
Supplementary information

Mis-splicing of a neuronal microexon promotes CPEB4 aggregation in ASD

In the format provided by the
authors and unedited

SUPPLEMENTARY INFORMATION

Mis-splicing of a neuronal microexon promotes CPEB4 aggregation in ASD

Carla Garcia-Cabau^{1,#}, Anna Bartomeu^{1,#}, Giulio Tesei², Kai Chit Cheung³, Julia Pose-Utrilla^{4,5}, Sara Picó^{4,5}, Andreea Balaceanu¹, Berta Duran-Arqué¹, Marcos Fernández-Alfara¹, Judit Martín¹, Cesare De Pace⁶, Lorena Ruiz-Pérez^{6,7,8}, Jesús García¹, Giuseppe Battaglia^{6,7,9}, José J. Lucas^{4,5}, Rubén Hervás³, Kresten Lindorff-Larsen², Raúl Méndez^{1,9,*}, Xavier Salvatella^{1,9,*}

¹Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.

²Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200, Copenhagen, Denmark.

³School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR.

⁴Center for Molecular Biology Severo Ochoa (CBM Severo Ochoa) CSIC/UAM, Madrid 28049, Spain.

⁵Networking Research Center on Neurodegenerative Diseases (CIBER-NED). Instituto de Salud Carlos III, Madrid 28029, Spain.

⁶Department of Chemistry and Institute for Physics of Living Systems, University College London.

⁷Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.

⁸Department of Applied Physics, University of Barcelona.

⁹Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

#equally contributed

*to whom correspondence should be addressed

raul.mendez@irbbarcelona.org

xavier.salvatella@irbbarcelona.org

Contents

- Supplementary Data
- Supplementary Methods
- Supplementary Figures 1-5
- Supplementary Video legends

nCPEB4 and Orb2B sequence alignment

CPEB4	1	MGDYGFGVLVQSNLTGKNSAFVPRFHPHLQPPHHHQNPATPSPAAFINNNTA	50
Orb2	1	-----MDSLKLP--KANSATSSASGSNSNLS	24
CPEB4	51	ANGSSAGSAWLFPAPATHNIQDEILGSEKA-----KSQQQE	86
Orb2	25	GSTSASASAATSPTSSGTAVGGILSGAPKSPPGGLGSSTPISVRFNANEES	74
CPEB4	87	QQDPLEK---QQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKI	133
Orb2	75	LDDILQSFHHSKHSPSGGASGG-----GDASPTSN-----	104
CPEB4	134	RIESPVLTGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTII	183
Orb2	105	-----LLG---MKNNGLG---VVGNCDSLSSSPSQPQMHAGSASLFG	141
CPEB4	184	NEDAS-----FFHQ---GGVPAASANNGALLFQNFPHHVSPGFGG	220
Orb2	142	NDEVSLRNNFMQAGGFFNRKSCGGLPNLNLNK-----	173
CPEB4	221	SFSPQIGPLSQHHPHHPHFQHHSQHQQQRRSPASPHPPPFTHRNAAFNQ	270
Orb2	174	--PPQLHQQQHQQQHQQQHQQQQQLHQHQQQQLSPNLSALHHHHHQQQQ	221
CPEB4	271	LPHLANNLNKPPSPWSSYQSPSPTPSSSWSPGGGGYGWGGSQGRDHRRG	320
Orb2	222	LRESGGS-----HSPS-----SPGGGG---GGSPYNGSQAG	249
CPEB4	321	L-NGGITPL-----NSIS-PLKKN-----FASNHIQ	344
Orb2	250	CSSGGISPIPPQMGVSPKYRRSISFPIKGNSTAIYGNMHMDGMGSGHMN	299
CPEB4	345	LQKYARPSSAFAPKSWMEDSLNRADNIFFPDR-----PRTFDMH---S	385
Orb2	300	IPTLSIGNGGGGGSTGMVSAGATGGGDAPYLGNSYGNMMTSNGQMHHGGG	349
CPEB4	386	LESSLIDIMRAENDTIKARTYGRRRGQSSLFPMEDGFLDDGRGDQPLH--	433
Orb2	350	LDNSLCDYMR--NMSLGGN--GGGDGSNSMSLMQDRMRVMG---GPKHLS	392
CPEB4	434	-----SG-----LGSPHCFSHQN-----GERV	450
Orb2	393	EADAMAIAASGNDPSVYLNALKMGSPSRLSPHSPHSPIQGGNGGNVGDGT	442

