Supplemental Material

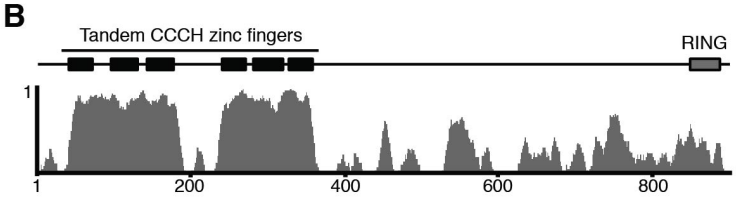
Control of a neuronal morphology program by an RNA-binding zinc finger protein Unkempt

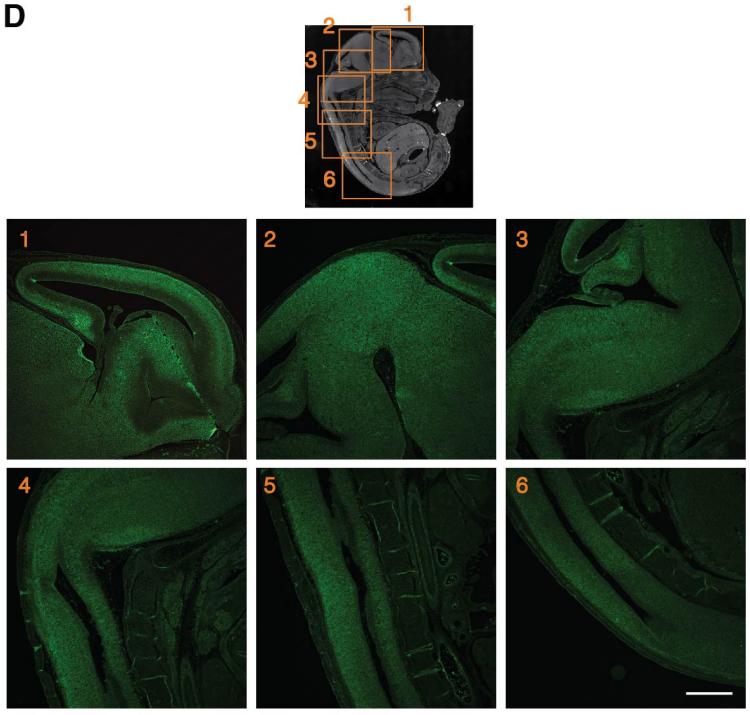
Jernej Murn, Kathi Zarnack, Yawei J. Yang, Omer Durak, Elisabeth A. Murphy, Sihem Cheloufi, Dilenny M. Gonzalez, Marianna Teplova, Tomaž Curk, Johannes Zuber, Dinshaw J. Patel, Jernej Ule, Nicholas M. Luscombe, Li-Huei Tsai, Christopher A. Walsh, Yang Shi

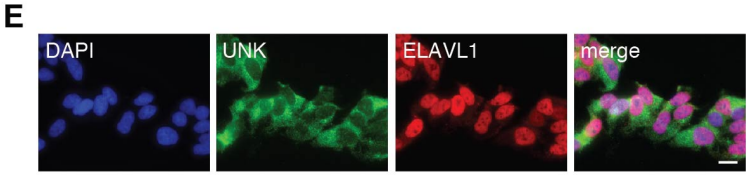
Murn_Supplemental Figure S1

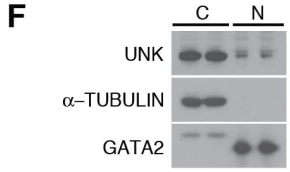
A

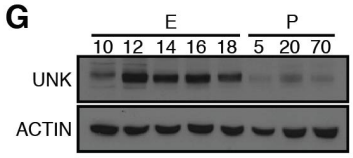
Organism	Unkempt ortholog (RefSeq Accession)	Protein length (AA)
Homo sapiens	NP_001073888.2	810
Mus musculus	NP_766157.1	810
Danio rerio	NP_956530.1	737
Branchiostoma floridae	XP_002595627.1	799
Strongylocentrotus purpuratus	XP_789013.3	825
Caenorhabditis elegans	NP_509350.2	675
Drosophila melanogaster	NP_001247254.1	672
Hydra magnipapillata	XP_002155288.2	701
Nematostella vectensis	XP_001640088.1	666
Amphimedon queenslandica	XP_003384546.1	674

