Depletion of the RNA-binding protein PURA triggers changes in posttranscriptional gene regulation and loss of P-bodies

Lena Molitor[1,*], Melina Klostermann[2,*], Sabrina Bacher[1], Nadine Spranger[1], Juliane Merl-Pham[3], Ejona Rusha[4], Daniel Tews[5], Anna Pertek[4], Marcel Proske[1,6], Anke Busch[7], Sarah Reschke[8], Sandra Burczyk[6], Regina Feederle[9], Stefanie M. Hauck[3], Helmut Blum[8], Micha Drukker[10,11], Pamela Fischer-Posovszky[5], Julian König[7], Kathi Zarnack[2,$], Dierk Niessing[1,6,$]

* Shared first authors.

$ Correspondence should be addressed to: Kathi Zarnack (kathi.zarnack@bmls.de), Dierk Niessing (niessing@helmholtz-muenchen.de)

Affiliations:

1 Institute of Structural Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
2 Buchmann Institute for Molecular Life Sciences (BMLS) and Faculty Biological Sciences, Goethe University Frankfurt, Frankfurt, Germany.
3 Metabolomics and Proteomics Core, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
4 iPSC Core Facility, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
5 Department of Pediatrics and Adolescent Medicine, Ulm University Medical Center, Ulm, Germany.
6 Institute of Pharmaceutical Biotechnology, Ulm University, Ulm, Germany
7 Institute of Molecular Biology (IMB), Mainz 55128, Germany.
8 LAFUGA Genomics, Genzentrum, Ludwig-Maximilians University Munich, Munich, Germany.
9 Monoclonal Antibody Core Facility, Institute for Diabetes and Obesity, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
10 Institute of Stem Cell Research, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
11 Division of Drug Discovery and Safety, Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, The Netherlands.

Abstract

The RNA-binding protein PURA is implicated in the rare, monogenetic, neurodevelopmental disorder PURA Syndrome. However, very little is known about the affected molecular pathways. PURA has been reported to bind DNA and RNA and to be connected to different cellular functions. Here, we show that PURA predominantly resides in the cytoplasm, where it binds to thousands of transcripts. Many of these transcripts change abundance in response to *PURA* depletion, in parts reflected in altered protein expression levels. A closer inspection of the regulated proteins indicates a role of PURA in immune responses, mitochondrial function, autophagy and processing (P)-body activity. Intriguingly, reduced PURA levels lower the expression of the integral P-body component LSM14A and strongly impair P-bodies in human cells. Furthermore, *PURA* knockdown results in stabilization of P-body-associated transcripts, whereas other mRNAs are decreased. Hence, reduced PURA levels, as reported in patients with PURA syndrome, are associated with the formation and composition of the phase-separated RNA processing machinery. Our study on the RNA-binding protein PURA provides a blueprint for the comprehensive understanding of the tight connection between P-body-associated RNA regulation and neurodevelopmental disorders.

Introduction

In recent years, an increasing number of RNA-binding proteins have been implicated in the pathology of human disorders, in particular in neuronal diseases (1,2). One of them is the purine rich element binding protein A (PURA, formerly known as Pur-alpha) (3).

In 2014, mutations in the *PURA* gene have been linked to the neurodevelopmental disorder PURA Syndrome (4,5). Patients with heterozygous *de novo* mutations in *PURA* show a range of symptoms including neurodevelopmental delay, intellectual disability, hypotonia, and epilepsy (6,7). For unknown reasons, the phenotypes of patients with PURA Syndrome vary considerably. In addition, mice with homozygous deletions of *PURA* from two independent studies show phenotypes partially resembling patients with PURA Syndrome, including tremor and movement abnormalities (8,9). While it is generally assumed that symptoms are caused by haploinsufficiency of the PURA protein (10,11), the affected molecular pathways remain largely elusive. Furthermore, mice with a heterozygous mutation in *PURA* show very little overlap with human symptoms (12), underlining the importance of PURA-related studies in human cells.

PURA was initially reported as a transcription factor binding to a specific sequence motif upstream of the *c-myc* gene (13,14). While this and other DNA targets of PURA have been described in the past (15), the majority of recent reports indicated an involvement in RNA-related processes. In particular, PURA was frequently found in cytoplasmic mRNA-transport complexes of neurons in mice (16-20) and localized to cytoplasmic stress granules in human cells (21). While these reports indicate that PURA associates with both DNA and RNA, it remains unclear which interaction of PURA is functionally more relevant.

PURA belongs to the PC4 family of proteins (22) that bind single-stranded (ss) DNA and RNA. NMR measurements and quantitative binding studies revealed that PURA interacts with ssDNA and ssRNA with comparable affinities and properties (23). Crystal structures of PURA from different species revealed two so-called PUR domains, which mediate PURA's dimerization, and indicated that there are no structural constraints favoring DNA or RNA binding (23,24). Thus, although we have a relatively clear understanding of PURA's structural and molecular properties (3), a link is missing that explains how its nucleic acid binding impacts cellular functions. Hence, the aim of this study was to identify target pathways of PURA that could be misregulated in patients with PURA Syndrome.

Material and Methods

Generation of a PURA specific antibody

Monoclonal IgG2-2a anti-PURA^{12D11} antibody was raised against the unstructured linker region between PUR I-II of human PURA (peptide sequence: Cys-AQLGPSQPPDLAQAQDEPRRA; Peps4LS GmbH, Heidelberg). Several antibodies were generated after immunization of Sprague Dawley rats with ovalbumin-coupled peptide and hybridoma fusion as described before (25). Antibody supernatants were validated for binding to biotinylated PURA peptide in a solid-phase immunoassay. Subsequently, the panel of antibody supernatants was tested by Western blot on total cell lysate from HeLa cells. Selected clones were subcloned twice by limiting dilution to obtain stable monoclonal cell lines. Finally, the best antibody was tested in Western blot experiments on total cell lysate from *PURA* KD and CTRL HeLa cells. Additionally, cross-reactivity to PURB was excluded using recombinantly expressed GST-PURA and GST-PURB in Western blot assays (**Supplementary Figure 3C**). For this, human PURA and PURB were cloned into the pOPIN-J expression vector and expressed in *E. coli* Rosetta 2 (DE3) cells. Proteins were purified via GSTrap FF column (Cytiva). The GST-Tag was removed via 3C-protease cleavage and sample passed through HiTrap Q FF and HiTrap Heparin FF column. Protein bound to the HiTrap Heparin FF column was eluted and sample further purified using a HiLoad 16/600 Superdex 75 size-exclusion chromatography column (Cytiva). Experiments in this work were performed with hybridoma supernatant clone 12D11 (IgG2a/ƙ).

Generation of inducible PURA overexpression HeLa cell line

PiggyBac plasmids (26) harboring GFP-P2A-FLAG-PURA under control of a doxycycline-inducible promoter were generated using Gibson assembly (Thermo Fisher Scientific) of three fragments containing overhangs according to manufacturer's instructions. Using these plasmids, overexpression cell lines were generated. The plasmid containing GFP-P2A-FLAG-PURA as well as a PiggyBac Helper plasmid (26) were transfected into cells using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. Transfected cells were incubated at 37 °C, 5% CO_2 for two to four days. Transfected cells were subsequently selected by addition of hygromycin (700 µg/ml) for at least ten days. GFP-P2A-FLAG-PURA overexpression was induced by addition of doxycycline (1 µg/ml) for 24 h.

siRNA-mediated *PURA* knockdown

siRNA-mediated knockdown of *PURA* in HeLa cells was performed using a predesigned PURA siRNA pool (Dharmacon, M-012136-01-000). One day prior to transfection, HeLa cells were plated at 20% confluency in 6-well plates. The next day, lipofection was performed using RNAiMAX (Thermo Fisher Scientific) according to manufacturer's instructions. Briefly, 20 µM siRNA pool was mixed with OptiMEM (Thermo Fisher Scientific) in one reaction and lipofectamine RNAiMAX was mixed with OptiMEM in a second reaction. Both reactions were incubated at room temperature (RT) for 5 min and mixed by flicking. Meanwhile, culture media on HeLa cells was

changed to 1 ml DMEM + 10% FBS lacking antibiotics. Contents of both reactions were mixed and further incubated for 20 min at RT. Finally, 400 µl of lipofection mix were added drop-wise onto HeLa cells. As a control, siGENOME Non-Targeting Pool #1 (Dharmacon, D-001206-13-05) was used as described for the PURA siRNA pool. The transfected cells were incubated for 48 h at 37 °C and 5% CO_2 and subsequently analyzed in downstream assays.

RNA sequencing

Four independent biological replicates of *PURA* KD and CTRL HeLa cells were prepared as described above. Control cells were only treated with lipofectamine RNAiMAX (Thermo Fisher) and OptiMEM (Thermo Fisher) without adding the non-targeting pool siRNA. After 48 h incubation with siRNA, cells were washed using PBS and subsequently collected in TRIzol Reagent (Thermo Fisher Scientific). Total RNA was extracted using Maxwell® RSC miRNA Tissue Kit (AS1460) on the Maxwell® RSC instrument according to the manufacturer's instructions. Briefly, shock frozen cells were resuspended in 500 µl Trizol. 6 µl 1-Thioglycerol were added to 300 µl Trizol suspension. The suspension was mixed vigorously and total RNA extracted by phase separation using RSC SimplyRNA Tissue (AS1340) setup. The concentration was measured on a Nanodrop spectrometer (Thermo Fisher Scientific) and the quality of the RNA samples checked on an Agilent Bioanalyzer (Agilent; RNA 6,000 Nano kit).

Libraries were generated with the Lexogen SENSE mRNA-seq Library Prep Kit V2 (Lexogen) with half of the volume described in the manual. Briefly, 100 ng of total RNA were bound to oligo-dT beads, reverse-transcribed on the beads and the second strand of cDNA was synthesized after ligation of an adapter. The eluted and adapted cDNA was amplified with barcoded primers during 11 cycles of PCR with the following program: 98 °C for 10 s, 65 °C for 20 s, 72 °C for 30 s and a final extension at 72 °C for 1 min (hold at 10 °C). The PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter). The quality of the libraries was validated with an Agilent Bioanalyzer (DNA 1,000 Kit). The finished, barcoded libraries were pooled and sequenced on an Illumina HiSeq1500 as 50-nt single-end reads with an average sequencing depth of 20 $x10^6$ reads per sample. Each library yielded on average 35 million reads.

Reads were aligned to the human genome (hg38.p12 from GENCODE) with STAR (version 2.7.6a) (27) with up to 4% mismatches and no multimapping allowed. Reads per gene were counted using htseq-count (version 0.11.3) (28) with default settings and GENCODE gene annotation (v31) (29). Differential expression analysis between *PURA* KD and CTRL samples was performed using DESeq2 (version 1.33.4) (30) with a significance cutoff at an adjusted P value < 0.01 (Benjamini-Hochberg correction). This yielded a total of 2,436 significantly differentially expressed transcripts, including 1,166 with increased and 1,270 with decreased expression upon *PURA* KD. For the heatmap in **Figure 1C**, library size-corrected RNA-seq read counts per sample were rlog-transformed (DESeq2, version 1.33.4) (30). For visualization, the counts were converted into row-wise z-scores.

Shot-gun proteomics

Generation of protein lysates

Four independent biological replicates of *PURA* KD and CTRL HeLa cells were prepared as described above. Cell samples were collected by scraping them in PBS and subsequent centrifugation at 500 x g for 3 min. Cell pellets were lysed in 100 µl RIPA buffer and subsequently centrifuged at 16,000 x g at 4 °C for 10 min. The supernatant was transferred to a new tube and the amount of protein in the lysate was determined by Bradford assay (ROTI®Quant, Carl Roth).

Sample preparation for mass spectrometric analysis

10 µg per sample were digested with Lys-C and trypsin using a modified FASP procedure (31,32). Briefly, after reduction and alkylation using DTT and IAA, the proteins were centrifuged on a 30 kDa cutoff filter device (Sartorius) and washed each thrice with UA buffer (8 M urea in 0.1 M Tris/HCl pH 8.5) and with 50 mM ammonium bicarbonate. The proteins were digested for 2 h at RT using 0.5 µg Lys-C (Wako Chemicals) and for 16 h at 37 °C using 1 µg trypsin (Promega). After centrifugation (10 min at 14 000 g), the eluted peptides were acidified with 0.5% TFA and stored at -20 °C.

LC-MS/MS measurements

LC-MS/MS analysis was performed on a QExactive HFX mass spectrometer (Thermo Fisher Scientific) online coupled to a Ultimate 3000 RSLC nano-HPLC (Dionex). Samples were automatically injected and loaded onto the C18 trap cartridge and after 5 min eluted and separated on the C18 analytical column (Acquity UPLC M-Class HSS T3 Column, 1.8 µm, 75 µm x 250 mm; Waters) by a 90 min non-linear acetonitrile gradient at a flow rate of 250 nl/min. MS spectra were recorded at a resolution of 60,000 with an AGC target of 3 x 10e6 and a maximum injection time of 30 ms from 300 to 1500 m/z. From the MS scan, the 15 most abundant peptide ions were selected for fragmentation via HCD with a normalized collision energy of 28, an isolation window of 1.6 m/z, and a dynamic exclusion of 30 s. MS/MS spectra were recorded at a resolution of 15,000 with a AGC target of 10e5 and a maximum injection time of 50 ms. Unassigned charges, and charges of +1 and >8 were excluded from precursor selection.