Supplementary Methods

Statistical analysis of Figure 4j

Statistical significance for differences in number of positive Proteostat foci between Control and TgCPEB4Δ4 mice was assessed using a generalized linear mixed model (family=Poisson(link='identity')) with mouse as random effect.

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [`'glmerMod'`]

Family: poisson (identity)
Formula: n.proteostat.pos.CPEB4.foci ~ genotype + (1 | mouse)
Data: x

AIC	BIC	logLik	deviance	df.resid
416.2	424.6	-205.1	410.2	120

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.71846	-0.49964	-0.03384	0.57701	2.07186

Random effects:

Groups	Name	Variance	Std.Dev.
mouse	(Intercept)	0.2434	0.4933

Number of obs: 123, groups: mouse, 7

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.4478	0.3334	4.342	1.41e-05 ***
genotypeTgCPEB4Δ4 (BP3)	1.1962	0.4557	2.625	0.00866 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)
gTCPEB4Δ4(B	-0.730

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: `glmer(formula = n.proteostat.pos.CPEB4.foci ~ genotype + (1 | mouse), data = x, family = poisson(link = "identity"))`

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)
TgCPEB4Δ4 (BP3) - Control (P3) == 0	1.1962	0.4557	2.625	0.00866 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- single-step method)

Protein sequences for *in vitro* experiments

The sequences of the N-terminal domain of nCPEB4 and mutants used for the *in vitro* experiments are the following.

nCPEB4

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PHHVSPGFGGSFSPQIGPLSQHHPHHPHFQHHHSQHQQQRRSPASPHPPPFTHRNAAFNQLPHLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR
PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSSLFPMED
GFLDDGRGDQPLHSGLGSPHCFSSHQNGE

H25S

MGDYGFGVLVQSNTGNKSAFPVRFSPHLQPPHHSQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PHSVSPGFGGSFSPQIGPLSQHHPHSPHFQHHSSQHQQQRRSPASPHPPPFTHRNAAFNQLPSLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR
PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSSLFPMED
GFLDDGRGDQPLSSGLGSPHCFSSHQNGE

H50S

MGDYGFGVLVQSNTGNKSAFPVRFSPHLQPPSHSQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFSQGGVPAASANNGALLFQNF
PHSVSPGFGGSFSPQIGPLSQHSPHSPHFQSHSSQHQQQRRSPASPHPPPFTHRNAAFNQLPSLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR
PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSSLFPMED
GFLDDGRGDQPLSSGLGSPHCFSSQNGE

H25SHC

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PSHVSPGFGGSFSPQIGPLSQSHPSHPSFQSHSSQSQQQRRSPASPHPPPFTSRNAAFNQLPHLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR
PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSSLFPMED
GFLDDGRGDQPLHSGLGSPHCFSSHQNGE

H50SHC

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PSSVSPGFGGSFSPQIGPLSQSSPSSPSFQSSSSQSQQQRRSPASPHPPPFTSRNAAFNQLPSLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR

PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSLFPMED
GFLDDGRGDQPLHSGLGSPHCFSHQNGE

nCPEB4Δ4

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PHHVSPGFGGSFSPQIGPLSQHHPHHPHFQHHHSQHQQRRSPASPHPPPFTHRNAAFNQLPHLANNLNK
PPSPWSSYQSPSPTPSSSWSPGGGYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYAR
PSSAFAPKSWMEDSLNRADNIFFPDRPRTFDMHSLESSLIDIMRAENDTIKGQSSLFPMEDGFLDDGRG
DQPLHSGLGSPHCFSHQNGE

