

Further Delineation of snRNA Neurodevelopmental Disorders: Novel Genetic Contributors and Refined Phenotypic Characterization

Marta Sevilla-Porras

marta.sevilla@upf.edu

Universitat Pompeu Fabra

Esther Nieto-Molina

Beatriz Morte

Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER)

Luis Perez-Jurado

Pompeu Fabra University <https://orcid.org/0000-0002-1988-3005>

Article

Keywords: small nuclear RNA (snRNA), neurodevelopmental disorders, genome sequencing, biallelic variants, DNA methylation episinature, facial gestalt

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9609995/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is a duality of interest

- 1 Further Delineation of snRNA
- 2 Neurodevelopmental Disorders: Novel Genetic
- 3 Contributors and Refined Phenotypic
- 4 Characterization.

6 Marta Sevilla-Porras^{1.2}, Esther Nieto-Molina^{1.3}, Zahara Medina¹, Alejandra Damián¹,
 7 Juan Ramón Tejedor^{1.4.5.6}, Carlos Ruiz-Arenas^{7.8}, Mario Cazalla⁹, Rosario
 8 Carmona^{1.3.10}, Jorge Amigo¹¹, Ibai Goicoechea¹², Raúl Tonda¹³, Noemí Toro-Barrios³,
 9 Virginia Aquino³, Emma Soengas-Gonda^{1.14.15}, Cinta Navarro-Moreno^{1.16}, Marcela
 10 Mena^{1.17}, June Corcuera^{1.18.19}, Verónica Martos-Gago¹, Míriam Alvarez-Barona^{1.11.20},
 11 María Barreda-Sánchez^{21.22.23}, Cristina Silván⁹, Jair Tenorio-Castano^{1.9}, Carolina
 12 Alves²⁴, Maria Isabel Alvarez-Mora^{1.25.26}, Laia Rodriguez-Revenga^{1.25.26}, Irene
 13 Madrigal^{1.25.26}, Maria Juliana Ballesta-Martinez^{21.22.27}, Vanesa López-González^{1.21.22},
 14 Belén Pérez^{1.28.29.30}, Montserrat Morales-Conejo^{1.31.32}, Irene Valenzuela³³, Marta
 15 Codina-Solà³³, Marta Alemany Albert³⁴, Virginia Ballesteros Cogollos³⁴, Raquel
 16 Rodríguez-López³⁴, Berta Almoguera^{1.35}, Isabel Lorda-Sanchez^{1.35}, Irene Lazaro-
 17 Rodríguez^{15.36}, Miguel A Martin^{1.32.37}, Lidia Fernández-caballero^{1.35}, Helena Gil-Peña^{5.38},
 18 Noelia García-González³⁸, Marta Agúndez Sarasola^{39.40}, Judith Armstrong^{1.41}, Loreto
 19 Martorell⁴², Dídac Casas-Alba^{1.42}, Carmen Fons^{43.44}, Roser Urreizti^{1.45}, Antonio F
 20 Martinez-Monseny⁴², Leticia Pias-Peleiteiro⁴², Francisco Palau^{1.46.47}, Lexuri
 21 Gerrikabeitia^{42.47}, Nelmar V Ortiz-Cabrera^{1.48}, Anna Duat-Rodríguez^{1.48}, Bárbara
 22 Fernandez-Garoz⁴⁸, Laura López-Marín^{1.48}, Beatriz Bernardino-Cuesta^{1.48}, Elena Anton-
 23 Martin⁴⁸, María José González-Gómez⁴⁸, Elena González-Alguacil^{1.48}, Carmen Gómez
 24 Lado⁴⁹, David Dacruz-Alvarez⁵⁰, Beatriz Sobrino^{1.11.20.51}, Victor Martinez-Glez⁵², Anna
 25 Ruiz⁵², Carmen Manso-Bazús⁵², Nino Spataro⁵², Neus Baena⁵², Juan Pablo Trujillo-
 26 Quintero⁵², Nuria Capdevila⁵², Anna Brunet-Vega⁵², Eugenio Zapata-Aldana⁵², Veronica
 27 A Seidel⁵³, Raquel Yahyaoui^{54.55.56}, Ignacio Blanco⁵⁷, Elisabeth Castellanos^{57.58}, Montse
 28 Pauta⁵⁷, María del Mar Rovira⁵⁷, Bárbara Masotto⁵⁷, Agusti Rodriguez-Palmero^{1.57.59},
 29 Aurora Pujol^{1.60.61}, Agatha Schlüter^{1.60}, Maria Palomares-Bralo^{1.9}, María Ángeles
 30 Gómez-Cano⁹, María Nieves-Moreno⁶², Emi Rikeros-Orozco⁹, Antonio Poyatos-
 31 Andujar^{63.64}, Inmaculada Medina Martínez⁶⁵, Fernando Santos-Simarro^{66.67}, Damià

32 Heine-Suñer^{66.67}, Susana R Avella-Klaassen⁶⁶, Ángeles Pérez-Granero⁶⁶, María Antonia
 33 Grimalt⁶⁸, Ignacio Arroyo Carrera⁶⁹, Andrea Sariego-Jamardo⁷⁰, Ana I Vega^{71.72}, José L
 34 Fernández-Luna^{1.70.72}, Mireia Del Toro^{1.73}, Flora Sánchez-Jiménez⁷⁴, Juan M
 35 Borreguero-León⁷⁴, Andrea Campo Barasoain⁷⁵, Joaquin Dopazo^{1.3.17}, Mario F
 36 Fraga^{1.4.5.76}, Feliciano Ramos^{77.78}, Jordi Rosell^{1.66.79}, Enrique Galán-Gomez^{80.81}, Salud
 37 Borrego^{1.82}, Luis Castaño^{1.18.83.84}, Francisco Barros^{1.11}, José M Millán^{1.16.85}, Encarna
 38 Guillén-Navarro^{1.22.27.42}, Carmen Ayuso^{1.35}, Ángel Carracedo^{1.11.20.51}, Pablo
 39 Lapunzina^{1.9}, Beatriz Morte¹, Luis Alberto Pérez-Jurado^{1.2.86}
 40

41 **Affiliations**

- 42 1. Center for Biomedical Network Research on Rare Diseases (CIBERER), Instituto
43 de Salud Carlos III, Madrid, Spain
- 44 2. Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra,
45 Barcelona, Spain
- 46 3. Plataforma Andaluza de Medicina Computacional, Fundación Pública Andaluza
47 Progreso y Salud, Sevilla, Spain
- 48 4. Centro de Investigación en Nanomateriales y Nanotecnología (CINN-CSIC),
49 Oviedo, Spain
- 50 5. Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo,
51 Spain
- 52 6. Instituto de Oncología del Principado de Asturias, Oviedo, Spain
- 53 7. CIMA-Universidad de Navarra, Pamplona, Spain
- 54 8. Institute of Health Research of Navarra, Pamplona, Spain
- 55 9. INGEMM-Institute of Medical and Molecular Genetics, Hospital Universitario La
56 Paz, Madrid, Spain
- 57 10. Instituto de Biomedicina de Sevilla (IBiS), Sevilla, Spain
- 58 11. Fundación Pública Galega de Medicina Xenómica (FPGMX), Servicio Galego de
59 Saúde (SERGAS), Santiago de Compostela, Spain
- 60 12. Department of Personalized Medicine, NASERTIC, Government of Navarra,
61 Pamplona, Spain
- 62 13. Center for Genomic Regulation, Centro Nacional de Análisis Genómico,
63 Barcelona, Spain
- 64 14. Hospital Universitario 12 de Octubre, Madrid, Spain
- 65 15. Unidad de Dismorfología y Genética (UDisGen), Hospital Universitario 12 de
66 Octubre, Madrid, Spain
- 67 16. Instituto de Investigación Sanitaria La Fe, Valencia, Spain

17. Institute of Biomedicine of Seville (IBiS), University Hospital Virgen del Rocio/CSIC/University of Sevilla, Sevilla, Spain
18. Center for Biomedical Research Network in Diabetes and Associated Metabolic Diseases (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain
19. Instituto de Investigación Sanitaria Biobizkaia, Barakaldo, Spain
20. Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain
21. Medical Genetics, Department of Pediatrics, Hospital Clínico Universitario Virgen de la Arrixaca, Murcia, Spain
22. Instituto Murciano de Investigación Biosanitaria (IMIB), Murcia, Spain
23. Universidad Católica de Murcia (UCAM), Murcia, Spain
24. FDNA Inc., Atlanta, GA, USA
25. Biochemistry and Molecular Genetics Service, Hospital Clinic Barcelona, Barcelona, Spain
26. Fundació de Recerca Clínic Barcelona-Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
27. Universidad de Murcia, Murcia, Spain
28. Centro de Diagnóstico de Enfermedades Moleculares (CEDEM), Universidad Autónoma de Madrid, Madrid, Spain
29. Centro de Biología Molecular, Madrid, Spain
30. IdiPAZ, Madrid, Spain
31. Internal Medicine Department, Reference Center (CSUR) for Congenital Metabolic Disorders, Hospital Universitario 12 de Octubre, Madrid, Spain
32. Mitochondrial and Neuromuscular Disorders Group, Instituto de Investigación Hospital 12 de Octubre (imas12), Madrid, Spain
33. Clinical and Molecular Genetics Area, Vall d'Hebron Hospital; Medicine Genetics Group, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain
34. Consorcio Hospital General Universitario de Valencia, Valencia, Spain

