The American Journal of Human Genetics, Volume *111*

Supplemental information

The clinical utility and diagnostic implementation of human subject cell transdifferentiation followed by RNA sequencing

Shenglan Li, Sen Zhao, Jefferson C. Sinson, Aleksandar Bajic, Jill A. Rosenfeld, Matthew B. Neeley, Mezthly Pena, Kim C. Worley, Lindsay C. Burrage, Monika Weisz-Hubshman, Shamika Ketkar, William J. Craigen, Gary D. Clark, Seema Lalani, Carlos A. Bacino, Keren Machol, Hsiao-Tuan Chao (趙孝端), Lorraine Potocki, Lisa Emrick, Jennifer Sheppard, My T.T. Nguyen, Anahita Khoramnia, Paula Patricia Hernandez, Sandesh CS. Nagamani, Zhandong Liu, Undiagnosed Diseases Network, Christine M. Eng, Brendan Lee, and Pengfei Liu

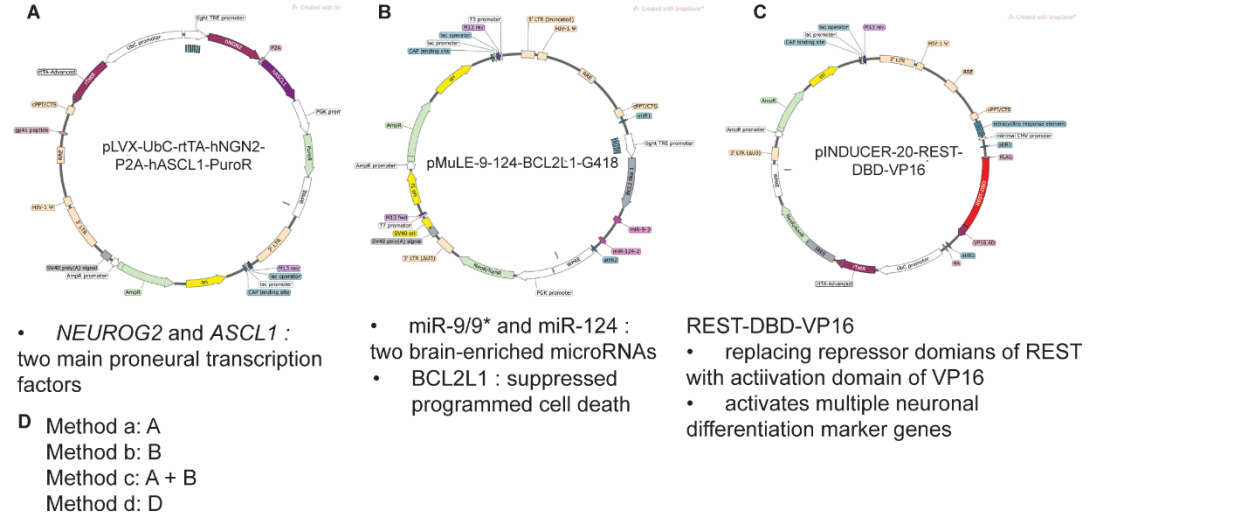

Figure S1. Four transdifferentiation methods used in this study. (A-C) Lentivirus vectors used in four transdifferentiation methods, including: **(A)** vector overexpressing the pro-neuronal transcription factors *NEUROG2* and *ASCL1*, **(B)** vector co-expressing the microRNA *miR-9/9**, *miR-124*, and the anti-apoptotic gene *BCL2L1*, and **(C)** vector DBD-REST-VP16, involving the replacement of *REST*/*NRSF* repressor domains with the activation domain of the viral activator VP16. **(D)** Four methods used for transdifferentiation.

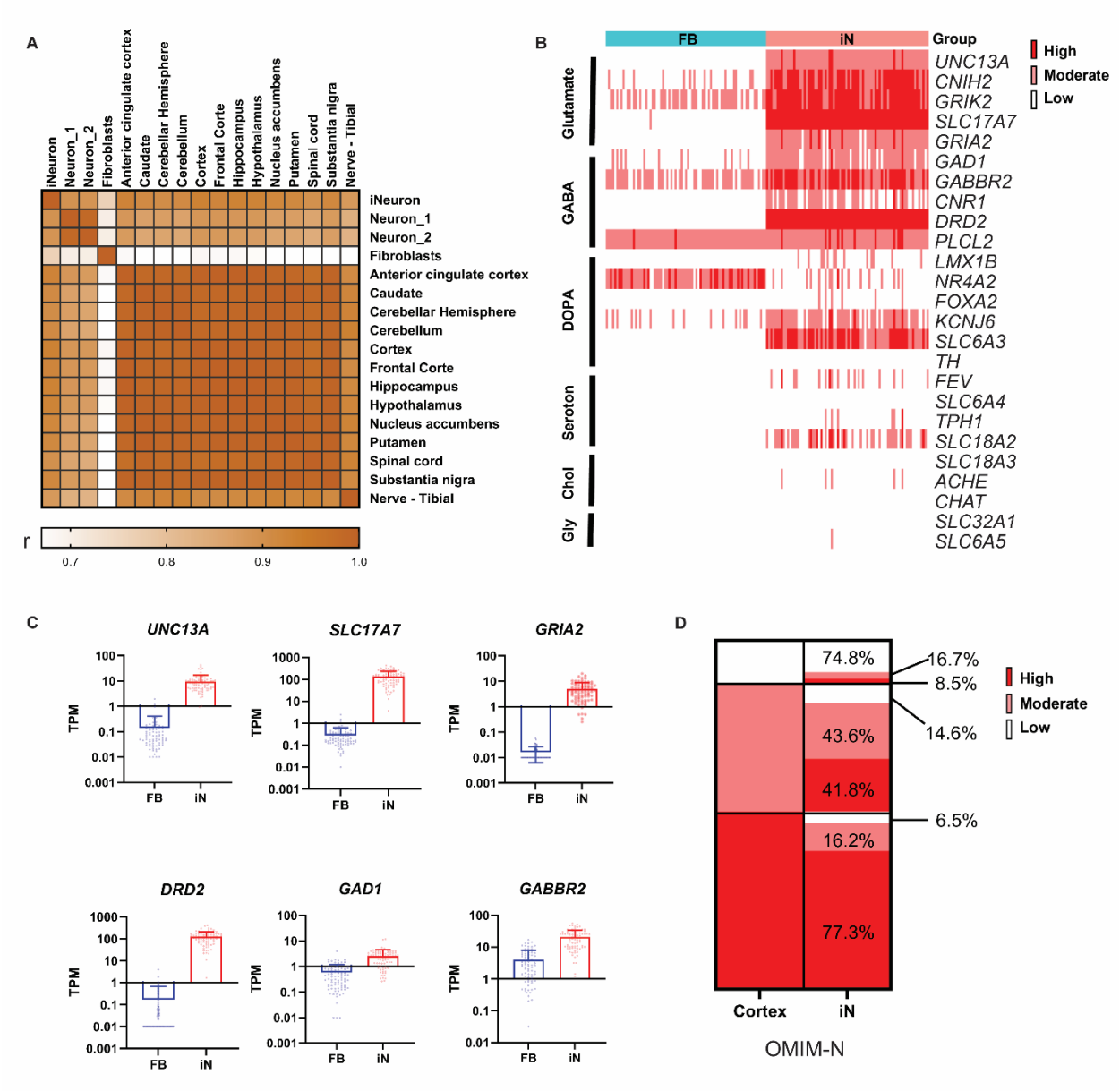

Figure S2. Transcriptomic characterizations of iNeurons. (A) Correlation heatmap comparing iNeurons with two iterations of iPSCs-derived neurons and a spectrum of neural tissues. Neuron_1 represents neurons differentiated from iPSCs, while Neuron_2 refers to neurons derived via overexpression of the neuronal transcription factor Neurog2 in iPSCs. **(B)** Heatmap showing the expression profiles of neuron subtype-specific genes across fibroblasts (FB) and iNeurons (iN). Genes related to glutamatergic neuron include Unc-13 homolog A (*UNC13A*),

