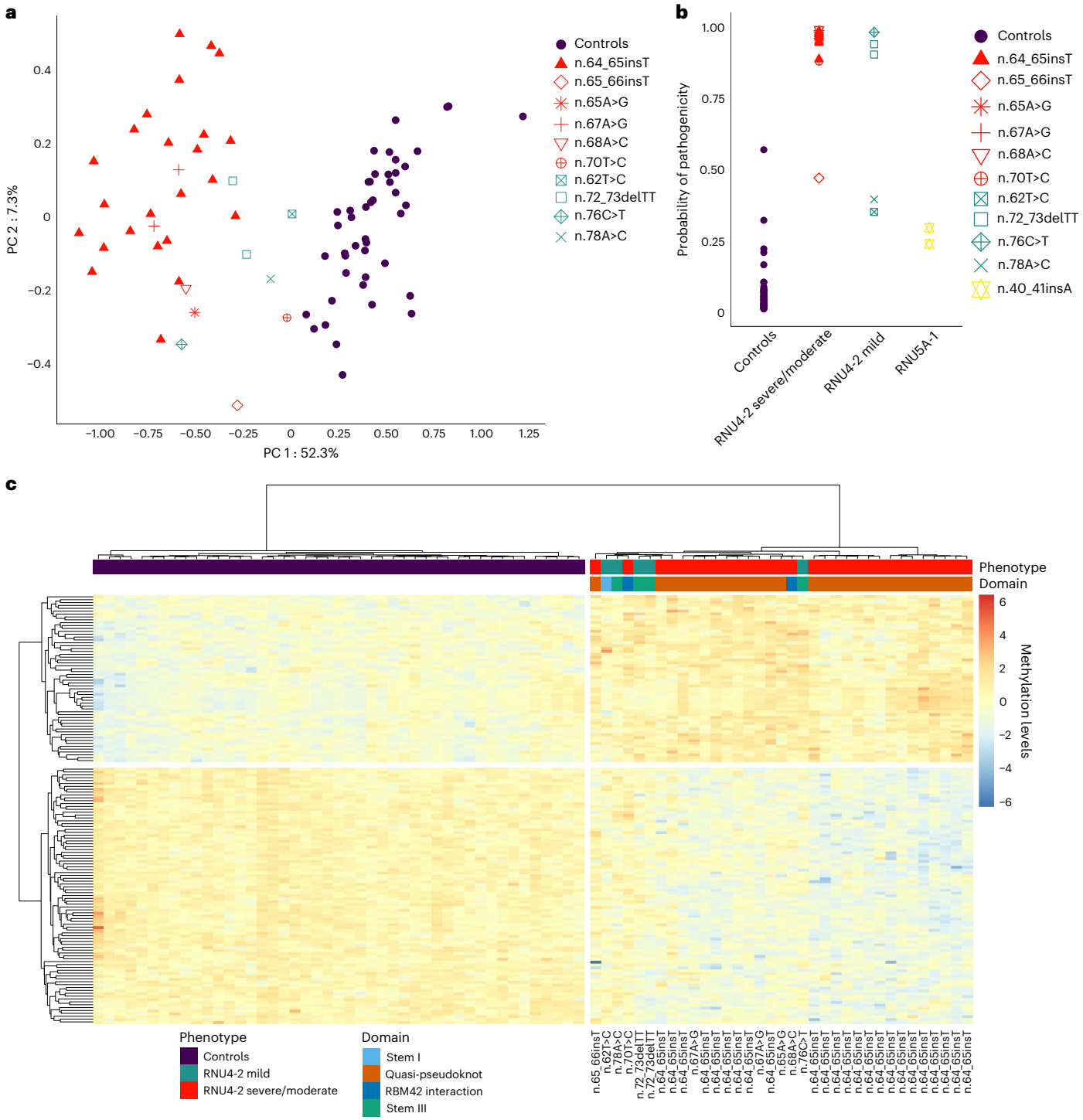


Fig. 7 | Identification of an episignature that discriminates patients with ReNU syndrome from controls and correlates with phenotypic severity. a, PCA of adjusted methylation levels at differentially methylated positions ($n = 147$), after correcting for expected methylation based on age, sex and estimated blood cell counts ($n = 80$ individuals in total). Different variants are represented by different shapes, while phenotype severity is indicated by color—purple for controls ($n = 45$), teal for mild phenotypes ($n = 5$) and red for moderate-to-severe phenotypes ($n = 30$). The percentage of variance explained is provided for each axis. **b**, Pathogenicity scores for each variant were obtained by fivefold cross-validation using a support vector machine predictor. Purple, controls; teal, *RNU4-2* variants, mild phenotypes; red, *RNU4-2* variants, moderate-to-severe phenotypes; yellow, *RNU5A-1* variants. **c**, Heatmap of adjusted methylation levels displays hierarchical clustering of controls and patients with *RNU4-2* LP/P variants. Blue indicates hypomethylated positions ($n = 89$), while red indicates hypermethylated positions ($n = 58$) with respect to expected methylation levels at equivalent age, sex and blood cell composition. Variants are colored according to the location within the distinct U4:U6 domains (stem I, light blue; quasi-pseudoknot, orange; RBM42 interaction region, blue; stem III, green). Phenotype severity is indicated by color (purple for controls, teal for mild phenotypes and red for moderate-to-severe phenotypes).

disruption of the U4/U6 duplex organization at the tri-snRNP stage, affecting the 5′SS introduction into the spliceosome's active site. The 5′SS, initially paired with U1 in the prespliceosome, is transferred to the U6 ACAGAGA box and U5 stem loop 1 in the U4/U6.U5 tri-snRNP. These interactions maintain the 5′SS in the active site during catalysis, marking the start of each intron[1,24]. When the 5′SS is transferred, it pairs

with the U6 ACAGAGA box to ensure its correct identification[25], which triggers molecular events that lead to the formation of the active site by Brr2. Before this transfer, the U6 ACAGAGA box is held as a flexible loop between the stem III and quasi-pseudoknot, and this organization is further stabilized by Snu66, SNRP27K and RBM42 (refs. 4,26–28). Pathogenic *RNU4-2* variants in—or close to—the quasi-pseudoknot possibly weaken or disrupt its structure and compromise its ability to maintain the ACAGAGA box at the right position for 5′SS recognition. Pathogenic variants in the stem III possibly alter Watson–Crick base pairs formed by U4 A78, C76 and C75 with U6 G34, G33 and U31 and affect U4/U6 stem III's stability. The stem III likely enhances 5′SS recognition fidelity by creating an energy barrier to extending the U6/5′SS helix after initial pairing with the ACAGAGA sequence[4,5]. Stem III disruption, necessary for Brr2 loading and active site formation, would weaken this barrier, allowing suboptimal 5′SS to extend the helix more easily and activate the spliceosome. In severe cases, increased 5′SS usage suggests a loss of specificity for intronic 5′SS motifs, potentially compensated by greater reliance on exonic 5′SS motifs via U5.

U5 loop I is crucial for 5′SS transfer. By interacting with the exonic sequences adjacent to 5′SS, it helps align the 5′ exon with the branch site and the 3′ exon during both steps of splicing catalysis[29,30]. Mutations in yeast U5 loop I result in aberrant 5′SS splicing[31]. However, our findings suggest variant-specific impacts, with *RNU5B-1* n.39C>G affecting 5′SS, *RNU5B-1* n.44A>G affecting the 3′SS and *RNU5A-1* n.40_41insA possibly altering both. These results align with the clinical variability observed in patients with *RNU5B-1* variants, including opposing head growth phenotypes associated with n.39C>G and n.44A>G. However, further studies on larger patient series are needed to confirm these observations and elucidate the underlying mechanisms. Overall, our findings underline the critical role of U4 and U5 RNA structures in maintaining splicing fidelity by preventing weak SS activation, with their destabilization reducing splicing accuracy.

Finally, this study focused on 50 HGNC-approved snRNA genes, while the hg38 reference genome includes 1,901 snRNA genes, most annotated as pseudogenes. A recent study suggests that *RNU2-2P*, annotated as a pseudogene, may be functional and linked to a new NDD[32]. While confirming the functionality of snRNA pseudogenes requires experimental validation[8], this discovery suggests that more snRNAs could contribute to genetic disorders.

In conclusion, this work emphasizes the critical role of de novo variants in snRNAs, particularly *RNU4-2*, in unsolved NDD. Moreover, we identify *RNU5B-1* and *RNU5A-1* as new NDD genes and provide valuable insights into fundamental aspects of spliceosome function.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02184-4.

References

1. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
2. Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
3. Vazquez-Arango, P. & O'Reilly, D. Variant snRNPs: new players within the spliceosome system. *RNA Biol.* **15**, 17–25 (2018).
4. Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5′ splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).
5. Artemyeva-Isman, O. V. & Porter, A. C. G. U5 snRNA interactions with exons ensure splicing precision. *Front. Genet.* **12**, 676971 (2021).
6. Perumal, K. & Reddy, R. The 3′ end formation in small RNAs. *Gene Expr.* **10**, 59–78 (2002).
7. Seal, R. L. et al. A guide to naming human non-coding RNA genes. *EMBO J.* **39**, e103777 (2020).
8. Mabin, J. W., Lewis, P. W., Brow, D. A. & Dvinge, H. Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA* **27**, 1186–1203 (2021).
9. Patel, S. B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* **36**, 6482–6493 (2008).
10. Will, C. L. & Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).
11. Morais, P., Adachi, H. & Yu, Y. T. Spliceosomal snRNA epitranscriptomics. *Front. Genet.* **12**, 652129 (2021).
12. Darzacq, X. et al. Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746–2756 (2002).
13. Chen, Y. et al. De novo variants in the RNU4-2 snRNA cause a frequent neurodevelopmental syndrome. *Nature* **632**, 832–840 (2024).
14. Greene, D. et al. Mutations in the U4 snRNA gene RNU4-2 cause one of the most prevalent monogenic neurodevelopmental disorders. *Nat. Med.* **30**, 2165–2169 (2024).
15. *The National Genomics Research Library v5.1* (Genomics England, 2020); https://figshare.com/articles/dataset/GenomicEnglandProtocol_pdf/4530893?file=22714349
16. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
17. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
18. PFMG2025 Contributors. PFMG2025—integrating genomic medicine into the national healthcare system in France. *Lancet Reg. Health Eur.* **50**, 101183 (2025).
19. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
20. Ellingford, J. M. et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* **14**, 73 (2022).
21. Wang, Y. et al. rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nat. Protoc.* **19**, 1083–1104 (2024).
22. Husson, T. et al. Episignatures in practice: independent evaluation of published episignatures for the molecular diagnostics of ten neurodevelopmental disorders. *Eur. J. Hum. Genet.* **32**, 190–199 (2024).
23. Prieto-Garcia, C. et al. Pathogenic proteotoxicity of cryptic splicing is alleviated by ubiquitination and ER-phagy. *Science* **386**, 768–776 (2024).
24. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
25. Kandels-Lewis, S. & Seraphin, B. Involvement of U6 snRNA in 5′ splice site selection. *Science* **262**, 2035–2039 (1993).
26. Zhang, Z. et al. Structural insights into the cross-exon to cross-intron spliceosome switch. *Nature* **630**, 1012–1019 (2024).
27. Sarka, K., Katzman, S. & Zahler, A. M. A role for SNU66 in maintaining 5′ splice site identity during spliceosome assembly. *RNA* **30**, 695–709 (2024).
28. Zahler, A. M. et al. SNRP-27, the *C. elegans* homolog of the tri-snRNP 27K protein, has a role in 5′ splice site positioning in the spliceosome. *RNA* **24**, 1314–1325 (2018).
29. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68**, 743–754 (1992).
30. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).

31. Newman, A. & Norman, C. Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell* **65**, 115–123 (1991).
32. Greene, D. et al. Mutations in the U2 snRNA gene *RNU2-2P* cause a severe neurodevelopmental disorder with prominent epilepsy. Preprint at *medRxiv* https://doi.org/10.1101/2024.09.03.24312863 (2024).

Caroline Nava[1,2,3,152] ✉, Benjamin Cogne[3,4,5,152], Amandine Santini[6,152], Elsa Leitão[7], François Lecoquierre[3,8], Yuyang Chen[9,10], Sarah L. Stenton[11,12], Thomas Besnard[4,5], Solveig Heide[2], Sarah Baer[13,14], Abhilasha Jakhar[15,16], Sonja Neuser[17], Boris Keren[2,3], Anne Faudet[2], Sylvie Forlani[1], Marie Faoucher[3,18,19,20], Kevin Uguen[3,21,22], Konrad Platzer[17], Alexandra Afenjar[23], Jean-Luc Alessandri[24], Stephanie Andres[25], Chloé Angelini[26], Bernard Aral[20,27], Benoit Arveiler[26,28], Tania Attie-Bitach[3,29,30], Marion Aubert Mucca[31], Guillaume Banneau[20,31], Tahsin Stefan Barakat[32,33], Giulia Barcia[3,29,30], Stéphanie Baulac[1], Claire Beneteau[20,26], Fouzia Benkerdou[2], Virginie Bernard[20], Stéphane Bézieau[4,5], Dominique Bonneau[34,35], Marie-Noelle Bonnet-Dupeyron[36], Simon Boussion[37], Odile Boute[37], Elise Brischoux-Boucher[38], Samantha J. Bryen[39,40], Julien Buratti[2], Tiffany Busa[41], Almuth Caliebe[42], Yline Capri[43], Kévin Cassinari[3,8], Roseline Caumes[37], Camille Cenni[44], Pascal Chambon[3,8], Perrine Charles[2], John Christodoulou[45,46], Cindy Colson[37], Solène Conrad[4], Auriane Cospain[18], Juliette Coursimault[8], Thomas Courtin[29,30], Madeline Couse[47], Charles Coutton[20,48,49], Isabelle Creveaux[20,50], Alissa M. D'Gama[51,52,53,54], Benjamin Dauriat[20,55], Jean-Madeleine de Sainte Agathe[2,3], Giulia Del Gobbo[56], Andrée Delahaye-Duriez[3,57,58,59], Julian Delanne[60,61], Anne-Sophie Denommé-Pichon[20,61,62], Anne Dieux-Coeslier[37], Laura Do Souto Ferreira[4], Martine Doco-Fenzy[4,20,63], Stephan Drukewitz[17], Véronique Duboc[64], Christèle Dubourg[3,18,19], Yannis Duffourd[61,62], David Dyment[56], Salima El Chehadeh[65], Monique Elmaleh[66], Laurence Faivre[60,61], Samuel Fennelly[26], Hanna Fischer[67], Mélanie Fradin[68], Camille Galludec Vaillant[2], Benjamin Ganne[20,69], Jamal Ghoumid[37], Himanshu Goel[70,71], Zeynep Gokce-Samar[72,73], Alice Goldenberg[8], Romain Gonfreville Robert[4], Svetlana Gorokhova[20,41,74], Louise Goujon[43], Victoria Granier[31], Mathilde Gras[2], John M. Greally[75], Bianca Greiten[42], Paul Gueguen[3,76,77], Anne-Marie Guerrot[8], Saurav Guha[78], Anne Guimier[29,30], Tobias B. Haack[79], Hamza Hadj Abdallah[3,29,30], Yosra Halleb[80], Radu Harbuz[20,48,49], Madeleine Harris[46], Julia Hentschel[17], Bénédicte Héron[81], Marc-Phillip Hitz[42,82], A. Micheil Innes[83,84], Vincent Jadas[85], Louis Januel[20,86], Nolwenn Jean-Marçais[68], Vaidehi Jobanputra[78,87], Florence Jobic[88], Ludmila Jornea[1], Céline Jost[60,61], Sophie Julia[31], Frank J. Kaiser[7,89], Daniel Kaschta[42], Sabine Kaya[7], Petra Ketteler[7,90], Bochra Khadija[7,91,92], Fabian Kilpert[7], Cordula Knopp[93], Florian Kraft[93], Ilona Krey[17], Marilyn Lackmy[94], Fanny Laffargue[95], Laetitia Lambert[96,97], Ryan Lamont[83,84], Vincent Laugel[13], Steven Laurie[98], Julie L. Lauzon[83,84], Louis Lebreton[20,99], Marine Lebrun[20,100], Marine Legendre[26], Eric Leguern[1,2], Daphné Lehalle[2], Elodie Lejeune[2], Gaetan Lesca[20,86,101], Marion Lesieur-Sebellin[3,29,30], Jonathan Levy[3,43], Agnès Linglart[102,103,104], Stanislas Lyonnet[29,30], Kevin Lüthy[7], Alan S. Ma[105,106], Corinne Mach[2], Jean-Louis Mandel[14], Lamisse Mansour-Hendili[3], Julien Marcadier[83], Victor Marin[20,99], Henri Margot[20,26,28], Valentine Marquet[20,55], Angèle May[8], Johannes A. Mayr[107], Catherine Meridda[108], Vincent Michaud[20,26,28], Caroline Michot[29,30], Gwenael Nadeau[109], Sophie Naudion[26], Laetitia Nguyen[2], Mathilde Nizon[4], Frédérique Nowak[110], Sylvie Odent[19,68], Valerie Olin[2], Ikeoluwa A. Osei-Owusu[11], Matthew Osmond[56], Katrin Õunap[111,112], Laurent Pasquier[68], Sandrine Passemard[59,113], Melissa Pauly[114], Olivier Patat[31], Marine Pensec[22], Laurence Perrin-Sabourin[43], Florence Petit[37], Christophe Philippe[20,115], Marc Planes[22], Annapurna Poduri[51,54,116], Céline Poirsier[117], Antoine Pouzet[43], Bradley Prince[83], Clément Prouteau[34], Aurora Pujol[118,119,120], Caroline Racine[60,61], Mélanie Rama[20,37], Francis Ramond[20,100], Kara Ranguin[121], Margaux Raway[38], André Reis[114,122], Mathilde Renaud[96,97], Nicole Revencu[123], Anne-Claire Richard[8], Lucile Riera-Navarro[64], Rocio Rius[39,40,45], Diana Rodriguez[81,124], Agustí Rodriguez-Palmero[125], Sophie Rondeau[3,29,30], Annika Roser-Unruh[126], Christelle Rougeot Jung[127], Hana Safraou[20,61,62], Véronique Satre[20,48,49], Pascale Saugier-Veber[3,8], Clément Sauvestre[26], Elise Schaefer[65], Wanqing Shao[54], Ina Schanze[128], Jan-Ulrich Schlump[129,130], Agatha Schlüter Martin[118,119], Caroline Schluth-Bolard[20,131,132], Sarah Schuhmann[114], Christopher Schröder[7], Monisha Sebastin[75], Sabine Sigaudy[41], Malte Spielmann[42], Marta Spodenkiewicz[133], Laura St Clair[105], Julie Steffann[3,29,30],