- 96 35. Department of Genetics & Genomics, Instituto de Investigación Sanitaria-
97 Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid
98 (IIS-FJD, UAM), Madrid, Spain
- 99 36. Department of Pediatric Endocrinology, Hospital Universitario 12 de Octubre,
100 Madrid, Spain
- 101 37. Genetic Service, Hospital Universitario 12 de Octubre, Madrid, Spain
- 102 38. Hospital Universitario Central de Asturias - HUCA (Servicio de Salud del
103 Principado de Asturias - SESPA), Oviedo, Spain
- 104 39. Hospital Alfredo Espinosa, Osakidetza, Urduliz Bizkaia, Spain
- 105 40. Neuroscience Department, Universidad del País Vasco - UPV/EHU, Spain
- 106 41. Genomics Unit, Clinical Laboratory Service, and Institut de Recerca Sant Joan
107 de Déu, Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain
- 108 42. Department of Medical Genetics, Genetics Unit, and Institut de Recerca Sant
109 Joan de Déu, Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona,
110 Spain
- 111 43. Pediatric Neurology Service, and Institut de Recerca Sant Joan de Déu, Hospital
112 Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain
- 113 44. Centro Coordinador ERN-Epicare
- 114 45. Clinical Biochemistry Department, and Institut de Recerca Sant Joan de Déu,
115 Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain
- 116 46. Neurogenetics & Molecular Medicine Research Group, Center for Genomic
117 Sciences in Medicine, and Institut de Recerca Sant Joan de Déu, Hospital Sant
118 Joan de Déu, Esplugues de Llobregat, Spain
- 119 47. University of Barcelona, Barcelona, Spain
- 120 48. Hospital Infantil Universitario Niño Jesús, Madrid, Spain
- 121 49. Neuropediatrics Unit, Complejo Hospitalario Universitario de Santiago, Santiago
122 de Compostela, Spain
- 123 50. Area Sanitaria Pontevedra - O Salnes, Pontevedra, Spain

- 124 51. Grupo de Medicina Xenómica, CIMUS, Universidade de Santiago de
125 Compostela, Santiago de Compostela, Spain
- 126 52. Center for Genomic Medicine, Parc Taulí Hospital Universitari, Institut
127 d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Universitat Autònoma de
128 Barcelona, Sabadell, Spain
- 129 53. Hospital General Universitario Gregorio Marañón, Madrid, Spain
- 130 54. Metabolopathy Laboratory, Hospital Regional Universitario de Málaga, Málaga,
131 Spain
- 132 55. Instituto de Investigación Biomédica de Málaga IBIMA-Plataforma BIONAND,
133 Málaga, Spain
- 134 56. Department of Biomedicine and Dentistry, Facultad de Ciencias Biomédicas y
135 Deporte Universidad Europea de Andalucía, Málaga, Spain
- 136 57. Genetic Service, Laboratori Clínic de la Metropolitana Nord Hospital Germans
137 Trias i Pujol, Badalona, Spain
- 138 58. Genomic Unit, Institut de Recerca Germans Trias, Badalona, Spain
- 139 59. Universitat Autònoma de Barcelona, Barcelona, Spain
- 140 60. Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute
141 (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain
- 142 61. Catalan Institution of Research and Advanced Studies (ICREA), Barcelona,
143 Catalonia, Spain
- 144 62. Department of Pediatric Ophthalmology, Hospital Universitario La Paz, Madrid,
145 Spain
- 146 63. Clinical Laboratory Management Unit, Hospital Universitario Virgen de las
147 Nieves, Granada, Spain
- 148 64. Instituto de Investigación Biosanitaria ibs GRANADA, Granada, Spain
- 149 65. Neuropediatrics Section, Department of Pediatrics, Hospital Universitario Virgen
150 de las Nieves, Granada, Spain

- 151 66. Molecular Diagnosis and Clinical Genetics Unit, Hospital Universitario Son
152 Espases, Palma, Spain
- 153 67. Health Genomics Group, Instituto de Investigación Sanitaria Illes Balears
154 (IdISBa), Palma, Spain
- 155 68. Pediatric Neurology Unit, Department of Pediatrics, Hospital Universitario Son
156 Espases, Palma, Spain
- 157 69. Pediatrics Service, Hospital San Pedro de Alcántara, Cáceres, Spain
- 158 70. Hospital Universitario Marqués de Valdecilla, Santander, Spain
- 159 71. Genetic Unit, Hospital Universitario Marqués de Valdecilla, Santander, Spain
- 160 72. Instituto de Investigación Valdecilla (IDIVAL), Santander, Spain
- 161 73. Pediatric Neurology Department, Vall d'Hebron University Hospital, Barcelona,
162 Spain
- 163 74. Clinical Biochemistry, Hospital Universitario Virgen Macarena, Sevilla, Spain
- 164 75. UGC Pediatric, Neuropediatrics Section, Hospital Universitario Virgen Macarena,
165 Sevilla, Spain
- 166 76. Instituto de Oncología del Principado de Asturias (IUOPA), Oviedo, Spain
- 167 77. Unidad de Genética Clínica, Servicio de Pediatría, Hospital Clínico Universitario
168 'Lozano Blesa', Zaragoza, Spain
- 169 78. Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain
- 170 79. Instituto de Investigación Sanitaria Illes Balears (IdISBa), Palma, Spain
- 171 80. Hospital Universitario de Badajoz, Badajoz, Spain
- 172 81. Facultad de Medicina, Universidad de Extremadura, Badajoz, Spain
- 173 82. Departamento de Medicina Maternofetal, Genética y Reproducción, Institute of
174 Biomedicine of Seville (IBiS), University Hospital Virgen del
175 Rocío/CSIC/University of Sevilla, Sevilla, Spain
- 176 83. Hospital Universitario Cruces, UPV/EHU, Barakaldo, Spain
- 177 84. Endo-ERN
- 178 85. Hospital Universitario y Politécnico La Fe, Valencia, Spain

179 86. Genetics Service, Hospital del Mar and Hospital del Mar Research Institute,
180 Barcelona, Spain

181
182
183

Abstract

Spliceosomal small nuclear RNA (snRNA) genes are an emerging cause of neurodevelopmental disorders (NDDs). We analyzed genome sequencing data from 1,708 probands with undiagnosed NDDs enrolled in the Spanish national rare disease programs and identified pathogenic variants in snRNA genes in 2.58% of individuals. We report de novo variants in three snRNA genes, *RNU4-2*, *RNU2-2* and *RNU5B-1*, and biallelic variants in *RNU4-2* and *RNU2-2*, including a deletion at the *RNU4-2* locus. Two individuals with biallelic variants in *RNU4-1* were identified, suggesting this gene as a potential candidate locus pending replication. Digenic combinations involving snRNA genes were enriched in individuals with NDDs (OR=2.08, p=0.04), suggesting potential digenic inheritance as an additional mutational mechanism. A domain-specific whole blood DNA methylation episinature was identified in individuals with de novo *RNU4-2* T-loop variants, while no robust episinature was identified for *RNU2-2* or *RNU5B-1*. A recognizable facial gestalt was identified in individuals with *RNU4-2* variants using Face2Gene, with strong discrimination from controls (AUC=0.948) and from individuals with *RNU2-2* variants (AUC=0.921). *RNU2-2* variants were predominantly associated with epileptic encephalopathy and *RNU5B-1* with a broader phenotype. A prenatal phenotype with intrauterine growth restriction was observed in 52% of individuals with *RNU4-2* variants, with a higher frequency in females (70%) than males (33%). These findings expand the genetic and phenotypic landscape of snRNA-related disorders and have implications for variant interpretation, genetic counseling, and prenatal surveillance.

Keywords: small nuclear RNA (snRNA), neurodevelopmental disorders, genome sequencing, biallelic variants, DNA methylation episinature, facial gestalt.

208 INTRODUCTION

209 Genome sequencing (GS) has transformed the diagnosis of neurodevelopmental
210 disorders (NDDs) by enabling systematic interrogation of non-coding regions of the
211 genome. One of the most relevant recent findings is the identification that variants in the
212 spliceosomal small nuclear RNA (snRNA) genes are a frequent cause of NDDs. snRNAs
213 are core components of the major spliceosome, where they mediate intron recognition
214 and catalysis of pre-mRNA splicing. Pathogenic variants in at least four genes—*RNU4-*
215 *2* (ReNU syndrome, OMIM#620823), *RNU2-2* (OMIM#621238), *RNU5A-1*
216 (OMIM#180691) and *RNU5B-1* (OMIM#621090)—have been linked to NDDs^{1–6}. *RNU4-*
217 *2* alone accounts for approximately 0.4% of all NDD individuals¹, being the most
218 prevalent monogenic cause identified to date, excluding fragile X syndrome⁷. *RNU2-2*
219 contributes an additional ~20% relative to this burden in its dominant form, with a
220 recessive form now estimated to be even more frequent^{3,8–11}. In all cases, pathogenic
221 variants affect functionally important domains of the respective snRNA. For dominant
222 forms, clustering within specific constrained regions is well established^{1,3–5}. For recessive
223 *RNU4-2*, saturation genome editing has confirmed that biallelic variants map to
224 functionally depleted positions in distinct structural elements from those causing
225 dominant disease^{9,10}; the structural basis of recessive *RNU2-2* variants is less
226 systematically characterized^{8,9,11}.