AMPA receptor auxiliary protein 2 (*CNIH2*), GluK2 (*GRIK2*), glutamate ionotropic receptor AMPA type subunit 2 (*GRIA2*) and vGLUT1 (*SLC17A7*); GABAergic neuron markers include glutamate decarboxylase 1 (*GAD1*), gamma-aminobutyric acid type B receptor subunit 2 (*GABBR2*), cannabinoid receptor 1 (*CNR1*), dopamine receptor D2 (*DRD2*), and phospholipase C like 2 (*PLCL2*); dopaminergic neuron markers are represented by LIM homeobox transcription factor 1 beta (*LMX1B*), nuclear receptor subfamily 4 group A member 2 (*NR4A2*), potassium voltage-gated channel subfamily J member 6 (*KCNJ6*), DAT1 (*SLC6A3*) and Tyrosine hydroxylase (*TH*); serotonergic neuron markers encompass ETS transcription factor (*FEV*), serotonin transporter 1 (*SLC6A4*), tryptophan hydroxylase (*TPH1*), and VMAT2 (*SLC18A2*); cholinergic neuron markers comprise vesicular acetylcholine transporter (*SLC18A3*), Acetylcholinesterase (*ACHE*) and choline O-acetyltransferase (*CHAT*); glycinergic neuron-specific genes: VGAT (*SLC32A1*) and GlyT-2 (*SLC6A5*). **(C)** Boxplots delineate the TPM expression levels of selecte neuron subtype-specific genes, namely *UNC13A*, *SLC17A* and *GRIA2* for glutamatergic; *DRD2*, *GAD1* and *GABBR2* for GABAergic. The data is presented as mean ± SD. **(D)** Comparative expression analysis of OMIM-N genes in the brain cortex (GTEx) and iNeurons, highlighting over 70% overlap for high- and low- expression genes. For moderate-expression genes, 43.6% are concordant and 41.8% are expressed higher in iNeurons. FB: fibroblasts, n=77; iN: iNeurons, n=91.

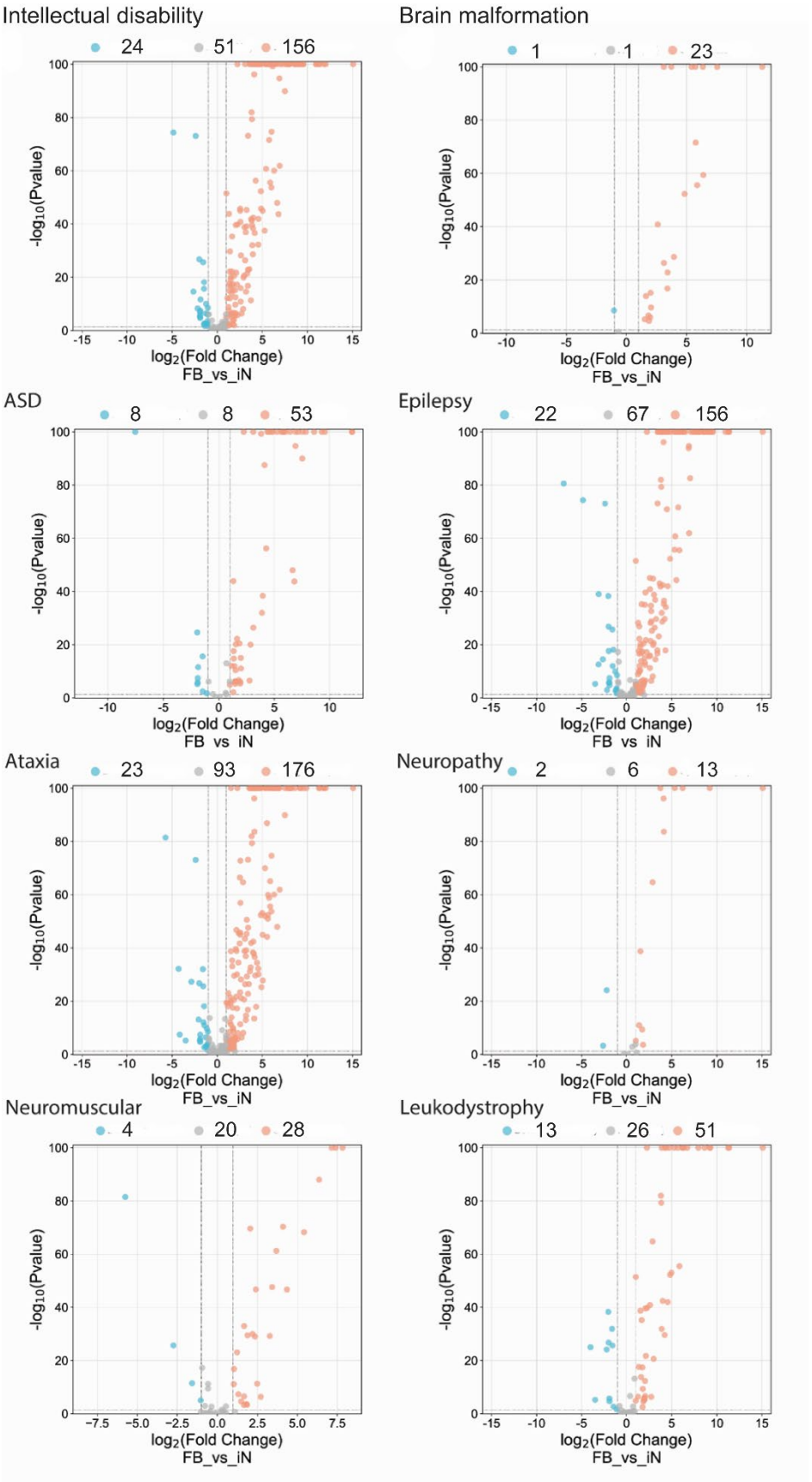
Intellectual disability
24 51 156
Brain malformation
1 1 23
ASD
8 8 53
Epilepsy
22 67 156
Ataxia
23 93 176
Neuropathy
2 6 13
Neuromuscular
4 20 28
Leukodystrophy
13 26 51
-log10(Pvalue)
log2(Fold Change)
FB_vs_iN
FB vs iN

Figure S3. Activation of low-expression genes in eight panels of genes that cause various neurological phenotypes. The panels include intellectual disability, brain malformation (BM), autism spectrum disorder (ASD), epilepsy, ataxia, neuropathy, neuromuscular disorder, and leukodystrophy.

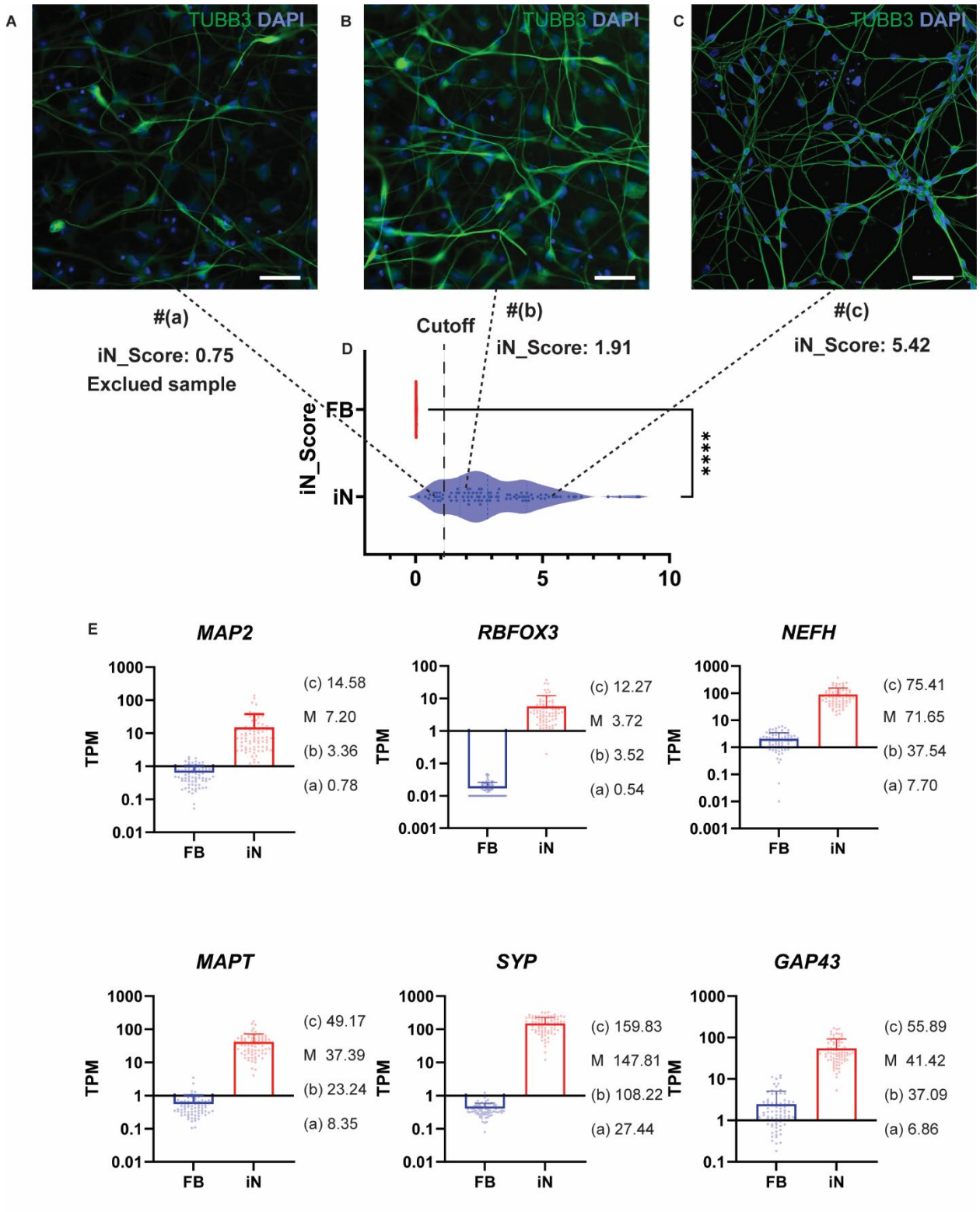

Figure S4. iN_Score and transdifferentiation efficiency in three representative iNeuron samples. (A-C) Immunostaining of Tubulin β-III (TUBB3) in iNeuron samples from individual