Radka Stoeva[80], Harald Surowy[7], Mark A. Tarnopolsky[134], Calina Todosi[135,136], Annick Toutain[137], Frédéric Tran Mau-Them[20,61,62,138], Astrid Unterlauft[139], Julien Van-Gils[26,28], Clémence Vanlerberghe[37], Georgia Vasileiou[114,122], Gabriella Vera[8], André Verdel[49], Alain Verloes[43,59,140], Yoann Vial[3,43,59], Cédric Vignal[43], Marie Vincent[4], Catherine Vincent-Delorme[37], Aline Vincent-Devulder[108], Antonio Vitobello[20,61,62], Sacha Weber[108], Marjolaine Willems[141], Khaoula Zaafrane-Khachnaoui[64], Pia Zacher[142], Lena Zeltner[143], Alban Ziegler[20,144], Wojciech P. Galej[145], Hélène Dollfus[65,132], Christel Thauvin[60,61,62], Kym M. Boycott[56,146], Pierre Marijon[3], Alban Lermine[3], Valérie Malan[3,29,30], Marlène Rio[29,30], Alma Kuechler[7], Bertrand Isidor[4], Séverine Drunat[3,43,59], Thomas Smol[3,37], Nicolas Chatron[20,86,101], Amélie Piton[14,20,131], Gael Nicolas[3,8], Matias Wagner[147,148,149], Rami Abou Jamra[17], Delphine Héron[2,140], Cyril Mignot[2,140], Pierre Blanc[3], Anne O'Donnell-Luria[11,12,150], Nicola Whiffin[9,10,11], Camille Charbonnier[151], Clément Charenton[15,16], Julien Thevenon[20,48,49,153] & Christel Depienne[3,7,153]

[1]Sorbonne Université, Institut du Cerveau—Paris Brain Institute—ICM, Inserm, CNRS, APHP, Hôpital Pitié-Salpêtrière, Paris, France. [2]Assistance Publique-Hôpitaux de Paris (APHP) Sorbonne Université, Département de Génétique Médicale, Hôpital Pitié-Salpêtrière, Paris, France. [3]Laboratoire SeqOIA, Paris, France. [4]Nantes Université, CHU de Nantes Service de Génétique Médicale, Nantes, France. [5]Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, Nantes, France. [6]Univ. Rouen Normandie, Normandie Univ., Inserm U1245, Rouen, France. [7]Institute of Human Genetics, University Hospital Essen, University Duisburg–Essen, Essen, Germany. [8]Univ. Rouen Normandie, Normandie Univ., Inserm U1245 and CHU Rouen, Department of Genetics and Reference Center for Developmental Abnormalities, Rouen, France. [9]Big Data Institute, University of Oxford, Oxford, UK. [10]Centre for Human Genetics, University of Oxford, Oxford, UK. [11]Broad Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [12]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [13]Service de Neuropédiatrie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France. [14]Institute of Genetics and Cellular and Molecular Biology (IGBMC), INSERM-U964, CNRS-UMR7104, University of Strasbourg, Illkirch, France. [15]CNRS, Inserm, Université de Strasbourg, IGBMC UMR 7104-UMR-S 1258, Illkirch, France. [16]Department of Integrated Structural Biology, IGBMC, Illkirch, France. [17]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany. [18]Laboratoire de Génétique Moléculaire et Génomique, FHU GenOMedS, CHU Rennes, Rennes, France. [19]University of Rennes, CNRS, INSERM, IGDR (Institut de Génétique et Développement de Rennes)-UMR 6290, ERL U1305, Rennes, France. [20]GCS AURAGEN, Lyon, France. [21]Univ. Brest, Inserm, EFS, UMR 1078, GGB, Brest, France. [22]Service de Génétique Médicale et Biologie de la Reproduction, CHRU de Brest, Brest, France. [23]Assistance Publique-Hôpitaux de Paris (APHP) Sorbonne Université, Centre de Référence Malformations et Maladies Congénitales du Cervelet et Déficiences Intellectuelles de Causes Rares, UF de Génétique Clinique, Hôpital Trousseau, Paris, France. [24]Service de Génétique, CHU de La Réunion, Saint-Pierre, France. [25]Medicover München Ost MVZ Humangenetik, Munich, Germany. [26]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France. [27]Laboratoire de Génétique Chromosomique et Moléculaire, Pôle Biologie, CHU de Dijon, Dijon, France. [28]INSERM U1211, University of Bordeaux, Bordeaux, France. [29]Assistance Publique-Hôpitaux de Paris (APHP), Service de Médecine Génomique des Maladies Rares, Hôpital Necker-Enfants malades, Paris, France. [30]Université Paris Cité, INSERM, IHU Imagine—Institut des maladies génétiques, Paris, France. [31]Service de Génétique Médicale, CHU Purpan, Toulouse, France. [32]Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [33]Discovery Unit, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [34]Department of Genetics, Angers University Hospital, Angers, France. [35]UMR CNRS 6214—INSERM 1083, Angers, France. [36]Consultations de Génétique, Centre Hospitalier de Valence, Valence, France. [37]Univ. Lille, CHU Lille, ULR7364 – RADEME, Lille, France. [38]Centre de Génétique Humaine—CHU Besançon, Université de Bourgogne-Franche-Comté, Besançon, France. [39]Centre for Population Genomics, Garvan Institute of Medical Research and UNSW Sydney, Sydney, New South Wales, Australia. [40]Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [41]Medical Genetics Department, Timone Children's Hospital, APHM, Marseille, France. [42]Institute of Human Genetics, University Hospitals Schleswig-Holstein, University of Lübeck and Kiel University, Lübeck, Kiel, Germany. [43]Assistance Publique-Hôpitaux de Paris (APHP), Département de Génétique, Hôpital Robert-Debré, Paris, France. [44]Service de Génétique Moléculaire, Chromosomique et Clinique, CHU de Nîmes, Nîmes, France. [45]Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia. [46]Murdoch Children's Research Institute, Parkville, Victoria, Australia. [47]Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada. [48]Service de Génétique, Génomique et Procréation, CHU Grenoble Alpes, Grenoble, France. [49]Université Grenoble Alpes, INSERM U 1209, CNRS UMR 5309, Institute for Advanced Biosciences, Grenoble, France. [50]CHU Clermont-Ferrand Department of Medical Biochemistry and Molecular Biology, Clermont-Ferrand, France. [51]Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital, Boston, MA, USA. [52]Division of Newborn Medicine, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA. [53]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [54]Children's Rare Disease Collaborative, Boston Children's Hospital, Boston, MA, USA. [55]Service de Génétique Médicale, Cytogénétique et Biologie de la Reproduction, CHU de Limoges, Limoges, France. [56]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada. [57]Hôpitaux Universitaires de Paris Seine-Saint-Denis-APHP, UF de médecine génomique et génétique Clinique, Hôpital Jean Verdier, Bondy, France. [58]UFR Santé Médecine et Biologie Humaine, Université Sorbonne Paris Nord, Bobigny, France. [59]NeuroDiderot, Inserm, Université Paris Cité, Paris, France. [60]Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, Centre de Référence Déficiences Intellectuelles de Causes Rares, FHU TRANSLAD, Institut GIMI, Dijon, France. [61]Inserm UMR1231 GAD, Université Bourgogne, Dijon, France. [62]Laboratoire de Génomique Médicale—Centre NEOMICS, CHU Dijon Bourgogne, Dijon, France. [63]Service de Génétique, CHU de Reims, Reims, France. [64]Université Côte d'Azur, Centre Hospitalier Universitaire de Nice, Inserm U1081, CNRS UMR7284, IRCAN, Nice, France. [65]Service de Génétique Médicale, Institut de Génétique Médicale d'Alsace (IGMA), CHU Strasbourg, Strasbourg, France. [66]Assistance Publique-Hôpitaux de Paris (AP-HP), DMU DREAM, Hôpital Robert Debré, Service de Radiologie Pédiatrique, Paris, France. [67]Sozialpädiatrisches Zentrum Konstanz, Konstanz, Germany. [68]Service de Génétique Clinique, Centre de Référence 'Anomalies du Développement et Syndromes Malformatifs' de l'Inter-région Ouest, FHU GenOMedS, CHU Rennes Hôpital Sud, Rennes, France. [69]Laboratoire de Génétique Chromosomique, CHU de Montpellier, Montpellier, France. [70]General Genetics Service, Hunter Genetics, Waratah, New South Wales, Australia. [71]School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, New South Wales, Australia. [72]Department of Epilepsy, Sleep and Pediatric Neurophysiology, Hospices Civils de Lyon, Lyon, France. [73]University of Lyon, Lyon, France. [74]Aix Marseille University, INSERM, Marseille Medical Genetics, U1251, Marseille, France. [75]Department of Pediatrics, Division of Pediatric Genetic, Medicine, Children's Hospital at Montefiore/Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY, USA. [76]Service de Génétique, CHU de Tours,

Tours, France. [77]Université de Tours, INSERM, Imaging Brain and Neuropsychiatry iBraiN U1253, Tours, France. [78]Molecular Diagnostics, New York Genome Center, New York City, NY, USA. [79]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. [80]Service de Génétique Médicale, Centre Hospitalier du Mans, Le Mans, France. [81]Assistance Publique-Hôpitaux de Paris (APHP) Sorbonne Université, Département de Neurologie Pédiatrique, Hôpital Armand Trousseau-La Roche Guyon, Fédération Hospitalo-Universitaire I2-D2, Paris, France. [82]Department of Medical Genetics, Carl von Ossietzky University, Oldenburg, Germany. [83]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada. [84]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada. [85]Service de Pédiatrie, CHMS Chambéry, CAMSP Chambéry, Chambéry, France. [86]Genetics Department, Hospices Civils de Lyon, Lyon, France. [87]Department of Pathology, Columbia University Irving Medical Center, New York City, NY, USA. [88]Clinical Genetics Unit, University Hospital of Amiens, Amiens, France. [89]Essen Center for Rare Diseases (EZSE), Essen, Germany. [90]Department of Pediatrics III, University Hospital Essen, Essen, Germany. [91]Higher Institute of Biotechnology of Monastir, University of Monastir, Monastir, Tunisia. [92]Cytogenetics, Molecular Genetics and Human Reproductive Biology Laboratory, Service of Genetics CHU Farhat Hached Sousse, Sousse, Tunisia. [93]Institute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University Hospital, Aachen, Germany. [94]Unité de Génétique Clinique, Centre de Compétences Maladies Rares Anomalies du Développement, CHRU de Pointe à Pitre, Guadeloupe, France. [95]Service de Génétique Médicale, CHU de Clermont-Ferrand, Clermont-Ferrand, France. [96]Service de Génétique Clinique, CHRU Nancy, Vandoeuvre les Nancy, France. [97]INSERM U1256-NGERE, Faculté de Médecine, Université de Lorraine, Vandoeuvre les Nancy, France. [98]Centro Nacional de Análisis Genómico (CNAG), Universitat de Barcelona (UB), Barcelona, Spain. [99]Service de Biochimie, Hôpital Pellegrin, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France. [100]Département de Génétique, Centre Hospitalier Universitaire de Saint-Etienne, Saint-Etienne, France. [101]Pathophysiology and Genetics of Neuron and Muscle (PGNM), UCBL, CNRS UMR5261—INSERM, U1315, Lyon, France. [102]Department of Endocrinology and Diabetology for Children, Assistance Publique-Hôpitaux de Paris (APHP), Paris, France. [103]Department of Adolescent Medicine, Bicetre Paris-Saclay University Hospital, Le Kremlin Bicetre, France. [104]Paris Saclay University, Paris, France. [105]Department of Clinical Genetics, Sydney Children's Hospitals Network Westmead, Sydney, New South Wales, Australia. [106]Specialty of Genomic Medicine, University of Sydney, Sydney, New South Wales, Australia. [107]University Children's Hospital, Salzburger Landesklinken (SALK) and Paracelsus Medical University, Salzburg, Austria. [108]Service de Génétique, CHU Caen, Caen, France. [109]Laboratoire de Cytogénétique, CH de Chambéry, Chambéry, France. [110]Health Technologies Institute, Inserm, Paris, France. [111]Department of Genetics and Personalized Medicine, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia. [112]Department of Clinical Genetics, Genetics and Personalized Medicine Clinic, Tartu University Hospital, Tartu, Estonia. [113]Assistance Publique-Hôpitaux de Paris (AP-HP), DMU INOV-RDB, Hôpital Robert Debré, Service de Neurologie Pédiatrique, Paris, France. [114]Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [115]Laboratoire de Génétique Médicale, CHR Metz-Thionville, Hôpital Mercy, Metz, France. [116]Department of Neurology, Harvard Medical School, Boston, MA, USA. [117]UF de Génétique clinique, CHU de Reims, Reims, France. [118]Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. [119]CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Madrid, Spain. [120]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain. [121]Unité de Génétique—Centre Hospitalier Universitaire de Guadeloupe, Guadeloupe, France. [122]Centre for Rare Diseases Erlangen, University Hospital Erlangen, Erlangen, Germany. [123]Center for Human Genetics, Cliniques universitaires Saint-Luc, Université Catholique de Louvain, Brussels, Belgium. [124]Reference Center for Rare Diseases and Intellectual Deficiencies of Rare Causes, Paris, France. [125]Paediatric Neurology Unit, Department of Pediatrics, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Barcelona, Spain. [126]LMU Klinikum—München, Munich, Germany. [127]Hospices Civils de Lyon, Hôpital Femme Mère Enfant, Service de Neuropédiatrie, Lyon, France. [128]Institute of Human Genetics, Magdeburg, Germany. [129]Department of Pediatrics, Centre for Neuromedicine, Gemeinschaftskrankenhaus Herdecke Gerhard-Kienle-Weg, Herdecke, Germany. [130]Department of Pediatrics, AMEOS Klinikum St. Clemens Oberhausen, Oberhausen, Germany. [131]Laboratoire de Diagnostic Génétique, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, Strasbourg, France. [132]UMRS 1112, INSERM, Université de Strasbourg, Strasbourg, France. [133]Department of Genetics, La Réunion University Hospital, Saint-Pierre, France. [134]Department of Pediatrics, McMaster Children's Hospital, McMaster University, Hamilton, Ontario, Canada. [135]Centre de Référence des Epilepsies Rares (CRéER), CHRU Nancy, Vandoeuvre les Nancy, France. [136]Service de Médecine Infantile, CHRU, Vandoeuvre les Nancy, France. [137]Génétique Médicale, Centre Hospitalier Universitaire; Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253, Tours, France. [138]INSERM UMR 1231, Génétique des Anomalies du Développement, Université de Bourgogne Franche-Comté, Dijon, France. [139]Department of Neurology, University of Leipzig Medical Center, Leipzig, Germany. [140]Centre de Référence Déficiences Intellectuelles de Causes Rares, Paris, France. [141]Service de Génétique Médicale, CHU de Montpellier, Institute for Neurosciences of Montpellier, University of Montpellier, INSERM, Montpellier, France. [142]Epilepsy Center Kleinwachau, Radeberg, Germany. [143]Department of Neurology, University Hospital Tübingen, Tübingen, Germany. [144]Service de Génétique, CRMR AnDDI-Rares, CHU Reims, Reims, France. [145]European Molecular Biology Laboratory, EMBL Grenoble, Grenoble, France. [146]Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada. [147]Division of Pediatric Neurology, Developmental Medicine and Social Pediatrics, Department of Pediatrics, Munich University Hospital, Munich, Germany. [148]Institute of Human Genetics, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. [149]Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany. [150]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [151]Univ. Rouen Normandie, Normandie Univ., Inserm U1245 and CHU Rouen, Department of Biostatistics and Reference Center for Developmental Abnormalities, Rouen, France. [152]These authors contributed equally: Caroline Nava, Benjamin Cogne, Amandine Santini. [153]These authors jointly supervised this work: Julien Thevenon, Christel Depienne. ✉e-mail: caroline.nava@aphp.fr; christel.depienne@uk-essen.de