227 Transcriptomic and epigenomic studies have begun to clarify the underlying
228 mechanisms. Dominant *RNU4-2* variants produce a reproducible signature of alternative
229 5' splice site selection detectable in blood and fibroblasts^{1,5}; dominant *RNU2-2* variants
230 affect branchpoint recognition, with subtle splicing alterations in blood^{3,11}; and *RNU5B-1*
231 variants impair loop I-mediated exon alignment, yielding a distinct 3' splice site signature
232^{4,5}. The recessive forms operate through different mechanisms: biallelic *RNU4-2* variants
233 lead to reduced transcript levels without producing the dominant alternative splice site

signature⁹, while biallelic *RNU2-2* variants result in reduced U2-2 snRNA abundance, with splicing defects detectable in fibroblasts but not blood^{9,12}. At the epigenomic level, dominant *RNU4-2* variants are associated with a reproducible DNA methylation episignature; by contrast, individuals with *RNU2-2* variants show only modest variant-specific methylation changes that do not constitute a robust episignature¹¹. Unlike *RNU4ATAC*-opathies of the minor spliceosome, in which biallelic variants cause widespread intron retention^{13,14}, most major spliceosome snRNA variants produce a more selective disruption of splicing fidelity, pointing to distinct underlying mechanisms^{3,5,15}.

Nevertheless, the genetic and clinical landscape of snRNA-related neurodevelopmental disorders remains incompletely defined. Current knowledge has been largely shaped by a dominant *de novo* model, while the contribution of alternative mutational mechanisms—including biallelic inheritance, structural variation, and potential oligogenic interactions—has not been systematically explored^{8–10}. In addition, key aspects of disease characterization, such as domain-specific epigenetic signatures, detailed genotype–phenotype correlations, and early developmental manifestations, remain limited.

To address these gaps, we leveraged the Spanish national programs for undiagnosed rare diseases—URD-Cat, ENOD-CIBERER, and IMPaCT-Genómica, coordinated through the CIBERER consortium^{16,17}—which combine GS with standardized phenotyping and data sharing across multiple centers. By analyzing 1,708 probands with undiagnosed NDDs, we expand the genetic architecture of snRNA-related disorders beyond the canonical *de novo* model, identifying pathogenic variants in 44 individuals: 38 with *de novo* variants across *RNU4-2*, *RNU2-2* and *RNU5B-1*, and 6 with biallelic variants in *RNU4-2* and *RNU2-2*, including a structural deletion at the *RNU4-2* locus. Two additional individuals carried biallelic variants in *RNU4-1*, proposed here as a novel candidate gene.

Through integrated genomic, epigenetic, and phenotypic analyses, we define a domain-specific DNA methylation epismutation for *RNU4-2*, delineate gene-specific clinical profiles supported by facial gestalt analysis, and identify a previously unrecognized prenatal phenotype. Together, these findings underscore the relevance of undiagnosed RD programs and contribute to refining the molecular and clinical framework of snRNA-related disorders, with direct implications for variant interpretation, genetic counseling, and diagnostic strategies.

SUBJECTS AND METHODS

Individuals enrolled in Spanish URD Programs

Inclusion criteria for individuals in the Spanish URD Programs (IMPaCT-Genómica, URD-Cat and ENoD-CIBERER) included: 1) a suspected genetic disorder that remained undiagnosed; 2) a complete diagnostic assessment with exome sequencing, currently considered standard-of-care in clinical genetics; 3) reanalysis of exome data if analysis was done >1 year ago (most individuals); and 4) informed consent to the storage and anonymous sharing of clinical and genomic data. There are no age or disease-type restrictions, but most individuals are congenital or pediatric-onset, and predominantly neurological disorders.

The aggregate data of these programs (until 12/2025) comprised 2,200 probands and 1,000 unaffected relatives (mostly parents), including 1,708 probands with NDDs. Phenotypic information from all individuals has been collected on a platform developed specifically for these programs, with standardized Human Phenotype Ontology (HPO) terms¹⁸ to ensure consistent clinical data collection and analysis. An average of 11.3 terms per individual was recorded.

Genome sequencing and data processing

1 µg of DNA was used for library preparation using the KAPA HyperPrep/ HyperPlus Kit (Roche Sequencing Solutions, Inc.) following the manufacturer's protocol. GS libraries were sequenced on the NovaSeq 6000 or NovaSeq X Plus (Illumina, Inc.) with 2x150 bp pair-end reads, achieving ≥30X mean coverage, >400 bp mean insert size, and >15X depth for ≥95% of the GRCh38. Details about processing are in supplementary material.

Reverse phenotyping

Clinical reevaluation was conducted at each participating institution for all individuals carrying *de novo* pathogenic variants in *RNU2-2*, *RNU4-2* or *RNU5B-1*, to ensure accurate and standardized phenotypic data collection. A total of 200 individual HPO terms were collected from the phenotypic data of the 38 individuals, and clinicians (neuropsychiatrist and/or clinical geneticist) were responsible for validating and expanding phenotypic annotations. To evaluate phenotypic variability among individuals with variants in the three gene groups, we performed unsupervised hierarchical clustering using HPO-coded features. Individuals with incomplete phenotypic annotation (<20% features) were excluded, resulting in a final cohort of 32 individuals: 22 with *RNU4-2*, 6 with *RNU2-2*, and 4 with *RNU5B-1* variants. Clustering was based on Euclidean distance and Ward's linkage, and results were visualized through a heatmap to illustrate shared and distinct phenotypic features.

We compared the frequency of individual HPO terms across genes to identify gene-specific phenotypic signatures. HPO terms were considered gene-specific if they showed $\geq 30\%$ difference in frequency between genes and occurred in $>50\%$ of individuals for the gene of interest.

Face-mask analysis

Frontal facial photographs of individuals with *de novo* dominant variants in *RNU4-2*, *RNU2-2* and *RNU5B-1* genes were analyzed using the Face2Gene Research application tool (FDNA Inc., USA) ¹⁹. The analysis included 15 individuals with *RNU4-2* variants, 10 individuals with *RNU2-2* variants, 4 individuals with *RNU5B-1* variants and a comparison cohort of 31 unaffected individuals. Controls were selected from unaffected individuals available within the Face2Gene Research platform database matched for sex, ethnicity, and age distribution (Table S1). Details of analysis in Supplementary material.

Genome-wide DNA methylation profiling

DNA samples were extracted from whole blood of 17 individuals with *RNU4-2* variants, 5 individuals with *RNU2-2* variants and 4 individuals with *RNU5B-1* variants using standard protocols (Table S2). Individuals' samples, along with DNA samples from 47 age- and sex-matched controls were processed and hybridized to the Illumina Infinium MethylationEPICv2 BeadChip arrays (San Diego, CA, USA) following the manufacturer's protocol. Additional details in Supplementary material.

RESULTS

Expanded genetic architecture of snRNA-related neurodevelopmental disorders

To systematically investigate the contribution of snRNA genes to neurodevelopmental disorders (NDDs), we analyzed 1,708 probands enrolled in Spanish URD programs using GS and detailed phenotypic data. We identified pathogenic variants in multiple snRNA genes, revealing an expanded mutational landscape that includes *de novo*, autosomal recessive, structural, and potentially oligogenic mechanisms.

We detected *de novo* pathogenic variants in 38 individuals involving *RNU4-2*, *RNU2-2* and *RNU5B-1*, all affecting functionally constrained regions (Table 1; Figure 1A–B; Table S3). Among these, 28 individuals carried heterozygous variants in *RNU4-2*, with a strong enrichment in the T-loop and Stem III domains, which form part of the U4/U6 interaction domain. The recurrent insertion n.64_65insT accounted for 70% of individuals, confirming a mutational hotspot. *De novo* occurrence was confirmed in all but two individuals both conceived via *in vitro* fertilization with anonymous gamete donors. Six individuals carried recurrent *de novo* variants in *RNU2-2* (n.35A>G and n.4G>A), previously reported³. Four unrelated individuals carried *de novo* variants in *RNU5B-1*, all clustering within the conserved loop I domain. We did not identify additional *de novo* pathogenic variants in other members of the RNU5 family or in the remaining candidate snRNA genes analyzed^{4,5}.

We systematically searched for biallelic variants across 19 snRNA genes selected based on high brain expression in GTEx (Figure S1). We identified six probands carrying rare biallelic variants in *RNU4-2* (3) and *RNU2-2* (3), and two additional individuals with biallelic variants in *RNU4-1*, supporting a broader mutational spectrum that includes recessive configurations (Figure 1A–B; Table S4–S5).

