Cell Reports, Volume *43*

Supplemental information

CGG repeats in the human *FMR1* gene regulate mRNA localization and cellular stress in developing neurons

Carissa L. Sirois, Yu Guo, Meng Li, Natalie E. Wolkoff, Tomer Korabelnikov, Soraya Sandoval, Jiyoun Lee, Minjie Shen, Amaya Contractor, Andre M.M. Sousa, Anita Bhattacharyya, and Xinyu Zhao

CGG Repeats in the human *FMR1* gene regulate mRNA localization and cellular stress in developing neurons

Sirois et al., 2024

Supplemental Figures and Legends

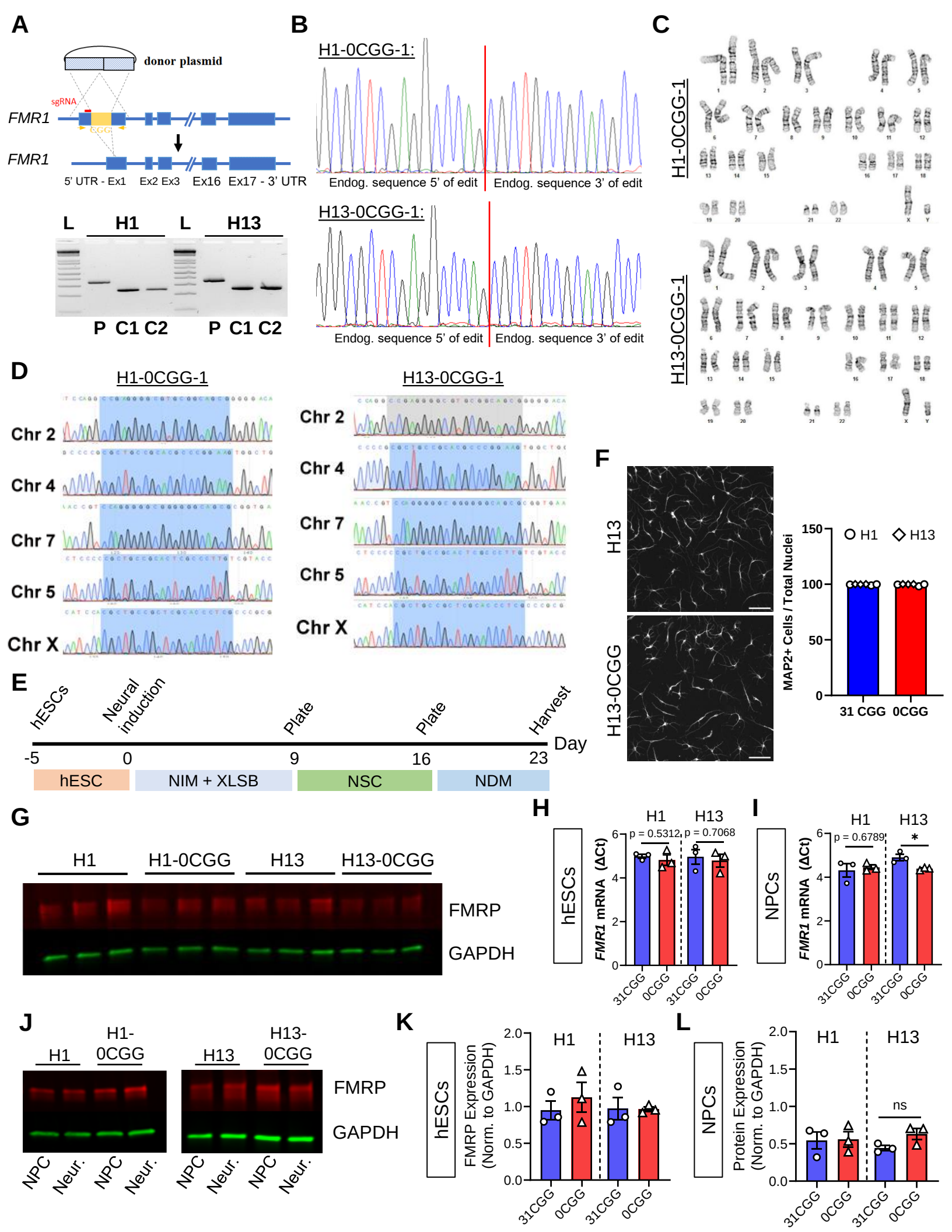

A
donor plasmid
sgRNA
FMR1
CGG
FMR1
5' UTR - Ex1
Ex2 Ex3
Ex16
Ex17 - 3' UTR
L
H1
L
H13
P C1 C2
P C1 C2
B
H1-0CGG-1:
Endog. sequence 5' of edit
Endog. sequence 3' of edit
H13-0CGG-1:
Endog. sequence 5' of edit
Endog. sequence 3' of edit
C
H1-0CGG-1
H13-0CGG-1
D
H1-0CGG-1
H13-0CGG-1
Chr 2
Chr 4
Chr 7
Chr 5
Chr X
Chr 2
Chr 4
Chr 7
Chr 5
Chr X
E
hESCs
Neural induction
Plate
Plate
Harvest
Day
-5
0
9
16
23
hESC
NIM + XLSB
NSC
NDM
F
H13
H13-0CGG
MAP2+ Cells / Total Nuclei
150
100
50
0
H1
H13
31 CGG
0CGG
G
H1
H1-0CGG
H13
H13-0CGG
FMRP
GAPDH
H
hESCs
FMR1 mRNA (ΔCt)
H1
H13
p = 0.5312
p = 0.7068
6
4
2
0
31CGG
0CGG
31CGG
0CGG
I
NPCs
FMR1 mRNA (ΔCt)
H1
H13
p = 0.6789
*
6
4
2
0
31CGG
0CGG
31CGG
0CGG
J
H1
H1-0CGG
H13
H13-0CGG
FMRP
GAPDH
NPC
Neur.
NPC
Neur.
NPC
Neur.
NPC
Neur.
K
hESCs
FMRP Expression (Norm. to GAPDH)
H1
H13
2.0
1.5
1.0
0.5
0.0
31CGG
0CGG
31CGG
0CGG
L
NPCs
Protein Expression (Norm. to GAPDH)
H1
H13
ns
2.0
1.5
1.0
0.5
0.0
31CGG
0CGG
31CGG
0CGG