Methods

Inclusion and ethics statement

This study complies with the ethical standards of each of the participating countries. Informed consent was obtained for all patients included in this study from their parents or legal guardians. A specific consent form was obtained from the families who consented to the publication of photographs. Patients/participants/samples were pseudonymized for the genetic study at each participating center. We collected information on the sex (but not gender) of the patients from the patients' clinical file. Grenoble-Alpes University Hospital (CHU Grenoble-Alpes, research 19814188) is the promoter of this research for the hospitals associated with the Auragen laboratory. Assistance Publique-Hôpitaux de Paris (AP-HP) is the promoter of this research for the hospitals associated with the SeqOIA laboratory (project ID: APHP241333). The study has received approval from the Ethics Committee of University Hospital Essen (reference 24-12010-BO) and approval from the Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations dans le domaine de la Santé (CESREES; reference 21082803 Bis/2038764). AP-HP has obtained authorization from the Commission Nationale de l'Informatique et des Libertés (reference HGTHGT/MFIMFI/AR2426865; request 924924336666) for the data processing activities related to this project. Part of the study has been approved by the CHU de Nantes Ethics Committee (number CCTIRS 14.556). Part of this research was ethically approved by CPP Ouest V (File 06/15) on 4 August 2015 (Ref MESR DC 2017 2987). For methylation analysis, DNA from all individuals (patients and controls) had been collected previously in the context of genetic analysis in a medical setting, following signature of a written, informed consent that includes a query on the use of leftovers in a research setting. Healthy controls consisted of individuals without NDD who underwent presymptomatic testing for other conditions and were found to be noncarriers or unaffected relatives of patients with a genetic disease among noncarriers of pathogenic variants. Samples used for the methylation study were stored within the genetics biological collection of the CRBi, Rouen, France, declared as DC 2008-711 (access authorization MCRBi/2024/02). The analysis of methylation profiles based on previously stored DNA in these conditions was approved by the CERDE ethics committee (notification E2023-13) from the Rouen University Hospital. Researchers and clinicians from participating centers contributing either data or intellectual input were involved at all stages of the study, from design and implementation to drafting and revising the manuscript, and are coauthors of the article.

List of snRNA genes and variant nomenclature

A list of 50 official gene symbols encoding functional snRNAs (Supplementary Table 3) was established from the HUGO gene nomenclature committee (https://www.genenames.org/). Information was retrieved from HGNC in December 2023 by applying the advanced filtering 'gd_locus_type = "RNA, small nuclear"' and restricting to genes with approved symbols. The coordinates (start and end positions) of genes and transcripts were in parallel retrieved from the National Institutes of Health Reference Sequence (curated subset downloaded from the University of California, Santa Cruz (UCSC) genome browser) and Ensembl. Of note, the start positions of transcripts from both entities differ for certain snRNA genes (for example, *RNU5F-1*), implying that variant nomenclature may vary depending on the transcript used to report them.

Patient cohorts

We initially identified the n.64_65insT variant in a single patient with developmental epileptic encephalopathy. This variant was prioritized because it was the strict de novo variant with the highest CADD score and was submitted to GeneMatcher[33]. Following the publication of the preprint by ref. 34 on 8 April 2024, we investigated variants in *RNU4-2* and 49 other snRNA genes in several diagnostic and research cohorts. Our inclusion criteria were as follows: (1) a de novo variant in any of the 50 snRNA-encoding genes with less than ten heterozygotes in gnomAD v4.1.0, or (2) a heterozygous variant in *RNU4-2* located within the critical 18-nucleotide region as defined in ref. 13. We then narrowed our search to *RNU5A-1* and *RNU5B-1* and investigated (3) de novo variants with less than ten heterozygotes in gnomAD v4.1.0 and/or (4) heterozygous *RNU5B-1* variants located in the U5 5′ loop I.

The main cohort is composed of 23,649 patients with rare disorders, including 15,073 patients with NDD and their parents, when available, who underwent genome sequencing as part of the diagnostic process in France (PFMG2025)[18] on one of the two national clinical sequencing laboratories, SeqOIA (https://laboratoire-seqoia.fr/) and Auragen (https://www.auragen.fr/) between 2019 and 2024. All de novo variants were visualized on IGV. The analysis of *RNU4-2* variants in this main cohort identified 80 patients. Furthermore, we collected data of 70 additional patients with de novo and/or pathogenic *RNU4-2* variants identified in either diagnostic or research contexts through national networks, established collaborations or GeneMatcher[33]. These additional cohorts included 42 patients from France, 20 from Germany, 5 from Canada, 1 from the Netherlands, 1 from Spain and 1 from the US. Thirty patients had genome sequencing, whereas in 40 patients, the variant was identified or confirmed by a targeted method—Sanger sequencing ($n = 35$) or next-generation sequencing of amplicons ($n = 5$). Among the patients diagnosed by Sanger sequencing, two had previously inconclusive exome analyses and were included in SOLVE-RD. Reads supporting the presence of n.64_65insT were identified in the exome data. None of the patients included in this study had been previously published, and we also checked that there were no duplicates for individuals with the same variant based on the individual's year of birth and initials.

The analysis of de novo variants in the other 49 snRNA genes in the PFMG cohort identified 36 patients with de novo variants in 17 genes (Supplementary Table 4 and Supplementary Note). A targeted search for variants in *RNU5B-1* and *RNU5A-1* in the GEL dataset (including both the 100,000 Genomes cohort (v18) and NHS-GMS (v3) cohort) identified five additional individuals with rare (<10 occurrences in gnomAD) de novo variants, five of 8,841 undiagnosed NDD probands and one of 21,816 non-NDD probands. In addition, three probands analyzed in duo had a rare variant located in the U5 5′ loop I absent from the single parent analyzed. In addition, five de novo variants in *RNU5B-1* were collected from the Broad Centre for Mendelian Genomics, UDN-Aus, the BCH Epilepsy Genetics Program and Care4Rare Canada.

Variants were reviewed using Alamut Visual Plus v1.11 (Sophia Genetics) and MobiDetails[35] (https://mobidetails.iurc.montp.inserm.fr/MD/).

Sanger sequencing

Sanger targeted sequencing was performed to screen for variants in *RNU4-2* and/or to perform segregation analysis. PCR amplification of *RNU4-2* was performed using the HotStarTaq Master Mix Kit (Qiagen, 203445) with the following primers: forward, 5′-AAATACGGCTGGTG GAGTGG-3′; reverse, 5′-TCACAGTACCCGCACAGAAC-3′, according to the manufacturer's instructions. Forward and reverse sequencing reactions were performed using the BrilliantDye Terminator v1.1 Cycle Sequencing Kit (Nimagen, BRD1-1000) or the BigDye Terminator v3.1 Sequencing Kit (Life Technologies, 4337457). ExoSAP-Purified sequencing products (ExoSAP-IT; Applied Biosystems, 78205) were run on Pop-7 polymer (Life Technologies, 4335615) using an ABI 3730 or 3730XL automated sequencer (Applied Biosystems). Sequences were analyzed using Geneious Prime 2019 (Biomatters) or Seqscape v2.6 software (Applied Biosystems).

Variant classification

We classified variants according to the ACMG/AMP criteria[19] using recommendations from ref. 20. The PS2 (or PM6 for patients who underwent targeted sequencing) criteria were applied for cases with

de novo inheritance. 'PM2 supporting' was applied for variants absent or very rare in gnomAD v4.1.0; PM1 for variants located in mutational hotspots—chr12(hg38): 120,291,825–120,291,842 for *RNU4-2* and chr15(hg38): 65,304,713–64,304,720 for *RNU5B-1*. We applied 'PS4 supporting' for variants identified in at least three patients, and 'PS4 moderate' for those found in at least six patients, either in this study or in ref. 13. PS4 was only applied for n.64_65insT. Finally, PS3 was applied when RNA-seq and/or methylation analyses supported pathogenicity.

Clinical data analyses

Clinical data were retrospectively collected from the referring physician using an anonymized Excel sheet. For patients aged 0–3 years, sitting and walking items were noted as 'too young' unless the clinician specifically noted their achievement. ID was noted as 'too young' unless the clinician assessed it as severe. Autism spectrum disorders were also noted as 'too young' unless the clinician could confirm or rule out the diagnosis. Categorical data for 44 selected clinical features from 143 patients with P and LP *RNU4-2* variants and 12 patients with *RNU5A-1* or *RNU5B-1* variants were converted to a 0–1 scale, with 0 representing a more favorable phenotype presentation and 1 representing a more severe phenotype. Hierarchical clustering was performed using the pheatmap R package, performing *z* score scaling for each row (across different patients), and ward.D2 clustering method keeping missing values. PCA was generated after replacing missing data with 0 and performing variable scaling. Microcephaly was defined as HC measurements less than the third percentile. We used charts established in ref. 36 to calculate the HC percentile at birth and define congenital microcephaly. Corresponding plots were generated with the 'Plotter: Preterm growth charts, 22–50 weeks' from the Canadian Pediatric Endocrine Group (https://cpeg-gcep.shinyapps.io/prem2013/). For additional HC measurements, reference chart data points were obtained from ref. 37. Male patients older than 21 years were plotted at age 21, and female patients older than 20 years were plotted at age 20, corresponding to the maxima for each sex. Fisher's tests (two-sided; 2 × 2, 2 × 3 or 2 × 4 contingency tables) adjusted for multiple comparisons using Bonferroni correction were used to compare clinical features in different U4 domains (n.64_65insT versus n.76C>T and n.64_65insT versus the other variant groups) for 41 clinical features.

Conservation and in silico predictions

The highest homologs to the human *RNU4-2* and *RNU5B-1* were obtained for *Ciona intestinalis*, *Ciona savignyi*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio* and *Mus musculus* by using BLAT on each of these genomes in Ensembl Release 112 (ref. 38). RNA sequences from *RNU4-2* and *RNU5B-1* were aligned to (1) their respective sequence homologs and (2) the sequence(s) of other U4- and U5-encoding genes expressed in the brain using Geneious Prime 2019 (Biomatters). The threshold for consensus was set to 100% identity, highlighting positions with 100% agreement between all sequences.

CADD PHRED scores and conservation in vertebrates (verphyloP) were calculated for P and LP patient variants and gnomAD v4.1.0 variants with CADD (v1.7)[39]. For each variant, in silico-mutated U4 RNA sequences were generated with seqkit mutate[40]. Bifold[41] was used to generate the multiple U4:U6 interactions and calculate the minimum free energy. Comparisons were performed by applying the Mann–Whitney *U* test, two-sided.

Expression of snRNAs in brain tissues

We used small RNA data for different human embryonic brain regions to inspect the expression level of selected snRNAs. These data were generated by the ENCODE Consortium[42]—diencephalon (GSE78292), temporal lobe (GSE78303), occipital lobe (GSE78298), frontal cortex (GSE78293), parietal lobe (GSE78299) and cerebellum (GSE78291). Tracks show unique read signals for plus and minus strands from the default anisogenic replicate. Expression of these genes in the brain using BrainVar was previously investigated[13].

RNA-seq

Peripheral blood mononuclear cells were isolated from 2 to 4 ml of EDTA-anticoagulated blood within 48 h of collection using UNISEP+ tubes (Eurobio Scientific, U-04). Cells were cultured in six-well plates (5.0×10^5 to 2.0×10^6 cells per well) in lymphocyte-stimulating medium (chromosome medium P; Euroclone, EKAMTB100) for 48–72 h at 37 °C (5% CO_2). After incubation, one well per sample was treated for 4–5 h with 1 mg ml^{-1} puromycin (Invivogen, ant-pr-5), an indirect NMD inhibitor. RNA was extracted using the NucleoSpin RNA Plus extraction kit (Macherey-Nagel, 740984.50) according to the manufacturer's instructions.

Stranded RNA-seq libraries were prepared from 100 ng of total RNA on the Magnis NGS Prep System (Agilent) using the SureSelect XT-HS2 kit (Human All Exon V8 capture probes, G9774C) with 12 and 10 PCR cycles for precapture and postcapture amplifications, respectively. RNA-seq was sequenced on an Illumina's NextSeq 550 (16 samples on HighOutput 2 × 75 bp) to obtain 25–30 million paired-end reads per sample.

