

# Spliceosome malfunction causes neurodevelopmental disorders with overlapping features

Dong Li,<sup>1,2,3</sup> Qin Wang,<sup>4</sup> Allan Bayat,<sup>5,6,7</sup> Mark R. Battig,<sup>1</sup> Yijing Zhou,<sup>4</sup> Daniëlle G.M. Bosch,<sup>8</sup> Gijs van Haaften,<sup>9</sup> Leslie Granger,<sup>10</sup> Andrea K. Petersen,<sup>10</sup> Luis A. Pérez-Jurado,<sup>11,12,13</sup> Gemma Aznar-Lain,<sup>13,14</sup> Anushree Aneja,<sup>4</sup> Miroslava Hancarova,<sup>15</sup> Sarka Bendova,<sup>15</sup> Martin Schwarz,<sup>15</sup> Radka Kremlikova Pourova,<sup>15</sup> Zdenek Sedlacek,<sup>15</sup> Beth A. Keena,<sup>2</sup> Michael E. March,<sup>1</sup> Cuiping Hou,<sup>1</sup> Nora O'Connor,<sup>1</sup> Elizabeth J. Bhoj,<sup>1,2,3</sup> Margaret H. Harr,<sup>1</sup> Gabrielle Lemire,<sup>16</sup> Kym M. Boycott,<sup>16</sup> Meghan Towne,<sup>17</sup> Megan Li,<sup>18</sup> Mark Tarnopolsky,<sup>19</sup> Lauren Brady,<sup>19</sup> Michael J. Parker,<sup>20</sup> Hanna Faghfoury,<sup>21</sup> Lea Kristin Parsley,<sup>22</sup> Emanuele Agolini,<sup>23</sup> Maria Lisa Dentici,<sup>24,25</sup> Antonio Novelli,<sup>23</sup> Meredith Wright,<sup>26</sup> Rachel Palmquist,<sup>27</sup> Khanh Lai,<sup>28</sup> Marcello Scala,<sup>29,30</sup> Pasquale Striano,<sup>29,30</sup> Michele Iacomino,<sup>31</sup> Federico Zara,<sup>31</sup> Annina Cooper,<sup>32</sup> Timothy J. Maarup,<sup>33</sup> Melissa Byler,<sup>34</sup> Robert Roger Lebel,<sup>34</sup> Tugce B. Balci,<sup>35</sup> Raymond Louie,<sup>36</sup> Michael Lyons,<sup>36</sup> Jessica Douglas,<sup>37</sup> Catherine Nowak,<sup>38</sup> Alexandra Afenjar,<sup>39</sup> Juliane Hoyer,<sup>40</sup> Boris Keren,<sup>41</sup> Saskia M. Maas,<sup>42</sup> Mahdi M. Motazacker,<sup>43</sup> Julian A. Martinez-Agosto,<sup>44</sup> Ahna M. Rabani,<sup>44</sup> Elizabeth M. McCormick,<sup>45</sup> Marni J. Falk,<sup>3,45</sup> Sarah M. Ruggiero,<sup>46,47</sup> Ingo Helbig,<sup>46,47,48,49</sup> Rikke S. Møller,<sup>5,50</sup> Lino Tessarollo,<sup>51</sup> Francesco Tomassoni Ardori,<sup>51</sup> Mary Ellen Palko,<sup>51</sup> Tzung-Chien Hsieh,<sup>52</sup> Peter M. Krawitz,<sup>52</sup> Mythily Ganapathi,<sup>53,54</sup> Bruce D. Gelb,<sup>55</sup> Vaidehi Jobanputra,<sup>53,54</sup> Ashley Wilson,<sup>53</sup> John Grealley,<sup>56</sup> Sébastien Jacquemont,<sup>57</sup> Khadijé Jizi,<sup>58</sup> Ange-Line Bruel,<sup>59,60,61</sup> Chloé Quelin,<sup>62</sup> Vinod K. Misra,<sup>63,64</sup> Erika Chick,<sup>63</sup> Corrado Romano,<sup>65,66</sup> Donatella Greco,<sup>67</sup> Alessia Arena,<sup>67</sup> Manuela Morleo,<sup>68,69</sup> Vincenzo Nigro,<sup>68,69</sup> Rie Seyama,<sup>70,71</sup> Yuri Uchiyama,<sup>70,72</sup> Naomichi Matsumoto,<sup>70</sup> Ryoji Taira,<sup>73</sup> Katsuya Tashiro,<sup>74</sup> Yasunari Sakai,<sup>73</sup> Gökhan Yigit,<sup>75,76</sup> Bernd Wollnik,<sup>75,76,77</sup> Michael Wagner,<sup>78</sup> Barbara Kutsche,<sup>78</sup> Anna C.E. Hurst,<sup>79</sup> Michelle L. Thompson,<sup>80</sup> Ryan Schmidt,<sup>81,82</sup> Linda Randolph,<sup>82,83</sup> Rebecca C. Spillmann,<sup>84</sup> Vandana Shashi,<sup>84</sup> Edward J. Higginbotham,<sup>85</sup> Dawn Cordeiro,<sup>86</sup> Amanda Carnevale,<sup>86</sup> Gregory Costain,<sup>86</sup> Tayyaba Khan,<sup>86</sup> Benoît Funalot,<sup>87</sup> Frederic Tran Mau-Them,<sup>59,60</sup> Luis Fernandez Garcia Moya,<sup>88</sup> Sixto García-Miñaur,<sup>88</sup> Matthew Osmond,<sup>16</sup> Lauren Chad,<sup>89</sup> Nada Quercia,<sup>90,91</sup> Diana Carrasco,<sup>92</sup> Chumei Li,<sup>93</sup> Amarilis Sanchez-Valle,<sup>94</sup> Meghan Kelley,<sup>94</sup> Mathilde Nizon,<sup>95,96</sup> Brynjar O. Jansson,<sup>97</sup> Patrick Sulem,<sup>97</sup> Kari Stefansson,<sup>97,98</sup> Svetlana Gorokhova,<sup>99,100</sup> Tiffany Busa,<sup>100</sup> Marlène Rio,<sup>101</sup> Hamza Hadj Habballah,<sup>101</sup> Marion Lesieur-Sebellin,<sup>101</sup> Jeanne Amiel,<sup>102,103</sup> Véronique Pingault,<sup>102,103,104</sup> Sandra Mercier,<sup>95,96</sup> Marie Vincent,<sup>95,96</sup> Christophe Philippe,<sup>59</sup> Clemence Fatus-Fauconnier,<sup>105</sup> Kathryn Friend,<sup>106</sup> Rebecca K. Halligan,<sup>107</sup> Sunita Biswas,<sup>107</sup> Jane Rosser,<sup>108</sup> Cheryl Shoubridge,<sup>109</sup> Mark Corbett,<sup>109</sup> Christopher Barnett,<sup>109,110</sup> Jozef Gecz,<sup>109,111</sup> Kathleen Leppig,<sup>112</sup> Anne Slavotinek,<sup>113</sup> Carlo Marcelis,<sup>114</sup> Rolph Pfundt,<sup>114</sup> Bert B.A. de Vries,<sup>114</sup> Marjon A. van Slegtenhorst,<sup>8</sup> Alice S. Brooks,<sup>8</sup> Benjamin Cogne,<sup>95,96,104</sup> Thomas Rambaud,<sup>104</sup> Zeynep Tümer,<sup>115,116</sup> Elaine H. Zackaj,<sup>2,3</sup> Naiara Akizu,<sup>4,117</sup> Yuanquan Song,<sup>4,117</sup> and Hakon Hakonarson<sup>1,2,3</sup>

<sup>1</sup>Center for Applied Genomics, and <sup>2</sup>Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>3</sup>Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA. <sup>4</sup>Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

<sup>5</sup>Department of Regional Health Research, University of Southern Denmark, Odense, Denmark. <sup>6</sup>Department for Epilepsy Genetics and Personalized Medicine, Danish Epilepsy Centre, Dianalund, Denmark.

<sup>7</sup>Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark. <sup>8</sup>Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>9</sup>Department of

Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. <sup>10</sup>Department of Genetics and Metabolism, Randall Children's Hospital at Legacy Emanuel Medical Center, Portland, Oregon, USA. <sup>11</sup>Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Madrid, Spain. <sup>12</sup>Genetic Service, Hospital del Mar Research Institute (IMIM),

Barcelona, Spain. <sup>13</sup>Universitat Pompeu Fabra, Barcelona, Spain. <sup>14</sup>Pediatric Neurology, Hospital del Mar Research Institute (IMIM), Barcelona, Spain. <sup>15</sup>Department of Biology and Medical Genetics, Charles

University Second Faculty of Medicine and University Hospital Motol, Prague, Czech Republic. <sup>16</sup>Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada.

<sup>17</sup>Ambry Genetics, Aliso Viejo, California, USA. <sup>18</sup>Invitae, San Francisco, California, USA. <sup>19</sup>Division of Neuromuscular and Neurometabolic Disorders, Department of Paediatrics, McMaster University Children's

Hospital, Hamilton, Ontario, Canada. <sup>20</sup>Department of Clinical Genetics, Sheffield Children's Hospital, Sheffield, United Kingdom. <sup>21</sup>University Health Network, Toronto, Ontario, Canada. <sup>22</sup>University of Illinois

College of Medicine, Mercy Health Systems, Rockford, Illinois, USA. <sup>23</sup>Laboratory of Medical Genetics, Translational Cyto-genomics Research Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

<sup>24</sup>Medical Genetics Unit, Academic Department of Pediatrics, IRCCS, Ospedale Pediatrico Bambino Gesù, Rome, Italy. <sup>25</sup>Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù,

IRCCS, Rome, Italy. <sup>26</sup>Rady Children's Institute for Genomic Medicine, San Diego, California, USA. <sup>27</sup>Division of Pediatric Neurology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake

City, Utah, USA. <sup>28</sup>Division of Pediatric Pulmonary and Sleep Medicine, University of Utah, Salt Lake City, Utah, USA. <sup>29</sup>Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and

Child Health, Università Degli Studi di Genova, Genoa, Italy. <sup>30</sup>Pediatric Neurology and Muscular Diseases Unit, and <sup>31</sup>Medical Genetics Unit, IRCCS, Istituto Giannina Gaslini, Genoa, Italy. <sup>32</sup>Department of

Genetics, Southern California Permanente Medical Group, Kaiser Permanente, San Diego, California, USA. <sup>33</sup>Department of Genetics, Kaiser Permanente, Los Angeles, California, USA. <sup>34</sup>Center for

Development, Behavior and Genetics, SUNY Upstate Medical University, Syracuse, New York, USA. <sup>35</sup>Division of Genetics, Department of Paediatrics, London Health Sciences Centre, London, Ontario, Canada.

<sup>36</sup>Greenwood Genetic Center, Greenwood, South Carolina, USA. <sup>37</sup>Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA. <sup>38</sup>Division of Genetics and Metabolism, Mass

General Hospital for Children, Boston, Massachusetts, USA. <sup>39</sup>APHP. SU, Reference Center for Intellectual Disabilities Caused by Rare Causes, Department of Genetics and Medical Embryology, Hôpital

Trousseau, Paris, France. <sup>40</sup>Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. <sup>41</sup>Department of Genetics, Hôpital Pitié-Salpêtrière, Paris, France.