Figure S1. Removal of the CGG repeats from the *FMR1* 5' UTR does not affect *FMR1* mRNA or FMRP protein levels (Related to Figure 1)

A. Top: A schematic illustration of the gene editing strategy used to precisely remove CGG repeats (yellow) from *FMR1* 5' UTR. Location of sgRNA binding site indicated by red line. Bottom: PCR showing removal of CGG repeats in gene-edited clone using the primers indicated in the schematic (yellow arrows). L: Ladder, P: Parent hESC line, C1: 0CGG-clone1, C2: 0CGG-clone 2. **B.** Sanger sequencing showing lack of CGG repeats in the 5' UTR of the *FMR1* gene in edited clones and lack of edits in endogenous sequence flanking the CGG repeats. Top: H1-0CGG, Bottom: H13-0CGG. Red line indicates where CGG repeats were located. **C.** Normal G-band karyotype in H1-0CGG (left) and H13-0CGG (right) hESCs. **D.** Lack of off-target gene editing at the top 5 predicted sgRNA binding sites (highlighted region). Left: H1-0CGG, Right: H13-0CGG. **E.** Schematic showing the timeline of the neural differentiation protocol used. **F.** Left: Representative images of H13 and H13-0CGG neurons labeled for the postmitotic neuron marker MAP2. Scale bar: 50 μm. Right: Quantification of MAP2+ neurons in 31CGG and 0CGG neuronal cultures at 1 week post-plating. n = 6 (3 independent differentiations each from N = 2 hESC lines). Statistics: two-tailed student's t-test. **G.** Western blot showing FMRP levels in parental control and isogenic 0CGG hESCS (n=3 technical replicates). **H.-I.** *FMR1* mRNA levels in 31CGG and 0CGG hESCs (H) and NPCs (I). n = 3 independent differentiations per line. Statistics: two-tailed student's t-test. *$p<0.05$ **J.** Representative Western blots showing FMRP levels in H1/H1-0CGG (left) and H13/H13-0CGG (right) NPCs and neurons. **K.-L.** Quantification of Western blots in G and H. n = 3 technical replicates (hESCs: 3 passages of hESCs; NPCs: 3 batches of differentiation). Statistics: two-tailed student's t-test. Error bars indicate SEM in all graphs.

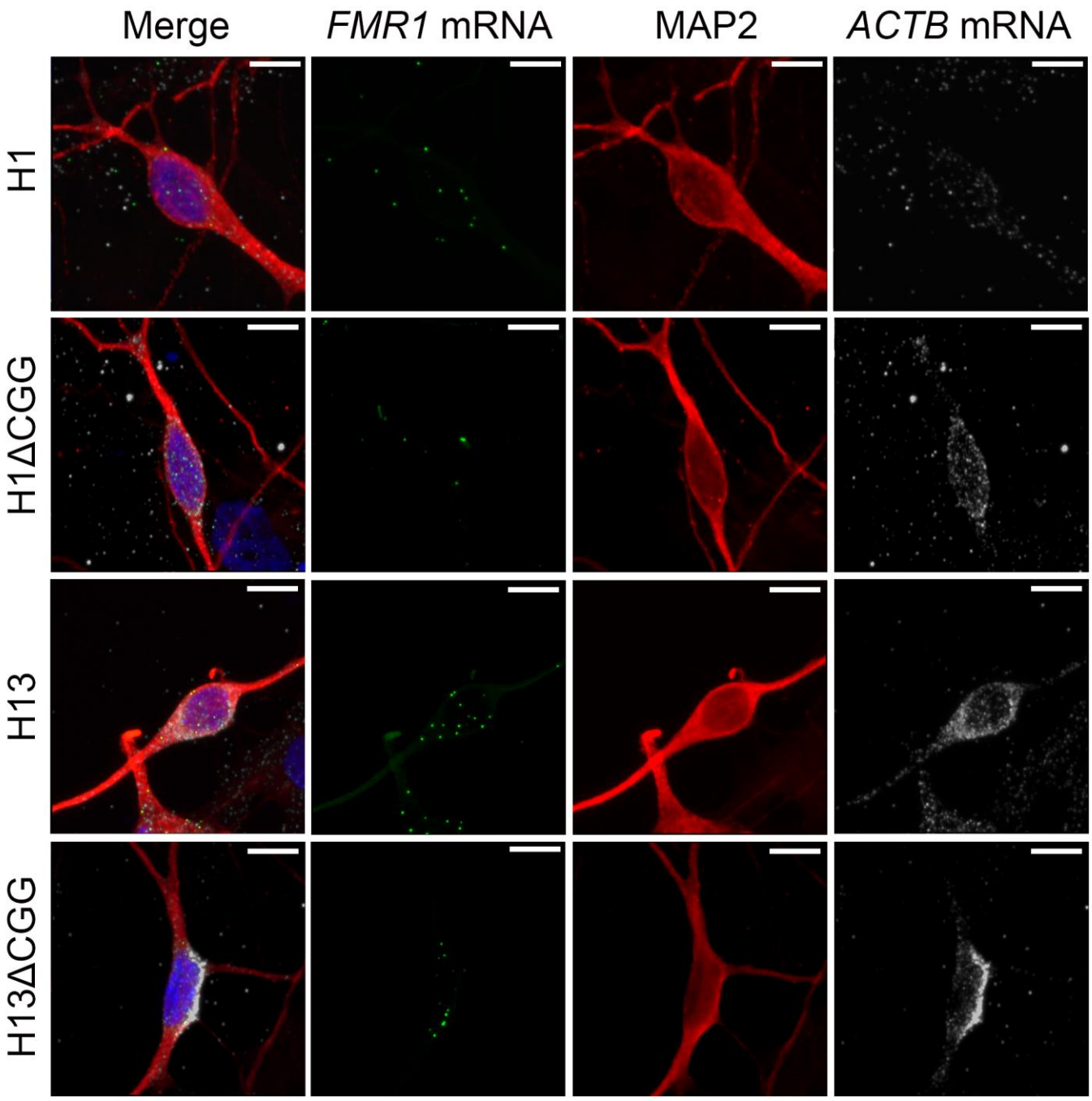


Figure S2. *FMR1* mRNA localization in 5 week neurons (Related to Figure 1)
Representative confocal images of *FMR1* and *ACTB* mRNA puncta in hESC-derived neurons stained for the post-mitotic neuron marker, MAP2, in 5 week neurons. Scale bar: 10 μm.