#(a), #(b), and #(c). **(D)** iN-Score of iNeuron samples from individual #(a), #(b), and #(c). **(E)** The expression levels of mature neuron genes microtubule associated protein 2 (*MAP2*), NeuN (*RBFOX3*), neurofilament heavy chain (*NEFH*), Tau (*MAPT*), Synaptophysin (*SYP*) and growth associated protein 43 (*GAP43*) in fibroblasts (n=77) and iNeurons (n=82). The gene expression in individual #(a), #(b), #(c) is listed on the right. M indicates the average expression of all iNeuron samples. The data is presented as mean ± SD.

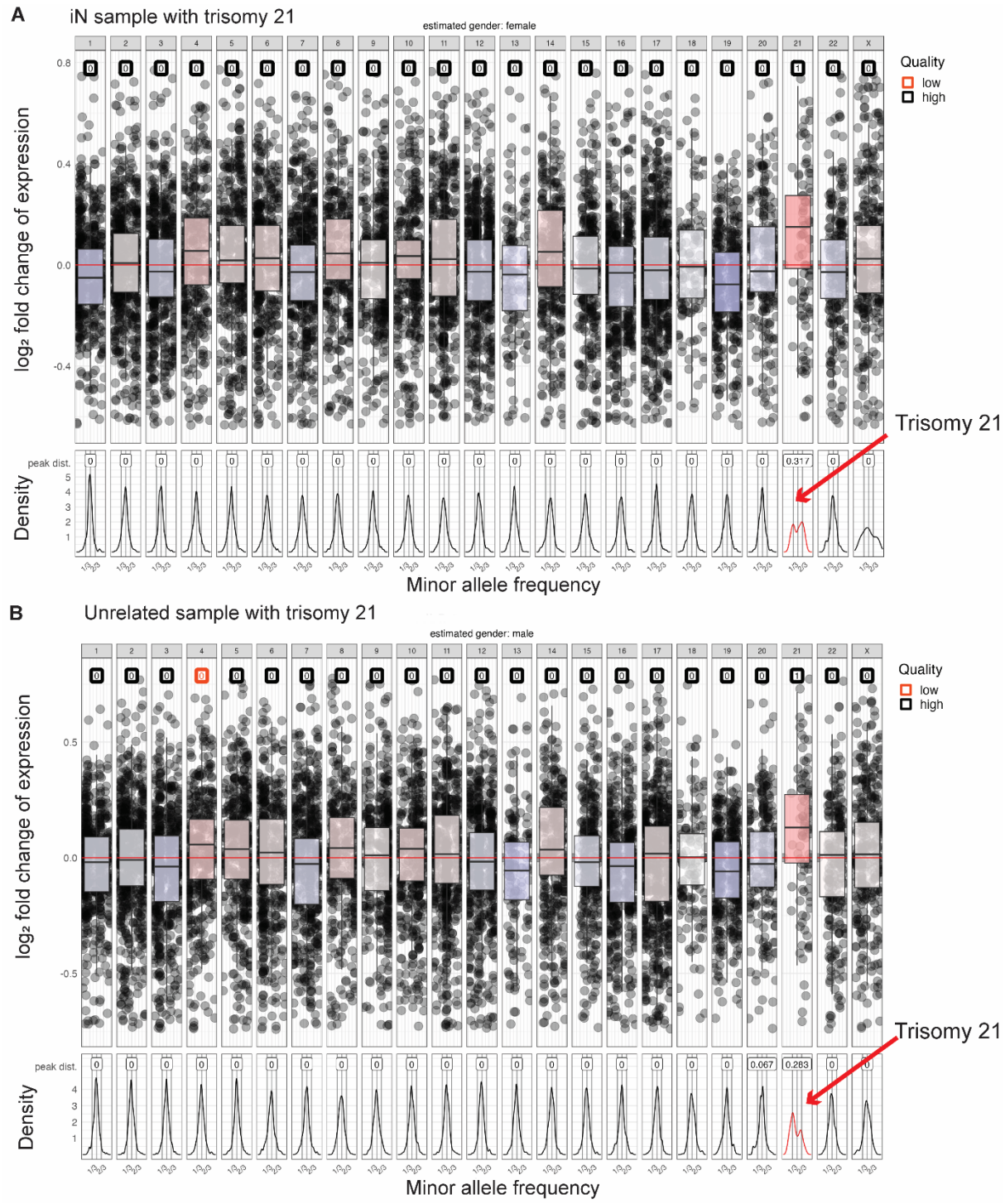

Figure S5. Detection of chromosomal aneuploidies using RNA-seq-based computational analysis. Among all the iNeuron samples included in this study, RNA-seq-based CNV analysis identified only one abnormality. This is illustrated as a gain of chromosome 21 in **(A).**

Diagnostic interpretation was performed with caution in this sample, and no molecular diagnosis was revealed. It is unclear whether the abnormality of chromosome 21 arose during or prior to the transdifferentiation. **(B)** denotes CNV plots from a positive control sample from an individual outside of this study with clinically confirmed trisomy 21. The experimental and bioinformatic RNA-seq analysis follow the same procedures as those described in this study.

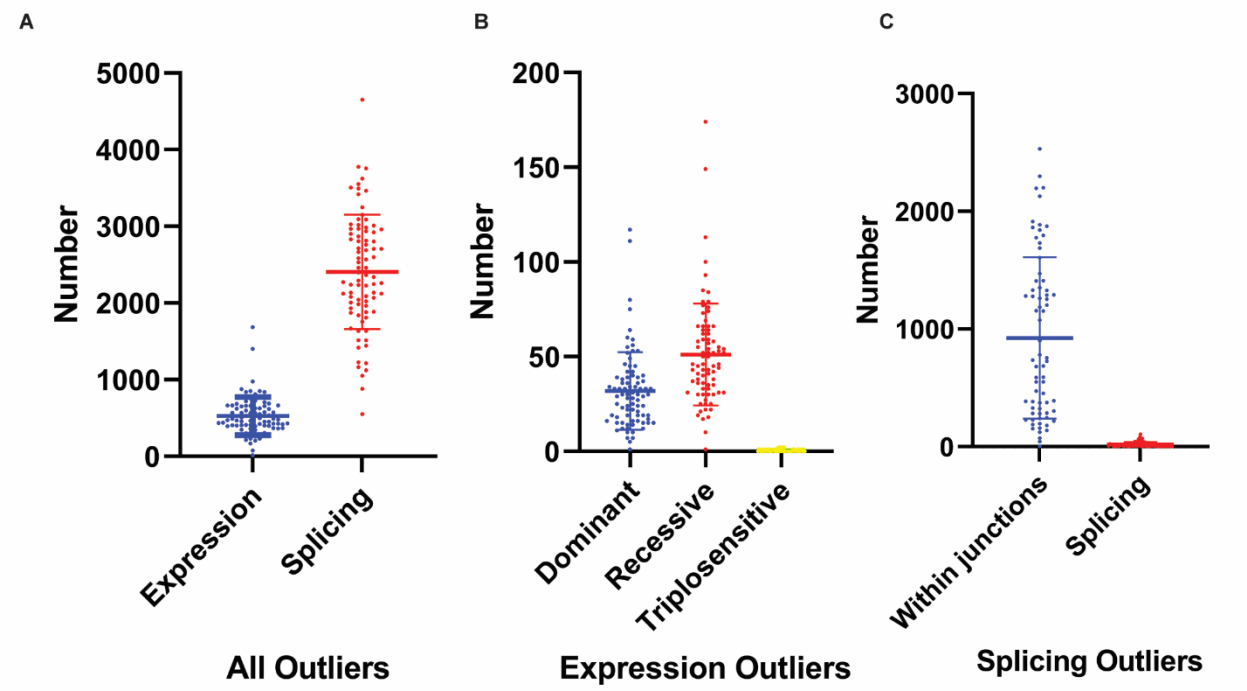

Figure S6. Expression and splicing outliers through filtration at each step. (A) Number of potential outliers identified in the initial step using our analytical workflow. **(B)** Number of expression outliers remaining after restricting to known disease-associated genes in ClinGen and OMIM. **(C)** Number of splicing outliers remaining after filtering DNA variant pairs and SpliceAI prediction. Data is presented as mean ± SD. Each dot represents an individual.