Quantitative MS analysis

Acquired raw data was analyzed in the Proteome Discoverer 2.4 SP1 software (Thermo Fisher Scientific; version 2.4.1.15) for peptide and protein identification via a database search (Sequest HT search engine) against the SwissProt Human database (Release 2020_02, 20435 sequences; 11,490,581 residues), considering full tryptic specificity, allowing for up to one missed tryptic cleavage sites, precursor mass tolerance 10 ppm, fragment mass tolerance 0.02 Da. Carbamidomethylation of cysteine was set as a static modification. Dynamic modifications included deamidation of asparagine and glutamine, oxidation of methionine, and a combination of methionine loss with acetylation on protein N-terminus. The Percolator algorithm (33) was used for validating peptide spectrum matches and peptides. Only top-scoring

identifications for each spectrum were accepted, additionally satisfying a false discovery rate < 1% (high confidence). The final list of proteins satisfying the strict parsimony principle included only protein groups passing an additional protein confidence false discovery rate < 5% (target/decoy concatenated search validation).

Quantification of proteins, after precursor recalibration, was based on intensity values (at RT apex) for all unique peptides per protein. Peptide abundance values were normalized on total peptide amount. Nested protein ratios were then calculated using the pairwise ratio-based approach, with a background-based Student's *t*-test as hypothesis test, resulting in *PURA* KD / CTRL ratios and corresponding significance values for the individual proteins. Since the ratios were strongly influenced by missing values, we corrected them using limma (3.49.4) (34) treating missing values as follows: If protein abundance was not measured in any of the samples in one condition it was set to zero, which results in infinite values for the calculated ratios. Peptides with two or more missing values in one condition were treated as non-quantifiable and no ratios were calculated. When only one value was missing per condition, the ratios were calculated on the other three samples.

P values were adjusted for multiple testing by Benjamini-Hochberg correction. We deemed proteins with a *P* value < 0.05 significant. In the heatmap in **Figure 1C**, protein abundance is shown after z-score normalization. The correlation of changes in transcript and protein levels (**Figure 1D**) was calculated on proteins with zero or one missing value per condition using Pearson correlation.

Quantitative real-time PCR (qPCR)

RNA was extracted using the High Pure RNA isolation kit (Roche Molecular Systems) according to manufacturer's instructions and following the general precautions required for RNA work (35). RNA was eluted in 50 µl DEPC water. Two rounds of DNase digestion using TURBO DNA-free kit (Thermo Fisher Scientific) according to manufacturer's instructions were performed to safely remove all remaining DNA. The amount of pure RNA was analyzed by measuring the OD_{260} using a NanoDrop microliter photometer. The extracted RNA was reverse transcribed using PrimeScript RT Master Mix (Takara) according to manufacturer's instructions. On the generated cDNA libraries, qPCR experiments were performed using SYBR Green Master Mix (Thermo Fisher Scientific) in a LightCycler 480 Instrument II (Roche Molecular Systems) in 96-well format. c_t values were used for the analysis of differential mRNA abundance between *PURA* KD and CTRL conditions. This analysis was done using the $\Delta\Delta c_t$ method (36) with the housekeeping genes *GAPDH* and *RPL32* for normalization. Changes were tested for significance using a paired Student's *t*-test on the $\Delta\Delta c_t$ values. All oligonucleotides used for qPCR experiments are listed in **Supplementary Table 5**.

Western blot experiments

Wet Western blot experiments were performed using Mini Blot module (Thermo Fisher Scientific) according to manufacturer's instructions. NuPage loading dye (Thermo Fisher Scientific) was added onto cell lysates and the mix was heated to 70 °C for 10 min prior to gel loading. The blotting module was filled with NuPAGE 1x SDS running buffer (Thermo Fisher Scientific) and 500 µl NuPAGE antioxidant (Thermo Fisher Scientific) was added to the front part of the chamber. Samples were cooled to RT and loaded onto NuPAGE 4-12% Bis-Tris pre-cast gradient gels (Thermo Fisher Scientific) in the buffer-filled blotting module. For protein molecular weight estimation, BlueStar Plus Prestained Protein Marker (Nippon Genetics) was used. The gel was run for 45 min at 200 V. The PVDF blotting membrane was activated in methanol for 30 s prior to usage. All other components were soaked in 1x transfer buffer (Thermo Fisher Scientific) with 20% methanol prior to blotting module assembly. The assembled module was inserted into the Mini Blot module (Thermo Fisher Scientific), the surrounding chamber was filled with MQ water and the blotting reaction was run for 90 min at 30 V. Subsequently, the membrane was blocked using 1% casein in PBST. This was incubated for 30 min at RT while rotating on tube revolver machine (neoLab). Then, the blocking solution was discarded and primary antibody (dilutions in **Supplementary Table 7**) in 0.5% casein in PBST was added. Primary antibody was incubated overnight at 4 °C rotating on a tube revolver device. On the next day the primary antibody solution was discarded, and blot was washed. Next, the secondary antibody (dilutions in **Supplementary Table 8**) in PBST was incubated on the membrane for 1 h at RT while rotating. Finally, the secondary antibody solution was discarded, and the membrane was washed. Then, the membrane was incubated with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare). 1:1 mix of solution A and solution B of the above-mentioned ECL kit was added onto the membrane. The membrane was analyzed using Fusion SL 4 device and software (Vilber Lourmat). Changes in fluorescence intensity between *PURA* KD and CTRL samples were tested for significance in R using a paired Student's *t*-test.

Immunofluorescence staining

Cells were cultivated in culture medium on coverslips (Thermo Fisher Scientific) in 12-well plates and subsequently washed using PBS and fixed using 3.7% formaldehyde in PBS. Fixing media was incubated for 10 min at RT in the cell culture dish. Permeabilization was induced by incubating the cells in 0.5% Triton X-100 in PBS for 5 min at RT. Afterwards, cells were washed twice using PBS and the coverslips were transferred to a pre-labelled parafilm that was placed onto wet Whatman paper to prevent drying. Cells were then blocked for 10 min in blocking buffer (1% donkey serum in PBST) and incubated with primary antibodies (**Supplementary Table 7**) in blocking buffer for 1 h at RT. Secondary antibodies (**Supplementary Table 8**) were diluted in blocking buffer and incubated for 1 h at RT. Washing steps after antibody incubation were performed with PBST. DNA was stained with DAPI at 0.5 µg/ml in PBS by incubation for 5 min at RT, followed by washing twice using PBS. Finally, cells

were mounted in ProLong Diamond Antifade (Thermo Fisher Scientific) and imaging was performed by fluorescence microscopy using an Axio.Observer.Z1 (Carl Zeiss AG) microscope. Immunofluorescence staining for P-bodies using the marker proteins DCP1A and LSM14A was performed on a LSM 880 Airyscan confocal microscope (Carl Zeiss AG).

Analysis of P-bodies

Quantification of granules in different cell lines was done using Fiji software. Per image, DCP1A- and LSM14A-positive granules were counted with the Analyze Particle Tool, after processing the corresponding channel with auto threshold. Granules were counted by selecting a particle size of 40-200 pixel units (μm^2) and a circularity of 0.1-1.00. In NHDF cells (**Supplementary Figure 10**), a particle size of 30-100 pixel units (μm^2) was selected. Cells were counted by quantifying the DAPI channel selecting a size of 1000-infinity pixel units (μm^2). Quantification was done of three biological replicated and for each replicate four confocal pictures (20x magnification) of each condition (CTRL and *PURA* KD) were analyzed.

Neural progenitor cell derivation from human induced pluripotent stem cells

The protocol for differentiation of neural progenitor cells was based on the generation of neurospheres (37). Briefly, the human induced pluripotent stem cell (hiPSC) line HMGU12 was harvested using StemMACS Passaging Solution (Miltenyi Biotec) and resuspended in DMEM/F-12 medium supplemented with 20% KSR, 1% NEAA, 1% GlutaMAX (Thermo Fisher Scientific), 10 µM SB431542, 5 µM dorsomorphin, 3 µM CHIR99021, 10 µM purmorphamine (Miltenyi Biotec), and 10 µM Y-27632 (Bio-Techne). The suspension was plated on an ultra-low attachment 6-well plate (Corning). Fresh medium was applied 24 h later without Y-27632. 48 h later, the basal medium was exchanged with N2B27-based medium containing a 1:1 mixture of DMEM-F-12 and Neurobasal A supplemented with 0.5% N-2, 1% B-27 minus vitamin A, 1% NEAA, and 1% GlutaMAX (Thermo Fisher Scientific), and the above-described small molecules. On day 5, N2B27-based medium was supplemented only with 50 µg/ml L-ascorbic acid, SB431542, and dorsomorphin. Approximately 24 h prior to replating, the medium was supplemented additionally with 5 ng/ml bFGF (Peprotech). On day 8, the neurospheres were mechanically dissociated and plated 1:6 on Matrigel-coated plates. Plated neurospheres were maintained for 7 days on N2B27 basal medium supplemented with Dual Smad inhibitors (Miltenyi Biotec (SB), Tocris (DM)), ascorbic acid and bFGF. On day 14, confluent neuroepithelial outgrowths were passaged in a 1:10 dilution using collagenase IV. The resulting neural progenitor cell (NPC) cultures were passaged every 7 days and maintained in N2B27 basal medium as described above with medium change every other day.

iCLIP experiments

iCLIP experiments for PURA were performed according to the published iCLIP2 protocol (38) with minor adaptions. Cells were grown to confluence, washed using PBS, crosslinked using Stratalinker 2400 (Vilber Lourmat) at 254 nm and at 300

mJ/cm^2 and lysed. Immunoprecipitation of PURA from cell lysates was either performed using anti-PURA^{12D11}-coated Dynabeads protein G (Thermo Fisher Scientific) or FLAG M2 beads (Sigma Aldrich). Subsequently, co-purified RNAs were dephosphorylated at the 3' end, the 3' adapter was ligated and the 5' end was labeled with radioactive isotopes (γ-^{32}P ATP; Hartmann Analytic). Samples were loaded onto a NuPAGE 4-12% Bis-Tris pre-cast SDS gradient gel (Thermo Fisher Scientific), transferred to a nitrocellulose membrane (GE Healthcare) and visualized using a Phosphoimager (Fujifilm). After cutting the correct region from the nitrocellulose membrane using a cutting mask, RNA was isolated by Phenol/Chloroform/Isoamylalkohol (pH=8.0, Sigma Aldrich) and the mix was transferred to a 2 ml Phase Lock Gel Heavy tube (Quantabio). After separation of the phases the RNA was precipitated by the addition of 0.75 µl GlycoBlueTM Coprecipitant (ThermoFisher), 40 µl 3 M sodium acetate pH 5.5 and 1 ml 100% ethanol. It was mixed and stored at -20 °C for at least two hours. Reverse transcription was performed on the resulting purified RNA using Superscript III (Thermo Fisher Scientific) according to manufacturer's instructions. Then cDNA was immobilized on MyONE SILANE beads (Thermo Fisher Scientific) as described by the manufacturer and the second adapter was ligated. The library was amplified for six PCR cycles with short primers followed by a ProNex size selection step (Promega) in a 1:2.95 (v/v) sample/bead ratio according to manufacturer's instructions. The size-selected library was amplified for 13/16 cycles (endogenous PURA, anti-PURA^{12D11}, HeLa cells; sample 1 & 4 and sample 2 & 3, respectively), 22 cycles (endogenous PURA, anti-PURA^{12D11}, NPCs), 12 cycles (overexpressed FLAG-PURA, anti-PURA^{12D11}, HeLa cells), 13/14 cycles (overexpressed FLAG-PURA, anti-FLAG, HeLa cells; sample 1 and sample 2, respectively) using Illumina primers (**Supplementary Table 6**). Afterwards, a second size selection was performed using ProNex chemistry (Promega) at 1/2.4 (v/v) sample/bead ratio according to manufacturer's instructions and cDNA was eluted in 20 µl MQ water. The final library was analyzed using a D1000 Chip (Agilent) in a Bioanalyzer (Agilent). Samples were equimolarly mixed before sequencing on an Illumina HiSeq 1500 platform with 75-110-nt single-end reads and 6.7-29.5 million reads per sample (**Supplementary Table 3**).

Processing of iCLIP reads

Basic quality controls were done with FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The first 15 nt of iCLIP reads hold a 6 nt sample barcode and 5+4 nt unique molecular identifiers (UMIs) surrounding the sample barcode. Based on the sequence qualities (Phred score) in this 15 nt barcode region, reads were filtered using the FASTX-Toolkit (v0.0.14) (http://hannonlab.cshl.edu/fastx toolkit/) and seqtk (v1.3) (https://github.com/lh3/seqtk/). The sample barcodes, found on positions 6 to 11 of the reads, were used to de-multiplex the set of all quality filtered reads using Flexbar (39) (v3.4.0). Afterwards, barcode regions and adapter sequences were trimmed from read ends using Flexbar, requiring a minimal overlap of 1 nt of read and adapter. UMIs were added to the read names and reads shorter than 15 nt were removed from further

analysis. Downstream analysis was done as described in Chapters 3.4 and 4.1 of Busch et al. (40). GENCODE (29) v31 genome assembly and annotation were used during mapping.

