

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

Caroline Nava<sup>1,2,\*#</sup>, Benjamin Cogne<sup>3,4,5\*</sup>, Amandine Santini<sup>6\*</sup>, Elsa Leitão<sup>7</sup>, François Lecoquierre<sup>8</sup>, Yuyang Chen<sup>9,10</sup>, Sarah L. Stenton<sup>11,12</sup>, Thomas Besnard<sup>3,4</sup>, Solveig Heide<sup>2</sup>, Sarah Baer<sup>13,14</sup>, Abhilasha Jakhar<sup>15,16</sup>, Sonja Neuser<sup>17</sup>, Boris Keren<sup>2,5</sup>, Anne Faudet<sup>2</sup>, Sylvie Forlani<sup>1</sup>, Marie Faoucher<sup>18,19</sup>, Kevin Uguen<sup>20,21</sup>, Konrad Platzer<sup>17</sup>, Alexandra Afenjar<sup>2,22</sup>, Jean-Luc Alessandri<sup>23</sup>, Stephanie Andres<sup>24</sup>, Chloé Angelini<sup>25</sup>, Bernard Aral<sup>26</sup>, Benoit Arveiler<sup>25,27</sup>, Tania Attie-Bitach<sup>28,29</sup>, Marion Aubert Mucca<sup>30</sup>, Guillaume Banneau<sup>30</sup>, Tahsin Stefan Barakat<sup>31,32</sup>, Giulia Barcia<sup>28,29</sup>, Stéphanie Baulac<sup>1</sup>, Claire Beneteau<sup>25</sup>, Fouzia Benkerdou<sup>2</sup>, Virginie Bernard<sup>33</sup>, Stéphane Bézieau<sup>3,4</sup>, Dominique Bonneau<sup>34,35</sup>, Marie-Noelle Bonnet-Dupeyron<sup>36</sup>, Simon Boussion<sup>37</sup>, Odile Boute<sup>37</sup>, Elise Brischoux-Boucher<sup>38</sup>, Samantha J. Bryen<sup>39,40</sup>, Julien Buratti<sup>2</sup>, Tiffany Busa<sup>41</sup>, Almuth Caliebe<sup>42</sup>, Yline Capri<sup>43</sup>, Kévin Cassinari<sup>8</sup>, Roseline Caumes<sup>37</sup>, Camille Cenni<sup>44</sup>, Pascal Chambon<sup>8</sup>, Perrine Charles<sup>2</sup>, John Christodoulou<sup>45,46</sup>, Cindy Colson<sup>37</sup>, Solène Conrad<sup>3</sup>, Auriane Cospain<sup>18</sup>, Juliette Coursimault<sup>8</sup>, Thomas Courtin<sup>28,29</sup>, Madeline Couse<sup>47</sup>, Charles Coutton<sup>48,49</sup>, Isabelle Creveaux<sup>50</sup>, Alissa M. D’Gama<sup>51,52,53,54</sup>, Benjamin Dauriat<sup>55</sup>, Jean-Madeleine de Sainte Agathe<sup>2</sup>, Giulia Del Gobbo<sup>56</sup>, Andree Delahaye-Duriez<sup>57,58,59</sup>, Julian Delanne<sup>60,61</sup>, Anne-Sophie Denommé-Pichon<sup>61,62</sup>, Anne Dieux-Coeslier<sup>37</sup>, Laura Do Souto Ferreira<sup>3</sup>, Martine Doco-Fenzy<sup>63,3</sup>, Stephan Drukewitz<sup>17</sup>, Véronique Duboc<sup>64</sup>, Christèle Dubourg<sup>18,19</sup>, Yannis Duffourd<sup>61,62</sup>, David Dymont<sup>56</sup>, Salima El Chehadeh<sup>65</sup>, Monique Elmaleh<sup>66</sup>, Laurence Faivre<sup>60,61</sup>, Samuel Fennelly<sup>25</sup>, Hanna Fischer<sup>67</sup>, Mélanie Fradin<sup>68</sup>, Camille Galludec Vaillant<sup>2</sup>, Benjamin Ganne<sup>69</sup>, Jamal Ghoumid<sup>37</sup>, Himanshu Goel<sup>70,71</sup>, Zeynep Gokce-Samar<sup>72,73</sup>, Alice Goldenberg<sup>8</sup>, Romain Gonfreville Robert<sup>3</sup>, Louise Goujon<sup>43</sup>, Victoria Granier<sup>30</sup>, Mathilde Gras<sup>2</sup>, John M. Grealley<sup>74</sup>, Bianca Greiten<sup>42</sup>, Paul Gueguen<sup>75,76</sup>, Anne-Marie Guerrot<sup>8</sup>, Saurav Guha<sup>77</sup>, Anne Guimier<sup>28,29</sup>, Tobias Haack<sup>78</sup>, Hamza HadjAbdallah<sup>28,29</sup>, Yosra Halleb<sup>79</sup>, Radu Harbuz<sup>48,49</sup>, Madeleine Harris<sup>46</sup>, Julia Hentschel<sup>17</sup>, Bénédicte Héron<sup>80</sup>, Marc-Phillip Hitz<sup>42,81</sup>, A. Micheil Innes<sup>82</sup>, Vincent Jadas<sup>83</sup>, Louis Januel<sup>84</sup>, Nolwenn Jean-Marçais<sup>68</sup>, Vaidehi Jobanputra<sup>77,85</sup>, Florence Jobic<sup>86</sup>, Ludmila Jornea<sup>1</sup>, Céline Jost<sup>60,61</sup>, Sophie Julia<sup>30</sup>, Frank J. Kaiser<sup>7,87</sup>, Daniel Kaschta<sup>42</sup>, Sabine Kaya<sup>7</sup>, Petra Ketteler<sup>7,88</sup>, Bochra Khadija<sup>89,90,7</sup>, Fabian Kilpert<sup>7</sup>, Cordula Knopp<sup>91</sup>, Florian Kraft<sup>91</sup>, Iлона Krey<sup>17</sup>, Marilyn Lackmy<sup>92</sup>, Fanny Laffargue<sup>93</sup>, Laetitia Lambert<sup>94,95</sup>, Ryan Lamont<sup>82,96</sup>, Vincent Laugel<sup>13</sup>, Steven Laurie<sup>97</sup>, Julie L. Lauzon<sup>82,96</sup>, Louis Lebreton<sup>98</sup>, Marine Lebrun<sup>99</sup>, Marine Legendre<sup>25</sup>, Eric Leguern<sup>1,2</sup>, Daphné Lehalle<sup>2</sup>, Elodie Lejeune<sup>2</sup>, Gaetan Lesca<sup>84,100</sup>, Marion Lesieur-Sebellin<sup>28,29</sup>, Jonathan Levy<sup>43</sup>, Agnès Linglart<sup>101,102,103</sup>, Stanislas Lyonnet<sup>28,29</sup>, Kevin Lüthy<sup>7</sup>, Alan S. Ma<sup>104,105</sup>, Corinne Mach<sup>2</sup>, Jean-Louis Mandel<sup>14</sup>, Lamisse Mansour-Hendili<sup>5</sup>, Julien Marcadier<sup>82</sup>, Victor Marin<sup>98</sup>, Henri Margot<sup>25,27</sup>, Valentine Marquet<sup>55</sup>, Angèle May<sup>8</sup>, Johannes A. Mayr<sup>106</sup>, Vincent Michaud<sup>25,27</sup>, Caroline Michot<sup>28,29</sup>, Gwenael Nadeau<sup>107</sup>, Sophie Naudion<sup>25</sup>, Laetitia Nguyen<sup>2</sup>, Mathilde Nizon<sup>3</sup>, Frédérique Nowak<sup>108</sup>, Sylvie Odent<sup>68</sup>, Valerie Olin<sup>2</sup>, Ikeoluwa A. Osei-Owusu<sup>11</sup>, Matthew Osmond<sup>56</sup>, Katrin Öunap<sup>109,110</sup>, Laurent Pasquier<sup>111</sup>, Sandrine Passemard<sup>112,113</sup>, Olivier Patat<sup>30</sup>, Marine Pensec<sup>21</sup>, Laurence Perrin-Sabourin<sup>43</sup>, Florence Petit<sup>37</sup>, Christophe Philippe<sup>114</sup>, Marc Planes<sup>21</sup>, Annapurna Poduri<sup>51,54,115</sup>, Céline Poirsier<sup>116</sup>, Antoine Pouzet<sup>43</sup>, Bradley Prince<sup>82</sup>, Clément Prouteau<sup>34,35</sup>, Aurora Pujol<sup>117,118,119</sup>, Caroline Racine<sup>60,61</sup>, Mélanie Rama<sup>37</sup>, Francis Ramond<sup>120</sup>, Kara Ranguin<sup>121</sup>, Margaux Raway<sup>38</sup>, Mathilde Renaud<sup>122,123</sup>, Nicole Revencu<sup>124</sup>, Anne-Claire Richard<sup>8</sup>, Lucile Riera-Navarro<sup>64</sup>, Rocio Rius<sup>45,39,40</sup>, Diana Rodriguez<sup>125,126</sup>, Agustí Rodriguez-Palmero<sup>127</sup>, Sophie Rondeau<sup>28,29</sup>, Annika Roser-Unruh<sup>128</sup>, Christelle Rougeot Jung<sup>129</sup>, Hana Safraru<sup>61,62</sup>, Véronique Satre<sup>48,49</sup>, Pascale Saugier-Veber<sup>8</sup>, Clément Sauvestre<sup>25</sup>, Elise Schaefer<sup>65</sup>, Wanqing Shao<sup>54</sup>, Ina Schanze<sup>130</sup>, Jan-Ulrich Schlump<sup>131</sup>, Agatha Schlüter Martin<sup>117,118</sup>, Caroline Schluth-Bolard<sup>132,133</sup>, Christopher Schröder<sup>7</sup>, Monisha Sebastin<sup>74</sup>, Sabine Sigaudy<sup>134</sup>, Malte Spielmann<sup>42</sup>, Marta Spodenkiewicz<sup>135</sup>, Laura St Clair<sup>104</sup>, Julie Steffann<sup>28,29</sup>, Radka Stoeva<sup>79</sup>, Harald Surowy<sup>7</sup>, Mark A. Tarnopolsky<sup>136</sup>, Calina Todosi<sup>137,138</sup>, Annick Toutain<sup>139</sup>, Frédéric Tran Mau-Them<sup>62,61,140</sup>, Astrid Unterlauff<sup>141</sup>, Julien Van-Gils<sup>25,27</sup>, Clémence Vanlerberghe<sup>37</sup>, Gabriella Vera<sup>8</sup>, André Verdel<sup>142</sup>, Alain Verloes<sup>43,113,143</sup>, Yoann Vial<sup>43,113</sup>, Cédric Vignal<sup>43</sup>, Marie Vincent<sup>3</sup>, Catherine Vincent-Delorme<sup>37</sup>, Antonio Vitobello<sup>62,61</sup>, Sacha Weber<sup>144</sup>, Marjolaine Willems<sup>145</sup>, Khaoula Zaafrane-Khachnaoui<sup>64</sup>, Pia Zacher<sup>146</sup>, Lena Zeltner<sup>147</sup>, Alban Ziegler<sup>148</sup>, Wojciech P. Galej<sup>149</sup>, Hélène Dollfus<sup>65,133</sup>, Christel Thauvin<sup>60,62,61</sup>, Kym M. Boycott<sup>56,150</sup>, Pierre Marijon<sup>5</sup>, Alban Lermine<sup>5</sup>, Valérie Malan<sup>28,29</sup>, Marlène Rio<sup>28,29</sup>, Alma Kuechler<sup>7</sup>, Bertrand Isidor<sup>3</sup>, Séverine Drunat<sup>43,113</sup>, Thomas Smol<sup>37</sup>, Nicolas Chatron<sup>84,100</sup>, Amélie Piton<sup>132,14</sup>, Gael Nicolas<sup>8</sup>, Matias Wagner<sup>151,152,153</sup>, Rami Abou Jamra<sup>17</sup>, Delphine Héron<sup>2,143</sup>, Cyril Mignot<sup>2,143</sup>, Pierre Blanc<sup>5</sup>, Anne O’Donnell-Luria<sup>11,12,154</sup>, Nicola Whiffin<sup>9,10,11</sup>, Camille Charbonnier<sup>155</sup>, Clément Charenton<sup>15,16</sup>, Julien Thevenon<sup>48,49,338</sup>, Christel Depienne<sup>78#</sup>