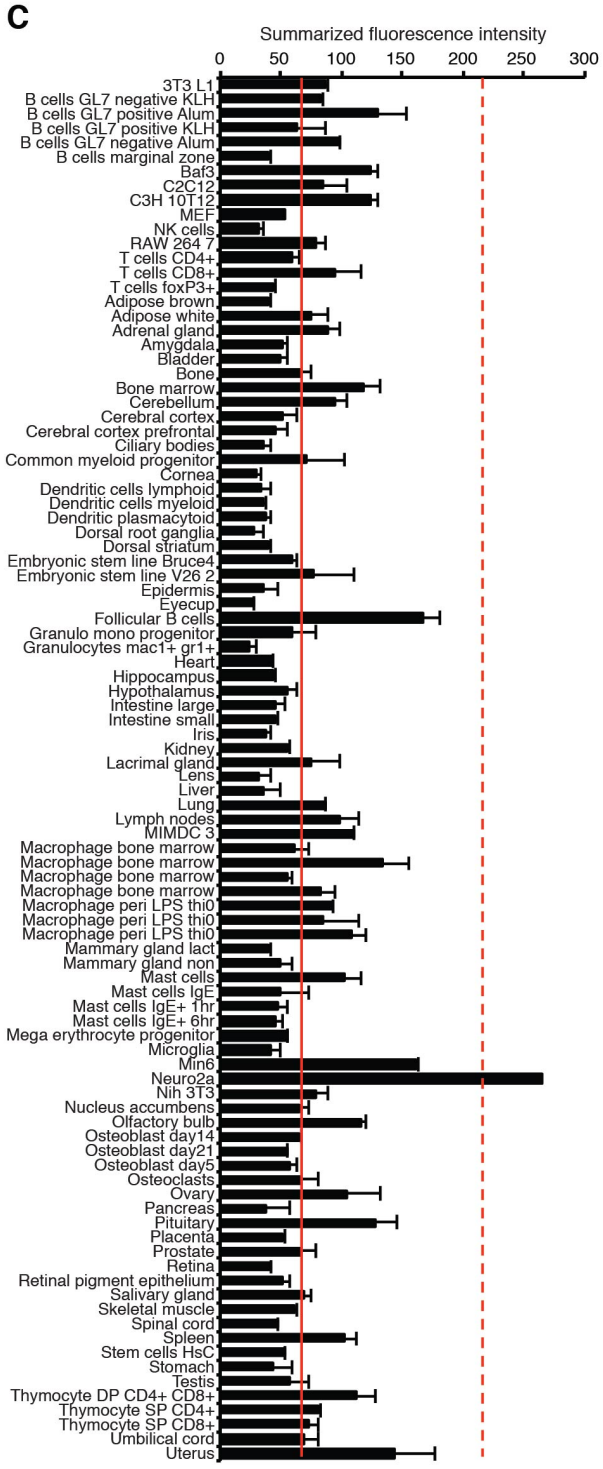


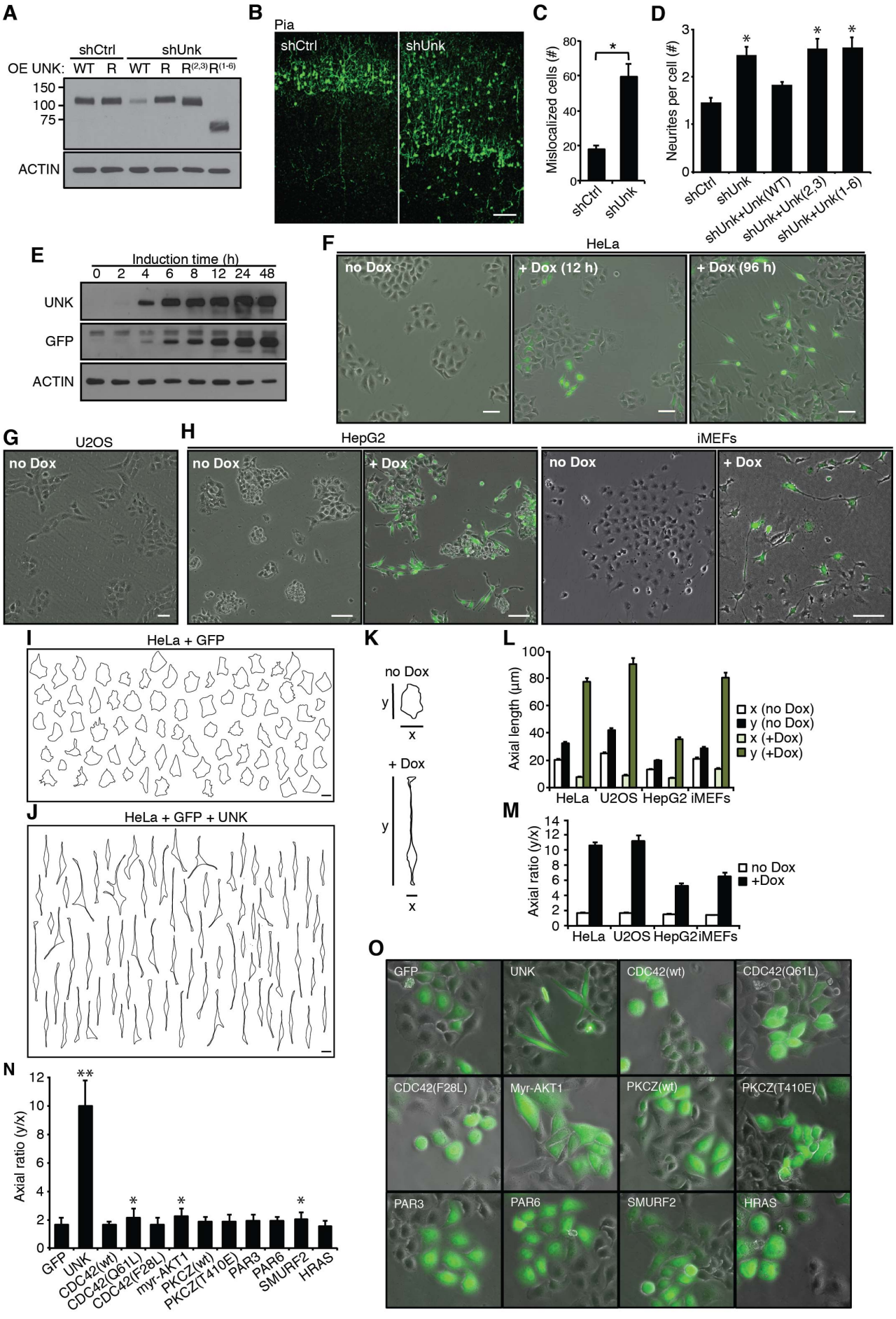




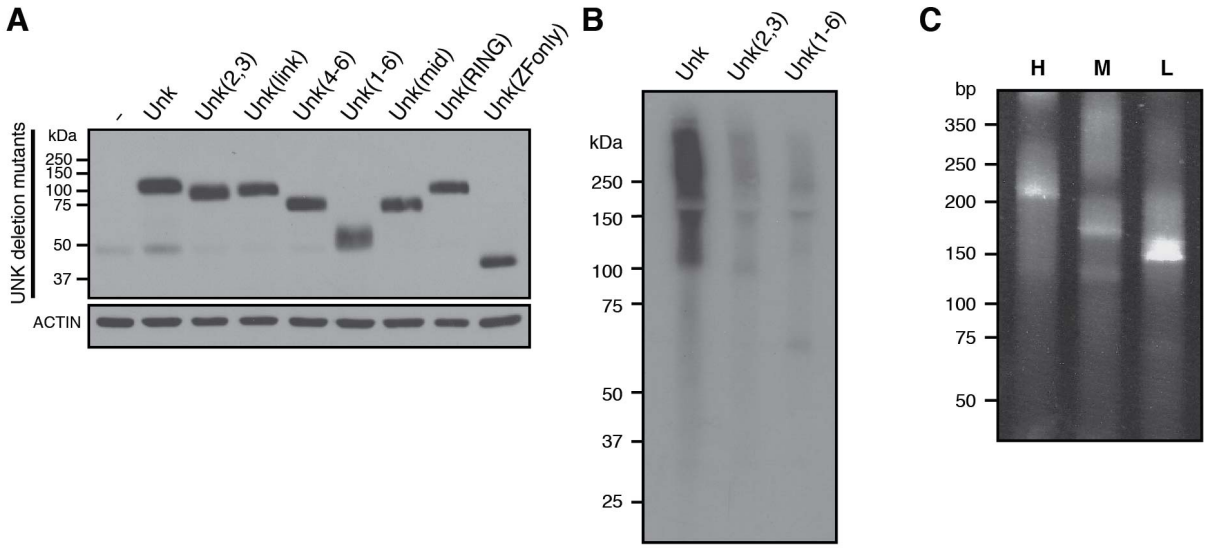




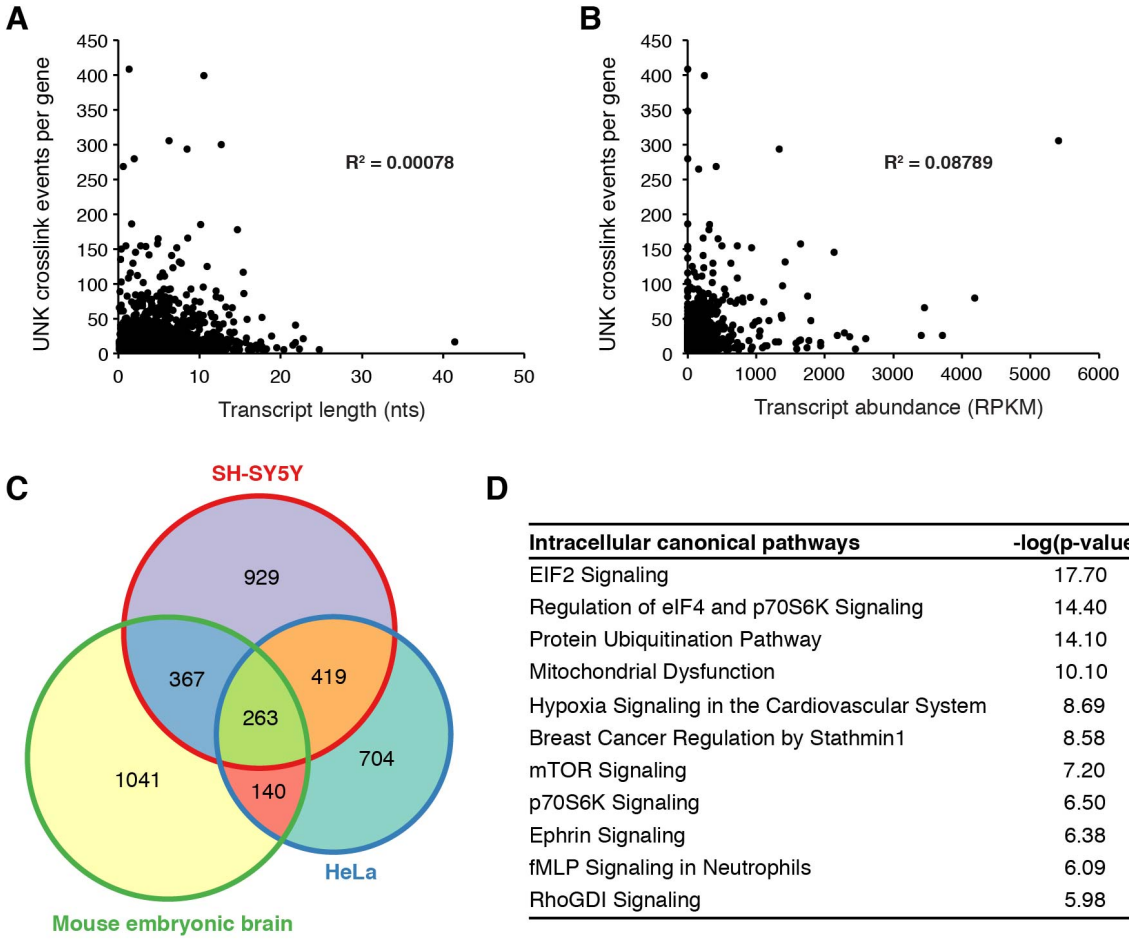



A
shCtrl
shUnk
OE UNK: WT R WT R R(2,3) R(1-6)
150
100
75
ACTIN
B
Pia
shCtrl
shUnk
C
Mislocalized cells (#)
shCtrl
shUnk
D
Neurites per cell (#)
shCtrl
shUnk
shUnk+Unk(WT)
shUnk+Unk(2,3)
shUnk+Unk(1-6)
E
Induction time (h)
0 2 4 6 8 12 24 48
UNK
GFP
ACTIN
F
HeLa
no Dox
+ Dox (12 h)
+ Dox (96 h)
G
U2OS
no Dox
H
HepG2
no Dox
+ Dox
iMEFs
no Dox
+ Dox
I
HeLa + GFP
J
HeLa + GFP + UNK
K
no Dox
y
x
+ Dox
y
x
L
Axial length (µm)
HeLa U2OS HepG2 iMEFs
x (no Dox)
y (no Dox)
x (+Dox)
y (+Dox)
M
Axial ratio (y/x)
HeLa U2OS HepG2 iMEFs
no Dox
+Dox
N
Axial ratio (y/x)
GFP
UNK
CDC42(wt)
CDC42(Q61L)
CDC42(F28L)
myr-AKT1
PKCZ(wt)
PKCZ(T410E)
PAR3
PAR6
SMURF2
HRAS
O
GFP
UNK
CDC42(wt)
CDC42(Q61L)
CDC42(F28L)
Myr-AKT1
PKCZ(wt)
PKCZ(T410E)
PAR3
PAR6
SMURF2
HRAS

Murn_Supplemental Figure S3



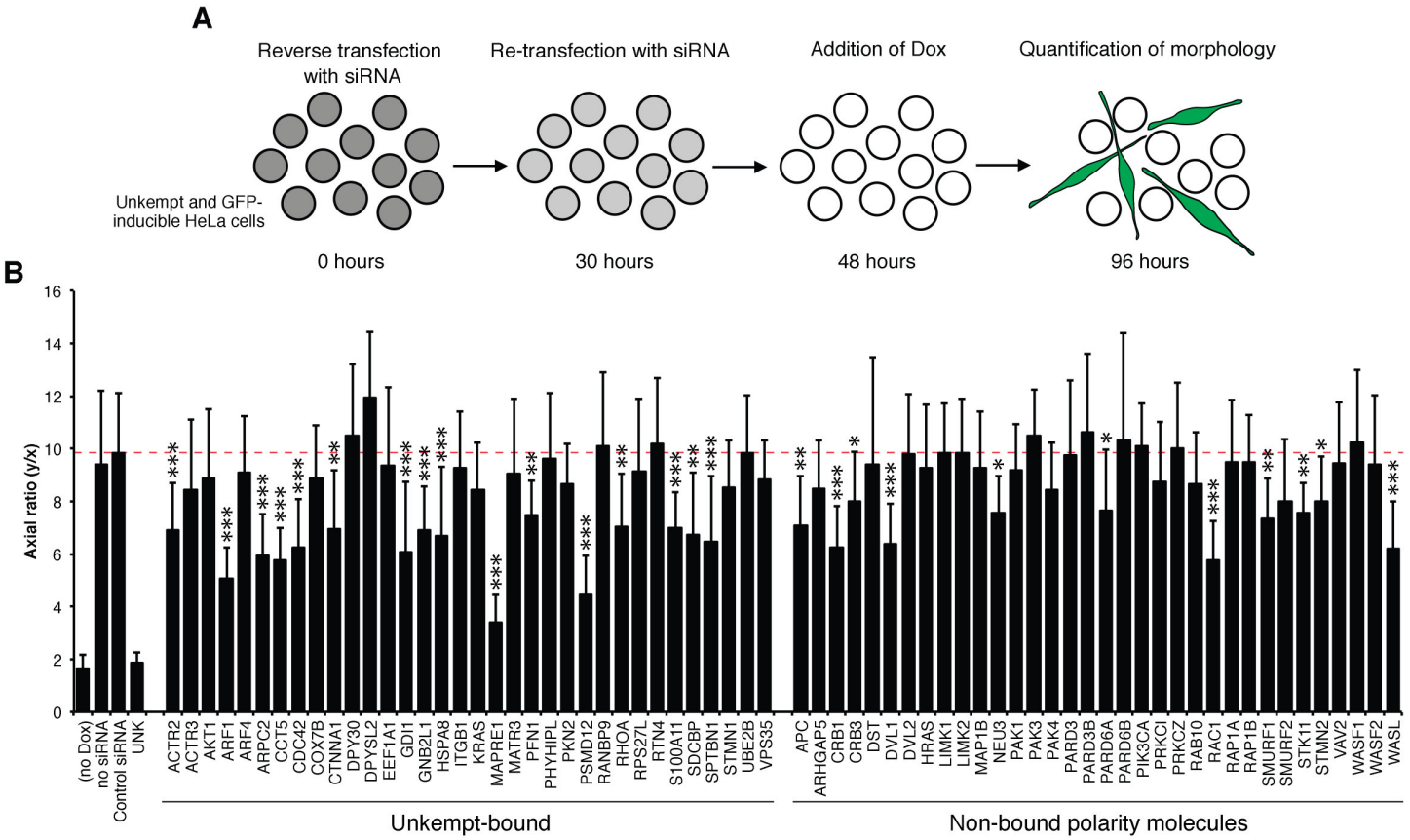
Murn_Supplemental Figure S4



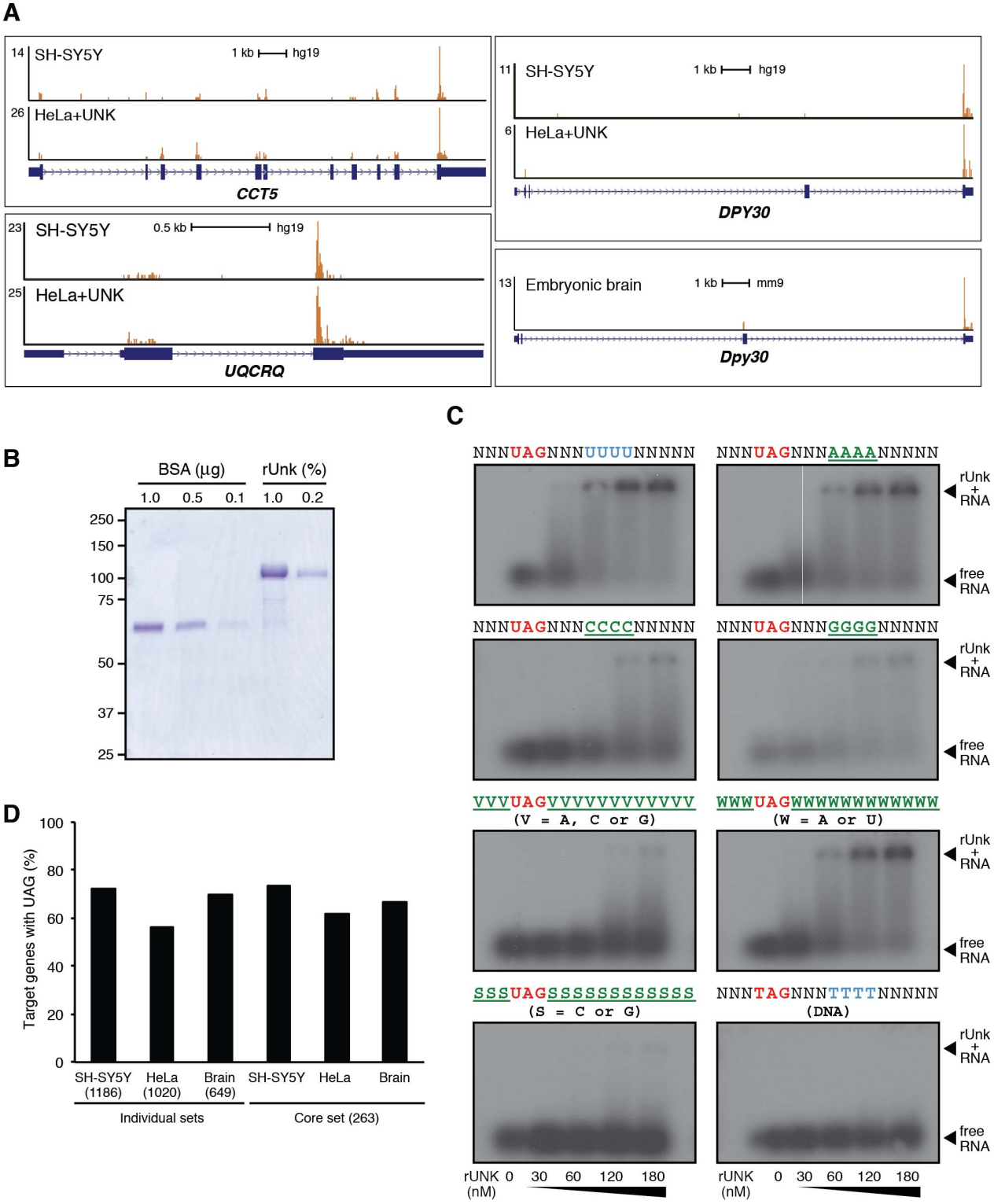
D

Intracellular canonical pathways	-log(p-value)
EIF2 Signaling	17.70
Regulation of eIF4 and p70S6K Signaling	14.40
Protein Ubiquitination Pathway	14.10
Mitochondrial Dysfunction	10.10
Hypoxia Signaling in the Cardiovascular System	8.69
Breast Cancer Regulation by Stathmin1	8.58
mTOR Signaling	7.20
p70S6K Signaling	6.50
Ephrin Signaling	6.38
fMLP Signaling in Neutrophils	6.09
RhoGDI Signaling	5.98