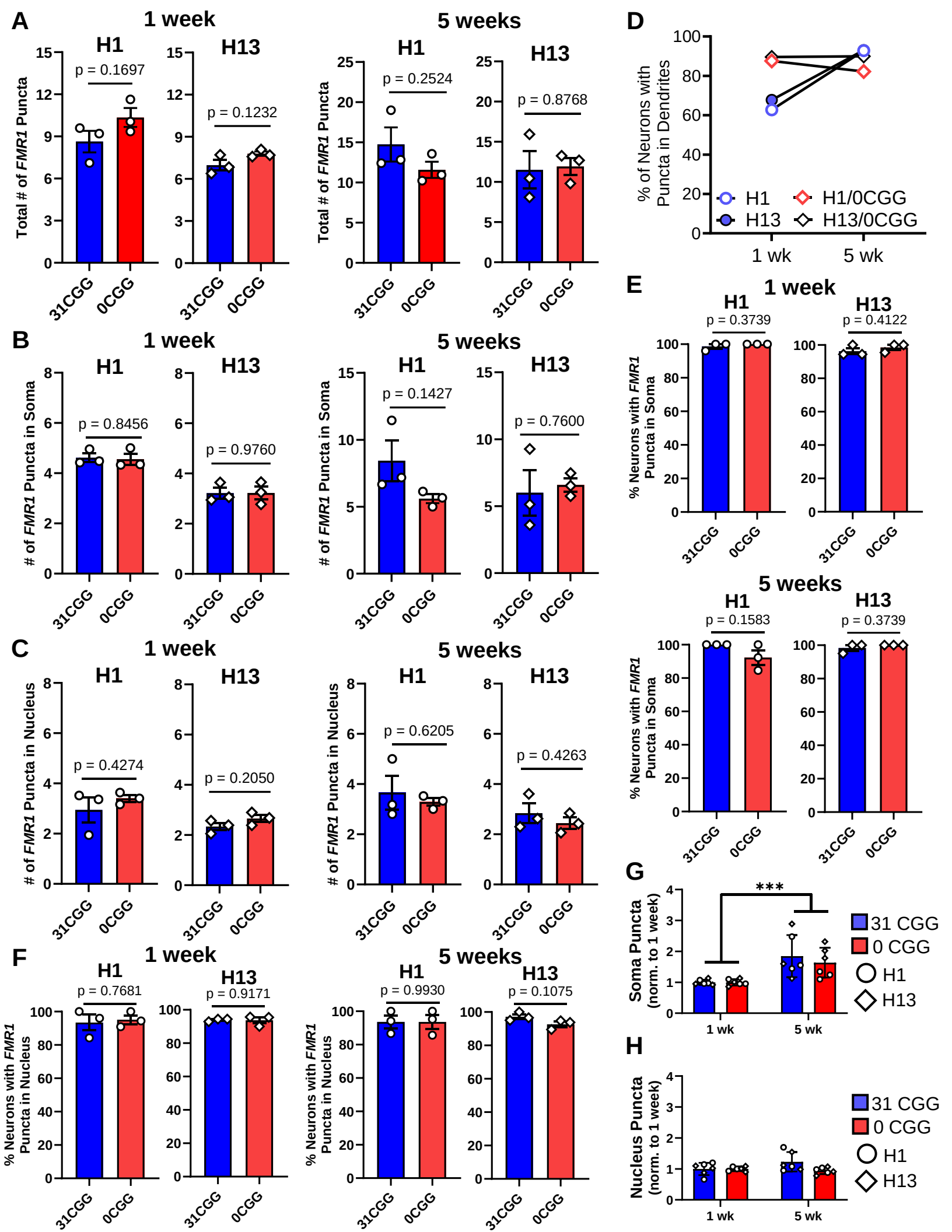

A
1 week
5 weeks
H1
H13
p = 0.1697
p = 0.1232
p = 0.2524
p = 0.8768
Total # of FMR1 Puncta
31CGG
0CGG
B
p = 0.8456
p = 0.9760
p = 0.1427
p = 0.7600
# of FMR1 Puncta in Soma
C
p = 0.4274
p = 0.2050
p = 0.6205
p = 0.4263
# of FMR1 Puncta in Nucleus
D
% of Neurons with Puncta in Dendrites
H1
H1/0CGG
H13
H13/0CGG
1 wk
5 wk
E
p = 0.3739
p = 0.4122
p = 0.1583
p = 0.3739
% Neurons with FMR1 Puncta in Soma
F
p = 0.7681
p = 0.9171
p = 0.9930
p = 0.1075
% Neurons with FMR1 Puncta in Nucleus
G
***
Soma Puncta (norm. to 1 week)
31 CGG
0 CGG
H
Nucleus Puncta (norm. to 1 week)

Figure S3. Quantification of *FMR1* mRNA in 1 and 5 week neurons (Related to Figure 1) **A.-C.** Quantification of *FMR1* puncta in 1 week (left) and 5 week (right) neurons, showing total number of puncta per neuron (A), the number of puncta in the soma of neurons (B), and the number of puncta in the nucleus of neurons (C). n = 3 technical replicates from N = 2 isogenic hESC lines. Each data point represents the average of $\geq$ 11 neurons. **D.** Percentage of neurons with at least one *FMR1* puncta localized to dendrites in 1 week versus 5 week neurons (related to Figure 1F,1G). **E.-F.** Percentage of cells with at least one *FMR1* puncta localized to the soma (E) and to the nucleus (F). n = 3 technical replicates from N = 2 isogenic hESC lines. Each data point represents the average of $\geq$ 11 neurons. Statistics: two-tailed student's t-test. **G, H.** Comparison of *FMR1* mRNA puncta in the soma (G) and nucleus (H) in 1 week versus 5 week neurons, with all data normalized to 1 week time point. n = 3 technical replicates N = 2 isogenic hESC lines. Each data point represents the average of $\geq$ 11 neurons. Statistics: two-way ANOVA ***p<0.005. Error bars indicate SEM in all graphs.