In *RNU4-2*, three individuals carried biallelic variants, including one individual with a SNV (n.119A>G) in trans with a 1.3-kb deletion encompassing the entire gene (Figure 1C), likely mediated by Non-Allelic Homologous Recombination (NHAR) between *RNU4-1* and *RNU4-2*, highlighting an additional mutational mechanism at this locus. Parental genome sequencing confirmed the segregation of each allele, with the deletion and the SNV inherited from different parents, supporting a compound heterozygous configuration. This represents, to our knowledge, the first structural alteration affecting *RNU4-2*.

Similarly, three individuals, including two affected siblings, carried biallelic variants in *RNU2-2* and presented with neurodevelopmental delay and epileptic phenotypes. In both *RNU4-2* and *RNU2-2*, all biallelic configurations were consistent with either homozygous or compound heterozygous states, supported by read-level evidence from GS, as the two variants were never observed within the same sequencing reads. The segregation and parental origin of each variant are detailed in Table 1. Notably, all corresponding SNVs had been previously reported in individuals with recessive snRNA-related disorders and were located outside the functional domains classically associated with dominant *de novo* disease, supporting distinct mutational landscapes in recessive contexts (Table 1, Table S6).

Two additional individuals carried compound heterozygous variants in *RNU4-1*, a gene not previously established as a cause of neurodevelopmental disorders, suggesting a potential but currently unconfirmed contribution in specific genetic contexts (Table 1, Table S6). The relatively higher allele counts of these variants in gnomAD compared to those in *RNU4-2* and *RNU2-2* are consistent with a recessive carrier state and do not preclude pathogenicity in a biallelic context, but formal functional validation is lacking, and these individuals should be considered as candidates pending replication. Notably, one of these individuals also carries a *de novo* deletion affecting *MED13L*, which is

considered the primary molecular diagnosis in this case, and therefore the contribution of the *RNU4-1* biallelic variants to the phenotype remains uncertain.

Not all biallelic configurations in *RNU4-2* resulted in NDD. The variant n.7G>A, previously reported in compound heterozygosity with NDD¹⁰ was identified in *trans* with n.18_19insA in a family with autosomal dominant retinal dystrophy without neurodevelopmental involvement²⁰, indicating that allelic context critically determines phenotypic outcome (Figure S2). These findings underscore that variant interpretation in snRNA genes must account for variant position, zygosity, and allelic combination.

We observed a modest enrichment of digenic variant combinations in individuals with NDDs compared to controls (OR = 2.08, *p* = 0.04; Figure S3). This analysis compared the frequency of individuals carrying rare variants (AF <0.001) in two or more distinct snRNA subunits between the NDD cohort (*n*=1,708) and a control group of 2,121 unaffected individuals, comprising 1,021 unaffected relatives from the URD programs and 1,100 individuals with no reported neurodevelopmental phenotype, using Fisher's exact test without correction for multiple comparisons given the exploratory nature of the analysis. This modest but significant enrichment suggests that oligogenic mechanisms involving different snRNA subunits may contribute to disease in a subset of individuals, although this observation requires replication in independent cohorts.

A domain-specific episignature defines *RNU4-2*-associated neurodevelopmental disorder

To further investigate the functional consequences of these variants, we evaluated their predicted effects on RNA secondary structure across genes. RNAfold predictions revealed variant-specific effects, supporting a role for structural context in modulating the impact of snRNA variants. In *RNU5B-1*, the recurrent insertion (n.42_43insA) induced a marked structural rearrangement, whereas single nucleotide substitutions produced only

local perturbations (Figure S4). Similarly, biallelic variants in *RNU4-1*, *RNU4-2* and *RNU2-2* were associated with distinct structural configurations, suggesting alternative modes of disruption. Together, these observations indicate that the functional impact of snRNA variants depends on their structural and domain context.

We next evaluated whether these differences were reflected at the epigenetic level through DNA methylation profiling. Individuals with *de novo* *RNU4-2* variants showed a robust and reproducible episignature, particularly for variants affecting the T-loop domain (Figure 2a-b, Table S7). This signature was supported by consistent methylation changes and accurate classification using a support vector machine model (5/6 validation *de novo* samples correctly classified) (Figure S6a-b), with the misclassified sample carrying a variant outside the main functional domain. The direction and magnitude of methylation changes were concordant with other reported episignatures⁵ (Figure S6c), supporting reproducibility of the underlying biological signal and a domain-specific epigenetic effect.

To further explore whether this episignature was shared by other inheritance models, we applied the *RNU4-2* classifier to an additional individual carrying biallelic variants in *RNU4-2* (harboring a deletion in *trans* with n.119A>G; U4-p32) (Figure S6b). This sample was not classified within the *RNU4-2* episignature, suggesting that this biallelic configuration may not recapitulate the epigenetic signature observed in individuals with *de novo* variants.

Functional enrichment analyses of genomic regions contributing to the *RNU4-2* episignature, indicated alterations in regulatory chromatin states, including enhancer and repressive-associated histone marks (H3K27me3 and H3K9me3) (Figure S6d-e). Transcription factor binding site analysis revealed enrichment for cohesin complex components (*SMC1A*, *SMC3* and *RAD21*) (Figure S6f), suggesting a potential link between spliceosomal dysfunction and chromatin organization. Differentially methylated

regions were also enriched for loci overlapping genes involved in neurodevelopment and synaptic function, including *ATG16L2*, *LRFN1* and *CALN1*(Table S8, Figures S8-14). A hypomethylated region overlapping the promoter of the *RNU6-672P* pseudogene suggests a potential downstream or compensatory regulatory response within the spliceosomal network (Figure 2c).

In contrast, no robust or reproducible epesignatures could be defined for *RNU2-2* or *RNU5B-1* (Figure S7). Notably, although most individuals with *RNU5B-1* variants did not show clear classification, one individual displayed slightly higher probabilities of assignment to the *RNU4-2* group compared to controls (Figure S6b), suggesting partial epigenetic convergence without supporting a consistent gene-specific signature.

Gene-specific clinical phenotypes in snRNA-related NDDs

To systematically assess whether individuals with *de novo* variants show gene-specific phenotypic patterns, we performed reverse phenotyping across the cohort, followed by unsupervised clustering based on Human Phenotype Ontology (HPO) terms. A total of 200 unique HPO terms were annotated (Table S9).

Unsupervised clustering revealed three principal phenotypic groups (Figure S15). Two clusters were predominantly composed of individuals with *RNU4-2* variants, whereas a third cluster included all individuals with *RNU2-2* variants together with most individuals with *RNU5B-1* variants. Within this cluster, individuals with *RNU2-2* variants showed a relatively homogeneous clinical profile, while individuals with *RNU5B-1* variants displayed greater phenotypic variability. Clustering was independent of age, supporting a primary role of genotype in phenotypic stratification.

Across all genes, shared features included global developmental delay, intellectual disability, motor delay, and speech impairment. However, gene-specific patterns were evident. *RNU4-2* variants were associated with craniofacial and neuroanatomical

abnormalities, including microcephaly, hypotelorism and corpus callosum hypoplasia. *RNU2-2* variants were more frequently associated with severe epileptic encephalopathy and developmental regression, with less distinctive craniofacial features, whereas *RNU5B-1* presented a broader and more heterogeneous phenotype (Figure 3a).

Within *RNU4-2*, variants clustered mainly in the T-loop domain. Individuals carrying T-loop variants tended to exhibit more severe clinical features compared to those with variants in Stem III, although this observation is limited by sample size.

Facial gestalt analysis supports gene-specific craniofacial profiles

To determine whether these gene-specific clinical differences were also reflected in facial morphology, we performed computational facial phenotyping using DeepGestalt and GestaltMatcher.

DeepGestalt multiclass classification showed partial separation between groups (mean accuracy 53.3%, exceeding the class-balanced random expectation of 39.2%) (Figure 4a). Binary analyses demonstrated robust discrimination of individuals with *RNU4-2* variants from both controls (AUC = 0.948, $p = 0.003$) and individuals with *RNU2-2* variants (AUC = 0.921, $p = 0.035$) (Figure 4b-c), supporting a consistent and recognizable facial gestalt. In contrast, individuals with *RNU2-2* variants were not reliably distinguished from controls, consistent with their less distinctive craniofacial phenotype.

GestaltMatcher analysis further supported these findings, with *RNU4-2* individuals clustering more closely together, whereas individuals with *RNU5B-1* variants appeared more dispersed across the facial phenotype space, consistent with their greater clinical heterogeneity (Figure 4d).

Prenatal manifestations expand the phenotypic spectrum of *RNU4-2*-associated disorder

Given the consistent clinical profile observed in *RNU4-2*, we next examined whether features of the disorder could be detected during prenatal development. A distinct prenatal phenotype was identified in individuals with *RNU4-2* variants, characterized by intrauterine growth restriction (IUGR), reduced birth weight, prematurity and mild ventriculomegaly, with a marked female predominance. IUGR was observed in 13 of 25 individuals with *RNU4-2* variants for whom prenatal data were available (52%), with a striking sex difference: 9 of 13 affected females (70%) versus 4 of 12 affected males (33%) presented with IUGR. Preterm birth was also more frequent in females, occurring in 4 of 13 (31%) compared to 1 of 12 males (9%). Mean birth weight was lower in females (2,483 $\pm$ 388 g) than in males (3,051 $\pm$ 483 g), and remained reduced after excluding preterm individuals (females: 2,591 $\pm$ 383 g; males: 3,117 $\pm$ 499 g). These features were not observed in individuals with *RNU2-2* or *RNU5B-1* variants, all of whom were born at term with no reported IUGR (Figure 3b, Table S10).

