**Neuron, Volume *110***

## Supplemental information

# Suppression of premature transcription termination leads to reduced mRNA isoform diversity and neurodegeneration

**Geneva R. LaForce, Jordan S. Farr, Jingyi Liu, Cydni Akesson, Evren Gumus, Otis Pinkard, Helen C. Miranda, Katherine Johnson, Thomas J. Sweet, Ping Ji, Ai Lin, Jeff Coller, Polyxeni Philippidou, Eric J. Wagner, and Ashleigh E. Schaffer**

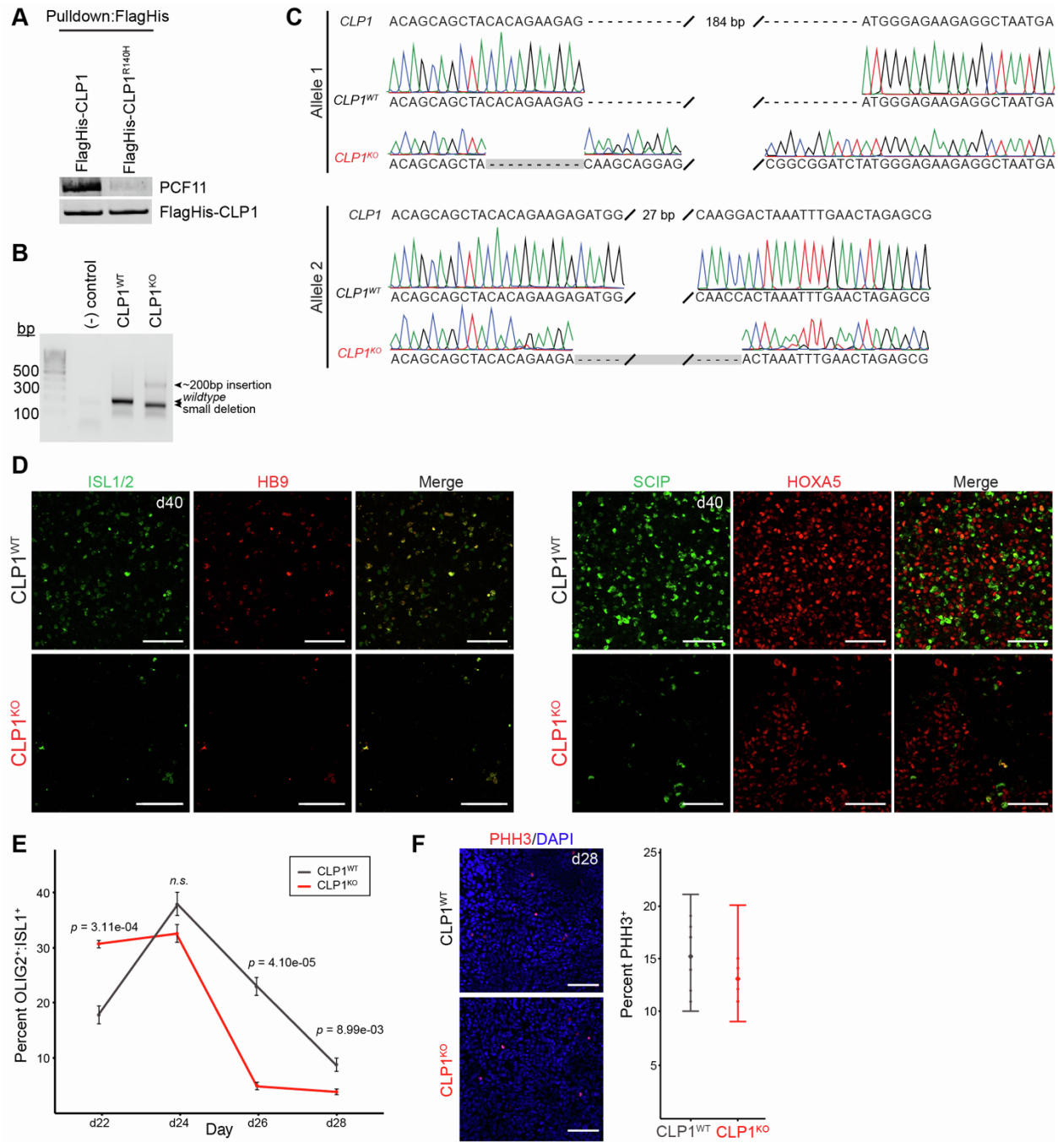
A
Pulldown:FlagHis
FlagHis-CLP1
FlagHis-CLP1R140H
PCF11
FlagHis-CLP1
B
(-) control
CLP1WT
CLP1KO
bp
500
300
100
~200bp insertion
wildtype
small deletion
C
Allele 1
CLP1 ACAGCAGCTACACAGAAGAG 184 bp ATGGGAGAAGAGGCTAATGA
CLP1WT ACAGCAGCTACACAGAAGAG ATGGGAGAAGAGGCTAATGA
CLP1KO ACAGCAGCTA CAAGCAGGAG CGGCGGATCTATGGGAGAAGAGGCTAATGA
Allele 2
CLP1 ACAGCAGCTACACAGAAGAGATGG 27 bp CAAGGACTAAATTTGAACTAGAGCG
CLP1WT ACAGCAGCTACACAGAAGAGATGG CAACCACTAAATTTGAACTAGAGCG
CLP1KO ACAGCAGCTACACAGAAGA ACTAAATTTGAACTAGAGCG
D
ISL1/2
HB9
Merge
SCIP
HOXA5
Merge
d40
CLP1WT
CLP1KO
E
Percent OLIG2+:ISL1+
CLP1WT
CLP1KO
n.s.
p = 3.11e-04
p = 4.10e-05
p = 8.99e-03
d22
d24
d26
d28
Day
F
PHH3/DAPI
d28
CLP1WT
CLP1KO
Percent PHH3+
CLP1WT
CLP1KO

**Figure S1. CLP1 knockout human embryonic stem cells have bi-allelic frameshift variants, and express mature motor neuron markers of phrenic identity upon directed differentiation; related to Figure 1.**