ΔHC

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PHHVSPGFGGSFSPQIGPPASPHPPPFTHRNAAFNQLPHLANNLNKPPSPWSSYQSPSPTPSSSWSPGGG
GYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYARPSSAFAPKSWMEDSLNRADNIFFP
PDRPRTFDMHSLESSLIDIMRAENDTIKARTYGRRRGQSSLFPMEDGFLDDGRGDQPLHSGLGSPHCFSH
QNGE

Δ4ΔHC

MGDYGFGVLVQSNTGNKSAFPVRFHPHLQPPHHHQNATPSPAAFINNNTAANGSSAGSAWLFPAPATHNI
QDEILGSEKAKSQQQEQQDPLEKQQLSPSPGQEAGILPETEKAKSEENQGDNSSENGNGKEKIRIESPVL
TGFDYQEATGLGTSTQPLTSSASSLTGFSNWSAAIAPSSSTIINEDASFFHQGGVPAASANNGALLFQNF
PHHVSPGFGGSFSPQIGPPASPHPPPFTHRNAAFNQLPHLANNLNKPPSPWSSYQSPSPTPSSSWSPGGG
GYGGWGSQGRDHRRGLNGGITPLNSISPLKKNFASNHIQLQKYARPSSAFAPKSWMEDSLNRADNIFFP
PDRPRTFDMHSLESSLIDIMRAENDTIKGQSSLFPMEDGFLDDGRGDQPLHSGLGSPHCFSHQNGE

Supplementary Figures

Fig. 1c

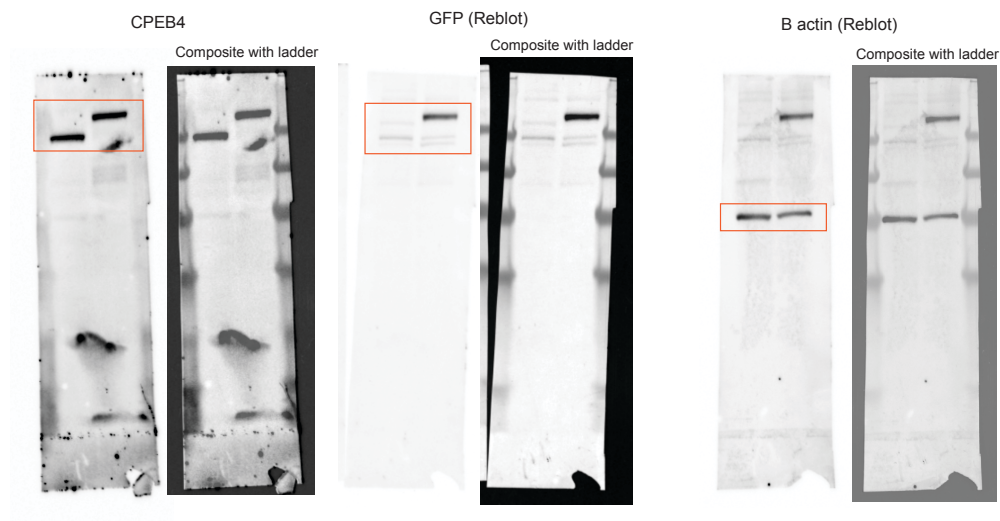
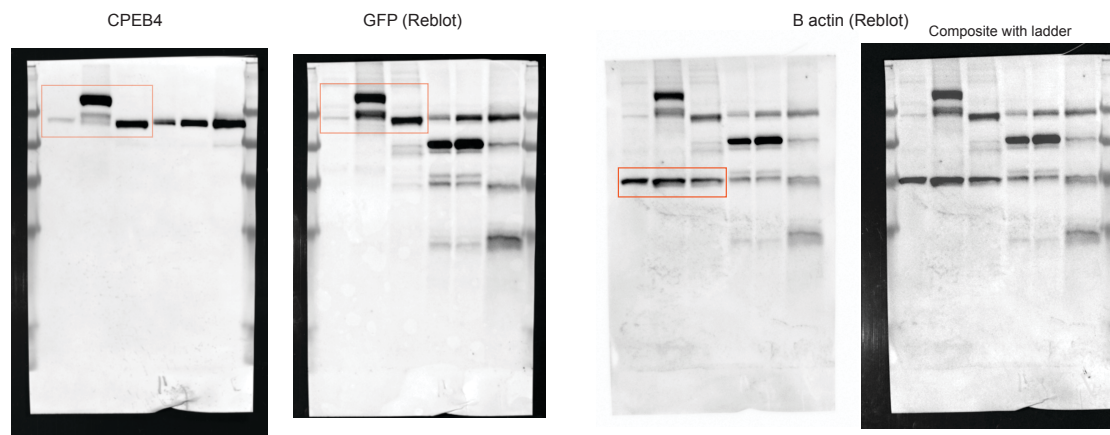