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

\*co-first authors

§co-last authors

#Correspondence to [caroline.nava@aphp.fr](mailto:caroline.nava@aphp.fr) or [christel.depienne@uk-essen.de](mailto:christel.depienne@uk-essen.de)

## Affiliations

<sup>1</sup>Sorbonne Université, Institut du Cerveau-Paris Brain Institute-ICM, Inserm, CNRS, Hôpital Pitié Salpêtrière, Paris, France, <sup>2</sup>Assistance Publique- Hôpitaux de Paris (APHP) Sorbonne Université, Département de Génétique, Hôpital Pitié-Salpêtrière, Paris, France, <sup>3</sup>Nantes Université, CHU de Nantes, Service de Génétique médicale, Nantes, France, <sup>4</sup>Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, Nantes, France, <sup>5</sup>laboratoire GCS SeqOIA, Paris, France, <sup>6</sup>Univ Rouen Normandie, Normandie Univ, Inserm U1245, Rouen, France, <sup>7</sup>Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen, Germany, Essen, Germany, <sup>8</sup>Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, Department of Genetics and Reference Center for Developmental Abnormalities, Rouen, France, <sup>9</sup>Big Data Institute, University of Oxford, Oxford, UK, <sup>10</sup>Centre for Human Genetics, University of Oxford, Oxford, UK, <sup>11</sup>Broad Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA, <sup>12</sup>Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA, <sup>13</sup>Service de neuropédiatrie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, <sup>14</sup>Institute of Genetics and Cellular and Molecular Biology (IGBMC), INSERM-U964, CNRS-UMR7104, University of Strasbourg, Illkirch, France, <sup>15</sup>CNRS, Inserm, Université de Strasbourg, IGBMC UMR 7104- UMR-S 1258, Illkirch, France, <sup>16</sup>Department of Integrated Structural Biology, IGBMC, Illkirch, France, <sup>17</sup>Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany, <sup>18</sup>Laboratoire de Génétique Moléculaire et Génomique, FHU GenOMedS, CHU Rennes, Rennes, France, <sup>19</sup>Univ Rennes, CNRS, INSERM, IGDR (Institut de Génétique et Développement de Rennes)-UMR 6290, Rennes, France, <sup>20</sup>Univ Brest, Inserm, EFS, UMR 1078, GGB, Brest, France, <sup>21</sup>Service de Génétique Médicale et Biologie de la Reproduction, CHRU de Brest, Brest, France, <sup>22</sup>Centre de Référence Malformations et maladies congénitales du cervelet et déficiences intellectuelles, Paris, France, <sup>23</sup>Service de Génétique, CHU de La Réunion, Saint-Pierre, France, <sup>24</sup>Medicover München Ost MVZ, Humangenetik, Munich, Germany, <sup>25</sup>Service de génétique médicale, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France, <sup>26</sup>Laboratoire de génétique chromosomique et moléculaire, Pôle Biologie, CHU de Dijon, Dijon, France, <sup>27</sup>INSERM U1211, University of Bordeaux, Bordeaux, France, <sup>28</sup>Assistance Publique- Hôpitaux de Paris (APHP), Service de Médecine Génomique des Maladies Rares, Hôpital Necker-Enfants malades, Paris, France, <sup>29</sup>Université Paris Cité, INSERM, IHU Imagine – Institut des maladies génétiques, Paris, France, <sup>30</sup>Service de Génétique médicale, CHU Purpan, Toulouse, Toulouse, France, <sup>31</sup>Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands, <sup>32</sup>Discovery Unit, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands, <sup>33</sup>GCS AURAGEN, Lyon, France, <sup>34</sup>Department of Genetics, Angers University Hospital, Angers, France, <sup>35</sup>UMR CNRS 6214 - INSERM 1083, Angers, France, <sup>36</sup>Consultations de Génétique, Centre Hospitalier de Valence, Valence, France, <sup>37</sup>Univ. Lille, CHU Lille, ULR7364 – RADEME, Lille, France, <sup>38</sup>Centre de Génétique Humaine - CHU Besançon, Université de Bourgogne-Franche-Comté, Besançon, France, <sup>39</sup>Centre for Population Genomics, Garvan Institute of Medical Research, and UNSW Sydney, Sydney, New South Wales, Australia, <sup>40</sup>Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia, <sup>41</sup>Service de génétique clinique, Hôpital Timone Enfants, APHM, Marseille, France, <sup>42</sup>Institute of Human Genetics, University Hospitals Schleswig-Holstein, University of Lübeck and Kiel University, Lübeck, Kiel, Germany, <sup>43</sup>Assistance Publique- Hôpitaux de Paris (APHP), Département de Génétique, Hôpital Robert-Debré, Paris, France, <sup>44</sup>Service de Génétique Moléculaire, Chromosomique et Clinique, CHU de Nîmes, Nîmes, France, <sup>45</sup>Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia, <sup>46</sup>Murdoch Children's Research Institute, Parkville, Victoria, Australia, <sup>47</sup>Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada, <sup>48</sup>Service de Génétique, Génomique et Procréation, CHU Grenoble Alpes, Grenoble, France, <sup>49</sup>Université Grenoble Alpes, INSERM U 1209, CNRS UMR 5309, Institut for Advanced Biosciences, Grenoble, France, <sup>50</sup>CHU Clermont-Ferrand, Department of Medical Biochemistry and Molecular Biology, Clermont-Ferrand, France, <sup>51</sup>Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital,

Boston, MA, USA, <sup>52</sup>Division of Newborn Medicine, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA, <sup>53</sup>Department of Pediatrics, Harvard Medical School, Boston, MA, USA, <sup>54</sup>Children's Rare Disease Collaborative, Boston Children's Hospital, Boston, MA, USA, <sup>55</sup>Service de génétique médicale, cytogénétique et biologie de la reproduction, CHU de Limoges, Limoges, France, <sup>56</sup>Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada, <sup>57</sup>Hôpitaux Universitaires de Paris Seine-Saint-Denis-APHP, UF de médecine génomique et génétique Clinique, Hôpital Jean Verdier, Bondy, France, <sup>58</sup>UFR Santé Médecine et Biologie Humaine, Université Sorbonne Paris Nord, Bobigny, France, <sup>59</sup>Inserm UMR1141 NeuroDiderot, Université Paris Cité, Paris, France, <sup>60</sup>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, <sup>61</sup>Inserm UMR1231 GAD, Université Bourgogne, Dijon, France, <sup>62</sup>Laboratoire de Génomique médicale – Centre NEOMICS, CHU Dijon Bourgogne, Dijon, France, <sup>63</sup>Service de Génétique, CHU de Reims, Reims, France, <sup>64</sup>Université Côte d'Azur, Centre Hospitalier Universitaire de Nice, Inserm U1081, CNRS UMR7284, IRCAN, Nice, France, <sup>65</sup>Service de génétique médicale, Institut de Génétique Médicale d'Alsace (IGMA), CHU Strasbourg, Strasbourg, France, <sup>66</sup>Assistance Publique - Hôpitaux de Paris (AP-HP), DMU DREAM, Hôpital Robert Debré, Service de Radiologie Pédiatrique, Paris, France, <sup>67</sup>Sozialpädiatrisches Zentrum Konstanz, Konstanz, Germany, <sup>68</sup>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, <sup>69</sup>Laboratoire de Génétique chromosomique, CHU de Montpellier, Montpellier, France, <sup>70</sup>General Genetics Service, Hunter Genetics, Waratah, New South Wales, Australia, <sup>71</sup>School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, New South Wales, Australia, <sup>72</sup>Department of Epilepsy, Sleep and Pediatric Neurophysiology, Hospices Civils de Lyon, Lyon, France, <sup>73</sup>University of Lyon, Lyon, France, <sup>74</sup>Department of Pediatrics, Division of Pediatric Genetic, Medicine, Children's Hospital at Montefiore/Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, New York, New York, USA, <sup>75</sup>Service de Génétique, CHU de Tours, Tours, France, <sup>76</sup>Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBrain U1253, Tours, France, <sup>77</sup>Molecular Diagnostics, New York Genome Center, New York, New York, USA, <sup>78</sup>Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany, <sup>79</sup>Service de Génétique Médicale, Centre Hospitalier du Mans, Le Mans, France, <sup>80</sup>Neurologie Pédiatrique, Hôpital Armand Trousseau-La Roche Guyon, Fédération Hospitalo-Universitaire I2-D2, Sorbonne-Université, AP-HP, Paris, France, <sup>81</sup>Institute of Medical Genetics, University Medicine Oldenburg, Carl von Ossietzky University, Oldenburg, Germany, <sup>82</sup>Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada, <sup>83</sup>service de Pédiatrie, CHMS Chambéry, CAMSP Chambéry, Chambéry, France, <sup>84</sup>Genetics Department, Hospices Civils de Lyon, Lyon, France, <sup>85</sup>Department of Pathology, Columbia University Irving Medical Center, New York, New York, USA, <sup>86</sup>Clinical Genetics Unit, University Hospital of Amiens, Amiens, France, <sup>87</sup>Essen Center for Rare Diseases (EZSE), Essen, Germany, <sup>88</sup>Department of Pediatrics III, University Hospital Essen, Essen, Germany, <sup>89</sup>Higher Institute of Biotechnology of Monastir, University of Monastir, Monastir, Tunisia, <sup>90</sup>Cytogenetics, Molecular Genetics and Human Reproductive Biology Laboratory, Service of Genetics CHU Farhat Hached Sousse, Sousse, Tunisia, <sup>91</sup>Institute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University Hospital, Aachen, Germany, <sup>92</sup>Unité de Génétique Clinique, Centre de Compétences Maladies Rares Anomalies du Développement, CHRU de Pointe à Pitre, Guadeloupe, France, <sup>93</sup>Service de Génétique Médicale, CHU de Clermont-Ferrand, Clermont-Ferrand, France, <sup>94</sup>Service de Génétique Clinique, CHRU Nancy, Nancy, France, <sup>95</sup>Université de Lorraine, INSERM UMR\_S1256, NGERE, Nancy, France, <sup>96</sup>Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada, <sup>97</sup>Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Catalonia, Spain, <sup>98</sup>Service de biochimie, Hôpital Pellegrin, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France, <sup>99</sup>Département de Génétique, Centre Hospitalier Universitaire de Saint-Etienne, Saint Etienne, France, <sup>100</sup>Pathophysiology and Genetics of Neuron and Muscle (PNMG), UCBL, CNRS UMR5261 - INSERM, U1315, Lyon, France, <sup>101</sup>Assistance Publique- Hôpitaux de Paris (APHP), Department of Endocrinology and Diabetology for Children, Paris, France, <sup>102</sup>Department of Adolescent Medicine, Bicetre Paris-Saclay University Hospital, Le Kremlin Bicetre, France, <sup>103</sup>Paris Saclay University, Paris, France, <sup>104</sup>Department of Clinical Genetics, Sydney Children's Hospitals Network Westmead, Sydney, New South Wales, Australia, <sup>105</sup>Specialty