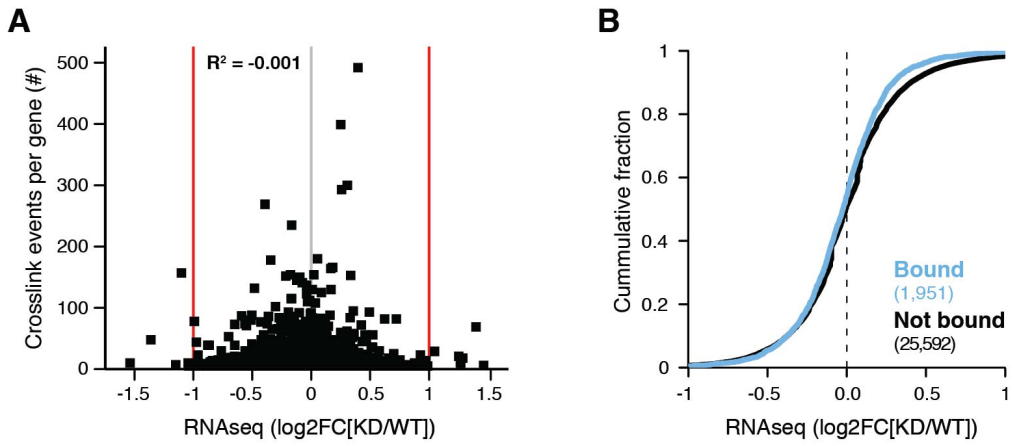
Murn_Supplemental Figure S5



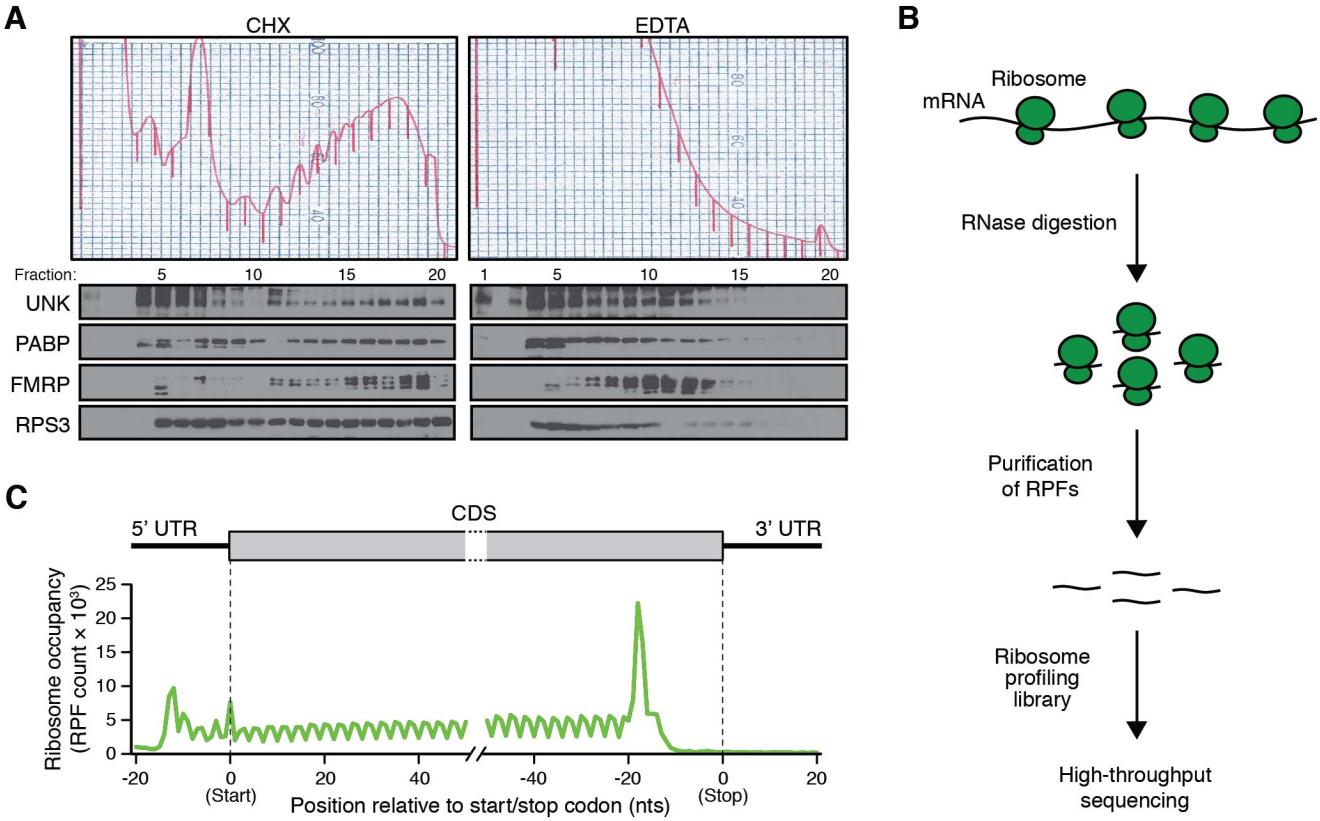
Murn_Supplemental Figure S6



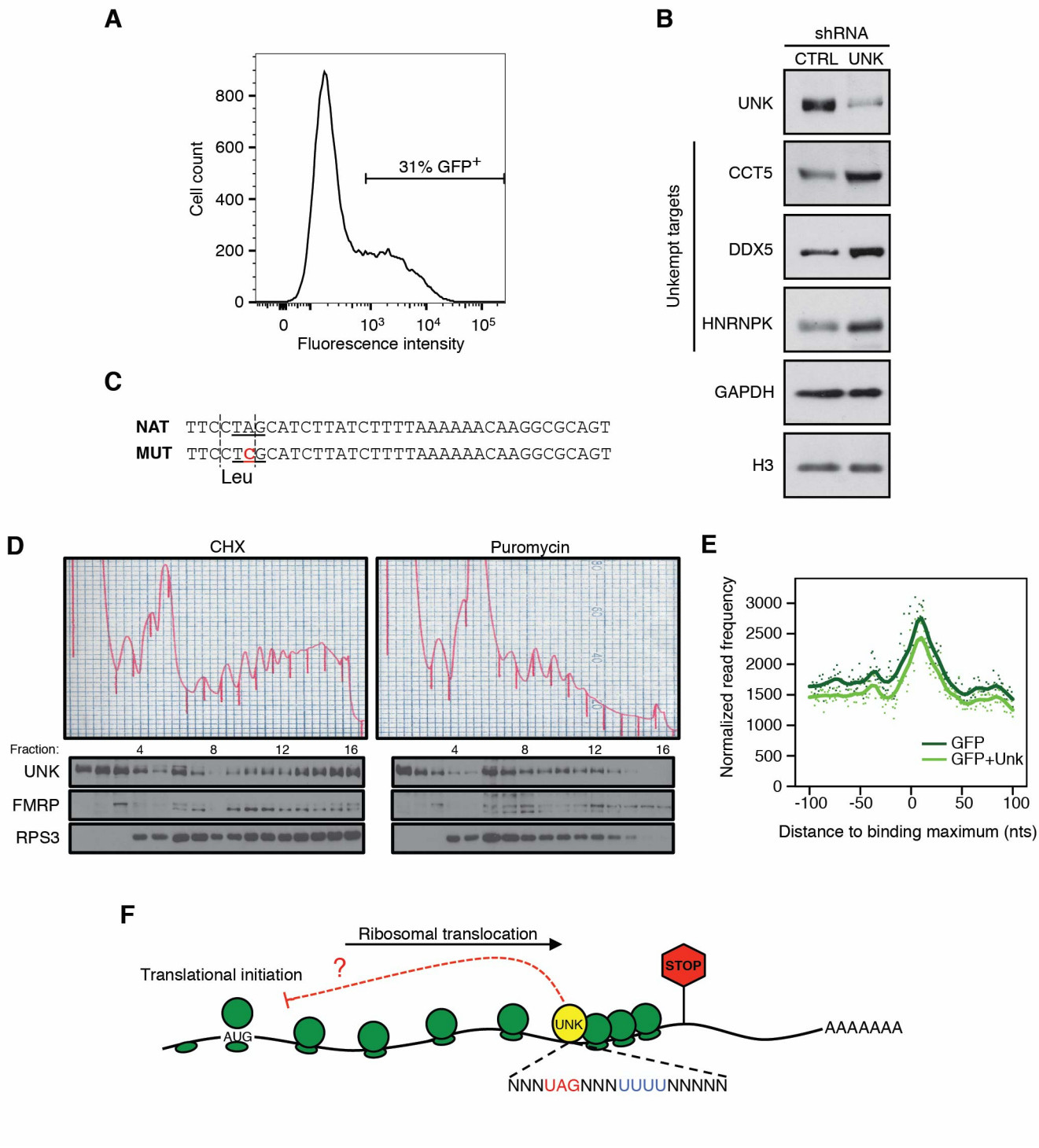
Murn_Supplemental Figure S7



Murn_Supplemental Figure S8



Murn_Supplemental Figure S9



Supplemental Figure Legends

Supplemental Figure S1. Conservation profile of Unkempt protein and expression pattern in the mouse. (*A*) List of all Unkempt orthologs compared in Fig. 1B in addition to the mouse (*Mus musculus*) and zebrafish (*Danio rerio*) orthologs used for the experiments in this study. The source organism for each Unkempt ortholog is listed along with the corresponding RefSeq protein accession number and protein length in amino acids (AA). Orthologs of the human Unkempt protein along with their percent sequence identity to the human ortholog (see Fig. 1B) were determined using BLAST bioinformatics tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). (*B*) Protein sequence similarity across all Unkempt orthologs shown in (*A*). Sketch of a hybrid Unkempt protein with the indicated structural motifs is shown on top of the matching protein sequence similarity profile. Multiple sequence alignment was performed using ClustalW program embedded in the Vector NTI Advance 11.0 package. Horizontal axis indicates the length of a hybrid protein in amino acids. (*C*) Expression of Unkempt in mouse tissues and cell lines. Expression values, presented as summarized fluorescence intensity from Affymetrix microarray analysis, are given for each indicated sample type. The dashed red line indicates three times the median value. The solid red line indicates the mean value. Note that the highest relative expression of Unkempt is seen in the mouse neuroblastoma line Neuro-2a. Data were downloaded from http://biogps.org/gene/217331/ (Wu et al. 2009). (*D*) Expression of Unkempt in E15 mouse embryos. Whole embryo sagittal section stained for Unkempt (top) as shown in Fig. 1E with numbered highlighted regions (orange rectangles) that are shown magnified and labeled correspondingly (bottom). Scale bar, 500 μm. (*E*) Localization of Unkempt (UNK) and a nuclear RNA-binding protein ELAVL1 in SH-SY5Y cells detected by immunofluorescence.

Cells were counterstained with DAPI to label the nuclei. Scale bar, 10 μm. (*F*) Partitioning of Unkempt (UNK) between the cytoplasmic (C) and nuclear (N) fractions. SH-SY5Y cells were fractionated into C and N fractions and the expression of Unkempt, α-Tubulin as a cytoplasmic marker, and GATA-2 as a nuclear marker was detected in each fraction by immunoblotting. Two biological replicates were analyzed. (*G*) Time-dependent expression of Unkempt in whole mouse brain detected by immunoblotting. Numbers indicate days of embryonic (E) and postnatal (P) development.