Together, these findings indicate that *RNU4-2*-associated disorder includes a previously underrecognized prenatal component and suggest that spliceosomal dysfunction may impact early human development. This pattern is consistent with disorders affecting the minor spliceosome, supporting a broader role of RNA splicing in prenatal growth and neurodevelopment^{21–23}.

DISCUSSION

Our data support a model in which both *de novo* and biallelic variants in snRNA genes contribute to neurodevelopmental disorders (NDDs) through distinct molecular mechanisms. Across genes, *de novo* variants consistently localize to highly constrained functional domains, whereas biallelic variants tend to occur outside these regions. This

positional segregation suggests fundamentally different modes of disruption, in line with transcriptomic studies indicating domain-specific effects for dominant variants and more variable, likely hypomorphic perturbations in recessive contexts^{1,5,9,10}.

A key finding of this study is the identification of a robust, domain-specific DNA methylation episinature associated with *de novo* *RNU4-2* variants. This signature is driven by alterations in the T-loop domain and provides functional evidence for domain-dependent pathogenicity at this locus. In contrast, the individual with biallelic *RNU4-2* variants (n=1) was not classified within the dominant episinature, which may suggest that this configuration does not recapitulate the epigenetic signature observed in *de novo* disease; however, this observation is based on a single individual and should be interpreted with caution pending validation in larger recessive cohorts. Similarly, the lack of consistent episinatures for *RNU2-2* and *RNU5B-1* further highlights the heterogeneity of downstream effects across genes and inheritance modes¹¹.

Beyond classification, epigenetic profiling also provides insight into downstream mechanisms. Enrichment for histone marks associated with enhancer and repressive chromatin states, together with overrepresentation of binding sites for cohesin complex components (*SMC1A*, *SMC3* and *RAD21*), suggests a link between spliceosomal dysfunction and chromatin organization. Notably, cohesin interacts directly with splicing factors and components of the U4/U6.U5 tri-snRNP complex^{24,25}, and actively regulates alternative splicing — a connection that may be relevant to the overlapping phenotypic features observed between snRNA-related disorders and Cornelia de Lange syndrome, in which cohesin dysfunction is causal²⁶. In addition, differentially methylated regions overlapping genes involved in neurodevelopment and synaptic function, as well as a hypomethylated region at the promoter of the *RNU6-672P* pseudogene, point to potential downstream or compensatory regulatory responses within the spliceosomal network.

From a genetic perspective, we further expand the architecture of snRNA-related disorders by identifying novel mutational configurations. In particular, we report the first structural alteration affecting *RNU4-2*, consisting of a complete gene deletion in trans with a previously reported variant (n.119A>G). In addition, we identified biallelic variants in *RNU4-1*, a gene not previously directly associated with neurodevelopmental disorders. Although *RNU4-1* has been less extensively studied than *RNU4-2*, recent evidence indicates that it is functionally expressed, supporting a potential role in spliceosomal activity despite lower expression levels in the brain ^{1,6} and suggesting that its contribution to disease may become evident in specific genetic contexts. Nevertheless, interpretation is further limited by the fact that one of the two individuals also carries a de novo deletion affecting *MED13L*, considered the primary molecular diagnosis in that case. Together with the absence of functional validation, *RNU4-1* should therefore be regarded as a candidate gene pending replication in independent cohorts.

Notably, a recent report has described a deletion encompassing both *RNU4-2* and *RNU4-1*, associated with atypical developmental phenotype and raising the hypothesis that reduced U4 snRNA dosage may contribute to disease through haploinsufficiency mechanisms ²⁷. However, the absence of a neurodevelopmental phenotype in the father of the affected individual — who carries the heterozygous deletion in isolation — suggests that haploinsufficiency of *RNU4-2* alone may be insufficient to cause a clinically relevant phenotype. Together, these observations point to a model in which disease manifestation may depend on the combined perturbation of multiple spliceosomal components. In this context, our data further support the possibility that digenic or oligogenic mechanisms involving snRNA genes may contribute to phenotypic variability, as we observe an enrichment in individuals with NDDs compared to non-NDD controls.

These molecular differences translate into clinically distinct phenotypes. Deep phenotypic characterization, supported by standardized HPO coding and automated facial analysis, confirms that snRNA-related disorders represent clinically distinct

entities. In line with these observations and previous reports ^{3,4,28}, distinct phenotypic patterns emerge across genes, with *RNU4-2* variants associated with more consistent craniofacial and neuroanatomical features, *RNU2-2* variants more frequently linked to severe epileptic encephalopathy and developmental regression, and individuals with *RNU5B-1* variants showing greater phenotypic heterogeneity.

Importantly, our study also identifies a prenatal phenotype associated with *RNU4-2* variants, characterized by intrauterine growth restriction (IUGR), reduced birth weight, and mild ventriculomegaly, with a notable female predominance. These findings suggest that *RNU4-2* disruption may affect early developmental processes beyond the central nervous system and indicate that this gene should be considered in cases of unexplained prenatal growth restriction, particularly in females and when standard genetic testing is inconclusive. Similar prenatal manifestations have been reported in disorders affecting components of the minor spliceosome, such as *RNU4ATAC-opathies* ^{21,22} supporting a broader role of spliceosomal dysfunction in early human development ²³. However, as prenatal data were collected retrospectively across multiple centers, prospective studies will be needed to fully characterize the prevalence and spectrum of prenatal manifestations in *RNU4-2*-associated disorder.

At a broader level, this study underscores the importance of genome-wide approaches for capturing noncoding variation and highlights the clinical utility of coordinated national programs combining standardized phenotyping and data sharing ^{16,17}, which identified pathogenic snRNA variants in 2.58% of previously unsolved individuals with NDDs. Altogether, our findings expand the genetic and phenotypic landscape of snRNA-related disorders and support a model of context-dependent spliceosomal dysfunction, with future transcriptomic and epigenomic studies needed to further define the impact of allelic and domain-specific effects.

Acknowledgements

We thank all participants and families involved mainly in “Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología en Medicina Genómica (IMPACT-Genómica)” and the undiagnosed programs “Programes de Malalties Rares no Diagnosticades de Catalunya (URD-Cat)” and “Programa de Enfermedades No Diagnosticadas (ENoD-CIBERER)”. IMPACT-Genómica was supported by Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and the European Union European Regional Development Fund (IMP/00009), URD-Cat was supported by the Department of Health of Catalonia (Grant SLT002/16/00174). ENoD-CIBERER Program was funded by the Biomedical Network Research Center for Rare Diseases-CIBER-ER-ISCIII.

Competing Interests

Luis A. Pérez-Jurado is stock option holder of Veritas Genetics SL. The rest of authors declare no competing financial and non-financial conflicts of interests.

Data Availability Statement

The data that support the findings of this study are available within the article and its supplementary files. Additional data are available from the corresponding author upon reasonable request.

Author Contribution Statement

M.S-P. and E.N-M. contributed to conceptualization, bioinformatic and statistical analyses, and manuscript writing. Z.M., A.D. and B.M. contributed to conceptualization, data analysis, and manuscript writing. L.A.P-J. and B.M. supervised the study, contributed to conceptualization, data interpretation, and manuscript writing. L.A.P-J., Á.C. and P.L. acquired funding and provided resources. C.R-A., M.C., R.C., J.A., I.G.,

R.T., N.T-B. and V.A. performed bioinformatic and statistical analyses. All remaining authors contributed to patient recruitment, clinical evaluation, phenotypic data collection, and critical revision of the manuscript.

Ethical approval

Participants were enrolled through the clinical services of participating Spanish hospitals under protocols subsequently approved by the Instituto de Salud Carlos III Research Ethics Committee (CEI-PI01_2022) and endorsed by the institutional review boards of the participating centers. Written informed consent was obtained from all participants at pre-test counseling sessions, allowing the sharing of pseudonymized clinical information with international collaborators and researchers. Additional written consent to publish clinical data and photographs was also collected in accordance with local ethical guidelines and best practices.