of Genomic Medicine, University of Sydney, Sydney, New South Wales, Australia, <sup>106</sup>University Children's Hospital, Salzburger Landeskliniken (SALK) and Paracelsus Medical University, Salzburg, Austria, <sup>107</sup>Laboratoire de Cytogénétique, CH de Chambéry, Chambéry, France, <sup>108</sup>Health Technologies Institute, Inserm, Paris, France, <sup>109</sup>Genetics and Personalized Medicine Clinic, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia, <sup>110</sup>Department of Clinical Genetics, Genetics and Personalized Medicine Clinic, Tartu University Hospital, Tartu, Estonia, <sup>111</sup>Clinical Genetics, Reference Center for Intellectual Disabilities, FHU GenOMedS, Rennes University Hospital, Rennes, France, <sup>112</sup>Assistance Publique - Hôpitaux de Paris (AP-HP), DMU INOV-RDB, Hôpital Robert Debré, Service de Neurologie Pédiatrique, Paris, France, <sup>113</sup>NeuroDiderot, Inserm, Université Paris Cité, Paris, France, <sup>114</sup>Laboratoire de Génétique Médicale, CHR Metz-Thionville, Hôpital Mercy, Metz, France, <sup>115</sup>Department of Neurology, Harvard Medical School, Boston, MA, USA, <sup>116</sup>UF de Génétique clinique, CHU de Reims, Reims, France, <sup>117</sup>Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain, <sup>118</sup>CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Madrid, Spain, <sup>119</sup>Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain, <sup>120</sup>Département de Génétique, Centre Hospitalier Universitaire de Saint-Etienne, Saint-Etienne, France, <sup>121</sup>Unité de génétique - Centre Hospitalier Universitaire de Guadeloupe, Guadeloupe, France, <sup>122</sup>Department of Neurology, CHRU Nancy, Nancy, France, <sup>123</sup>Department of Medical Genetics, CHRU de Nancy, Nancy, France, <sup>124</sup>Center for Human Genetics, Cliniques universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium, <sup>125</sup>Assistance Publique- Hôpitaux de Paris (APHP) Sorbonne Université, Department of Pediatric Neurology, Armand-Trousseau Hospital, Paris, France, <sup>126</sup>Reference Center for Rare Diseases and Intellectual Deficiencies of Rare Causes, Paris, France, <sup>127</sup>Paediatric Neurology Unit, Department of Pediatrics, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Barcelona, Spain, <sup>128</sup>LMU Klinikum - München, Munich, Germany, <sup>129</sup>Hospices Civils de Lyon, Hôpital Femme Mère Enfant, Service de Neuropédiatrie, Bron cedex F-69677, Lyon, France, <sup>130</sup>Institute of Human Genetics, Magdeburg, Germany, <sup>131</sup>Department of Pediatrics, Centre for Neuromedicine, Gemeinschaftskrankenhaus Herdecke Gerhard-Kienle-Weg, Herdecke, Germany, <sup>132</sup>Laboratoire de Diagnostic Génétique, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, <sup>133</sup>UMRS 1112, Université de Strasbourg, Strasbourg, France, <sup>134</sup>Département de Génétique Médicale, Unité de Génétique Clinique Prénatale, CHU de Marseille-Hôpital de La Timone, Marseille, France, <sup>135</sup>Department of Genetics, La Réunion University Hospital, Saint-Pierre, France, <sup>136</sup>Department of Pediatrics, McMaster Children's Hospital, McMaster University, Hamilton, Ontario, Canada, <sup>137</sup>Centre de Référence des Epilepsies Rares (CRÉER) Centre Hospitalier Régional Universitaire, Nancy, France, <sup>138</sup>Laboratoire de Génétique, Centre Hospitalier Régional Universitaire, Nancy, France, <sup>139</sup>Génétique Médicale, Centre Hospitalier Universitaire; Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBrain U1253, Tours, France, <sup>140</sup>INSERM UMR 1231, Génétique des Anomalies du Développement, Université de Bourgogne Franche-Comté, Dijon, France, <sup>141</sup>Department of Neurology, University of Leipzig Medical Center, Leipzig, Germany, <sup>142</sup>Université Grenoble Alpes, INSERM U 1209, CNRS UMR 5309, Team RNA, Epigenetics & Stress, Institut for Advanced Biosciences, Grenoble, France, <sup>143</sup>Centre de Référence Déficiences Intellectuelles de Causes Rares, Paris, France, <sup>144</sup>Service de Génétique, CHU Caen, Caen, France, <sup>145</sup>Service de Génétique Médicale, CHU de Montpellier, Institute for Neurosciences of Montpellier, Univ Montpellier, INSERM, Montpellier, France, <sup>146</sup>Epilepsy Center Kleinwachau, Radeberg, Germany, <sup>147</sup>Department of Neurology, University Hospital Tübingen, Tübingen, Germany, <sup>148</sup>Service de Génétique, CRMR AnDDI-Rares, CHU Reims, Reims, France, <sup>149</sup>European Molecular Biology Laboratory, EMBL Grenoble, Grenoble, France, <sup>150</sup>Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada, <sup>151</sup>Division of Pediatric Neurology, Developmental Medicine and Social Pediatrics, Department of Pediatrics, Munich University Hospital, Munich, Germany, <sup>152</sup>Institute of Human Genetics, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany, <sup>153</sup>Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany, <sup>154</sup>Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, <sup>155</sup>Univ Rouen Normandie, Normandie Univ, Inserm U1245 and CHU Rouen, department of Biostatistics and Reference Center for Developmental Abnormalities, Rouen, France.

## Abstract

Variants in *RNU4-2*, encoding the small nuclear RNA (snRNA) U4, were recently identified as a major cause of neurodevelopmental disorders (ReNU syndrome). Here, we investigated *de novo* variants in 50 snRNAs in a French cohort of 23,649 individuals with rare disorders and collected data of additional patients through an international collaboration. Altogether, we identified 133 probands with pathogenic or likely pathogenic variants in *RNU4-2* and 15 individuals with *de novo* and/or recurrent variants in constrained regions of *RNU5B-1*, one of five genes encoding U5. These variants cluster in evolutionarily conserved regions of U4 and U5 essential for splicing. *RNU4-2* variants affecting stem III are associated with milder phenotypes than those in the T-loop (quasi-pseudoknot). Phaseable variants associated with severe phenotypes occurred on the maternal allele. Individuals with *RNU4-2* variants show specific defects in alternative 5' splice site usage, correlating with variant location and clinical severity. Additionally, we report an epismutation associated with severe ReNU syndrome. This study further highlights the importance of *de novo* variants in snRNAs and establishes *RNU5B-1* as a new neurodevelopmental disorder gene.

## Introduction

The splicing of pre-mRNA into mature mRNA in eukaryotic cells consists of excising introns and ligating exons. The two transesterification reactions necessary for this process are carried out by a large ribonucleoprotein (RNP) complex called the spliceosome<sup>1,2</sup>. This 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<sup>3</sup>. 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 and branch points, while U4, U5, and U6 form the tri-snRNP complex that is recruited to assemble a precatalytic spliceosome complex. U6 is initially maintained in an inactive conformation by pairing with U4. Dissociation of the U4/U6 interaction allows U6 to interact with U2 and form the catalytic site<sup>4</sup>. U5 is responsible for aligning the exons for ligation by binding to the 5' and 3' splice sites, ensuring accurate exon joining<sup>5</sup>.

Spliceosomal snRNAs are ubiquitously expressed and encoded by distinct single-exon genes transcribed by polymerase II and/or polymerase III<sup>6</sup>. Genes encoding snRNAs U1, U2, U4, U5 and U6 in human genomes are present in multiple copies, some of which are functional and others are pseudogenes<sup>7,8</sup>. After transcription, snRNAs undergo extensive processing steps essential for their stability and function. These include 5'-capping, 3'-end processing, nuclear export, binding to the Sm protein/SMN complex via their conserved Sm site to form the core snRNP structure, nuclear re-import, and nucleotide modifications (2'-O-methylation, and pseudouridylation) guided by small Cajal body-specific RNAs (scaRNAs)<sup>9-12</sup>.

A recent landmark discovery has implicated *de novo* variants in *RNU4-2*, one of two functional genes encoding U4, as a major cause of a neurodevelopmental disorder (NDD) called ReNU syndrome (OMIM#620851)<sup>13,14</sup>. Genome sequencing is necessary to detect these variants, which are typically not captured by current exome sequencing. The identification of variants in *RNU4-2* was facilitated by the high recurrence of a single base insertion (n.64\_65insT), which represented 78% of all pathogenic variants identified in patients. This variant is statistically enriched in the Genomics England (GEL)<sup>15</sup> NDD cohort while absent from gnomAD<sup>16</sup> and highly depleted in UK Biobank<sup>17</sup>. The analysis of 28 other brain-

expressed snRNA genes did not reveal any enrichment of *de novo* variants similar to *RNU4-2*, although 14 regions of 13 genes appear more evolutionary constrained<sup>13</sup>. These results raise the question 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 snRNA genes in a French cohort of 23,649 patients with various rare disorders and collected variants in additional cohorts via international collaborations. Using these data, we implicated one further snRNA in NDDs and more comprehensively defined ReNU syndrome. Additionally, we investigated *RNU4-2* splicing profiles by RNA sequencing and performed DNA methylation studies, identifying a ReNU syndrome-associated episinature.

## Results

### Analysis of *RNU4-2* variants in multiple cohorts of patients with rare diseases

We investigated *de novo* variants in *RNU4-2* (NR\_003137.2) and/or rare variants (<10 alleles in gnomADv4.1.0) located in the 18-bp critical region defined by Chen *et al.*<sup>13</sup> in the Plan France Médecine Génomique 2025 (PFMG) cohort comprising 23,649 patients with rare disorders (15,073 with NDD). This analysis revealed 73 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 of 60 patients with *RNU4-2* variants identified through genome sequencing data reanalysis (22 patients) or targeted sequencing (38 patients including one monozygotic twin pair). Variants occurred *de novo* in 44/45 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, 138 individuals had 22 distinct *RNU4-2* variants. Ninety-five patients (43 males, 52 females including the twin; 69%) had the recurrent n.64\_65insT insertion. Seven other variants were recurrent: n.76C>T (*n*=9), 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 (2 patients each). Fourteen *de novo* variants were identified in single patients. All but three variants clustered in the 18-bp critical

region spanning nucleotides 62 to 79 (chr12(hg38):120,291,825-120,291,842)<sup>13</sup>. This region overlaps four distinct domains in the U4/U6 structure<sup>4</sup> (Fig. 1): stem I (U4 n.62), T-loop/quasi-pseudoknot (n.63-67), RBM42 interaction region (n.68-70) and stem III (n.72-79). The remaining variants were located in the 5' stem loop between stem I and stem II (n.45\_46insT) or in the 3' stem loop (n.92C>G, n.111C>T). Eighteen variants were absent from gnomADv4.1.0, AllofUs, and TOPMed\_freeze10 while three were present at very low frequency (n.66A>G, n.76C>T once each; n.76del: 3 times; Supplementary Fig. 1A; Supplementary Table 1). Two variants (both located outside of the 18-bp critical region) had higher occurrences: n.92C>G and n.111C>T, 16 and 82 times respectively in gnomADv4.1.0, AllofUs and TOPMed combined. We classified 18 variants (in 134 individuals; 133 probands) as pathogenic (P) or likely pathogenic (LP) using ACMG/AMP criteria<sup>18,19</sup> and four variants (n.45\_46insT, n.92C>G, n.111C>T, n.76del) as of uncertain significance (methods).

We assessed CADD PHRED scores, nucleotide conservation, and *in silico* structure predictions for the 18 LP/P variants and compared these features with variants in *RNU4-2* present in gnomADv4.1.0. While no differences were found in CADD scores or vertebrate conservation (verPhyloP), LP/P variants had a greater predicted effect on U4/U6 interaction compared to variants observed  $\geq 10$  times in gnomAD (Supplementary Fig. 2; Supplementary Table 2). However, due to the large overlap, these predictions cannot be used to predict variant pathogenicity.

### ***De novo* analysis in snRNA genes reveals variants in brain-expressed U5 genes**

We next analysed *de novo* variants in 49 additional genes encoding snRNAs (Supplementary Table 3) in the PFMG cohort. This analysis revealed 12 rare *de novo* and/or recurrent alterations in eight genes in 15 unrelated patients. Notably, eight variants (in 11 patients) were located in genes encoding U5 snRNAs (two in *RNU5A-1* (NR\_002756.2), four in *RNU5B-1* (NR\_002757.3), one in *RNU5E-1* (NR\_002754.2) and one in *RNU5F-1* (NR\_002753.5)). Three variants (*RNU5A-1* n.40\_41insA, *RNU5B-1* n.39C>G and n.44A>G) were recurrent and identified in two individuals each. In seven patients, the variant was located in the highly conserved 5' loop I (Fig. 2), which is depleted in variants in gnomADv4.1.0 (Supplementary Fig. 1B) and in the UK Biobank<sup>13</sup>. Four *de novo* variants were located in other parts of U5, the internal loops 1 and 2 (*RNU5B-1* n.24G>C, and n.74T>C), the variable 3' stem loop II (*RNU5F-1* n.115C>A), or the Sm site (*RNU5E-1* n.90\_91insA). The individual with *RNU5F-1* n.115C>A had a *de novo* heterozygous frameshift variant in *KDM5B*. All other patients



remained negative after genomic variant analysis, suggesting the variants in U5 encoding genes as possible underlying genetic causes.