Supplemental Figure S2. Unkempt-dependent early neuronal morphology *in vivo* and *in vitro*. (*A*) Knockdown efficiency of the shRNA constructs and RNAi-sensitivity of the wild-type (WT) but not the RNAi-resistant wild-type (R) or mutant ($R^{(2,3)}$, $R^{(1\text{-}6)}$) Unkempt proteins. (*B*) Maintained presence of mislocalized neurons postnatally. Cortical sections of mouse embryos electroporated at E14.5 with the indicated constructs and analyzed at P30. Scale bar, 30 μm. (*C*) Quantification of mislocalized GFP-positive cells at P30, as shown in (*B*). Average numbers of cells from ten cortical sections are shown. $*P < 0.0001$. (*D*) Quantification of the morphologies of migrating cortical neurons electroporated at E14.5 with the indicated constructs and analyzed at E19. Shown are average numbers of primary neurites per targeted cell for each condition. Error bars represent S.D. $*P < 0.001$. (*E*,*F*) Time-dependent effect of Unkempt on morphological transformation of HeLa cells. (*E*) Time-dependent expression of Unkempt and GFP upon induction with Dox. Time in hours post addition of Dox to cells in culture is indicated above the immunoblot. (*F*) Changes in morphology of GFP- and Unkempt-inducible cells incubated without Dox (left) or with Dox for 12 (middle) and 96 hours (right). Scale bars, 50 μm. (*G*) Inducible U2OS cells grown in the absence of Dox. Scale bar, 30 μm. (*H*) Cells of diverse

lineages, including the human hepatoma line HepG2 (left) and immortalized mouse embryonic fibroblasts (iMEFs, right) are morphologically transformed by Unkempt. GFP and Unk-inducible HepG2 cells and iMEFs were incubated without or with Dox for 72 hours, as indicated. Scale bars, 100 μm. (*I*,*J*) Individual outlines of GFP-induced (*I*) and GFP and Unkempt-induced (*J*) HeLa cells shown overlaid in Fig. 2K. Ectopic expression of Unkempt induces essentially all cells to adopt an early neuronal-like morphology. The displayed orientations of cells were used for the overlay. Scale bars, 20 μm. (*K*) The morphology of each cell was evaluated by an axial ratio where the length of the absolute longest cellular axis (y) was divided by the length of the longest axis perpendicular to the y-axis (x). (*L*) Lengths of x- and y-axes and (*M*) axial ratios of the indicated cells incubated with or without Dox for 72 hours. At least 50 cells were quantified for each condition. Error bars represent S.E.M. (*N*) HeLa cells engineered for Dox-inducible expression of GFP alone or GFP and either of the indicated proteins (polarity components or Unkempt) were incubated with Dox for 48 hours and quantified by calculating the axial ratios (y/x). Fifty GFP-positive cells of each stable cell line were analyzed and compared to GFP control. Error bars represent S.D. *, $P < 0.01$; **, $P <$ 1e-10, determined using a two-tailed Student's t test. (*O*) Representative images of cells quantified in (*N*).

Supplemental Figure S3. Deletion mutants of Unkempt and the iCLIP experiment. (*A*) Expression of Unkempt mutant proteins in inducible HeLa cells at 72 hours of incubation with Dox. ACTIN was used as a loading control. See also Fig. 3A,B. (*B*) Binding of full-length Unkempt but not its zinc finger-deletion mutants to RNA in HeLa cells. Autoradiogram indicating the intensity of a signal emitted by the labeled material IP-ed using anti-UNK antibodies from lysates of HeLa cells inducibly expressing either wild-type Unkempt or deletion

mutants Unk(2,3) or Unk(1-6) at 24 hours of incubation with Dox. All lysates were treated with RNase I at low concentration. See also Fig. 3A,D, and Supplemental Materials and Methods. (*C*) Analysis of PCR-amplified iCLIP cDNA libraries using gel electrophoresis. RNA recovered from the membrane (Fig. 3D) was reverse transcribed, the obtained cDNA size-purified in three fractions (high [H], medium [M], and low [L]), and, in turn, PCR-amplified. Numbers indicate the size of PCR products in base pairs (bp). See also Supplemental Materials and Methods.

Supplemental Figure S4. Analysis of Unkempt iCLIP results and identities of the targeted mRNAs. (*A*,*B*) Binding of Unkempt to RNA is independent of transcript length or abundance. The collapsed number of crosslink events per gene from replicate iCLIP experiments was correlated with the annotated transcript length (*A*) or transcript abundance (determined by RNA-seq and expressed as RPKM) (*B*). Shown is the analysis of data from SH-SY5Y cells. (*C*) Overlap between Unkempt target transcripts in SH-SY5Y cells, HeLa cells, and mouse embryonic brain. Weighted Venn diagram displaying the total number of unique and common transcripts. See also Supplemental Table S2D. (*D*) Most highly enriched signaling pathways among the 1,186 Unkempt targets in SH-SY5Y cells. The analysis was carried out using the IPA software (Ingenuity Systems, www.ingenuity.com) focusing on intracellular and nervous system-related canonical pathways. Top four enriched pathways are listed for each category, ranked by their significance (p-values; Fisher's exact test). Comprehensive lists of different subsets of canonical pathways enriched in SH-SY5Y cells, HeLa cells, and mouse embryonic brain are given in Supplemental Table S4A-C. See also Supplemental Materials and Methods.

Supplemental Figure S5. Loss- and gain-of-function screens in Unkempt-induced cell morphogenesis. (*A*) Outline of the RNAi screening protocol. The expression of each gene was silenced individually by two rounds of siRNA transfection before induction of GFP and Unkempt and quantitation of cell morphology. (*B*) The impact of gene depletion on cell morphology was quantified by calculating the axial ratios of at least twenty GFP-positive cells for each indicated gene. The results were compared to non-targeting control siRNA. The dashed red line indicates the mean value of the control experiment. *P = < 0.01; ** P < 0.001; *** P < 0.0001, determined using a two-tailed Student's t test. Error bars represent S.D. See also Supplemental Table S5. (*C*) Binding sites in mRNA regions of Unkempt-target genes used for the overexpression screen. The crosslink sites (blue) and the conserved UAG motifs (red) are indicated. Point mutations introduced to the critical UAG motifs to abolish the binding by Unkempt but preserve the encoded amino acid sequence are shown in green. (*D*) Outline of the gain-of-function screening protocol. Plasmids for the expression of RFP and the selected target or non-target gene were transfected into the inducible HeLa cells before induction of GFP and Unkempt with Dox. (*E*) Morphologies of RFP- and GFP-expressing cells were quantified as in (*B*). The dashed red line indicates the mean value of the control experiment. *P < 0.01; ** P < 0.001; *** P < 0.0001, determined using a two-tailed Student's t test. Error bars represent S.D.

Supplemental Figure S6. *In vitro* binding of Unkempt to RNA and global occurrence of the UAG trimer. (*A*) Unkempt contacts its target transcripts commonly at just one dominant binding site. Snapshots from the UCSC genome browser of different Unkempt target genes depict binding positions in the indicated sample types. Note the maintained binding specificity of endogenous Unkempt in SH-SY5Y cells and ectopically expressed Unkempt in HeLa cells.

(*B*) Full-length recombinant Unkempt protein (rUnk). Coomassie-stained denaturing gel indicating the size and purity of mouse rUnk. Bovine serum albumin (BSA) was loaded in parallel in different amounts to aid in quantification of rUnk. See also Supplemental Materials and Methods. (*C*) Additional RNA-EMSA assays probing the requirement for binding of the U-rich motif, the influence of the nucleotides surrounding the UAG trimer, and binding of a DNA sequence harboring both Unkempt-binding motifs. Mutations with respect to the initial RNA sequence (top left) are highlighted in green and underlined. Nanomolar concentrations of rUnk used in all assays are indicated below the panels. See also Fig. 5E. (*D*) Fraction of Unkempt target transcripts containing the UAG motif within the binding sites. Shown is the percentage of all transcripts bound in each cell line (individual sets; total number of Unkempt targets in each cell line is indicated) and the percentage of the 263 core targets harboring the UAG motif in each cell line. See also Supplemental Table S2.

Supplemental Figure S7. Analysis of changes in transcript abundance. mRNA-binding by Unkempt has no significant effect on steady-state levels of the majority of transcripts. (*A*) Correlation of changes in the abundance of Unkempt target transcripts determined by RNA-seq analysis comparing wild-type (WT) and knockdown (KD) SH-SY5Y cells with total number of crosslink events per Unkempt target transcript. Vertical lines indicate no change (gray) or a two-fold change in transcript abundance (red). (*B*) Cumulative distribution of RNA-seq changes upon depletion of Unkempt in SH-SY5Y cells. Plotted are distributions for Unkempt-target (Bound), and non-target genes (Not bound). The number of genes in each category is indicated in parentheses.

Supplemental Figure S8. Association of Unkempt with polyribosomes and the ribosome profiling experiment. (*A*) Sucrose gradient fractionation of mouse embryonic brain lysates followed by immunoblot analysis of individual fractions for the indicated proteins. Cells were harvested either in the presence of cycloheximide (CHX) or EDTA to stabilize or dissociate ribosomes, respectively. The poly(A) binding protein (PABP), FMRP, ribosomal protein S3 (RPS3), but not ACTIN, associate with polyribosomes and were used as controls. See also Fig. 5A. (*B*) Outline of the ribosome profiling experiment. Ribosome-protected mRNA fragments (RPFs) were isolated from GFP- or GFP and Unkempt-inducible HeLa cells at 18 hours of treatment with Dox and converted into cDNA libraries that were in turn subjected to high-throughput sequencing. (*C*) Features of translation captured by ribosome profiling in HeLa cells. Shown is the ribosome occupancy around the beginning and end of the CDS of a composite mRNA, combining data from all quantified genes. Nucleotide positions were binned in 5-nucleotide windows.

Supplemental Figure S9. Unkempt inhibits translation of its bound mRNAs. (*A*) Percentage of cells in a total population of inducible HeLa cells that successfully express the transgene upon induction by Dox. GFP and Unkempt-inducible HeLa cells were treated with Dox for 48 hours and examined by FACS analysis for the expression of GFP. (*B*) Increased abundance of Unkempt target-encoded proteins upon depletion of Unkempt. SH-SY5Y cells stably expressing either non-targeting control (CTRL) or Unkempt-targeting (UNK) shRNA were analyzed by immunoblotting for the expression of the indicated proteins. Two non-targeted proteins, GAPDH and histone H3, served as loading controls. (*C*) DNA sequence corresponding to the Unkempt-binding site located in the mouse *Dpy30* transcript used to extend in-frame the 3' end of the

firefly luciferase protein. The native sequence (NAT) extension differs from the mutant (MUT) version in a single nucleotide within the critical UAG motif (underlined) such that the amino acid (Leu) encoded by the affected codon (vertical dashed lines) is preserved. See also Fig. 6I,J. (*D*) Sucrose gradient fractionation of SH-SY5Y cell lysates before (CHX, left) or after 30 minutes of incubation in the presence puromycin (right). Western blots were used to compare the distribution of UNK, FMRP, and RPS3 (bottom). Absorbance traces of total RNA distribution and gradient fractions are shown (top panels). (*E*) Unkempt binds mRNAs immediately upstream of endogenously paused ribosomes. The plot depicts a global enrichment of RPFs in the vicinity of Unkempt binding sites in cells inducibly expressing either GFP alone (dark green) or GFP and Unkempt (light green). The correlative analysis within 100 nucleotides from Unkempt binding site maximum is shown. (*F*) Model for the action of Unkempt as a translational inhibitor. Unkempt (depicted in yellow) binds to its recognition element consisting of a UAG motif and a U/A-rich motif (represented as WWW) within the CDS region, preferentially immediately upstream of the endogenously paused ribosomes (depicted in green) and with high frequency in the proximity of the stop codon (red). Red dashed line indicates a hypothetical interference of Unkempt with translation initiation (AUG, start codon).

Supplemental Table Legends

Supplemental Table S1. Mapping statistics of iCLIP sequence reads. cDNA libraries were prepared from Unkempt-bound RNA in four, three, and five independent iCLIP experiments performed with SH-SY5Y cells, HeLa cells, and mouse embryonic brain, respectively, and sequenced using the HiSeq 2000 system (Illumina). In parallel, control iCLIP libraries were prepared from sample aliquots of SH-SY5Y cells (in triplicates) and embryonic brain (in duplicates) using normal rabbit IgG antibodies (sc-2027). Replicate-specific and total numbers of sequencing reads belonging to the indicated read categories are listed for each sample type. Reads from SH-SY5Y and HeLa cells were mapped to the hg19 build of the human genome, and reads from the embryonic brain were mapped to the mm9 build of the mouse genome. See also Supplemental Materials and Methods.