Supplementary information

Supplementary information is available on European Journal of Human Genetics website

TABLE & FIGURES

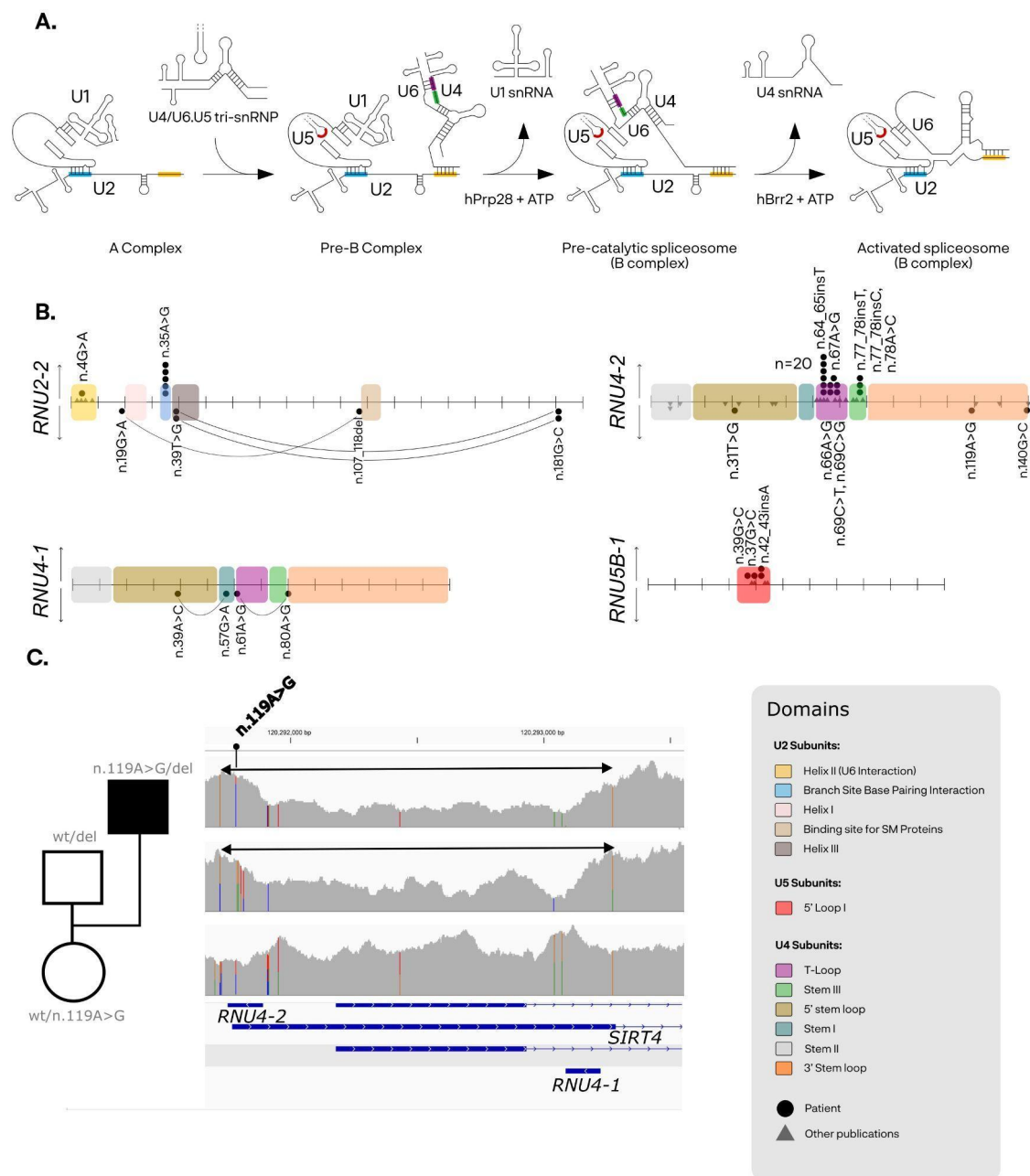
Table 1. Pathogenic variants detected in snRNA genes in NDD individuals from the Spanish URD

Programs. The table includes variant-related information, such as HGVS nomenclature, chromosomal location (GRCh38) and altered gene domain. It is also indicated the total number of variant carriers among Spanish undiagnosed NDDs, inheritance pattern and genotypes. Except for n.66A>G, which has an allele count (AC) of 1, all variants have ACs of 0 in gnomAD v4.1. Variants n.69G>C and n.77_78insC have not been previously described. *Previously reported. **This individual is also a carrier of a de novo deletion encompassing MED13L (GRCh38:12:115999405-116018153). ***This individual is also a carrier of a genomic duplication encompassing HNRNPH1 and RUFY1, reported as likely pathogenic ²⁹(Individual 3).
?: Unknown: non-paternal, maternal sample unavailable.

Gene	Variant	Position (GRCh38)	Domain	AC gnomADv4	Probands	Dominant de novo	Recessive biallelic	Genotype
RNU2-2	n.4G>A	chr11:62841806	U6 binding domain	0	1*	1	-	het
	n.35A>G	chr11:62841775	Branch site recognition domain	0	5*	5	-	het
RNU4-2	n.64_65insT	chr12:120291839	T-loop (quasi-pseudoknot)	0	20*	19+?	-	het
	n.66A>G	chr12:120291838	T-loop (quasi-pseudoknot)	1	2 (MZT)	2	-	het
	n.67A>G	chr12:120291837	T-loop (quasi-pseudoknot)	0	1	?	-	het
	n.69C>T	chr12:120291835	T-loop (RBM42 interaction)	0	1*	1	-	het
	n.69C>G	chr12:120291835	T-loop (RBM42 interaction)	0	1	1	-	het
	n.77_78insC	chr12:120291826	Stem III	0	1	1	-	het
	n.77_78insT	chr12:120291826	Stem III	0	1	1	-	het
	n.78A>C	chr12:120291826	Stem III	0	1	1	-	het
RNU5B-1	n.37G>C	chr15:65304713	5' loop I	0	1	1	-	het
	n.39C>G	chr15:65304715	5' loop I	0	1	1	-	het
	n.42_43insA	chr15:65304718	5' loop I	0	2	2	-	het
					38			
RNU2-2	n.19G>A	chr11:62841791	Branch site recognition domain	1	1	-	pat	comp het
	n.107_118del	chr11:62841679-62841690	Binding site SM proteins	4		-	mat	
	n.39T>G	chr11:62841771	Stem Ia	0	2 (Sibs)	-	pat	comp het

	n.181G>C	chr11:62841629	-	6		-	mat	
RNU4-1	n.39A>C	chr12:120293199	5' Stem loop	138	1	-	pat	comp het
	n.57G>A	chr12:120293181	Stem I	3		-	mat	
	n.61A>G	chr12:120293177	Stem I	35	1**	-	pat	comp het
	n.80A>G	chr12:120293158	Stem III	12		-	mat	
RNU4-2	n.31T>G	chr12:120291873	5' stem loop	0	1***	-	mat & pat	hom
	n.119A>G	chr12:120291785	3' Stem loop	8	1	-	mat	comp het
	1.778kdel	chr12:120291763-120293127	Entire U4-1 & U4-2 genes	0		-	pat	
	n.140G>C	chr12:120291764	3' Stem loop	1	1	-	Mat & pat	hom
					8			

629



632 **Figure 1 Schematic representation of pathogenic variants detected in snRNA genes in NDD**
633 **individuals from the Spanish URD Programs.** A. Schematic representation of the major spliceosome
634 complex illustrating the dynamic transitions through different intermediate stages: A complex, pre-B
635 complex, pre-catalytic B complex and activated spliceosome B complex. Highlighted in colour are the
636 domains of the snRNA genes (U2, U4, and U5) where the identified mutations are located. Adapted from
637 (Boesler et al. 2016 ³⁰) B. Variant distribution across snRNA genes in our cohort. The upper panel shows
638 de novo variants, while the lower panel displays biallelic variants. Lines connect variants found in the same
639 individual, representing combinatorial configurations. Colors indicate the structural domain in which each

640 variant is located. Circles represent individuals from our cohort, whereas triangles correspond to previously
641 reported variants from published studies. **C.** Visualization of the deletion identified in individual U4-p32,
642 which spans both RNU4-2 and RNU4-1 Genome browser views of the individual and both parents are shown
643 for comparison. The black pin marks the single nucleotide variant (SNV) located within the affected region.
644 The black arrows marks the deletion.