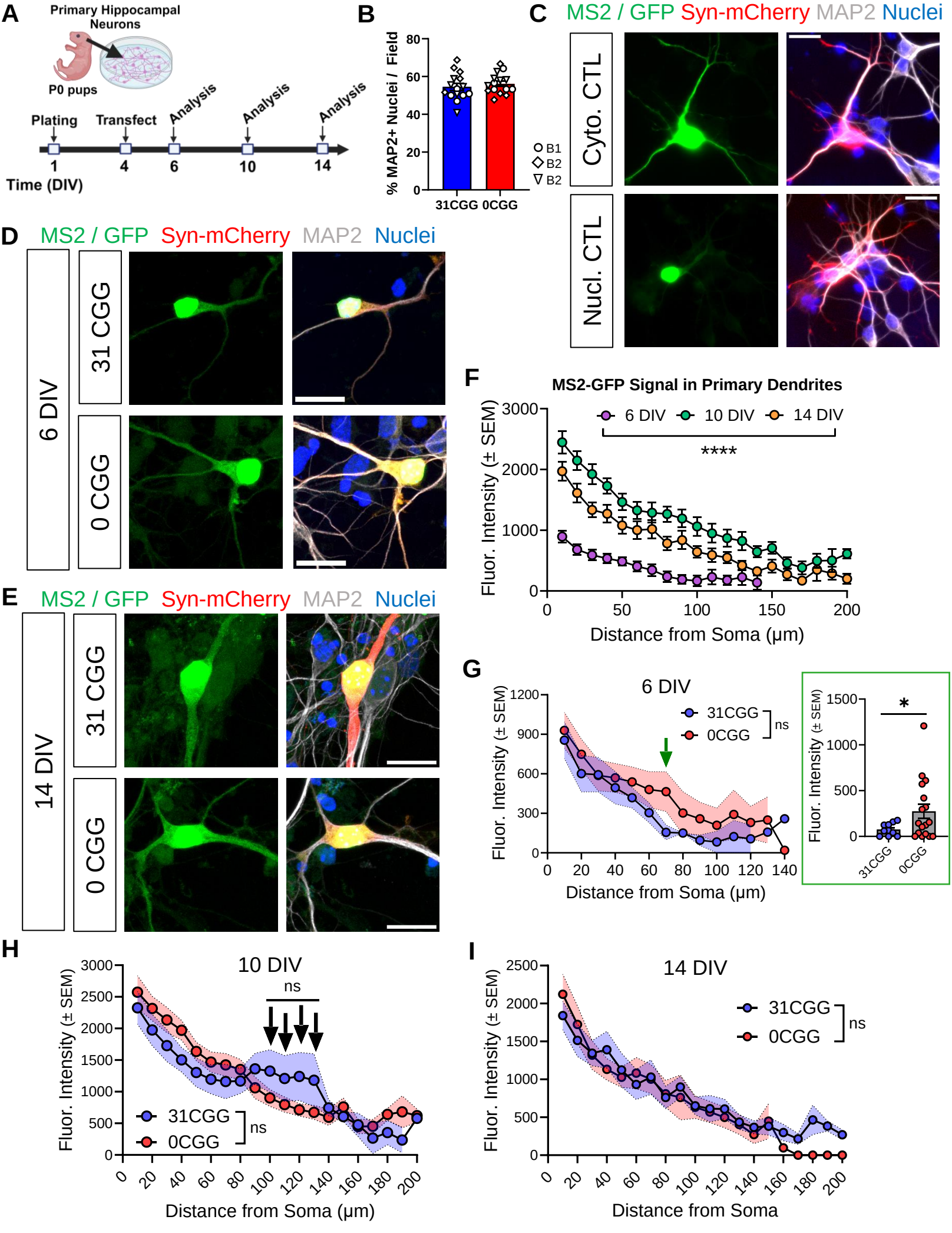

A
Primary Hippocampal Neurons
P0 pups
Plating
Transfect
Analysis
Analysis
Analysis
1
4
6
10
14
Time (DIV)
B
% MAP2+ Nuclei / Field
31CGG
0CGG
B1
B2
B2
C
MS2 / GFP Syn-mCherry MAP2 Nuclei
Cyto. CTL
Nucl. CTL
D
MS2 / GFP Syn-mCherry MAP2 Nuclei
6 DIV
31 CGG
0 CGG
E
MS2 / GFP Syn-mCherry MAP2 Nuclei
14 DIV
31 CGG
0 CGG
F
MS2-GFP Signal in Primary Dendrites
6 DIV
10 DIV
14 DIV
****
Fluor. Intensity (± SEM)
Distance from Soma (µm)
G
6 DIV
31CGG
0CGG
ns
Fluor. Intensity (± SEM)
Distance from Soma (µm)
*
H
10 DIV
ns
31CGG
0CGG
ns
Fluor. Intensity (± SEM)
Distance from Soma (µm)
I
14 DIV
31CGG
0CGG
ns
Fluor. Intensity (± SEM)
Distance from Soma

Figure S4. The CGG repeats in human *FMR1* 5' UTR affects mRNA localization in mouse neurons (Related to Figure 2)

A. A schematic diagram indicating the timing of transfection and analysis of mouse neurons in S3B-H. **B.** Quantification of MAP2+ neurons in primary neuron cultures. Each dot represents one field imaged at 20x. n = 15 fields from N = 3 independent neuron isolations (5 fields per isolation). Statistics: two-tailed student's t-test. **C.** MCP-GFP protein is localized to the cytoplasm when co-transfected with a cytoplasmic-targeted MS2 mRNA (top) and localized to the nucleus in the absence of MS2 mRNA (bottom). **D, E.** Representative confocal images of neurons at 6 days *in vitro* (DIV) (**D**) and at 14 DIV (**E**). **F.** Comparison of GFP signal along primary dendrites at 3 time points. 31CGG and 0CGG neurons were combined into one group for each time point. Two-way ANOVA, significant main effect of time in culture, ****p<0.001 **G.** Left: comparison of GFP signal along primary dendrites in 31CGG versus 0CGG conditions at 6 DIV. Right: bar graph showing differences in GFP intensity at 70 μm from the soma (indicated by green arrow in line graph). Statistics: left: two-way ANOVA, right: Welch's t-test, *p<0.05 **H, I.** Comparison of GFP signal along primary dendrites in 31CGG versus 0CGG conditions at 10 DIV (**H**), and 14 DIV (**I**) post-transfection. Statistics: two-way ANOVA. Arrows in (**H**) indicate non-significant differences between 31CGG and 0CGG, as determined by Welch's t-test. For all graphs, n = 13-19 cells per condition from N = 1 neuron isolation/transfection. Scale bars: 20 μm. Error bars (B, F) and shaded areas (G-I) indicate SEM.