Fig. 1e

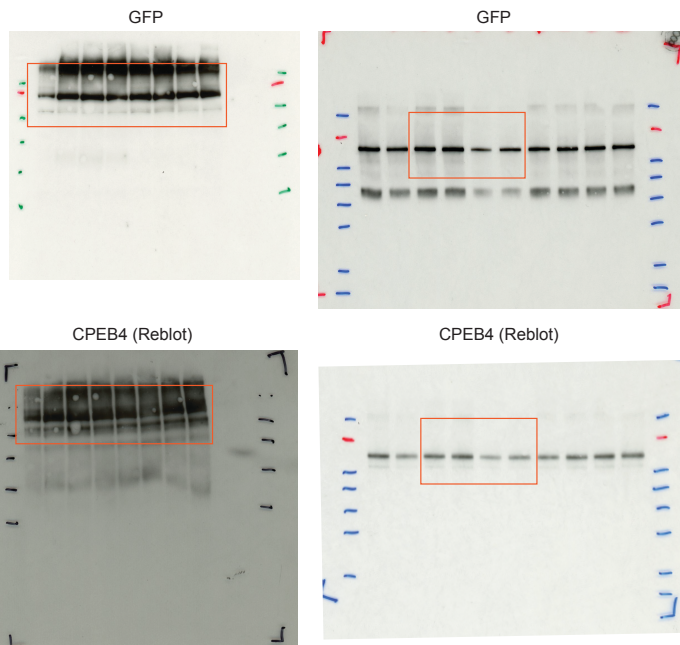


ED Fig. 1c

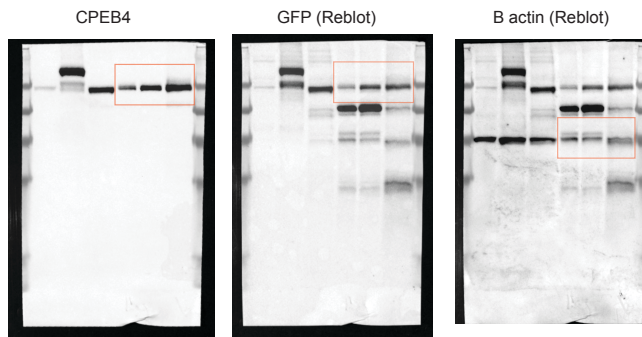


Supplementary Figure 1 | Source data - Uncropped gels and blots from Figure 1 and Extended Data Figure 1.