<sup>42</sup>Department of Human Genetics, Academic Medical Center, and <sup>43</sup>Laboratory of Genome Diagnostics, Department of Human Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam,

The Netherlands. <sup>44</sup>Division of Medical Genetics, Department of Pediatrics, UCLA, Los Angeles, California, USA. <sup>45</sup>Mitochondrial Medicine Frontier Program, Division of Human Genetics, Department of

Pediatrics,<sup>46</sup>Division of Neurology, and<sup>47</sup>The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>48</sup>Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>49</sup>Department of Biomedical and Health Informatics (DBHI), Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>50</sup>Department of Epilepsy Genetics and Personalized Medicine, Danish Epilepsy Centre, Dianalund, Denmark. <sup>51</sup>Mouse Cancer Genetics Program, Center for Cancer Research, National Cancer Institute (NCI), Frederick, Maryland, USA. <sup>52</sup>Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany. <sup>53</sup>New York Genome Center, New York, New York, USA. <sup>54</sup>Department of Pathology, Columbia University Irving Medical Center, New York, New York, USA. <sup>55</sup>Mindich Child Health and Development Institute and the Departments of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine, New York, New York, USA. <sup>56</sup>Department of Genetics, Albert Einstein College of Medicine, Bronx, New York, USA. <sup>57</sup>Division of Genetics and Genomics, CHU Ste-Justine Hospital and CHU Sainte-Justine Research Centre, University of Montreal, Montreal, Quebec, Canada. <sup>58</sup>Division of Genetics and Genomics, CHU Ste-Justine Hospital and CHU Sainte-Justine Research Centre, University of Montreal, Montreal, Quebec, Canada. <sup>59</sup>INSERM UMR 1231, Genetics of Developmental Anomalies, Université de Bourgogne Franche-Comté, Dijon, France. <sup>60</sup>UF Innovation en Diagnostic Génomique des Maladies Rares, CHU Dijon Bourgogne, Dijon, France. <sup>61</sup>FHU-TRANSLAD, Fédération Hospitalo-Universitaire Translational Medicine in Developmental Anomalies, CHU Dijon Bourgogne, Dijon, France. <sup>62</sup>Medical Genetics Department, Centre de Référence Maladies Rares CLAD-Ouest, CHU Hôpital Sud, Rennes, France. <sup>63</sup>Division of Genetic, Genomic, and Metabolic Disorders, Children's Hospital of Michigan, Detroit, Michigan, USA. <sup>64</sup>Central Michigan University College of Medicine, Discipline of Pediatrics, Mount Pleasant, Michigan, USA. <sup>65</sup>Research Unit of Rare Diseases and Neurodevelopmental Disorders, Oasi Research Institute-IRCCS, Troina, Italy. <sup>66</sup>Medical Genetics, Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy. <sup>67</sup>Oasi Research Institute-IRCCS, Troina, Italy. <sup>68</sup>Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Naples, Italy. <sup>69</sup>Department of Precision Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy. <sup>70</sup>Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. <sup>71</sup>Department of Obstetrics and Gynecology, Juntendo University, Tokyo, Japan. <sup>72</sup>Department of Rare Disease Genomics, Yokohama City University Hospital, Yokohama, Japan. <sup>73</sup>Department of Pediatrics, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. <sup>74</sup>Department of Pediatrics, Karatsu Red Cross Hospital, Saga, Japan. <sup>75</sup>Institute of Human Genetics, University Medical Center Göttingen, Göttingen, Germany. <sup>76</sup>DZHK (German Center for Cardiovascular Research), partner site Göttingen, Göttingen, Germany. <sup>77</sup>Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany. <sup>78</sup>Kinderzentrum Oldenburg, Sozialpädiatrisches Zentrum, Diakonisches Werk Oldenburg, Oldenburg, Germany. <sup>79</sup>Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA. <sup>80</sup>HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. <sup>81</sup>Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, California, USA. <sup>82</sup>Keck School of Medicine of the University of Southern California, Los Angeles, California, USA. <sup>83</sup>Division of Medical Genetics, Children's Hospital Los Angeles, California, USA. <sup>84</sup>Department of Pediatrics–Medical Genetics, Duke University School of Medicine, Durham, North Carolina, USA. <sup>85</sup>Genome Diagnostics, Department of Paediatric Laboratory Medicine, and <sup>86</sup>Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada. <sup>87</sup>Department of Genetics, Hôpital Henri-Mondor APHP and CHI Creteil, University Paris Est Creteil, IMRB, Inserm U.955, Creteil, France. <sup>88</sup>Institute of Medical and Molecular Genetics (INGEMM), Hospital Universitario La Paz, Madrid, Spain. <sup>89</sup>Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. <sup>90</sup>Department of Genetic Counselling, Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Ottawa, Ontario, Canada. <sup>91</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. <sup>92</sup>Department of Clinical Genetics, Cook Children's Hospital, Fort Worth, Texas, USA. <sup>93</sup>Division of Genetics, Department of Paediatrics, McMaster University, Hamilton, Ontario, Canada. <sup>94</sup>Division of Genetics and Metabolism, Department of Pediatrics, University of South Florida, Tampa, Florida, USA. <sup>95</sup>Nantes Université, CHU Nantes, Medical Genetics Department, Nantes, France. <sup>96</sup>Nantes Université, CNRS, INSERM, l'Institut du Thorax, Nantes, France. <sup>97</sup>deCODE genetics/Amgen Inc., Reykjavik, Iceland. <sup>98</sup>Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. <sup>99</sup>Aix Marseille University, Inserm, U1251-MMG, Marseille Medical Genetics, Marseille, France. <sup>100</sup>Department of Medical Genetics, Timone Hospital, APHM, Marseille, France. <sup>101</sup>Department of Genomic Medicine of Rare Disorders, Necker Hospital, APHP Center, University Paris Cité, Paris, France. <sup>102</sup>Rare Disease Genetics Department, APHP, Hôpital Necker, Paris, France. <sup>103</sup>Université Paris Cité, Inserm, Institut Imagine, Embryology and Genetics of Malformations Laboratory, Paris, France. <sup>104</sup>Laboratoire de Biologie Médicale Multi-Sites SeqOIA (laboratoire-seqoia.fr), Paris, France. <sup>105</sup>Reference Center for Hereditary Metabolic Diseases, CHU Dijon Bourgogne, Dijon, France. <sup>106</sup>Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia. <sup>107</sup>Metabolic Clinic, and <sup>108</sup>Department of General Medicine, Women's and Children's Hospital, Adelaide, South Australia, Australia. <sup>109</sup>Adelaide Medical School and Robinson Research Institute, The University of Adelaide, South Australia, Australia. <sup>110</sup>Pediatric and Reproductive Genetics Unit, Women's and Children's Hospital, North Adelaide, South Australia, Australia. <sup>111</sup>South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. <sup>112</sup>Genetic Services, Kaiser Permanente of Washington, Seattle, Washington, USA. <sup>113</sup>Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA. <sup>114</sup>Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. <sup>115</sup>Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark. <sup>116</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. <sup>117</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

**Pre-mRNA splicing is a highly coordinated process. While its dysregulation has been linked to neurological deficits, our understanding of the underlying molecular and cellular mechanisms remains limited. We implicated pathogenic variants in *U2AF2* and *PRPF19*, encoding spliceosome subunits in neurodevelopmental disorders (NDDs), by identifying 46 unrelated individuals with 23 de novo *U2AF2* missense variants (including 7 recurrent variants in 30 individuals) and 6 individuals with de novo *PRPF19* variants. Eight *U2AF2* variants dysregulated splicing of a model substrate. Neuritegenesis was reduced in human neurons differentiated from human pluripotent stem cells carrying two *U2AF2* hyper-recurrent variants. Neural loss of function (LoF) of the *Drosophila* orthologs *U2af50* and *Prp19* led to lethality, abnormal mushroom body (MB) patterning, and social deficits, which were differentially rescued by wild-type and mutant *U2AF2* or *PRPF19*. Transcriptome profiling revealed splicing substrates or effectors (including Rbfox1, a third splicing factor), which rescued MB defects in *U2af50*-deficient flies. Upon reanalysis of negative clinical exomes followed by data sharing, we further identified 6 patients with NDD who carried *RBFOX1* missense variants which, by in vitro testing, showed LoF. Our study implicates 3 splicing factors as NDD-causative genes and establishes a genetic network with hierarchy underlying human brain development and function.**

**Authorship note:** DL, QW, and AB contributed equally to this work and are co-first authors. DL and YS contributed equally to this work and are co-senior authors.

**Conflict of interest:** ML is an employee and shareholder of Invitae, and M Towne is an employee of Ambray Genetics.

**Copyright:** © 2023, Li et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

**Submitted:** April 6, 2023; **Accepted:** October 31, 2023; **Published:** January 2, 2024.

**Reference information:** *J Clin Invest.* 2024;134(1):e171235.

<https://doi.org/10.1172/JCI171235>.

## Introduction

Neurodevelopmental disorders (NDDs) are a heterogeneous group of neurological and related neuropsychiatric conditions including intellectual disability (ID), autism spectrum disorder (ASD), and developmental and epileptic encephalopathies that manifest during childhood (1–3). Over the past 2 decades, there has been an impressive expansion with respect to gene discoveries for NDDs, revealing more than 1,500 genes in different signaling pathways (3, 4), including many transcriptional regulators such as DNA/histone modifiers (5–7) and chromatin-regulatory protein complexes (8–11). De novo variants are increasingly appreciated as playing a substantial causal role in the development of these disorders, and clinical laboratories now routinely report de novo variants in candidate genes not definitively associated with NDD (2, 12), which further facilitates the gene discovery effort. Over the past decade, a genetic diagnosis was obtained, on average, in one-third of individuals with NDD, which is a fundamental step forward in offering biological insights into underlying molecular mechanisms and providing individuals who have an undiagnosed disorder with a prognosis, counseling for recurrence risk, and precision medicine considerations. However, a mechanistic understanding of how these NDD-associated genes are networked and how their malfunction leads to neurodevelopmental or cognitive impairments are major unanswered questions.

Most protein-coding genes in humans are transcribed as pre-mRNAs that contain a series of exons and introns. Following transcription, the removal of introns during the process of pre-mRNA splicing is required before the nascent transcript is translated into a protein. Alternative splicing is a highly coordinated and precise process that involves *cis*-acting exonic and intronic elements (i.e., consensus motif sequences) and numerous *trans* pre-mRNA-binding protein factors, including heterogeneous nuclear ribonucleoproteins (hnRNPs), uridine-rich small nuclear RNPs (snRNPs), the PRP19 complex (also known as the nineteen complex [NTC]), and multiple RNA-binding proteins (RBPs) (13–15). It is widely acknowledged that aberrant splicing of many NDD-related genes due to mutations or dysregulation of *cis*- or *trans*-acting elements/factors has profound influences on neuronal differentiation, neuronal patterning, and synaptic function (16–18), however, germline variants in core spliceosome components have rarely been implicated in NDDs (19–23).

Here, using integration of clinical phenotyping, exome/genome sequencing, protein structure analysis, modeling in flies and human pluripotent stem cells (hPSCs), and transcriptomics, we mapped the genetic architecture of 3 NDD-associated genes, determined their roles in neurodevelopment and behavior, and

validated the pathogenicity of these human variants. We first identified 46 and 6 individuals with undiagnosed disorders, who harbored mostly de novo heterozygous variants in U2 small nuclear RNA auxiliary factor 2 (*U2AF2*), and pre-mRNA processing factor 19 (*PRPF19*), respectively, which encode spliceosomal subunits. We demonstrated that these pathogenic variants lead to converging neurodevelopmental phenotypes, including, but not limited to, developmental delay, ID, and autism. Transcriptome profiling of fly brains deficient in their orthologs *U2af50* and *Prp19*, hPSCs with CRISPR/Cas9 knockin, and patient-derived induced pluripotent stem cells (iPSCs) uncovered potential splicing substrates and downstream effectors, including RNA-binding Fox protein 1 (*Rbfox1*), which encodes an RBP that recognizes specific intronic sequences (UGCAUG and GCAUG) and regulates neuronal alternative splicing (24, 25). Overexpression of the downstream effectors partially rescued the neural defects in *U2af50*-deficient flies, establishing a genetic network of NDD-associated susceptibilities with hierarchy. Importantly, we subsequently identified 6 individuals carrying de novo missense variants in *RBFOX1*, who exhibited similar neurodevelopmental features. Therefore, by combining fly and human genetics, our work reveals a mechanistic network of NDD-associated genes involved in the spliceosome machinery and implicated in phenotypically overlapping disorders and identifies an evolutionarily conserved genetic hotspot for orchestrating neurodevelopment and function.

## Results

*U2AF2 variant carriers have a neurodevelopmental phenotype.* Phenotype and genotype data from 46 unrelated individuals with an NDD and a heterozygous missense variant in *U2AF2* (OMIM 191318) were ascertained through GeneMatcher and Matchmaker Exchange (26, 27). The cohort consisted of 27 female and 19 male probands spanning from 6 months to 24 years of age at the time of data collection, including 1 individual (individual 3) who was recruited from the Deciphering Developmental Disorders (DDD) study ([www.ddduk.org](http://www.ddduk.org)) and 1 (individual 27) who was described elsewhere (28).