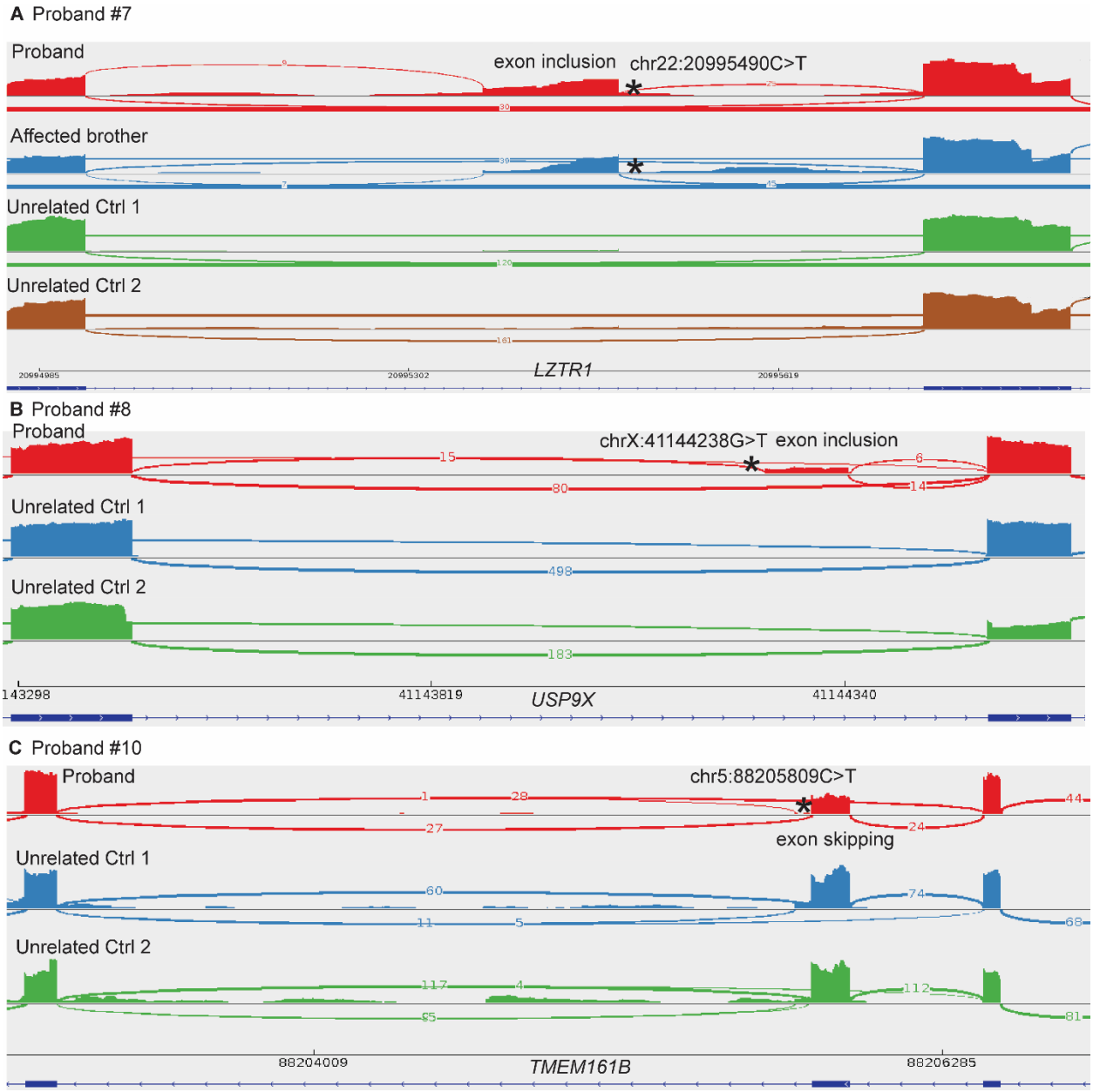

Figure S7. Detection and characterization of aberrant splicing events in *LZTR1* (Proband #7), *USP9X* (Proband #8), and *TMEM161B* (Proband #10). (A) The presence of an intronic variant in *LZTR1* (NM_006767.4: c.1943-256C>T) leads to 117nt and 99nt cryptic exon inclusion (r.1942_1943ins[1942+342_1943-262/1942+360_1943-262]) in both the proband and his affected brother. **(B)** A *de novo* variant in *USP9X* (NM_001039591.3: c.1315-284G>T) causes 106nt and 110nt cryptic exon inclusion (r.1314_1315ins[1315-281_1315-176/1315-

281_1315-172]). The variant also results in aberrant expression (fold change 0.36). **(C)** A variant in *TMEM161B* (NM_153354.5: c.800+5G>A) leads to skipping of exon 8 (141nt) (r.660_800del). Variants are indicated by asterisk.