Supplemental Table S2. Unkempt target genes in SH-SY5Y cells, HeLa cells, and mouse embryonic brain. (*A-C*) Unkempt targets in individual sample types. The identity of each target is described by the approved unique Gene Symbol (column B), the corresponding Ensembl Gene ID (column C), and human (hg19) or mouse (mm9) genomic coordinates detailing the identity of the chromosome (column D), base-pair position (start and end of the gene on the chromosome; columns E and F), and the orientation of the DNA strand harboring the gene (column G). The length of the longest encoded mature transcript (cDNA_length) is given in column H. iCLIP-related information includes the total number of crosslink events within the strongest binding site on a target (column I), the number of replicates in which the strongest (column J) or any (column

K) binding site is observed, the number of UAG motifs found within the strongest binding site (column L), and whether the UAG motif is found in any binding site on a target transcript (column M). Targets are ranked by the maximum binding site iCLIP tag count in descending order (column A). (*D*) Core subset of Unkempt targets. The 263 targets commonly bound in all three cell types were defined with relaxed stringency regarding the reproducibility of binding in each sample type. Human orthologs of targets bound in mouse brain were used to define the overlap with Unkempt targets in SH-SY5Y and HeLa cells. The annotated target information includes gene and transcript categories as listed in (*A-C*) for human and mouse Unkempt orthologs along with the iCLIP-related data as in (*A-C*) for each sample type. Targets are ranked by reproducibility of binding in SH-SY5Y cells in descending order.

Supplemental Table S3. Gene Ontology (GO) analysis of Unkempt targets. Functional annotation of each target set was carried out with DAVID Bioinformatics Resources package using lists of background genes as described in Supplemental Materials and Methods. Top ten statistically over-represented GO categories (Biological Process (BP), level 4; column B) in each target set (column A) are shown ranked by descending p-values (column D). Target count (column C) and target identities (column E) are listed for each GO category. See also Supplemental Materials and Methods.

Supplemental Table S4. Pathway analysis of Unkempt targets. Ingenuity Pathway Analysis software was used to identify over-represented canonical pathways (column A) in SH-SY5Y cells (A), HeLa cells (B), and mouse embryonic brain (C). Pathways are ranked by descending p-

values (column B) and target molecules belonging to each pathway are listed in column C. Listed are all pathways with –log(p-value) > 3. See also Supplemental Materials and Methods.

Supplemental Table S5. Genes interrogated by the RNAi screen in Unkempt-induced cell morphogenesis. Listed are the 70 genes targeted by siRNAs used in the loss-of-function screen (see Supplemental Fig. S5B). Column E indicates whether a gene-encoded transcript is (Y) or is not (N) targeted by Unkempt in HeLa cells. Genes that have previously been associated with various aspects of cell morphology are designated (Y) in column F. Column G lists sequences targeted by each siRNA. 'Multiple' indicates cases where esiRNAs or a pool of siRNAs was used. Column H lists the efficiency of knockdown (KD) determined by RT-qPCR using primers listed in Supplemental Materials and Methods. The source information for each RNAi reagent is provided in columns I and J.

Supplemental Material and Methods

Cell culture

HeLa, U2OS, iMEFs, HepG2, and SH-SY5Y cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, and 100 ug/mL streptomycin at 37 °C and 5% CO_2. HeLa S3 were grown in Minimum Essential Medium (MEM; Sigma M0518) supplemented with 5% fetal bovine serum, 2 g/L sodium bicarbonate, 2 mM L-glutamine, 100 U/mL penicillin, and 100 ug/mL streptomycin in suspension (spinner) culture at 37 °C.

Plasmid constructs

Silencing vectors and RNAi-resistant Unkempt

Lentiviral vectors (pLKO.1) harboring shRNAs targeting human Unkempt were obtained from Sigma (SHCLNG-NM_001080419). To knockdown Unkempt in mouse neurons, we modified the pSilencer1.0 GFP vector (Shen et al. 2010) by inserting between the ApaI and EcoRI sites the shRNAs targeting mouse Unkempt mRNA. The targeted sequences ($shUnk^W$: CCCACTCTTCGTGCAGCACAA, shUnk: GGTATCACCTTCGTTACTA) were identified by the Broad Institute RNAi Consortium.

RNAi-resistant mouse Unkempt proteins (Unk(WT), Unk(2,3), and Unk(1-6)) were generated by site-directed mutagenesis to mutate the shUnk-targeted sequence into 5' GATACCATTTGAGATATTA 3'. The created N-terminally tagged (Flag-HA) RNAi-resistant Unkempt was cloned into the pCAGGS-IRES-RFP vector (gift from Francois Guillemot, MRC NIMR, UK) using ClaI and NheI restriction sites.

Doxycycline-inducible vectors and cell lines

To create Dox-inducible cell lines, we first transduced HeLa, U2OS, HepG2, and iMEF cells with MSCV-rtTA3-IRES-EcoR-PGK-Neo (pRIEN), a retrovirus we constructed in the MSCV backbone (Clontech) to co-express an improved reverse transactivator (rtTA3), the ecotropic receptor, and a neomycin resistance marker. Following G418 selection, stably rtTA3-expressing cells were subsequently infected ecotropically with a second retrovirus expressing a puromycin resistance gene and a TREtight-driven transcript encoding GFP alone (pTt-IGPP) or GFP and either full-length mouse Unkempt protein, any of the mouse Unkempt-deletion mutants, or any of the polarity molecules analyzed in Supplemental Fig. S2N, O (pTtight-X-IGPP, where X is an ortholog or mutant of Unkempt, or a polarity molecule; see also Fig. 3A), which we constructed in a self-inactivating retroviral backbone (pQCXIX, Clontech). The ORFs encoding polarity components were subcloned into the pTt-IGPP vector from the following source plasmids obtained from Addgene: pK-myc-Par3b (#19388), pK-myc-Par6C (#15474), FLAG.PKCzeta.T410E (#10804), FLAG.PKCzeta (#10799), pCMV5B-Flag-Smurf2_C716A (#11747), pLNCX_myr_AKT1 (#17245), pEXV3-Myc-H-RAS_V12 (#39528), pRK5myc_Cdc42_L61 (#15906), and pRK5myc_Cdc42_wt (#15905). To induce transgene expression, double-selected cells were treated with doxycycline (Sigma-Aldrich) at 1 ug/ml.

Plasmids for inducible expression of Unkempt-mutant proteins

Unkempt-deletion mutants were generated by mutagenesis of the pTtight-Flag-HA-Unk-IGPP plasmid according to the following procedure. The plasmid was PCR-amplified with the mutating primer pairs (listed below) using PfuTurbo DNA Polymerase (Agilent Technologies) for 18 cycles according to the QuickChange XL Site-Directed Mutagenesis Kit manual (Agilent Technologies). One µl of 10 mM dNTP and 0.5 µl of T4 DNA polymerase (New England BioLabs) were added directly to the PCR reaction, which was then mixed by pipetting and

incubated for 10 minutes at 37 °C. The reaction was purified using QIAquick PCR Purification Kit (Qiagen) and eluted in 35 µl of water. PCR products were simultaneously 5'-phosphorylated and ligated by adding 4 µl of 10 × T4 DNA Ligase Buffer, 1 µl of T4 PNK, and 1 µl of T4 DNA Ligase (all New England BioLabs), mixing, and incubating at 16 °C overnight. Following the ligation, 1 µl of DpnI enzyme (New England BioLabs) was added, the reaction was mixed and incubated for 1 hour at 37 °C. The reaction was transformed into chemically competent bacteria and the mutated clones were selected according to standard procedures. The following oligonucleotide primer pairs were used to create Unkempt-deletion mutants:

Unk(2,3); F: CTCCGCTCCCCTGTTTACGA, R: TGTAGCCTCGTCATACTTCGTG

Unk(link); F: ATGAGGCCCGTGGGCAAAAGC, R: CTCCAGGCCATGGAGGCCTTGCAGA

Unk(4-6); F: TTAAGTGATGATGTGCAGCCTTCC, R: AGCCCCAGCTGACTGGCCTT

Unk(1-6); F: TTAAGTGATGATGTGCAGCC, R: CATTGGATATCTGGTGAAGG

Unk(mid); F: TTGTTGCAGAGCTCCGCGCCTGTGA, R: GCCAGGGGCAAGGCTGGACTC

Unk(RING); F: CCCAGCCGGGCCCATGCCCT, R: CTGCATGTGGAACACAGCCT

Unk(ZFonly); F: CCCAGCCGGGCCCATGCCCT, R: GCCAGGGGCAAGGCTGGACTC

Hybrid luciferase constructs

To extend the 3' end of the sequence encoding the firefly luciferase, the pGL4.35[luc2P/9XGAL4UAS/Hygro] Vector (Promega) was first mutated according to the protocol described in QuickChange II Site-Directed Mutagenesis Kit manual (Agilent Technologies) using an oligonucleotide primer pair (forward: CCAGCGCCAGGATCAACGCGTAAGGCCGCGACTCTAGAG; reverse:

CTCTAGAGTCGCGGCCTTACGCGTTGATCCTGGCGCTGG) to introduce a unique MluI restriction site immediately upstream of the stop codon. This mutation converted the most C-terminal amino acid residue of the firefly luciferase protein, Val, into Ala. The introduced MluI site was used to insert in-frame the oligonucleotide sequences corresponding to Unkempt-binding sites located within target genes *S100A11* and *Dpy30*. The inserted sequences are shown in Fig. 6H and Supplemental Fig. S9C.

Dual luciferase assays with RT-qPCR analysis

Dual luciferase assays were carried out by co-transfecting 500 ng of each hybrid luciferase construct and 20 ng of the control pRL-TK Vector (Promega) into shCTRL- or shUNK-expressing SH-SY5Y cells using Lipofectamine 2000 (Life Technologies). Twelve hours after transfection, cells were harvested and processed using Dual-Luciferase Reporter Assay System according to the manufacturer's instructions (Promega). The emitted luminescence was detected using SpectraMax L Luminescence Microplate Reader (Molecular Devices).

RNA was extracted from samples equivalent to those used for the Dual luciferase assays using Trizol Reagent according to the manufacturer's instructions (Life Technologies). Contaminant DNA was removed from the isolated RNA by using the TURBO DNA-*free* Kit (Life Technologies). cDNA was prepared from equal amounts of RNA using Random Hexamers and Superscript III Reverse Transcriptase (both Life Technologies) according to the manufacturer's instructions. qPCR was performed using LightCycler 480 SYBR Green I Master and the oligonucleotide primers listed below to amplify the cDNA on the LightCycler 480 Instrument II (both Roche) at the annealing temperature of 63 °C. Relative hybrid firefly

luciferase mRNA levels were normalized to relative expression levels of the *RPS18* gene that was used as an internal control. Oligonucleotide primer pairs used for RT-qPCR were:

Firefly luciferase; F: GAGCTATTCTTGCGCAGCTT, R: CCTCACCTACCTCCTTGCTG

RPS18; F: GATGGGCGGCGGAAAATAG, R: GCGTGGATTCTGCATAATGGT

SDS-PAGE and Western Blotting

Whole-cell lysates, cytoplasmic and nuclear fractions, or TCA-precipitated proteins from each fraction of the sucrose gradient (400 µl) were run on 10% or 15% SDS-polyacrylamide gels and transferred to supported nitrocellulose membrane (Bio-Rad) by standard methods. Membranes were blocked for 1 hour in 5% non-fat dry milk in 1 x TBS with 0.1% Tween-20 (TBST), rinsed, and incubated with primary antibody diluted in 3% BSA in TBST overnight at 4°C. The following primary antibodies were used: anti-UNK (HPA023636, Sigma-Aldrich), anti-β-ACTIN-Peroxidase (A3854, Sigma-Aldrich), anti-α-Tubulin (sc-5286), anti-GATA2 (sc-9008), anti-RPS3 (sc-376008), anti-CCT5 (sc-374554), anti-DDX5 (sc-81350), anti-HNRNPK (sc-28380), anti-GFP (sc-8334), anti-GAPDH (sc-32233; all Santa Cruz Biotechnology), anti-FMRP (ab17722), anti-PABP (ab21060), anti-H3 (ab1791; all Abcam), and anti-HA (MMS-101P, Covance). Blots were washed in TBST, incubated with HRP-conjugated secondary antibodies in 5% milk in TBST for 1 hour (except for anti-β-ACTIN-Peroxidase antibody), and washed again. HRP signal was detected by Enhanced ChemiLuminescence (Perkin Elmer).