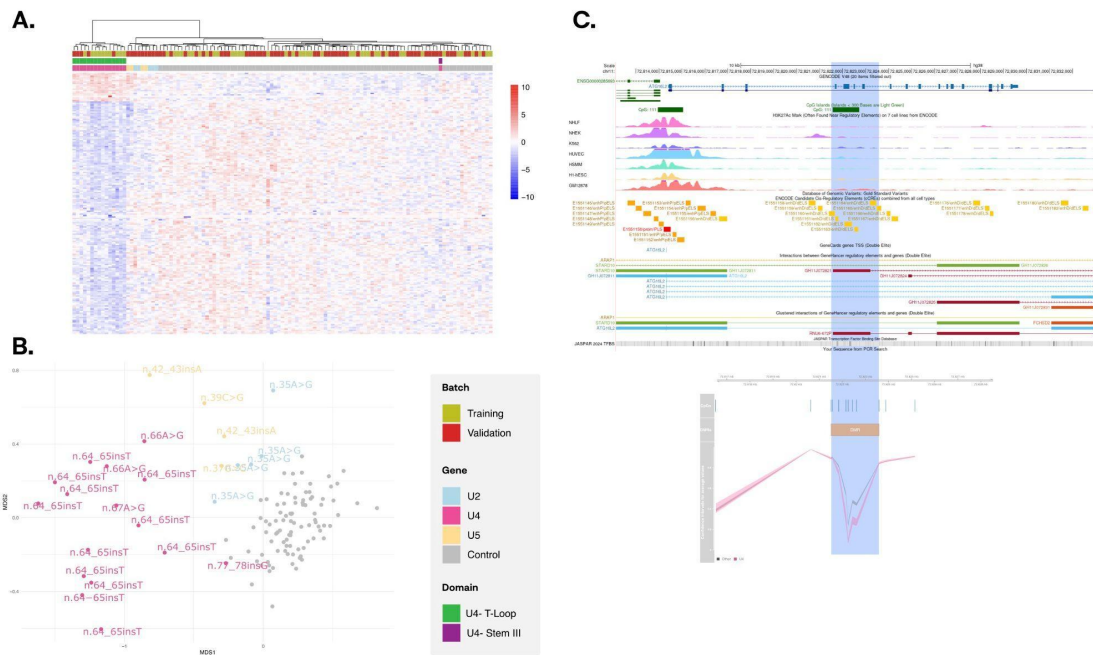
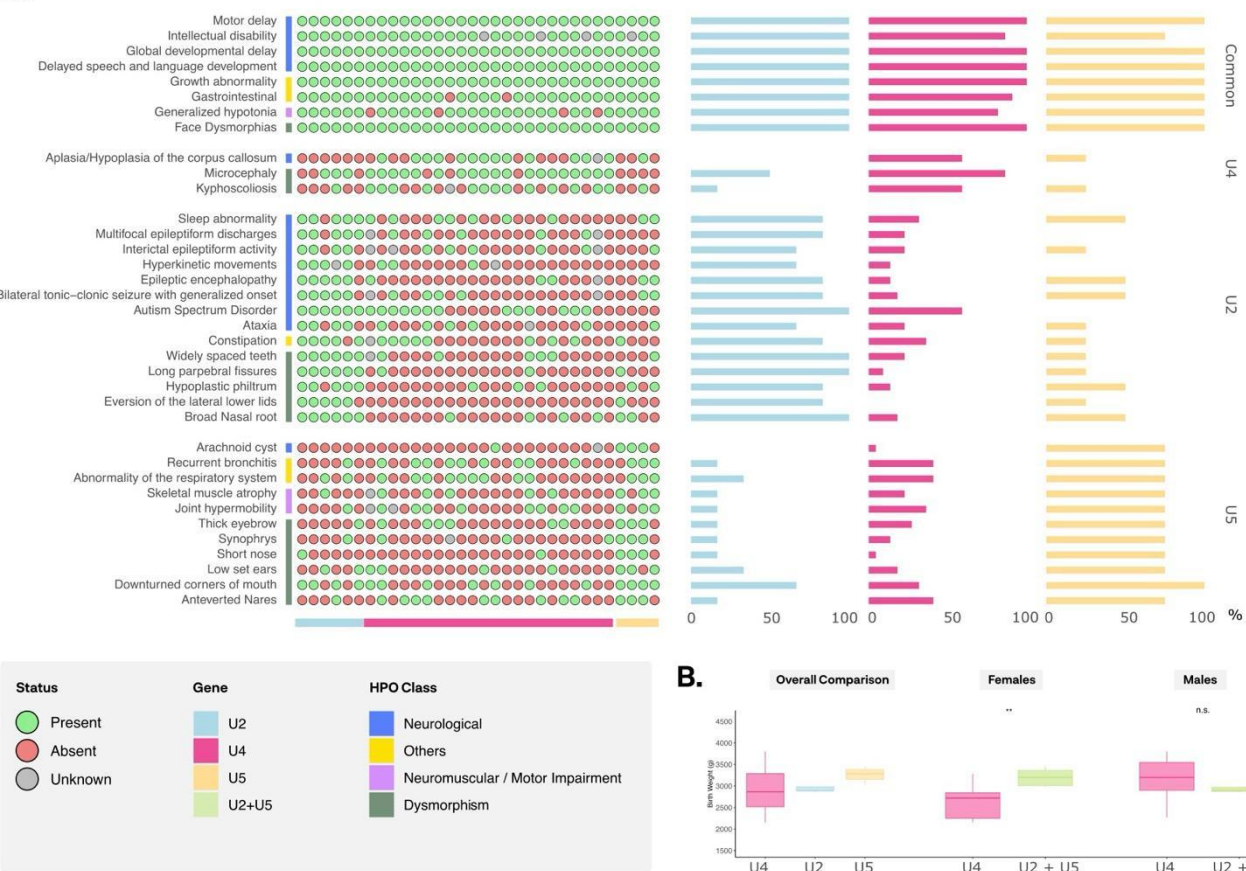


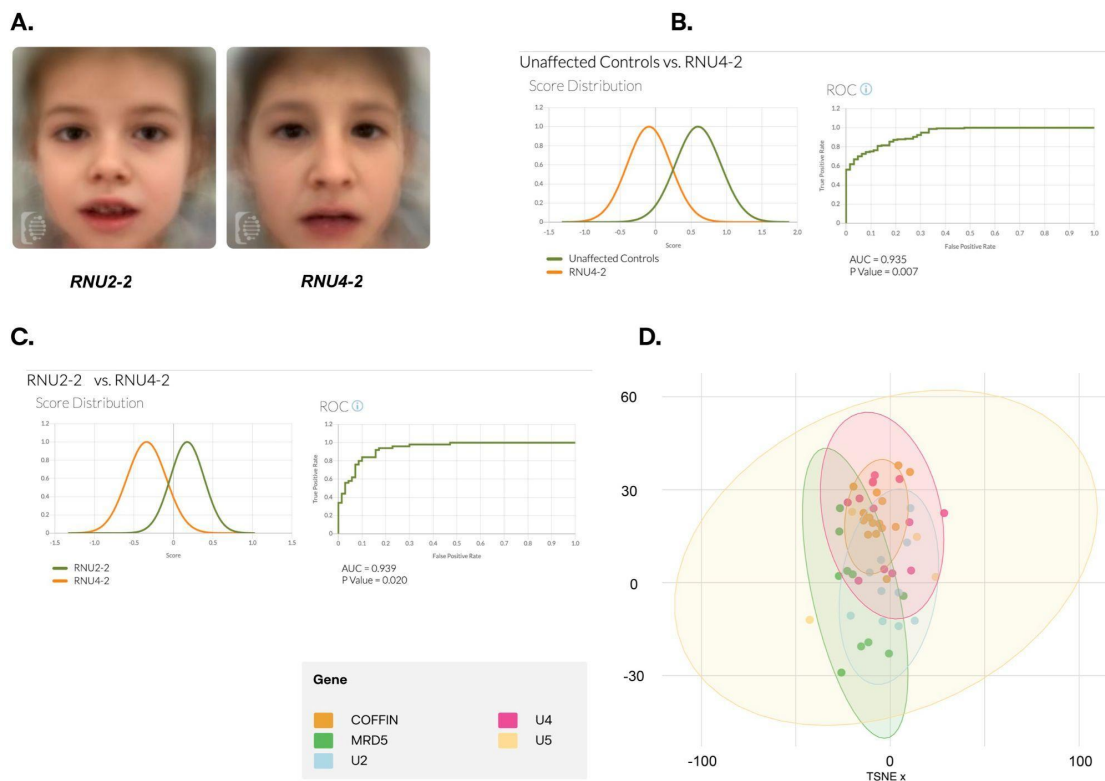
Figure 2 DNA methylation analyses in individuals with de novo RNU4-2 variants. **A.** Hierarchical clustering analysis using RNU4-2 (U4)-specific probes. Rows correspond to selected probes, while columns represent individual samples. Methylation levels range from blue (hypomethylated) to red (hypermethylated). RNU2-2 (U2), RNU4-2 (U4), RNU5B-1 (U5), and control groups are labelled in blue, pink, yellow, and grey, respectively. Samples are also categorized into training (red) and validation (green) cohorts. **B.** Multidimensional Scaling (MDS) plot of selected probes, with U2, U4, U5, and control groups represented in blue, pink, yellow, and grey, respectively. **C.** Differentially methylated region (DMR) identified in the promoter of RNU6-672P. Upper panel: UCSC Genome Browser screenshot displaying the genomic context of the DMR, including annotated genes (blue), CpG islands (green), H3K27Ac marks, ENCODE cis regulatory elements (transcription start sites (TSS; red), enhancers (E; yellow), and weak enhancers (WE; pale yellow)), GeneHancer annotations and JASPAR Transcription factor binding site. Lower panel: plot showing methylation levels (mean $\pm$ confidence intervals) for each CpG probe within the DMR in U4 individuals (pink) and others (gray).

A.



661 **Figure 3 Phenotypic spectrum and birth weight differences in individuals with de novo snRNA**
662 **variants.** A. Left panel: Human Phenotype Ontology (HPO) representation of the most common phenotypic
663 features observed in individuals with variants in each snRNA gene (RNU2-2 (U2), RNU4-2 (U4), and
664 RNU5B-1 (U5), along with the specific HPO terms unique to each group. Phenotypes are categorized into
665 major classes: neurological (blue), neuromuscular/motor impairment (purple), dysmorphic features (green),
666 and other (yellow). Right panel: Overall HPO frequency distribution across each group (U2, U4, and U5). B.
667 Birth weight distribution stratified by sex. After excluding preterm individuals, individuals with RNU4-2
668 variants were compared to those with variants in RNU2-2 or RNU5B-1 (grouped as RNU2-2 + RNU5B-1),
669 using two-sample t-tests for each comparison. $P < 0.05$ (*); not significant (n.s.).