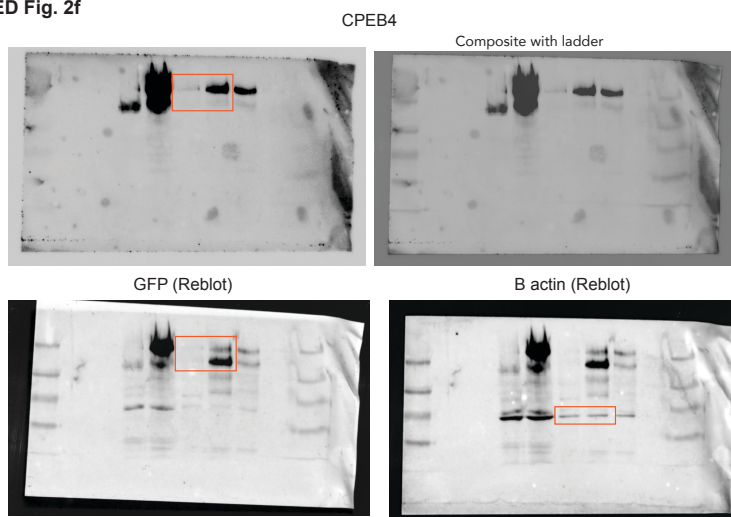
ED Fig. 2b



ED Fig. 2d

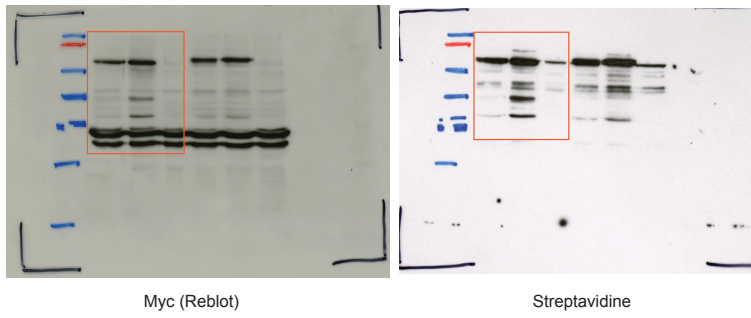


ED Fig. 2f

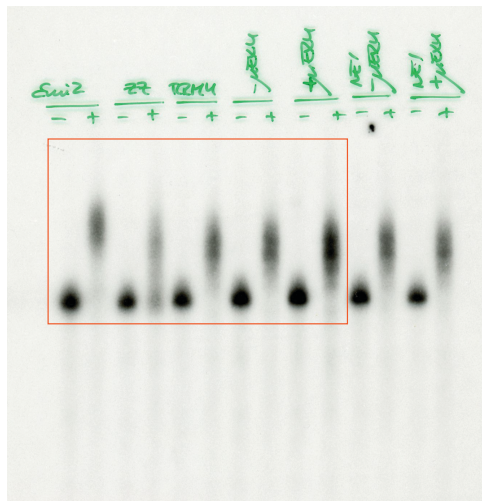


Supplementary Figure 2 | Source data - Uncropped gels and blots from Extended Data Figure 2.

ED Fig. 4f



ED Fig. 4h



Supplementary Figure 3 | Source data - Uncropped gels and blots from Extended Data Figure 4.

Fig. 4g

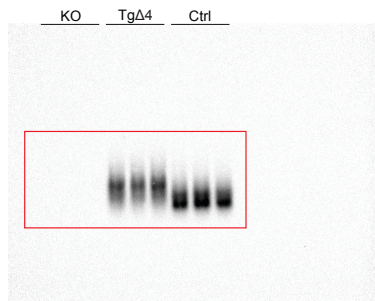
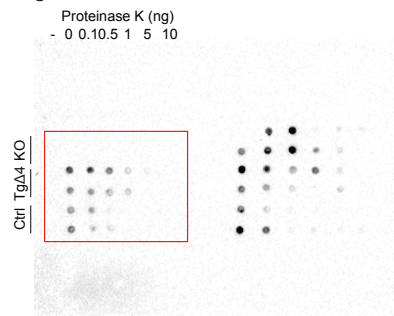
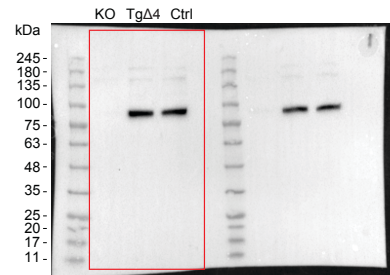