Fastq files were aligned on the GRCh38 reference genome with STAR (v.2.7.11a) in two-pass mode using Ensemble transcripts (v.106). Quality control was performed with fastqc (v.0.11.3) and fastp (v.0.23.4). CIBERSORTx (v1.0) was used to estimate the relative abundance of blood cells using the LM22 signature matrix file[43]. One *RNU4-2* sample was removed because of a low proportion of activated T $CD4^+$ cells (1/38; Supplementary Fig. 5a). To generate the *RNU4-2* splicing signature, we used lymphocytes from 19 *RNU4-2* samples and 21 controls not treated by puromycin by using rMATS-turbo (v.4.3.0)[21] with the following parameters: -t paired –anchorLength 1 –libType fr-firststrand–novelSS –variable-read-length –allow-clipping. Controls were matched on the following criteria: library preparation kit, sequencing flow cell and culture time. Python scripts were used to filter rmats output files with the following filters: mean coverage >7, false discovery rate (FDR) < 0.1, deltaPSI > 0.05. PCA was performed using the sklearn Python library using PSI values from significant alternatively spliced exons, keeping only (1) the most significant call when several were called impacting the same exon and (2) events affecting genes with approved HGNC symbols. The significant calls from the 5′SS signature were used to perform an additional PCA with three testing samples (n.45_46insT, n.62C>T and n.64_65insT). The study of NMD impact was performed by comparing nine patients with *RNU4-2* P/LP variants (five n.64_65insT, n.67A>G, n.70T>C, two n.76C>T) against 22 controls, all treated with puromycin. The splicing study of the *RNU5B-1* (two n.39C>G and two n.44A>G) and *RNU5A-1* (n.40_41insA and n.39del) variants was performed by comparing each variant to the same 21 controls as for *RNU4-2* using the same rMATS-turbo parameters except for the singletons variants n.40_41insA and n.39del for which an FDR threshold of 0.01 was used. To ensure the specificity of the signal, we included additional 20 controls, 19 *RNU4-2* and other U5 variants, not involved in the statistical analysis, for PCA visualization. Raw spliceAI scores were obtained from MobiDetails[35,44]. Sashimi plots were made using rmats2sahimi and boxplots with seaborn (v0.13.2). Consensus nucleotide sequences were generated using Logomaker (v0.8). Scripts used for RNA-seq analysis are available on Zenodo using the following link: https://doi.org/10.5281/zenodo.13868501 (ref. 45).

Epigenome-wide analysis and DNA methylation signature

Genomic DNA was extracted from whole blood and subjected to bisulfite conversion. DNA methylation profile was then derived using Infinium MethylationEPIC v2.0 BeadChips (Illumina, 20087708), in accordance with the manufacturer's protocol. Patients and negative controls were balanced across 24 arrays and within each array row to reduce technical biases. DNA methylation arrays were generated at the

ASGARD-Rouen genomic platform (University of Rouen and Rouen University Hospital) on an Illumina NextSeq 550 scanner. Raw IDAT data were processed and normalized using the default Meffil R package protocol along with all other samples included in the 24 arrays to better estimate the variability of methylation signals within and across arrays[46]. One *RNU4-2* sample failed default quality controls and was excluded from further steps. The remaining samples were functionally normalized together as advocated in the Meffil documentation, with random effect adjustment on array and sentrix row as well as fixed effect adjustment on the first two PCs, before computing β values.

Several predictions were obtained from methylation values to apply additional quality control and normalization steps. Sex predictions were extracted from the standard Meffil normalized object. No inconsistencies between reported and predicted sex were noted (Supplementary Fig. 6a). Blood cell counts were estimated with the meffil.cell.count.estimates function. PCA of predicted blood cell counts showed a good overlap of positive and negative controls in terms of overall blood cell composition (Supplementary Fig. 6b). DNA methylation age was predicted with the DNAmAge function from the methylclock R package[47]. The Horvath and skinHorvath clocks both displayed a very strong correlation with actual age at blood sample on our dataset (Pearson correlation $r = 0.97$; Supplementary Fig. 6c,d).

The set of differentially methylated probes was identified with the meffil.ewas function on the subset of controls and P or LP *RNU4-2* variant carriers. To correct for well-known confounders, the differential analysis accounted for skinHorvath age, sex and predicted blood cell composition. Manhattan and volcano plots are given in Extended Data Fig. 9. After filtering the P value at 10^{-7} (Bonferroni-corrected threshold for an effective number of approximately 500,000 independent tests) and the average methylation difference between positive and negative controls at 0.05, adjusted methylation levels were visualized through PCA and heatmap representations (pheatmap package with the Euclidean distance and Ward aggregation method). Namely, a baseline methylation level model adjusting for skinHorvath age, sex and cell blood composition was fitted on negative control samples for each probe. Adjusted methylation levels were computed for each sample from this model by correcting each β value for the expected baseline level according to this model. Phenotype classification into mild/moderate and severe subtypes was derived independently from the episignature discovery and a posteriori added to these graphical representations.

Finally, the robustness of the signature was challenged through fivefold cross-validation. The dataset was split into five random and equal-sized blocks. Each block was used in turn as a validation set, while the remaining four blocks were used as a training set to run a new differential analysis based on controls and moderate-to-severe phenotypes. An SVM model was trained on each cross-validation training set and applied to the test set to derive unbiased sensitivity and specificity estimations overall, by phenotype class and variant type, along with 95% binomial CIs. For *RNU5A-1* variants, the pathogenicity score was derived from a prediction model based on the complete training set.

Combined analysis of *KMT2D*, *KMT2A* and *RNU4-2* signatures was done similarly to the main *RNU4-2* analysis. Epic v1 and Epic v2 samples were imported and normalized separately with standard Meffil functions. Baseline methylation models were fitted separately on Epic v1 and Epic v2 positive and negative control samples. Adjusted methylation profiles of CpG positions belonging to the union of published *KMT2D* and *KMT2A* signatures, as well as the *RNU4-2* signature, were then combined and represented on a heatmap using the pheatmap package with Euclidean distance and Ward aggregation method for columns.

Variant impact

Structural analysis of variants and corresponding figures was performed using the PyMol v3.0.0 visualization software[48] on published coordinates of the human tri-snRNP structure—Protein Data Bank (PDB) IDs: 6QW6 (ref. 4) and 8Q7N ref. 49.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Variant details have been submitted to ClinVar (P/LP variants: SUB15052639, https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB15052639; VUS: SUB15154590, https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB15154590). RNA-seq data and methylation have been deposited in the European Genome–Phenome Archive (EGA, http://www.ebi.ac.uk/ega), which is hosted at the EBI. RNA-seq data are available under the study accession EGAS50000000889 (https://ega-archive.org/studies/EGAS50000000889). Methylation data are accessible under the study accession EGAS00001008070 (https://ega-archive.org/studies/EGAS00001008070). Both are subjected to a data processing agreement. Requests will be evaluated by a data accessibility committee to ensure that data access complies with ethical and legal standards respective to the corresponding projects. Individual genome data could not be made publicly available due to ethical considerations. Controlled access to human genome data is necessary to protect the privacy of participants and to ensure compliance with ethical and legal standards, particularly the General Data Protection Regulation in Europe. Data access to individual genome data from the PFMG2025 with other researchers is submitted to current data protection and regulations in France and is only possible through the Collecteur Analyseur de Données (CAD). More information on data access and the CAD structure can be obtained on the PFMG2025 website (https://pfmg2025.fr/le-plan/collecteur-analyseur-de-donnees-cad/). Research on de-identified patient data from the GEL 100,000 Genomes Project and NHS-GMS dataset is possible in the GEL Research Environment under a collaborative agreement. More information on data access can be obtained by email (research-network@genomicsengland.co.uk) or on the GEL website (https://www.genomicsengland.co.uk/research). All other data supporting the findings described in this paper are available in the article, Supplementary Tables 1–23, Extended Data Figs. 1–10 or Supplementary Information. We also used data from Ensembl Release 112 and data from the ENCODE Consortium: bigwig files with the plus/minus strand signals of unique reads from the default anisogenic replicate from the following tissues: diencephalon (https://www.encodeproject.org/experiments/ENCSR000AFR/), parietal lobe (https://www.encodeproject.org/experiments/ENCSR000AFY/), occipital lobe (https://www.encodeproject.org/experiments/ENCSR000AFX/), frontal cortex (https://www.encodeproject.org/experiments/ENCSR000AFS/), temporal lobe (https://www.encodeproject.org/experiments/ENCSR000AGD/) and cerebellum (https://www.encodeproject.org/experiments/ENCSR000AFQ/).

Code availability

The analyses were conducted using existing software and packages, including STAR aligner (v.2.7.11a), CIBERSORTx, rMATS (v.4.3.0), rmats2sahimi, seaborn, Logomaker, Meffil R package, methyclock R package, ggplot2 (v3.3.6), pheatmap (v1.0.12), stats (v4.2.0), factoextra (v1.0.7) and PyMol (v3.0.0). Free energy for RNA secondary structure was calculated using bifold from RNAstructure (v6.5; https://rna.urmc.rochester.edu/RNAstructure.html). Preterm HC graphs were plotted using the following link: https://cpeg-gcep.shinyapps.io/prem2013/. Custom scripts used for RNA-seq analysis are available on GitHub and Zenodo using the following links: https://github.com/benjamin-cogne/RNU4-2_transcriptomics/tree/v2.0 and https://doi.org/10.5281/zenodo.13868501, respectively.

References

33. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
34. Chen, Y. et al. De novo variants in the non-coding spliceosomal snRNA gene *RNU4-2* are a frequent cause of syndromic neurodevelopmental disorders. Preprint at *medRxiv* https://doi.org/10.1101/2024.04.07.24305438 (2024).
35. Baux, D. et al. MobiDetails: online DNA variants interpretation. *Eur. J. Hum. Genet.* **29**, 356–360 (2021).
36. Fenton, T. R. & Kim, J. H. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr.* **13**, 59 (2013).
37. Rollins, J. D., Collins, J. S. & Holden, K. R. United States head circumference growth reference charts: birth to 21 years. *J. Pediatr.* **156**, 907–913 (2010).
38. Harrison, P. W. et al. Ensembl 2024. *Nucleic Acids Res.* **52**, D891–D899 (2024).
39. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
40. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
41. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
42. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
43. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
44. De Sainte Agathe, J. M. et al. SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. *Hum. Genomics* **17**, 7 (2023).
45. benjamin-cogne. benjamin-cogne/RNU4-2_transcriptomics: splicing signature for RNU4-2, RNU5B-1 and RNU5A-1 (v2.0). *Zenodo* https://doi.org/10.5281/zenodo.13868501 (2025).
46. Min, J. L., Hemani, G., Davey Smith, G., Relton, C. & Suderman, M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989 (2018).
47. Pelegi-Siso, D., de Prado, P., Ronkainen, J., Bustamante, M. & Gonzalez, J. R. methylclock: a Bioconductor package to estimate DNA methylation age. *Bioinformatics* **37**, 1759–1760 (2021).
48. Schrödinger, L. & DeLano, W. PyMOL. www.pymol.org/pymol (2020).
49. Zhang, Z. et al. Cryo-EM analyses of dimerized spliceosomes provide new insights into the functions of B complex proteins. *EMBO J.* **43**, 1065–1088 (2024).

Acknowledgements

We thank the patients and their families for their participation in this study. Patients included in this study were diagnosed through multiple studies, including PFMG2025; DEFIDIAG, a study sponsored by INSERM and supported by the French Ministry of Health in the framework of French initiative for genomic medicine (PFMG2025); Solve-RD, a project that has received funding from the European Union's Horizon 2020 research and innovation program (grant 779257); Deutsche Forschungsgemeinschaft (DFG; projects 458099954 (to C. Depienne) and 37144118 (to R.A.J. and S. Neuser)); Inserm as part of the 2022 MESSIDORE program (project Inserm-MESSIDORE 19), and the European Joint Programme on Rare Diseases GENOMIT I6478-B by the Austrian Science Fund FWF and by the Centre for Population Genomics (Garvan Institute of Medical Research and Murdoch Children's Research Institute), which was funded in part by a Medical Research Future Fund (MRFF) Genomics Health Futures Mission grant (2008820). Part of the results have been supported by the RNU-SPLICE project, financed by the health philanthropic program of Mutuelles AXA, dedicated to supporting innovative research projects in France (to C.N.). Part of this study was supported by the 'Priority Research Programme on Rare Diseases' of the French Investments for the Future Programme, project MultiOmixCare. The structural interpretation was conducted by C. Charenton and A.J. as part of the ERC Starting Grant project 'SPLIFEM'. A.S. received a grant from Region Normandie and GIRCI Nord-Ouest (FHU-A2M2P). This research was made possible through access to data in the National Genomic Research Library, which is managed by GEL (a wholly owned company of the Department of Health and Social Care). We thank P. O'Donovan, M. Sato and Z. Mustafa from GEL for their help with Airlock requests. The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. Cases identified by the Broad Center for Mendelian Genomics (to A.O'D.-L. and S.L.S.) were supported by the National Human Genome Research Institute (grants UM1HG008900 and U01HG0011755). We thank the Care4Rare Canada Consortium for their support of a subset of this work. The Care4Rare Canada Consortium is funded by Genome Canada and the Ontario Genomics Institute (OGI-147), the Canadian Institutes of Health Research, Ontario Research Fund, Genome Alberta, Genome British Columbia, Genome Quebec and the Children's Hospital of Eastern Ontario Foundation. K.M.B. is supported by a CIHR Foundation (grant FDN-154279) and a Tier 1 Canada Research Chair in Rare Disease Precision Health. G.D.G. is supported by a CIHR Fellowship award (MFE-491710). A.M.I., J.L.L. and R.L. obtained support for sequencing through Translational Implementation of Genomics for Rare Disease, funded by Genome Canada, Genome Alberta, Alberta Innovates and the Alberta Children's Hospital Foundation. The research at Boston Children's Hospital was supported by the BCH Children's Rare Disease Collaborative and the Robinson Fund for Transformative Research in Epilepsy, and A.M.D'G. was supported by a BCH Office of Faculty Development/Basic & Clinical Translational Research Executive Committees Faculty Career Development Fellowship. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. The UDN-Aus acknowledges financial support from the Australian Government's MRFF (grant MRF2007567). We thank H. Thomson and E. Martin for their contributions to patient data collection in UDN-Aus. Genome sequencing in Spanish patients was supported by the Undiagnosed Rare Diseases Program of Catalonia (PERIS SLT002/16/00174) from the Autonomous Government of Catalonia, the Biomedical Research Networking Center on Rare Diseases (CIBERER, ACCI19-759), the Hesperia Foundation (Royal House of Spain), and La Marató de TV3 Foundation with project 202006-30 and IMPACT-Genomica (IMP/00009) to A.P. T.B.H. received funding from the German Research Foundation (DFG; research grants 418081722, 433158657 and EJP-RD Artemis—542553983) and the European Commission (Recon4IMD—GAP-101080997). We thank the CERCA Program/Generalitat de Catalunya for institutional support. We thank M.D. Biase, A. Wilson, A. Rehman, A. Thomas-Wilson, S. Phadke and A. Vabhyankar from the New York Genome Center for their help with genome analysis and for providing anonymized patient data. We thank A. Lavillaureix, C. Quélin, F. Prieur and F. Demurger for their contribution to the establishment of *KTM2A*/*KTM2D* episignatures. This study uses resources generated by the ENCODE Consortium and the T. Gingeras lab (CSHL). N.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society

(grant 220134/Z/20/Z), a research prize from the Lister Institute. Y. Chen is supported by a studentship from Novo Nordisk. P.Z.'s work was supported by Stiftung Michael through the assistance of the Canger-Janz Fellowship. T.S.B. was supported by ZonMw Vidi (grant 09150172110002) and acknowledges support from Stichting 12q. K.Õ. was supported by the Estonian Research Council (grants PRG471 and PRG2040). S.L.S. was supported by a Manton Center for Orphan Disease Research postdoctoral fellowship. S. Baulac received a grant from la Fondation pour la Recherche médicale (EQU202203014861). The Chair in Genomic Medicine awarded to J. Christodoulou is generously supported by The Royal Children's Hospital Foundation. A. Kuechler, F.J.K., A.R. and G. Vasileiou from Germany, T.S.B. from the Netherlands, K.Õ. from Estonia, A.-S.D.-P., L.F., G.L., J.V.-G., M. Willems and S.O. from France and N.R. from Belgium are members of the European Reference Network (ERN) ITHACA, whose coordinator is A. Verloes. G.L., N.C., L. Januel and Z.G.-S. are members of ERN EpiCARE. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute. Open access was supported by AP-HP.