The affected individuals had substantial overlapping features, which are summarized in Table 1 and described in more detail in Supplemental Table 1 (supplemental material available online with this article; <https://doi.org/10.1172/JCI171235DS1>). All individuals except 2 (individuals 27 and 33, who died before 6 months of age) had developmental delay ranging from mild to severe, and most of them had global developmental delay involving speech and language, cognition, motor skills, and social behavior. ID was present in 30 of 34 (88%) of the individuals assessed, with a severity ranging from mild ( $n = 22$ ) to severe ( $n = 8$ ) ID. Four study participants (individuals 7, 11, 30, and 36) had borderline ID, with an IQ ranging from 74 to 79. Of the 10 remaining individuals, 7 were too young to perform formal IQ testing, and contact was lost with 3 individuals. Various neurological features were observed in the cohort. Hypotonia was identified in 25 of 37 probands (68%). Seizures occurred in 23 of 41 participants (56%), 15 of whom had febrile seizures (65%). Brain MRI was performed in 34 probands, and 14 (41%) had abnormal imaging findings, including delayed myelination, ventricle asymmetry, atrophy of the cerebral hemisphere, absent olfactory bulbs and tracts, dysmorphic inferior

**Table 1. Summary of the clinical characteristics associated with variants in 3 different genes**

Clinical findings	U2AF2 (n = 46)	PRPF19 (n = 6)	RBFOX1 (n = 6)
	Number of individuals/assessed (and percentage)		
Sex	27 female and 19 male	2 female and 4 male	4 female and 2 male
Developmental delay	44/44 (100%), mild to severe	5/5 (100%)	6/6 (100%), mild
Motor delay	40/42 (95%)	5/5 (100%)	6/6 (100%)
Speech delay	42/42 (100%)	5/5 (100%)	6/6 (100%)
Intellectual disability	30/34 (88%), mild to severe	3/4 (75%), mild to moderate	4/6 (67%)
Hypotonia	25/37 (68%)	3/4 (75%)	3/6 (50%)
Seizures/epilepsy	23/41 (56%)	1/4 (25%)	2/6 (33%)
Febrile seizures	15/23 (65%)	–	1/2 (50%)
Behavioral problem	25/39 (64%)	4/5 (80%)	6/6 (100%)
Autism or autistic features	11/25 (44%)	4/5 (80%)	3/6 (50%)
Brain malformation	14/34 (41%)	0/2 (0%)	1/4 (25%)
Hypoplastic or agenesis of corpus callosum	9/14 (64%)	–	–
Vision abnormalities	20/37 (54%)	3/5 (60%)	3/5 (60%)
Hearing abnormalities	7/38 (18%)	0/5 (0%)	0/6 (0%)
Cardiac abnormalities	10/42 (24%)	1/5 (20%)	1/6 (17%)
Facial dysmorphisms	45/45 (100%)	5/5 (100%)	3/6 (50%)

frontal gyri, and, the most prevalent finding, a thin or poorly developed corpus collosum (9 of 14 individuals). A formal diagnosis of ASD was made in 6 individuals, with 5 additional individuals having autistic features (11 of 39), and a number of different behavioral disturbances (i.e., anxiety, attention deficit disorder [ADD], attention deficit hyperactivity disorder [ADHD], aggression, obsession, stubbornness, tics, etc.) were observed in 22 of 39 individuals (56%), with a total of 64% (25 of 39) of the participants identified as having behavioral issues. Ten individuals had postnatal short stature (height  $\leq -2$  SD; 10 of 40), 5 individuals had microcephaly (head circumference [HC]  $\leq -2$  SD), 3 individuals had macrocephaly (HC  $\geq +2$  SD), and 4 individuals had obesity. Vision abnormality was a prominent feature in this cohort (20 of 37), and hearing loss was less pervasive (7 of 38). In 11 of 42 individuals (26%), heart abnormalities were reported. In 27 of 37 individuals, additional congenital abnormalities were observed and involved skeletal defects affecting the spine and thorax, such as scoliosis, pectus excavatum, butterfly vertebrae, and limb abnormalities, including pes cavus, pes planus, club foot, clinodactyly, brachydactyly, and tapered fingers. Gastrointestinal, genitourinary, metabolic, and endocrine systems were less frequently affected (Supplemental Table 1).

Clinical photographs of 19 participants show several shared craniofacial features, including a prominent/broad forehead, a high anterior hairline, deep-set eyes, hypertelorism, short palpebral fissures, downslanting palpebral fissures, a broad nasal root, a narrow nasal bridge with an upturned nose, thin upper lips, micrognathia, a wide mouth, wide-spaced teeth, and a short neck (Figure 1A), as illustrated in Figure 1B (bottom panel). Moreover, GestaltMatcher (29) analysis revealed that 11 of 19 photos are within a cluster (Supplemental Figure 1A). Notably, 9 of 11 individuals within the cluster carry a variant affecting the residue Arg149 or Arg150, which was also confirmed with the *t*-distributed stochastic neighbor embedding (*t*-SNE) cluster analysis (Supplemental Figure 1B).

These data reveal a neurodevelopmental phenotype associated with *U2AF2* de novo missense variants, with core features of global developmental, speech, and motor delays, mild-to-severe ID, hypotonia, seizures, frequent brain malformations, autistic features, behavioral disturbances, and a shared facial gestalt.

*De novo U2AF2 variants cluster in RNA recognition motifs.* The *U2AF2* variants (NM\_001012478.1) of the 46 probands were identified via exome or genome sequencing and confirmed to be de novo except in 2 instances due to lack of parental DNA. We identified 23 unique variants, of which 22 were missense and 1 was an inframe deletion (Figure 1C). Seven variants

[c.445C>T, p.(Arg149Trp), c.446G>A, p.(Arg149Gln), c.448C>T, p.(Arg150Cys), c.449G>A, p.(Arg150His), c.457G>A, p.(Val153Met), c.556G>A, p.(Val186Met) and c.794G>A, p.(Gly265Asp)] were recurrent and were identified in 30 individuals (65% of participants), representing the hotspots of the mutation spectrum (Figure 1C). Of note, the variants c.445C>T, p.(Arg149Trp) and c.448C>T, p.(Arg150Cys) were hyper-recurrent in 11 and 7 individuals (Supplemental Table 2), respectively, who all exhibited the consistent clinical phenotypes (Supplemental Table 1). All identified variants were absent in population genomics resources (i.e., 1000 Genomes Project, ESP6500SI, and gnomAD, version 2.1.2) and in and in-house data set of more than 10,000 exomes. All variants were predicted to be deleterious by multiple bioinformatics prediction algorithms (MutationTaster, Combined Annotation-Dependent Depletion [CADD], Mendelian Clinically Applicable Pathogenicity [M-CAP], etc.) (Supplemental Table 2). All but 1 variant in this cohort clustered within or near 2 critical RNA recognition motifs (RRMs) (Figure 1C and Supplemental Figure 2A) and all variants affecting residues that were predicted to be highly intolerant to variation, as calculated by the MetaDome server (Figure 1C and Supplemental Table 2). Together, we identified 23 *U2AF2* variants associated with an NDD, including 21 de novo and 7 recurrent variants, in the RRM regions required for RNA recognition.

*U2AF2 variants impair pre-mRNA splicing.* *U2AF2* is an essential pre-mRNA splicing factor for the earliest stage of spliceosome assembly (30). The *U2AF2* RRM regions serve the major function of recognizing polypyrimidine (Py) tract signals marking 3' splice sites (31). Considering the locations of the variants in the *U2AF2* RRM regions, we performed a broader test that included 16 identified and 2 previously described variants (DDD study, ref. 32) distributed across *U2AF2* using a well-characterized minigene splicing reporter assay, which consists of 3 exons and 2 intervening Py tracts (uridine-poor [py] and uridine-rich [PY]) with different splicing efficiencies. We hypothesized that variants occurring physically in close contact with the RNA Py tract would alter the splicing



**Figure 1. Clinical photographs of affected individuals with *U2AF2* variants demonstrating a shared facial gestalt and molecular analyses demonstrating *U2AF2* variants alter pre-mRNA splicing.** (A) Individual identifiers correlated with those in Supplemental Table 1. Shared craniofacial features include a prominent/broad forehead, a high anterior hairline, deep-set eyes, short palpebral fissures, downslanting palpebral fissures, a broad nasal root, a narrow nasal bridge with upturned nose, a thin upper lip, micrognathia, a wide mouth, wide-spaced teeth, and a short neck. (B) Two average faces generated from controls (top) and all 19 available photos of *U2AF2* individuals (bottom) summarizing an identifiable facial gestalt. (C) An intolerance landscape plot generated by MetaDome for *U2AF2* variant (NM\_001012478.1) analysis (top panel) and a lollipop plot (middle panel) with a schematic outline of the *U2AF2* protein domains (lower panel) showing 7 recurrent variants [p.(Arg149Trp), p.(Arg149Gln), p.(Arg150Cys), p.(Val153Met), p.(Arg150His), p.(Val186Met) and p.(Gly265Asp)] and other conserved variants identified in individuals 1–46 and in the DDD study [p.(Pro157Leu) and p.(Thr252Ile)]. (D) Eight *U2AF2* variants reduced expression of the longer isoform in the minigene splicing assay, indicative of the pathogenicity of these variants. Normalized ratios are illustrated by the box-and-whisker plot at the bottom (minimum to maximum, showing all the points).  $n = 3$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , by ordinary 1-way ANOVA with Dunnett's multiple comparisons test.

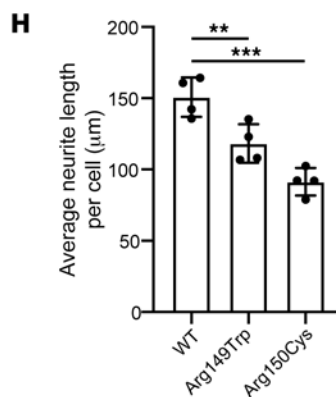
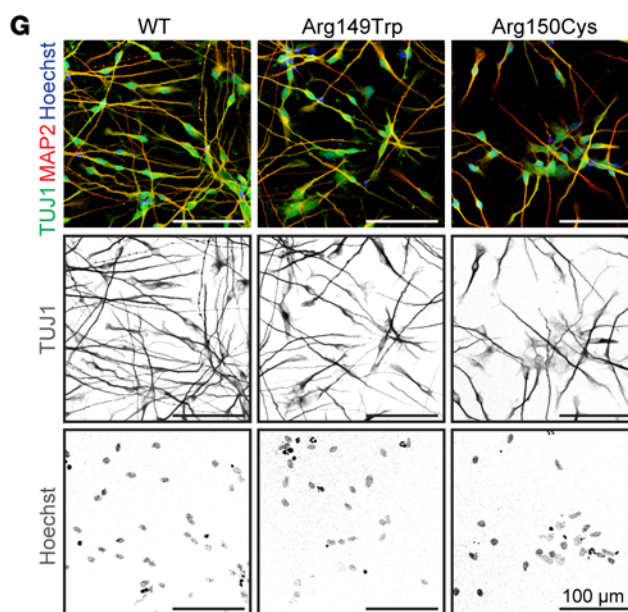
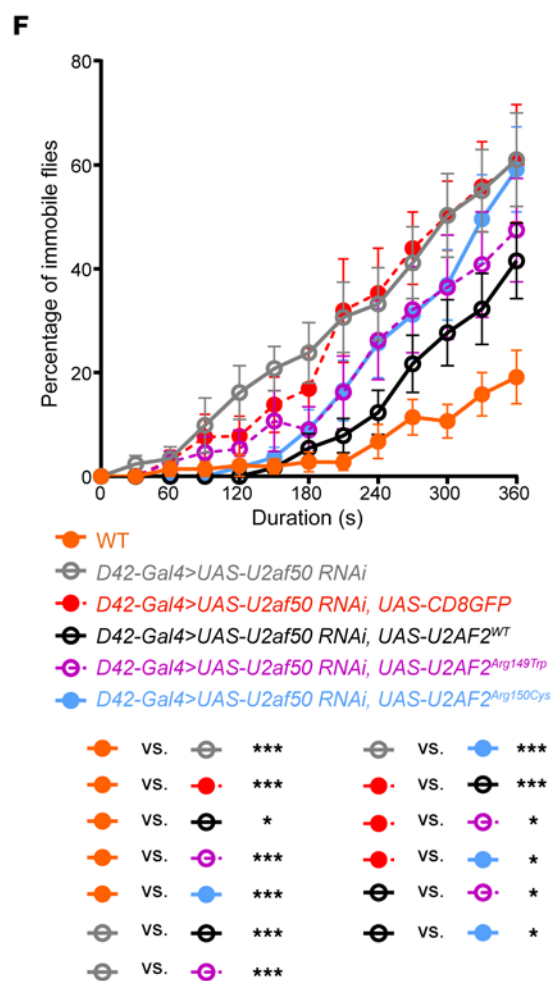
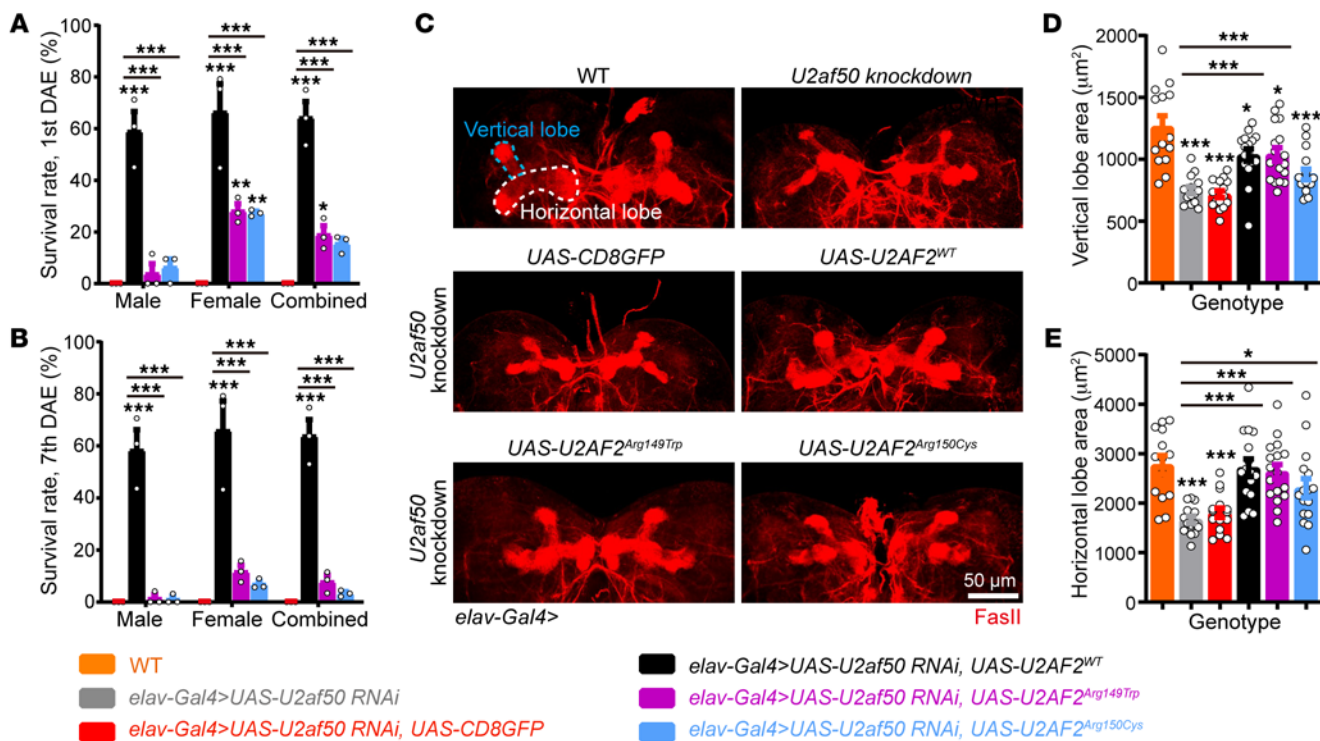
pattern of the *pyPY* transcript. Indeed, we observed that coexpression of 8 of 18 mutant constructs and the *pyPY* minigene in HEK293T cells significantly reduced *py* splicing (Figure 1D), i.e., Arg146Gly, Arg149Trp, Arg150Cys, Arg150His, Pro157Leu, Arg203Cys, Asp215Gly, and Try232His. The abundance of all mutant *U2AF2* proteins was comparable to WT (Supplemental Figure 2B and Supplemental Table 2). In addition, RNA-Seq demonstrated that endogenous *U2AF2* expression levels were unchanged in 4 human iPSC lines reprogrammed from lymphoblastoid cells from individuals 1 (Arg149Trp), 4 (Arg149Trp), 6 (Arg150Cys), and 13 (Lys329del), relative to individual 1's unaffected father and an unrelated control iPSC ( $|\log FC| = 0.19$  and  $P = 0.24$ ).