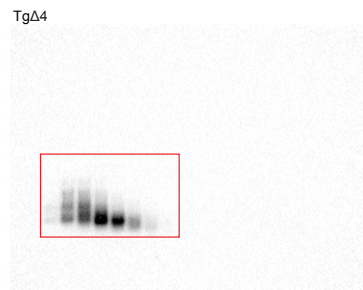
Fig. 4h



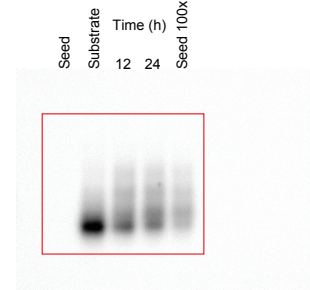
ED Fig. 6f



ED Fig. 6g



ED Fig. 6h



Supplementary Figure 4 | Source data - Uncropped gels and blots from Figure 4 and Extended Data Figure 6.

measurements. Scale bar: 10 μm . **b)** ^1H , ^{15}N -CP-HSQC spectrum of nCPEB4-NTD in the absence (light green) and presence (dark green) of 2 M urea. n=1 spectrum. **c)** DLS intensity and volume fraction of an NMR sample. n=3 measurements. **d)** Time evolution of the multimers followed by DLS (mean $\pm$ SD). Left: diffusion coefficient (D). Center: Z-average. Right: polydispersity index (PDI). n=3 measurements of the same sample. **e)** Micrograph obtained via liquid-phase TEM showing a population of multimers. Scale bar: 50 nm. **f)** Left: chromatogram profile of the normalized absorbance at 280 nm and the molar mass integration of the SEC-MALS experiment. Right: boxplots (line, median; box, quartiles; whiskers, 1.5x interquartile range) for each peak of the chromatogram corresponding to the molar mass distribution and the deconvoluted number of monomers, taking into account that one nCPEB4-NTD molecule weighs 48444.87 Da. Representative of n=2 independent measurements. **g)** Calculation of diffusion coefficients (D) by NMR, fitting the gradient strength-dependent decay in intensity in the absence (light green) and presence (dark green) of 2 M urea. The hydrodynamic diameters (d_h) were calculated by using the values of dioxane as reference. n=1 measurement. **h,i)** Superimposition of the aliphatic region (*h*) and aromatic region (*i*) of the ^1H , ^{13}C -HSQC spectra in the absence (light green) and presence (black) of 4 M urea. The specific residue types (one letter code) and ^1H position within the side chain (Greek letters) are indicated with arrows. n=1 spectrum. **j)** ^1H , ^{13}C -HSQC traces of identified amino acids (one letter code) side chain atoms (as Greek letters) in the absence (light green) and presence (black) of 4 M urea. The inset value indicates the ^{13}C chemical shift of the trace (δC in ppm). n=1 spectrum. **k)** Superimposition of the $\text{N}_\epsilon/\text{H}_\epsilon$ Arg resonances region of the ^1H , ^{15}N -CP-HSQC spectra in the absence (light green) and presence (black) of 4 M urea. n=1 spectrum. **l)** Schematic representation of the monomer-multimer-condensate equilibrium of nCPEB4-NTD in solution, indicating the interactions stabilizing each state.

Supplementary Video legends

Supplementary Video 1. *Ex vivo* live-cell imaging movie of a striatal neuron from mEGFP-CPEB4 knock-in mice pre (0 min) and post (3-27 min) KCl stimulation. Scale bar, 10 μ m.