A. CLP1 p.R140H has reduced affinity for PCF11 compared to wild type (Schaffer et al., 2014).
B. PCR amplification of *CLP1* after CRISPR/Cas9 demonstrates the absence of a *wildtype* allele and the presence of a bi-allelic mutation consisting of a ~200-base pair (bp) insertion and a small deletion.
C. Sanger sequencing chromatograms from H9 human embryonic stem cells (hESCs) subjected to CRISPR/Cas9 genome editing to induce bi-allelic indels in exon 2 of *CLP1.* Unedited isogenic controls ($CLP1^{WT}$) have the reference genotype, while CLP1 knockout ($CLP1^{KO}$) hESCs have a compound heterozygous mutation in *CLP1* consisting of a 37-base pair (bp) deletion and a 204-bp insertion.
D. Immunofluorescence staining for motor neuron identity markers ISL1/2, HB9, SCIP and HOXA5 in $CLP1^{KO}$ and isogenic $CLP1^{WT}$ hESC-derived motor neuron cultures at day (d) 35 of differentiation. Scale bar: 50 μm.
E. $CLP1^{KO}$ motor neuron cultures contain a higher percentage of OLIG2 and ISL1/2 co-positive cells at day (d) 22 of differentiation but lower percentages of OLIG2 and ISL1/2 co-positive cells at later time points compared to $CLP1^{WT}$ motor neuron cultures. *n.s.*, non-significant ($p > 0.05$); two-sided Student's t-test. Representative images shown from three clonal hESC-derived motor neuron lines per genotype. Scale bar: 50 μm.
F. Analysis of total proliferation by p-Histone H3 (PHH3) immunostaining in $CLP1^{WT}$ and $CLP1^{KO}$ motor neuron cultures at day (d) 28 showed similar levels of proliferation. two-sided Student's t-test. Representative images shown from three clonal hESC-derived motor neuron lines per genotype. Scale bar: 50 μm.

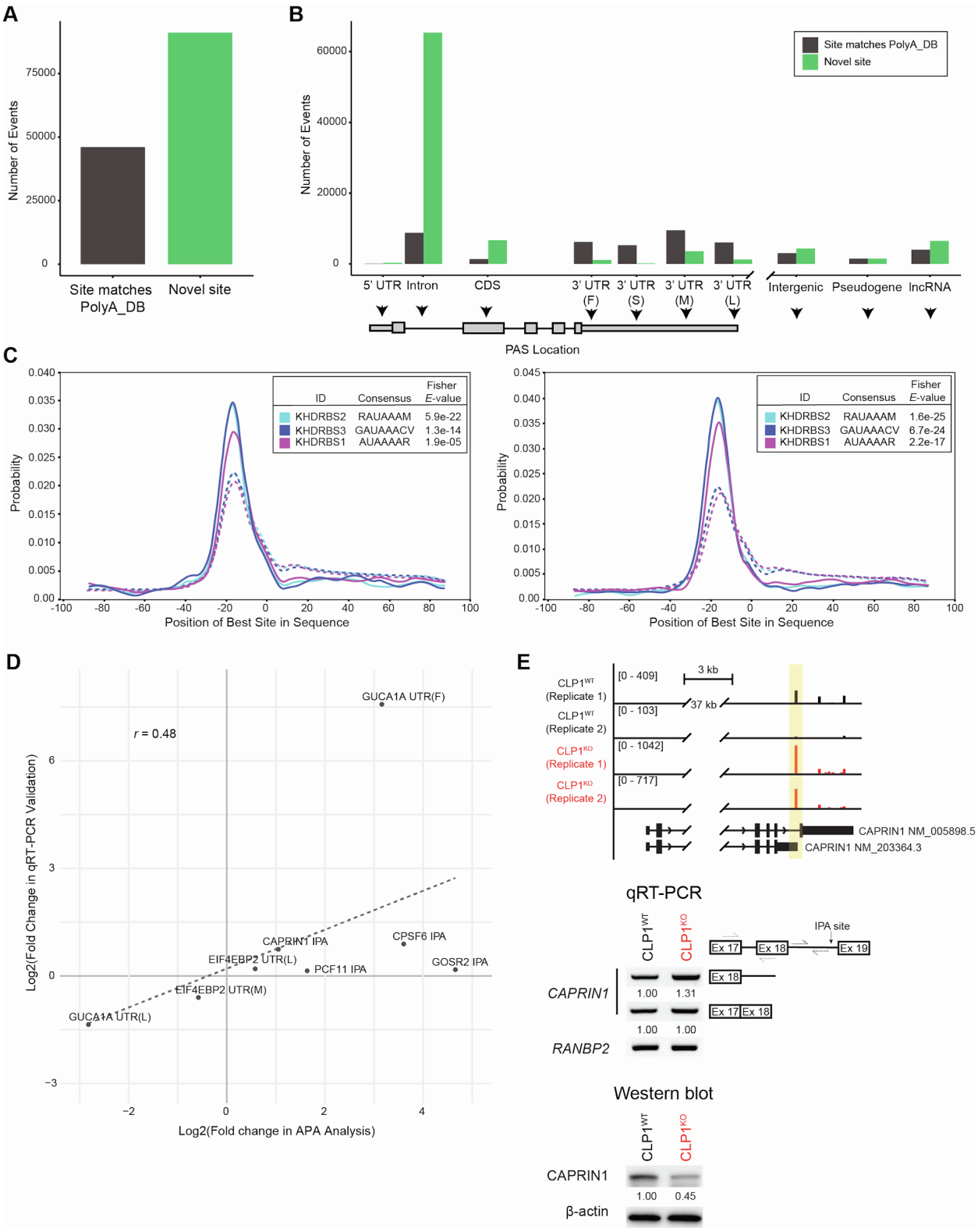
A
Number of Events
75000
50000
25000
0
Site matches PolyA_DB
Novel site
B
Number of Events
60000
40000
20000
0
Site matches PolyA_DB
Novel site
5' UTR
Intron
CDS
3' UTR (F)
3' UTR (S)
3' UTR (M)
3' UTR (L)
Intergenic
Pseudogene
lncRNA
PAS Location
C
Probability
ID Consensus Fisher E-value
KHDRBS2 RAUAAAM 5.9e-22
KHDRBS3 GAUAAACV 1.3e-14
KHDRBS1 AUAAAAR 1.9e-05
Position of Best Site in Sequence
ID Consensus Fisher E-value
KHDRBS2 RAUAAAM 1.6e-25
KHDRBS3 GAUAAACV 6.7e-24
KHDRBS1 AUAAAAR 2.2e-17
Position of Best Site in Sequence
D
r = 0.48
GUCA1A UTR(F)
CAPRIN1 IPA
CPSF6 IPA
EIF4EBP2 UTR(L)
PCF11 IPA
GOSR2 IPA
EIF4EBP2 UTR(M)
GUCA1A UTR(L)
Log2(Fold Change in qRT-PCR Validation)
Log2(Fold change in APA Analysis)
E
CLP1WT (Replicate 1)
CLP1WT (Replicate 2)
CLP1KO (Replicate 1)
CLP1KO (Replicate 2)
[0 - 409]
3 kb
37 kb
[0 - 103]
[0 - 1042]
[0 - 717]
CAPRIN1 NM_005898.5
CAPRIN1 NM_203364.3
qRT-PCR
CLP1WT
CLP1KO
IPA site
Ex 17
Ex 18
Ex 19
CAPRIN1
1.00
1.31
RANBP2
1.00
1.00
Western blot
CAPRIN1
1.00
0.45
β-actin

**Figure S2. A motor neuron-specific database of poly(A) sites enables quantitative understanding of polyadenylation at thousands of previously unannotated loci; related to Figure 2.**