A

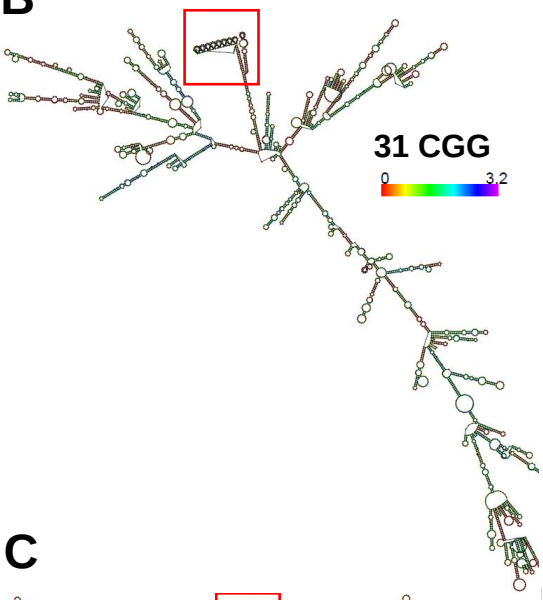
31 CGG

```
CTCAGTCAGG CGCTCAGCTC CGTTTCGGTT TCACTTCCGG TGGAGGCCCG CCTCTGAGCG GGCGGCGGGC CGACGGCGAG CGCGGGCGGC GGCGGTGACG
GAGGCGCCGC TGCCAGGGGG CGTGCGGCAG CGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG
GCGGCGGCGG CGGCGGCGGC GGCGGCTGGG CCTCGAGCGC CCGCAGCCCA CCTCTCGGGG GCGGGCTCCC GGCGCTAGCA GGGCTGAAGA GAAGATGGAG
GAGCTGGTGG TGGAAGTGCG GGGCTCCAAT GGCGCTTTCT ACAAGGCATT TGTAAAGGAT GTTCATGAAG ATTCAATAAC AGTTGCATTT GAAAACAACT
GGCAGCCTGA TAGGCAGATT CCATTTCATG ATGTCAGATT CCCACCTCCT GTAGGTTATA ATAAAGATAT AAATGAAAGT GATGAAGTTG AGGTGTATTC
CAGAGCAAAT GAAAAAGAGC CTTGCTGTTG GTGGTTAGCT AAAGTGAGGA TGATAAAGGG TGAGTTTTAT GTGATAGAAT ATGCAGCATG TGATGCAACT
```

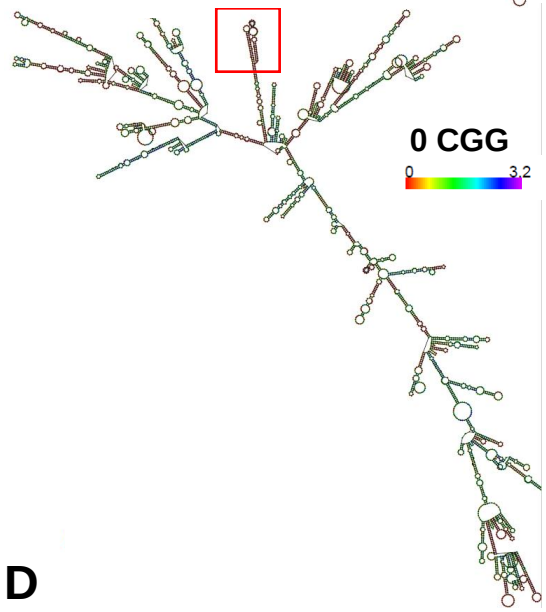
0 CGG

```
CTCAGTCAGG CGCTCAGCTC CGTTTCGGTT TCACTTCCGG TGGAGGCCCG CCTCTGAGCG GGCGGCGGGC CGACGGCGAG CGCGGGCGGC GGCGGTGACG
GAGGCGCCGC TGCCAGGGGG CGTGCGGCAG CGCTGGGCCT CGAGCGCCCG CAGCCCACCT CTCGGGGGCG GGCTCCCGGC GCTAGCAGGG CTGAAGAGAA
GATGGAGGAG CTGGTGGTGG AAGTGCGGGG CTCCAATGGC GCTTTCTACA AGGCATTTGT AAAGGATGTT CATGAAGATT CAATAACAGT TGCATTTGAA
AACAACTGGC AGCCTGATAG GCAGATTCCA TTTCATGATG TCAGATTCCC ACCTCCTGTA GGTTATAATA AAGATATAAA TGAAAGTGAT GAAGTTGAGG
TGTATTCCAG AGCAAATGAA AAAGAGCCTT GCTGTTGGTG GTTAGCTAAA GTGAGGATGA TAAAGGGTGA GTTTTATGTG ATAGAATATG CAGCATGTGA
```