*U2af50* knockdown leads to lethality and defects in neural morphology and function. Because all of the *U2AF2* patients have neurologic deficits and *U2AF2* is highly expressed in the prenatal and postnatal brain (33), we used an in vivo system to study the physiological significance of the variants. For this, we studied *U2af50*, the *Drosophila* ortholog of *U2AF2*, by assessing the functional effects of *U2af50* knockdown. We first expressed *U2af50 RNAiBL27542* under the control of *elav-Gal4*, a pan-neural driver and found that most offspring died before the end of the pupal stage, with very few surviving to adulthood, but all died shortly after eclosion (Figure 2, A and B). We confirmed the lethality in neural *U2af50*-knockdown flies by crossing *elav-Gal4* with 2 other independent RNAi strains, *U2af50 RNAiv24176* and *U2af50 RNAiBL55153*. Compared with WT controls, neural *U2af50* knockdown using the *elav-Gal4* driver led to a 49% reduction in *U2af50* expression based on RNA-Seq data ( $|\log FC| = 0.98$  and  $P_{adj} = 1.34 \times 10^{-46}$ ). Although the body lengths were comparable between WT and neural *U2af50*-knockdown larvae (Supplemental Figure 3, A and B), we observed a modest decrease in the larval brain area after *U2af50* knockdown (Supplemental Figure 3, C and D). To further dissect whether cell mitosis was responsible for the reduced brain area, we performed immunostaining with the phosphorylated histone H3 (p-H3) antibody, a mitotic marker

(34). We observed that the number of p-H3-positive cells in the brain lobes was significantly decreased in *U2af50 RNAi*-expressing larvae relative to controls (Supplemental Figure 3, C and E). Moreover, in *U2af50*-knockdown brains, the mushroom body (MB), which is the primary learning and memory center, exhibited severe structural defects. Compared with the WT brains, knocking down *U2af50* led to an approximately 40% reduction in both vertical and horizontal lobes of the MB (Figure 2, C–E). To determine if *U2af50* knockdown directly affected MB development, or if the decrease was secondary to generalized brain malformations, we induced expression of *U2af50 RNAi* under the control of *OK107-Gal4*, a MB-specific driver (35). Unlike pan-neural knockdown, knocking down *U2af50* specifically in the MB was not lethal, and in the larval brain the MB showed mild morphological deficits (Supplemental Figure 3, F–H), suggesting that *U2af50* was critical for MB development via mitotic and potentially postmitotic mechanisms. To investigate whether the MB-specific knockdown had an effect on fly behavior, we next subjected the mushroom-specific *U2af50*-knockdown virgin females to the social space assay (36) and found that, while WT flies were dispersed throughout the triangle arena, the *U2af50*-knockdown flies were clustered together and much closer to their nearest neighbor (Supplemental Figure 3, I–K), suggesting disrupted social functioning of adult MB neurons as an impairment in learning and memory, which corroborated with some of the behavioral features observed in patients.

As noted above, 23 individuals had childhood febrile seizures and/or generalized tonic clonic seizures accompanied by motor delay. To test whether knocking down *U2af50* in motor neurons is sufficient to trigger seizure in flies, we crossed the *U2af50 RNAi* fly stock with *D42-Gal4*, a motor neuron driver (37). The motor neuron *U2af50*-knockdown flies were viable and displayed normal motor performance in the negative geotaxis assay (a climbing test), suggesting that their overall locomotor function was not attenuated (Supplemental Figure 3L). We next bathed the fly vials in hot water and recorded the heat-induced paralysis every 30 seconds up to 6 minutes. All flies behaved normally at the onset of the assay but gradually fell to the bottom of the tube, lying on their back with legs twitching. Interestingly, *U2af50*-knockdown flies showed a substantially shorter latency and higher percentage of paralysis. At 6 minutes, 60% of *U2af50*-knockdown flies were immobile, whereas only 20% of the WT flies showed the heat-induced paralysis (Figure 2F), paralleling the frequent seizures seen in patients (56%). Taken together, these phenotypes observed suggest that *U2af50* played an essential role in the *Drosophila* nervous system, and, most important, they mimic some key brain malformations observed in patients carrying *U2AF2* variants.

*U2AF2* variants are pathogenic and exert partial LoF. To investigate the functional impact of the identified *U2AF2* missense variants, we generated transgenic flies with inducible expression of 2 recurrent variants that were found in 11 (*U2AF2*<sup>Arg149Trp</sup>) and 7 (*U2AF2*<sup>Arg150Cys</sup>) individuals, respectively. We first examined whether *U2AF2*<sup>WT</sup> and its variants could rescue the lethality phenotype in neural-specific *U2af50*-knockdown flies. We crossed *UAS-U2AF50*<sup>WT</sup> *UAS-U2af50 RNAi/TM6B* with *elav-Gal4* *UAS-Dcr-2* flies. In the adult offspring expressing *U2AF2*<sup>WT</sup>, 60% also expressed *U2af50 RNAi* (the remaining flies express TM6B instead), and these flies survived beyond 7 days after eclosion, compared with the dose-



**Figure 2. *Drosophila* U2af50-knockdown, hPSCs, and rescue studies with 2 hyper-recurrent variants.** (A and B) Flies expressing CD8GFP in the setting of *U2af50* knockdown with the *elav-Gal4* driver died before or soon after eclosion. Expression of U2AF2<sup>WT</sup> drastically increased the survival rate on the first day after eclosion (DAE) (A), and almost all the flies were still healthy on the seventh DAE (B), whereas variants only slightly increased the survival rate at the first DAE.  $n = 3$  groups, with 17–51 flies per group for males and 21–77 flies per group for females. Data indicate the mean  $\pm$  SEM. \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$ , by 2-way ANOVA followed by Tukey's test. (C–E) Pan-neuronal *U2af50* knockdown led to a decrease in the MB area in larvae, whereas U2AF2<sup>WT</sup> and variants partially (vertical lobe) or fully (horizontal lobe) rescued the phenotype.  $n = 14$ –17 brains. (C) Representative images of MBs. Scale bar: 50  $\mu$ m. (D) Quantification of the vertical lobe area. (E) Quantification of the horizontal lobe area. Data indicate the mean  $\pm$  SEM. \* $P < 0.05$  and \*\*\* $P < 0.001$ , by 1-way ANOVA followed by Dunnett's test. (F) *U2af50* knockdown by the *D42-Gal4* driver resulted in increased heat-induced paralysis. Expressing U2AF2<sup>WT</sup> and variants in *U2af50* knockdown partially attenuated the phenotype.  $n = 6$ –13 groups, with 7–14 flies per group. Data indicate the mean  $\pm$  SEM. \* $P < 0.05$  and \*\*\* $P < 0.001$ , by 2-way ANOVA followed by Tukey's test. (G) Representative immunofluorescence images of induced neurons on day 2. Scale bars: 100  $\mu$ m. (H) Quantification of the average neurite length per cell normalized by the Hoechst-labeled cell number. Graphs show the mean  $\pm$  SD of 4 biological replicates. \*\* $P < 0.01$  and \*\*\* $P < 0.001$ , by ordinary 1-way ANOVA with Dunnett's multiple-comparison test.

controlled RNAi that showed minimal survival (Figure 2, A and B). These data demonstrated that the expression of U2AF2<sup>WT</sup> dramatically increased the survival rate of *U2af50*-knockdown flies. In contrast, expressing either U2AF2<sup>Arg149Trp</sup> or U2AF2<sup>Arg150Cys</sup> only slightly elevated the survival rate of both males and females (Figure 2, A and B). Second, we asked whether the expression of WT or U2AF2 variants could rescue the morphological defects in pan-neuronal *U2af50*-knockdown brains. We found that the expression of U2AF2<sup>WT</sup> as well as U2AF2<sup>Arg149Trp</sup> or U2AF2<sup>Arg150Cys</sup> could fully rescue the proliferation deficit induced by *U2af50* RNAi (Supplemental Figure 3, C and E). Moreover, U2AF2<sup>WT</sup> or U2AF2<sup>Arg149Trp</sup> not only fully rescued the larval brain hemisphere area (Supplemental Figure 3D), but also restored the horizontal lobe area of the MB (Figure 2E). In addition, the vertical lobe area was also partially rescued (Figure 2D). Comparatively, expression of U2AF2<sup>Arg150Cys</sup> did not rescue the brain area, but modestly rescued the area of both horizontal and vertical lobes (Figure 2, C–E). Third, in the heat-induced seizure paradigm, U2AF2<sup>WT</sup> in the setting of *U2af50*-knockdown remarkably reduced the percentage of immobile flies, while the 2 variants only partially rescued the phenotype (Figure 2F). These results strongly suggest that, while WT U2AF2 could largely replace *U2af50* in flies, the 2 variants were partial LoF alleles. It is worth noting that knocking down *U2af50* in fly larval sensory neurons using *ppk-Gal4* (38) resulted in partially thinned/fragmented/degenerative dendrites that could be fully rescued by human WT U2AF2 and variants (Supplemental Figure 3M), suggesting the diverse function of U2AF2/U2af50 in neural circuit assembly and its indispensable role in postmitotic neurons.

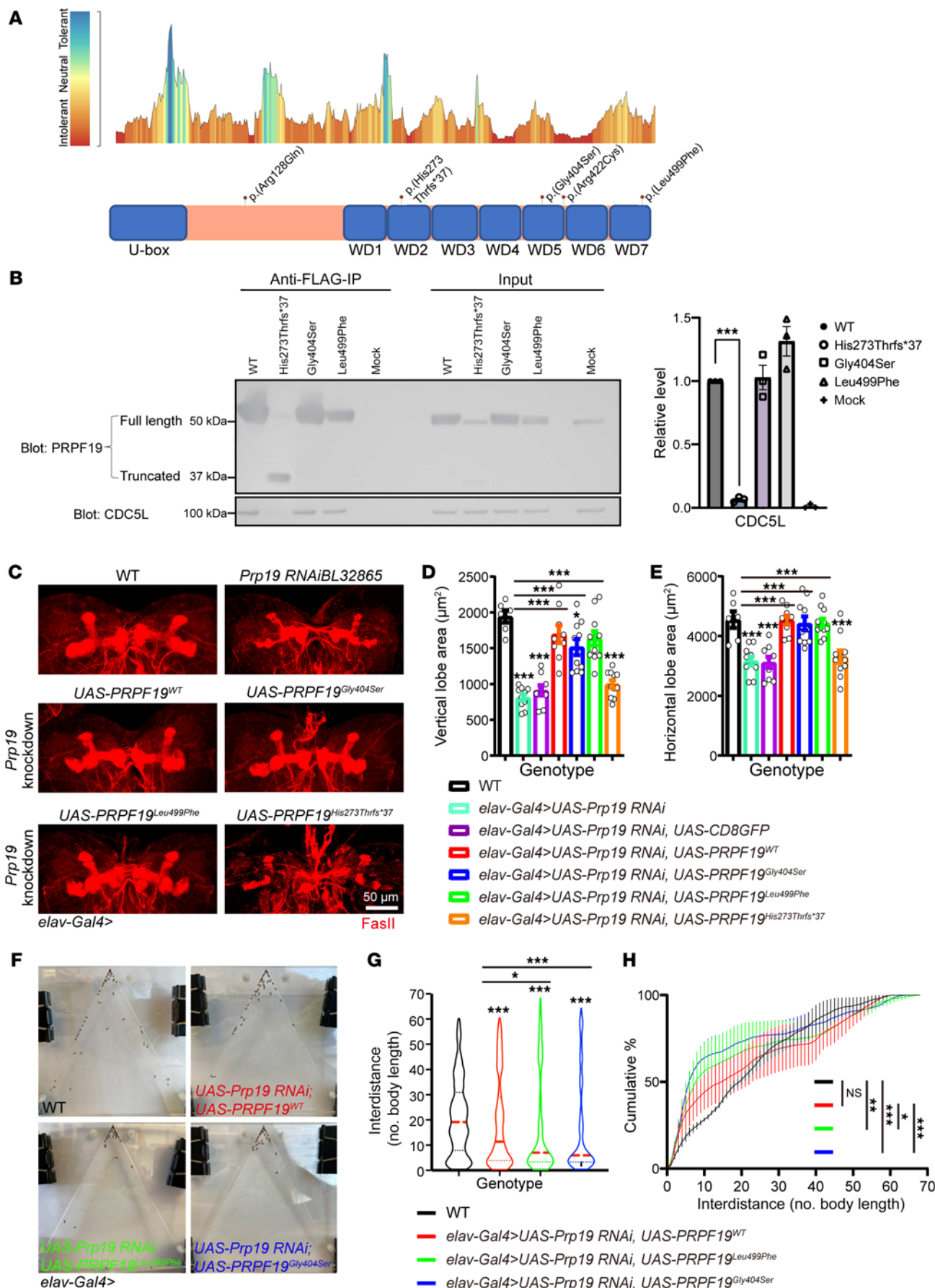
**Neurite growth defect in knockin isogenic hPSCs.** To unveil the effect of patient variants in cortical neuron differentiation, we introduced the U2AF2<sup>Arg149Trp</sup> or U2AF2<sup>Arg150Cys</sup> hyper-recurrent variants into H9 hPSCs using CRISPR/Cas9-based genome editing to generate an isogenic model. The knockin was confirmed by Sanger sequencing and RNA-Seq with the heterozygous state of