A. Annotation of novel polyadenylation sites (PASs) reveals ⅔ of those used in motor neurons are not previously annotated in the PolyA_DB database (91,089 novel sites, 46,032 PolyA_DB-matching sites).
B. Subdividing novel and previously annotated polyadenylation sites (PASs) in stem cell-derived motor neurons by intragenic position reveals a majority of novel PASs are located within introns. 3' untranslated region (UTR) subdivisions are "F", "S", "M", and "L" for first, single, middle, and last, respectively. Intergenic = 10 kb beyond the annotated end of genes.
C. Positional motif enrichment analysis of polyadenylation sites with increased (left) and decreased (right) usage in $CLP1^{KO}$ compared to $CLP^{WT}$ hESC-derived motor neurons. Plots are centered on the cleavage site, the AU rich element in the -20 position indicates the polyadenylation signal.
D. Validated alternative polyadenylation events correlate in fold change by PAC-Seq and qRT-PCR (n = 8 target genes, Pearson $r$ = 0.48).
E. Genome browser tracks (top) of $CLP1^{KO}$ and $CLP1^{WT}$ hESC-derived motor neuron 3'-end sequencing reads mapped to *CAPRIN1* at a differentially used intronic polyadenylation (IPA) site. Increased usage of IPA site (yellow). qRT-PCR validation (middle) of IPA site and total gene expression. Representative gel images shown with qRT-PCR values below, n = 3. Western blot (bottom) for CAPRIN1 expression in $CLP1^{KO}$ and isogenic $CLP1^{WT}$ H9 hESC-derived motor neurons. Beta actin, loading control.

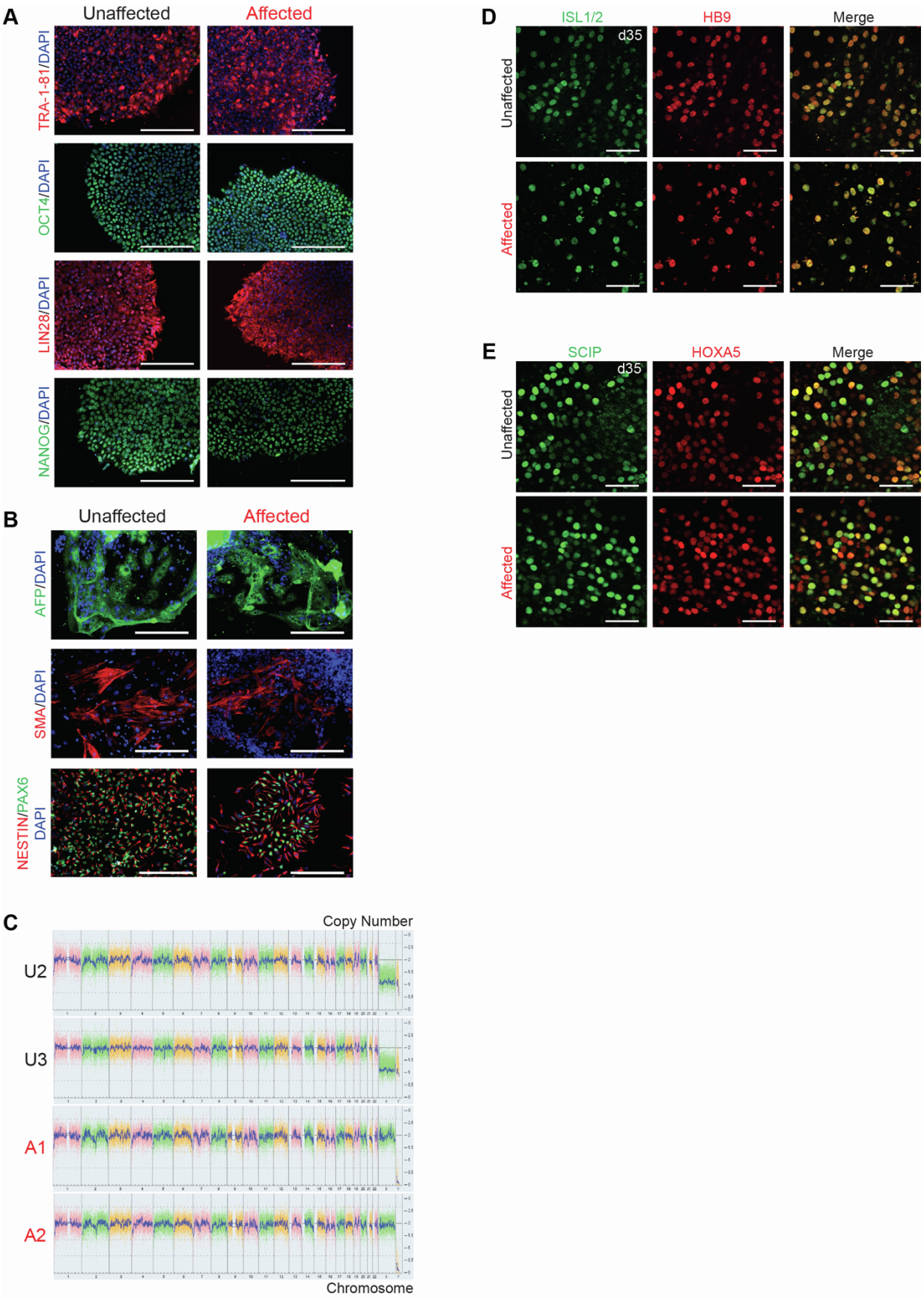
A
Unaffected
Affected
TRA-1-81/DAPI
OCT4/DAPI
LIN28/DAPI
NANOG/DAPI
B
Unaffected
Affected
AFP/DAPI
SMA/DAPI
NESTIN/PAX6
DAPI
C
Copy Number
U2
U3
A1
A2
Chromosome
D
ISL1/2
HB9
Merge
d35
Unaffected
Affected
E
SCIP
HOXA5
Merge
d35
Unaffected
Affected

**Figure S3. CLP1 p.R140H does not alter fibroblast reprogramming or motor neuron cell type specification; related to Figure 3.**

A. Induced pluripotent stem cells derived from a patient with PCH10 (Affected) and their parent (Unaffected) immunostained for pluripotency markers TRA-1-81, OCT4, LIN28, and NANOG, and counterstained with DAPI. n = 2 clonal replicates per individual, representative images shown. Scale bar: 100 μm.
B. Differentiated Affected and Unaffected iPSCs stained for markers of the endoderm, α-fetoprotein (AFP); mesoderm, smooth muscle actin (SMA); ectoderm, NESTIN and PAX6; counterstained with DAPI. n = 3 clonal replicates per individual, representative images shown. Scale bar: 100 μm.
C. Chromosome copy number analysis of Affected (A1, A2) and Unaffected (U2, U3) iPSCs. n = 2 clonal replicates per individual shown.
D. Immunofluorescence staining for motor neuron specific transcription factors ISL1/2 and HB9 in Affected and Unaffected motor neuron cultures at day (d) 35 of differentiation. Representative images shown from two clonal iPSC-derived motor neuron lines per individual. Scale bar: 50 μm.
E. SCIP and HOXA5 immunostaining in Unaffected and Affected motor neuron cultures. n = 2 clonal replicates per individual, representative images shown. Scale bar: 50 μm.