Author contributions

C.N. and J.T. performed genome analyses in the PFMG cohort. B.C., A.S., E. Leitão and F. Lecoquierre developed bioinformatics pipelines and performed data analyses. Y. Chen and N.W. performed data analyses in the GEL/NHS cohorts. S.L.S. and A.O'D.-L. contributed data from other cohorts. C. Charenton and A.J. performed structural analysis. B.C. and T. Besnard performed RNA-seq experiments. A.S. and C. Charbonnier performed the methylation (episignature) analysis. C.N., J.T. and C. Depienne conceived the study, performed data analyses and supervised the project. B.C., A.S., C. Charenton and E. Leitão made the final figures. S.H., S. Baer, A.J., S. Neuser, B. Keren, A.F., S. Forlani, M. Faoucher, K.U., K.P., A.A., J.-L.A., S.A., C.A., B. Aral, B. Arveiler, T.A.-B., M.A.M., G. Banneau, T.S.B., G. Barcia, S. Baulac, C.B., F.B., V.B., S. Bézieau, D.B., M.-N.B.-D., S. Boussion, O.B., E.B.-B., S.J.B., J.B., T. Busa, A. Caliebe, Y. Capri, K.C., R.C., C. Cenni, P. Chambon, P. Charles, J. Christodoulou, C. Colson, S.C., A. Cospain, J. Coursimault, T. Courtin, M.C., C. Coutton, I.C., A.M.D'G., B.D., J.-M.d.S.A., G.D.G., A.D.-D., J.D., A.-S.D.-P., A.D.-C., L.D.S.F., M.D.-F., S. Drukewitz, V.D., C. Dubourg, Y.D., D.D., S.E.C., M.E., L.F., S. Fennelly, H.F., M. Fradin, C.G.V., B. Ganne, J.G., H.G., Z.G.-S., A. Goldenberg, R.G.R., S. Gorokhova, L.G., V.G., M.G., J.M.G., B. Greiten, P.G., A.-M.G., S. Guha, A. Guimier, T.B.H., H.H.A., Y.H., R.H., M.H., J.H., B.H., M.-P.H., A.M.I., V. Jadas, L. Januel, N.J.-M., V. Jobanputra, F.J., L. Jornea, C.J., S. Julia, F.J.K., D.K., S.K., P.K., B. Khadija, F. Kilpert, C.K., F. Kraft, I.K., M. Lackmy, F. Laffargue, L. Lambert, R. Lamont, V.L., S. Laurie, J.L.L., L. Lebreton, M. Lebrun, M. Legendre, E. Leguern, D.L., E. Lejeune, G.L., M.L.-S., J. Levy, A. Linglart, S. Lyonnet, K.L., A.S.M., C. Mach, J.-L.M., L.M.-H., J. Marcadier, V. Marin, H.M., V. Marquet, A. May, J.A.M., C. Meridda, V. Michaud, C. Michot, G. Nadeau, S. Naudion, L.N., M.N., F.N., S.O., V.O., I.A.O.-O., M.O., K.Õ., L. Pasquier, S.P., M. Pauly, O.P., M. Pensec, L.P.-S., F.P., C. Philippe, M. Planes, A. Poduri, C. Poirsier, A. Pouzet, B.P., C. Prouteau, A. Pujol, C. Racine, M. Rama, F.R., K.R., M. Raway, A.R., M. Renaud, N.R., A.-C.R., L.R.-N., R.R., D.R., A.R.-P., S.R., A.R.-U., C.R.J., H. Safraou, V.S., P.S.-V., C. Sauvestre, E. Schaefer, W.S., I.S., J.-U.S., A.S.M., C.S.-B., S. Schuhmann, C. Schröder, M. Sebastin, S. Sigaudy, M. Spielmann, M. Spodenkiewicz, L.S.C., J.S., R.S., H. Surowy, M.A.T., C. Todosi, A.T., F.T.M.-T., A.U., J.V.-G., C. Vanlerberghe, G. Vasileiou, G. Vera, A. Verdel, A. Verloes, Y.V., C. Vignal, M.V., C.V.-D., A.V.-D., A.V., S.W., M. Willems, K.Z.-K., P.Z., L.Z., A.Z., W.P.G., H.D., C. Thauvin, K.M.B., P.M., A. Lermine, V. Malan, M. Rio, A. Kuechler, B.I., S. Drunat, T. Smol, N.C., A. Piton, G. Nicolas, M. Wagner, R.A.J., D.H., C. Mignot and P.B. contributed molecular data, clinical data and/or intellectual content. C.N., B.C., A.S., C. Charbonnier, C. Charenton and C. Depienne drafted the paper. All authors reviewed and approved the final paper.

Competing interests

N.W. receives research funding from Novo Nordisk and has consulted for ArgoBio Studio. A.O'D.-L. is on the scientific advisory board for Congenica, was a paid consultant for Tome Biosciences and Ono Pharma USA, and at present for Addition Therapeutics, and received reagents from PacBio to support rare disease research. V. Jobanputra has served as a consultant to Illumina and received consulting fees from the company. All other authors declare no competing interests.

Additional information

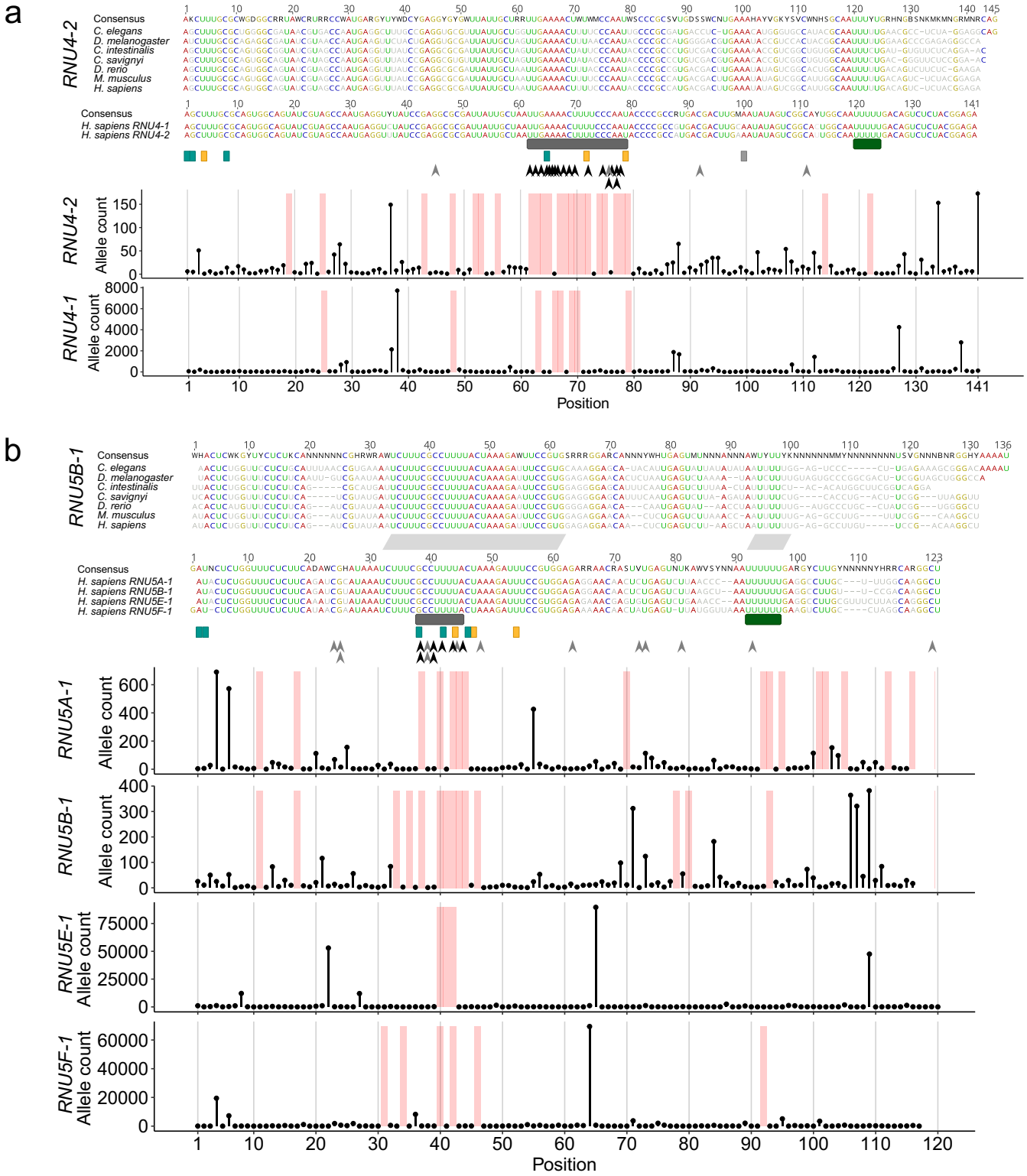
Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02184-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02184-4.

Correspondence and requests for materials should be addressed to Caroline Nava or Christel Depienne.

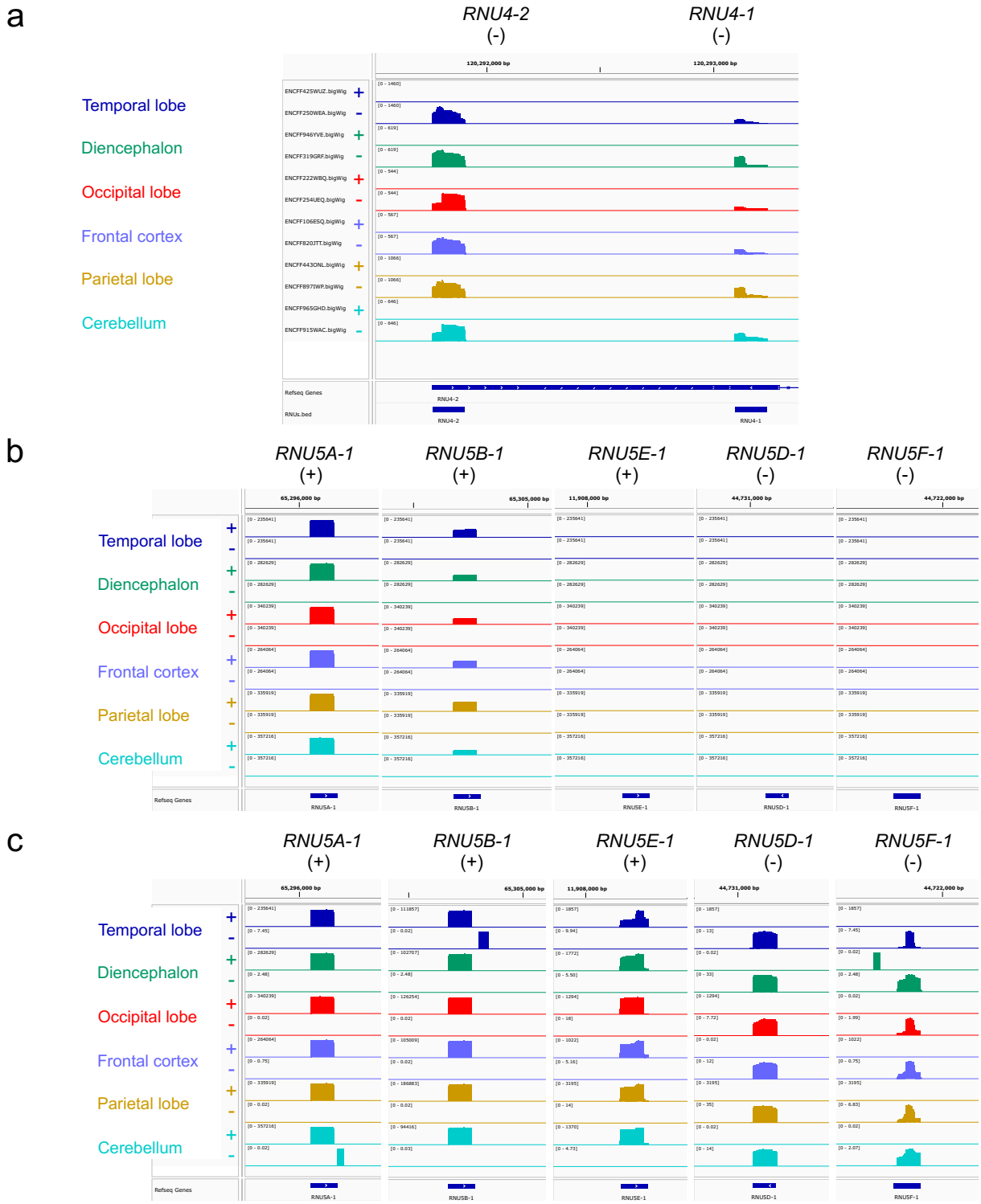
Peer review information *Nature Genetics* thanks Dong Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-025-02184-4