In parallel, we collected information on rare *de novo* variants in *RNU5A-1* and *RNU5B-1* in data from the 100,000 Genomes Project (GEL) and NHS GMS cohorts available within Genomics England ([Supplementary Table 5](#)). This analysis identified five *de novo* variants in *RNU5B-1* in six NDD probands (n.37G>C; n.37G>T; n.42\_43insA twice, n.44A>G; n.64G>A). This was compared to only a single non-NDD individual in GEL / NHS GMS having a *de novo* variant in *RNU5B-1* (n.59G>C; 6/12,724 undiagnosed NDD vs 1/30,058 non-NDD; Fisher's *p*-value = 0.0036). Additionally, we collected data of five patients with *de novo* variants (n.39C>G: 3 patients; n.42\_43insA; n.44A>G) from additional cohorts, including the Broad Centre for Mendelian Genomics (two patients), the BCH Epilepsy Genetics Program, the Australian undiagnosed network, and Care4Rare Canada (one patient each).

Altogether, 13 NDD probands had five distinct *de novo* variants in the 5' loop I of *RNU5B-1* ([Supplementary Table 6](#)). These variants recurrently affected the same nucleotide positions: n.37G>C and n.37G>T, each found in one individual; n.39G>C (five patients); n.42\_43insA and n.44A>G (three patients each). This led us to define a critical region in *RNU5B-1* spanning chr15(hg38):65,304,713-64,304,720. Three additional patients analysed in duo in the GEL/NHS GMS cohort and one patient in the PFMG had variants located within this critical region that were absent from the available sequenced parent. In contrast, no individuals with non-NDD phenotypes in GEL/NHS GMS had *de novo* or duo variants in this region (8/12,724 undiagnosed NDD vs 0/30,058 non-NDD in GEL/NHS GMS; Fisher's *p*-value =  $6.1 \times 10^{-5}$ ). In total, we identified 17 NDD individuals with variants in this critical region of *RNU5B-1*. No other *de novo* variants were identified in *RNU5A-1* in NDD probands. *RNU5A-1* or *RNU5B-1* are the main genes encoding U5 in the brain ([Supplementary Fig. 3](#)).

### **Most pathogenic *RNU4* and *RNU5* variants occur *de novo* 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. 4](#)). We could reliably determine the parental origin of the mutations in 45 trios. The variant was assigned to the maternal allele in 42 cases and to the paternal allele in only three cases. Notably, all 34 n.64\_65insT variants were phased to the maternal allele, consistent with observations from Chen *et al.*<sup>13</sup>. Among the variants assigned to the paternal allele, two were

classified as likely pathogenic (n.62T>C, n.68A>C) and one (n.92C>G) was of uncertain significance. All three paternally derived variants were SNVs, while this variant type only represented 22% (10/45) of phased variants (Fisher's  $p$ -value = 0.0085).

Among the phaseable variants located in the 5' loop I *RNU5B-1*, five (n.39C>G, n.42\_43insA and three n.44A>G) were phased to the maternal allele, while two (n.39C>G, 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, 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 (quasi-pseudoknot) and stem III differ in severity**

Clinical data were available for 129 patients with P/LP variants in *RNU4-2* (63 males, 66 females excluding the monozygotic twin who had an identical phenotype to her sister). The median age at study entry was 9 years (range: 4m-45y). All patients had NDD with intellectual disability (ID) of variable severity, ranging from mild (8%), moderate (29%) to severe/profound (63%).

We investigated genotype-phenotype correlations to understand the phenotypic variability of *RNU4-2*-related disorders. We first performed an unsupervised clustering of clinical features, which revealed two separate clusters with different phenotypic severity (Fig. 3A). Most *RNU4-2* variants in stem III (71%, 12/17) were in the mild phenotype cluster, whereas most variants in the T-loop and RBM42 interacting region were in the higher severity cluster (95%, 106/111). The same finding was observed when principal component analysis (PCA) was performed. Variants in stem III and variants in the T-loop separated on the first PC axis, accounting for 12.3% of the variance (Fig. 3B-C; Supplementary Fig. 5). These results suggest that a large part of the phenotypic variability is due to the location of *RNU4-2* variants in different U4 functional domains.

Clinical data were available for 91 patients with *RNU4-2* c.64\_65insT (Table 1; Supplementary Table 7; Supplementary Fig. 6). Prenatal findings were observed in 50/82 (61%) of cases, predominantly intrauterine growth restriction (IUGR) (28%) and cerebral abnormalities (32%), mainly ventriculomegaly (18%). Of the foetuses with abnormalities, 62% had isolated signs, while 38% exhibited two or more signs, including IUGR, cerebral abnormalities, talipes equinovarus, and renal abnormalities. Neonatal findings were frequent

(90%), with hypotonia (69%) and feeding difficulties (57%) being the most common. Microcephaly (head circumference (HC) < 3<sup>rd</sup> percentile) was present at birth in 26%, and at last examination in 74% ([Supplementary Fig. 7](#)). Short stature was noted in 58% of individuals. All patients older than three years ( $n=73$ ) exhibited developmental delay. Most could walk, with a median walking age of 32 months (range 13 months to 12 years), whereas 14% did not reach this milestone. Most patients were non-verbal (60%) or could only speak a few words (35%). The majority had severe/profound ID (77%), with 22% having moderate ID and one patient having mild ID. Behavioural disturbances were common, with 83% displaying autistic features and/or midline stereotypies reminiscent of Rett syndrome in some patients. Epilepsy affected 57%, with an additional 9% experiencing a single seizure. Seizures typically started between 15 months and 8 years (median onset at 33 months) and were usually generalized, rare, fever-sensitive, and responsive to antiepileptic medications. However, five patients were diagnosed with developmental and epileptic encephalopathy; 14 experienced status epilepticus, and seven had drug-resistant epilepsy. Nystagmus and cerebellar ataxia were observed in 36% and 23%, respectively. Brain MRI abnormalities were prevalent (90%), with the most common findings being enlarged ventricles (60%) and a thin corpus callosum (34%). Less common findings included heterotopia ( $n=7$ ), delayed myelination or hypomyelination ( $n=7$ ), and abnormal gyration ( $n=4$ ). Cardiac and genitourinary or renal abnormalities were less common, affecting 19% and 18% of patients respectively, and were generally benign. Skeletal abnormalities, including osteopenia or fractures ( $n=18$ ) and hip dysplasia ( $n=9$ ), were seen in 40% of cases. Dysmorphic features could suggest Pitt-Hopkins syndrome. Strabismus, and drooling were also common. Feeding difficulties affected 69%, failure to thrive 55%, and constipation 53%. Acrocyanosis or vasomotor disorders were present in 16 patients, blood count anomalies in 12, and hypothyroidism in seven.

The phenotype of the patients with variants in U4 T-loop and RBM42 interaction region was indistinguishable from that of patients with the n.64\_65insT patients. Patients with the recurrent n.66A>G ( $n=5$ ) and n.67A>G variants ( $n=5$ ) had a similar phenotype, characterized by neonatal hypotonia (5/5 and 3/5), microcephaly (5/5 for both variants), epilepsy in about half of them (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 non-verbal.

Patients with *RNU4-2* n.76C>T variant ( $n=8$ ) exhibited a distinct clinical profile from patients with the n.64\_65insT (Table 1; Supplementary Table 7). They had less neonatal findings (Fisher's  $p=6.58E-04$ ), specially hypotonia ( $p=0.0208$ ), presented less severe ID ( $p=1.06E-06$ ) and developmental delay ( $p=6.63E-04$ ), were more proficient in their language abilities ( $p=1.06E-06$ ), and rarely showed brain MRI abnormalities ( $p=2.97E-03$ ). All patients could walk, four of them achieving this milestone at a normal age (median walking age: 19m [12-33m]), and all could speak, with simple sentences ( $n=4$ ) or normal language skills ( $n=4$ ). IUGR was observed in only 1/8 patients. Microcephaly was noted in 2/8 patients, and short stature in 1/8. Two out of five patients had autistic features. Six patients had fever-sensitive generalized epilepsy, well-controlled with anti-seizure medication, while two others had a single febrile seizure. None had nystagmus or ataxia, and brain MRI was normal in 5/6 cases, with one showing moderate diffuse cortico-subcortical atrophy. Congenital malformations were rare, and the dysmorphic features were distinct from those seen in patients with the recurrent variant.

Similarly, patients with other variants in the stem III ( $n=9$ ) exhibited a mild/moderate phenotype than patients with the n.64\_65insT variant showing less severe intellectual disability ( $p=1.52E-04$ ) and developmental delay ( $p=3.10E-02$ ), and with improved language abilities ( $p=7.79E-06$ ). All patients could walk and speak, with varying degrees of language development (2: normal language, 6: simple sentences and 1: few words). ID was mild in three and moderate in six, with autistic features in 2/5 cases. Fever-sensitive epilepsy was common (7/9), but well controlled with anti-seizure medication. Brain MRI was normal in 5/7 patients.

### ***RNU5* variants are associated with NDD with variable malformations**

Detailed clinical data were available for 9 out of 15 NDD patients with *RNU5B-1* LP variants (Supplementary Fig. 8; Supplementary Table 8). 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 (-15.75/-12.25 dioptries). 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

subjects with n.39C>G had macrocephaly. HPO terms enriched in *RNU5B-1* cases from GEL include seizures, macrocephaly and dystonia ([Supplementary Table 9](#)).

The three patients with *RNU5A-1* variants 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 agenesis; the third had cardiac malformations and marfanoid habitus. The two patients with n.40\_41insA had seizures. Head circumference was normal in all.

### **Pathogenic variants in *RNU4-2* lead to alternative 5' splice site usage**

Chen et al. reported specific alternative 5' splice site (5'SS) abnormalities in the blood of individuals with *RNU4-2* variants.<sup>13</sup> To confirm and extend this observation, we performed RNA sequencing on short-term lymphocyte cultures from 19 *RNU4-2* affected individuals, (10 with n.64\_65insT and 9 with other variants) and 21 control individuals with other NDDs. Using rMATS<sup>20</sup> with an FDR of 0.1 and an absolute deltaPSI difference > 0.05, we detected 35 significant aberrant splicing events affecting 3' splice sites (3'SS), 126 for mutually exclusive exons, 100 for intronic retention, 121 for exon skipping and 111 for 5'SS ([Supplementary Tables 10-14](#)). We then extracted PSI values of significantly altered exons for each splicing category and we performed a principal component analysis using matrices with samples as columns and PSI values as rows. We observed no separation between affected and control individuals for 3'SS and intronic retention, and only a slight separation for mutually exclusive exons and exon skipping ([Supplementary Fig. 9](#)). The most striking effect was for 5'SS. We observed a cluster of patient samples with severe phenotypes (c.64\_65insT, n.67A>G, n.68A>C and n.70T>C). Patients with a mild phenotype appeared intermediate between severe cases and controls (n.72\_73del, n.75C>G, n.76C>T) ([Fig. 4A](#)). These results suggest a common signature of 5'SS usage in *RNU4-2* probands, with distinct profiles correlating with disease severity.

We visually characterised 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. 4B-D](#); [Supplementary Table 15](#); [Supplementary Fig. 10](#)). Decreased 5'SS consistently show 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. 4B](#)). In

contrast, increased 5'SS events associated with severe phenotypes have significantly lower SpliceAI scores (median=0.52, t-test  $p=0.005$ ) compared to shared sites (median=0.80). Additionally, only five 5'SS events were absent from controls (mean supporting reads in controls  $<3$ ; 5/69; 7.2%), suggesting that the main effect of U4 variants is a shift in existing alternative isoforms rather than the use of new cryptic splice sites. The consensus of 5'SS usage revealed a consistent decrease in A/A/G nucleotides replaced by C/T at positions +3/+4/+5 (Fig. 4E). This effect is associated with an increase in the dependence of A/G in position -2/-1 that is particularly visible for decreased 5'SS events restricted to severe phenotypes (2/19 'AG' for severe variants vs. 26/50 'AG' for shared variants; 2-tailed Fisher's test,  $p=0.0008$ ). This suggests that 5'SS used only in severe patients are more reliant on the end of the exon and less on the sequence of the intron. Altogether, we identified a specific splicing pattern associated with *RNU4-2* variants that can distinguish affected samples based on disease severity and variant location.

### Identification of a specific *RNU4-2* epismature

Finally, we investigated whether *RNU4-2* variants are associated with specific DNA methylation profiles. We analysed methylation profiles of 23 patients with P/LP variants in *RNU4-2* and 35 controls (healthy, age-matched individuals). Following adjustment on age, sex and blood cell composition, epigenome-wide analysis led to the identification of 90 probes with uncorrected  $p$ -value  $< 10^{-7}$  and  $|\Delta\beta| > 5\%$ . PCA and heatmap representations of adjusted methylation levels showed a good separation between moderate-to-severe *RNU4-2* phenotypes and controls while mild NDD phenotypes tend to cluster with the controls (Fig. 5). The first PC axis, separating moderate to severe patients from controls captures 54% of the residual variance of methylation levels, after adjustment for age, sex and cell composition (Supplementary Fig. 11), which is of the order of magnitude previously observed for *ATRX*, *KMT2D* or *KMT2A* epismatures.<sup>21</sup> After five-fold cross-validation, the signature obtains a sensitivity of 0.68 (95% CI [0.47 – 0.85]), and a specificity of 0.97 (95% CI [0.85, 0.999]) (Supplementary Fig. 12).