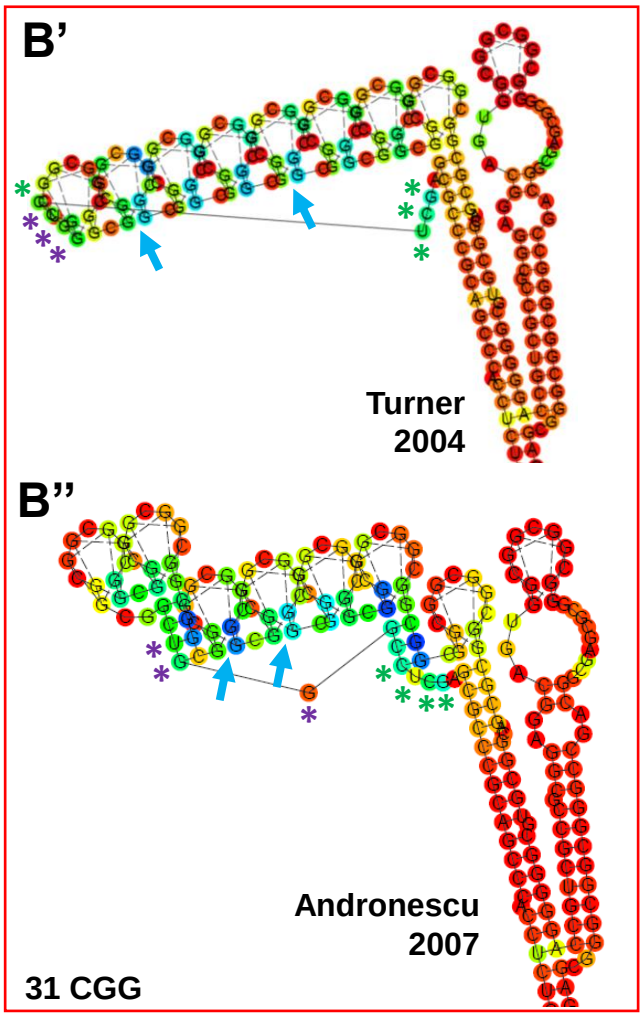
B

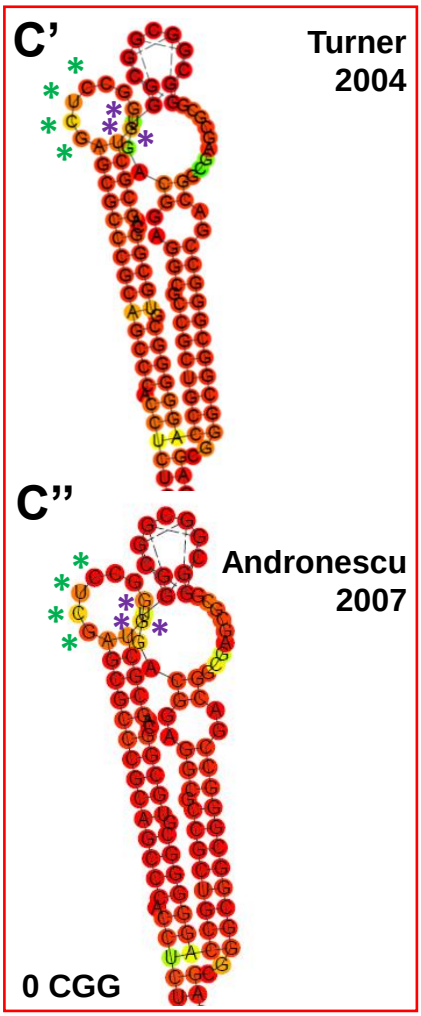


C



	# of QGRS Found
31 CGG	23
0 CGG	16





D

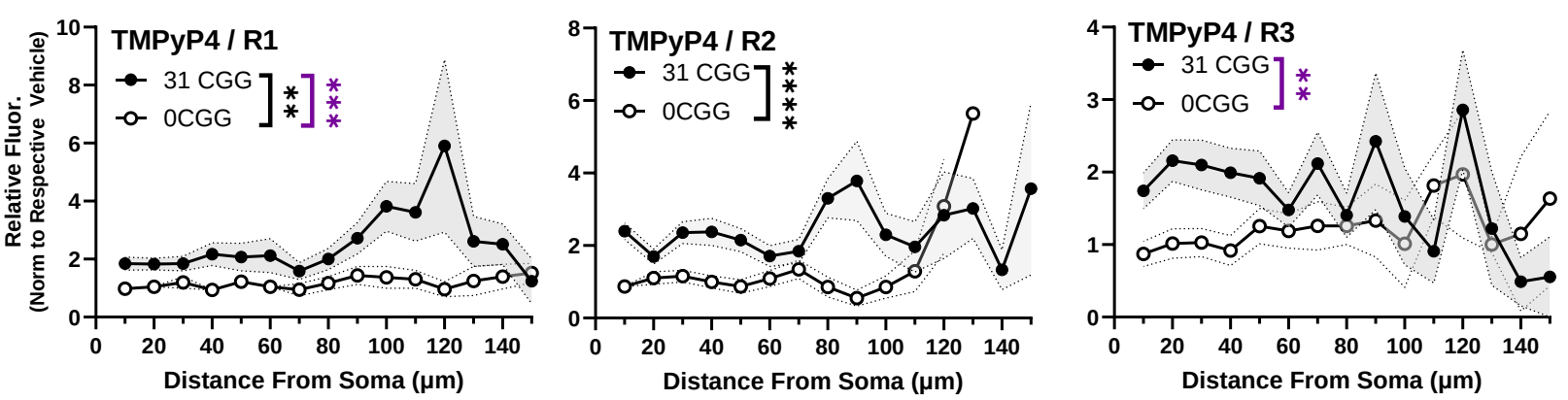
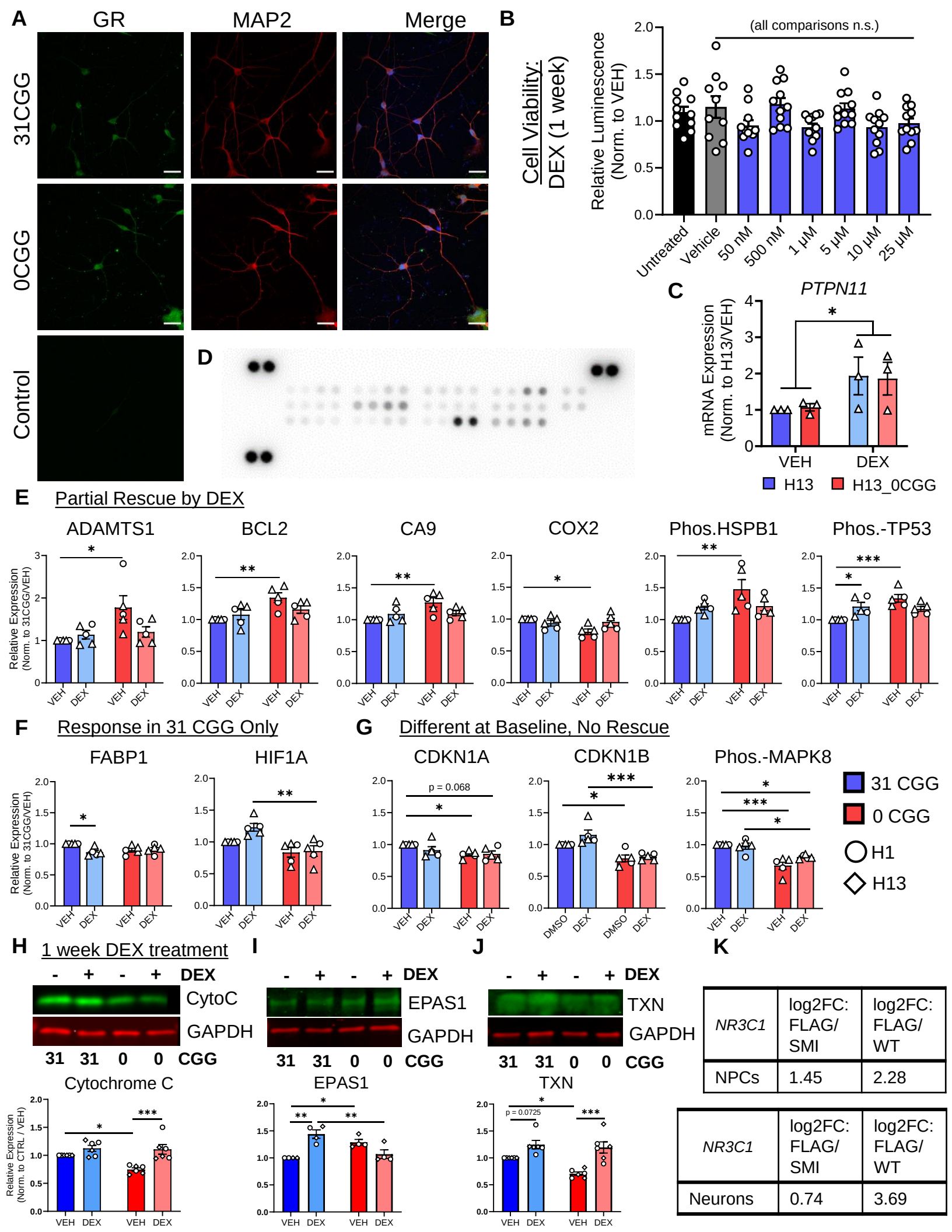