Binding site definition

Using the procedure described in (40), we defined 5-nt wide binding sites on the iCLIP data for the endogenous PURA (anti-PURA^{12D11}) from HeLa cells. In brief, the processed crosslink events of the four biological replicates were merged into two pseudo-replicates and subjected to peak calling by PureCLIP (version 1.3.1) (41) with default parameters. To define binding sites, PureCLIP sites closer than 5 nt were merged into regions, and isolated PureCLIP sites without an adjacent site within 4 nt were discarded. Binding site centers were defined iteratively by the position with the highest number of crosslink events and enlarged by 2 nt on both sides to obtain 5-nt binding sites. Binding site centers were required to harbor the maximum PureCLIP score within the binding site. Furthermore, binding sites were filtered for reproducibility by requiring a binding site to be supported by a sufficient number of crosslink events in at least two out of four replicates. The threshold for sufficient crosslink event coverage, using the 0.05 percentile, was determined as described in (40).

Genomic location of binding sites and crosslink events

We mapped binding sites or crosslink events to the gene and transcript annotation from GENCODE (v31, genome version hg38.p12) (29). Genes of gene level 3 were included only when no genes of higher level were overlapping. Similarly, transcripts of transcript level NA were included only when no transcripts of levels 1-3 were annotated for the gene. When a binding site overlapped with two genes, one was chosen at random. For the assignment of PURA binding sites to transcript regions, when binding sites overlapped with different regions from different transcripts, the region was chosen by a hierarchical rule with 3'UTR > 5'UTR > CDS > intron, which we established from visual evaluation of the crosslink event and binding site distribution. We divided the PURA-bound mRNAs into groups according to the location of the strongest PURA binding site on the transcript (3'UTR, n = 2,472; 5'UTR n = 102; CDS, n = 1,600; intron, n = 124; noncoding RNA = 391). The metaprofiles of the distribution of crosslink events along the transcript regions for mRNAs with strongest binding sites in the 3'UTR or CDS (**Figure 4C**) was calculated and visualized in R using the Bioconductor package cliProfiler (http://bioconductor.org/packages/release/bioc/html/cliProfiler.html).

Comparison of crosslink patterns

We compared the distribution of PURA crosslink events from the iCLIP data for endogenous PURA (anti-PURA^{12D11}) from HeLa cells to the two iCLIP datasets in which FLAG-tagged PURA was overexpressed in HeLa cells and immunoprecipitated with either anti-PURA^{12D11} or an anti-FLAG antibody using the following methods: First, we assessed differences in the iCLIP signal using metaprofiles in a 65-nt window around all PURA binding sites identified for endogenous PURA in HeLa cells

(**Supplementary Figure 5B**). In order to account for expression level and other differences, the crosslink events in each window were scaled to the minimum (set to 0) and maximum (set to 1) therein ("min-max normalization"). Second, modified from the approaches suggested in StoatyDive (42), we used unsupervised clustering to separate the crosslink patterns from all three experiments in a 205-nt window around 1,000 randomly selected PURA binding sites (**Supplementary Figure 5D**). For this, the crosslink patterns were subjected to min-max normalization within this window, followed by spline-smoothing using the smooth.spline function (stats package version 4.1.1, spar = 0.5) and inflated dimensions (dim = 1000). *k*-means clustering was performed on the smoothed patterns with k = 4. To visualize the crosslink patterns in 3'UTRs, we generated heatmaps of the first 300 nt of 1,563 3'UTRs with intermediated PURA crosslink coverage (10^2-10^6 crosslink events per window). The heatmaps show the min-max normalized, spline-smoothed (spar = 0.5, dim = 500) crosslink patterns for the three iCLIP experiments (**Figure 4D, Supplementary Figure 5E**).

Motif analysis

As PURA predominantly acts on mature RNAs, all motif analyses were performed on the sequences of processed transcripts. Binding sites were transferred from genomic to transcriptomic coordinates with the Bioconductor package GenomicFeatures (version 1.45.2) (43) using GENCODE transcript annotation (v31, genome version hg38.p12) (29). 50,952 binding sites could be unambiguously assigned and were used for further analyses. Sequences around binding sites were extracted with Biostrings::getSeq (version 2.61.2) (https://bioconductor.org/packages/release/bioc/html/Biostrings.html). For sequence logos, a 51-nt window centered at the binding site was used and logos are generated with ggseqlogo (version 0.1). 23,956 out of 50,952 (47.0%) were centered on a G and 31,004 out of 50,952 (60.8%) enclosed three or more purines.

To search for enriched motifs of 5-nt length (5-mers) within and around the PURA binding sites, we calculated the frequencies of all overlapping 5-mers with Biostrings::oligonucleotideFrequency (version 2.61.2) (https://bioconductor.org/packages/release/bioc/html/Biostrings.html) in three windows: (i) [-7 nt; 7 nt], i.e., including the binding sites plus 5 nt flanking on either side, and (ii) [-27 nt; -8 nt] and [8 nt; 27 nt], i.e., representing the 20 nt on either side of the binding site-containing windows. For comparison, we randomly selected 50,952 5-nt windows from expressed transcripts (i.e, harboring at least one PURA crosslink event).

Accessibility prediction

We used RNAplfold (from the ViennaRNA package, version 2.4.17) (44) to predict the accessibility in and around PURA binding sites (n = 48,803) in comparison to randomly selected positions (n = 9,737). In brief, we used 501-nt windows either around PURA binding sites or randomly selected and predicted the probability for each single nucleotide to be single-stranded. RNAplfold was set to a sliding window w = 100 nt and a maximum span l = 30 nt. As the obtained unpaired probabilities follow a bimodal

distribution, we transferred them to log-odds to obtain a bell-shaped distribution: log(odds ratio) = log(prob/1-prob). These were further converted into z-scores using the mean log-odds ratios of the binding sites versus the mean and standard deviation of the background regions (drawing 1,000 regions for 1,000 times). *P* values were calculated from the z-scores using 2*pnorm(-abs(z_score) and then adjusted for multiple testing with Benjamini Hochberg correction.

Comparison to published transcriptomes

We used published processed RNA-seq data comparing RNAs in stress granules (human osteosarcoma U-2 OS cells) or P-bodies (human embryonic kidney HEK293 cells) to total RNA (45,46), taking RNAs with FDR < 0.01 and $\log_2$-transformed fold change (l2fc) > 0 to belong to the respective P-body or stress granule transcriptome.

To determine the overlap of PURA-bound RNAs with dendritically localized RNAs, we used data from (47) that summarizes eight studies describing the dendritic transcriptome in mouse neurons from enrichment in RNA-seq data of dendrites compared to complete neurons. We used only RNAs that were found enriched in at least three of the eight studies (n = 397). These were matched to their human orthologs using the Ensembl database (version GRCh38.p13) (48) keeping only one-to-one orthologs (n = 4,867). Only RNAs expressed in HeLa cells (baseMean > 0 in our RNA-seq data) were considered for the overlap (n = 294).

Significance of overlaps was tested using Fisher's exact test, and differences in fold change distributions between groups were tested using Welch *t*-test.

Functional enrichment analyses

We performed functional enrichment analysis using hypeR (version 1.9.1) on the gene sets from REACTOME (version 7.4.1) and Gene Ontology (version 7.4.1) databases (49,50).

Results

Loss of PURA causes widespread changes in gene expression

Most studies in the past had focused on individual protein or nucleic acid targets of PURA. In order to globally assess the impact of PURA on gene expression, we depleted *PURA* in HeLa cells using siRNAs (**Supplementary Figure 1**). By performing transcriptomic analyses, we observed widespread changes in transcript abundance upon *PURA* knockdown (KD), relative to control. 2,436 genes were significantly down- or up-regulated (false discovery rate [FDR] < 0.01, DESeq2, (30), **Figure 1A, Supplementary Table 1**). Gene ontology analysis revealed that the regulated genes encoded for diverse functional categories, including mitochondrial functions, autophagy, vesicle-mediated transport, membrane trafficking and RNA metabolism (**Supplementary Figure 2A-C**). The most significantly downregulated transcript was the mRNA encoding the transcription factor CUX1, which is involved in the control of neuronal differentiation by specifically regulating dendrite development, branching, and spine formation in cortical neurons (51).

To test whether the changes in the transcriptome correlated with changes at the protein level, we performed mass-spectrometric analysis of protein extracts from *PURA* KD and control HeLa cells. We detected a total of 5,606 proteins with good confidence (peptide FDR < 1% and protein FDR < 5%) and observed a significant up- and down-regulation upon *PURA* KD of 79 and 64 proteins, respectively (FDR < 0.05, **Figure 1B, C Supplementary Figure 1F, G, Supplementary Table 2**). Of these, 45 proteins showed significant changes also on transcript levels and the vast majority of them changed in consistent direction in both RNA expression and protein abundance (41 out of 45, **Figure 1C, D, Supplementary Figure 2D**). More generally, the protein levels tended to follow the underlying transcript levels also in cases where these changes did not reach significance (Pearson correlation coefficient [*R*] = 0.16, *P* value = 2.2e-16, **Figure 1D**), indicating that the *PURA* KD altered transcript abundance rather than translation efficiency.

Among the 41 factors that showed consistent changes in both transcript and protein levels were the mitochondrial lipid transfer protein STARD7 that is involved in the delivery of phosphatidylcholine to the mitochondria (52) and the cytoskeleton associated Catenin alpha-1 (CTNNA1) that links actin and adherence junction components (53) (**Figure 1C, D**). Furthermore, coherent downregulation upon *PURA* KD was observed for the autophagy receptor Sequestome 1 (SQSTM1; also known as p62) consistent with (54), and the essential P-body component LSM14A. The downregulation of SQSTM1 and LSM14A was orthogonally confirmed by Western blot analysis (**Figure 1E-H**).

Altogether, we find that loss of PURA results in widespread changes in gene expression, affecting central functionalities of the cell such as mitochondrial function, autophagy and granule assembly. The quantitative comparison of transcriptome and proteome suggests that the majority of changes originate at the RNA level.

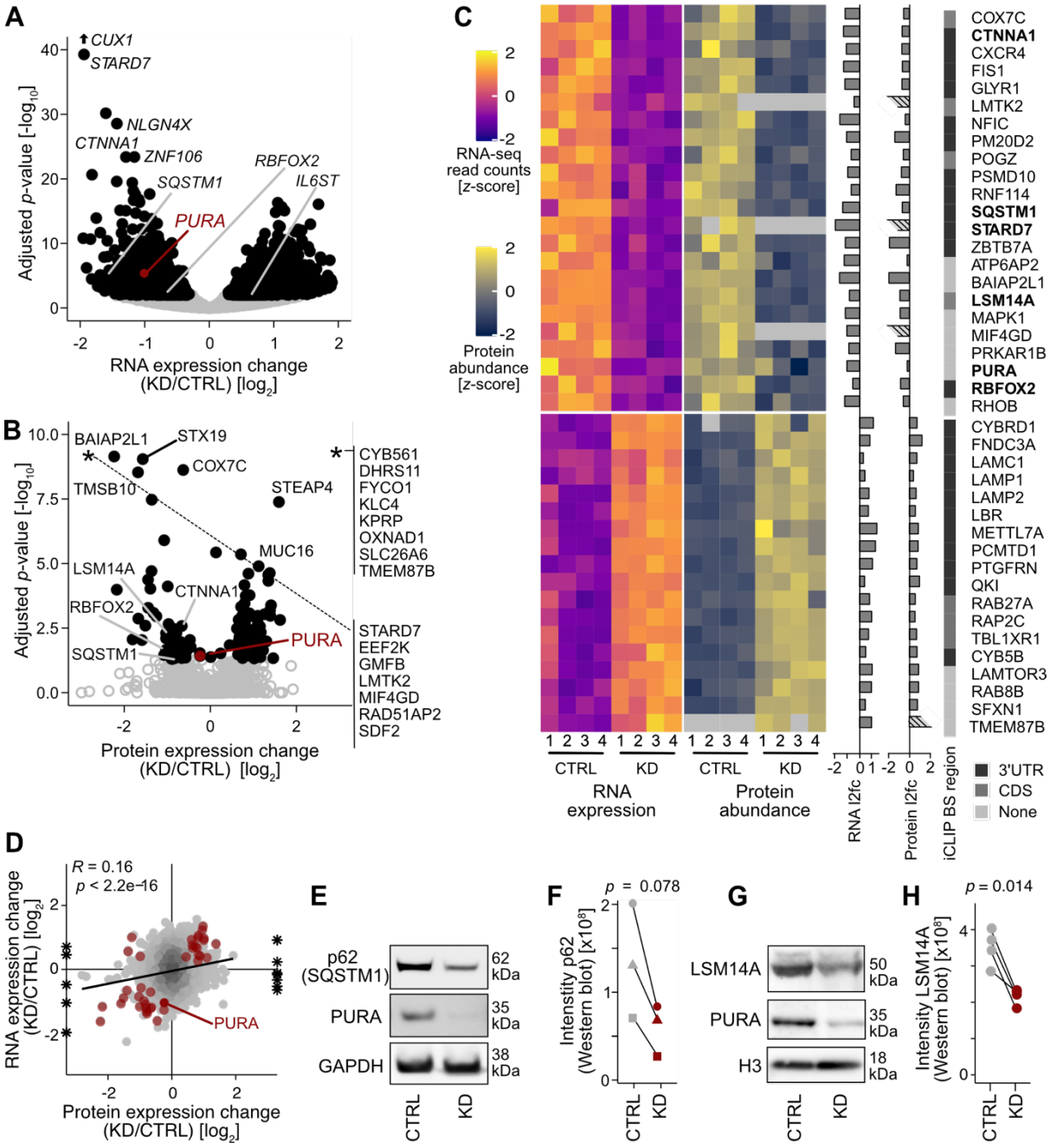

Figure 1. Loss of PURA results in widespread changes in gene and protein expression. (A) Differentially expressed genes in *PURA* knockdown (KD) vs. control (CTRL) conditions are shown in a volcano plot giving the RNA expression change [$\log_2$] against the significance of these changes as adjusted *P* value [$\log_{10}$] (black: 2,436 significantly changing genes with FDR < 0.01). **(B)** Changes in protein abundance upon *PURA* KD are shown in a volcano plot giving the protein expression change [$\log_2$] against the significance of these changes as adjusted *P* value [$\log_{10}$] (black: 143 significantly changing proteins with FDR < 0.05). **(C)** 41 genes are consistently regulated by PURA on RNA and protein levels. Heatmap shows *z*-score-normalized RNA-seq reads and peptide counts of all replicates with associated $\log_2$-transformed fold change (l2fc) and PURA-bound transcript region (iCLIP). Shading indicates |l2fc| > 2. **(D)** Protein abundances follow RNA levels. Scatter plot compares