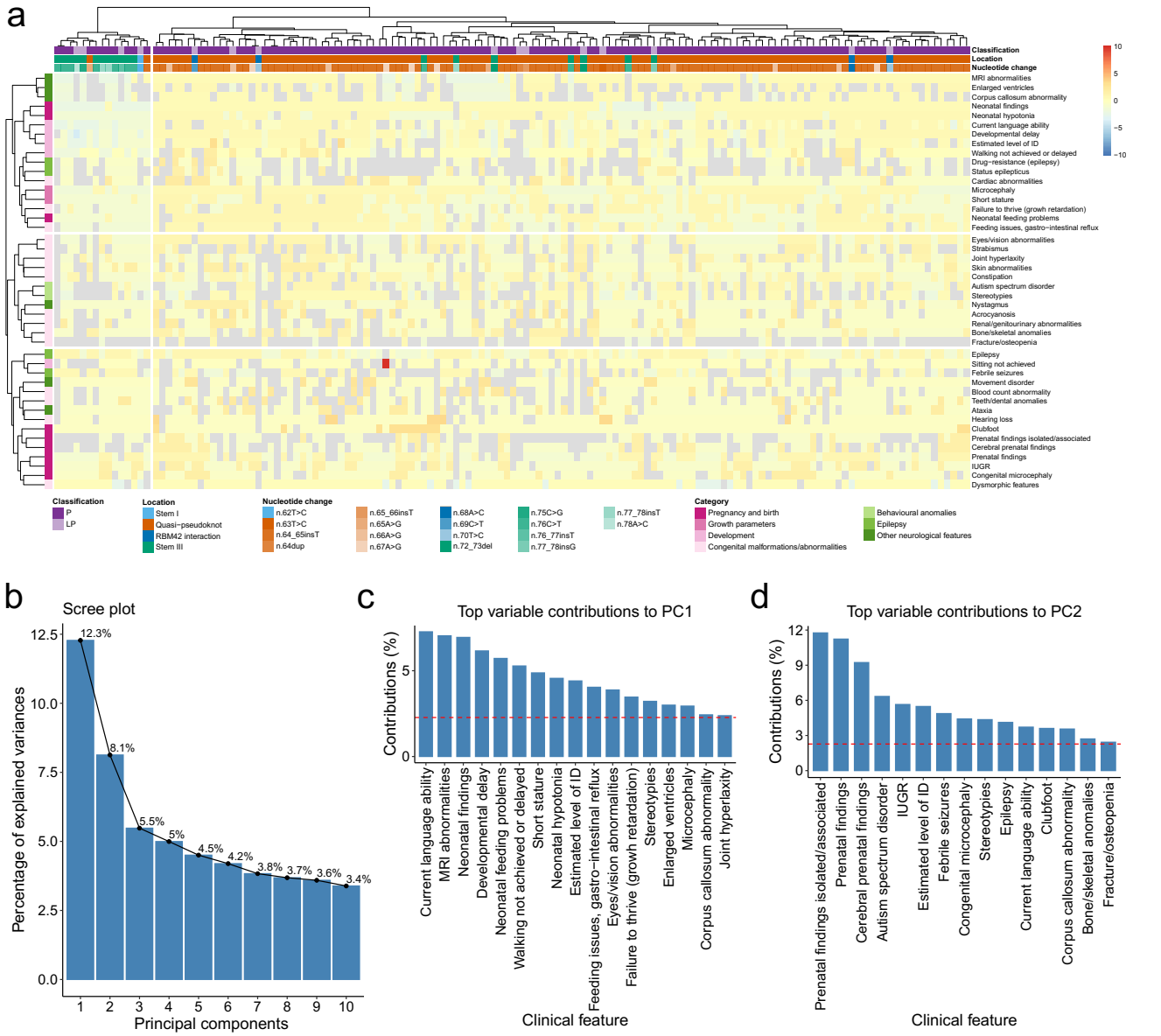
Extended Data Fig. 1 | Conservation and constraints of genes encoding U4 and U5. a, Correspondence between the alignment of *RNU4-2* sequences from distinct species (top), the alignment of human *RNU4-2* and *RNU4-1* (middle) and the allele counts from *RNU4-2* and *RNU4-1* variants in gnomAD v4.1.0 (bottom). The 18-bp critical region from ref. 13 is highlighted in gray and the Sm site in dark green. **b**, Correspondence between the alignment of *RNU5B-1* sequences from distinct species (top), the alignment of human *RNU5A-1*, *RNU5B-1*, *RNU5E-1* and *RNU5F-1* (middle) and their allele counts in gnomAD v4.1.0 (bottom). The 5′ loop I and Sm regions are highlighted in gray and dark green, respectively. The threshold for consensus is 100% identity. Nucleotides in red, blue, yellow and green are shown only for positions with 100% agreement between all sequences. Other nucleotides are in black (consensus, also using IUPAC codes) or gray (sequences). Arrowheads indicate variants from this study (pathogenic and likely pathogenic in black, variants of uncertain significance (VUS) in gray). Pseudouridine (yellow), 2′-O-methyl residues (teal); N6-methyladenosine (gray). Regions without variants in gnomAD v4.1.0 are shaded light red.

https://doi.org/10.1038/s41588-025-02184-4



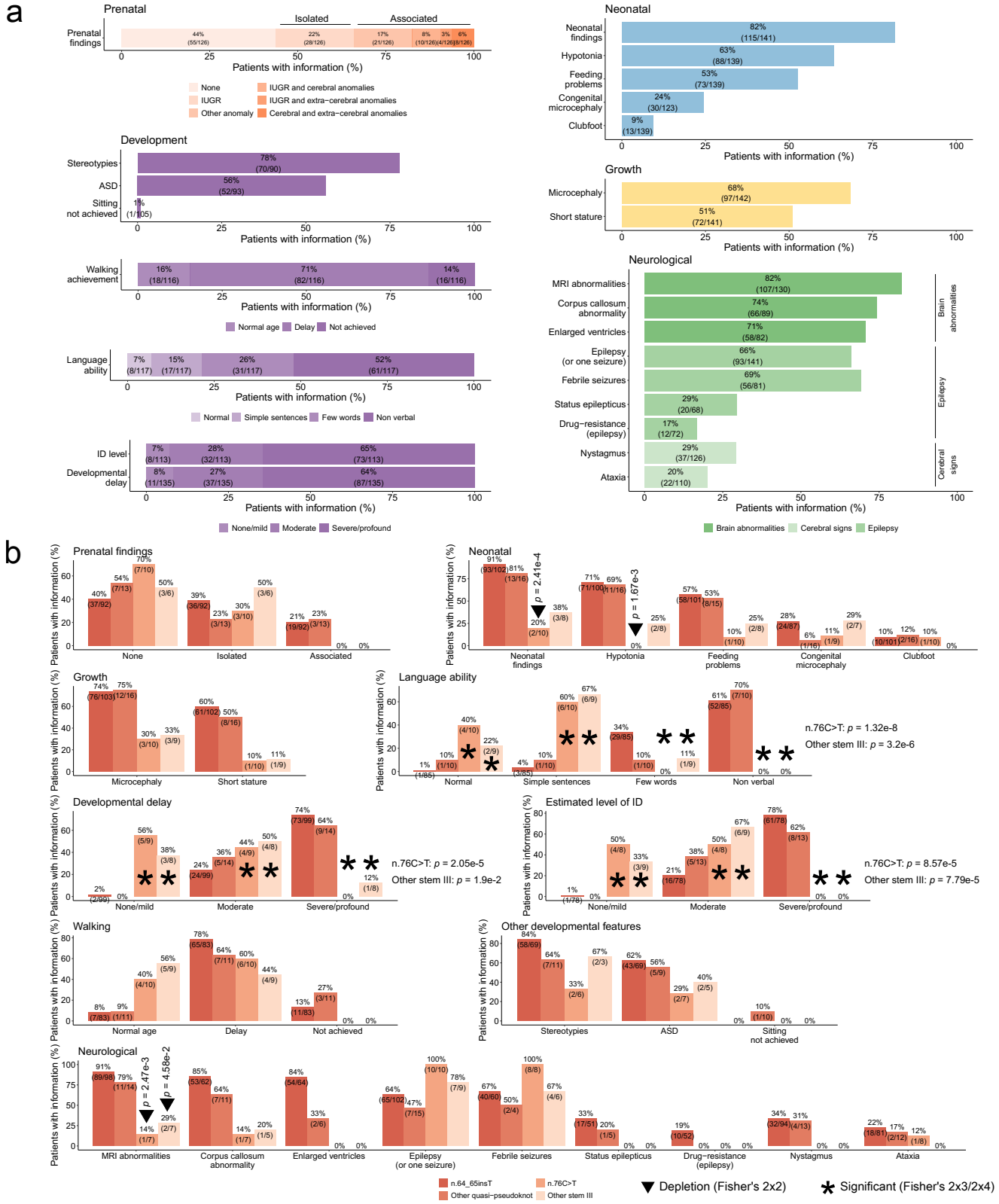
Extended Data Fig. 2 | Expression of the different genes encoding U4 and U5 genes in multiple brain regions. a, *RNU4-2* is more highly expressed than *RNU4-1* in all brain regions studied. Minus strand tracks (−) were auto-scaled, and each of their maximum was set to the plus strand (+). **b**,**c**, *RNU5A-1* and *RNU5B-1* are both highly expressed in the brain, while *RNU5E-1* is much less expressed. The expression of *RNU5D-1* and *RNU5F-1* in the brain is negligible. **b**, Plus strand track maximum at *RNU5A-1* from each tissue was set to the minus strand and kept for all genes. **c**, Auto-scale was allowed for each tissue and gene. Small RNA data were generated by the ENCODE Consortium for different human embryonic brain regions[38]—diencephalon (GSE78292), temporal lobe (GSE78303), occipital lobe (GSE78298), frontal cortex (GSE78293), parietal lobe (GSE78299) and cerebellum (GSE78291). Tracks show unique read signals for plus and minus strands from the default anisogenic replicate.

https://doi.org/10.1038/s41588-025-02184-4



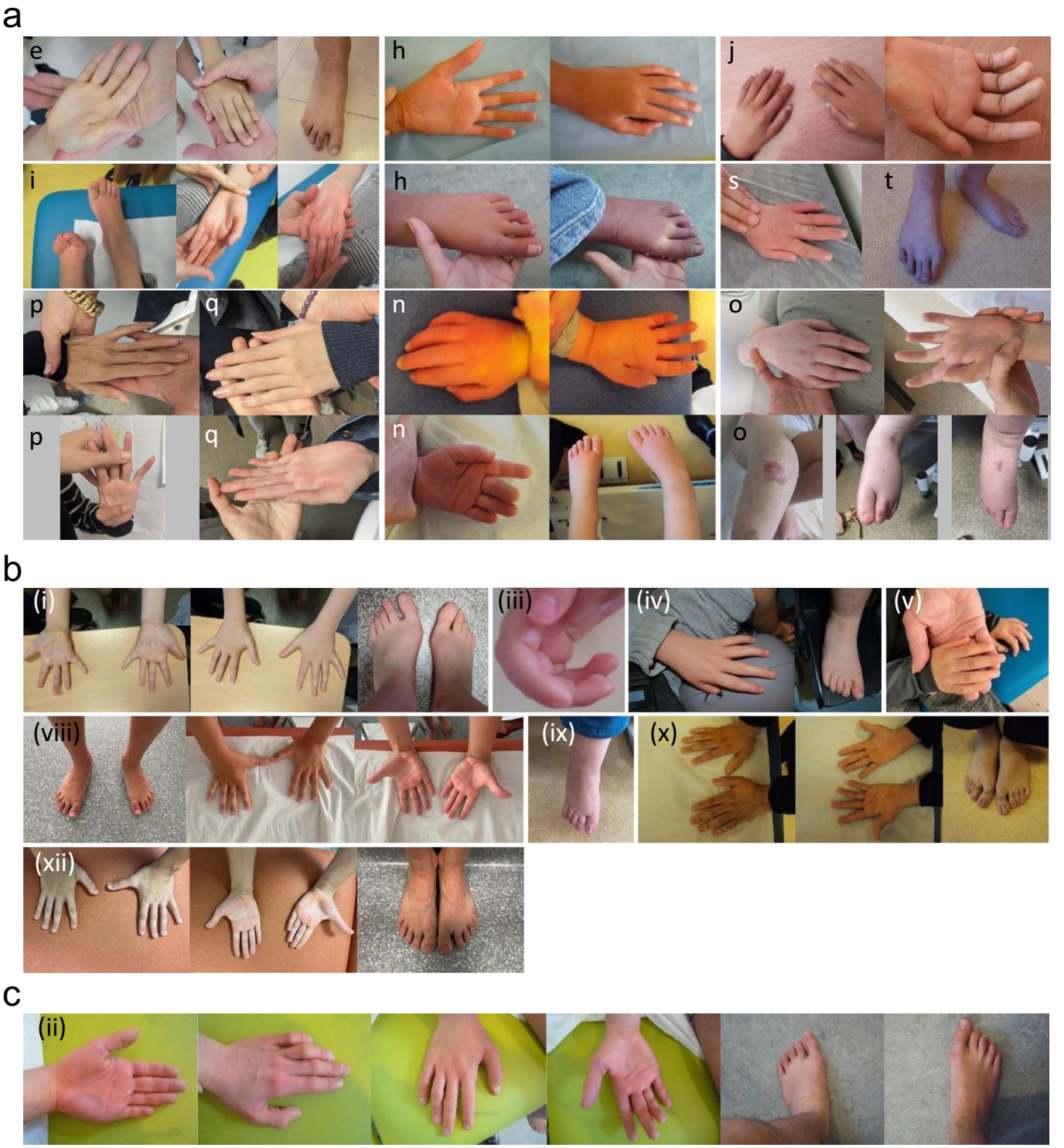
Extended Data Fig. 3 | *RNU4-2* genotype–phenotype correlations. **a**, Hierarchical clustering of the clinical features ($n = 44$, rows) of patients with pathogenic (P) and likely pathogenic (LP) *RNU4-2* variants ($n = 129$, columns). Categorical data were converted to 0–1 scale, and values were z-score scaled for each row. Blue–yellow–red scale depicts z scores. Lower values indicate a more favorable phenotype, while higher values represent a more severe phenotype. Missing values are shown in gray. Columns are colored based on the variant classification (purple, pathogenic; light purple, likely pathogenic), its location within the distinct U4:U6 domains (stem I, light blue; quasi-pseudoknot, orange; RBM42 interaction region, blue; stem III, green) and the nucleotide change (color shades related to their position within the respective U4:U6 domain). Rows are colored on the category of the clinical feature (shades of pink and green). **b**–**d**, Details of the principal component analysis of clinical features associated with *RNU4-2* LP/P variants presented in Fig. 3a. **b**, Percentage of explained variance by the first 10 principal components (PC). **c**, Top clinical features contributing to PC1. **d**, Top clinical features contributing to PC2. The horizontal red line represents the expected level of contribution if the contributions were uniform. Only variables with values above the red line are shown.

https://doi.org/10.1038/s41588-025-02184-4



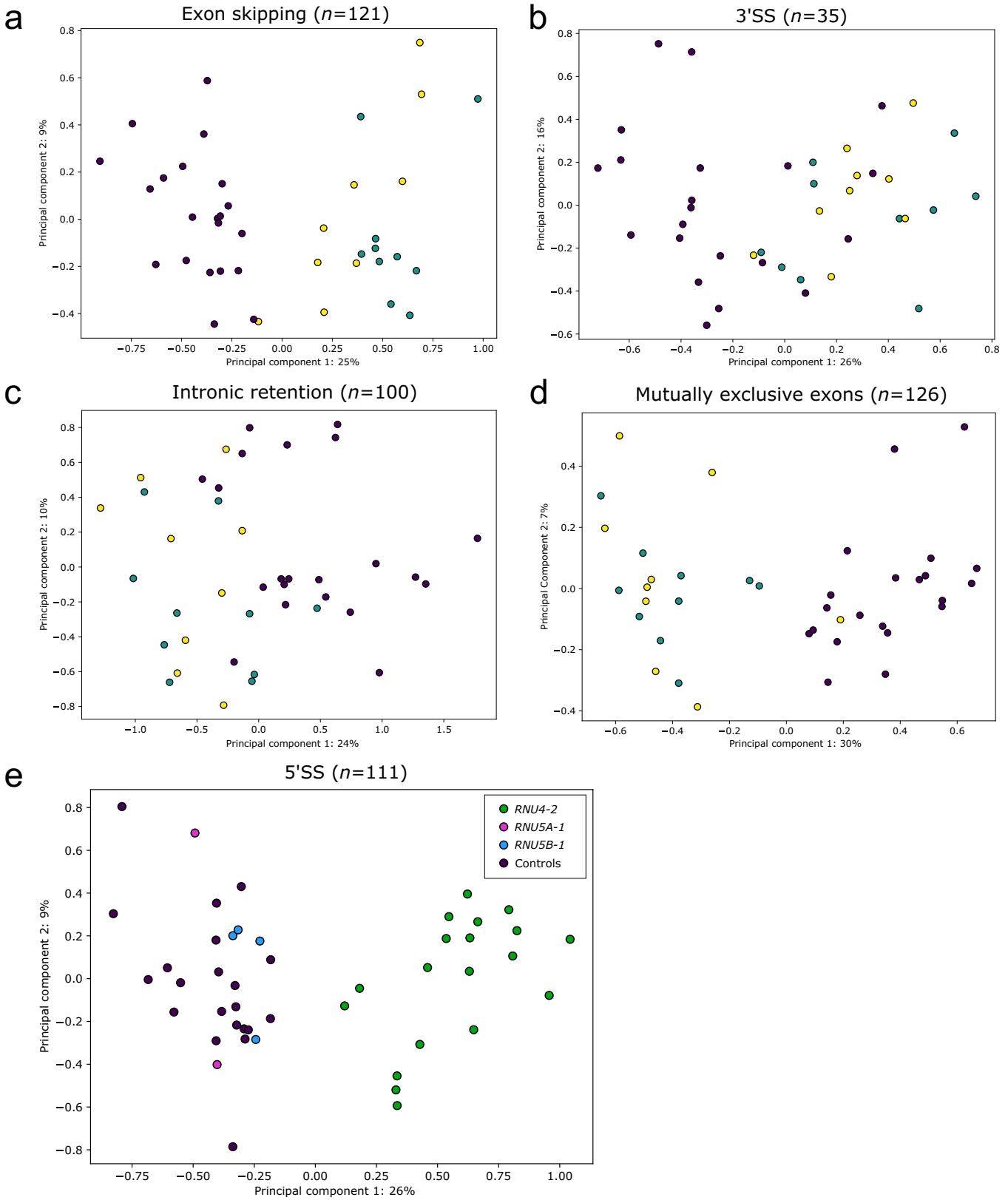
Extended Data Fig. 4 | Overview of the clinical characteristics of patients with *RNU4-2* LP/P variants. a, Aggregated clinical features of the whole cohort. **b**, Comparison of phenotypes related to *RNU4-2* variants in the T-loop and stem III domains. The *P* values were calculated using Fisher's exact tests (two-sided 2 × 2, 2 × 3 or 2 × 4 contingency tables) to compare 41 phenotypes between patients with n.64_65insT variants to those in the other three variant groups. Multiple comparisons were adjusted using the Bonferroni correction. Triangles, depletion (2 × 2 contingency tables); asterisks, significant difference (2 × 3 or 2 × 4 contingency tables). The percentage of patients with the feature, followed by the numerator (number of affected patients) and denominator (total assessed), are shown directly in the bars. For full details on all tests and patients numbers, please refer to Supplementary Table 8.