670



671

672 **Figure 4 Facial analysis of individuals with de novo snRNA variants.** **A.** Composite facial masks
 673 generated using the FDNA system for individuals with variants in RNU4-2(U4), RNU2-2(U2), and unaffected
 674 individuals. **B-C.** Binary comparison results. **D.** t-SNE visualization illustrating the clustering of facial
 675 phenotypes across the cohort.

676

678

References

- 679 Chen Y, Dawes R, Kim HC, et al. De novo variants in the RNU4-2 snRNA
 680 cause a frequent neurodevelopmental syndrome. *Nature*. 2024;632(8026):832-
 681 840. doi:10.1038/s41586-024-07773-7
- 682 Greene D, Thys C, Berry IR, et al. Mutations in the U4 snRNA gene RNU4-2
 683 cause one of the most prevalent monogenic neurodevelopmental disorders. *Nat*
 684 *Med*. 2024;30(8):2165-2169. doi:10.1038/s41591-024-03085-5
- 685 Greene D, De Wispelaere K, Lees J, et al. Mutations in the small nuclear RNA
 686 gene RNU2-2 cause a severe neurodevelopmental disorder with prominent
 687 epilepsy. *Nat Genet*. 2025;57(6):1367-1373. doi:10.1038/s41588-025-02159-5
- 688 Jackson A, Thaker N, Blakes A, et al. Analysis of R-loop forming regions
 689 identifies RNU2-2 and RNU5B-1 as neurodevelopmental disorder genes. *Nat*
 690 *Genet*. 2025;57(6):1362-1366. doi:10.1038/s41588-025-02209-y
- 691 Nava C, Cogne B, Santini A, et al. Dominant variants in major spliceosome U4
 692 and U5 small nuclear RNA genes cause neurodevelopmental disorders through
 693 splicing disruption. *Nat Genet*. 2025;57(6):1374-1388. doi:10.1038/s41588-025-
 694 02184-4
- 695 Antonarakis SE. Small nuclear RNA genes in Mendelian disorders. *Nat Genet*.
 696 2026;58(1):28-38. doi:10.1038/s41588-025-02440-7
- 697 Gillentine MA, Wang T, Eichler EE. Estimating the Prevalence of De Novo
 698 Monogenic Neurodevelopmental Disorders from Large Cohort Studies.
 699 *Biomedicines*. 2022;10(11):2865. doi:10.3390/biomedicines10112865

780 Greene D, Mendez R, Lees J, et al. Biallelic variants in RNU2-2 cause the most
701 prevalent known recessive neurodevelopmental disorder. *Nat Genet.*
702 2026;58(4):774-781. doi:10.1038/s41588-026-02539-5

793 Rius R, Blakes AJM, Chen Y, et al. Biallelic variants in the noncoding RNA gene
704 RNU4-2 cause a recessive neurodevelopmental syndrome with distinct white
705 matter changes. *Nat Genet.* 2026;58(4):761-773. doi:10.1038/s41588-026-
706 02554-6

707 De Jonghe J, Kim HC, Adedeji A, et al. Saturation editing of RNU4-2 reveals
708 distinct dominant and recessive disorders. *Nature*. Published online 2026.
709 doi:10.1038/s41586-026-10334-9

710 Leitão E, Santini A, Cogne B, et al. Systematic analysis of snRNA genes
711 reveals frequent RNU2-2 variants in dominant and recessive developmental
712 and epileptic encephalopathies. *Nat Genet.* 2026;58(4):782-797.
713 doi:10.1038/s41588-026-02547-5

714 Henry OJ, Pacheco NP, Duba I, et al. Phenotypic and transcriptomic
715 characterisation of a novel biallelic RNU2-2 developmental
716 and epileptic encephalopathy. *medRxiv*. Published online January 1,
717 2026:2026.02.19.26345867. doi:10.64898/2026.02.19.26345867

718 He H, Liyanarachchi S, Akagi K, et al. Mutations in U4atac snRNA, a
719 component of the minor spliceosome, in the developmental disorder MOPD I.
720 *Science*. 2011;332(6026):238-240. doi:10.1126/science.1200587

721 Merico D, Roifman M, Braunschweig U, et al. Compound heterozygous
722 mutations in the noncoding RNU4ATAC cause Roifman Syndrome by disrupting
723 minor intron splicing. *Nat Commun*. 2015;6:8718. doi:10.1038/ncomms9718

715. Schot R, Ferraro F, Geeven G, Diderich KEM, Barakat TS. Re-analysis of whole
725 genome sequencing ends a diagnostic odyssey: Case report of an RNU4-2
726 related neurodevelopmental disorder. *Clin Genet.* 2024;106(4):512-517.
727 doi:<https://doi.org/10.1111/cge.14574>
716. Luque J, Mendes I, Gómez B, et al. CIBERER: Spanish national network for
729 research on rare diseases: A highly productive collaborative initiative. *Clin*
730 *Genet.* 2022;101(5-6):481-493. doi:10.1111/cge.14113
717. Bullich G, Matalonga L, Pujadas M, et al. Systematic Collaborative Reanalysis
732 of Genomic Data Improves Diagnostic Yield in Neurologic Rare Diseases. *The*
733 *Journal of Molecular Diagnostics.* 2022;24(5):529-542.
734 doi:<https://doi.org/10.1016/j.jmoldx.2022.02.003>
718. Gargano MA, Matentzoglou N, Coleman B, et al. The Human Phenotype
736 Ontology in 2024: phenotypes around the world. *Nucleic Acids Res.*
737 2024;52(D1):D1333-D1346. doi:10.1093/nar/gkad1005
719. Gurovich Y, Hanani Y, Bar O, et al. Identifying facial phenotypes of genetic
739 disorders using deep learning. *Nat Med.* 2019;25(1):60-64. doi:10.1038/s41591-
740 018-0279-0
720. Quinodoz M, Rodenburg K, Cvackova Z, et al. De novo and inherited dominant
742 variants in U4 and U6 snRNA genes cause retinitis pigmentosa. *Nat Genet.*
743 2026;58(1):169-179. doi:10.1038/s41588-025-02451-4
721. Goumy C, Gay-Bellile M, Salaun G, et al. A novel 2q14.1q14.3 deletion
745 involving GLI2 and RNU4ATAC genes associated with partial corpus callosum
746 agenesis and severe intrauterine growth retardation. *Birth Defects Res A Clin*
747 *Mol Teratol.* 2016;106(9):793-797. doi:<https://doi.org/10.1002/bdra.23535>

722. Putoux A, Alqahtani A, Pinson L, et al. Refining the phenotypical and mutational
749 spectrum of Taybi-Linder syndrome. *Clin Genet*. 2016;90(6):550-555.
750 doi:10.1111/cge.12781
723. Chen YS, He JF, Quan T, Li SB, Li DZ. Hydrocephalus and Growth Retardation:
752 A Fetal RNU4ATAC-opathy Missed by Whole-Exome Sequencing. *Mol*
753 *Syndromol*. 2023;13(6):522-526. doi:10.1159/000524501
724. Kim JS, He X, Liu J, et al. Systematic proteomics of endogenous human
755 cohesin reveals an interaction with diverse splicing factors and RNA-binding
756 proteins required for mitotic progression. *Journal of Biological Chemistry*.
757 2019;294(22):8760-8772. doi:10.1074/jbc.RA119.007832
725. Singh AK, Chen Q, Nguyen C, Meerzaman D, Singer DS. Cohesin regulates
759 alternative splicing. *Sci Adv*. 2025;9(9):eade3876. doi:10.1126/sciadv.ade3876
726. Deardorff MA, Noon SE, Krantz ID. Cornelia de Lange Syndrome. 日本小児放射
761 線学会. 2020;18(1):35-39. Accessed August 6, 2025.
762 <https://www.ncbi.nlm.nih.gov/books/NBK1104/>
727. Abaji M, Gerard B, Nasca L, et al. First Identification of a Heterozygous RNU4-2
764 and RNU4-1 Deletion Associated With Fetal Urogenital and Anorectal
765 Malformations. *Prenat Diagn*. 2026;46(4):607-609.
766 doi:<https://doi.org/10.1002/pd.70138>
728. Valenzuela I, Codina-Solà M, Vazquez E, et al. Deep phenotyping of 11
768 individuals with pathogenic variants in RNU4-2 reveals a clinically recognizable
769 syndrome. *Genet Med*. 2024;26(12):101288. doi:10.1016/j.gim.2024.101288
729. Reichert SC, Li R, A. Turner S, et al. HNRNPH1-related syndromic intellectual
771 disability: Seven additional cases suggestive of a distinct syndromic

772 neurodevelopmental syndrome. *Clin Genet.* 2020;98(1):91-98.
773 doi:<https://doi.org/10.1111/cge.13765>
774 30. Boesler C, Rigo N, Anokhina MM, et al. A spliceosome intermediate with loosely
775 associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat*
776 *Commun.* 2016;7(1):11997. doi:10.1038/ncomms11997

777

778

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SevillaMSupplementalTables.xlsx](#)
- [Supplementarymaterial.pdf](#)