Immunocytochemistry and immunohistochemistry

Cultured cells were fixed in 4% paraformaldehyde for 15 minutes at room temperature, permeabilized in 0.5% Triton X-100, blocked in 5% goat serum, and probed with primary

antibodies by incubating at 4°C for 24 hours. Brain sections were prepared by drop-fixing embryonic brains overnight in 4% formaldehyde and then immersing them in 30% sucrose/PBS solution at 4°C. Brains were embedded in O.C.T. compound (Electron Microscopy Sciences) and sliced into 20 - 30 μm sections using cryostat. Cryosections were rehydrated with PBS and blocked for 1 – 2 hours with blocking solution (1×PBS, 10% donkey normal serum, 0.3% Triton-X100). Following blocking, the cryosections were incubated with anti-GFP antibody (GFP-1020, Aves Labs) overnight at 4ºC. Incubation with secondary antibodies was performed for 1 hour at room temperature. Finally, cryosections were mounted using ProLong Gold Antifade Reagent (Invitrogen). Brain sections of P30 mice were sliced into 50 - 200 μm sections using vibratome, permeabilized in a mix of 0.04% Tween-20 and 0.1% Triton X-100, blocked, and incubated with anti-GFP antibody as above. Whole mouse E15 embryo sagittal sections (Zyagen) were processed according to the standard protocol (Abcam). Cells or whole mouse embryo sections were immunostained with antibodies recognizing GFP (ab290, Abcam), UNK (HPA023636 or HPA027962, Sigma-Aldrich), ELAVL1 (sc-5261, Santa Cruz Biotechnology), TUJ1 (MMS-435P, Covance), NeuN (MAB337, Millipore), and Cux1 (sc-13024, Santa Cruz Biotechnology). Samples were incubated with fluorochrome-conjugated secondary antibodies and counterstained with DAPI or Hoechst 33258 (Sigma-Aldrich). Live cell images were taken with Eclipse TE2000-U inverted microscope (Nikon), and images of immunostained tissue specimens were taken with LSM510 or LSM700 confocal microscopes (Zeiss). Z-stack images of GFP-positive neurons were taken with 0.5 to 1 μm steps. Sections of mouse brain were morphologically matched before comparing experimental to control conditions.

Mice and *in utero* electroporation

Swiss Webster pregnant mice were purchased from Taconic (Hudson, NY) and allowed to recover before the surgery. *In utero* electroporations were performed as described (Xie et al. 2007). Briefly, pregnant dams at E14.5 were anesthetized by intraperitoneal injection of Ketamine (1%) and Xylazine (2 mg/ml), the uterine horns were exposed, and the plasmids stained with Fast Green (Sigma) were microinjected into the lateral ventricles of embryos. Five pulses of current (50 ms on / 950 ms off) were delivered across the head of the embryos. In the electroporated DNA mixture, the shRNA plasmid concentration was 1.5 to 2-fold higher than that of the rescue vector. The numbers of animals and litters on which the data in Fig. 2A,B are based were as follows: shCtrl (8 animals, 4 litters), shUnk (9 animals, 5 litters), shUnk+Unk(WT) (7 animals, 4 litters), shUnk+Unk(2,3) (4 animals, 2 litters), and shUnk+Unk(1-6) (6 animals, 3 litters). All animal surgical experiments were carried out by adhering to the protocols approved by the IACUCs of the Harvard Medical School and Boston Children's Hospital or The Institutional Animal Care and Use Committee of Massachusetts Institute of Technology.

Neuron culture and plasmid transfection

Primary cortical neurons were isolated from E14.5 mouse embryos and dissociated using the Papain Dissociation System (Worthington Biochemical Corporation). Cells were grown on Poly-L-Ornithine coated plates (Sigma) in Neurobasal Medium (Invitrogen) with 0.6% glucose, B-27 Supplement (Invitrogen), N-2 Supplement (Invitrogen), Penicillin-Streptomycin (Invitrogen), L-glutamine, and transfected with pSilencer RNAi plasmids one hour after plating using the CalPhos transfection kit (Clontech). All experiments were analyzed in a double-blind manner using an unpaired t-test.

Overlays of cellular outlines for morphological comparisons

After 72 hours of incubation with Dox, the inducible HeLa, U2OS, and HepG2 cells were imaged and the outlines of GFP-positive cells were drawn using the Adobe Illustrator software (Adobe). The outlines of different cells were aligned by the absolute longest cellular process or the most acuminate part of each outline, and overlaid by superimposing the cell bodies.

RNAi and overexpression screens in inducible HeLa cells

For the RNAi screen, GFP- and Unkempt-inducible HeLa cells were reverse-transfected at 3.5×10^4 cells per well in 12-well plates using DharmaFECT 1 transfection reagent (GE Healthcare) according to the manufacturer's instructions at a final concentration of 20 nM siRNA. Thirty hours later, cells were re-transfected and incubated for a further 18 hours. Cells were then treated with Dox for 48 hours and the morphologies of at least 20 GFP-positive cells were quantified for each targeted gene by calculating their axial ratios (see Supplemental Fig. S6A,B). Our measured knockdown efficiencies were largely in agreement with those reported by Qiagen for HeLa cells, and were determined at 48 hours of treatment with Dox in duplicate experiments by RT-qPCR using the following primer pairs:

ACTR2; F: CATCACGGTTGGAACGAGAAC, R: GTGGGTCTTCAATGCGGATCT

ACTR3; F: TGCAGGCTGTTCTTGCCTTAG, R: CACATACCCTTCAGCCACAGG

APC; F: TCTTCAGTGCCTCAACTTGC, R: GGAGACAGAATGGAGGTGCT

ARPC2; F: GGAACCTCCTCTGGAGCTGAA, R: CGTGTGGATCAGGTTGATGGT

CCT5; F: AATGCCGTCCTCACTGTAGCA, R: CCACAATCACGCCCTTAATCA

CDC42; F: TTGTTTGTTGTGTAGGATATCAGGA, R: CCCGGTGGAGAAGCTGAG

DPY30; F: GACAAACGCAGGTTGCAGAAA, R: CGAGTTGGCAAAGACTGGAGA

DPYSL2; F: TGAAGGATCACGGGGTAAATTCC, R:
AATATCCCGGATCACACTCAGT

GAPDH; F: CCATGAGAAGTATGACAACAGCC, R: GGGTGCTAAGCAGTTGGTG

HSPA8; F: TGATGCCGCAAAGAATCAAGT, R: CCACCATAAAGGGCCAATGTT

LIMK1; F: AGTACTGGTGCGACAGGGAG, R: GCCAGAGGATCTATGATGGC

LIMK2; F: AGAGCTTCCCATCCTTCTCA, R: GGACTGTCAACGAAACCTGG

MAP1B; F: CCATCTCCCTTTGACATCTTGG, R: CGCCATTGATGAGCATATTGAA

MAPRE1; F: CAGGGAACAGCATGTCCATA, R: CCTGGATCAATGAGTCTCTGC

PAK1; F: GTCCTCTTTGGGTTCGCTGTA, R: CCAGTGACCACAAAACGACTAT

PARD3; F: AATGATGGCGACCTTCGAAAT, R: CAGGAACCACATGGAACCAAA

PARD3B; F: CCTGCTGATACCCAGCCAAG, R: TCTGCGTGGAATCTGGAACAA

PARD6A; F: GATGGTTGCCAACAGCCATAA, R: AGAGGGAGGACCTGTCAAACG

PARD6B; F: TTGGAGCTGAATTTCGTCGGT, R: AGCCTACCAAAACGTCAACATT

PFN1; F: GATCACCGAACATTTCTGGC, R: AAACGTTCGTCAACATCACG

RAB10; F: TGCAAGGGAGCATGGTATTAGG, R: TGGGCTCTTTTACAGGGGTCTT

RAC1; F: ATGTCCGTGCAAAGTGGTATC, R: CTCGGATCGCTTCGTCAAACA

RAP1B; F: AGCAAGACAATGGAACAACTGT, R: TGCCGCACTAGGTCATAAAAG

RHOA; F: GATTGGCGCTTTTGGGTACAT, R: AGCAGCTCTCGTAGCCATTTC

RPS27L; F: ATCCGTCCTTGGAAGAGGAAAA, R: CTGTCTGAGCATGGCTGAAAAC

S100A11; F: TCTCCAGCCCTACAGAGACT, R: GGCAGCTAGTTCTGTATTCA

STK11; F: GACCTGCTGAAAGGGATGCTT, R: CTTCAGCCGGAGGATGTTTCT

STMN1; F: AAATGGCTGCCAAACTGGAAC, R: TCAGTCTCGTCAGCAGGGTCT

STMN2; F: GCTCTTGCTTTTACCCGGAAC, R: AGGCACGTTTGTTGATTTGCT

WASF1; F: CCAAGAGCACCTCATGACAGG, R: AAATGAGAGGCTGGGCCATTA

WASL; F: TGTCCTGCTCTGGACGAGAT, R: GTTGGTGGTGTAGACTCTTGG

For the overexpression screen, ORF regions of targeted molecules that displayed single, well-defined Unkempt-binding sites were cloned into the pCAGGS-IRES-RFP vector and point mutations were introduced into the critical UAG motifs using site-directed mutagenesis so as to preserve the encoded amino acid sequence but abolish the binding by Unkempt (see Supplemental Fig. S5C). We selected Unkempt-targeted molecules that significantly affected the induction of bipolar shape in the RNAi screen (ARF1, ARPC2, CCT5, CDC42, HSPA8, MAPRE1, RHOA, S100A11), controls that showed no effect (DPY30, RPS27L), and non-targeted polarity components (HRAS, PRKCZ, RAC1; Supplemental Fig. S5E). To test the effect of overexpression of these molecules on morphological transformation, we seeded GFP- and Unkempt-inducible HeLa cells at 8×10^4 cells per well in 6-well plates and transfected the plasmids using Lipofectamine 2000 transfection reagent (Life Technologies) according to the manufacturer's instructions. Twelve hours later, the transfected cells were treated with Dox for 48 hours and the morphologies of at least 20 GFP- and RFP-positive cells were quantified for each overexpressed gene by calculating their axial ratios (see Supplemental Fig. S5D,E).

Individual-nucleotide resolution UV-crosslinking and immunoprecipitation (iCLIP)

iCLIP experiments on SH-SY5Y cells, HeLa cells, and mouse embryonic brain were carried out in replicates as described in Konig et al., 2011 (Konig et al. 2011). See also Fig. 3C,D, and Supplemental Fig. S3.

SH-SY5Y and GFP and Unkempt-inducible HeLa cells were grown in 10 cm plates and harvested at 80% confluency. Prior to harvest, the inducible HeLa cells were treated with Dox for 24 hours to induce expression of Unkempt. Whole brains of E15.5 CD1 mouse embryos were processed as described (Ule et al. 2005). Briefly, ten whole embryonic brains were isolated and transferred to 6 ml of PBS in a 15 ml tube on ice. Brains were triturated first through a 5 ml serological pipette and then through a 200 µl pipette tip fitted to a 5 ml serological pipette to partially dissociate the tissue. The suspension was transferred to a 10 cm petri dish and irradiated with UV-light (254 nm) three times at 100 mJ/cm^2 on ice, mixing suspension after each irradiation pulse. The irradiated suspension was aliquoted into three 2 ml tubes and centrifuged at 2500 rpm for 3 minutes at 4 °C. The supernatant was removed and the cell pellets frozen at -80 °C until use.

Immunoprecipitation of the cross-linked Unkempt-RNA complexes from all three sample types was carried out using anti-Unkempt antibody (HPA023636, Sigma-Aldrich). Control reactions were performed using normal rabbit IgG antibodies (sc-2027). See also Supplemental Table S1.

The complete iCLIP experiment, including deep sequencing of the prepared cDNA libraries, was repeated in four, three, and five replicates for SH-SY5Y cells, HeLa cells, and mouse embryonic brain, respectively. Numbers of the obtained sequence reads, including mapping statistics, are given in Supplemental Table S1.

RNA-seq library preparation

SH-SY5Y cells were transfected at 40% confluence using Opti-MEM I Reduced Serum Medium and Lipofectamine 2000 (both Life Technologies) with a pool of siRNA targeting *UNK*

(siGENOME SMART pool, M-022950-01) or a non-targeting siRNA pool (siGENOME Non-Targeting siRNA Pool #1; D-001206-13; both Thermo Scientific Dharmacon) at 25 nM. After 36 hours of transfection, cells were re-transfected with the above siRNA pools and incubated for an additional 24 hours. Total RNA extracted from triplicates of each control and knockdown cells was DNase-treated and purified using RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA quality was assessed with the 2100 Bioanalyzer (Agilent Technologies). Sequencing library preparation was carried out using TruSeq RNA Sample Preparation Kit v2 (Illumina).