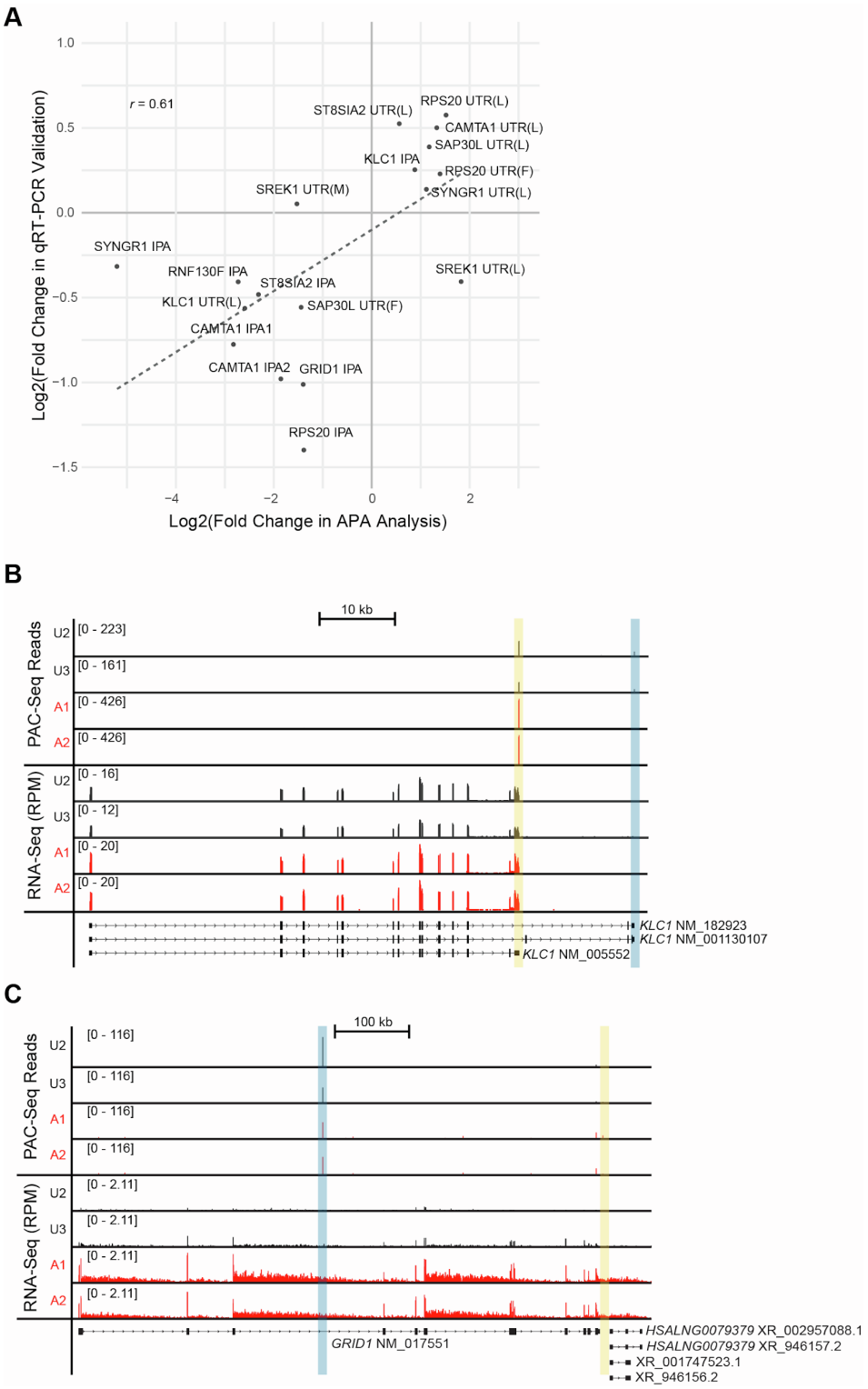
A
r = 0.61
ST8SIA2 UTR(L)
RPS20 UTR(L)
CAMTA1 UTR(L)
SAP30L UTR(L)
KLC1 IPA
RPS20 UTR(F)
SYNGR1 UTR(L)
SREK1 UTR(M)
SYNGR1 IPA
RNF130F IPA
ST8SIA2 IPA
SREK1 UTR(L)
KLC1 UTR(L)
SAP30L UTR(F)
CAMTA1 IPA1
CAMTA1 IPA2
GRID1 IPA
RPS20 IPA
Log2(Fold Change in qRT-PCR Validation)
Log2(Fold Change in APA Analysis)
B
10 kb
PAC-Seq Reads
RNA-Seq (RPM)
U2
U3
A1
A2
[0 - 223]
[0 - 161]
[0 - 426]
[0 - 426]
[0 - 16]
[0 - 12]
[0 - 20]
[0 - 20]
KLC1 NM_182923
KLC1 NM_001130107
KLC1 NM_005552
C
100 kb
[0 - 116]
[0 - 2.11]
GRID1 NM_017551
HSALNG0079379 XR_002957088.1
HSALNG0079379 XR_946157.2
XR_001747523.1
XR_946156.2

**Figure S4. Validation of differential polyadenylation site usage in CLP1. p.R140H motor neurons; related to Figure 4.**

A. qRT-PCR validations of alternative polyadenylation events show concordance with the bioinformatic findings in PCH10 patient-derived motor neurons (Pearson *r* = 0.61).
B. Genome browser tracks of Unaffected (U2, U3) and Affected (A1, A2) iPSC-derived motor neuron 3'-end sequencing and RNA-sequencing reads mapped to *KLC1* at differentially used intronic polyadenylation (IPA) and distal 3' UTR sites. Increased usage of PAS (yellow), decreased usage of PAS (blue). qRT-PCR validation for these two events is shown in (A), n = 3.
C. Genome browser tracks of Unaffected (U2, U3) and Affected (A1, A2) iPSC-derived motor neuron 3'-end sequencing and RNA-sequencing reads mapped to *GRID1* at a novel, differentially used intronic polyadenylation (IPA) site. Decreased usage of PAS (blue), increased RNA-seq read density (yellow). qRT-PCR validation for the IPA event shown in (A), n = 3.