https://doi.org/10.1038/s41588-025-02184-4



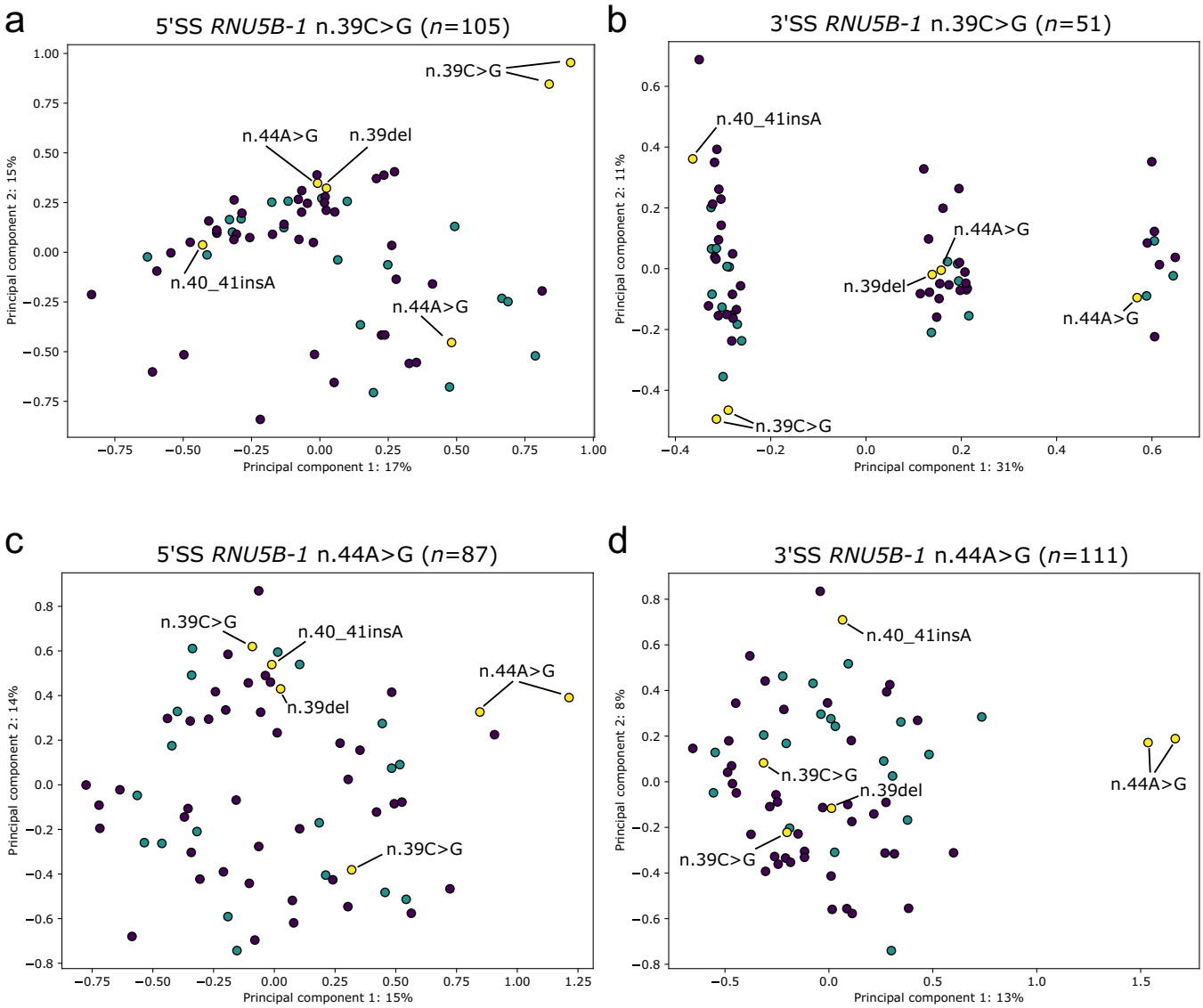
Extended Data Fig. 5 | Photographs of hands and feet of patients with *RNU4-2* or *RNU5B-1* LP/P variants. a, Individuals with the *RNU4-2* n.64_65insT variant. Please note the presence of vasomotor disorders (h,n), foot edema (o) and long fingers (h,i,p,q). The patient identification letters are the same as in Fig. 4. **b**, Individuals with other variants in *RNU4-2*. (i), n.65A>G; (iii) and (iv), n.66A>G; (v), n.67A>G; (viii)–(x), n.76C>T; (xii), n.77_78insT. The patient identification letters are the same as in Fig. 5a. **c**, Individual with the *RNU5B-1* n.39C>G variant. The patient identification letters are the same as in Fig. 5b.

https://doi.org/10.1038/s41588-025-02184-4



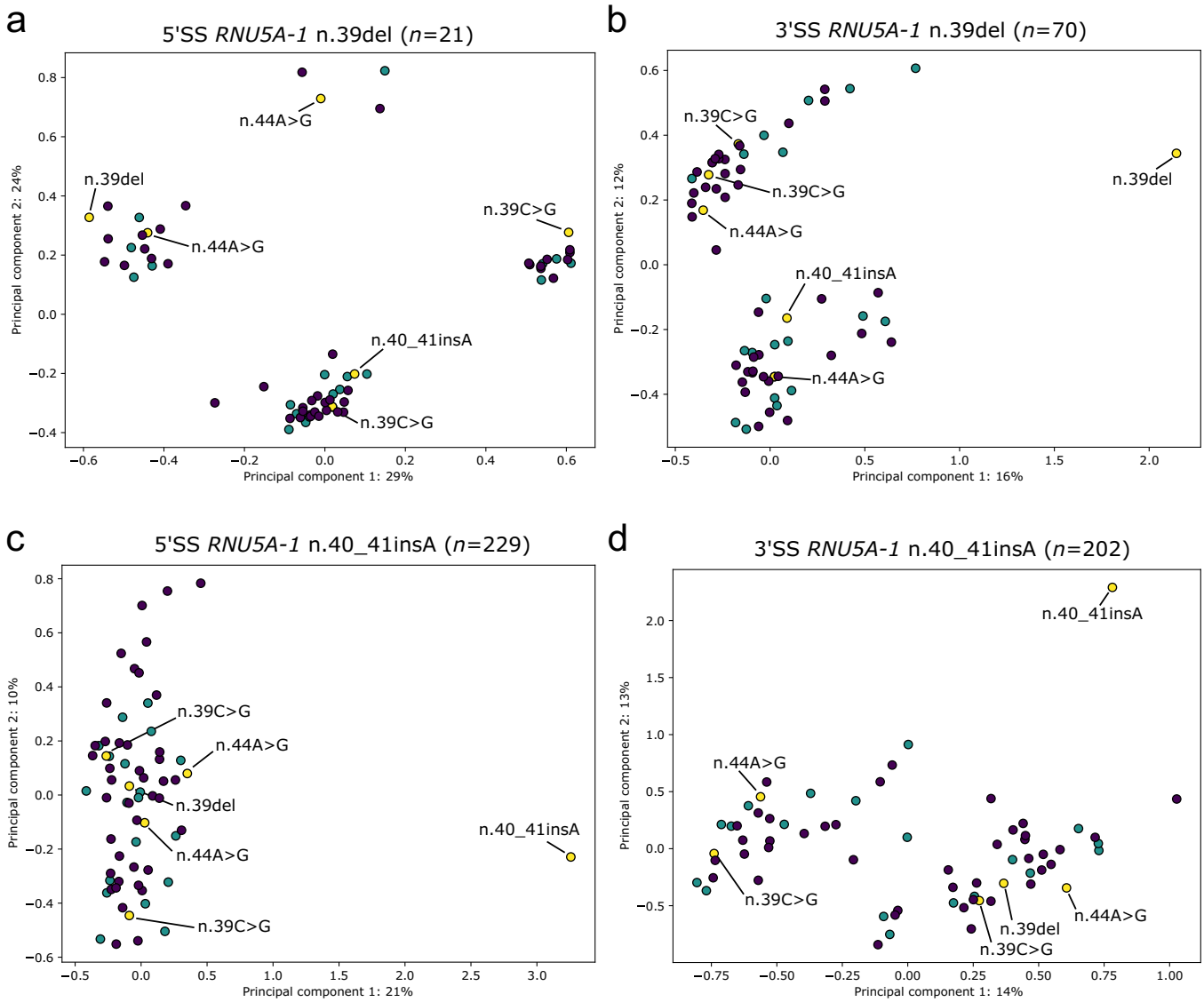
Extended Data Fig. 6 | Principal component analysis using PSI values of significant splicing events detected using rMATS. a–**d**, PCA of splicing events other than those altering the 5′SS in patients with *RNU4-2* variants. PCA was performed using PSI values of significant calls (FDR < 0.1) with a |deltaPSI| >0.05 for exon skipping (**a**, $n = 121$), alternative 3′ splice sites (**b**, $n = 35$, 3′SS), intronic retention (**c**, $n = 100$), or mutually exclusive exons (**d**, $n = 126$). Purple, controls; teal, *RNU4-2* n.64_65insT; yellow, other variants. **e**, PCA showing the *RNU4-2* 5′SS signature (111 events) applied to U5 variants. *RNU5B-1* ($n = 4$) and *RNU5A-1* ($n = 2$) variants cluster with controls, indicating that they do not share the 5′SS signature of *RNU4-2*. Purple, controls; green, *RNU4-2* variants; blue, *RNU5B-1* variants; pink, *RNU5A-1* variants.

https://doi.org/10.1038/s41588-025-02184-4



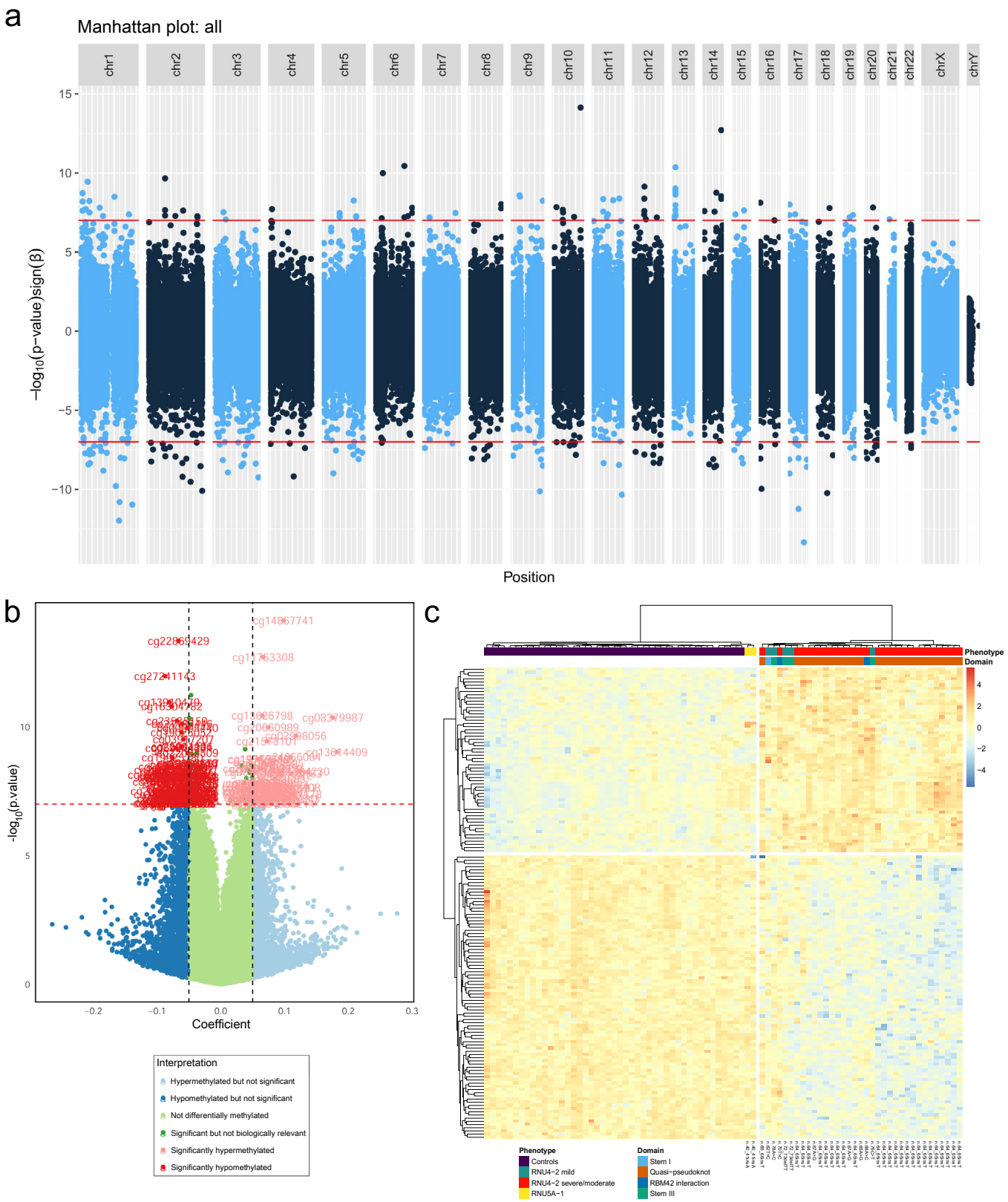
Extended Data Fig. 7 | Splicing analysis of 5′SS and 3′SS events in patients with *RNU5B-1* variants. a–**d**, Significant alternative 5′SS and 3′SS events (FDR < 0.1) were called for variants n.39C>G ($n = 2$) and n.44A>G ($n = 2$) independently against 21 controls using rMATS. PCAs were performed using PSI values of these calls with additional 20 controls, 19 *RNU4-2* and other U5 variants, including n.39del and n.40_41insA in *RNU5A-1*. PCAs were performed using PSI values of 105 5′SS events for n.39C>G (**a**), 51 3′SS events for n.39C>G (**b**), 87 5′SS events for n.44A>G (**c**), and 111 3′SS events for n.44A>G (**d**). Purple, controls; teal, *RNU4-2*; yellow, U5 variants.