changes in RNA (y-axis) and protein abundance (x-axis) for 4,911 genes that were measured in both assays. Targets with significant changes in both data sets (n = 45) are marked in red. Linear regression line and Pearson correlation coefficient with corresponding *P* value are given. Infinite l2fc values from protein measurements are shown as asterisks. **(E-H)** SQSTM1 and LSM14A levels are significantly reduced upon *PURA* KD. Representative Western blots (E, G) and quantifications of replicates (F, H) are shown for SQSTM1 and LSM14A in CTRL and *PURA* KD conditions. *PURA* KD was confirmed by detection with anti-PURA^{12D11}. GAPDH and H3 served as loading controls for SQSTM1 and LSM14A respectively (n = 3 [SQSTM1] or 4 [LSM14A] biological replicates, paired Student's *t*-test).

PURA localizes to the cytoplasm in multiple cell lines

The impact of PURA on gene expression could potentially arise from different cellular mechanisms. To understand at which level PURA regulates its targets, we investigated the subcellular localization of PURA in different cell lines. To ensure reliable detection, we raised a monoclonal antibody against the human PURA protein. In the light of earlier studies in which antibodies cross-reacted between PURA and its close paralog PURB (18), we selected a peptide in the unstructured linker region between PUR repeats I and II as epitope (**Supplementary Figure 3A**). This sequence differs between PURA and its close paralog PURB (**Supplementary Figure 3B**) but not from the mouse PURA paralog. We confirmed in Western blot experiments that the obtained in-house antibody (anti-PURA^{12D11}) specifically recognized PURA and showed no cross-reactivity against PURB (**Supplementary Figure 3C**).

Using anti-PURA^{12D11} in immunofluorescence staining, we assessed the localization of the endogenous PURA protein in five immortalized cell lines from a diverse set of tissue origins including epithelial, neuronal, kidney, lung and liver tissue. We found that in all cell lines, PURA located predominantly in the cytoplasm (**Figure 2**). This observation suggests that PURA is primarily involved in cytoplasmic regulatory processes and less likely to bind DNA or regulate nuclear events. Therefore, we decided to elucidate PURA's cytoplasmic function further.

PURA acts as a global cytoplasmic RNA-binding protein

Since PURA was previously found to bind RNA *in vitro* (24), we next investigated whether PURA directly binds to the RNAs it regulates. We therefore performed individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) experiments, which allow to map the RNA binding profile of an RBP in living cells (38,55). Endogenous PURA was crosslinked to its target RNAs in HeLa cells using UV radiation (254 nm) and subsequently immunoprecipitated with anti-PURA^{12D11} (**Figure 3A, B**). In total, we obtained more than 20 million unique PURA crosslink events from four biological replicates (**Supplementary Table 3**).

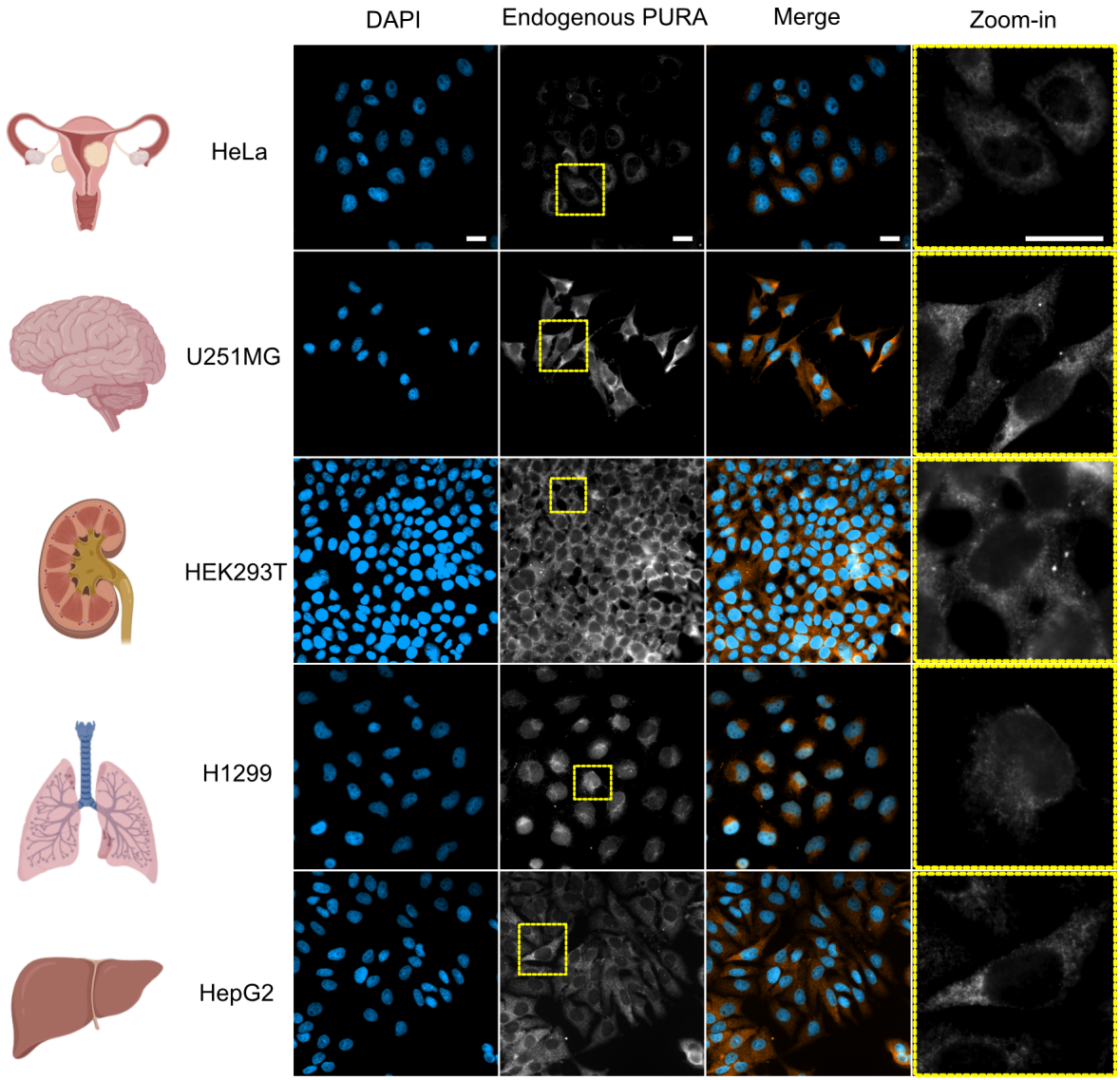

Figure 2. PURA is predominantly located in the cytoplasm. Immortalized cell lines from different tissue origins were analyzed for PURA's subcellular localization by immunofluorescence staining. Cells were stained with DAPI (blue) and anti-PURA^{12D11} (white). In the merged image, PURA staining is shown in orange for better visualization. Scale bars, 20 µm.

Assessing the global distribution of crosslink events before binding site definition, we observed that only a small fraction of PURA crosslink events occurred within introns (**Supplementary Figure 4A**). The relative depletion of introns in the iCLIP experiments suggests that PURA predominantly binds mature RNA, consistent with cytoplasmic mRNA binding.

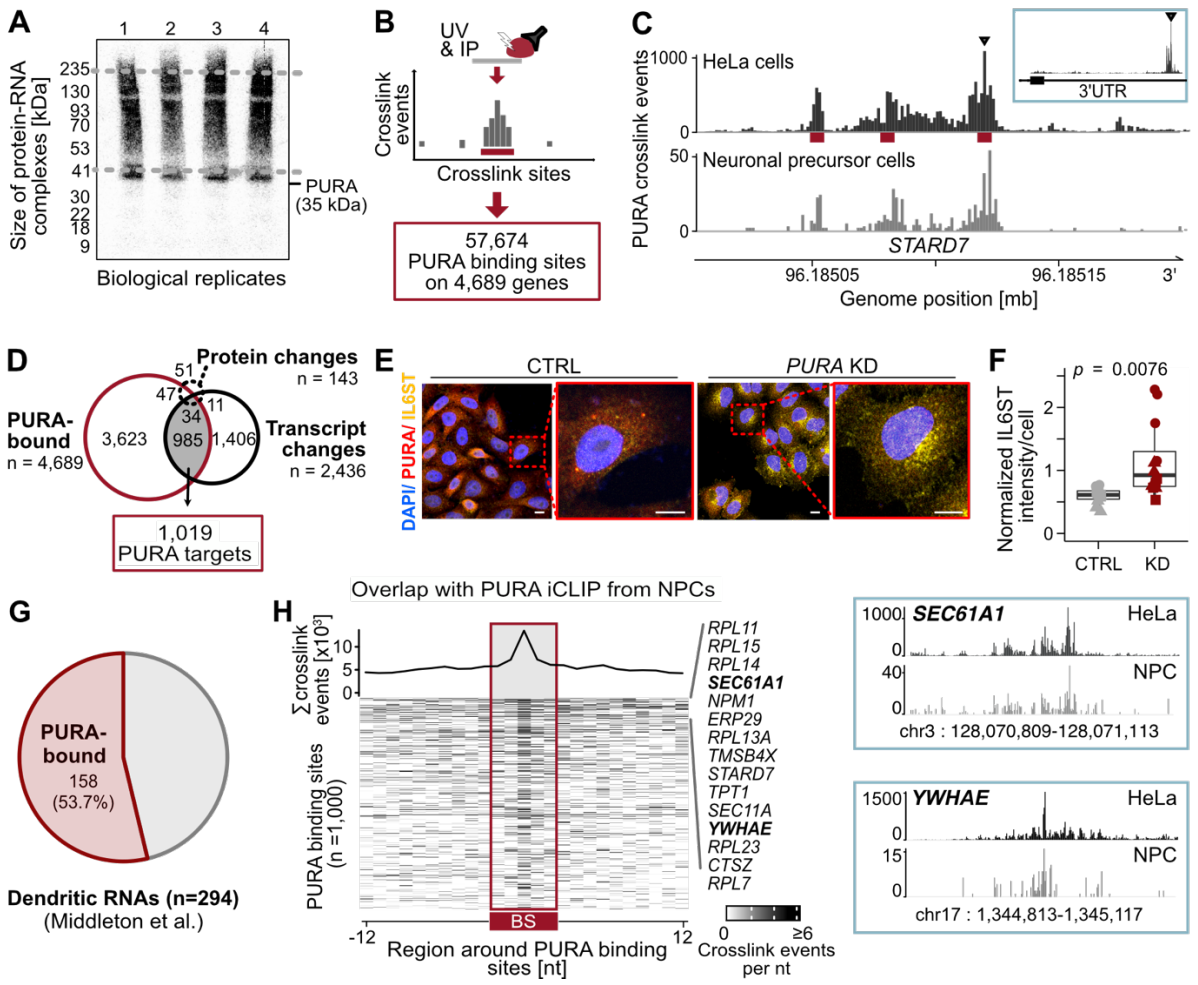

Figure 3. PURA globally binds RNAs via specific binding sites and regulates a subset of its target transcripts. (A) Autoradiogramm with radiolabeled RNA after immunoprecipitation of endogenous PURA (anti-PURA^{12D11}) from HeLa cell replicates shows crosslinked PURA-RNA complexes. Co-purified RNA fragments of variable lengths result in a smear upstream of the expected molecular weight of PURA (35 kDa). Four biological replicates are shown. **(B)** Schematic depiction of iCLIP experiment that allowed for the identification of 57,674 binding sites for endogenous PURA in HeLa cells. **(C)** Genome browser view shows the pileup of crosslink events of endogenous PURA from HeLa cells (top) and neural progenitor cells (bottom) at binding sites (red) in the 3'UTR of the *STARD7* transcript. **(D)** Venn diagram depicts overlap of PURA binding in iCLIP experiments (n = 4,689) with genes regulated at the RNA (n = 2,436) or protein level (n = 143). 1,019 PURA-bound transcripts change abundance upon *PURA* KD (odds ratio 4.14 [3.78, 4.52] (95% confidence interval [CI]), *P* value < 2.2e-16, Fisher`s exact test). **(E)** *PURA* KD leads to increased IL6ST expression and intracellular foci formation in HeLa cells. Immunofluorescence staining of PURA (anti-PURA^{12D11}, red) and IL6ST (green), as well as DAPI staining (blue) in CTRL and *PURA* KD HeLa cells. Scale bars, 10 µm. **(F)** Quantification of IL6ST-positive foci in CTRL and *PURA* KD HeLa cells (n = 3 biological replicates with 4 technical replicates each, Student`s *t*-test). **(G)** Pie chart shows overlap of PURA-bound transcripts with dendritically localized RNAs identified in at least three out of

five independent studies that are summarized in (47) (n = 294). **(H)** PURA crosslink events from neural progenitor cells (NPCs) enrich in PURA binding sites identified in HeLa cells. Metaprofile (top) and heatmap (bottom) display crosslink events per nucleotide of endogenous PURA in NPCs for 1,000 binding sites with highest signal in the NPC iCLIP data. Host gene names are given for top 17 binding sites. Genome browser views (right) illustrate exemplary binding sites on the *SEC61A1* and *YWHAE* transcripts in the HeLa and NPC iCLIP data.