patient variants and synonymous variants that prevent recutting with Cas9 (Supplemental Figure 4A). After validation of genome integrity by genotyping and preservation of pluripotency markers by RNA-Seq, knockin hPSCs and WT cells (with synonymous variants) were further differentiated using the NGN2-mediated neuronal induction approach (39). We started to observe neurite outgrowth from both WT and 2 mutant clones within 48 hours. Compared with WT cells, neurons differentiated from mutant hPSCs demonstrated decreased neurite length (Figure 2, G and H). We confirmed that the neurite growth deficit was most likely not caused by protein mislocalization, as the variants did not change the localization of U2AF2 in the nucleus (Supplemental Figure 4B), or by a general defect in proliferation (Figure 2G). Together, these data suggest that U2AF2 has an important role in neurogenesis and neuritogenesis and a partial LoF effect for both variants, which would ultimately alter neuronal connectivity and network activity.

**PRPF19 variants are associated with NDD and autism.** The spliceosomal complex has more than 200 proteins (40), but only a few have been associated with NDDs or neurodegenerative disorders (19–22). Consequently, we considered genes involved in the splicing process as candidates for NDD and reanalyzed research or clinical sequencing data of a cohort of 245 unresolved NDD cases. In 1 previously negative clinical exome, we identified a de novo variant, c.1210G>A, p.(Gly404Ser), in *PRPF19* (P1; NM\_014502.5; OMIM 608330). By GeneMatcher, we identified 5 more patients with heterozygous de novo *PRPF19* variants: c.1495C>T, p.(Leu499Phe) (P2), a 400 kb duplication spanning *PRPF19* (P3), c.1264C>T, p.(Arg422Cys) (P4), c.383G>A, p.(Arg128Gln) (P5), and c.816delT, p.(His273Thrfs\*37) (P6) (Figure 3A and Supplemental Table 3). Individuals P1–P5 all have speech/language and motor delays, and individuals P1–P4 have a formal diagnosis of autism. Moreover, hypotonia was present in 3 individuals (P1, P2, and P4). Three individuals (P1, P4, and P5) had ID and 1 was normal (P3). One individual was too young for ID assessment (P2), and P6 was a fetus with multiple congenital malformations (Table 1 and Supplemental Table 3), substantially overlapping the core features of *U2AF2*-related NDD. All variants were confirmed by Sanger sequencing and were absent in population genomics resources. Prediction of their deleterious effects is summarized in Supplemental Table 2. *PRPF19* is intolerant to both LoF (probability of LoF intolerance [pLI] = 1) and missense ( $Z = 3.89$ ) variants according to the gnomAD population database.

*PRPF19* encodes pre-mRNA processing factor 19, which functions in splicing and more specifically in the catalytic activation and structural stabilization of the spliceosome (41–43). The Prp19 complex, also known as NTC, has PRPF19 as one of the core subunits (i.e., PRPF19, CDC5L, and BCAS2) assembling the other Prp19 complex-associated proteins via C-terminal WD40 domains (44–47). Spliceosome assembly and activation is a dynamic process: the U1 and U2 snRNPs (including U2AF2) mark an intron boundary, then the pre-mRNA binds to U4/U6-U5 tri-snRNP, followed by the release of both U1 and U4 with the arrival of the Prp19 complex to activate the spliceosome (15). However, some reports suggested that the Prp19 complex is recruited to the early spliceosome independent of the tri-snRNP, and that U2AF2 may be involved in the process (48). The 2 missense *PRPF19* vari-





**Figure 3. Modeling PRPF19 in in vitro cellular and in vivo fly models.** (A) An intolerance landscape plot generated by MetaDome for *PRPF19* variant (NM\_014502.5) analysis (top panel) and a schematic outline of the PRPF19 protein domains (lower panel). (B) Western blot and co-immunoprecipitation analysis of overexpressed PRPF19<sup>WT</sup>, PRPF19<sup>Gly404Ser</sup>, PRPF19<sup>Leu499Phe</sup>, and PRPF19<sup>His273Thrfs\*37</sup> with FLAG tag. Graph shows the mean  $\pm$  SEM.  $n = 3$  independent experiments.  $***P < 0.001$ , by 2-tailed, paired  $t$  test. (C–E) Pan-neural *Prp19* knockdown led to reduced MB area. Expression of human PRPF19<sup>WT</sup>, PRPF19<sup>Gly404Ser</sup>, or PRPF19<sup>Leu499Phe</sup> fully or partially rescued the area, while PRPF19<sup>His273Thrfs\*37</sup> failed to attenuate the structural defects.  $n = 6$ –11 brains. (C) Representative images. Scale bar: 50  $\mu$ m. (D) Quantification of the vertical lobe area. (E) Quantification of the horizontal lobe area. Data indicate the mean  $\pm$  SEM.  $*P < 0.05$  and  $***P < 0.001$ , by 1-way ANOVA (D and E). (F–H) In the social behavior assay, flies expressing the 2 *PRPF19* missense variants in the setting of *Prp19* knockdown displayed disrupted social behavior as revealed by their aberrant distribution in the device and reduced interdistance. In comparison, expression of PRPF19<sup>WT</sup> in the setting of *Prp19* knockdown partially restored the impaired social behavior.  $n = 3$  groups, with 26–42 flies per group. (F) Representative images showing fly distribution in the social behavior assay. (G) Quantification of the distance between any of the 2 flies in 1 trial.  $*P < 0.05$  and  $***P < 0.001$ , by 2-tailed, unpaired  $t$  test. (H) Cumulative frequency of fly interdistance in the device.  $*P < 0.05$ ,  $**P < 0.01$ , and  $***P < 0.001$ , by 2-way ANOVA followed by Sidák's test.

ants (Gly404Ser and Arg422Cys) are in the WD40 beta-propeller architecture 5 and 1 (Leu499Phe) in WD40 beta-propeller architecture 7 (Figure 3A). 3D structural modeling (Supplemental Figure 2A, PDB: 4LG8) showed that these substitutions, whose side chains are bigger or smaller than WT residues (49), are expected to destabilize the beta-propeller folding and in turn impair association with NTC-associated proteins. We induced expression of the WT and 3 PRPF19 variants in HEK293T cells, and, as expected, cells transfected with PRPF19<sup>His273Thrfs\*37</sup> showed a complete loss of binding with CDC5L (Figure 3B), suggesting that the variant p.(His273Thrfs\*37) is a complete LoF allele.

*PRPF19* pathogenic variants lead to social behavior deficits. To dissect the pathophysiological mechanism of how *PRPF19* variants contribute to disease, we analyzed fly ortholog *Prp19* with 2 independent RNAi lines (*Prp19* RNAiBL32865 and *Prp19* RNAiV41438). RNA-Seq analysis showed a knockdown efficiency of 77% ( $|\log_{2}FC| = 2.09$  and  $P_{adj} = 1.56 \times 10^{-15}$ ). As expected, all pan-neural *Prp19* knockdown flies died before eclosion (Supplemental Figure 5A) and resulted in a generalized reduction in brain cell proliferation (Supplemental Figure 5B). Of note, overexpression of PRPF19<sup>WT</sup> significantly increased cell proliferation (Supplemental Figure 5B), consistent with the phenotype observed in participant P3, who has an increased HC (97%;  $Z = 1.95$ ) and a tall stature. Since individuals carrying *PRPF19* missense variants have motor deficits, we next examined the locomotor function in pan-neural *Prp19* RNAi larvae in the open-field locomotion assay. We assessed the total crawling length, distance, and speed and found that RNAi larvae had overt hypoactivity (Supplemental Figure 5C). Pan-neural *Prp19* RNAi also led to structural morphology defects in the MB at the larval stage (Figure 3C). Similar to the phenotype caused by *U2af50* knockdown, both the vertical and horizontal lobes of the MB exhibited remarkable area reduction in *Prp19* RNAi larvae (Figure 3, D and E). The expression of PRPF19<sup>WT</sup> or either of the 2 missense variants (PRPF19<sup>Gly404Ser</sup> and PRPF19<sup>Leu499Phe</sup>) could fully or partially rescue the structural defect in the MB and survival rate, with no significant difference in the rescue efficacy of the missense variants compared with WT. However, the truncated form of human PRPF19 [p.(His273Thrfs\*37)] failed to rescue these phenotypes (Figure 3, C–E, and Supplemental Figure 5A), corroborating the fact that the variant c.816delT, p.(His273Thrfs\*37) is indeed a complete LoF allele. Moreover, compared with PRPF19<sup>WT</sup> and PRPF19<sup>Leu499Phe</sup>, which fully rescued the proliferation deficit, PRPF19<sup>Gly404Ser</sup> had reduced rescuing of larval brain proliferation (Supplemental Figure 5B). Nevertheless, PRPF19<sup>WT</sup> and both PRPF19<sup>Gly404Ser</sup> and PRPF19<sup>Leu499Phe</sup> variants fully rescued the lethality in *Prp19* pan-neural knockdown flies (Supplemental Figure 5A), allowing us to conduct the social space assay to further test the effect of these 2 de novo missense variants on social behavior related to the autism diagnosis in patients. We found that female flies expressing either PRPF19 variant exhibited a reduction in social space as determined by the decreased interdistance when compared with flies expressing PRPF19<sup>WT</sup>, which were similar to WT flies (Figure 3, F–H). This result indicated that, although the 2 *PRPF19* missense variants drastically extended the lifespan of *Prp19*-knockdown flies, they did not completely rescue their social behaviors. Thus, Prp19 was required for normal social behaviors of flies, and de novo *PRPF19* variants affected neurons regulating social functions and most likely caused neurophysiological defects.

*Dysregulation of transcripts associated with neurodevelopmental processes.* Since U2AF2 and PRPF19 have converging functions in splicing and variants of both genes cause overlapping NDDs with similar features, we hypothesized that there may be common downstream effectors regulated by these genes. To test this hypothesis, we performed transcriptome profiling of brain tissues harvested at the third instar larval stage from WT controls and *elav-Gal4* driving knockdown flies for *U2af50* and *Prp19*. In *U2af50* RNAi fly brains, we identified 1,219 differentially expressed genes (DEGs) ( $|\log_{2}FC| \geq 0.7$  and  $P_{adj} \leq 0.01$ ; Supplemental Figure 6, A and B), of which 70.6% were upregulated ( $n = 861$ ) and 29.4% were downregulated ( $n = 358$ ). DEGs were subsequently analyzed through the DAVID Functional Annotation Resource, and we found significant enrichment of downregulated DEGs in 9 Gene Ontology (GO) terms, which included axon guidance, negative regulation of neuroblast proliferation, neuron differentiation, and others (Supplemental Figure 6C). Similarly, we identified 1,498 DEGs in *Prp19* RNAi, with 1,085 being upregulated and 413 downregulated. Enrichment analysis similarly revealed 14 significant GO terms ( $|\log_{2}FC| \geq 1$  and  $P_{adj} \leq 0.01$ ; Supplemental Figure 6, D and E), including neuron differentiation, regulation of neurogenesis, regulation of neuroblast proliferation, and axon guidance (Supplemental Figure 6F). These results suggest that both genes play essential roles in neurodevelopmental processes, which could be mapped to fly MB development and patient phenotypes, especially brain function and neurodevelopment.