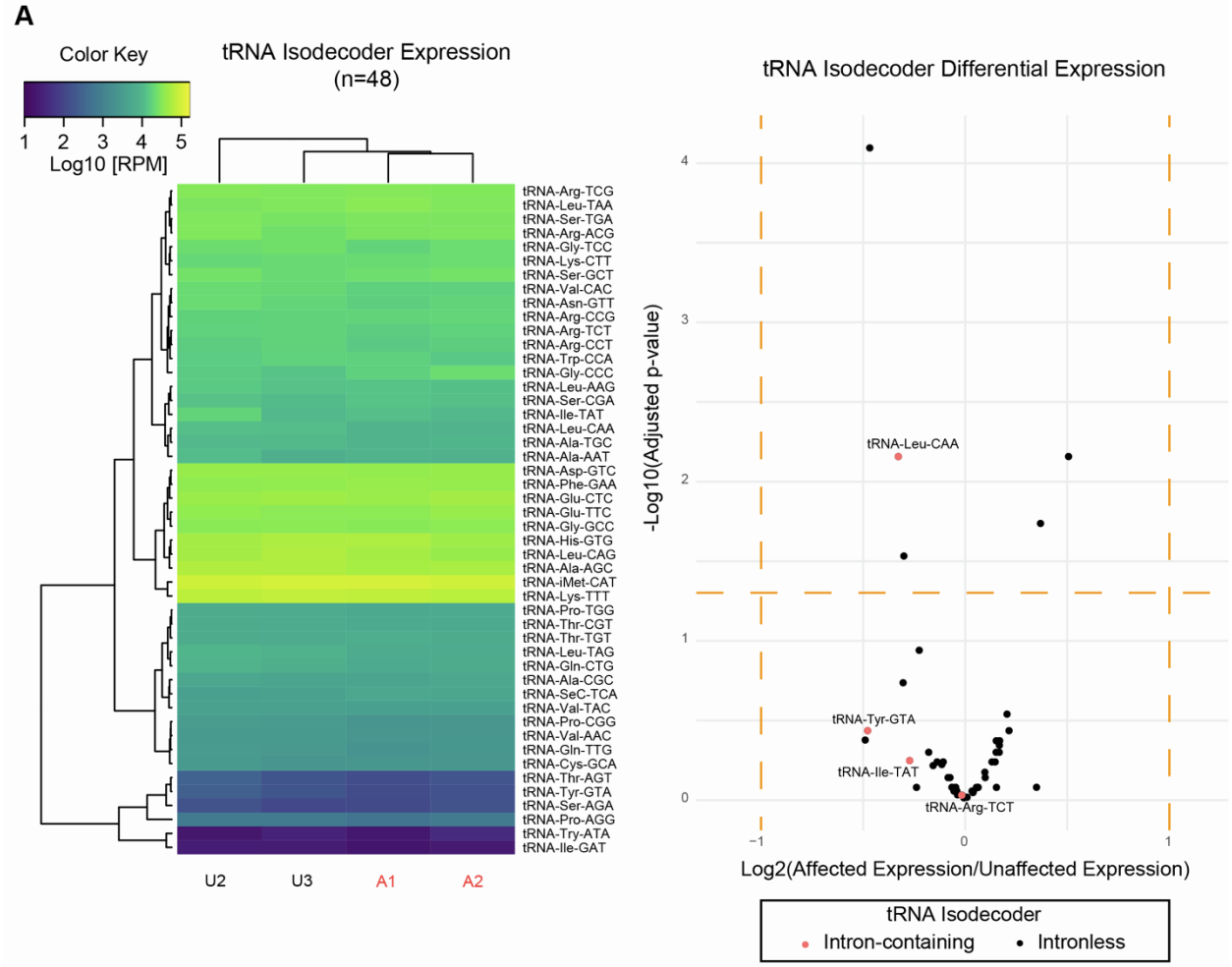
A
Color Key
1 2 3 4 5
Log10 [RPM]
tRNA Isodecoder Expression
(n=48)
tRNA-Arg-TCG
tRNA-Leu-TAA
tRNA-Ser-TGA
tRNA-Arg-ACG
tRNA-Gly-TCC
tRNA-Lys-CTT
tRNA-Ser-GCT
tRNA-Val-CAC
tRNA-Asn-GTT
tRNA-Arg-CCG
tRNA-Arg-TCT
tRNA-Arg-CCT
tRNA-Trp-CCA
tRNA-Gly-CCC
tRNA-Leu-AAG
tRNA-Ser-CGA
tRNA-Ile-TAT
tRNA-Leu-CAA
tRNA-Ala-TGC
tRNA-Ala-AAT
tRNA-Asp-GTC
tRNA-Phe-GAA
tRNA-Glu-CTC
tRNA-Glu-TTC
tRNA-Gly-GCC
tRNA-His-GTG
tRNA-Leu-CAG
tRNA-Ala-AGC
tRNA-iMet-CAT
tRNA-Lys-TTT
tRNA-Pro-TGG
tRNA-Thr-CGT
tRNA-Thr-TGT
tRNA-Leu-TAG
tRNA-Gln-CTG
tRNA-Ala-CGC
tRNA-SeC-TCA
tRNA-Val-TAC
tRNA-Pro-CGG
tRNA-Val-AAC
tRNA-Gln-TTG
tRNA-Cys-GCA
tRNA-Thr-AGT
tRNA-Tyr-GTA
tRNA-Ser-AGA
tRNA-Pro-AGG
tRNA-Try-ATA
tRNA-Ile-GAT
U2 U3 A1 A2
tRNA Isodecoder Differential Expression
-Log10(Adjusted p-value)
4
3
2
1
0
tRNA-Leu-CAA
tRNA-Tyr-GTA
tRNA-Ile-TAT
tRNA-Arg-TCT
−1
0
1
Log2(Affected Expression/Unaffected Expression)
tRNA Isodecoder
Intron-containing
Intronless
B
Exon Probes
tRNA-Tyr-GTA
tRNA-Leu-CAA
tRNA-Ile-TAT
U2 U3 A1 A2
nts
150
80
50
25
21
17
pre-tRNA
mature tRNA
U6
Percent
mature tRNA:
40.6 43.4 44.6 30.0
45.9 46.2 53.7 47.5
38.3 39.6 45.6 30.0
C
Intron Probe
pre-tRNA
Ile-TAT-1-1
U2 U3 A1 A2
nts
150
80
50
25
21
17
pre-tRNA
U6

**Figure S5. PCH10 motor neurons do not have altered tRNA expression or detectable intron fragment accumulation; related to Figure 4.**

A. Pontocerebellar hypoplasia (PCH10) induced pluripotent stem cell derived-motor neurons (Affected, A1, A2) show nearly identical tRNA isodecoder gene expression levels to familial controls (Unaffected, U2, U3). Heat map of tRNA isodecoder reads per million (RPM) detected by tRNA sequencing (right) from two independent induced pluripotent stem cell (iPSC) clones per individual. Volcano plot of differential expression showing 5 of 48 isodecoders had a marginal change in Affected motor neurons (left).
B. Northern blots for intron-containing tRNAs in PCH10 iPSC-derived motor neurons (Affected, A1, A2) compared to familial controls (Unaffected, U2, U3) revealed no change in the ratio of spliced to pre-tRNAs or tRNA exon fragment accumulation. n = 2 independent clones per individual. Quantification below.
C. Northern blot for tRNA IleTAT using an intron probe found no differences in pre-tRNA levels or intron fragment accumulation in PCH10 iPSC-derived motor neurons (Affected, A1, A2) compared to familial controls (Unaffected, U2, U3). n = 2 independent clones per individual.