Figure S5. Predicted secondary structure of *FMR1* mRNA with or without 31 CGG repeats (Related to Figure 3)

A. Putative G quadruplex structures in the *FMR1* 5' UTR (yellow) predicted using QGRS Mapper (Kikin et al., 2006). The blue box highlights CGG repeats in the 31 CGG (top) sequence. Red boxes highlight the sequence immediately upstream of the CGG repeats, with predicted alterations in G4 structure in 0CGG repeats. Table at right indicates the total of number of quadruplex forming G-rich sequences (QGRS) in the *FMR1* gene with 31 CGG and 0 CGG repeats.

B, C. Predicted structure of entire *FMR1* mRNA sequence containing 31 CGG (**B**) or 0 CGG (**C**) repeats using RNAfold software (Lorenz et al., 2011) using the "Turner (2004)" prediction settings. **B', C'.** Zoom in on 5' UTR sequence highlighted by red box in **B,C**, which shows detailed predicted changes to the *FMR1* mRNA structure when CGG repeats are removed (**C'**) compared to 31 CGG repeats (**B'**). **B", C".** Predicted structure of 5' UTR using alternative prediction settings ("Anronescu 2007") of 31 CGG (**B"**) and 0 CGG (**C"**) repeats. Blue arrows indicate examples of predicted G-quadruplex structures formed by CGG repeats. Nucleotides highlighted by purple and green stars in **B', B', C', C"** correspond to the same nucleotides as the colored boxes highlighted in **A.**

D. Comparison of the fold change in GFP signal in 31CGG-TMPyP4 neurons (Figure 3C) to the fold change in GFP signal in 0CGG-TMPyP4 neurons (Figure 3E). Graphs represent each of the three batches of primary neurons transfected and treated with TMPyP4. n = 24-26 neurons (R1), 27-29 neurons (R2), 19-20 neurons (R3). Statistics: mixed effects model (REML), black indicates a significant main effect of genotype, purple indicates a significant interaction between genotype and distance from soma. **p<0.01, ***p<0.005, ****p<0.001



NR3C1	log2FC: FLAG/ SMI	log2FC: FLAG/ WT
NPCs	1.45	2.28

NR3C1	log2FC: FLAG/ SMI	log2FC: FLAG/ WT
Neurons	0.74	3.69

Figure S6. Altered cellular stress proteins in 0CGG neurons compared to 31CGG neurons (Related to Figure 5, Figure 6, Table S3)

A. ICC images showing expression of GR in hESC-derived neurons at 1 week. Top, 31CGG (H1), Middle, H1-0CGG, Bottom, Control: No primary antibody control. Scale bar: 25 µm **B.** Results from Cell Titer GLO assay showing no dramatic changes in cell viability in H9 neurons treated with DEX for 1 week (n = 11 technical replicates). **C.** DEX treatment induces expression of *PTPN11* gene Top: timeline for DEX treatment. Bottom: Quantification of qRT-PCR data from n = 3 batches of differentiation. **D.** Representative image of the protein arrays used to assay cellular stress proteins in Figure 5, S6. **E**. Proteins on cellular stress arrays that were significantly different at baseline between 31 CGG and 0 CGG that were no longer significantly different following DEX treatment ("partial rescue"). **F.** Proteins on cellular stress array that showed a significant response in 31 CGG neurons but not 0 CGG neurons. **G.** Proteins on cellular stress array that were significantly different at baseline in 31 CGG and 0 CGG neurons but that were not rescued by DEX. **H.-J.** Quantification of cellular stress proteins in neurons treated with DEX for 1 week: Cytochrome C (**H**), EPAS1 (**I**), and TXN (**J**). n = 2-3 independent batches of differentiation from N = 2 hESC lines. **K.** FMRP CLIPseq log2 fold change values from Li et al., 2020 for *NR3C1* mRNA. Statistics: B: Brown-Forsythe and Welch ANOVA tests with Dunnett's T3 multiple comparisons test; D: 2-way ANOVA *p<0.05 (main effect of treatment); E-J: 2-way ANOVA with Tukey's multiple comparison's test *p<0.05, **p<0.01, ***p<0.005. All error bars indicate SEM.