We defined PURA binding sites from the iCLIP crosslink events by peak calling and merging adjacent peaks into binding sites of 5-nt width. This approach allowed us to identify 57,674 PURA binding sites in the transcripts of 4,689 genes (supported by at least 2 of the 4 replicates, **Figure 3B, C, Supplementary Figure 4, Supplementary Table 4**). The strength of PURA binding at these sites, measured as relative number of crosslink events inside the binding site, was highly reproducible between replicates (Pearson correlation coefficient [R] 0.24 – 0.71, *P* value < 2 x10^{-16}, **Supplementary Figure 4B**).

In a control experiment, we overexpressed FLAG-tagged PURA in HeLa cells and confirmed that comparable profiles were obtained from immunoprecipitation with either anti-PURA^{12D11} or an anti-FLAG antibody (**Supplementary Figure 5A**), underlining the specificity of our anti-PURA^{12D11} antibody. The PURA crosslink events from these two experiments accumulated in the defined binding sites similarly to those of the endogenous iCLIP experiment (**Supplementary Figure 5B-E**). However, the PURA overexpression led to an increased occurrence of crosslink events outside the binding sites thereby lowering the signal-noise ratio. This indicated unspecific RNA binding by the unphysiologically abundant ectopic PURA. We therefore decided to focus on the RNA binding profiles detected for the endogenous PURA protein.

Out of 41 proteins whose levels consistently changed at both mRNA and protein level in response to *PURA* KD, 33 were independently identified by iCLIP (80.5%; odds ratio 0.024 [0.02, 0.04], 95% confidence interval [CI], *P* value < 2.2e-16, Fisher`s exact test; **Figure 1C**). Since RNA-seq is much more sensitive than shot-gun proteomics and thereby also detects less abundant targets, we extended the list of putative direct targets to all 1,019 transcripts that were directly bound by PURA and significantly changed in RNA abundance upon *PURA* KD (odds ratio 4.14 [3.78, 4.52], 95% CI, *P* value < 2.2e-16, Fisher`s exact test; **Figure 3D**). To assess a representative target that did not pass the threshold for a reliable quantification in our proteomics data, we analyzed interleukin-6 receptor subunit beta (IL6ST) using fluorescence microscopy. We observed an increased signal intensity and accumulation of IL6ST in the perinuclear region upon *PURA* KD in both HeLa and normal human dermal fibroblast (NHDF) cells, supporting a role of PURA in IL6ST protein expression (**Figure 3E, F, Supplementary Figures 4C**). This indicates that our approach of an extended target list based on iCLIP and RNA-seq experiments is valid. Further targets with potential functional connections to brain development included the *CUX1*, *NLGN4X* and

ZNF106 mRNAs. The direct binding to these transcripts connects PURA to factors involved in neuronal signaling and neurological diseases (56-60).

PURA was repeatedly found to be a component of neural transport granules (16-20). Of note, even though HeLa cells are of non-neural origin, more than half of previously reported dendritically localized RNAs showed direct PURA binding in our iCLIP data (**Figure 3G, Supplementary Figure 6A, B**) (47). To test whether PURA shows similar RNA binding in neural cells, we used human induced pluripotent stem cells and differentiated them into neural progenitor cells (NPCs). We validated the NPC state by observing rosette morphology of the cells and testing the expression of several known markers of NPC differentiation (**Supplementary Figure 6C, D**). Although we obtained only a limited iCLIP dataset from the scarce material, the crosslink events for endogenous PURA in NPCs enriched within the PURA binding sites identified from HeLa cells, indicating a similar RNA binding behavior in both cell types (**Figure 3C, H**). Interestingly, the bound targets in NPCs included several neurodevelopmentally relevant transcripts, such as *RBFOX2* and *CTNNA1*, which were also regulated by *PURA* KD (**Figure 1A-C**).

Altogether, we identified more than 50,000 binding sites and around 1,000 regulated target transcripts of endogenous PURA in HeLa cells. The observed RNA binding pattern was independent of the antibody used and comparable between HeLa cells and NPCs. Hence, these results indicate that the binding sites identified in HeLa cells reflect the *bona fide* binding of PURA in living cells.

PURA binds in coding sequences and 3'UTRs

In total, 47.1% of all expressed genes (transcripts per million [TPM] > 10) harbored at least one PURA binding site, underlining the widespread binding of PURA throughout the transcriptome (**Supplementary Figure 4D, E**). Among the 4,689 PURA-bound transcripts, we predominantly found protein-coding mRNAs (n = 4,478, 95.5%). These encoded for several functional categories that were also enriched in the PURA-regulated targets, including nervous system development, mitochondrial functions, membrane trafficking and RNA metabolism (**Supplementary Figure 4F, G**). PURA also showed binding to 99 long non-coding RNAs (lncRNAs; **Figure 4A**), including the previously reported PURA target lncRNAs *RN7SL1* (20,61), which has been implicated in transcriptional control, as well as *NEAT1* and *MALAT1*, two architectural RNAs forming distinct nuclear bodies (62) (**Supplementary Figure 4H**).

Within the protein-coding transcripts, the PURA binding sites were almost equally distributed between the coding sequences (CDS, n = 25,692) and the 3' untranslated regions (3'UTRs, n = 25,320; **Figure 4B, C**), which together accounted for 86.4% of all binding sites. In contrast, 5'UTRs and intronic regions were almost completely depleted of PURA binding sites (3.3% 5'UTRs, 1.8% introns). Transcripts with PURA binding in the 3'UTR or CDS showed the same overall changes in expression upon *PURA* KD (**Supplementary Figure 7**). The 41 targets which were changing in

abundance at both the RNA and protein levels were four times more often bound by PURA in the 3'UTR compared to the CDS (relative to all PURA-bound genes changing at the RNA level, odds ratio 4.07 [2.08, 7.95], 95% CI, *P* value = 1.81e-05, Fisher`s exact test, **Figure 1C**). This observation suggests that 3'UTR binding sites convey more potent regulation than CDS binding sites. Within the 3'UTRs, PURA generally displayed one or more defined peaks of binding, as seen in *STARD7* mRNA (**Figure 3C, 4D**).

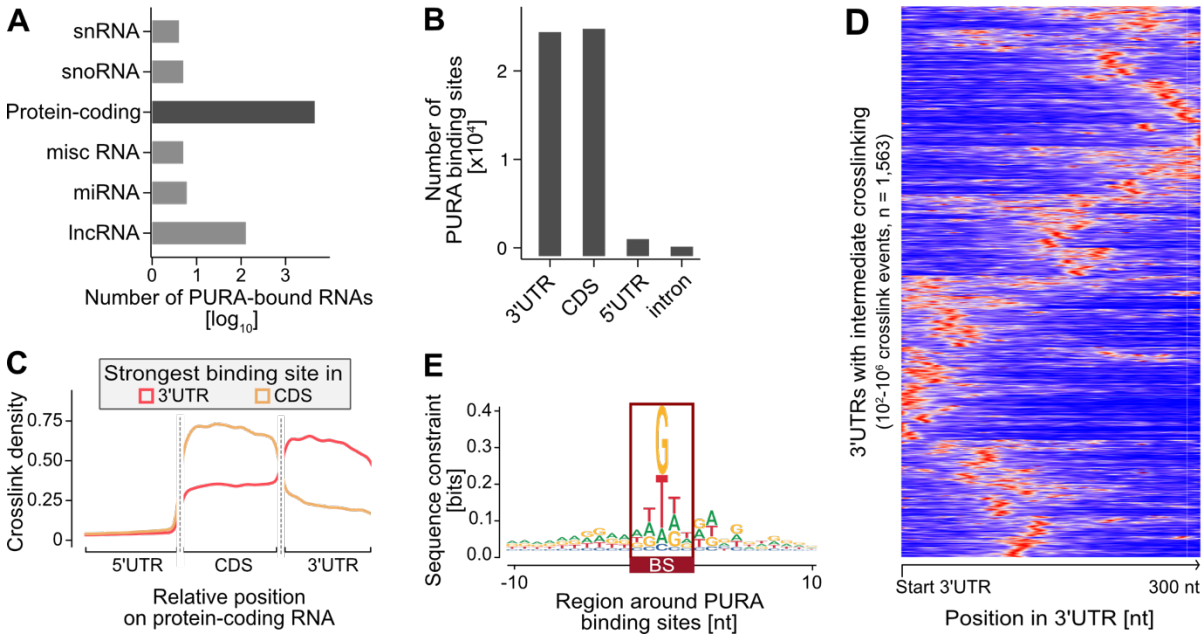

Figure 4. PURA binds to CDS and 3'UTRs of mRNAs. (A) Transcript types bound by endogenous PURA in HeLa cells show PURA's preference for binding to protein-coding RNAs. Bars indicate the number of bound RNAs of each type on a $\log_{10}$ scale. **(B)** PURA binding sites per transcript region in protein-coding genes. Most binding sites are located in 3'UTR or CDS. Bars show total number of binding sites per region. **(C)** Metaprofile of PURA crosslink events along transcripts with strongest PURA binding site in the 3'UTR (red) or CDS (beige). Smoothed density of crosslink events is shown against scaled positions within the respective region. **(D)** PURA displays one or more distinct peaks in the 3'UTR of bound transcripts. Heatmap shows normalized, smoothed distribution of PURA crosslink events in the first 300 nt of 1,563 exemplary 3'UTRs with intermediate crosslink signal (10^2-10^6 crosslink events per window). **(E)** Sequence logo of the nucleotide composition in a 21-nt window around PURA binding sites reveals purine enrichment. Y-axis shows sequences constraint in bits.

Mapping of the binding sites enabled us to revisit the sequence and structural preferences of PURA's RNA binding. We therefore predicted the local folding probability in a 500-nt window around the binding sites using RNAplfold (44). This analysis proposes a propensity of PURA's RNA binding sites to be single-stranded (**Supplementary Figure 8A**), contrasting past reports about PURA's double-stranded nucleic acid binding preference and unwinding ability (23,63). The increased single-strandedness coincided with a strong enrichment for purines (A and G) within and around the PURA binding sites, such that 47.0% of binding sites were centered on a guanine and 60.8% enclosed three or more purines (**Figure 4E, Supplementary Figure 8B**). More globally, the binding sites and their surrounding sequences were enriched for *k*-mers comprising purine combinations like GAAGA and AAGAA (**Supplementary Figure 8C**). We also observed an upstream enrichment for TTTTT which likely reflects the UV crosslinking bias towards uridines (64).

Together, these observations underline a specificity of PURA to bind single-stranded, purine-rich regions, primarily in CDS and 3'UTRs.

PURA binds to P-body transcripts and localizes to P-bodies

Under unstressed conditions, we observed the PURA immunofluorescence signal in cytoplasmic foci (**Figure 2**). To obtain a first indication what kind of foci these might be, we overlapped the PURA-bound RNAs with transcripts that were enriched in processing (P-)bodies (45) or stress granules (SGs) (46). In line with the previously reported PURA localization to SGs (21), more than half of the SG-enriched transcripts were bound by PURA (51.9%, *P* value < 2.2×10^{-16}, Fisher`s exact test; **Figure 5A**). Surprisingly, however, we detected an almost equally strong overlap with transcripts enriched in P-bodies (45) (45%, *P* value < $2.2 \times 10^{-16}$, Fisher`s exact test; **Figure 5A**), indicating that PURA might be linked to these granules as well.