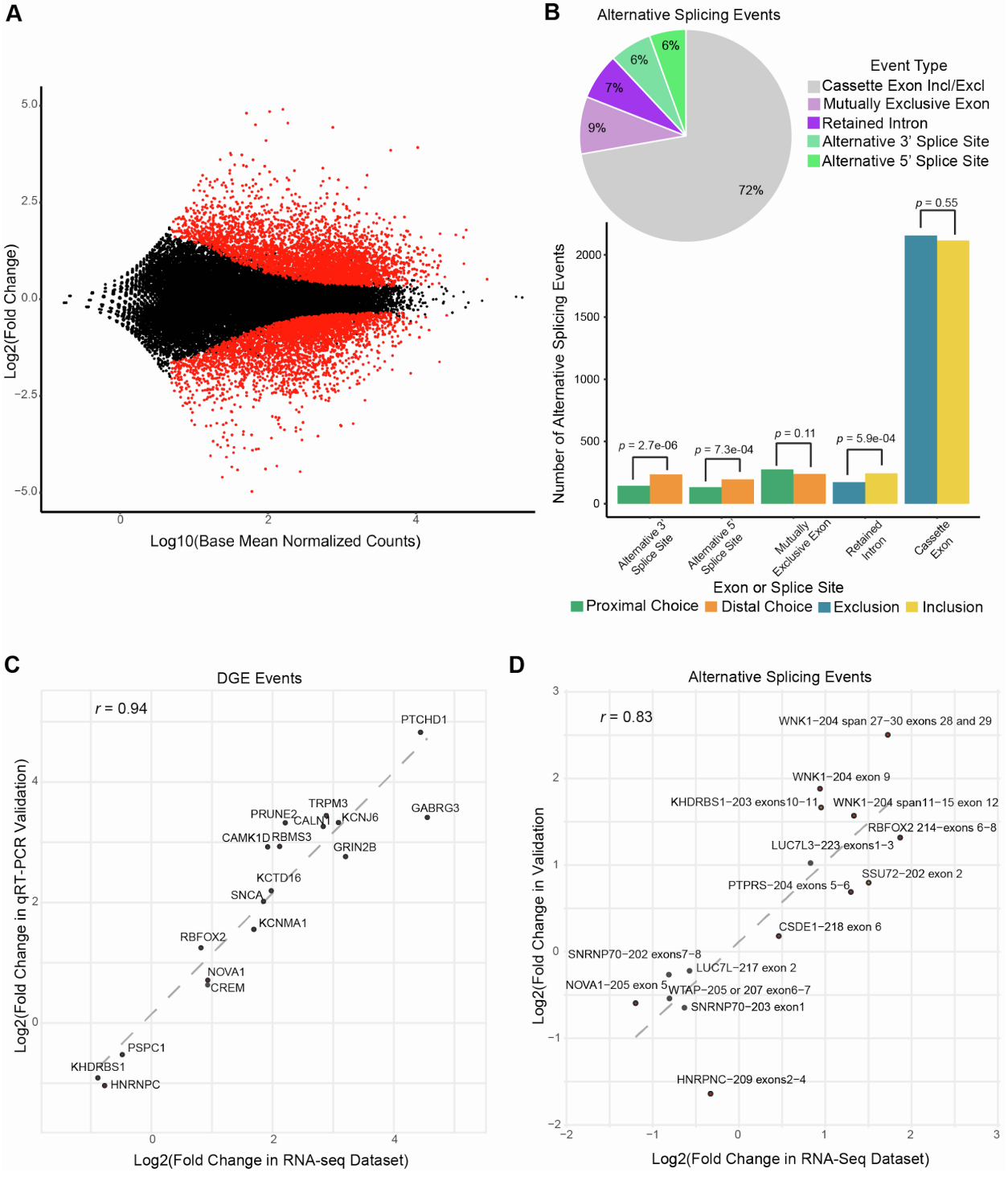
A
Log2(Fold Change)
Log10(Base Mean Normalized Counts)
B
Alternative Splicing Events
Event Type
Cassette Exon Incl/Excl
Mutually Exclusive Exon
Retained Intron
Alternative 3' Splice Site
Alternative 5' Splice Site
72%
9%
7%
6%
6%
Number of Alternative Splicing Events
p = 2.7e-06
p = 7.3e-04
p = 0.11
p = 5.9e-04
p = 0.55
Alternative 3' Splice Site
Alternative 5' Splice Site
Mutually Exclusive Exon
Retained Intron
Cassette Exon
Exon or Splice Site
Proximal Choice
Distal Choice
Exclusion
Inclusion
C
DGE Events
r = 0.94
Log2(Fold Change in qRT-PCR Validation)
Log2(Fold Change in RNA-seq Dataset)
PTCHD1
GABRG3
TRPM3
PRUNE2
CALN1
KCNJ6
CAMK1D
RBMS3
GRIN2B
KCTD16
SNCA
KCNMA1
RBFOX2
NOVA1
CREM
PSPC1
KHDRBS1
HNRNPC
D
Alternative Splicing Events
r = 0.83
Log2(Fold Change in Validation)
Log2(Fold Change in RNA-Seq Dataset)
WNK1−204 span 27−30 exons 28 and 29
WNK1−204 exon 9
KHDRBS1−203 exons10−11
WNK1−204 span11−15 exon 12
RBFOX2 214−exons 6−8
LUC7L3−223 exons1−3
SSU72−202 exon 2
PTPRS−204 exons 5−6
CSDE1−218 exon 6
SNRNP70−202 exons7−8
LUC7L−217 exon 2
NOVA1−205 exon 5
WTAP−205 or 207 exon6−7
SNRNP70−203 exon1
HNRNPC−209 exons2−4

**Figure S6. Overview of differential gene expression and splicing events, validated by PCR-based methods; related to Figure 5.**

A. Thousands of genes show differential expression in induced pluripotent stem cell derived motor neurons from patients with pontocerebellar hypoplasia type 10 (Affected) across a broad range of base mean expression without visible bias toward up- or downregulation.
B. Breakdown of differential splicing events by type (two-sided Binomial Test). Cassette exon inclusion/exclusion events make up the vast majority of alternative splicing events in the Affected motor neurons and do not show a preference for inclusion or exclusion; however, intron retention and distal splicing choice is slightly elevated. Incl, inclusive; Excl, exclusive.
C. Validated differentially expressed genes strongly correlate in fold change by total RNA-seq and qRT-PCR (n = 18 target genes, Pearson $r = 0.94$). DGE, differential gene expression.
D. Splicing changes detected in Affected motor neurons by bioinformatic analysis correlated strongly with RT-PCR/Tapestation validation (n = 15 target genes, Pearson $r = 0.83$). Hyphenated numbers correspond to Ensembl isoform identifiers.