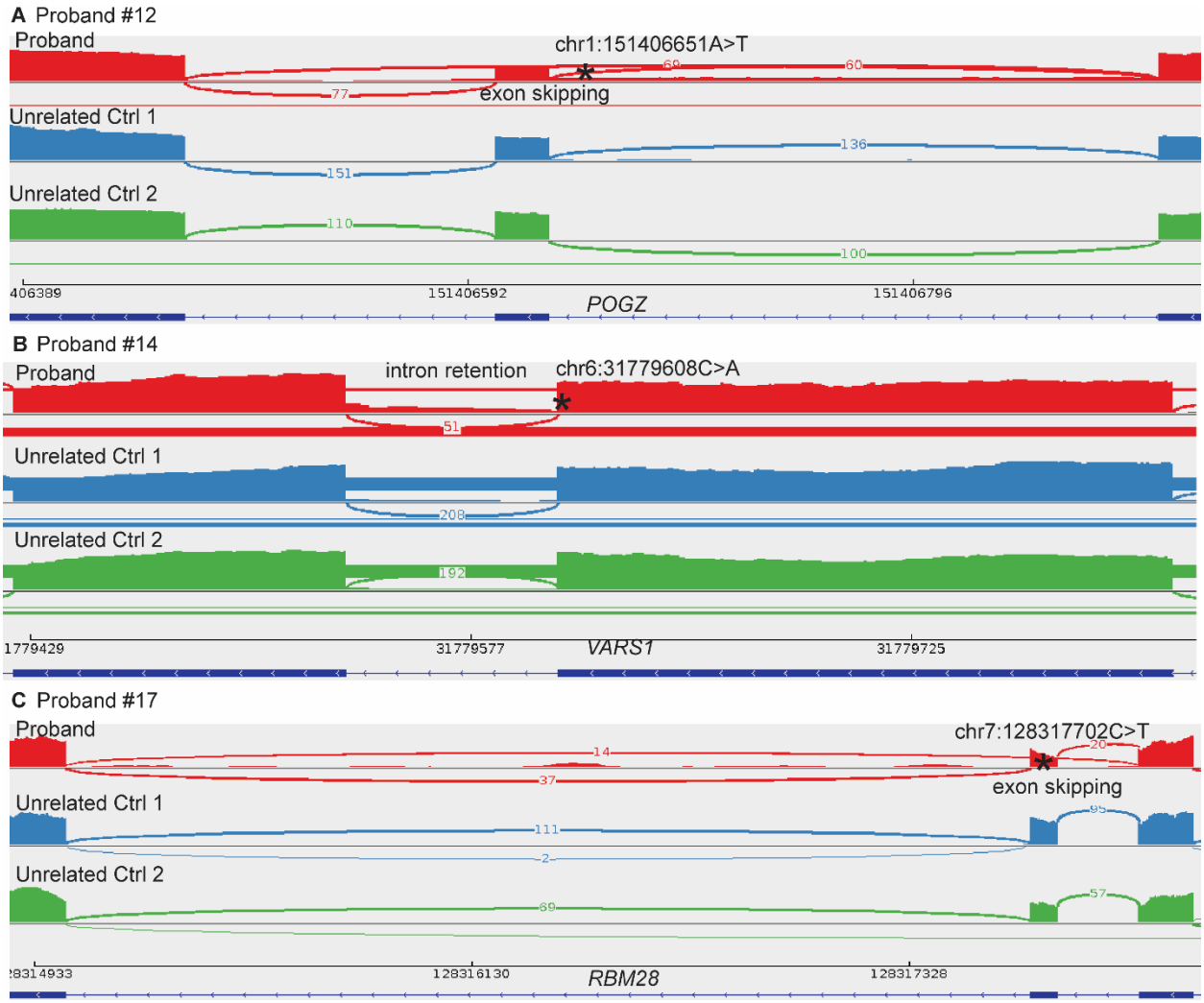

Figure S8. Identification and characterization of aberrant splicing events in *POGZ* (Proband #12), *VARS1* (Proband #14), and *RBM28* (Proband #17). (A) A *de novo* variant in *POGZ* (NM_015100.4: c.2546-20T>A) causes skipping of exon 18 (25nt) and retention of introns 17-18 (446nt) (r.2546_2570del/2545_2571ins[2545+1_2546-21,a,2546-19_2571-1]). **(B)** A variant in *VARS1* (NM_006295.3: c.3288G>T) leads to entire intron 27 (71nt) retention (r.3288delins[u,3288+1_3289-1]). **(C)** A variant in *RBM28* (NM_018077.3: c.1745G>A) results in skipping of exon 16 (75nt, r.1714_1788del). Variants are indicated by asterisk.

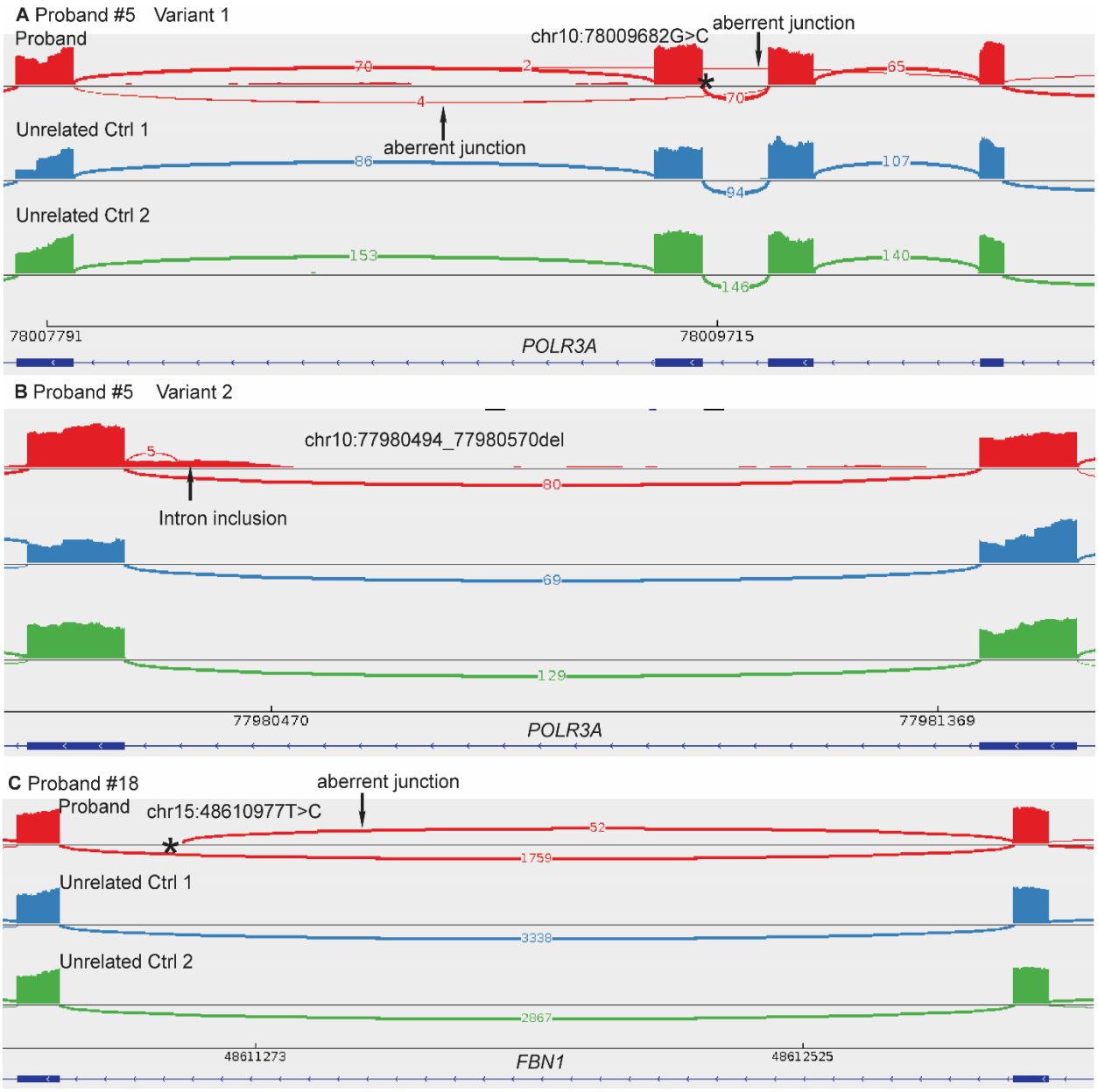

Figure S9. Detection and characterization of splicing events in two probands, #5 and #18. These events were identified manually and not flagged by the DROP pipeline, likely due to the limited presence of aberrant junctions caused by nonsense-mediated decay (NMD). **(A)** The presence of an intronic variant in *POLR3A* (NM_007055.4: c.1771-7C>G) results in skipping of exon 14 (139nt) and exons13-14 (267nt) (r.1771_1909del/1643_1909del). **(B)** Another 76nt intronic deletion in *POLR3A* (NM_007055.4: c.3892-297_3892-221del) causes 227nt intron

inclusion at splice acceptor and 154nt cryptic exon inclusion from intron 29 (r.3891_3892ins[3892-1_3892-227/3892-74_3892-227]). **(C)** An intronic variant in *FBN1* (NM_000138.5: c.248-151A>G) leads to inclusion of cryptic exon (131nt) with premature stop codon likely causing NMD (r.247_248ins[248-282_248-152]). Variants are indicated by asterisk or black box (deletion).


A Proband #6 Variant 1
Proband ES
Proband fibroblast RNA-Seq
chr18:10714931C>T
Only 1 junction read in fibroblast
Proband iNeuron RNA-Seq
aberrent junction
misalignment
PIEZO2
PIEZO2
B Proband #6 Variant 2
misalignment
Proband ES
chr18:10795003C>T
Proband fibroblast RNA-Seq
No junction read in fibroblast
Proband iNeuron RNA-Seq
aberrent junction
PIEZO2
PIEZO2
right exon

Figure S10. Aberrant splicing events caused by misalignment and NMD in Proband #6. This proband has compound heterozygous variants in *PIEZO2* (NM_022068.4: c.5257-1G>A and c.1528-1G>A). The first DNA variant C>T (c.5257-1G>A) functions equivalent to moving the splice site to the left by 1nt (r.5258del), resulting in delC. But because there are 3 Cs, STAR treated the junction shift as an indel **(A)**. The second DNA variant (c.1528-1G>A) results in a CATC sequence, which is misaligned from the right exon to the left. This variant also results in shifting of the splice site by 1nt (r.1528del) **(B)**. Limited junction reads are observed in fibroblast RNA-Seq data.