### Discussion

Despite extensive genetic testing, 40-60% of patients with NDDs of suspected genetic origin remain unsolved. This diagnostic gap is due to several challenges including difficulties in interpreting variants in non-coding genomic regions. The recent discovery of *RNU4-2* variants

as a major cause of NDDs, overlooked until 2024, highlights the significant role that non-coding genes likely play in undiagnosed cases. Here, we analysed genes encoding functional snRNAs in a large French cohort of patients who underwent genome sequencing as part of routine diagnosis and identified 74 patients with pathogenic or likely pathogenic variants in *RNU4-2* (0.49% of subjects with NDD) and seven (0.046%) with variants in the 5' loop I of *RNU5A-1* or *RNU5B-1*. We also collected data of 73 additional patients from distinct cohorts. Altogether, we observed that pathogenic variants cluster in evolutionarily conserved, key functional regions of U4 and U5 involved in splicing. Specifically, *RNU4-2* variants cluster in two main U4 regions, the T-loop/quasi-pseudoknot and stem III, while variants in *RNU5B-1* all are in the conserved loop I, the part of U5 that pairs with the end of the exon adjacent to the 5'SS.<sup>5</sup> The three *de novo* variants in *RNU5A-1* are also located in this constrained region. A few additional *de novo* variants in other domains or snRNA genes were also identified but their clinical relevance remains unclear.

This study provides a comprehensive overview of *RNU4-2*-related phenotypes based on a large series of patients, and shows distinct clinical outcomes depending on variant location. Variants in the T-loop, including n.64\_65insT, are associated with more severe phenotypes, while variants in the stem III lead to milder forms. The n.76C>T variant, previously suggested to cause a milder phenotype from single case observations<sup>13,14</sup>, and c.72\_73del, support the idea of a continuum of *RNU4-2*-related phenotypes, with inherited variants also possibly contributing to the genetic aetiology of NDDs.

We observed prenatal manifestations in 60% of individuals with *RNU4-2* pathogenic variants, mainly isolated cerebral abnormalities (corpus callosum anomalies and enlarged ventricles) and/or IUGR. Considering the high incidence of *RNU4-2* variants, these findings have noteworthy implications for prenatal diagnosis, emphasising the need for genome sequencing, or adding *RNU4-2* analysis, in prenatal genetic testing.

A striking observation in line with previous findings<sup>13</sup> is the predominant maternal origin of *RNU4-2* variants, possibly explained by negative selection of variants severely affecting splicing in the male germline. However, paternal transmission of less severe variants is possible, as evidenced by three 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,<sup>11,12</sup> though 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 blood cells of affected patients, with detected events correlating with phenotype severity. Variants in the T-loop and stem III indeed show distinct, partially overlapping transcriptional signatures, which could be used to interpret variants of uncertain significance. Furthermore, the more severe NDD phenotypes associated with *RNU4-2* variants in the T loop also display a specific methylation profile (episignature). Given the widespread use of exome sequencing in routine diagnostics, these transcriptional and epigenetic signatures could help diagnose additional ReNU syndrome cases worldwide.

From a molecular standpoint, U4 variants may disrupt spliceosome function at various stages: U4 snRNP biogenesis, U4/U6 di-snRNP, U4/U6.U5 tri-snRNP assembly, or spliceosome activation. However, considering their location and associated transcriptional alterations, they more likely disrupt the organization of the U4/U6 duplex at the tri-snRNP stage and affect the introduction of the 5'SS into the spliceosome's active site. The 5'SS is initially paired with U1 in the pre-spliceosome and then transferred to the U6 ACAGAGA box and U5 stem loop 1 located in U4/U6.U5 tri-snRNP. These interactions maintain the 5'SS in the active site during catalysis, marking the start of each intron<sup>1,22</sup>. When the 5'SS is transferred, it pairs with the U6 ACAGAGA box to ensure its correct identification<sup>23</sup>. Proper pairing 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 quasipseudoknot. Snu66, SNRP27K, and RBM42 proteins further stabilize this organization, ensuring the ACAGAGA box can properly recognize the 5'SS.<sup>4,24-26</sup>

Pathogenic *RNU4-2* variants fall under two main categories. Variants in -or close to- the quasi-pseudoknot (n.62-67) possibly weaken if not disrupt its structure and compromise its ability to maintain the ACAGAGA box at the right position for 5'SS recognition. A similar effect could apply to variants (n.68-70) located in nearby single stranded U4 region interacting with RBM42. The second variant category (n.73-79) likely affects U4/U6 stem III's stability by altering Watson Crick base pairs formed by U4's A78, C76 and C75 with U6's G34, G33 and U31. The role of stem III in 5'SS recognition remains unclear. Besides orienting the ACAGAGA box, stem III may enhance fidelity by imposing an energy barrier to the extension



of a U6/5'SS helix stemming from the initial pairing of the 5'SS with the ACAGAGA sequence<sup>4,5</sup>. Stem III disruption is required prior to Brr2 loading onto U4 for active site formation. Hence, we hypothesize that a weakened stem III may allow sub-optimal 5'SS to more easily extend the U6/5'SS helix, promoting spliceosome activation. The increased 5'SS in the severe cases suggests that the spliceosome loses specificity for the intronic 5'SS sequence motif, which seems to be compensated by more contribution from the exonic 5'SS sequence motif by U5.

U5 loop I plays an important role during 5'SS transfer by interacting with the exonic sequences adjacent to 5'SS. These interactions help alignment of the 5'exon with respect to the branch site and the 3'exon during both steps splicing catalysis<sup>27,28</sup>. Mutations in yeast U5 loop I result in the activation of the aberrant 5'SS, making it an important determinant of 5'SS specificity<sup>29</sup>. *RNU5* variants may therefore modify the specificity of U5 loop I towards pre-mRNA, resulting in similar effects as *RNU4-2* variants. Future studies should analyse whether there is a similar or distinct splicing signature associated with variants in *RNU5A-1* and *RNU5B-1* to that present in *RNU4-2*. Together, these RNA structures formed by U4 and U5 build a conformation that prevents weak 5'SS-triggered activation of the spliceosome, promoting the usage of optimal 5'SS. Destabilization of these structures would make the spliceosome more prone to error, leading to a decrease in splicing fidelity.

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

## Methods

### Inclusion & Ethics statement

This study complies with the ethical standards of each of the participating countries. An informed consent was obtained for 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. Patient/participant/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. The study has received the approval of the ethics committee of University Hospital Essen (24-12010-BO). 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.

### List of snRNA genes

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.

### 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<sup>30</sup>. Following the publication of the

preprint by Chen et al. on April 8, 2024<sup>31</sup>, we investigated variants in *RNU4-2* and 49 other snRNA genes in several diagnostic and research cohorts. Our inclusion criteria were: 1) a *de novo* variant in any of the 50 snRNA genes with less than 10 heterozygotes in gnomADv4.1.0, or 2) a heterozygous variant in *RNU4-2* located within the critical 18-nucleotide region as defined by Chen et al. 2024.<sup>13</sup> We then narrowed our search to *RNU5A-1* and *RNU5B-1*, and investigated 3) *de novo* variants with less than 10 heterozygotes in gnomADv4.1.0 and/or 4) heterozygous variants in *RNU5B-1* located in the 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 (Plan France Médecine Génomique 2025, PFMG2025) 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 78 patients. Furthermore, we collected data of patients with *de novo* and/or pathogenic *RNU4-2* variants identified in either diagnostic or research contexts through national networks, established collaborations, or GeneMatcher<sup>30</sup>. These additional cohorts included 40 patients from France, 17 patients from Germany, one patient from the Netherlands, one from Spain, and one from the US. Twenty-two patients had genome sequencing in diagnostic ( $n=5$ ) or research ( $n=15$ ) contexts, whereas in 38 patients, the variant was identified or confirmed by a targeted method: Sanger sequencing ( $n=31$ ) or next generation sequencing of amplicons ( $n=5$ ). Among the patients who were diagnosed by Sanger sequencing, two patients previously had inconclusive exome analysis 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 six patients with *de novo* variants *RNU5B-1*, three patients with *de novo* variants in *RNU5A-1*, and six patients with *de novo* variants in a single gene: *RNU4-1*, *RNU5E-1*, *RNU5F-1*, *RNU6atac*, *RNVU1-22*, or *RNVU1-27* (Supplementary Table 4). A targeted search for variants in *RNU5B-1* and *RNU5A-1* in the Genomics England dataset (including both the 100,000 Genomes cohort (v18) and National Health Service Genomic Medicine Service (NHS-GMS v3) cohort) identified six additional individuals with rare ( $< 10$  occurrences in gnomAD)

*de novo* variants, five out of 8,841 undiagnosed NDD proband and one out of 21,816 non-NDD probands. In addition, three probands analysed in duo had a rare variant in the 5' loop I absent from the single parent analysed. In addition, five *de novo* variants in *RNU5B-1* were collected from the Broad Centre for Mendelian Genomics, Undiagnosed Disease Networks (UDN), the BCH Epilepsy Genetics Program, and Care4Rare Canada.

## **Variant classification**

We classified variant according to ACMG/AMP criteria<sup>18</sup> using recommendations from Ellingford et al.<sup>19</sup> The PS2 (or PM6 for patients who underwent targeted sequencing) criteria was applied for all cases with *de novo* inheritance. We applied PS3 when RNAseq and/or methylation analyses supported pathogenicity or 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*. PM2 supporting was applied for variants absent from gnomAD v4.1.0, and PS4 supporting was applied for recurrent variants (at least 3 patients in this study or in Chen et al.<sup>13</sup>). This led to classify *de novo* variants located in the critical region as LP. Recurrent variants with > 3 occurrences and absent from databases were classified as P. Only variants for which inheritance was unknown (even located in critical region) or variants occurring outside of critical regions (even if *de novo*) were classified as variants of uncertain significance (VUS).

## **Clinical data analyses**

Clinical data were retrospectively collected from the referring physician using an anonymized Excel sheet. Categorical data for 44 selected clinical features from 129 patients with pathogenic and likely pathogenic *RNU4-2* variants were converted to a 0-1 scale, with 0 representing a more favourable phenotype presentation and 1 a more severe phenotype. Hierarchical clustering was performed using 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 by 0 and performing variable scaling. Plots concerning to head of circumference (HC) measurements at birth were generated with the “Plotter: Preterm growth charts, 22-50 weeks” from the Canadian Pediatric Endocrine Group (<https://cpeg-gcep.shinyapps.io/prem2013/>).<sup>32</sup> For additional HC measurements, reference chart data points were obtained from Rollins et al. 2010.<sup>33</sup> 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 vs. n.76C>T and n.64\_65insT vs. the other variants groups) for the 39 clinical features.

## Conservation and *in silico* predictions

The highest homolog 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.<sup>34</sup> RNA sequences from *RNU4-2* and *RNU5B-1* were aligned to i) their respective sequence homologs and 2) the sequence(s) of other U4- and U5-encoding genes expressed in the brain using Geneious Prime® 2019.2.3 (Geneious Alignment). The threshold for consensus was set to 100% identical, highlighting positions with 100% agreement between all sequences.

CADD PHRED scores and conservation in vertebrate (verphyloP) were calculated for pathogenic and likely pathogenic patient variants and gnomADv4.1.0 variants with CADD v1.7<sup>35</sup>. For each variant, *in silico*-mutated U4 RNA sequences were generated with seqkit mutate<sup>36</sup>. Bifold<sup>37</sup> was used to generate the multiple U4:U6 interactions and calculate the minimum free energy. Comparisons were performed by applying 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<sup>38</sup>: diencephalon (GEO:GSE78292), temporal lobe (GEO:GSE78303), occipital lobe (GEO:GSE78298), frontal cortex (GEO:GSE78293), parietal lobe (GEO:GSE78299), cerebellum (GEO:GSE78291). Tracks show unique read signals for plus and minus strand from the default anisogenic replicate. Expression of these genes in the brain using BrainVar was previously investigated.<sup>13</sup>

## RNA-sequencing

Peripheral blood mononuclear cells (PBMCs) were isolated from 2 to 4 ml of EDTA-anticoagulated blood within 48 hours of collection using UNISEP+ tubes (EUROBIO). Cells

were cultured in 6-well plates ( $5.0 \times 10^5$  to  $2.0 \times 10^6$  cells per well) in lymphocyte-stimulating medium (Chromosome Medium P, EUROCLONE) for 48 to 72 hours at 37°C (5% CO<sub>2</sub>). After incubation, one well per sample was treated with 1mg/ml puromycin solution for 4-5 hours to inhibit Nonsense-Mediated Decay (NMD). RNA was extracted using the NucleoSpin RNA Plus extraction kit (MACHEREY-NAGEL) according to the manufacturer's instructions.