Ribosome profiling

To determine the effect of Unkempt on the rate of protein synthesis, we performed ribosome profiling experiment on inducible HeLa cells essentially as described (Ingolia et al. 2012), but using the linker and the primers as for the iCLIP experiments (Konig et al. 2011).

GFP-only or GFP and Unkempt-inducible Hela cells were grown in 15 cm dishes and harvested at 70% confluency. Prior to harvest, cells were treated with Dox for 18 hour to induce the expression of the transgenes. Cells were then washed in ice-cold PBS, lysed for 10 minutes on ice in lysis buffer, triturated by passing twice through a syringe fitted with a 26-gauge needle, and spun at 20,000 g for 10 minutes at 4 °C.

Cell lysate was digested with RNase I for 45 mintues at room temperature followed by the addition of SUPERase· IN to stop nuclease digestion. The lysate was underlayed with 1 M sucrose and spun in a TLA 100.3 rotor at 70,000 rpm for 4 hours at 4 °C. Ribosomal pellet was resuspended in Trizol reagent (Life Technologies) and RNA extracted according to the manufacturer's instructions.

The extracted RNA was size-selected by denaturing PAGE, retaining only fragments between 26 and 34 nts, and 3' end-dephosphorylated with T4 PNK for 20 minutes at 37 °C. Dephosphorylated RNA was precipitated, resuspended, and ligated to a pre-adenylated linker (L3-App) as described for the iCLIP procedure (Konig et al. 2011). The ligation mix was precipitated, resuspended, and the ligation products (between 47 and 55 nts) were purified by denaturing PAGE.

Reverse transcription (RT) was carried out using barcoded Rclip primers used for the iCLIP experiments (Konig et al. 2011) by resuspending the ligation products in 8.5 µl water, 0.5 µl Rclip primer (100 µM; Rclip primer with a different barcode was used for each sample), and 1 µl dNTP (10 mM), with SuperScript III Reverse Transcriptase (Life Technologies) according to the manufacturer's instructions, including the treatment with RNase H. The reaction mix was precipitated and the produced cDNA (between 79 and 87 nts) was purified by denaturing PAGE.

RT products were circularized, cut, and precipitated as described for the iCLIP procedure (Konig et al. 2011). PCR amplification (12-15 cycles) was carried out using P5/P3 Solexa primers and Accuprime Supermix 1 enzyme (Life Technologies). The prepared cDNA libraries were purified using the Agencourt AMPure XP system (Beckman Coulter).

Quality control of cDNA libraries and High-throughput sequencing

iCLIP, RNA-seq, and ribosome profiling cDNA libraries were analyzed by non-denaturing PAGE and the 2100 Bioanalyzer (Agilent Technologies), quantified with the Qubit 2.0 Fluorometer (Life Technologies), pooled by library and cell type (up to six libraries per pool), and sequenced using the HiSeq 2000 system (Illumina). All libraries were processed by single-

end sequencing with read lengths of 50 bases for iCLIP and ribosome profiling libraries, and 100 bases for RNA-seq libraries.

Gene ontology analysis

Unkempt targets in SH-SY5Y cells, HeLa cells, and mouse embryonic brain were analyzed using the online tool DAVID (Huang da et al. 2009) to find enriched gene ontology (GO) categories in each set of targets. Lists of background genes used for the analyses were defined by setting a minimum gene expression level threshold to our RNA-seq data for SH-SY5Y cells (RPKM ≥ 1; 10,879 genes), or to the ENCODE expression datasets for HeLa S3 cells (RPKM ≥ 5; ENCODE Caltech RNA-seq; 7,407 genes) and mouse embryonic brain (RPKM ≥ 1; GSM1000572; 12,992 genes). We performed a GO analysis of the core set of Unkempt targets by using as a background genes in common to all three cell type-specific background gene lists (5,240 genes). See also Supplemental Table S3. Human orthologs of mouse genes were identified using the web-based bioDBnet application (Mudunuri et al. 2009).

Canonical pathway analysis

The enrichment of canonical pathways in Unkempt-target genes in SH-SY5Y cells, HeLa cells, and mouse embryonic brain was determined using the web-based Ingenuity Pathway Analysis program (Ingenuity Systems; www.ingenuity.com). For each cell line, the list of corresponding background genes as defined under GO analysis (see above) was uploaded by Ensembl Gene IDs. Numerical cutoffs were assigned specifically to Unkempt target genes (Supplemental Tables S2A-C) and the analysis was run under default settings by the core analysis module taking cell line-specific background genes as a reference set. Pathway

enrichment for each target set was analyzed by considering either all available pathways or by focusing on specific pathway subsets, namely pathways implicated in intracellular and second messenger signaling, or pathways related to neurotransmitter and other neuronal signaling. See also Supplemental Table S4.

Preparation of recombinant Unkempt

Recombinant mouse Unkempt protein was prepared by immunoaffinity purification according to an adaptation of previously published protocols (Nakatani and Ogryzko 2003; Abmayr et al. 2006). HeLa S3 cells were engineered for Dox-inducible expression of GFP and Flag-HA tagged mouse Unkempt as above and grown in suspension culture at 8 liters to a maximal density of 5×10^5 cells/ml. Prior to harvest, the cells were treated with Dox for 24 hours to induce the transgenes. Cells were pelleted, washed with ice-cold PBS, and resuspended in 5 × pellet volume of ice-cold hypotonic buffer (10 mM Tris, pH 7.3, 10 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM PMSF, 10 mM 2-mercaptoethanol, EDTA-free protease inhibitors [Roche]) supplemented with 30 μM $ZnCl_2$. The suspension was incubated on ice for 10 minutes, spun at 2,500 rpm for 10 minutes at 4 °C, resuspended in 1 × pellet volume of ice-cold hypotonic buffer supplemented with 0.3 % NP-40, and homogenized with 12 strokes in a Dounce homogenizer with a loose pestle (Wheaton). The lysate was spun at 4,000 rpm for 10 minutes at 4 °C, the supernatant (cytoplasmic extract) was transferred to a fresh tube and supplemented with 0.11 × volume of 10 × cytoplasmic extract buffer (0.3 M Tris, pH 7.3, 1.4 M KCl, 0.03 M $MgCl_2$). The extract was mixed and centrifuged 25,000 g for 30 minutes at 4 °C. The supernatant was transferred to a new tube, adjusted to 10% with glycerol, and the tagged Unkempt was immunoprecipitated using ANTI-FLAG M2 Affinity Gel (Sigma; A2220), rotating the tube end-

over-end for 2 hours at 4 °C. The tube was spun at 3,500 rpm for 5 minutes at 4 °C and the affinity gel was washed three times with high-salt buffer (50 mM Tris, pH 7.3, 500 mM KCl, 5 mM $MgCl_2$, 30 μM $ZnCl_2$, 10 % glycerol, 0.5 % NP-40), spinning at 3.500 rpm for 1 minute at 4 °C between each wash. Subsequently, the affinity gel was washed twice with storage buffer (20 mM Tris, pH 7.3, 100 mM KCl, 50% glycerol, 30 μM $ZnCl_2$). The affinity gel was resuspended in storage buffer supplemented with 0.2 mg/ml Flag peptide (Sigma; F3290) and recombinant Unkempt was eluted by rotating the suspension end-over-end for 30 minutes at 4 °C. Affinity gel was removed using a Micro Bio-Spin Chromatography Column (Bio-Rad; 732-6204) and the concentration of recombinant Unkempt was estimated by denaturing PAGE followed by Coomassie-staining. The recombinant Unkempt was stored at -20 °C for up to 2 months without any signs of aggregation or degradation.

Radiolabeling of RNA and EMSA assays

Synthetic oligoribonucleotides (18- or 19-mers; Integrated DNA Technologies) at 1 μM were combined with 2 μl of [γ-^{32}P]-ATP (3000 Ci/mmol; PerkinElmer), 5 μl of 10 × T4 PNK reaction buffer, and 20 U of T4 polynucleotide kinase (New England BioLabs) in a 50-μl reaction volume and incubated for 30 minutes at 37 °C. The enzyme was heat-inactivated by incubating the reaction for 20 minutes at 65 °C and the radiolabeled oligoribonucleotides were purified using Micro Bio-Spin P-6 Gel Columns (Bio-Rad) according to the manufacturer's instructions.

EMSA assays were performed by combining radiolabeled RNA at a final concentration of 50 nM with varying concentrations of recombinant Unkempt (0 – 180 nM) in 10 μl binding buffer (20 mM Tris, pH 7.4, 100 mM KCl, 1 mM $MgCl_2$, 10 % glycerol, 1 mM DTT, and 0.1

mM $ZnCl_2$). The reactions were mixed by gentle pipetting, incubated for 15 minutes at room temperature, and loaded on a 1.7 % gel made of low melting temperature agarose (NuSieve GTG Agarose, Lonza) in 0.5 × TBE buffer. Novex Hi-Density TBE Sample Buffer (5X) (Life Technologies) was loaded in a separate well on the gel to allow for assessment of RNA migration. Electrophoresis was run at 80 V for 35 minutes at room temperature in 0.5 × TBE buffer. After the electrophoresis, the gel was dried in a sandwich assembly with single sheets of nylon membrane (Hybond $–N^+$, GE Healthcare) in direct contact with the gel (above and below), followed on each side by four sheets of Whatman paper (Chromatography paper, GE Healthcare), followed by several sheets of paper towel, and covered with a glass plate to evenly distribute the pressure of a weight (about 4,000 g) that was placed on top of the sandwich assembly. The gel was dried in this assembly for three hours and then exposed to autoradiography film (HyBlot CL, Denville Scientific) overnight at -80 °C. No radioactive signal was detected on either nylon membrane after drying, indicating that neither RNA nor protein diffused out of the gel during drying.

Polyribosome profiling and nuclease sensitivity assay

Mouse embryonic brain polyribosomes were prepared as follows. Whole brain of E14.5 CD1 mouse embryos was isolated and transferred directly to fresh ice-cold lysis buffer (gradient buffer [20 mM HEPES-KOH, pH 7.4, 150 mM NaCl, 5mM $MgCl_2$] supplemented with EDTA-free protease inhibitors [Roche], RNasin ribonuclease inhibitor [Promega], cycloheximide [CHX; 100 μg/ml final concentration], and 0.5 mM DTT). Forty whole brains were homogenized in 10 ml of lysis buffer with 12 strokes in a Dounce homogenizer with a loose pestle (Wheaton). Detergent (NP-40) was added at a final concentration of 1% and the homogenized brains were

incubated on ice for 10 minutes. The obtained lysate was aliquoted into 1.5 ml microcentrifuge tubes (1 ml per tube, meaning approximately four whole brains per sample) and spun at 20,000 g for 10 minutes at 4 °C.

Polyribosomes from SH-SY5Y cells were prepared by treating 60%-70% confluent cells grown in a 15 cm tissue culture dish with CHX (100 μg/ml final concentration) for 15 minutes. Cells were rinsed twice with PBS, the dish was placed on a mix of water and ice, and 1.2 ml of ice-cold lysis buffer containing 1% NP-40 was evenly distributed onto the cells. Following incubation on ice for 5 minutes, cells were scraped into a 1.5 ml tube, passed twice through a syringe fitted with a 23-gauge needle, and spun at 20,000 g for 10 minutes at 4 °C.

Puromycin-induced run-off experiment was carried out as above but with pre-incubation of cells for 30 minutes with 1 mM puromycin (Sigma-Aldrich) followed by treatment with CHX for 15 minutes.

For EDTA-control experiments, $MgCl_2$ in the lysis buffer was substituted with 30 mM EDTA.

Polyribosome profiles were generated according to the published protocol (Darnell et al. 2011). The supernatant (1 ml) was layered on top of a 15%-50% sucrose gradient prepared in gradient buffer in polyallomer ultracentrifuge tubes (Beckman 331372). The loaded tubes we spun at 40,000 rpm for 2 hours at 4 °C using a Beckman SW41 rotor and twenty fractions were collected. Fraction collection and recording of the absorbance at 254 nm was carried out using BR-188 Density Gradient Fractionation System (Brandel).