https://doi.org/10.1038/s41588-025-02184-4



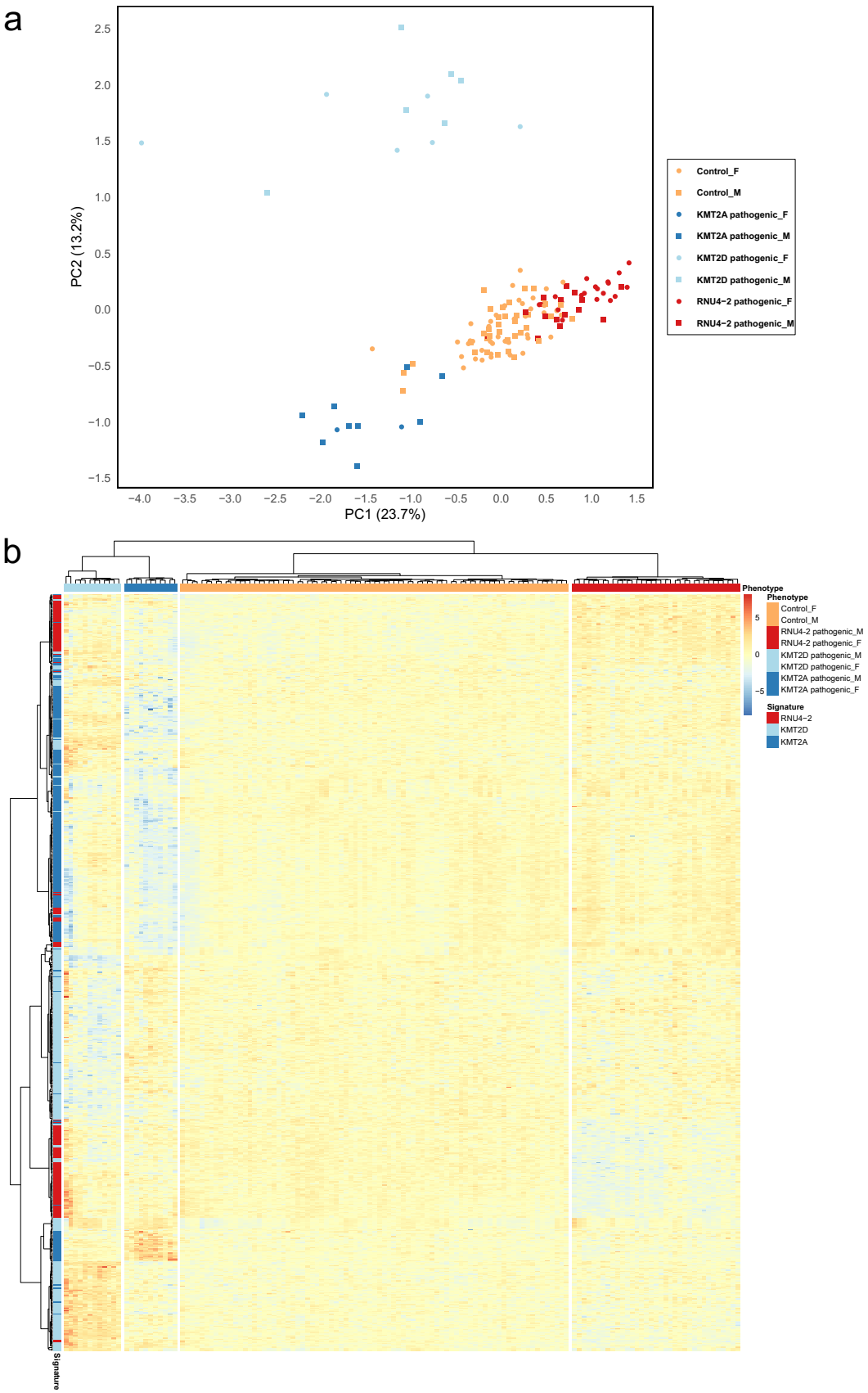
Extended Data Fig. 8 | Splicing analysis of 5′SS and 3′SS events in patients with *RNU5A-1* variants. a–**d**, Significant alternative 5′SS and 3′SS events (FDR < 0.1) were called for variants n.39del ($n = 1$) and n.40_41insA ($n = 1$) independently against 21 controls using rMATS. PCAs were performed using PSI values of these calls with additional 20 controls, 19 *RNU4-2* and other U5 variants, including n.39C>G and n.44A>G in *RNU5B-1*. PCAs were performed using PSI values of 21 5′SS events for n.39del (**a**), 70 3′SS events for n.39del (**b**), 229 5′SS events for n.40_41insA (**c**) and 202 3′SS events for n.40_41insA (**d**). Purple, controls; teal, *RNU4-2*; yellow, U5 variants.

https://doi.org/10.1038/s41588-025-02184-4



Extended Data Fig. 9 | Epigenetic study. a,**b**, Manhattan plot (**a**) and volcano plot (**b**) of the epigenome-wide association analysis (45 controls, 35 *RNU4-2* pathogenic variant carriers). **c**, Heatmap of adjusted methylation levels with the addition of both *RNU5A-1* n.40_41insA variants (n = 45 controls, 30 *RNU4-2* pathogenic variant carriers with moderate-to-severe phenotype, 5 with mild phenotype, 2 *RNU5A-1* n.40_41insA variant carriers, 147 differentially methylation positions).

https://doi.org/10.1038/s41588-025-02184-4



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41588-025-02184-4

Extended Data Fig. 10 | Comparison of *RNU4-2*, *KMT2A* and *KMT2D* episignatures. a, PCA of adjusted methylation levels, after correction for expected methylation based on age, sex and estimated blood cell counts showing no overlap between *RNU4-2* (147 probes), and the union of *KMT2A* (287 probes) and *KMT2D* (348 probes) episignatures. The compared groups are color-coded—orange for controls, red for *RNU4-2* P/LP variants, dark blue for *KMT2A* P/LP variants and light blue for *KMT2D* P/LP variants. The percentage of variance explained is provided on each axis. **b**, Heatmap of adjusted methylation levels on the union of *RNU4-2* (147 probes), and the union of *KMT2A* (287 probes) and *KMT2D* (348 probes) episignatures with the addition of two *RNU5A-1* n.40_41insA, 12 *KMT2D* and 11 *KMT2A* variant carriers. The compared subject groups in columns and signatures of origin in rows are color-coded similarly—orange for controls (columns only), red for *RNU4-2* P/LP variants, dark blue for *KMT2A* P/LP variants and light blue for *KMT2D* P/LP variants.

nature portfolio

Corresponding author(s): Caroline Nava, Christel Depienne

Last updated by author(s): Mar 27, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Variants in genes encoding snRNA in the PFMG were accessed using custom scripts. gLEAVES, the system used for genome analysis on the SeqIOA platform, which is restricted to registered and accredited users, was used to investigate inheritance of variants and visualize bam files. Genomics England data were analysed using custom scripts (Y.C./N.W) within the Genomics England secure research environment. Variants in RNU4-2 and RNU5B-1 were collected through the French Filières (Défisciences, AnDDI-Rares), the European reference networks (ERN) EpiCARE and ITHACA, and Genematcher (https://genematcher.org/). We additionally used data from gnomAD v4.1.0, All of Us (accessed via the publicly available data browser https://databrowser.researchallofus.org/ on 28 March 2023), and TOPMED_freeze10.

Data analysis	The following software and analysis tool were used: Geneious Prime® 2019.2.3, STAR aligner v.2.7.11a, CIBERSORTx v1.0, rMATS v.4.3.0, rmats2sahimi, seaborn v0.13.2, Logomaker v0.8, Meffil R package, methyclock R package, ggplot2 v3.3.6, pheatmap v1.0.12, stats v4.2.0, factoextra v1.0.7, and PyMol Version 3.0.0. Free energy for RNA secondary structure was calculated using bifold from RNAstructure v6.5 (https://rna.urmc.rochester.edu/RNAstructure.html). CADD PHRED scores were calculated using CADD v1.7. Preterm head circumference graphs were plotted using https://cpeg-gcep.shinyapps.io/prem2013/. Variants were reviewed using Alamut Visual Plus v1.11 and MobiDetails (https://mobidetails.iurc.montp.inserm.fr/MD/). Bam files were visualized with IGV 2.18.2. Custom scripts used for RNU4-2 transcriptomics are available on GitHub using the following link: https://github.com/benjamin-cogne/RNU4-2_transcriptomics (https://doi.org/10.5281/zenodo.13868502). We also used data from Ensembl Release 112 and data from ENCODE Consortium: bigwig files with the plus/minus strand signals of unique reads from the default anisogenic replicate from the following tissues: diencephalon (https://www.encodeproject.org/experiments/ENCSR000AFR/), parietal lobe (https://www.encodeproject.org/experiments/ENCSR000AFY/), occipital lobe, (https://www.encodeproject.org/experiments/ENCSR000AFX/), frontal cortex (https://www.encodeproject.org/experiments/ENCSR000AFS/), temporal lobe (https://www.encodeproject.org/experiments/ENCSR000AGD/), cerebellum (https://www.encodeproject.org/experiments/ENCSR000AFQ/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Variant details have been submitted to ClinVar (Pathogenic/ likely pathogenic variants: SUB15052639; https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB15052639; VUS: SUB15154590, https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB15154590;). RNAseq data and methylation have been deposited in the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega), which is hosted at the EBI. RNAseq data are available under the Study Accession Number EGAS50000000889 (https://ega-archive.org/studies/EGAS50000000889). Methylation data are accessible under the Study Accession Number EGAS00001008070 (https://ega-archive.org/studies/EGAS00001008070). Both are subjected to a Data Processing Agreement. Requests will be evaluated by a Data accessibility committee to ensure that data access complies with ethical and legal standards respective to the corresponding projects. Individual genome data could not be made publicly available due to ethical considerations. Controlled access to human genome data is necessary to protect the privacy of participants and to ensure compliance with ethical and legal standards, particularly the General Data Protection Regulation (GDPR) in Europe. Data access to individual genome data from the PFMG2025 with other researchers is submitted to current data protection and regulations in France and is only possible through the Collecteur Analyseur de Données (CAD). More information on data access and the CAD structure can be obtained on the PFMG2025 website: https://pfmg2025.fr/le-plan/collecteur-analyseur-de-donnees-cad/. Research on de-identified patient data from the Genomics England 100,000 Genomes Project and NHS GMS dataset is possible in the Genomics England Research Environment under a collaborative agreement. More information on data access can be obtained by email (research-network@genomicsengland.co.uk) or on the Genomics England website: https://www.genomicsengland.co.uk/research. All other data supporting the findings described in this manuscript are available in the article, Supplementary Tables, Extended Data Figures or Supplementary Information files. We also used data from Ensembl Release 112 and data from ENCODE Consortium: bigwig files with the plus/minus strand signals of unique reads from the default anisogenic replicate from the following tissues: diencephalon (https://www.encodeproject.org/experiments/ENCSR000AFR/), parietal lobe (https://www.encodeproject.org/experiments/ENCSR000AFY/), occipital lobe, (https://www.encodeproject.org/experiments/ENCSR000AFX/), frontal cortex (https://www.encodeproject.org/experiments/ENCSR000AFS/), temporal lobe (https://www.encodeproject.org/experiments/ENCSR000AGD/), cerebellum (https://www.encodeproject.org/experiments/ENCSR000AFQ/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We collected data on the patients' sex from their clinical records and referring physician. We did not collect data about gender.
Reporting on race, ethnicity, or other socially relevant groupings	We collected data about geographic origin and reported ancestry of a subset of participants but these data were not used in any analysis.
Population characteristics	Our study report variants in RNU4-2 in 150 patients (77 females, 73 males), variants in RNU5A-1 in three patients (2 males, one female) and variants in RNU5B-1 in 19 patients (6 females, 5 males, 8 unknown). The age of patients with RNU4-2 variants (at the time of inclusion) range from 4 months to 45 years. The age of patients with RNU5B-1 patients range from 3 to 36 years.
Recruitment	The main cohort is composed of patients with rare disorders and their parents when available, who underwent genome sequencing as part of the diagnostic process in France (Plan France Médecine Génomique 2025, PFMG2025). Analysis of 50 snRNA genes lead to further study 116 participants with rare variants in one of these genes (80 with variants in RNU4-2, 36 with variants in additional snRNA genes) from this cohort and 83 additional patients (70 with variants in RNU4-2, 13 with variants in RNU5B-1) from other cohorts. These cohorts include the Genomics England project, the Broad Centre for Mendelian Genomics, Undiagnosed Disease Networks (UDN), the BCH Epilepsy Genetics Program, and Care4Rare Canada for replication of the findings related to RNU5B-1 .
Ethics oversight	This study complies with the ethical standards of each of the participating countries. An informed consent was obtained for

Ethics oversight	all patients included in this study, from their parent or legal guardian. A specific consent form was obtained from the families who consented to publication of photographs. Patients/participants/samples were pseudonymized for the genetic study at each participating center. We collected information on the sex (but not gender) of the patients from the patients' clinical file. Grenoble-Alpes University Hospital (CHU Grenoble-Alpes, research 19814188) is the promoter of this research for the hospitals associated with the Auragen laboratory. Assistance Publique - Hôpitaux de Paris (AP-HP) is the promoter of this research for the hospitals associated with the SeqOIA laboratory (project ID: APHP241333). The study has received an approval from the ethics committee of University Hospital Essen (reference 24-12010-BO) and an approval of the Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations dans le domaine de la Santé (CESREES ; reference 21082803 Bis / 2038764)). AP-HP has obtained an authorization from the Commission Nationale de l'Informatique et des Libertés (CNIL; reference HGTHGT/MFIMFI/AR2426865; request no. 924924336666) for the data processing activities related to this project. Part of the study has been approved by the CHU de Nantes-ethics committee (number CCTIRS: 14.556). This research was ethically approved by CPP Ouest V (File 06/15) on 04/08/2015 (Ref MESR DC 2017 2987). For methylation analysis, DNA from all individuals (patients and controls) had been collected previously in the context of genetic analysis in a medical setting, following signature of a written, informed consent that includes a query on the use of leftovers in a research setting. Healthy controls consisted in individuals without NDD who underwent pre-symptomatic testing for other conditions and were found to be non-carriers or unaffected relatives of patients with a genetic disease, among non-carriers of pathogenic variants. Samples used for the methylation study were stored within the genetics biological collection of the CRBi, Rouen, France, declared as DC 2008-711 (access authorization n° MCRBi/2024/02). The analysis of methylation profiles based on previously stored DNA in these conditions was approved by the CERDE ethics committee (notification n° E2023-13) from the Rouen University Hospital. Researchers and clinicians from participating centers contributing either data or intellectual input were involved at all stages of the study from design, implementation, drafting, and revising the manuscript, and are coauthors of the article.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The main cohort (PFMG) is composed of 23,649 patients with rare disorders, including 15,073 patients with NDD and their parents. This number correspond to all genomes present in the PFMG2025 cohort at the time of the study. Sample size calculation was not performed, as the study was exploratory and not designed to test a predefined hypothesis.
Data exclusions	BAM files were visualized in IGV, and only confirmed de novo variants were included in further analysis. Variants resulting from mismapped reads (regions with multiple variants) or 'de novo' variants also detectable in reads of the parents were discarded.
Replication	We queried data from Genomics England, which included 8,841 undiagnosed NDD probands and 21,816 non-NDD probands, to replicate findings related to variants in RNU5A-1 and RNU5B-1. Variants in RNU5B-1 were replicated in many additional cohorts whereas no additional variants in RNU5A-1 was identified beyond the PFMG2025 cohort.
Randomization	Not applicable (observational study)
Blinding	Not applicable (observational study)

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	na
Novel plant genotypes	na
Authentication	na