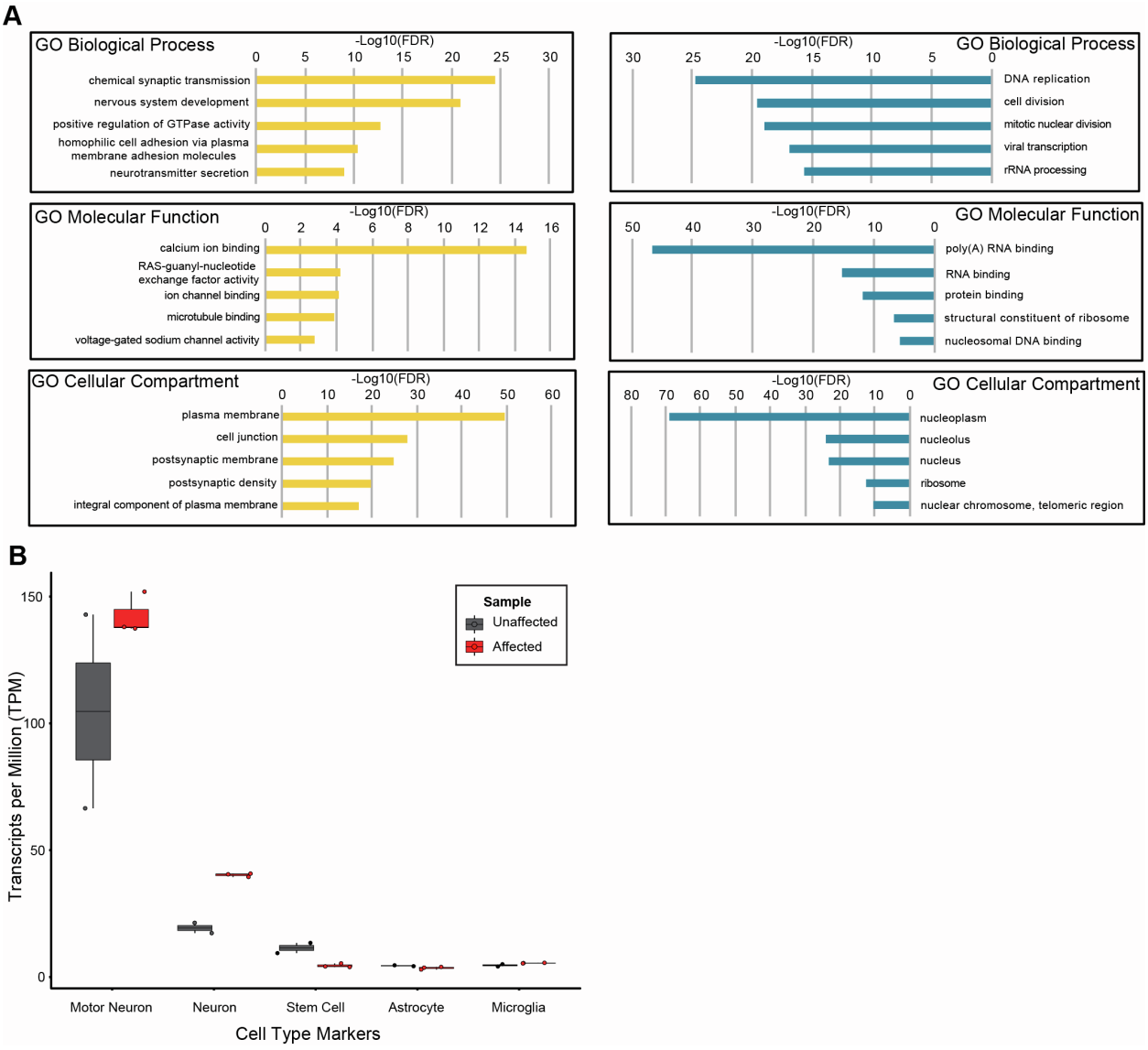


**Figure S7. Gene ontology (GO) terms enriched in induced pluripotent stem cell-derived motor neurons from patients with pontocerebellar hypoplasia type 10 (Affected) motor neurons are not caused by changes in cell type composition; related to Figure 6.**

A. Top five biological processes, molecular function, and cellular compartment GO terms enriched from genes differentially expressed in Affected motor neurons, ranked by false discovery rate (FDR). Terms enriched in upregulated genes (yellow) and downregulated genes (blue), respectively.

B. Geometric mean expression of cell type specific marker genes produces no statistically significant differences in cell type abundance and shows that motor neuron markers are expressed at a much higher proportion than markers for other potentially non-motor neuron cell types.