To test for a direct association, we used immunofluorescence staining to overlay PURA with the P-body marker proteins DCP1A and LSM14A. Indeed, we identified a strong colocalization of PURA with both marker proteins (**Figure 5B, C, Supplementary Figure 9**). This indicates that under basal growth conditions, PURA mainly localizes to P-bodies in HeLa cells.

PURA depletion reduces the number of P-bodies per cell

Based on the strong association of PURA with P-bodies, we wondered whether PURA binding directly affects the abundance of the P-body-associated RNAs. Intriguingly, additional bioinformatics analysis revealed that P-body-enriched transcripts were selectively stabilized by experimental downregulation of *PURA* (**Figure 5D**). Therefore, we hypothesized that P-bodies might be impaired in the absence of PURA. In contrast, SG-associated RNAs showed an inverse behavior, and the remaining RNAs were not affected.

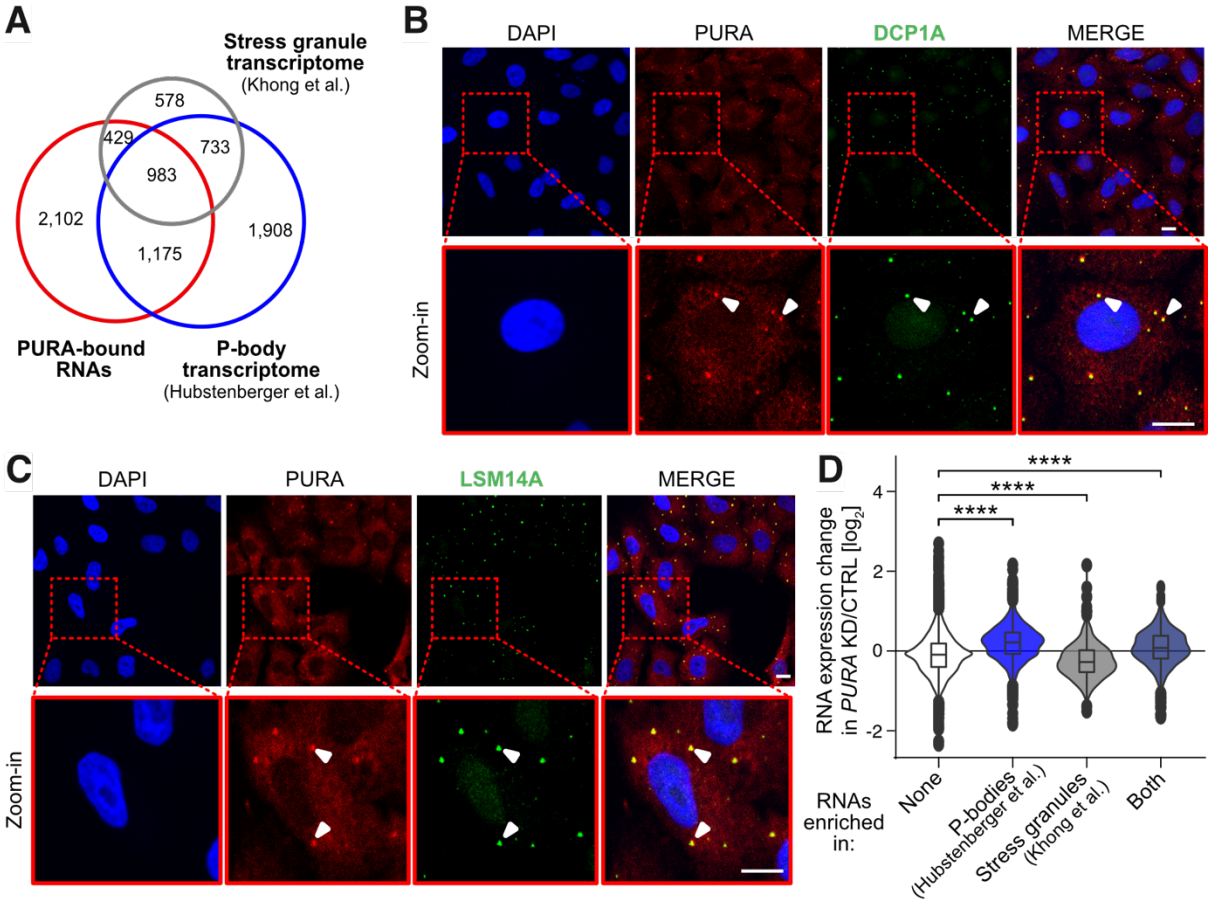

Figure 5. PURA localizes to P-bodies and P-body RNAs are upregulated upon *PURA* KD. (A) Venn diagram depicts overlap of PURA-bound RNAs (red) with the transcriptomes of P-bodies (blue, n = 4,799, odds ratio 4.06 [3.78, 4.37] (95% CI), *P* value < 2.2x10e-16, Fisher`s exact test) (45) and SGs (grey, n = 2,416, odds ratio 12.08 [11.07, 13.1], *P* value < 2.2x10e-16, Fisher`s exact test) (46). **(B)** Confocal micrographs of PURA (red), P-body marker DCP1A (green) and colocalized signal (yellow). Nuclei were stained with DAPI (blue). Arrowheads indicate examples of P-bodies with overlapping staining. Scale bars, 10 µm. **(C)** Staining as in (B) using LSM14A as P-body marker (green). Colocalization (yellow) of PURA (red) and LSM14A (green). Scale bars, 10 µm. **(D)** P-body RNAs are upregulated in *PURA* KD, while SG RNAs are downregulated. Violin plot of RNA expression changes (log_2 fold change, *PURA* KD/CTRL) for RNAs enriched in P-bodies (blue), SGs (grey) or both (blue-grey) *versus* all other RNAs (white) (*P* value < 0.0001, Student`s *t*-test).

In order to focus on PURA's role in the physiological context of unstressed cells, we took a closer look at P-body core components and were intrigued to find that PURA directly bound to *LSM14A* mRNA (**Figure 6A**). LSM14A was already identified in this study to be strongly downregulated upon *PURA* KD at the transcript as well as protein level (**Figure 1C**). This was orthogonally validated by a 60% reduction in LSM14A protein in Western blot analysis after *PURA* KD, relative to control HeLa cells (*P* value = 0.014, Student's *t*-test, **Figure 1G, H**).

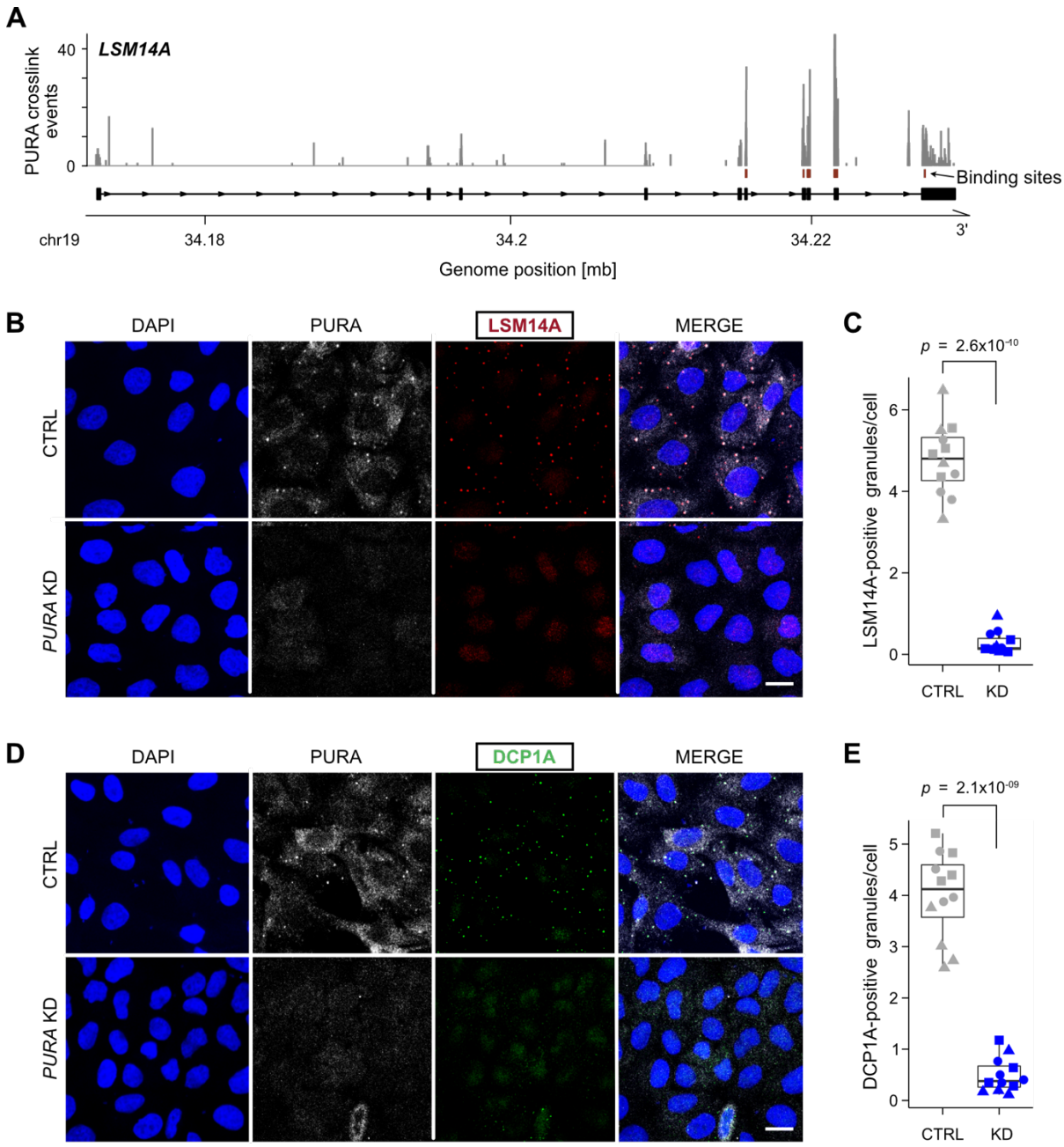

Figure 6. *PURA* KD leads to a loss of P-bodies. (A) Genome browser view shows the pileup of crosslink events of endogenous PURA from HeLa cells at binding sites (red) in the *LSM14A* transcript. **(B)** Immunofluorescence staining of PURA (white) together with the P-body protein LSM14A (red) and DAPI (blue) as a nuclear stain in CTRL (top) and *PURA* KD (bottom) HeLa cells. Scale bars, 20 µm. **(C)** ImageJ analysis of LSM14A-positive granules on approximately 400 cells per replicate (n = 3) comparing WT and *PURA* KD conditions. Shown is the number of LSM14A-positive granules per cell for all replicates. 3 biological replicates are distinguished by dot shape, with each containing 4 technical replicates. *P* value from Student's *t*-test on all replicates. **(D)** Immunofluorescence staining of PURA (white) together with the P-body marker DCP1A (green) and DAPI (blue) as a nuclear stain in CTRL (top) and *PURA* KD (bottom) HeLa cells. Scale bars, 20 µm. **(E)** ImageJ analysis of DCP1A-positive

granules per cell on approximately 400 cells per replicate (n=3) comparing WT and *PURA* KD conditions. Visualization of replicates and *P* value as in (C).

Strikingly, when analyzing LSM14A-positive granules in *PURA* KD cells using immunofluorescence staining, we observed a drastic decrease in granule numbers (**Figure 6B, C**). Although several proteins modulate the appearance of P-bodies, LSM14A is one of a handful of currently known factors that have been shown to be essential for P-body formation (65). To test whether the PURA-dependent reduction of LSM14A globally impairs P-bodies, we visualized their occurrence using immunofluorescence staining for the P-body marker DCP1A. Remarkably, *PURA* KD led to a dramatic reduction in P-body number in HeLa cells and normal human dermal fibroblasts (NHDF; **Figure 6D, E, Supplementary Figure 10**). Importantly, in contrast to LSM14A, the protein levels of the P-body marker DCP1A remained unchanged upon *PURA* KD (**Supplementary Figure 9A**), suggesting that PURA acts selectively by modulating the LSM14A levels in human mesenchymal and epithelial cells.

Together, our data suggest that PURA regulates *LSM14A* mRNA and concomitant protein levels and the abundance of P-bodies. Intriguingly, PURA depletion leads to an almost complete loss of P-bodies. Given the important role of P-bodies in cellular RNA surveillance and maintenance, our findings reveal a putative link between the neuronal PURA Syndrome and the cytoplasmic control of mRNA levels and P-body condensation.

Discussion

PURA is a cytoplasmic RNA-binding protein

The PURA protein was initially described as a transcription factor with binding preferences for purine-rich single-stranded DNA regions (13,14). While this might be one of its functions, PURA was subsequently shown to associate with cellular RNAs and to bind single-stranded DNA and RNA sequences *in vitro* in the same way and with similar affinities (23). In this study, we developed a paralog-specific antibody to demonstrate that the vast majority of PURA protein resides in the cytoplasm in a range of different cell lines. This is in line with more recent studies describing PURA in cytoplasmic RBP granules (17-19,21). It also underlines the need to study PURA's function in processes other than DNA regulation.