Stranded RNA-Seq libraries were prepared from 100ng of total RNA on the Magnis NGS Prep System (Agilent) using SureSelect XT-HS2 kit (Human All Exon V8 capture probes) with 12 and 10 PCR-cycles for pre-capture and post-capture amplifications, respectively. RNA-Seq were sequenced on an Illumina's NextSeq 550 (16 samples on HighOutput 2x75bp) to obtain 25 to 30 million paired-end reads per sample.

Fastq files were aligned on 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). Controls were matched on following criteria: library preparation kit, sequencing flowcell and culture time. CIBERSORTx was used to estimate relative abundance of blood cells using LM22 signature matrix file<sup>39</sup>. One *RNU4-2* sample was removed because of a low proportion of activated T CD4+ cells (1/38). Abnormal splicing events were called using 19 *RNU4-2* samples and 21 controls by using rMATS (v.4.3.0)<sup>20</sup> with following parameters: -t paired -anchorLength 1 -libType fr-firststrand -task both -novelSS -variable-read-length -allow-clipping. Python scripts were used to filter rmats output files with following filters: mean coverage > 7, FDR < 0.1, deltaPSI > 0.05. Principal component analysis was performed using sklearn python library using PSI values from significant A5SS, keeping only the most significant A5SS if several were called in the same exon. Raw spliceAI scores were obtained from MobiDetails<sup>40,41</sup>. Sashimi plots were made using rmats2sahimi and boxplots with seaborn. Consensus nucleotide sequences were generated using Logomaker.

## **Epigenome-wide analysis and DNA methylation signature**

Genomic DNA was extracted from whole blood and bisulfite converted. DNA methylation profile was then derived using Illumina's Infinium EPIC array v2.0, in accordance with the manufacturer's protocol. Patients and negative controls were balanced across sixteen arrays and each array rows in order to reduce technical biases. DNA methylation arrays were generated at the ASGARD-Rouen genomic platform (University of Rouen and Rouen University Hospital, Rouen France) on an Illumina NextSeq550 scanner. Raw IDAT data were

processed and normalised using the default Meffil R package protocol along with all other samples included in the sixteen arrays in order to better estimate the variability of methylation signals within and across arrays<sup>42</sup>. One *RNU4-2* sample failed default quality controls and was excluded from further steps. Remaining samples were functionally normalised 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  $\beta$ -values.

Several predictions were obtained from methylation values to apply additional QC and normalization steps. Sex predictions were extracted from the standard Meffil normalized object. No inconsistencies between reported and predicted sex were noted ([Supplementary Fig. 11A](#)). 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. 11B](#)). DNA methylation age was predicted with the `DNAMAge` function from the `methylock` R package<sup>43</sup>. The Horvath and skinHorvath clocks both displayed very strong correlation with actual age at blood sample on our dataset (Pearson correlation  $r=0.97$  ([Supplementary Fig. 11C-D](#))).

The set of differentially methylated probes was identified with the `meffil.ewas` function on the subset of controls and pathogenic or likely pathogenic *RNU4-2* variant carriers. To correct for well-known confounders, the analysis was adjusted on age at blood sample, sex and predicted blood cell composition. Manhattan and volcano plots are given in [Supplementary Fig. 12A-B](#). After filtering on p-value and average methylation difference between positive and negative controls, residuals of a linear regression model fitted to adjust probe  $\beta$ -values on age, sex and predicted cell composition, were visualized through principal component analysis and heatmap representations. 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 5-fold cross-validation. The dataset was split into five random and equal-sized blocks. Each block was used in turn as validation set, while the remaining four blocks were used as 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, and by phenotype class, along with 95% binomial confidence intervals.

## **Variant impact**

Structural analysis of variants and corresponding figures were performed using the PyMol v3.0.0 visualisation software<sup>44</sup> on published coordinates of the human tri-snRNP structure: PDB 6QW6<sup>4</sup> and PDB 8Q7N<sup>45</sup>.

## **Data availability**

Individual genome, RNAseq and methylation data could not be made publicly available due to ethical considerations. Controlled access to human sequences is necessary to protect the privacy of participants and to ensure that the use of the data conforms to ethical and legal standards, particularly in relation to data protection laws in France, Germany, and Europe. For inquiries regarding data access, please contact the corresponding authors by email. Other data supporting the findings described in this manuscript are available in the article and its 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/>).

## **Code availability statement**

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 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>). Preterm head circumference graphs were plotted using <https://cpeg-gcep.shinyapps.io/prem2013/>. Custom scripts used for *RNU4-2* RNAseq analysis are



available on GitHub using the following link: [https://github.com/benjamin-cogne/RNU4-2\\_transcriptomics](https://github.com/benjamin-cogne/RNU4-2_transcriptomics) (<https://doi.org/10.5281/zenodo.13868502>).

## 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**, 6482-93 (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**, 2746-56 (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. doi:10.6084/m9.figshare.4530893/7. (2020).
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. 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**, 405-24 (2015).
19. Ellingford, J.M. *et al.* Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med* **14**, 73 (2022).
20. 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).
21. 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).

22. Will, C.L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**(2011).
23. Kandels-Lewis, S. & Seraphin, B. Involvement of U6 snRNA in 5' splice site selection. *Science* **262**, 2035-9 (1993).
24. Zhang, Z. *et al.* Structural insights into the cross-exon to cross-intron spliceosome switch. *Nature* **630**, 1012-1019 (2024).
25. 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).
26. 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).
27. Newman, A.J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743-54 (1992).
28. Sontheimer, E.J. & Steitz, J.A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989-96 (1993).
29. Newman, A. & Norman, C. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell* **65**, 115-23 (1991).
30. 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**, 928-30 (2015).
31. Chen, Y. *et al.* De novo variants in the non-coding spliceosomal snRNA gene RNU4-2 are a frequent cause of syndromic neurodevelopmental disorders. *medRxiv* (2024).
32. 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).
33. 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 e2 (2010).
34. Harrison, P.W. *et al.* Ensembl 2024. *Nucleic Acids Res* **52**, D891-D899 (2024).
35. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
36. 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).
37. Reuter, J.S. & Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
38. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
39. Newman, A.M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019).
40. Baux, D. *et al.* MobiDetails: online DNA variants interpretation. *Eur J Hum Genet* **29**, 356-360 (2021).
41. de Sainte Agathe, J.M. *et al.* SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. *Hum Genomics* **17**, 7 (2023).
42. 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).
43. 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).
44. Schrödinger, L., DeLano, W. PyMOL. (2020).
45. 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).

## **Author contribution**

Be.Co., Am.Sa., El.Le., and Fr.Le., developed bioinformatics pipelines and performed data analyses. Y.C. and N.W. performed data analyses in the GEL/NHS cohorts. S.L.S. and A.O'D-L contributed data from other cohorts. Cl.Ch. performed structural analysis. Ca.Na., Ju.Th., and Ch.De. conceived the study, performed data analyses, and supervised the project. All other authors contributed molecular and/or clinical data. Ca.Na., Be.Co., Ca.Ch., Cl.Ch. and Ch.De. drafted the article. All authors reviewed and approved the final manuscript.

## **Competing interest**

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, Ono Pharma USA Inc. and at present for Addition Therapeutics, and received reagents from PacBio to support rare disease research. All other authors declare no competing interests.

## **Acknowledgments**

We thank the patients and their families for their participation in this study. This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). We thank P. O'Donovan, Mitra Sato and Zainab Mustafa from Genomics England 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. We thank the Care4Rare Canada Consortium for support of a subset of this work. The Chair in Genomic Medicine awarded to JC is generously supported by The Royal Children's Hospital Foundation. We also thank Miranda Di Biase, Ashley Wilson, Atteeq Rehman, Amanda Thomas-Wilson, Shruti Phadke, Avinash Vabhyankar from the New York Genome Center for their help with genome analysis and for providing anonymized patient data. A.K. and F.J.K. from Essen, Germany, T.S.B from the Netherlands, K.Õ. from Estonia, A.-S.D.-P., L.F., G.L., J. V-G., M.W. and S.O. from France, and N.R. from Belgium are members of the European Reference Network (ERN) ITHACA whose EU coordinator is A.V. Support

for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

## **Funding**

Patients included in this study were diagnosed through multiple studies including Plan France Médecine Génomique 2025 (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 programme under grant agreement No 779257; Deutsche Forschungsgemeinschaft (DFG) – Project number 458099954 to Ch.De. and Project number 37144118 to R.A.J and So.N.; Inserm as part of the 2022 MESSIDORE program (project number Inserm-MESSIDORE N°19); the European Joint Programme on Rare Diseases GENOMIT I6478-B by the Austrian Science Fund FWF; by the Centre for Population Genomics (Garvan Institute of Medical Research and Murdoch Children's Research Institute) and was funded in part by a Medical Research Future Fund (MRFF) Genomics Health Futures Mission grant (2008820). 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 Australian Undiagnosed Diseases Network (UDN-Aus) acknowledges financial support from the Australian Government's Medical Research Future Fund (MRFF grant reference number MRF2007567). Cases identified by the Broad Center for Mendelian Genomics (AODL and SLS) were supported by the National Human Genome Research Institute (NHGRI) grants UM1HG008900 and U01HG0011755. 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 Children's Hospital of Eastern Ontario Foundation. K.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). AMI, JL and RL obtained support for sequencing through Translational Implementation of Genomics for Rare Disease (TIGeR), 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. was supported by a BCH Office of Faculty Development/Basic & Clinical Translational