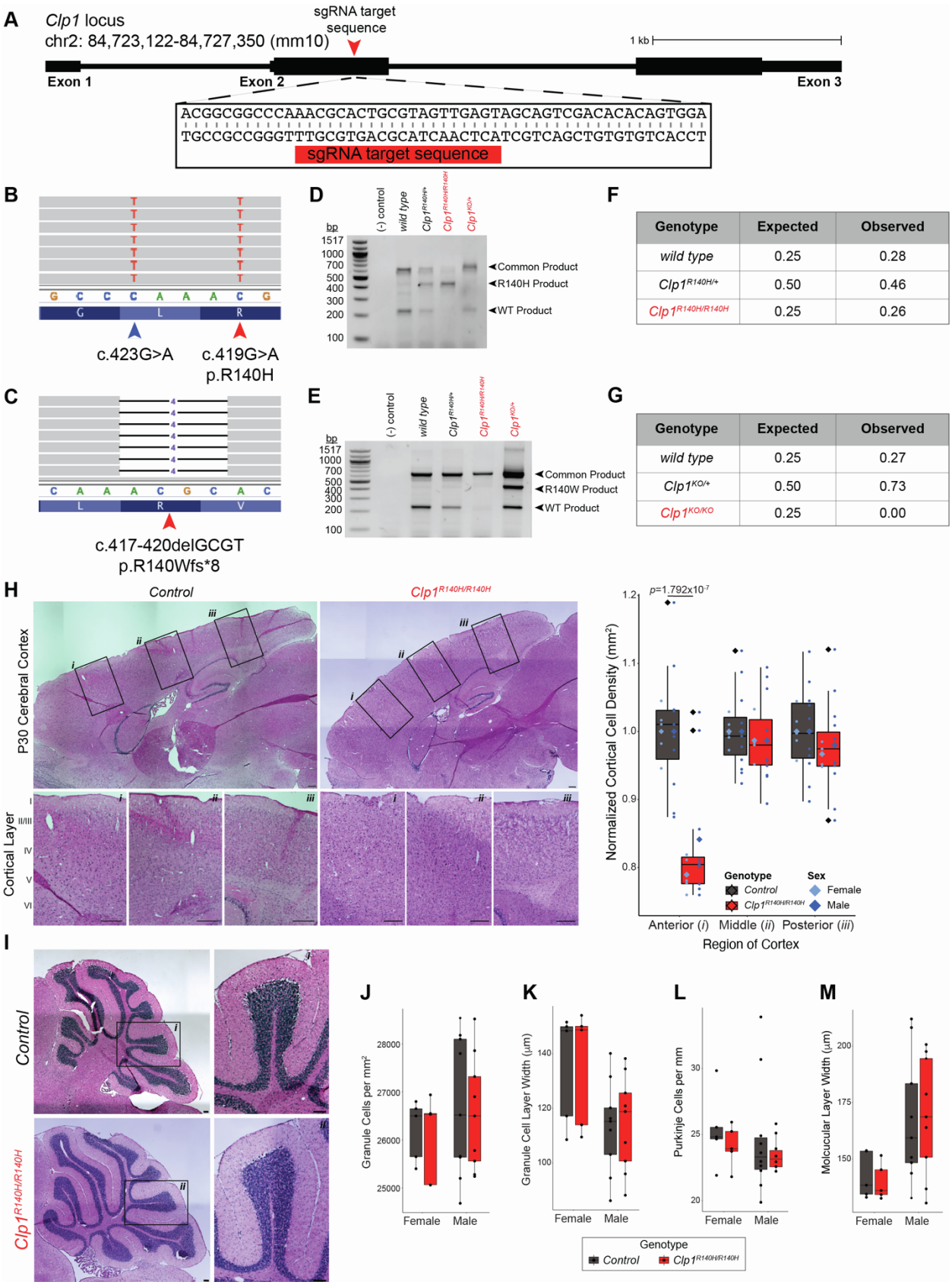
A
Clp1 locus
chr2: 84,723,122-84,727,350 (mm10)
sgRNA target sequence
1 kb
Exon 1
Exon 2
Exon 3
ACGGCGGCCCAAACGCACTGCGTAGTTGAGTAGCAGTCGACACACAGTGGA
TGCCGCCGGGTTTGCGTGACGCATCAACTCATCGTCAGCTGTGTGTCACCT
sgRNA target sequence
B
c.423G>A
c.419G>A
p.R140H
C
c.417-420delGCGT
p.R140Wfs*8
D
(-) control
wild type
Clp1R140H/+
Clp1R140H/R140H
Clp1KO/+
bp
1517
1000
700
500
400
300
200
100
Common Product
R140H Product
WT Product
E
(-) control
wild type
Clp1R140H/+
Clp1R140H/R140H
Clp1KO/+
bp
1517
1000
700
500
400
300
200
100
Common Product
R140W Product
WT Product
F
Genotype | Expected | Observed
wild type | 0.25 | 0.28
Clp1R140H/+ | 0.50 | 0.46
Clp1R140H/R140H | 0.25 | 0.26
G
Genotype | Expected | Observed
wild type | 0.25 | 0.27
Clp1KO/+ | 0.50 | 0.73
Clp1KO/KO | 0.25 | 0.00
H
Control
Clp1R140H/R140H
P30 Cerebral Cortex
Cortical Layer
I
II/III
IV
V
VI
p=1.792x10-7
Normalized Cortical Cell Density (mm2)
Genotype
Control
Clp1R140H/R140H
Sex
Female
Male
Anterior (i) Middle (ii) Posterior (iii)
Region of Cortex
I
Control
Clp1R140H/R140H
J
Granule Cells per mm2
Female
Male
K
Granule Cell Layer Width (μm)
Female
Male
L
Purkinje Cells per mm
Female
Male
M
Molecular Layer Width (μm)
Female
Male
Genotype
Control
Clp1R140H/R140H

**Figure S8. Generation of Clp1 p.R140H mutant mice and characterization of *Clp1*$^{R140H/R140H}$ brain features; related to Figure 7.**

A. Strategy for targeting the murine *Clp1* locus with CRISPR/Cas9 genome editing to introduce p.R140H. Scale bar: 1 kb.

B, C. Confirmation of edited *Clp1* alleles, and mosaicism, determined by next generation sequencing. Representative sequencing reads in B demonstrate c.419G>A (p.R140H, *Clp1*$^{R10H/R140H}$) (red arrow) and synonymous PAM site c.423G>A variant (blue arrow). Representative sequencing reads in C demonstrate c.417-420delGCGT (p.R140Wfs*8) (red arrow), resulting in a Clp1 null allele (*Clp1*$^{KO/KO}$).

D, E. Allele-specific polymerase chain reaction (AS-PCR) for *Clp1* variants (see **Data File S1L**). Representative gel image in D demonstrates primer specificity for Clp1 p.R140H. Representative gel image in E demonstrates primer specificity for Clp1 p.R140Wfs*8.

F. Litter ratio analysis for *Clp1*$^{R140H/R140H}$ mice demonstrates expected ratios of *wild type*, *Clp1*$^{R140H/+}$, and *Clp1*$^{R140H/R140H}$ offspring, n = 50 mice.

G. Litter ratio analysis for *Clp1*$^{KO/KO}$ mice are consistent with embryonic lethality, n = 22 mice.

H. Hematoxylin and eosin histochemistry of sagittal sections from postnatal day (P) 30 control and *Clp1*$^{R140H/R140H}$ mutant brains. Scale bar: 500 μm. Boxes *i-iii* represent areas of the isocortex assessed for changes in cortical cell density. Scale bar: 200 μm. Quantification, right. Boxplot represents the distribution of normalized cortical cell density for *n* = 3 mice per genotype, black diamonds represent outliers, colored circles represent the cell density for 5 matched sections from individual mice divided by sex, colored diamonds represent the average cell density of 5 matched sections from sex-matched mice. Whiskers extend the full range of the data, excluding outliers (>1.5x interquartile range); two-sided Student's t-test.

I. Hematoxylin and eosin histochemistry of sagittal brain sections from postnatal day (P) 30 control (top) and *Clp1*$^{R140H/R140H}$ mutant mice (bottom). Scale bar: 500 μm. Boxes *i* and *ii* are representative areas of the cerebellar folia assessed for changes in morphology and cell density. Scale bar: 200 μm.

J-M. Quantification of cerebellar folia morphology and cell density in control and *Clp1*$^{R140H/R140H}$ littermates show no differences. (J) Granule cell density, (K) Granule cell layer width, (L) Purkinje cell density, (M) Molecular layer width. Boxplots represent the distribution of measurements for each feature for *n* = 3 mice per genotype, individual points represent measurements collected for 5 matched sections from individual mice divided by sex. Whiskers extend the full range of the data, excluding outliers (>1.5x interquartile range); two-sided Student's t-test.