To date, a physiological RNA binding function of PURA had only been shown for a few specific target RNAs like the lncRNA *RN7SL1* (20,61) or the circular RNAs circSamD4 (66) and circCwc27 (67). Most other studies were based on either *in vitro* binding assays, colocalization studies or enrichment in membrane-less organelles (reviewed in (3). Here, we provide the first transcriptome-wide analysis of PURA's RNA binding in living cells. Our iCLIP data show that PURA binds to almost half of all expressed mRNAs, thereby bearing the potential to regulate a multitude of different cellular processes (see below).

In line with the first description of PURA as purine-rich binding element (13,14,68,69), we identified a binding preference for purine-rich RNA regions. Similar to previous studies (23,24), we did not observe a distinct sequence motif. A degenerate sequence preference is characteristic for many RBPs (70) and might suggest that PURA's RNA recognition is driven by additional factors, such as RNA secondary structure or co-factor binding.

PURA regulates mRNAs associated with mitochondria, neural transport granules and autophagy

Our integration of multiscale omics data revealed widespread changes in expression upon loss of PURA in cells. Most prominently, we identify almost 2,500 dysregulated genes with significantly changed transcript levels upon *PURA* KD. Despite the inherent detection limits of mass spectrometry, we find that the proteome largely follows the transcript-level changes, indicating that translational efficiency is not globally altered. We hence conclude that most likely not translation but other RNA-regulatory mechanisms are primarily affected in the *PURA* KD cells. These may include RNA stability or degradation as well as RNA transport and subcellular localization, among others.

Integration with transcriptome-wide RNA-binding events revealed over 1,000 putative direct targets of PURA. The molecular roles of these targets link PURA to various processes including interleukin signaling, autophagy, mitochondrial homeostasis and neurodevelopment. One interesting target that was directly bound by PURA and upregulated upon *PURA* KD is the *IL6ST* mRNA encoding interleukin-6 cytokine family

signal transducer (IL6ST; also known as gp130). IL6ST is a signal transducer shared by many cytokines, including interleukin 6, ciliary neurotrophic factor, leukemia inhibitory factor, and oncostatin M (71). In patients with PURA Syndrome, an overexpression of IL6ST upon *PURA* haploinsufficiency may contribute to a compromised immune response that may be responsible for the severe infections reported by affected families (personal communication).

Another direct PURA target with potential links to the phenotype of patients is the *SQSTM1* mRNA encoding the autophagy receptor Sequestome 1 (SQSTM1; also known as p62). In a zebrafish model for C9 ALS, RNA-triggered toxicity could be rescued through ectopic overexpression of PURA, and this rescue could be related to the upregulation of SQSTM1 by PURA (54). Interestingly, we observed that PURA directly binds to *SQSTM1* mRNA and multiple other targets from the autophagy pathway. This suggests that PURA may act as a modulator of autophagy in human cells, which has long been proposed as a neuroprotective mechanism in cellular surveillance (72).

Furthermore, several of the newly identified PURA targets that are downregulated upon *PURA* KD are involved in mitochondrial functions, including the *STARD7*, *COX7C* and *FIS1* mRNAs. Indeed, mitochondrial respiration was already described to be impaired upon *PURA* depletion (73). Since mitochondrial dysfunctions are associated with various neuronal disorders (74), it is tempting to speculate that mitochondrial defects caused by *PURA* haploinsufficiency may contribute to the neurodevelopmental phenotype of PURA Syndrome patients. This may converge with direct effects of PURA on neurodevelopmentally relevant targets. In line with this, PURA has been associated with neuronal RNA transport in different studies in mouse and human cells. For instance, a misdistribution of the dendritic marker protein MAP2 was reported in one *PURA* KO mouse model (8). The direct PURA targets identified in our study will be relevant in establishing the neuronal and neurodevelopmental functions of PURA in the future.

PURA depletion leads to an upregulation of P-body-associated transcripts and a complete loss of P-bodies

Cytoplasmic RBPs are known to play important roles in the formation of cytoplasmic membrane-less organelles such as P-bodies and stress granules (75). While omics studies revealed a broad view on some of their transcriptomes and proteomes (45,46), only a small number of proteins are considered essential for P-body formation, including DDX6, 4E-T and LSM14A (65).

Here, we show that PURA is enriched in P-bodies and binds to 45% of the P-body transcriptome. Interestingly, upon *PURA* KD P-body-enriched transcripts are globally upregulated. In addition, we find that depletion of PURA in different human cell lines leads to a loss of P-bodies. We can link this observation to PURA binding to *LSM14A* RNA and down-regulation of its RNA and protein product. Since the LSM14A protein is an essential P-body factor, its PURA-dependent depletion most likely triggers the observed loss of P-bodies.

P-bodies were initially proposed as a location of RNA depletion (76). Seemingly in line with this, we find RNAs of the P-body transcriptome globally upregulated upon *PURA* KD and P-body loss. However, more recently P-bodies were found to act in RNA storage rather than depletion (45). Moreover, SG and P-bodies were shown to partially share their proteomes and transcriptomes (45) and are suggested to dynamically interact with each other (77,78). We now show that PURA is one of the proteins shared by both granule types. The transcriptomic changes in *PURA* KD show a differential behavior for P-body versus SG RNAs, further suggesting a role of PURA in the dynamic composition of cytoplasmic phase-separated condensates.

Implications for understanding the etiology of the PURA Syndrome

As PURA Syndrome is described to result from a haploinsufficiency and thereby lowered cellular PURA levels, our findings on PURA-dependent, LSM14A-mediated P-body formation could directly relate to a molecular dysfunction in patients. This indicates that altered granule dynamics could play a role in the patho-mechanism of PURA Syndrome. Similarly, the observed changes in the sequestome degradation machinery and the mitochondrial proteome may be in functional correlation to symptoms of PURA Syndrome patients. Among the most downregulated targets was also the transcription factor CUX1 that is involved in the control of neuronal differentiation (51). This and other factors allow for connecting the neurodevelopmental delay of patients to our observations of misregulated factors.

While the above-mentioned potential links are highly suggestive, it should be considered that the spectrum of symptoms of PURA Syndrome patients is broad and variable (6,7). Furthermore, the understanding of the medical condition of patients has only begun to be investigated, rendering it very difficult to directly and safely correlate molecular PURA-dependent events to symptoms of patients with haploinsufficiency in the *PURA* gene. This consideration is even more important as not all molecular changes are likely to have an impact on the symptoms observed in humans. While this study opened the door for an unbiased and systematic understanding of molecular PURA-dependent pathways, it will require more research on both ends, clinical and non-clinical, to convincingly connect our findings with the symptoms in PURA Syndrome patients.

Conflict of interest

The authors confirm that they have no conflict of interest to declare.

Data availability

All sequencing data is available in the Gene Expression Omnibus (GEO) under the SuperSeries accession number GSE193905. The collection includes the RNA-seq data from *PURA* KD and CTRL HeLa cells (GSE193900) as well as the iCLIP data for endogenous PURA (anti-PURA^{12D11} antibody) in HeLa cells (GSE193901) and NPC (GSE193902) as well as for overexpressed FLAG-PURA in HeLa cells precipitated with anti-PURA^{12D11} (GSE193904) and anti-FLAG antibody (GSE193903).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (79) partner repository with the dataset identifier PXD030266.

Scripts used to process the files are accessible under the GitHub repository located at: https://github.com/ZarnackGroup/Publications/blob/main/Molitor_et_al_2022/

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) via the Research Unit FOR2333 to K.Z., J.K. and D.N. (ZA 881/3-1 to K.Z., KO 4566/5-1 to J.K., and Ni1110/5-1 and 7-1 to D.N.), a Heisenberg professorship to P.F.-P. (project number 398707781), a research grant to D.T. (TE012-2/2), the SFB902 subproject B13 to K.Z, and the SPP1935 to D.N. (Ni1110/8-1). This work was also supported by the Science Award of the Care-for-Rare Foundation to D.N..

Acknowledgements

The authors would like to thank the members of the Niessing and Zarnack labs for support and discussion. We thank Saskia Hutten for help with immunofluorescence experiments, Annika Niedner-Boblenz and Nadine Körtel for initial help in setting up iCLIP experiments and Stefan Krebs for help in planning and performing the sequencing runs. We also thank Jana Tretter, Verena Kirchner, Elke Pertler, Deniz Yavru, Hania Ahmed, Mirko Brüggemann, You Zhou, Robert Janowski and Vera Roman for advice and technical assistance. We would like to thank Ruth Brack-Werner for providing us with additional cell lines, Robert Schneider for using their BioAnalyzer and the confocal microscopy facility at HMGU for using their microscopes. We also thank Eran Hornstein for helpful comments and discussions.

Author contributions

L.M. performed iCLIP, *PURA* knockdown, immunofluorescence (IF) and Western blot experiments, validated the anti-PURA^{12D11} antibody and analyzed data. M.K. performed most bioinformatics analyses of iCLIP, RNA-seq and shotgun proteomics experiments. S.B. performed and quantified IF experiments of PURA and P-body marker proteins and IL6ST as well as Western blots. J.M.-P. prepared, measured and analyzed shot-gun proteomics samples. N.S. performed IF experiments for the PURA cell atlas. E.R. differentiated hiPSCs to NPCs and validated NPCs for subsequent iCLIP experiments. D.T. performed Western blot experiments for SQSTM1. A.P. reprogrammed and validated the HMGU12 hiPS cell line. M.P. assisted with IF experiments for P-body marker DCP1A. S.R. performed RNA extraction and library preparation for RNA-seq samples. H.B. planed and supervised sequencing of iCLIP and RNA-seq libraries. A.B. performed initial analysis of iCLIP data sets. S.B. performed Western blots and sequence alignments. The antibody facility headed by R.F. generated the in-house anti-PURA^{12D11} antibody. S.H. planed and supervised shot-gun proteomics experiments. M.D. planed and supervised reprogramming and differentiation experiments of hiPSCs. P.F.-P. planed and supervised the validation of autophagy marker SQSTM1. J.K. planed and supervised iCLIP experiments. Study was designed by L.M., M.K., K.Z. and D.N.. K.Z. supervised most of the bioinformatics analyses. D.N. supervised most of the experimental work. L.M., M.K., K.Z. and D.N. wrote the manuscript with comments from all co-authors.

References

1. Gebauer, F., Schwarzl, T., Valcarcel, J. and Hentze, M.W. (2021) RNA-binding proteins in human genetic disease. *Nat Rev Genet*, **22**, 185-198.
2. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat Rev Genet*, **15**, 829-845.
3. Molitor, L., Bacher, S., Burczyk, S. and Niessing, D. (2021) The Molecular Function of PURA and Its Implications in Neurological Diseases. *Frontiers in Genetics*, **12**.
4. Hunt, D., Leventer, R.J., Simons, C., Taft, R., Swoboda, K.J., Gawne-Cain, M., study, D.D.D., Magee, A.C., Turnpenny, P.D. and Baralle, D. (2014) Whole exome sequencing in family trios reveals de novo mutations in PURA as a cause of severe neurodevelopmental delay and learning disability. *J Med Genet*, **51**, 806-813.
5. Lalani, S.R., Zhang, J., Schaaf, C.P., Brown, C.W., Magoulas, P., Tsai, A.C., El-Gharbawy, A., Wierenga, K.J., Bartholomew, D., Fong, C.T. *et al.* (2014) Mutations in PURA Cause Profound Neonatal Hypotonia, Seizures, and Encephalopathy in 5q31.3 Microdeletion Syndrome. *Am J Hum Genet*, **95**, 579-583.
6. Reijnders, M.R.F., Janowski, R., Alvi, M., Self, J.E., van Essen, T.J., Vreeburg, M., Rouhl, R.P.W., Stevens, S.J.C., Stegmann, A.P.A., Schieving, J. *et al.* (2018) PURA syndrome: clinical delineation and genotype-phenotype study in 32 individuals with review of published literature. *J Med Genet*, **55**, 104-113.