*Shared splicing alteration by U2af50 and Prp19 knockdown.* To further probe the involvement of the 2 genes in mRNA splicing and prioritize potential common downstream targets, we analyzed differentially used exons (diffUEs) by 3 algorithms, DEXSeq (50), limma (51), and JunctionSeq (52), and retained candidates supported by at least 2 algorithms. Analysis of *U2af50* RNAi RNA-Seq

data with moderate knockdown efficiency (49%) revealed 234 diffUEs in 185 unique genes (Supplemental Table 4), and *Prp19* RNAi (77% knockdown efficiency) led to 1,090 diffUEs in 727 unique genes (Supplemental Table 5). The numbers of affected exons are consistent with the different knockdown efficiencies and also with the respective functions of *U2af50* and *Prp19* orthologs in splice site selectivity (31) and general spliceosome activation (53). Gene set enrichment analysis of these 727 genes similarly demonstrated that the significant gene categories hit many neurodevelopmental processes, such as axon guidance, dendrite morphogenesis, MB development, and axonogenesis (Supplemental Figure 6G). We intersected significant diffUEs across the 2 data sets and identified 117 shared diffUEs. Four exon candidates with increased inclusion were then selected for real-time RT-qPCR validation, due to the implication of their human orthologous genes in human rare disorders. These exons belong to the genes *Iswi*, *Brm*, *RpS19a*, and *Rbfox1*, and increased inclusion of all 4 and 3 exons within each gene were confirmed in *U2af50* and *Prp19* RNAi, respectively (Figure 4, A and B), suggesting that they are potential substrates of the spliceosome.

We hypothesized that increased inclusions of noncanonical exons (i.e., alternative first exons) decrease the expression of canonical transcript(s), and compensating the diminished expression of the canonical isoform in fly pan-neurons could alleviate the MB defects. Remarkably, 3 genes tested, *Iswi*, *Rbfox1*, and *RpS19a*, could rescue the MB defects observed in flies expressing *U2af50* RNAi to a different extent. Expressing *Iswi*, *Rbfox1*, or *RpS19a* in the setting of *U2af50* knockdown significantly increased the area of the horizontal lobe of the MB. Strikingly, expression of *Iswi* in *U2af50* RNAi fully restored the area of both vertical and horizontal lobes (Figure 4, C-E). However, *Iswi* and *Rbfox1* failed to rescue the mushroom defects in *Prp19*-knockdown flies (Supplemental Figure 7). These results demonstrate both hierarchy and specificity in *U2AF2* and *PRPF19* substrates or downstream effectors, depicting the complexity of the genetic network mediating neural development.

*Mis-splicing events further revealed by hPSC RNA-Seq.* To isolate the effect of a single variant change, we performed transcriptomic profiling of the hPSC isogenic model with 24 samples (10 colonies of Arg149Trp, 4 colonies of Arg150Cys, and 10 WT colonies; Supplemental Figure 6H). Interestingly, we found 62 diffUE events ( $|\log_2FC| > 0.7$  and  $P_{adj} \leq 0.01$ ) in 39 unique genes (Supplemental Table 6 and Supplemental Figure 6, I-K). Gene enrichment analysis of these 39 genes revealed that only RNA binding under molecular functions was enriched ( $P_{adj} = 0.017$ ). Real-time reverse transcription quantitative PCR (RT-qPCR) confirmed 3 randomly selected exon targets, i.e., *RPL37*, *DNAJC21*, and *OTX2* (Supplemental Figure 6, I-K). These data suggest that the 2 pathogenic variants (Arg149Trp and Arg150Cys) residing in the interaction surface with RNA could result in aberrant mRNA splicing in stem cells, a relevant cell type to model the pathophysiology.

*Missense variants in RBFOX1 associated with a mild NDD.* Guided by these functional data, and in particular our results showing that RBFOX1 acted as a potential substrate and downstream effector of spliceosome malfunction, reanalysis of clinical negative exome data indeed revealed a de novo variant (c.353G>A, p.Arg-118Gln) in *RBFOX1* (NM\_018723.4; OMIM 605104) in a patient with mild global developmental delay, hypotonia, ataxic gait,

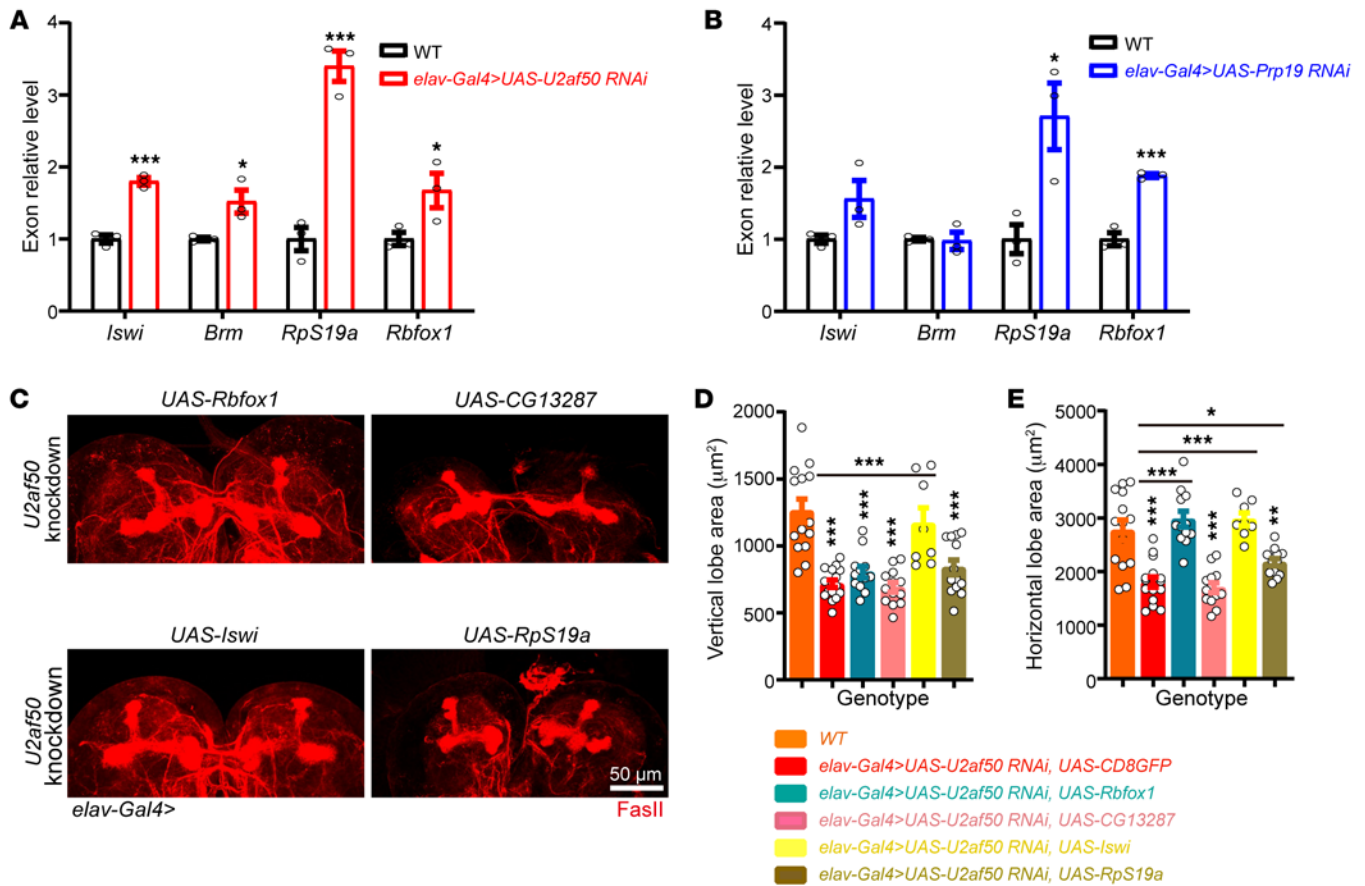
delayed myelination, and brain frontal volume loss, overlapping with the clinical symptoms observed in patients with *U2AF2* variants. Through GeneMatcher, we recruited 5 additional patients. All 6 patients presented with mild neurological symptoms ranging from normal intelligence and a mild DD affecting verbal and motor skills on the mild end of the spectrum to mild-to-moderate ID, autism spectrum disorder, and seizures on the severe end of the spectrum (Table 1 and Supplemental Table 7). All *RBFOX1* variants were confirmed to be de novo, with the p.(Arg118Gln) variant being recurrent in 4 unrelated individuals (Figure 5A), and other 2 individuals having the c.460G>T, p.(Gly154Cys) and c.559G>A, p.(Glu187Lys) variant. RBFOX1 is not directly involved in the core spliceosome complexes but has a single RRM domain that binds to specific sequences [5'-(U)GCAUG-3'] to promote inclusion or exclusion of exons (24, 25) and is expressed mainly in the nervous system, heart, and muscle (25, 54). All identified missense variants occurred in the only RRM domain, and multiple algorithms predicted potential deleterious effects (Supplemental Table 2). Moreover, inspection of the protein structure (Supplemental Figure 2A, PDB: 2N82) demonstrated that all 3 missense variants are in a region of close contact with RNA, as for many *U2AF2* variants, suggesting that the RNA-binding affinity may be impaired.

To functionally test the pathogenicity of these variants, we utilized an in vitro minigene system sensing the alternative splicing activity of *Rbfox1*. Since expression of 2 *TrkB* isoforms by the *TrkB* locus is regulated by *Rbfox1* (55), we generated a 164 kb bacterial artificial chromosome (BAC) containing part of the mouse *TrkB* locus including the alternatively spliced exons (Figure 5B). Cotransfection of the minigene and either the murine *Rbfox1* or the human RBFOX1 in HEK293T cells caused a significant upregulation in the levels of the *TrkB* full-length (*TrkB.FL*) specific isoform (Figure 5C). As expected (25), a known mouse *Rbfox1* pathogenic variant (F158A) did not change *TrkB.FL* expression levels (Figure 5C). Interestingly, RBFOX1<sup>Arg118Gln</sup>, RBFOX1<sup>Gly154Cys</sup>, and RBFOX1<sup>Glu187Lys</sup> expressed at a level similar to that in the human WT construct did not modulate *TrkB.FL* isoform expression (Figure 5, C and D). Altogether, these data suggest that these missense variants uncovered in the RRM domain of RBFOX1 impaired its splicing function, and these pathogenic de novo missense variants in the RRM domain were associated with a mild NDD (Supplemental Table 2).

## Discussion

We describe 3 NDDs caused by de novo heterozygous variants in *U2AF2*, *PRPF19*, and *RBFOX1*, all of which are involved in the pre-mRNA splicing machinery. *U2AF2* and *PRPF19* encode core spliceosome subunits to guide the early stage of splice-site choice and spliceosome activation, respectively. *RBFOX1* encodes a tissue-specific splicing factor that regulates splicing through binding to the pre-mRNA *cis*-acting element 5'-(U)GCAUG-3'. Variants in *U2AF2* and *RBFOX1* clustered predominantly in RNA recognition motifs, while a majority of *PRPF19* variants were in the WD40 domain. Remarkably, all of the patients converge on similar neurodevelopmental phenotypes (Table 1), including mild-to-severe developmental delays, intellectual disability, hypotonia, seizures, and autism.