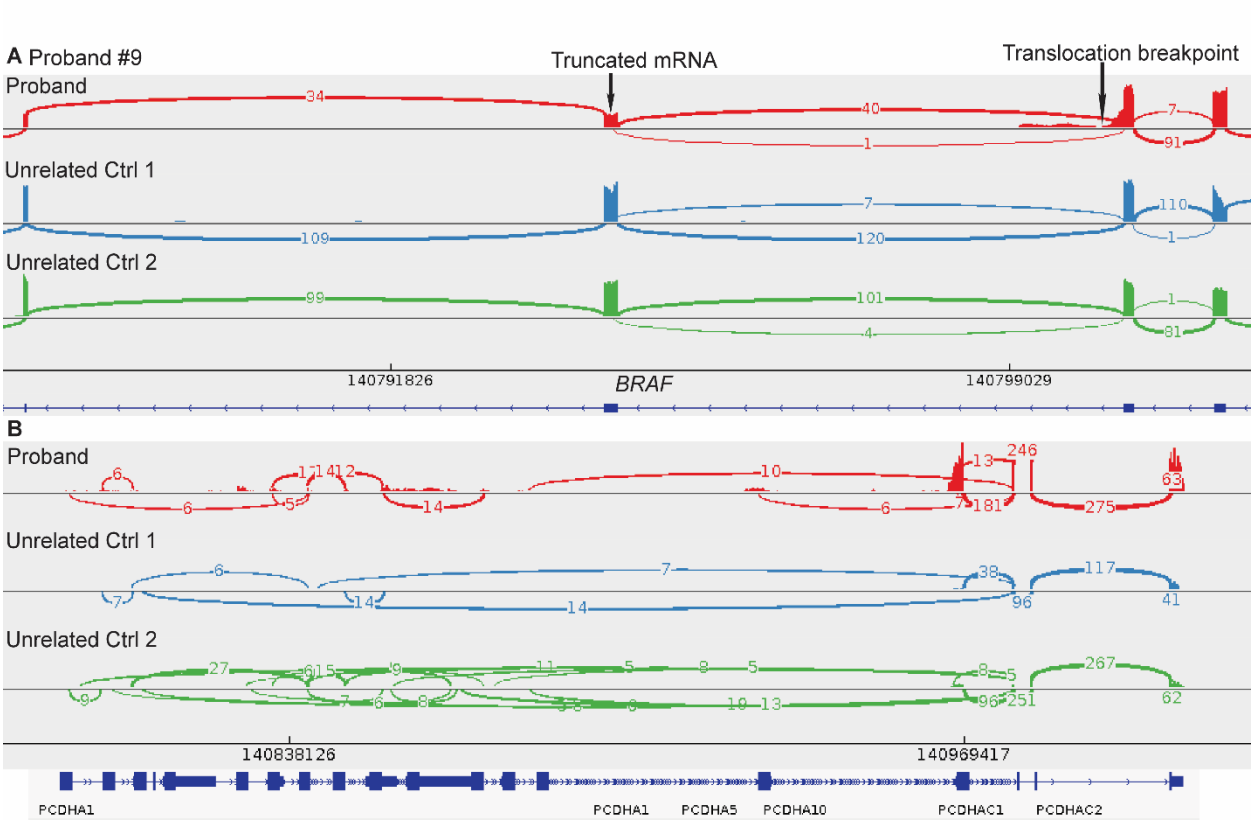

Figure S11. Translocation of t(5;7)(q31.3;q34) in Proband #9. (A) t(5;7)(q31.3;q34) translocation leads to fusion gene of 5' *BRAF* with sequence from 7q34 with uncertain breakpoints; and fusion gene of 5' *BRAF* plus 1225nt intron retention with uncertain sequence from 7q34. **(B)** The translocation also causes the disruption of the protocadherin-alpha gene cluster.

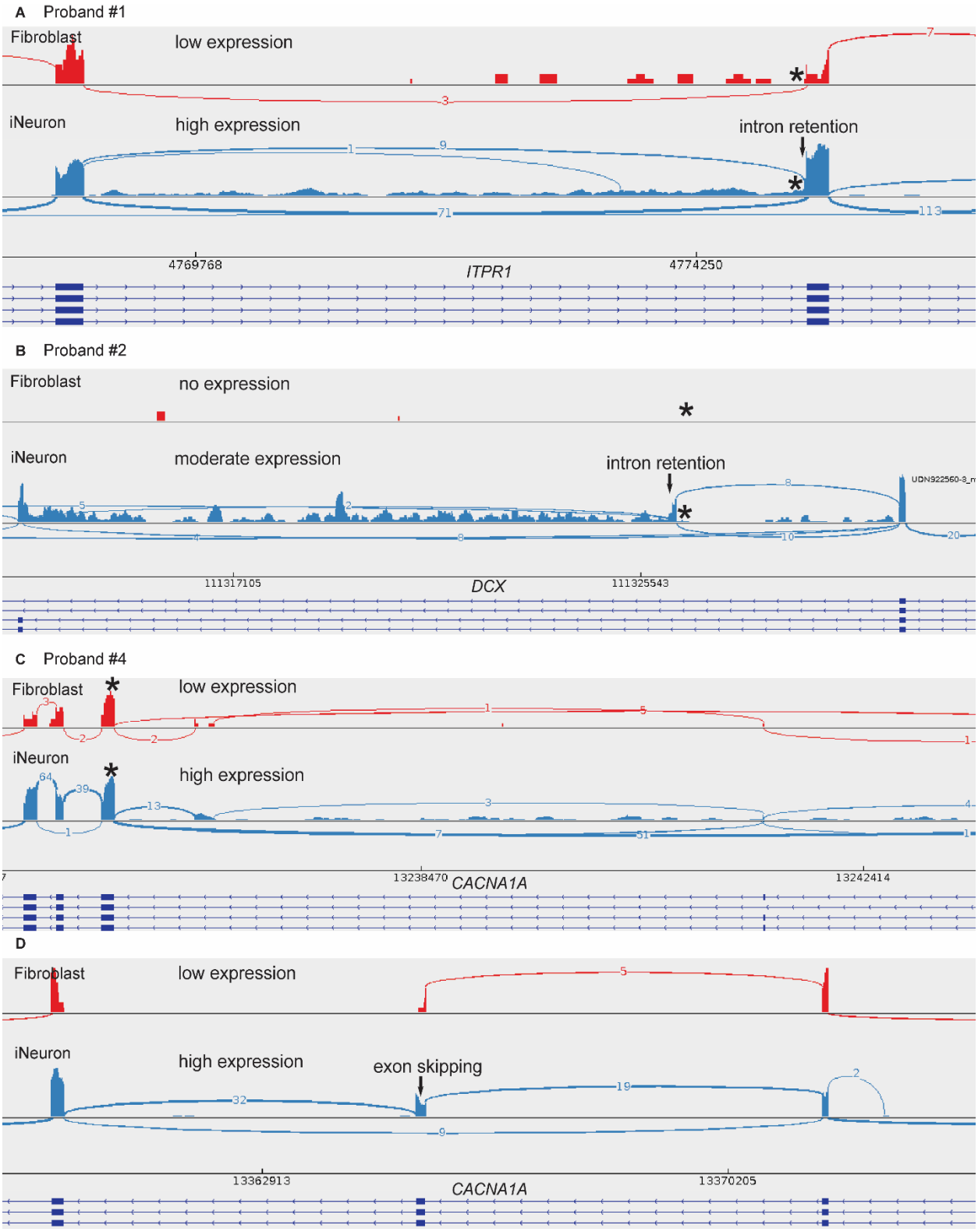

Figure S12. Detection of aberrant events in iNeuron RNA-Seq data from patient #1, #2, and #4. (A) 15-nucleotide retention from intron 45 and complete intron 45 retention are uniquely

observed in iNeuron due to the low expression of *ITPR1* in fibroblast. **(B)** Inclusion of a 13,549 bp cryptic exon from intron is exclusively observed in iNeuron as *DCX* is not expressed in fibroblast. **(C)** Skewed expression of the variant allele and exon skipping are only detected in iNeuron owing to the limited expression of *CACNA1A* in fibroblast.