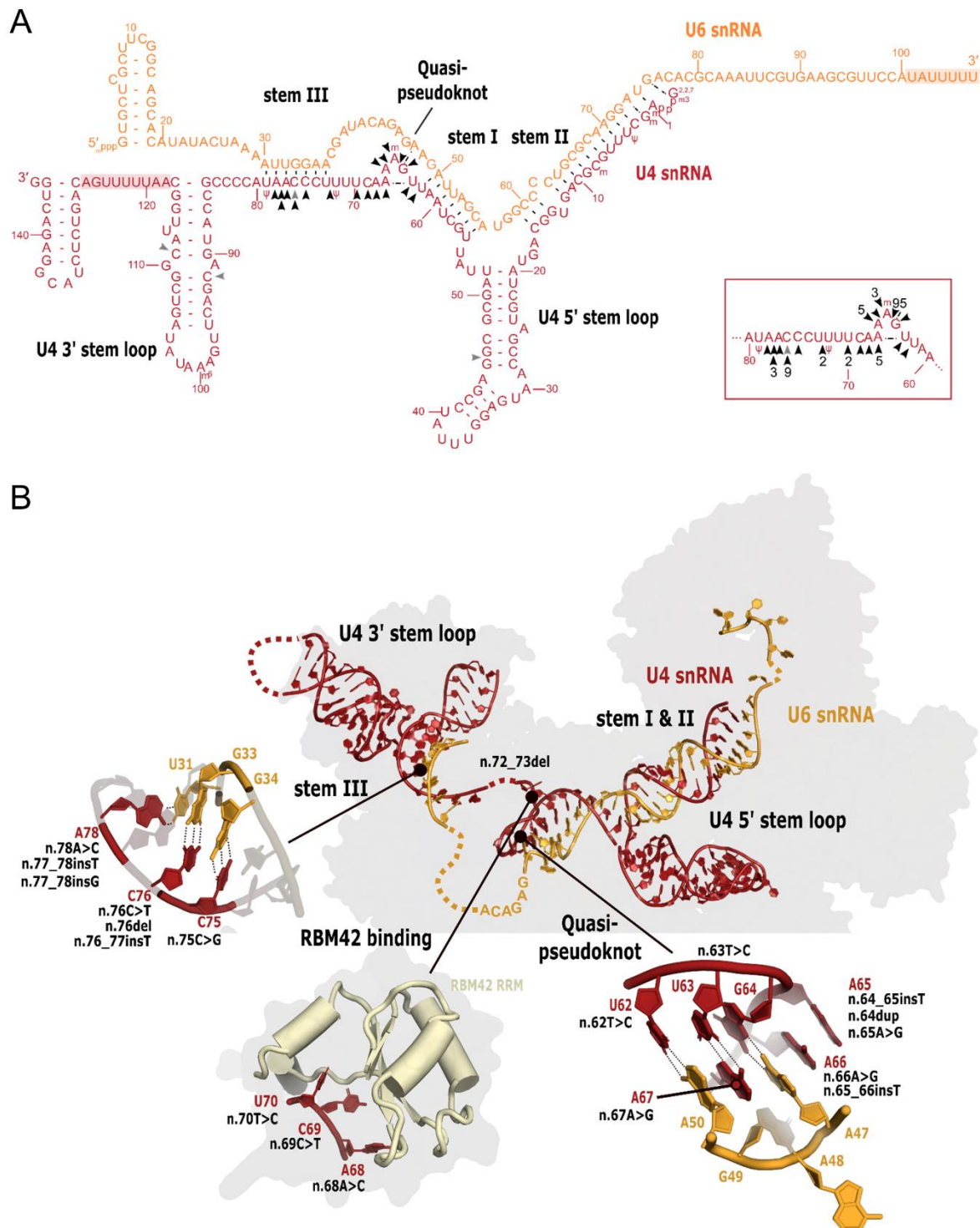
Research Executive Committees Faculty Career Development Fellowship. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. Genome sequencing in Spanish patients was supported by the Undiagnosed Rare Diseases Program of Catalonia (URDCat; 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 to C.C. and A.P. and the Instituto de Salud Carlos III co-funded by the Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea, una manera de hacer Europa (FIS PI20/00758 and PI23/01090) to C.C., and IMPACT-Genomica (IMP/00009) to A.P. We thank the CERCA Program/Generalitat de Catalunya for institutional support. This study uses resources generated by the ENCODE Consortium and Thomas Gingeras lab (CSHL). N.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant no. 220134/Z/20/Z), a research prize from the Lister Institute. Y.C. is supported by a studentship from Novo Nordisk. P.Z.'s work was being supported by Stiftung Michael through the assistance of the Canger-Janz-Fellowship. TSB was supported by ZonMw Vidi, grant 09150172110002), and acknowledges support from Stichting 12q. KÕ was supported by Estonian Research Council grants PRG471 and PRG2040. SLS was supported by a Manton Center for Orphan Disease Research postdoctoral fellowship. SB received a grant from la Fondation pour la Recherche médicale (EQU202203014861).

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

		Total	n.64_65insT	n.76C>T	Other variants in the quasi-pseudoknot or RBM42 interaction	Other variants in the stem III
Patients with clinical data		134*	95	9	20	9
		129*	91	8	20	9
Prenatal findings		64/113 (56.64%)	50/82 (61%)	2/8 (25%)	9/17 (53%)	3/6 (50%)
- IUGR		38/125 (30.4%)	25/89 (28%)	1/8 (12%)	9/20 (45%)	3/8 (38%)
- Cerebral abnormalities		27/88 (30.68%)	25/79 (32%)	0/1 (0%)	2/6 (33%)	0/2 (0%)
Neonatal findings		103/127 (81.1%)	82/91 (90%)	1/8 (12%)	17/20 (85%)	3/8 (38%)
- Neonatal hypotonia		78/124 (62.9%)	61/88 (69%)	0/8 (0%)	15/20 (75%)	2/8 (25%)
- Neonatal feeding problems		66/125 (52.8%)	51/90 (57%)	1/8 (12%)	12/19 (63%)	2/8 (25%)
Congenital microcephaly		26/109 (23.85%)	20/76 (26%)	1/7 (14%)	3/19 (16%)	2/7 (29%)
Microcephaly		87/127 (68.5%)	67/91 (74%)	2/8 (25%)	15/19 (79%)	3/9 (33%)
Short stature		63/126 (50%)	52/90 (58%)	1/8 (12%)	9/19 (47%)	1/9 (11%)
Walking	not achieved	15/104 (14.42%)	10/74 (14%)	0/8 (0%)	5/13 (38%)	0/9 (0%)
	delay	72/104 (69.23%)	57/74 (77%)	4/8 (50%)	7/13 (54%)	4/9 (44%)
	normal age	17/104 (16.35%)	7/74 (9%)	4/8 (50%)	1/13 (8%)	5/9 (56%)
Language ability	non verbal	54/104 (51.92%)	45/75 (60%)	0/8 (0%)	9/12 (75%)	0/9 (0%)
	few words	28/104 (26.92%)	26/75 (35%)	0/8 (0%)	1/12 (8%)	1/9 (11%)
	simple sentences	14/104 (13.46%)	3/75 (4%)	4/8 (50%)	1/12 (8%)	6/9 (67%)
	normal	8/104 (7.69%)	1/75 (1%)	4/8 (50%)	1/12 (8%)	2/9 (22%)
Developmental delay	severe	77/121 (63.64%)	63/88 (72%)	0/7 (0%)	13/18 (72%)	1/8 (12%)
	moderate	35/121 (28.93%)	23/88 (26%)	3/7 (43%)	5/18 (28%)	4/8 (50%)
	mild	8/121 (6.61%)	2/88 (2%)	4/7 (57%)	0/18 (0%)	2/8 (25%)
	no	1/121 (0.83%)	0/88 (0%)	0/7 (0%)	0/18 (0%)	1/8 (12%)
Estimated level of ID	severe	65/103 (63.11%)	53/69 (77%)	0/8 (0%)	12/17 (71%)	0/9 (0%)
	moderate	30/103 (29.13%)	15/69 (22%)	4/8 (50%)	5/17 (29%)	6/9 (67%)
	mild	8/103 (7.77%)	1/69 (1%)	4/8 (50%)	0/17 (0%)	3/9 (33%)
Autism spectrum disorder		49/83 (59.04%)	40/62 (65%)	2/5 (40%)	5/11 (45%)	2/5 (40%)
Stereotypies		67/86 (77.91%)	55/66 (83%)	2/5 (40%)	8/12 (67%)	2/3 (67%)
Epilepsy		75/126 (59.52%)	51/90 (57%)	6/8 (75%)	11/19 (58%)	7/9 (78%)
Febrile seizures		53/88 (60.23%)	38/67 (57%)	7/7 (100%)	4/8 (50%)	4/6 (67%)
Status epilepticus		18/65 (27.69%)	15/48 (31%)	0/5 (0%)	3/9 (33%)	0/3 (0%)
Drug-resistance		10/69 (14.49%)	8/48 (17%)	0/6 (0%)	2/9 (22%)	0/6 (0%)
Abnormal brain MRI		96/118 (81.36%)	78/86 (91%)	1/7 (14%)	15/18 (83%)	2/7 (29%)
- Enlarged ventricles		51	47	0	4	0
- Corpus callosum abnormality		57	44	1	11	1
Cardiac		20/102 (19.61%)	14/76 (18%)	1/7 (14%)	4/14 (29%)	1/5 (20%)
Renal/genitourinary abnormality		20/100 (20%)	14/74 (19%)	1/4 (25%)	4/16 (25%)	1/6 (17%)
Bone/skeletal abnormality		40/111 (36.04%)	33/82 (40%)	1/7 (14%)	4/16 (25%)	2/6 (33%)
Eyes/vision abnormality		62/112 (55.36%)	51/84 (61%)	0/5 (0%)	7/15 (47%)	4/8 (50%)
Hearing loss		8/113 (7.08%)	5/80 (6%)	0/6 (0%)	3/18 (17%)	0/9 (0%)
Teeth/dental abnormality		17/91 (18.68%)	10/65 (15%)	1/5 (20%)	4/15 (27%)	2/6 (33%)
Skin abnormality		26/104 (25%)	18/75 (24%)	3/5 (60%)	4/17 (24%)	1/7 (14%)
Dysmorphic features		103/119 (86.55%)	74/85 (87%)	6/7 (86%)	16/18 (89%)	7/9 (78%)
Feeding issues		73/113 (64.6%)	57/81 (70%)	0/5 (0%)	13/18 (72%)	3/9 (33%)
Failure to thrive		60/114 (52.63%)	47/83 (57%)	1/5 (20%)	10/18 (56%)	2/8 (25%)
Constipation		55/110 (50%)	43/79 (54%)	2/6 (33%)	8/18 (44%)	2/7 (29%)
Joint hyperlaxity		40/102 (39.22%)	31/74 (42%)	0/5 (0%)	6/15 (40%)	3/8 (38%)
Acrocyanosis		18/99 (18.18%)	16/72 (22%)	0/4 (0%)	2/16 (12%)	0/7 (0%)
Blood count abnormality		14/85 (16.47%)	12/63 (19%)	0/5 (0%)	1/12 (8%)	1/5 (20%)

IUGR: intrauterine growth restriction; ID: intellectual disability; MRI: magnetic resonance imaging; \* Includes one patient with a variant (n.62T>C) located in stem loop I. A more detailed table with statistical tests is available as Supplementary Table 7.

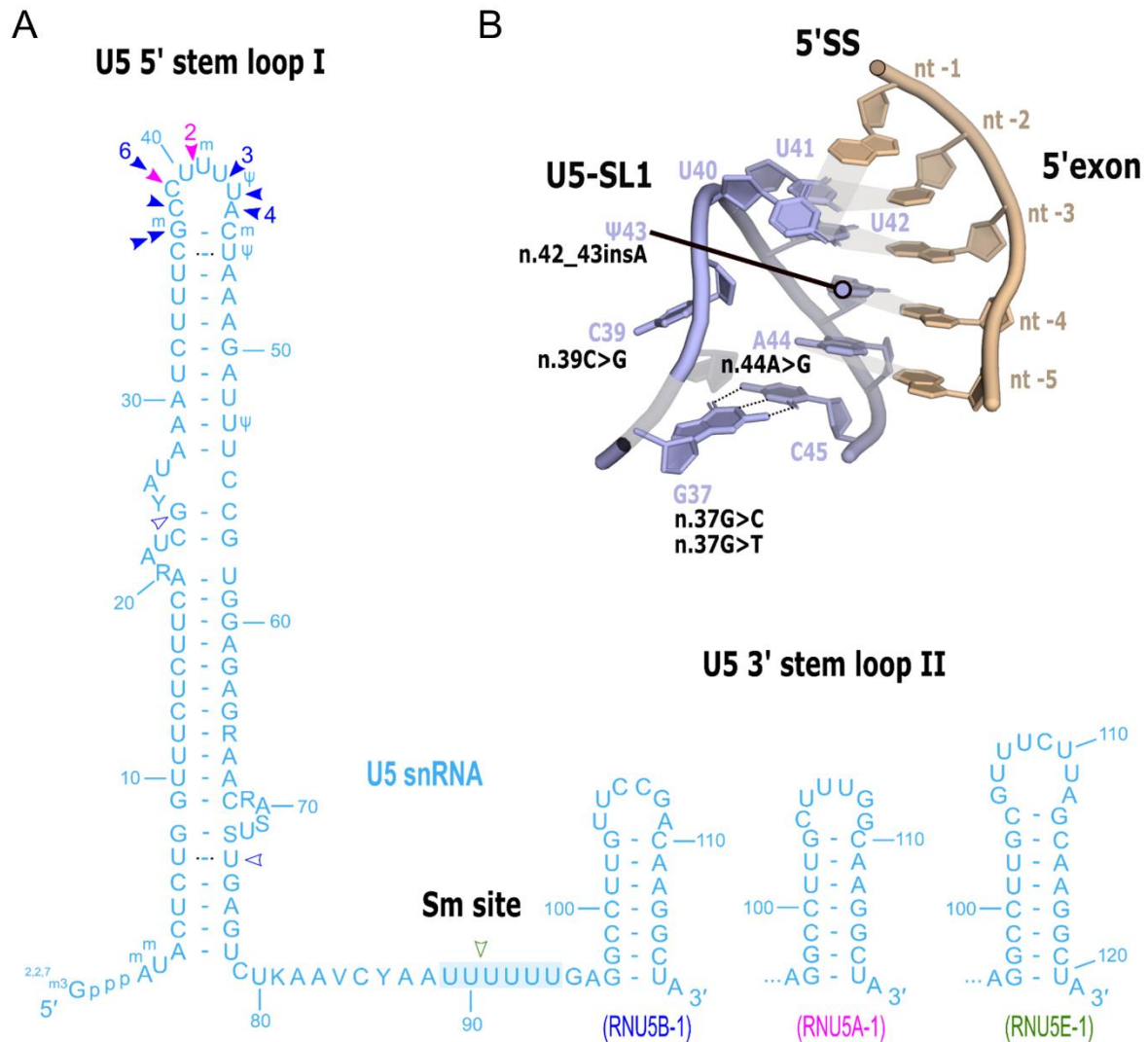
## Figures



**Figure 1: Overview of *RNU4-2* variants identified in this study.** (A) Two-dimensional predicted structure of the interaction between U4 (red) and U6 (orange) small nuclear RNAs showing distinct domains. Arrows indicate variants from this study (pathogenic and likely pathogenic in black, variants of uncertain significance (VUS) in grey). Numbers in black in the



zoom-in represent the number of patients with the variant, for nucleotide changes appearing more than once. The four nucleotide differences between *RNU4-2* and *RNU4-1* are shown using IUPAC codes. Red and orange numbers refer to the numbering of nucleotides from each snRNA.  $\Psi$ : pseudouridine, m: 2'-O-methyl residues; m6: N6-methyladenosine; <sup>2,2,7</sup>m3Gppp: 2,2,7-trimethylguanosine cap; mpppG: 5' guanosine triphosphate cap with gamma-monomethyl phosphate. Shaded regions: Lsm/Sm protein binding sites. **(B)** Organisation of the U4-U6 duplex at the tri snRNP stage (PDB 6QW6) and close-up views of stem III, RBM42 binding and quasipseudoknot regions. Interactions stabilizing these structures as well as mutations potentially affecting their stability are represented.