A group of rare Mendelian diseases caused by pathogenic variants in genes encoding core spliceosomal complexes have



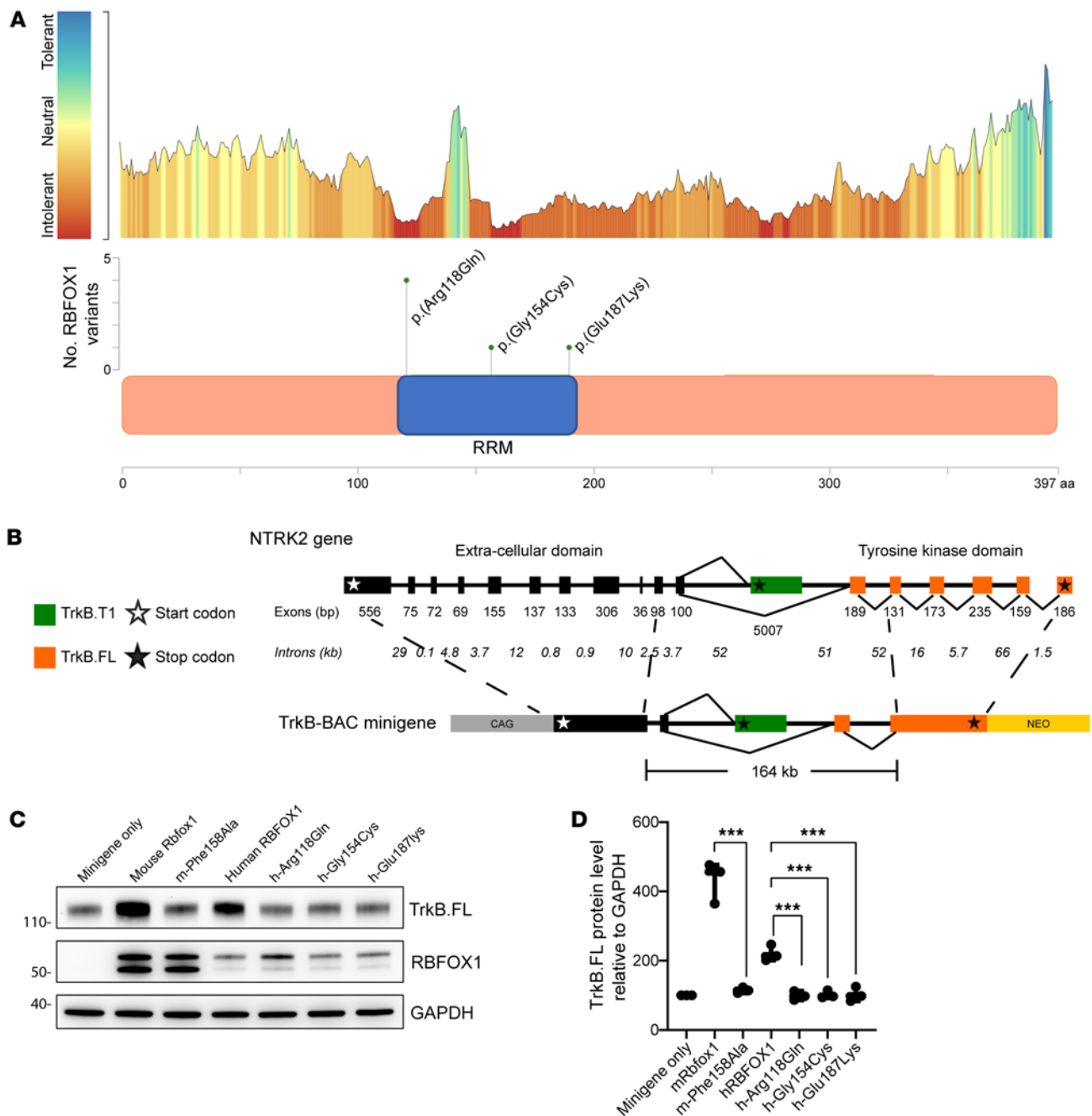
**Figure 4. RT-qPCR confirms potential substrates or downstream effectors that robustly rescue MB defects in *U2af50*-knockdown brains.** (A and B) RT-qPCR data confirmed 4 exon candidates from fly brain RNA-Seq and show that *Iswi*, *Rbfox1*, and *RpS19a* might be the shared downstream splicing targets of *U2af50* and *Prp19*.  $n = 3$ . (A) Compared with WT, increased exon inclusions in *Iswi*, *RpS19a*, and *Rbfox1* were significantly upregulated in *U2af50*-knockdown brains.  $*P < 0.05$  and  $***P < 0.001$ , by 2-tailed, unpaired *t* test. (B) The expression level of the same differentially used exons in *Iswi*, *RpS19a*, and *Rbfox1* was also upregulated in *Prp19*-knockdown brains. Data indicate the mean  $\pm$  SEM.  $*P < 0.05$  and  $***P < 0.001$ , by 2-tailed, unpaired *t* test. (C–E) The expression of *Rbfox1*, *Iswi*, or *RpS19a* fully or partially rescues the decreased MB area in *U2af50*-knockdown brains.  $n = 8–16$ . (C) Representative images. Scale bar: 50  $\mu\text{m}$ . (D) Quantification of the vertical lobe area.  $***P < 0.001$ , by 1-way ANOVA followed by Dunnett's multiple-comparison test. (E) Quantification of the horizontal lobe area. Data indicate the mean  $\pm$  SEM.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ , by 1-way ANOVA followed by Dunnett's multiple-comparison test.

been characterized as spliceosomopathies, e.g., SF3B4 involved in the U2 snRNP in Nager syndrome, an acrofacial dysostosis; SNRPB of the U1/2/5/4/6 snRNP in cerebrotostomandibular syndrome, characterized by Robin sequence and rib defects; and U4/U6 snRNP subunits (PRPF3, PRPF4, and PRPF31) and U5 snRNP subunits (PRPF6, PRPF8, and SNRNP200) in retinitis pigmentosa. Most recently, syndromic intellectual disability has been described as an emerging phenotype linking variants in 2 genes, *PUF60* and *FAM50A*, which encode spliceosomal components of the U2 snRNP and the C-stage spliceosome complexes, respectively (22, 56). Our study, by combining human genetics and in vitro molecular and hPSC cellular models, together with the fly in vivo model provides strong evidence supporting a critical role for the spliceosome in neurodevelopment (Supplemental Table 2).

Despite the major role of U2AF2 in pre-mRNA splicing, facilitating the recognition of exon-intron boundaries and identification of the 3' splice site during the spliceosome assembly, until recently, U2AF2 passenger somatic variants have only been loosely linked to cancer (57). The identification of missense variants in

*U2AF2*, coupled with the extensive in vitro and in vivo functional data presented here, has broadened our understanding of its role in neurodevelopment and in splicing biology. It is worth noting that, while this work was in progress, a large NDD cohort study (including the DDD data set) and a few case reports nominated an association between *U2AF2* and NDD but with limited clinical information and without functional validation (58–61). It has been suggested that partial RNAi knockdown of the fly *U2af50* results in reduced neuroblasts in the larval brain associated with blunted neuroblast lineages and abnormal neuroblast shape (62). We further demonstrated that 2 hyper-recurrent *U2AF2* variants reduced neurogenesis in the hPSC isogenic model and resulted in MB patterning defects, seizures, and social deficits in the fly model.

PRPF19 is an indispensable core subunit of the dynamic hPrp19 complex. Knockout of its mouse ortholog has been shown to cause lethality at an early stage of embryonic development (63). We also found that depleting *Drosophila dPrp19* with pan-neural or motor neuron-specific Gal4 caused lethality before eclosion. Even though our PRPF19 sample size was modest ( $n = 5$ ), we dis-



**Figure 5. The identified *RBFOX1* missense variants alter the gene-splicing pattern.** (A) Intolerance landscape plot generated by MetaDome for *RBFOX1* (NM\_018723.4) variant analysis (top panel) and a lollipop plot (middle panel) with a schematic outline of the *RBFOX1* protein domains (lower panel) showing 1 recurrent variant, p.(Arg118Gln), and other conserved variants identified in individuals 5 and 6. (B) Schematic representation of the murine *TrkB* gene (top) and the *TrkB*-BAC minigene (bottom). The exons in black are commonly expressed in both the full-length (*TrkB.FL*) and the truncated T1 (*TrkB.T1*) isoforms. The *TrkB.T1* isoform was generated by including the specific T1 exon (green), whereas the *TrkB.FL* isoform was generated by including exons in orange. The 164 kb *TrkB*-BAC minigene includes upstream, the transmembrane coding exon and, downstream, in addition to the *TrkB.T1* coding exon, the 2 exons encoding the juxtamembrane region preceding the tyrosine kinase region. The cDNAs coding for the missing extracellular domain and the tyrosine kinase region of *TrkB* were fused inframe to an upstream exon (98 bp) and a downstream exon (131 bp) of the BAC region (dashed lines). A neomycin resistance cassette (NEO) is present in the minigene for selection, while a synthetic CAG promoter drives the minigene expression. (C) Representative immunoblot analysis of the clonal cell line expressing the *TrkB*-BAC minigene transiently transfected with plasmid vectors expressing the mouse *Rbfox1* and the human *RBFOX1* (as positive controls), the F158A mutant (as a negative control), and the de novo *RBFOX1* variants. *Ntrk2* (*TrkB*) protein levels were tested 48 hours after transfection with an antibody against the *TrkB* intracellular domain to specifically detect the full-length protein (*TrkB.FL*). (D) Immunoblot quantification analysis of *TrkB.FL* protein levels from 3 independent experiments, as in C.  $n = 3$ . Data indicate the mean  $\pm$  SEM. \*\*\* $P < 0.001$ , by ordinary 1-way ANOVA with Dunnett's multiple-comparison test.

covered 3 types of variants and 2 pathogenic mechanisms besides LoF. The whole gene duplication resulted in elevated proliferation in brain mitotic neurons, which correlated with an NDD phenotype with macrocephaly and tall stature in participant P3. The frameshift variant truncated the WD40 domain, leading to a complete loss of hPrp19 complex formation with CDC5L, which highly suggested that its canonical splicing function may have been lost. In the fly rescue study, the truncated form failed to reinstate normal MB patterning. This may explain why the fetus (P4) with p.(His273Thrfs\*37) has the most severe congenital abnormalities (Supplemental Table 3). Absence of any other LoF alleles (including intragenic deletion) in publicly available population genomic data sets and our in-house genomic data suggests that *PRPF19* is an essential gene for human embryonic development, and the fetus P4 is in that sense unique. By contrast, 2 human missense variants could largely replace *dPrp19* and rescue the MB structural defect and adult lethality phenotypes. However, they failed to restore normal social interaction. In addition to the essential role in splicing, *PRPF19* could form different complexes that are implicated in the DNA damage response, transcription elongation, mRNA export, and protein degradation (45). Whether other functions of *PRPF19* contribute to the phenotypes observed in our study should be further investigated. Nevertheless, by characterizing 4 variants (2 missense, 1 LoF, 1 microduplication) at different cellular levels, we demonstrated the crucial function of *PRPF19* in neurodevelopment and provide a promising tool to study the effect of pathogenic variants in vivo. This highlights the finding that our fly model is robust for interrogating not only general neural developmental defects, but also specific and more subtle defects that are spatiotemporally dependent.

It is noteworthy that *PRPF19* missense variants cause clustering or reduce inter-fly distance. Similarly, knockdown of *U2af50* in the MB resulted in increased social aggregation in adult female flies. This suggests that the spliceosome complex is critical for normal development of the MB, which is the center for integrating and processing diverse cues, especially the visual and olfactory inputs that are most important for fly social behavior. Loss of *U2af50* or *Prp19* may lead to compromised MB function in neural processing and in the regulation of higher-order social behavior, mimicking the patients with impaired intellectual functions. Interestingly, different from the *elav-GAL4>UAS-ISWI* RNAi flies, which exhibited increased social space in relation to genetic controls (64), our results resembled what was observed in *Drosophila* neurologin 4-deficient (dNlg4-deficient) flies (65), which also show an increased tendency to aggregate. This corroborates the wide range of genetic causes for ID or ASD and their spectrum of social deficits.

Our work aims to establish a genetic network associated with NDD, through gene discovery, functional validation, and mechanistic inquiry. We have successfully pinpointed several critical nodes encompassing the U2AF2/PRPF19/RBFOX1 axis. We have revealed hierarchy among the various nodes, and even within the spliceosome complex, with U2AF2 and PRPF19 being closer to the master regulator and RBFOX1 being their substrate and executor. This represents the first layer of the nodes. We have further identified an additional substrate for the spliceosome complex, namely *Iswi*/SMARCA5/SMARCA1, and downstream effectors

such as Rps19a, representing layers 2 and 3 of the nodes. The rescue efficiency of these factors correlates well with their position in the hierarchy. A pressing question is how mutations in a general machinery, such as the spliceosome, lead to relatively specific malfunctions in the nervous system. We would like to postulate at least a few possible scenarios. The expression of the spliceosome complex subunits may be unevenly distributed across tissue types, and there may be enriched expression of U2AF2/PRPF19/RBFOX1 in the brain. There may also be a possible enrichment of the substrates in the nervous system, especially during the critical time window of neurodevelopment. Alternatively, the composition of coregulators or the presence of modulators may differ among cell types. Intriguingly, it was reported that stringent conservation of alternative splicing is present in only a subset of tissues including the brain and is frequently lineage specific (66). Concomitantly, alternative splicing is also the highest in the brain compared with other tissues (67, 68), and is critical for producing neural circuit diversity and specificity. It may be possible that the dysfunction of mutated spliceosome proteins in the brain is less efficiently compensated, given its more stringent configuration. The high energy consumption in the brain might be another potential contributor, which might affect the dynamics, activity, and throughput of the spliceosome, especially considering the fact that the spliceosome is an energy-intensive machinery (69). During brain development, multiple genes have to be expressed in a regulated and coordinated fashion. A defect in spliceosomes will perturb this machinery, leading to a developmental stagnation of the brain's finer functions. A key step forward will be to systematically dissect more nodes in the network, characterize their spatiotemporal expression, and map out their functional interactions.

In summary, our study implicates the spliceosome in neurodevelopment. It has been widely acknowledged that alternative pre-mRNA splicing is an important mechanism for regulating gene expression for multiple essential genes in the CNS, where it both affects neuronal differentiation and development and controls functions in the mature neurons. The data described in this study advance our understanding of the mechanistic underpinnings of neurodevelopment and neuropsychiatric traits and identify 3 genes to add to the causal gene list in NDDs.

## Methods

**Statistics.** The data shown represent the results of at least 3 separate experiments. All results are presented as the mean  $\pm$  SEM or the mean  $\pm$  SD. Comparisons between 2 groups were analyzed by *t* test, and comparisons between 3 groups were done by 1-way ANOVA followed by Dunnett's test or 2-way ANOVA followed by Turkey's or Šidák's test using GraphPad Prism (GraphPad Software). A *P* value of less than 0.05 was considered statistically significant.