7. Johannesen, K.M., Gardella, E., Gjerulfsen, C.E., Bayat, A., Rouhl, R.P.W., Reijnders, M., Whalen, S., Keren, B., Buratti, J., Courtin, T. *et al.* (2021) PURA-Related Developmental and Epileptic Encephalopathy: Phenotypic and Genotypic Spectrum. *Neurol Genet*, **7**, e613.
8. Hokkanen, S., Feldmann, H.M., Ding, H., Jung, C.K., Bojarski, L., Renner-Muller, I., Schuller, U., Kretzschmar, H., Wolf, E. and Herms, J. (2012) Lack of Pur-alpha alters postnatal brain development and causes megalencephaly. *Hum Mol Genet*, **21**, 473-484.
9. Khalili, K., Del Valle, L., Muralidharan, V., Gault, W.J., Darbinian, N., Otte, J., Meier, E., Johnson, E.M., Daniel, D.C., Kinoshita, Y. *et al.* (2003) Puralpha is essential for postnatal brain development and developmentally coupled cellular proliferation as revealed by genetic inactivation in the mouse. *Mol Cell Biol*, **23**, 6857-6875.
10. Deciphering Developmental Disorders, S. (2015) Large-scale discovery of novel genetic causes of developmental disorders. *Nature*, **519**, 223-228.
11. Huang, N., Lee, I., Marcotte, E.M. and Hurles, M.E. (2010) Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*, **6**, e1001154.
12. Barbe, M.F., Krueger, J.J., Loomis, R., Otte, J. and Gordon, J. (2016) Memory deficits, gait ataxia and neuronal loss in the hippocampus and cerebellum in mice that are heterozygous for Pur-alpha. *Neuroscience*, **337**, 177-190.
13. Bergemann, A.D. and Johnson, E.M. (1992) The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. *Mol Cell Biol*, **12**, 1257-1265.
14. Bergemann, A.D., Ma, Z.W. and Johnson, E.M. (1992) Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. *Mol Cell Biol*, **12**, 5673-5682.
15. Daniel, D.C. and Johnson, E.M. (2018) PURA, the gene encoding Pur-alpha, member of an ancient nucleic acid-binding protein family with mammalian neurological functions. *Gene*, **643**, 133-143.
16. El Fatimy, R., Davidovic, L., Tremblay, S., Jaglin, X., Dury, A., Robert, C., De Koninck, P. and Khandjian, E.W. (2016) Tracking the Fragile X Mental Retardation Protein in a Highly Ordered Neuronal RiboNucleoParticles Population: A Link between Stalled Polyribosomes and RNA Granules. *PLoS Genet*, **12**, e1006192.
17. Johnson, E.M., Kinoshita, Y., Weinreb, D.B., Wortman, M.J., Simon, R., Khalili, K., Winckler, B. and Gordon, J. (2006) Role of Pur alpha in targeting mRNA to sites of translation in hippocampal neuronal dendrites. *J Neurosci Res*, **83**, 929-943.
18. Kanai, Y., Dohmae, N. and Hirokawa, N. (2004) Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron*, **43**, 513-525.
19. Mitsumori, K., Takei, Y. and Hirokawa, N. (2017) Components of RNA granules affect their localization and dynamics in neuronal dendrites. *Mol Biol Cell*, **28**, 1412-1417.
20. Ohashi, S., Koike, K., Omori, A., Ichinose, S., Ohara, S., Kobayashi, S., Sato, T.A. and Anzai, K. (2002) Identification of mRNA/protein (mRNP) complexes containing Pur-alpha, mStaufen, fragile X protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. *J Biol Chem*, **277**, 37804-37810.
21. Daigle, J.G., Krishnamurthy, K., Ramesh, N., Casci, I., Monaghan, J., McAvoy, K., Godfrey, E.W., Daniel, D.C., Johnson, E.M., Monahan, Z. *et al.* (2016) Pur-alpha regulates cytoplasmic stress granule dynamics and ameliorates FUS toxicity. *Acta Neuropathol*, **131**, 605-620.

22. Janowski, R. and Niessing, D. (2020) The large family of PC4-like domains - similar folds and functions throughout all kingdoms of life. *RNA Biol*, 1-11.
23. Weber, J., Bao, H., Hartlmuller, C., Wang, Z., Windhager, A., Janowski, R., Madl, T., Jin, P. and Niessing, D. (2016) Structural basis of nucleic-acid recognition and double-strand unwinding by the essential neuronal protein Pur-alpha. *Elife*, **5**.
24. Graebsch, A., Roche, S. and Niessing, D. (2009) X-ray structure of Pur-alpha reveals a Whirly-like fold and an unusual nucleic-acid binding surface. *Proc Natl Acad Sci USA*, **106**, 18521-18526.
25. Feederle, R., Gerber, J.K., Middleton, A., Northrup, E., Kist, R., Kremmer, E. and Peters, H. (2016) Generation of Pax1/PAX1-Specific Monoclonal Antibodies. *Monoclon Antib Immunodiagn Immunother*, **35**, 259-262.
26. Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S., Simon, L., Muller, N.S., Straub, T., Brooks, K.E. *et al.* (2017) GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proc Natl Acad Sci U S A*, **114**, E9579-E9588.
27. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
28. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.
29. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*, **47**, D766-D773.
30. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
31. Wisniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Methods*, **6**, 359-362.
32. Grosche, A., Hauser, A., Lepper, M.F., Mayo, R., von Toerne, C., Merl-Pham, J. and Hauck, S.M. (2016) The Proteome of Native Adult Muller Glial Cells From Murine Retina. *Mol Cell Proteomics*, **15**, 462-480.
33. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S. and MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*, **4**, 923-925.
34. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47.
35. Edelmann, F.T., Niedner, A. and Niessing, D. (2014) Production of pure and functional RNA for in vitro reconstitution experiments. *Methods*, **65**, 333-341.
36. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, **29**, e45.
37. Reinhardt, P., Glatza, M., Hemmer, K., Tsytsyura, Y., Thiel, C.S., Hoing, S., Moritz, S., Parga, J.A., Wagner, L., Bruder, J.M. *et al.* (2013) Derivation and expansion using only small molecules of human neural progenitors for neurodegenerative disease modeling. *PLoS One*, **8**, e59252.
38. Buchbender, A., Mutter, H., Sutandy, F.X.R., Kortel, N., Hanel, H., Busch, A., Ebersberger, S. and Konig, J. (2020) Improved library preparation with the new iCLIP2 protocol. *Methods*, **178**, 33-48.

39. Roehr, J.T., Dieterich, C. and Reinert, K. (2017) Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics*, **33**, 2941-2942.
40. Busch, A., Bruggemann, M., Ebersberger, S. and Zarnack, K. (2020) iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. *Methods*, **178**, 49-62.
41. Krakau, S., Richard, H. and Marsico, A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol*, **18**, 240.
42. Heyl, F. and Backofen, R. (2021) StoatyDive: Evaluation and classification of peak profiles for sequencing data. *Gigascience*, **10**.
43. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol*, **9**, e1003118.
44. Bernhart, S.H., Hofacker, I.L. and Stadler, P.F. (2006) Local RNA base pairing probabilities in large sequences. *Bioinformatics*, **22**, 614-615.
45. Hubstenberger, A., Courel, M., Benard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.B., Munier, A., Fradet, M. *et al.* (2017) P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell*, **68**, 144-157 e145.
46. Khong, A., Matheny, T., Jain, S., Mitchell, S.F., Wheeler, J.R. and Parker, R. (2017) The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol Cell*, **68**, 808-820 e805.
47. Middleton, S.A., Eberwine, J. and Kim, J. (2019) Comprehensive catalog of dendritically localized mRNA isoforms from sub-cellular sequencing of single mouse neurons. *BMC Biol*, **17**, 5.
48. Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res*, **49**, D884-D891.
49. Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L. and Hermjakob, H. (2017) Reactome pathway analysis: a high-performance in-memory approach. *BMC bioinformatics*, **18**, 142.
50. The Gene Ontology, C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*, **47**, D330-D338.
51. Cubelos, B., Sebastian-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., Bovolenta, P. *et al.* (2010) Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron*, **66**, 523-535.
52. Horibata, Y. and Sugimoto, H. (2010) StarD7 mediates the intracellular trafficking of phosphatidylcholine to mitochondria. *J Biol Chem*, **285**, 7358-7365.
53. Escobar, D.J., Desai, R., Ishiyama, N., Folmsbee, S.S., Novak, M.N., Flozak, A.S., Daugherty, R.L., Mo, R., Nanavati, D., Sarpal, R. *et al.* (2015) alpha-Catenin phosphorylation promotes intercellular adhesion through a dual-kinase mechanism. *J Cell Sci*, **128**, 1150-1165.
54. Swinnen, B., Bento-Abreu, A., Gendron, T.F., Boeynaems, S., Bogaert, E., Nuyts, R., Timmers, M., Scheveneels, W., Hersmus, N., Wang, J. *et al.* (2018) A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathol*, **135**, 427-443.
55. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, **17**, 909-915.

56. Rodriguez-Tornos, F.M., Briz, C.G., Weiss, L.A., Sebastian-Serrano, A., Ares, S., Navarrete, M., Frangeul, L., Galazo, M., Jabaudon, D., Esteban, J.A. *et al.* (2016) Cux1 Enables Interhemispheric Connections of Layer II/III Neurons by Regulating Kv1-Dependent Firing. *Neuron*, **89**, 494-506.
57. Shi, L., Chang, X., Zhang, P., Coba, M.P., Lu, W. and Wang, K. (2013) The functional genetic link of NLGN4X knockdown and neurodevelopment in neural stem cells. *Hum Mol Genet*, **22**, 3749-3760.
58. Anderson, D.M., Cannavino, J., Li, H., Anderson, K.M., Nelson, B.R., McAnally, J., Bezprozvannaya, S., Liu, Y., Lin, W., Liu, N. *et al.* (2016) Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. *Proc Natl Acad Sci U S A*, **113**, E4494-4503.
59. Joyce, P.I., Fratta, P., Landman, A.S., McGoldrick, P., Wackerhage, H., Groves, M., Busam, B.S., Galino, J., Corrochano, S., Beskina, O.A. *et al.* (2016) Deficiency of the zinc finger protein ZFP106 causes motor and sensory neurodegeneration. *Hum Mol Genet*, **25**, 291-307.
60. Celona, B., Dollen, J.V., Vatsavayai, S.C., Kashima, R., Johnson, J.R., Tang, A.A., Hata, A., Miller, B.L., Huang, E.J., Krogan, N.J. *et al.* (2017) Suppression of C9orf72 RNA repeat-induced neurotoxicity by the ALS-associated RNA-binding protein Zfp106. *Elife*, **6**.
61. Tretiakova, A., Gallia, G.L., Shcherbik, N., Jameson, B., Johnson, E.M., Amini, S. and Khalili, K. (1998) Association of Puralpha with RNAs homologous to 7 SL determines its binding ability to the myelin basic protein promoter DNA sequence. *J Biol Chem*, **273**, 22241-22247.
62. West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y. and Kingston, R.E. (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell*, **55**, 791-802.
63. Darbinian, N., Gallia, G.L. and Khalili, K. (2001) Helix-destabilizing properties of the human single-stranded DNA- and RNA-binding protein Puralpha. *J Cell Biochem*, **80**, 589-595.
64. Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M. and Ule, J. (2015) hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature*, **519**, 491-494.
65. Ayache, J., Benard, M., Ernoult-Lange, M., Minshall, N., Standart, N., Kress, M. and Weil, D. (2015) P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol Biol Cell*, **26**, 2579-2595.
66. Pandey, P.R., Yang, J.H., Tsitsipatis, D., Panda, A.C., Noh, J.H., Kim, K.M., Munk, R., Nicholson, T., Hanniford, D., Argibay, D. *et al.* (2020) circSamd4 represses myogenic transcriptional activity of PUR proteins. *Nucleic Acids Res*, **48**, 3789-3805.
67. Song, C., Zhang, Y., Huang, W., Shi, J., Huang, Q., Jiang, M., Qiu, Y., Wang, T., Chen, H. and Wang, H. (2021) Circular RNA Cwc27 contributes to Alzheimer's disease pathogenesis by repressing Pur-alpha activity. *Cell Death Differ*.
68. Haas, S., Gordon, J. and Khalili, K. (1993) A developmentally regulated DNA-binding protein from mouse brain stimulates myelin basic protein gene expression. *Mol Cell Biol*, **13**, 3103-3112.
69. Haas, S., Thatikunta, P., Steplewski, A., Johnson, E.M., Khalili, K. and Amini, S. (1995) A 39-kD DNA-binding protein from mouse brain stimulates transcription of myelin basic protein gene in oligodendrocytic cells. *J Cell Biol*, **130**, 1171-1179.
70. Singh, R. and Valcarcel, J. (2005) Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol*, **12**, 645-653.

71. Jones, S.A. and Jenkins, B.J. (2018) Recent insights into targeting the IL-6 cytokine family in inflammatory diseases and cancer. *Nat Rev Immunol*, **18**, 773-789.
72. Bar-Yosef, T., Damri, O. and Agam, G. (2019) Dual Role of Autophagy in Diseases of the Central Nervous System. *Front Cell Neurosci*, **13**, 196.
73. Sun, Y., Gao, J., Jing, Z., Zhao, Y., Sun, Y. and Zhao, X. (2020) PURalpha Promotes the Transcriptional Activation of PCK2 in Oesophageal Squamous Cell Carcinoma Cells. *Genes (Basel)*, **11**.
74. Ortiz-Gonzalez, X.R. (2021) Mitochondrial Dysfunction: A Common Denominator in Neurodevelopmental Disorders? *Dev Neurosci*, **43**, 222-229.
75. Youn, J.Y., Dyakov, B.J.A., Zhang, J., Knight, J.D.R., Vernon, R.M., Forman-Kay, J.D. and Gingras, A.C. (2019) Properties of Stress Granule and P-Body Proteomes. *Mol Cell*, **76**, 286-294.
76. Sheth, U. and Parker, R. (2003) Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science*, **300**, 805-808.
77. Riggs, C.L., Kedersha, N., Ivanov, P. and Anderson, P. (2020) Mammalian stress granules and P bodies at a glance. *J Cell Sci*, **133**.
78. Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M.J., Scheuner, D., Kaufman, R.J., Golan, D.E. and Anderson, P. (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J Cell Biol*, **169**, 871-884.
79. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M. *et al.* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*, **47**, D442-D450.