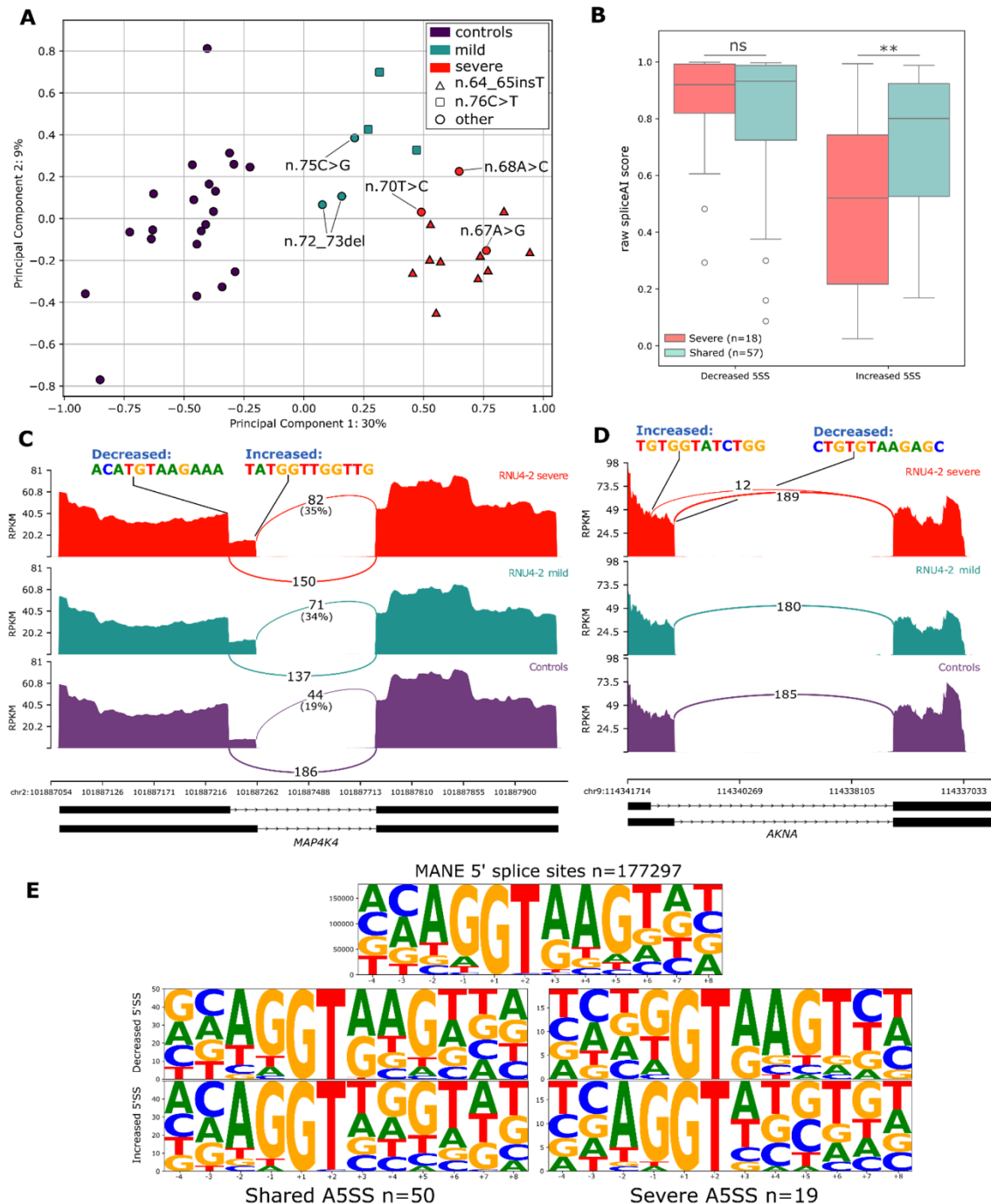


**Figure 2: Overview of *RNU5A-1*, *RNU5B-1* and *RNU5E-1* variants identified in this study.**

(A) Two-dimensional predicted structure of U5 (blue) small nuclear RNA showing distinct domains. Arrows indicate variants from this study (pink: *RNU5A-1*, dark blue: *RNU5B-1*, green: *RNU5E-1*; pathogenic and likely pathogenic with filled color, variants of uncertain significance (VUS) filled in white). Numbers near the arrows represent the number of patients with the variant, for nucleotide changes appearing more than once. Nucleotide differences between *RNU5A-1*, *RNU5B-1* and *RNU5E-1* are shown using IUPAC codes, except for the highly variable 3' stem loop II, for which we show separate loop. Blue numbers refer to the numbering of nucleotides from each snRNA. Ψ: pseudouridine; m: 2'-O-methyl residues; <sup>2,2,7</sup>m<sup>3</sup>Gppp: 2,2,7-trimethylguanosine cap. Shaded region: Sm site. (B) 5' exon recognition by the U5 stem loop I (SL1) at the B-complex stage (PDB 8Q7N).

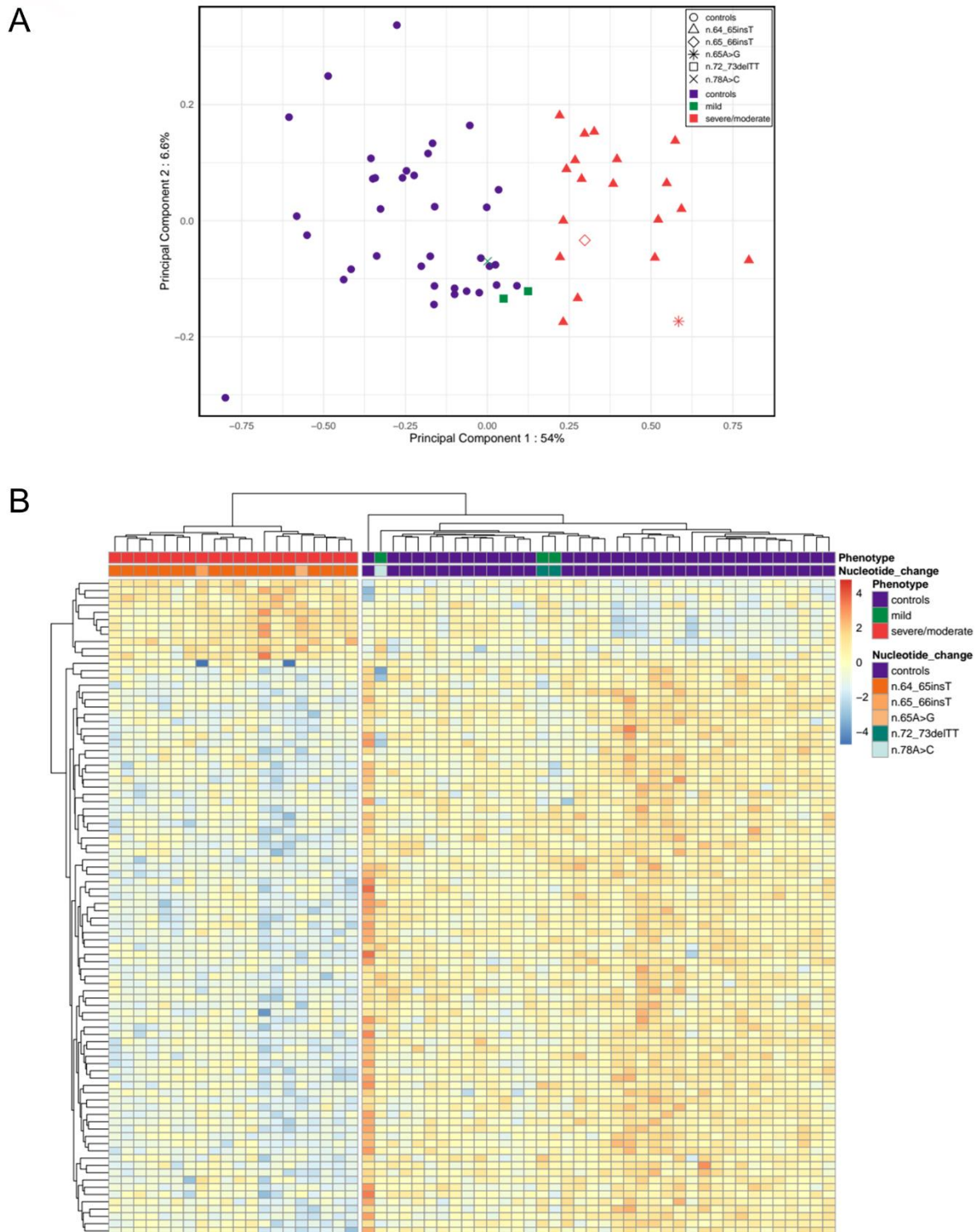


appear for variants other than n.64\_65insT. Colored according to the location within the distinct U4:U6 domains: stem I, light blue; quasi-pseudoknot, orange; RBM42 interaction region, blue; stem III, green). Triangles: P variants, squares: LP variants. (C) Contributions of the clinical features for the PCA.



**Figure 4: RNA-Seq identifies an alternative 5'SS pattern which differentiates severe from mild *RNU4-2*-related phenotypes.** (A) Principal component analysis using PSI values of significant 5'SS detected using rMATS ( $n=111$ ). Colors distinguish controls (purple) from mildly affected (green) and severely affected individuals (red). (B) Box plot showing raw spliceAI scores of the decreased 5'SS site and the increased 5'SS, for the events shared between mild and severe individuals and for those only detected in severe individuals. SpliceAI scores for severe and shared 5'SS were not statistically different for decreased sites ( $p=0.62$ ) but were significant for increased sites ( $p=0.005$ ) using t-test. Box plot elements are defined as follows:

center line: median; box limits: upper and lower quartiles; whiskers:  $1.5\times$  interquartile range; points: outliers. **(C)** Sashimi plot showing aggregated coverage and splicing-supporting reads for ‘mild’ *RNU4-2* variants (n.75C>G, n.76C>T, n.72\_73del), severe variants (n.64\_65insT, n.67A>G, n.68A>C, n.70T>C) and controls, on *MAP4K4* gene. **(D)** Sashimi plot showing on *AKNA* the usage of a 5'SS only for severe variants. **(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).



**Figure 5: *RNU4-2* epismutation differentiates moderate-to-severe NDD phenotypes from controls.** (A) Principal component analysis of residual methylation levels after adjustment for age, sex and estimated blood cell counts. Different variations are represented by different shapes. Phenotypes are distinguished by color, blue for controls, green for mild phenotypes, red for moderate to severe phenotypes. Percentage of explained variance is added to each axis. (B) Heatmap of residual methylation levels displays hierarchical clustering of controls and

*RNU4-2* patients. Blue indicates hypo-methylated positions while red indicates hyper-methylated positions with respect to expected methylation levels at equivalent age, sex and blood cell composition.