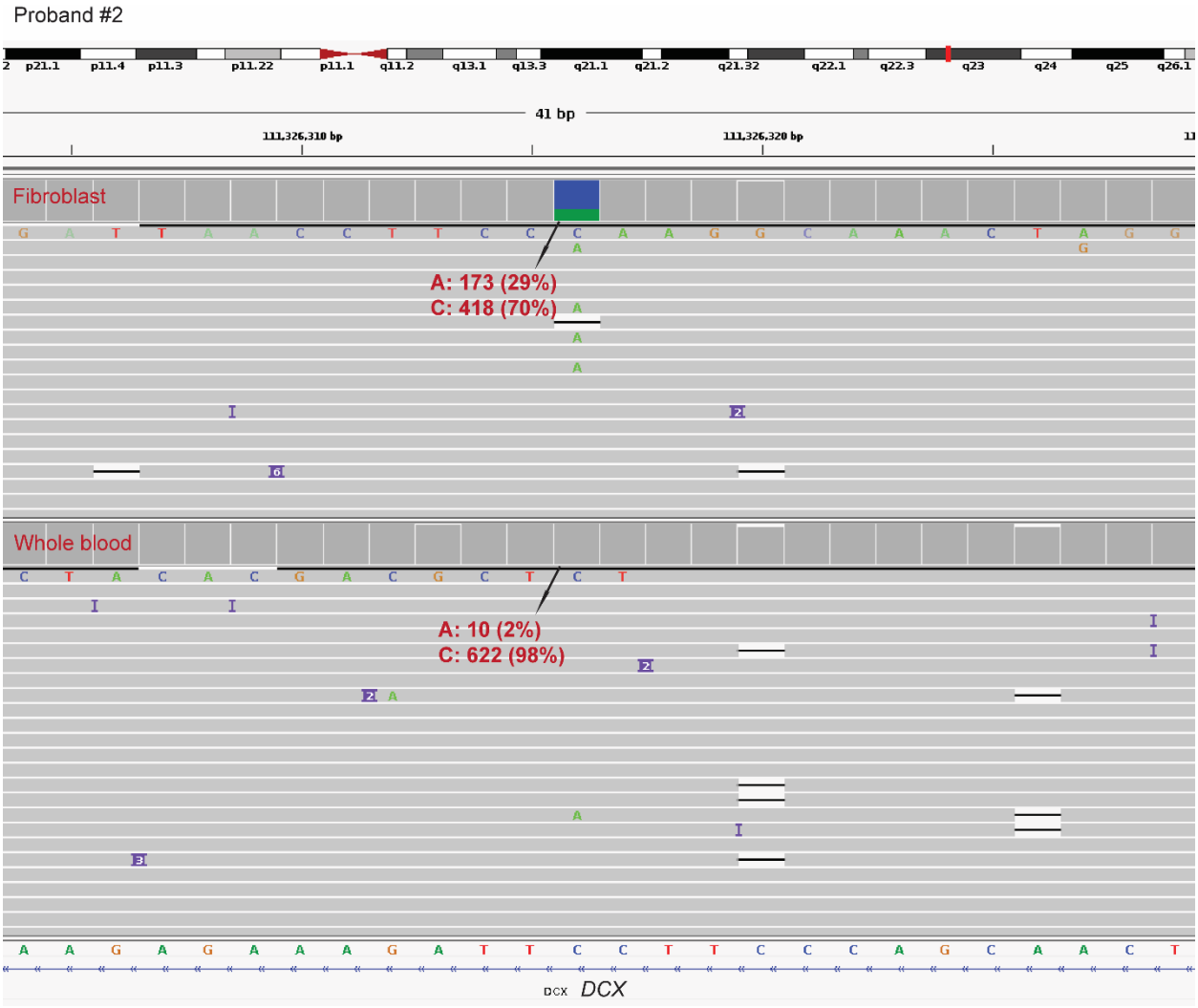

Figure S13. Amplicon-based next-generation sequencing on the fibroblast and whole blood from the mother of Proband #2. result revealed a variant fraction of 29% in fibroblast and 2% in whole blood.

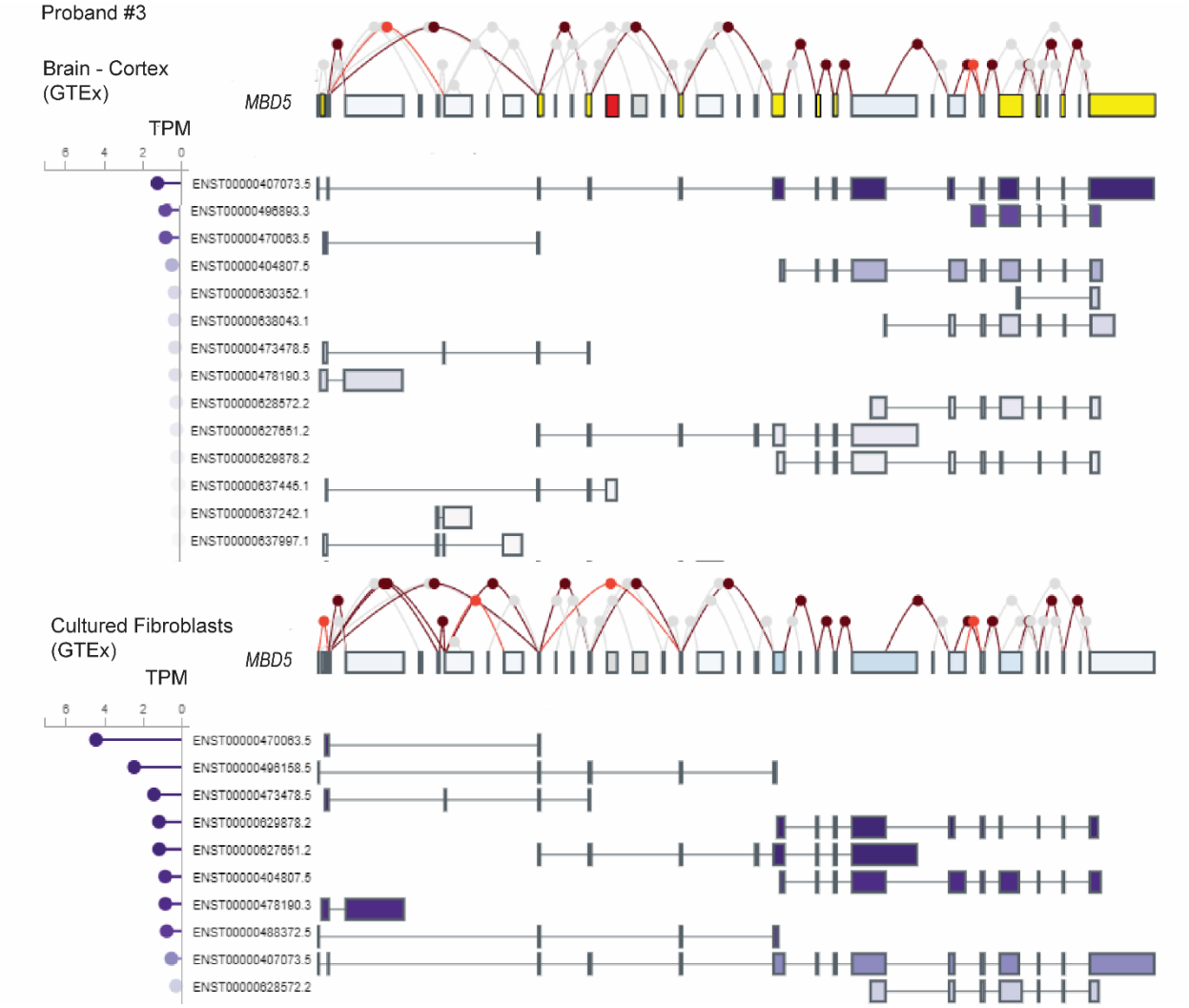

Figure S14. Identification of 12 neuron-enriched exons in *MBD5* genes. Although iNeurons and fibroblasts have similar *MBD5* expression, these exons exhibit higher expression in iNeurons and brain cortex (GTEx consortium) when compared to cultured fibroblasts. Neuron-specific exons are represented by red, while neuron-enriched exons are denoted by yellow.

Table S1. Low-expressed genes associated with neurological disorders in CATs

Panels (Gene Number)	Proportion of Genes with TPM≤1 (%)			
	FB (N=77)	FB (GTEx)	WB (GTEx)	Muscle (GTEx)
Neurological OMIM (2721)	20.7	19.8	34.9	24.3
Intellectual Disability (1379)	16.8	15.3	32.7	21.4
Brain Malformation (163)	14.7	14.1	36.8	22.1
Autism Spectrum Disorder (301)	22.9	21.7	36.0	23.7
Epilepsy (1150)	21.4	19.6	32.6	23.6
Ataxia (1282)	22.8	21.4	36.8	26.1
Neuropathy (105)	20.0	14.7	32.4	21.6
Neuromuscular Disorder (226)	23.1	18.8	32.1	7.1
Leukodystrophy (709)	12.7	10.4	25.5	13.9

Table S2. Activation of low-expressed genes associated with neurological disorders using four transdifferention (21 days' induction) methods

Panels	Proportion of Genes (%)			
	a	b	c	d
Neuronal Markers	89.2	48.2	63.3	81.9
Neurological OMIM	29.6	19.3	19.6	24.8
Intellectual Disability	39.6	26.0	27.2	38.5
Brain Malformation	52.2	27.9	32.4	52.1
Autism Spectrum Disorder	38.8	31.5	25.9	41.5
Epilepsy	37.0	201.1	23.0	33.1
Ataxia	29.8	20.9	18.7	27.9
Neuropathy	43.4	33.5	33.4	38.4
Neuromuscular Disorders	17.8	21.7	9.8	15.9
Leukodystrophy	36.6	21.4	29.7	33.2

a) overexpression of the pro-neuronal transcription factors *NEUROG2* and *ASCL1*; b) co-expression of the microRNA *miR-9/9**, *miR-124*, and the anti-apoptotic gene *BCL2L1*; c) a and b; d) overexpression of DBD-REST-VP16, involving the replacement of *REST*/*NRSF* repressor domains with the activation domain of the viral activator VP16.