Nuclease sensitivity assays were performed essentially as described (Darnell et al. 2011). Briefly, sucrose fractions of whole brain polyribosomes containing Unkempt were pooled, diluted 1:1 with gradient buffer supplemented with 1 mM $CaCl_2$, and mixed for 20 minutes in the

presence of 1000 U/ml micrococcal nuclease (MNase), 2500 U/ml RNase I, or no enzyme at room temperature. EGTA (2 mM) was added to the sample treated with MNase to stop the reaction, and all samples were placed on ice. Following high-speed sedimentation for 2 hours at 280,000 g at 4 °C, supernatants were precipitated with trichloroacetic acid (TCA) and pellets were washed in PBS, restored to the original volume, and TCA-precipitated. All samples, including input to the spin, were analyzed by Western blotting.

Processing and genomic mapping of high-throughput sequencing reads

iCLIP data:

Genomic alignment using Bowtie (Langmead et al. 2009) and further processing of the iCLIP data was done using the human genome version hg19/NCBI37 and the mouse genome version mm9/MGSCv37 with annotations taken from Ensembl (version 60) (Flicek et al. 2011) as described previously (Konig et al. 2010). The iCLIP library contained a 4-nt experimental barcode flanked by a total of 5 nts of random barcode, which allowed multiplexing and the removal of PCR duplicates (Konig et al. 2011). See also Supplemental Table S1.

Detailed genomic analyses of the iCLIP data from human samples were based on GENCODE gene annotations (version 18) (Harrow et al. 2012). For assessing the genomic distribution of iCLIP crosslink nucleotides, we used the following hierarchy: ncRNA > CDS > 3' UTR > 5' UTR > intron > other > intergenic (Fig. 4A).

RNA-seq data:

The RNA-seq data from SH-SY5Y control and UNK knockdown samples (3 biological replicates each; 145,674,045 reads in total) were mapped to the human genome version NCBI37/hg19 using the splice-aware alignment algorithm Tophat (version 1.4.1; default

parameters for non-strand-specific data) (Trapnell et al. 2009). Between 83% and 87% of the reads mapped uniquely to the human genome and were kept for all following analyses.

Ribosome profiling:

The ribosome profiling data from HeLa cells expressing either GFP alone or GFP and UNK (two or three biological replicates, respectively) were processed as follows. All reads started with an interspersed arrangement of a 5-nt random barcode and a 4-nt experimental barcode (barcoded oligonucleotides Rclip#1-6 (Konig et al. 2011)). Before mapping, reads were trimmed to remove the 5' barcodes as well as 3' adapter sequences originating from oligonucleotide P3Solexa (5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'). The resulting fragments were further filtered for a fragment length of 25-35 nts.

Trimmed and filtered reads were then mapped to the human genome version NCBI37/hg19 using Tophat2 (version 2.0.6; parameters --segment-length 21 --library-type fr-firststrand (Kim and Salzberg 2011), keeping only uniquely mapping reads. Upon mapping, the first nt of ribosome-protected fragment (RPF) was extracted as the ribosome-occupied position.

Analyses of differential transcript abundance and RPKM calculation (RNA-seq)

Using the RNA-seq data from *UNK* knockdown and control SH-SY5Y cells, differential transcript abundance was assessed using DESeq2 (Anders and Huber 2010) based on exon coordinates from GENCODE gene annotations (version 18; support levels 1 and 2 only). 6,782 genes showed significant differential regulation upon *UNK* depletion (adjusted p-value < 0.05), including 3,428 up- and 3,354 down-regulated genes. *UNK* itself shows a significant 2-fold down-regulation (adjusted p-value = 9e-55).

RPKM calculation was performed using rpkmforgenes.py (http://sandberg.cmb.ki.se/media/data/rnaseq/instructions-rpkmforgenes.html) excluding non-coding genes (parameters -readcount -a ensGene.txt –onlycoding) (Ramskold et al. 2009).

Identification and characterization of UNK binding sites (iCLIP data)

Identification of significant binding sites was performed on the collapsed replicates of each sample type as described (Konig et al. 2010), applying a flank size of 15 nts and a false discovery rate of 5%. Binding sites were further filtered keeping only binding sites identified (i) by at least five crosslink events, (ii) overlapping only with mature transcript regions, and (iii) being unambiguously assigned to only one transcript. This procedure identified a total of 3,979, 2,846 and 3,112 binding sites from SH-SY5Y cells, HeLa cells and mouse brain samples, respectively. To call target transcripts in the different sample types, we further applied a reproducibility filter, considering only binding sites that were identified by crosslink events from multiple replicates (3/4 replicates for SY5Y, 2/3 for HeLa, and 3/5 for mouse brain samples), identifying a total of 1,186, 1,020 and 649 target genes, respectively. The reproducibility filter was omitted for identifying a core set of 263 UNK targets.

Within each binding site, we identified the position of the maximum (i.e. the nucleotide with the highest number of crosslink events within the cluster; the first was taken in case multiple nucleotides within the cluster had equal counts) and searched for the presence of UAG trimers within the 20 nucleotides upstream of the maximum. In order to evaluate the increased occurrence of UAG-containing Unkempt binding sites in the target transcripts, we performed 100 permutations of the sequence within this 20-nt window at all binding sites to obtain a mean background frequency. This analysis yielded significantly less UAG-containing targets for all

three sample types, as follows [cell type, (% observed mRNAs with UAG-containing binding sites, % obtained after randomization, p-value)]: SH-SY5Y (72.2, 37.0, 1.18e-145), HeLa (56.4, 26.2, 9.66e-146), mouse brain (69.8, 39.9, 1.72e-75). The same nucleotide position of the maximum was also used for assessing the positioning of UAG trimers (Fig. 5D) and ribosome occupancy. For integration with the ribosome profiling data, the iCLIP data from SH-SY5Y cells were further restricted to clusters mapping to the longest protein-coding transcripts (support levels 1 and 2 only) for each protein-coding gene (see below).

Finally, we tested several different consensus sequence expressions for their capacity to predict Unkempt binding sites in the CDS of the bound genes, assessing their sensitivity (% of all observed Unk binding sites detected by the expression) and specificity (% of true positives among the predicted binding sites). We obtained the following results [sequence, specificity and sensitivity]: UAGNNUUU, 23.0% and 11.1%; UAGNNNUUU, 21.9% and 9.9%; UAGNNWWWW, 12.2% and 20.7%; UAGNNNWWWW, 12.1% and 24.2%; UAGNNNNWWWW, 11.3% and 20.7%.

Analyses of differential ribosome occupancy (ribosome profiling data)

We used the ribosome profiling data from HeLa cells expressing either GFP alone or GFP and UNK to identify changes in ribosome occupancy upon UNK overexpression. Gene/transcript assignment was done based on GENCODE gene annotations (version 18), considering the longest protein-coding transcript (support levels 1 and 2 only) for each protein-coding gene, as described (Olshen et al. 2013). Differential analyses was performed with DESeq2 using the total number of RPFs per gene in the different replicates. In total, 1,815 genes showed a significant

change in ribosome occupancy (adjusted p-value < 0.05), including 767 and 1,048 genes with increased and decreased occupancy, respectively.

Availability of high-throughput sequencing data to reviewers

The RNA-seq, iCLIP and ribosome profiling data from this study are available to the reviewers for anonymous download at the following locations:

A) RNA-seq data:

We performed RNA-seq experiments to assess changes in transcript abundance upon knockdown of UNK in human SH-SY5Y cells. We processed three biological replicates from either siCTRL- or siUNK-transfected cells, generating 100-nt single-end reads on a HiSeq 2000 system (Illumina).

The fastq files with the raw sequence reads can be retrieved under the following URLs:

RNA-seq control (siCTRL), replicates 1-3:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_GL1.R1.fastq.gz
http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_GL2.R1.fastq.gz
http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_GL3.R1.fastq.gz

RNA-seq Unk knockdown (siUnk), replicates 1-3:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_UU1.R1.fastq.gz
http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_UU2.R1.fastq.gz
http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_UU3.R1.fastq.gz

In addition, we submitted the RNA-seq data to ArrayExpress under the accession number E-MTAB-2277.

B) iCLIP data:

We mapped binding sites of Unkempt under three different conditions: in human SH-SY5Y cells, in HeLa cells with doxycycline-inducible expression of UNK and GFP, and in murine E15 embryonic brain samples. We processed between three and five biological replicates per condition, generating 50-nt single-end reads on a HiSeq 2000 system (Illumina).

For convenient viewing of the data, we provide bedGraph files containing the collapsed iCLIP data from the different conditions after genomic mapping and random barcode evaluation, showing the number of crosslink events at each nucleotide within the UNK binding sites:

iCLIP SH-SY5Y cells:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Unk_binding_sites_SH-SY5Y.bed

iCLIP HeLa cells:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Unk_binding_sites_HeLa.bed

iCLIP embryonic brain:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Unk_binding_sites_mouse_brain.bed

Furthermore, the fastq files containing the raw sequence reads can be found under the following links:

SH-SY5Y cells (4 replicates):

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_SH-SY5Y_exp_Hs_NNNACCTNN_20110920_MuE_2.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_SH-SY5Y_exp_Hs_NNNTTGTNN_20111117_MC134SHHe_2.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_SH-SY5Y_exp_Hs_NNNGGTTNN_20111117_MC134SHHe_1.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_SH-SY5Y_exp_Hs_NNNTTGTNN_20110920_MuE_1.fq.gz

HeLa cells (3 replicates):

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_HeLa_exp_Hs_NNNCAATNN_20111117_MC134SHHe_3.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_HeLa_exp_Hs_NNNGGCGNN_20111117_MC134SHHe_4.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_HeLa_exp_Hs_NNNCCGGNN_20111117_MC134SHHe_5.fq.gz

Embryonic brain (5 replicates):

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_brain_exp_Mm_NNNGGCGNN_20111011_MuNE_1.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_brain_exp_Mm_NNNTTGTNN_20111011_MuNE_3.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_brain_exp_Mm_NNNGGTTNN_20111011_MuNE_2.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_brain_exp_Mm_NNNTTGTNN_20111123_MC137NSH_5.fq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/iCLIP_Unk_brain_exp_Mm_NNNGGTTNN_20111123_MC137NSH_4.fq.gz

In addition, we submitted the iCLIP data to ArrayExpress under the accession number E-MTAB-2279.

C) Ribosome profiling data:

We performed ribosome profiling experiments to assess changes in ribosome occupancy upon doxycycline-inducible expression of UNK and GFP in human HeLa cells. We processed three biological replicates of UNK and GFP expressing cells plus two replicates of control HeLa cells (expressing only GFP), generating 50-nt single-end reads on a HiSeq 2000 system (Illumina).

The fastq files with the raw sequence reads can be retrieved under the following URLs:

Ribosome profiling control (GFP expression only), replicates 1 and 2:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_Lane2_HeLa_1.R1.fastq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_Lane2_HeLa_2.R1.fastq.gz

Ribosome profiling UNK + GFP expression, replicates 1-3:

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_Lane2_HeLa_4.R1.fastq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_Lane2_HeLa_5.R1.fastq.gz

http://www.ebi.ac.uk/luscombe-srv/Data_Murn_etal/Sample_Lane2_HeLa_6.R1.fastq.gz

In addition, we submitted the ribosome profiling data to ArrayExpress under the accession number E-MTAB-2278.

Supplemental References

Abmayr SM, Yao T, Parmely T, Workman JL. 2006. Preparation of nuclear and cytoplasmic extracts from mammalian cells. *Curr Protoc Mol Biol* **Chapter 12**: Unit 12 11.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW et al. 2011. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**: 247-261.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S et al. 2011. Ensembl 2011. *Nucleic Acids Res* **39**: D800-806.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760-1774.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**: 1534-1550.

Kim D, Salzberg SL. 2011. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* **12**: R72.

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909-915.

-. 2011. iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp*.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Mudunuri U, Che A, Yi M, Stephens RM. 2009. bioDBnet: the biological database network. *Bioinformatics* **25**: 555-556.

Nakatani Y, Ogryzko V. 2003. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol* **370**: 430-444.

Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS. 2013. Assessing gene-level translational control from ribosome profiling. *Bioinformatics* **29**: 2995-3002.

Ramskold D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol* **5**: e1000598.

Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, Eyaid W, Bodell A, Barry B, Gleason D, Allen K et al. 2010. Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. *Nat Genet* **42**: 245-249.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Ule J, Jensen K, Mele A, Darnell RB. 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**: 376-386.

Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW, 3rd et al. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology* **10**: R130.

Xie Z, Moy LY, Sanada K, Zhou Y, Buchman JJ, Tsai LH. 2007. Cep120 and TACCs control interkinetic nuclear migration and the neural progenitor pool. *Neuron* **56**: 79-93.