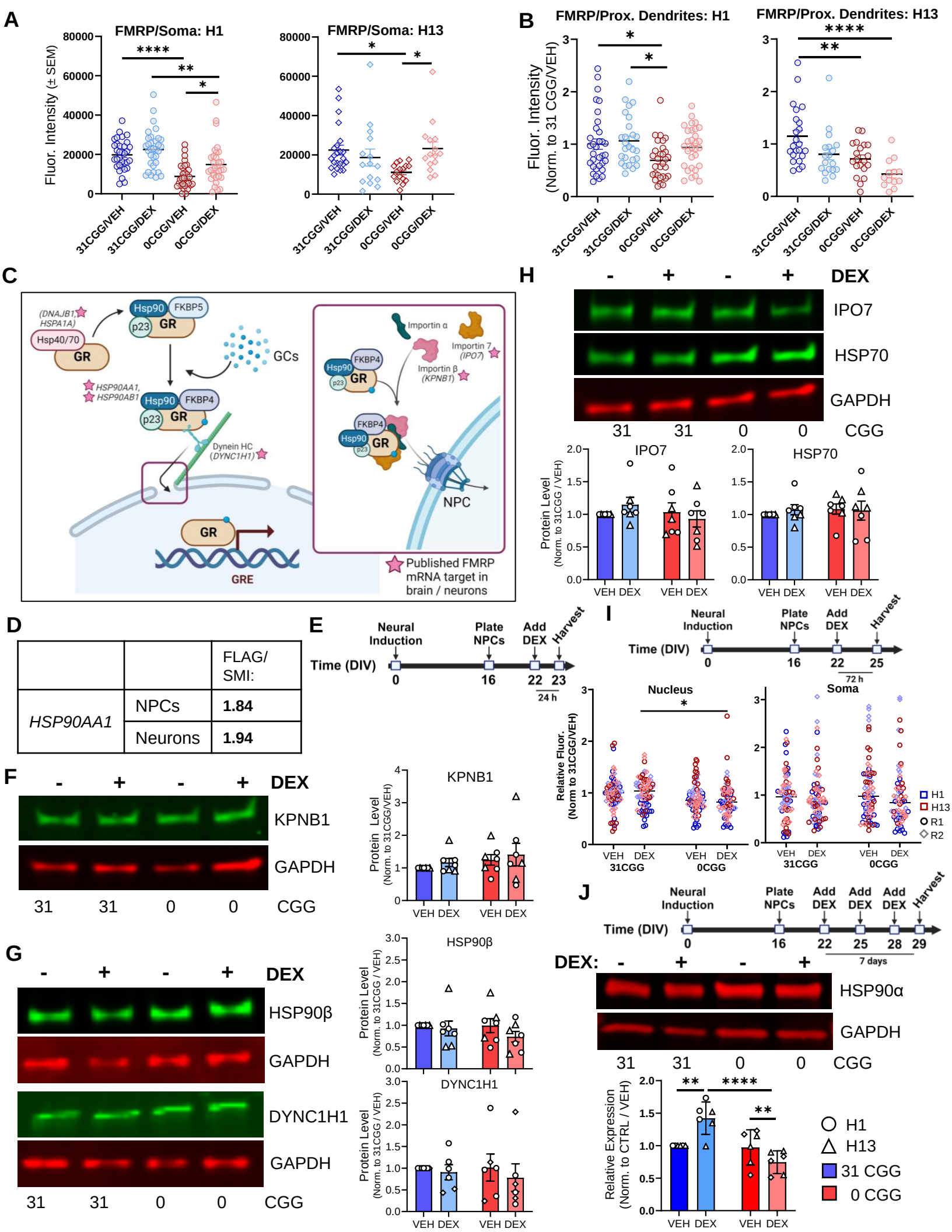

A
FMRP/Soma: H1
FMRP/Soma: H13
Fluor. Intensity (± SEM)
31CGG/VEH
31CGG/DEX
0CGG/VEH
0CGG/DEX
B
FMRP/Prox. Dendrites: H1
FMRP/Prox. Dendrites: H13
Fluor. Intensity (Norm. to 31 CGG/VEH)
C
Hsp90
FKBP5
p23
GR
Hsp40/70
(DNAJB1, HSPA1A)
GCs
HSP90AA1, HSP90AB1
FKBP4
Dynein HC (DYNC1H1)
Importin α
Importin 7 (IPO7)
Importin β (KPNB1)
NPC
GRE
Published FMRP mRNA target in brain / neurons
D
FLAG/ SMI:
HSP90AA1
NPCs
1.84
Neurons
1.94
E
Neural Induction
Plate NPCs
Add DEX
Harvest
Time (DIV)
24 h
F
DEX
KPNB1
GAPDH
CGG
Protein Level (Norm. to 31CGG/VEH)
G
HSP90β
DYNC1H1
H
IPO7
HSP70
Protein Level (Norm. to 31CGG / VEH)
I
Nucleus
Soma
Relative Fluor. (Norm to 31CGG/VEH)
72 h
H1
H13
R1
R2
J
7 days
HSP90α
Relative Expression (Norm. to CTRL / VEH)
31 CGG
0 CGG
VEH
DEX

Figure S7. Assessment of protein levels of FMRP mRNA targets involved in GR translocation (Related to Figure 7)

A. FMRP signal in the soma: the same data as in Figure 7D, with each cell line in a separate graph. **B.** FMRP signal in proximal dendrites: the same data as in Figure 7F, with each cell line in a separate graph. Statistics (A,B): two-way ANOVA followed by Tukey's multiple comparison test. *p<0.05, **p<0.01, ****p<0.001 **C.** Schematic illustrating the multiple chaperones involved in GR nuclear translocation, highlighting proteins whose mRNAs are published FMRP targets (pink star). **D.** CLIPseq log2 fold change values from Li et al. 2020 for *HSP90AA1*. **E.** Schematic showing the timing of DEX treatment for the Western blots in F-H. **F-H.** Western blots of 31CGG and 0CGG neurons treated with DEX for GR chaperones KPNB1 (F), HSP90β (G, top), DYNC1H1 (G, bottom), IPO7 (H, top), and HSP70 (H, bottom). Left: representative Western blots from H13 and H13-0CGG. Right: Quantification of protein levels from Western blots. n = 6-7 technical replicates from N = 2 isogenic pairs of cells (2-4 independent batches of differentiation and DEX treatment per line. Statistics: two-way ANOVA. **I.** GR signal intensity in neurons treated with DEX for 72 hours. Top: Schematic showing the timing of DEX treatment. Bottom left: GR signal in the nucleus. Bottom right: GR signal in the soma. Statistics: two-way ANOVA followed by Tukey's multiple comparison test. *p<0.05 **J.** Western blot of 31CGG and 0CGG neurons treated with DEX for 1 week for GR chaperone HSP90α. Top: Schematic showing the timing of DEX treatment. Middle: Representative Western blots from H13 and H13-0CGG. Bottom: Quantification of protein levels from Western blots. n = 6 technical replicates from N = 2 isogenic pairs of cells (3 independent batches of differentiation and DEX treatment per line) Statistics: two-way ANOVA followed by Tukey's multiple comparison test. **p<0.01, ****p<0.001. All error bars indicate SEM.