Table S3. Activation of low-expressed genes associated with neurological disorders at different induction time point

Panels	Proportion of Genes (%)		
	a-D14	a-D21	a-D28
Neuronal Markers	70.5	89.2	89.2
Neurological OMIM	22.2	29.6	28.7
Intellectual Disability	30.1	39.6	39.1
Brain Malformation	44.3	52.2	48.5
Autism Spectrum Disorder	33.0	38.8	40.1
Epilepsy	27.9	37.0	36.8
Ataxia	23.0	29.8	29.3
Neuropathy	38.4	43.4	47.5
Neuromuscular Disorders	13.8	17.8	15.9
Leukodystrophy	29.2	36.6	34.4

a) overexpression of the pro-neuronal transcription factors *NEUROG2* and *ASCL1*. D14: 14 days after induction; D21: 21 days after induction; D28: 28 days after induction.

Table S4. Activated and actionable genes associated with neurological disorders in iNeurons compared to whole blood

Category	N of low-expression genes in whole blood	Activated and Actionable in iN	
		N	%
Neurological OMIM	909	520	57.2
Intellectual Disability Panel	449	293	65.2
Brain Malformation Panel	60	44	73.3
Autism Spectrum Disorder Panel	108	67	62.3
Epilepsy Panel	373	215	57.6
Ataxia Panel	470	268	57.0
Neuropathy Panel	33	19	57.6
Neuromuscular Disorders Panel	72	35	48.6
Leukodystrophy Panel	178	135	75.8

Table S5. Successful transdifferentiation replication in 3 samples (out of 4) with low *ASCL-NEUROG2* expression.

Sample	Repeat	iN_Score	*ASCL1*	*NEUROG2*
#1	R1	Failed	3.86	3.28
	R2	Failed	366.17	456.43
#2	R1	Failed	53.94	50.06
	R2	Passed	639.59	855.04
#3	R1	Failed	79.95	76.30
	R2	Passed	175.92	216.40
#4	R1	Failed	94.62	95.29
	R2	Passed	781.93	999.31

Supplemental Data Files 1. Gene Ontology (GO) terms and pathways identified in iNeurons through functional enrichment analysis.

Supplemental Data Files 2. Human Phenotype Ontology (HPO) description for individuals with a molecular diagnosis.

Supplemental Data Files 3. Expression of neuron-enriched/specific exons from OMIM-N genes.

Members of the Undiagnosed Diseases Network

Maria T. Acosta

David R. Adams

Raquel L. Alvarez

Justin Alvey

Aimee Allworth

Ashley Andrews

Euan A. Ashley

Ben Afzali

Carlos A. Bacino

Guney Bademci

Ashok Balasubramanyam

Dustin Baldridge

Jim Bale

Michael Bamshad

Deborah Barbouth

Pinar Bayrak-Toydemir

Anita Beck

Alan H. Beggs

Edward Behrens

Gill Bejerano

Hugo J. Bellen

Jimmy Bennett

Jonathan A. Bernstein

Gerard T. Berry

Anna Bican

Stephanie Bivona

Elizabeth Blue

John Bohnsack

Devon Bonner

Lorenzo Botto

Lauren C. Briere

Gabrielle Brown

Elizabeth A. Burke

Lindsay C. Burrage

Manish J. Butte

Peter Byers

William E. Byrd

John Carey

Olveen Carrasquillo

Thomas Cassini

Ta Chen Peter Chang

Sirisak Chanprasert

Hsiao-Tuan Chao

Ivan Chinn

Gary D. Clark

Terra R. Coakley

Laurel A. Cobban

Joy D. Cogan

Matthew Coggins

F. Sessions Cole

Heather A. Colley

Heidi Cope

Rosario Corona

William J. Craigen

Andrew B. Crouse

Michael Cunningham

Precilla D’Souza

Hongzheng Dai

Surendra Dasari

Joie Davis

Jyoti G. Dayal

Margaret Delgado

Esteban C. Dell'Angelica

Katrina Dipple

Daniel Doherty

Naghmeh Dorrani

Argenia L. Doss

Emilie D. Douine

Dawn Earl

David J. Eckstein

Lisa T. Emrick

Christine M. Eng

Marni Falk

Elizabeth L. Fieg

Paul G. Fisher

Brent L. Fogel

Irman Forghani

Jiayu Fu

William A. Gahl

Ian Glass

Page C. Goddard

Rena A. Godfrey

Alana Grajewski

Andrea Gropman

Meghan C. Halley

Rizwan Hamid

Neal Hanchard

Kelly Hassey

Nichole Hayes

Frances High

Anne Hing

Fuki M. Hisama

Ingrid A. Holm

Jason Hom

Martha Horike-Pyne

Alden Huang

Yan Huang

Sarah Hutchison

Wendy Introne

Rosario Isasi

Kosuke Izumi

Gail P. Jarvik

Jeffrey Jarvik

Suman Jayadev

Orpa Jean-Marie

Vaidehi Jobanputra

Emerald Kaitryn

Shamika Ketkar

Dana Kiley

Gonench Kilich

Shilpa N. Kobren

Isaac S. Kohane

Jennefer N. Kohler

Susan Korrick

Deborah Krakow

Donna M. Krasnewich

Elijah Kravets

Seema R. Lalani

Byron Lam

Christina Lam

Brendan C. Lanpher

Ian R. Lanza

Kimberly LeBlanc

Brendan H. Lee

Roy Levitt

Richard A. Lewis

Pengfei Liu

Xue Zhong Liu

Nicola Longo

Sandra K. Loo

Joseph Loscalzo

Richard L. Maas

Ellen F. Macnamara

Calum A. MacRae

Valerie V. Maduro

AudreyStephannie Maghiro

Rachel Mahoney

May Christine V. Malicdan

Laura A. Mamounas

Teri A. Manolio

Rong Mao

Ronit Marom

Gabor Marth

Beth A. Martin

Martin G. Martin

Julian A. Martínez-Agosto

Shruti Marwaha

Jacob McCauley

Allyn McConkie-Rosell

Alexa T. McCray

Elisabeth McGee

Matthew Might

Danny Miller

Ghayda Mirzaa

Eva Morava

Paolo Moretti

Marie Morimoto

John J. Mulvihill

Mariko Nakano-Okuno

Stanley F. Nelson

Shirley Nieves-Rodriguez

Donna Novacic

Devin Oglesbee

James P. Orengo

Laura Pace

Stephen Pak

J. Carl Pallais

Jeanette C. Papp

Neil H. Parker

Leoyklang Petcharet

John A. Phillips III

Jennifer E. Posey

Lorraine Potocki

Barbara N. Pusey Swerdzewski

Aaron Quinlan

Deepak A. Rao

Anna Raper

Wendy Raskind

Genecee Renteria

Chloe M. Reuter

Lynette Rives

Amy K. Robertson

Lance H. Rodan

Jill A. Rosenfeld

Elizabeth Rosenthal

Francis Rossignol

Maura Ruzhnikov

Marla Sabaii

Ralph Sacco

Jacinda B. Sampson

Mario Saporta

Judy Schaechter

Timothy Schedl

Kelly Schoch

Daryl A. Scott

Elaine Seto

Prashant Sharma

Vandana Shashi

Emily Shelkowitz

Sam Sheppeard

Jimann Shin

Edwin K. Silverman

Janet S. Sinsheimer

Kathy Sisco

Edward C. Smith

Kevin S. Smith

Lilianna Solnica-Krezel

Ben Solomon

Rebecca C. Spillmann

Andrew Stergachis

Joan M. Stoler

Kathleen Sullivan

Jennifer A. Sullivan

Shirley Sutton

David A. Sweetser

Virginia Sybert

Holly K. Tabor

Queenie K.-G. Tan

Amelia L. M. Tan

Arjun Tarakad

Mustafa Tekin

Fred Telischi

Willa Thorson

Cynthia J. Tifft

Camilo Toro

Alyssa A. Tran

Rachel A. Ungar

Tiina K. Urv

Adeline Vanderver

Matt Velinder

Dave Viskochil

Tiphanie P. Vogel

Colleen E. Wahl

Melissa Walker

Nicole M. Walley

Jennifer Wambach

Jijun Wan

Lee-kai Wang

Michael F. Wangler

Patricia A. Ward

Daniel Wegner

Monika Weisz Hubshman

Mark Wener

Tara Wenger

Monte Westerfield

Matthew T. Wheeler

Jordan Whitlock

Lynne A. Wolfe

Kim Worley

Shinya Yamamoto

Zhe Zhang

Stephan Zuchner