**Study approval.** Informed consent was obtained from all the families according to protocols approved by local IRBs and human research ethics committees. Specific written consent was also retained for all individuals whose photographs appear in the manuscript.

**Data availability.** RNA-Seq data generated from flies and hPSCs are available in the NCBI's Gene Expression Omnibus (GEO) database (GEO GSE246137). Supporting data for all values underlying the data presented in the graphs are provided in the Supplemental Supporting Data Values file.

## Author contributions

DL, YS, and HH conceptualized the study. DL, AB, DGMB, GVH, LG, AKP, LAP, GA, MH, SB, MS, RKP, ZS, EHZ, BAK, EJB, MHH, GL, KMB, M Tarnopolsky, ML, M Towne, LB, MJP, HF, LKP, EA, MLD, AN, MW, RP, KL, MS, PS, MI, FZ, AC, TJM, MB, RRL, TBB, RL, ML, JD, CN, AA, JH, BK, SMM, MMM, JAM, AMR, EMM, MJF, SMR, IH, RSM, MG, BDG, VJ, AW, JG, SJ, KJ, AB, CQ, VKM, EC, CR, DG, AA, MM, VN, RS, YU, NM, RT, KT, YS, GY, BW, MW, BK, ACEH, MLT, RS, LR, RCS, VS, EJH, DC, AC, GC, TK, BF, FTM, LFGM, SG, MO, LC, NQ, DC, CL, AS, MK, MN, BOJ, PS, KS, JA, VP, SM, MV, CP, CF, KF, RKH, SB, JR, CS, MC, CB, JG, KL, AS, CM, RP, BBADV, MAVS, ASB, BC, TR, ZT, and HH conducted molecular and clinical evaluation of the affected individuals. DL, QW, MRB, YZ, AA, MEM, CH, LT, FT, MEP, NA, and YS performed functional assessment and validation. DL and NO were responsible for project administration. TH and PMK conducted the GestaltMatcher analysis. DL, QW, and YS wrote the original draft of the manuscript. All authors reviewed and edited the manuscript. DL, YS, and HH supervised the study. The order of co-first authors was determined on the basis of their contributions to the study and drafting of the manuscript, and a consensus agreement was reached to ensure mutual understanding and commitment.

## Acknowledgments

We thank the patients and their families for participating in this research. We acknowledge Clara L. Kielkopf (University of Rochester, Rochester, New York, USA) for providing the *ppyPY* minigene construct. This study was supported in part by a CHOP Roberts Collaborative Functional Genomics Rapid grant (to DL, MB, and YS); Institutional Development Funds (to HH); an Eagles Autism Foundation grant (to DL); NIH grant R01NS107392 (to YS); and a Pennsylvania Department of Health grant 4100088540 (to YS). Work for U2AF2 individual 30 was supported by the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director under Award U01HG007672 (to Duke University). Work for U2AF2 individual 27 was conducted as part of the Clinical Sequencing Evidence-Generating Research (CSER) consortium, which is funded by the National Human Genome

Research Institute (NHGRI) (U01HG007301) with co-funding from the National Institute on Minority Health and Health Disparities (NIMHD) and the NCI, NIH. Work for U2AF2 individual 28 was conducted at the University of Alabama at Birmingham and the HudsonAlpha Institute for Biotechnology, and funded by the state of Alabama. Work for U2AF2 individual 45 was supported in part by the Telethon Undiagnosed Diseases Program (TUDP) (GSP15001). Work for PRPF19 individual P2 was supported by NIH U01 award HG009610 (to BDG). Additional support includes grant NU22-07-00165 from the Czech Ministry of Health (to ZS); the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) under Germany's Excellence Strategy (EXC 2067/1-390729940); and the DZHK (German Centre for Cardiovascular Research; partner site Göttingen) (to BW); the Dutch Organization for Health Research and Development ZON-MW grant 912-12-109 (to BBADV); the Novo Nordisk Foundation grant NNF20SA0064340 (to AB); the Japan Agency for Medical Research and Development (AMED) under grant numbers JP22ek0109486, JP22ek0109549, and JP22ek0109493 (to NM); JSPS KAKENHI under grant number JP21K15907 (to YU); and the Takeda Science Foundation (to NM). The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant HICF-1009-003). This study makes use of DECIPHER (<https://www.deciphergenomics.org>), which is funded by Wellcome (grant 223718/Z/21/Z) and hosted by EMBL-EBI. See *Nature* PMID: 25533962 or [www.ddduk.org/access.html](http://www.ddduk.org/access.html) for full acknowledgment. The illustration in the Graphical Abstract was generated in part using Biorender.com.

Address correspondence to: Dong Li, Assistant Professor of Pediatrics, The Perelman School of Medicine University of Pennsylvania, Center for Applied Genomics, The Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Room 10161, Philadelphia, Pennsylvania 19104-4318, USA. Phone: 267.649.1269; Email: Lid2@chop.edu. Or to: Yuanquan Song, Associate Professor of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania 19104-4318, USA. Email: SONGY2@chop.edu

- Vissers LE, et al. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 2016;17(1):9-18.
- Vissers LE, et al. A de novo paradigm for mental retardation. *Nat Genet.* 2010;42(12):1109-1112.
- Ropers HH. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet.* 2010;11:161-187.
- Leblond CS, et al. Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Mol Cell Neurosci.* 2021;113:103623.
- Bryant L, et al. Histone H3.3 beyond cancer: germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Sci Adv.* 2020;6(49):eabc9207.
- Sheppard SE, et al. Expanding the genotypic and phenotypic spectrum in a diverse cohort of 104 individuals with Wiedemann-Steiner syndrome. *Am J Med Genet A.* 2021;185(6):1649-1665.
- Sobering AK, et al. Variants in PHF8 cause a spectrum of X-linked neurodevelopmental disorders and facial dysmorphism. *HGG Adv.* 2022;3(3):100102.
- Li D, et al. Further supporting SMARCC2-related neurodevelopmental disorder through exome analysis and reanalysis in two patients. *Am J Med Genet A.* 2022;188(3):878-882.
- Li D, et al. Pathogenic variants in *SMARCA5*, a chromatin remodeler, cause a range of syndromic neurodevelopmental features. *Sci Adv.* 2021;7(20):eabf2066.
- Drivas TG, et al. A second cohort of CHD3 patients expands the molecular mechanisms known to cause Snijders Blok-Campeau syndrome. *Eur J Hum Genet.* 2020;28(10):1422-1431.
- Li D, et al. The variability of *SMARCA4*-related Coffin-Siris syndrome: Do nonsense candidate variants add to milder phenotypes? *Am J Med Genet A.* 2020;182(9):2058-2067.
- Thiffault I, et al. On the verge of diagnosis: detection, reporting, and investigation of de novo variants in novel genes identified by clinical sequencing. *Hum Mutat.* 2018;39(11):1505-1516.
- Gebauer F, et al. RNA-binding proteins in human genetic disease. *Nat Rev Genet.* 2021;22(3):185-198.
- Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol.* 2014;15(2):108-121.
- Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol.* 2017;18(11):655-670.
- Lalli MA, et al. High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res.* 2020;30(9):1317-1331.
- Marchetto MC, et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell.*

- 2010;143(4):527–539.
18. Van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. *Annu Rev Genet.* 2011;45:81–104.
  19. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet.* 2016;17(1):19–32.
  20. Wang T, et al. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat Commun.* 2020;11(1):4932.
  21. Kalscheuer VM, et al. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet.* 2003;35(4):313–315.
  22. Lee YR, et al. Mutations in FAM50A suggest that Armfield XLID syndrome is a spliceosomopathy. *Nat Commun.* 2020;11(1):3698.
  23. Von Elsner L, et al. Biallelic FRA10AC1 variants cause a neurodevelopmental disorder with growth retardation. *Brain.* 2022;145(4):1551–1563.
  24. Modafferi EF, Black DL. A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. *Mol Cell Biol.* 1997;17(11):6537–6545.
  25. Jin Y, et al. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* 2003;22(4):905–912.
  26. Sobreira N, et al. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat.* 2015;36(10):928–930.
  27. Philippakis AA, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat.* 2015;36(10):915–921.
  28. Bowling KM, et al. Genome sequencing as a first-line diagnostic test for hospitalized infants. *Genet Med.* 2022;24(4):851–861.
  29. Hsieh TC, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. *Nat Genet.* 2022;54(3):349–357.
  30. Zamore PD, Green MR. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A.* 1989;86(23):9243–9247.
  31. Singh R, et al. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science.* 1995;268(5214):1173–1176.
  32. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature.* 2017;542(7642):433–438.
  33. Miller JA, et al. Transcriptional landscape of the prenatal human brain. *Nature.* 2014;508(7495):199–206.
  34. Schlegelmilch K, et al. Yap1 acts downstream of  $\alpha$ -catenin to control epidermal proliferation. *Cell.* 2011;144(5):782–795.
  35. King I, et al. Drosophila tao controls mushroom body development and ethanol-stimulated behavior through par-1. *J Neurosci.* 2011;31(3):1139–1148.
  36. Simon AF, et al. A simple assay to study social behavior in Drosophila: measurement of social space within a group. *Genes Brain Behav.* 2012;11(2):243–252.
  37. Parkes TL, et al. Extension of Drosophila lifespan by overexpression of human SOD1 in motor neurons. *Nat Genet.* 1998;19(2):171–174.
  38. Han C, et al. Enhancer-driven membrane markers for analysis of nonautonomous mechanisms reveal neuron-glia interactions in Drosophila. *Proc Natl Acad Sci U S A.* 2011;108(23):9673–9678.
  39. Zhang Y, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron.* 2013;78(5):785–798.
  40. Cvitkovic I, Jurica MS. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.* 2013;41(database issue):D132–D141.
  41. Tarn WY, et al. Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. *Proc Natl Acad Sci U S A.* 1993;90(22):10821–10825.
  42. Tarn WY, et al. Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J.* 1994;13(10):2421–2431.
  43. Chan SP, et al. The Prp19p-associated complex in spliceosome activation. *Science.* 2003;302(5643):279–282.
  44. Ohi MD, Gould KL. Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA.* 2002;8(6):798–815.
  45. Chanarat S, Strasser K. Splicing and beyond: the many faces of the Prp19 complex. *Biochim Biophys Acta.* 2013;1833(10):2126–2134.
  46. Rasche N, et al. Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J.* 2012;31(6):1591–1604.
  47. De Moura TR, et al. Prp19/Pso4 is an autoinhibited ubiquitin ligase activated by stepwise assembly of three splicing factors. *Mol Cell.* 2018;69(6):979–992.
  48. David CJ, et al. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* 2011;25(9):972–983.
  49. Venselaar H, et al. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics.* 2010;11:548.
  50. Anders S, et al. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012;22(10):2008–2017.
  51. Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
  52. Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* 2016;44(15):e127.
  53. Hogg R, et al. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans.* 2010;38(4):1110–1115.
  54. Underwood JG, et al. Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol.* 2005;25(22):10005–10016.
  55. Tomassoni-Ardori F, et al. Rbfox1 up-regulation impairs BDNF-dependent hippocampal LTP by dysregulating TrkB isoform expression levels. *Elife.* 2019;8:e49673.
  56. El Chehadeh S, et al. Dominant variants in the splicing factor PUF60 cause a recognizable syndrome with intellectual disability, heart defects and short stature. *Eur J Hum Genet.* 2016;25(1):43–51.
  57. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64–69.
  58. Kaplanis J, et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature.* 2020;586(7831):757–762.
  59. Hiraide T, et al. Global developmental delay, systemic dysmorphism and epilepsy in a patient with a de novo U2AF2 variant. *J Hum Genet.* 2021;66(12):1185–1187.
  60. Kittock CM, et al. U2AF2 variant in a patient with developmental delay, dysmorphic features, and epilepsy. *Am J Med Genet A.* 2023;191(7):1968–1972.
  61. Wang X, et al. A presumed missense variant in the U2AF2 gene causes exon skipping in neurodevelopmental diseases. *J Hum Genet.* 2023;68(6):375–382.
  62. Neumuller RA, et al. Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell.* 2011;8(5):580–593.
  63. Fortschegger K, et al. Early embryonic lethality of mice lacking the essential protein SNEV. *Mol Cell Biol.* 2007;27(8):3123–3130.
  64. Gong NN, et al. The chromatin remodeler ISWI acts during Drosophila development to regulate adult sleep. *Sci Adv.* 2021;7(8):eabe2597.
  65. Corthals K, et al. Neuroligins Nlg2 and Nlg4 affect social behavior in *Drosophila melanogaster*. *Front Psychiatry.* 2017;8:113.
  66. Merkin J, et al. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science.* 2012;338(6114):1593–1599.
  67. Li Q, et al. Neuronal regulation of alternative pre-mRNA splicing. *Nat Rev Neurosci.* 2007;8(11):819–831.
  68. Raj B, Blencowe BJ. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron.* 2015;87(1):14–27.
  69. Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell.* 1998;92(3):315–326.