DNA and RNA binding mediate the toxicity of arginine-rich peptides encoded by *C9ORF72* GGGGCC repeats

V. Lafarga[1]*, O. Sirozh[1]*, I. Díaz-López[2], M. Hisaoka[3], E. Zarzuela[4], J. Boskovic[5], B. Jovanovic[3], R. Fernandez-Leiro[6], J. Muñoz[4], G. Stoecklin[3], I. Ventoso[2], O. Fernandez-Capetillo[1,7]

[1]Genomic Instability Group, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

[2]Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Departamento de Biología Molecular, Universidad Autónoma de Madrid (UAM), 28049 Madrid, Spain.

[3]Division of Biochemistry, Center for Biomedicine and Medical Technology Mannheim (CBTM), Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany; Center for Molecular Biology of Heidelberg University (ZMBH), German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, National Center for Tumor Diseases (NCT), partner site, 69120 Heidelberg, Germany.

[4]ProteoRed-ISCIII. Proteomics Unit, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

[5]Electron Microscopy Unit, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

[6]Genomic Integrity and Structural Biology Group, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

[7]Science for Life Laboratory, Division of Genome Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institute, S-171 21 Stockholm, Sweden.

*These authors contributed equally to this work.

Correspondence should be addressed to O.F (ofernandez@cnio.es)

Expanded intronic GGGGCC sequences in *C9ORF72* found in patients of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are translated into several dipeptide-repeat polypeptides (DPRs), from which the two that contain arginine, (PR)n and (GR)n, accumulate at nucleoli and kill cells. While this toxicity plays an important role in ALS/FTD, its mechanism remains unknown. We here show that (PR)n polypeptides bind DNA and RNA, so that any reaction involving nucleic acids is impaired by the DPRs. Consistently, (PR)n-induced cell death is rescued by addition of non-coding oligonucleotides. Interestingly, the effects of (PR)n are mimicked by protamine, a sperm-specific arginine-rich protein with affinity for nucleic acids, and the toxicity of either protamine or (PR)n DPRs is rescued by the anticoagulant heparin. We propose that generalized coating of nucleic acids by arginine-rich peptides accounts for the toxicity of *C9ORF72* mutations in ALS/FTD.

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are devastating human neurodegenerative diseases which lead to the patients´ death within 5 years of diagnosis[1]. Both pathologies share common genetic sources as evidenced by the frequent co-occurrence of the disease in siblings or even within the same patient. While many disease-associated mutations have been discovered, a large fraction occur in RNA-binding factors such as *FUS*, *TAF15* or several hnRNPs[2]. In addition, a general observation in ALS/FTD patients is the accumulation of cytoplasmic aggregates of RNA-binding proteins such as TDP-43[3], suggesting a general alteration of RNA metabolism in these cells [4]. Interestingly, the most frequent mutation linked to ALS/FTD occurs in *C9ORF72*[5,6], a gene whose protein product is involved in autophagy and in principle not related to RNA biology[7].

C9ORF72 mutations involve the expansion of a GGGGCC hexanucleotide within the first intron of the gene, which is amplified to hundreds or even thousands of copies in the patients[5,6]. Several hypotheses have been proposed to explain the pathogenicity of *C9ORF72* mutations including a reduction in C9ORF72 protein expression[8] or the toxicity of repeat-containing RNAs through sequestering RNA-binding proteins[9-11]. Besides these effects in *cis*, sense and antisense GGGGCC expansions can be translated into dipeptide-repeats (DPR) through repeat-associated non-AUG (RAN) translation[12-14]. The expression of these DPRs, particularly the two that contain arginine, (PR)n and (GR)n, is toxic in cells and animal models[15-21]. Significantly, synthetic (PR)n and (GR)n peptides added to culture media enter cells, accumulate at nucleoli, and kill cells[15].

While the toxicity of arginine-rich DPRs provides a mechanistic model to explain *C9ORF72* pathology in ALS/FTD, the mechanisms of this phenomenon remain poorly understood.

Proteomic and genetic studies have revealed that arginine-rich DPRs have a particular impact on mRNA translation[21-24]. To further evaluate this aspect, we performed a proteomic analysis of ribosomes purified from HeLa cells stably expressing streptavidin-binding peptide (SBP)-tagged RPS9 upon exposure to synthetic $(PR)_{20}$ peptides (**Fig. S1a**). Besides changes in specific factors, the most prominent observation was a generalized decrease in the abundance of L ribosomal proteins (RPLs) from the 60S subunit on the ribosomal fractions isolated from $(PR)_{20}$-treated cells (**Fig. S1b, Table S1**), which was not due to a reduction in protein levels of these factors caused by the polypeptide (**Fig. S1c**). Since ribosome purification was performed by pulling down RPS9, a component of the 40S ribosomal subunit, the observed reduction in 60S factors suggested an effect of the DPR on the assembly of 80S ribosomes. Accordingly, ribosome profiling from HeLa cells treated with $(PR)_{20}$ revealed an accumulation of polysome halfmers (**Fig. S2a**), indicative of an inefficient assembly of 40S and 60S subunits at sites of translation initiation. However, $(PR)_{20}$ did not affect the *in vitro* assembly of purified 40S and 60S subunits into 80S particles in the absence of mRNA, as evaluated by electron microscopy, arguing that the DPR did not directly affect the 40S or 60S subunits (**Fig. S2b,c**).

To further evaluate the role of the DPRs in translation, we performed *in vitro* translation reactions in rabbit reticulocyte lysates. Consistent with previous work[22], *in vitro* translation of a luciferase mRNA was impaired by $(PR)_{20}$ in a dose-dependent manner (**Fig. S3a,b**). While prior observations indicated that the effect of (PR)n peptides on translation occurs through preventing the access of translation initiation factors to mRNA[22], $(PR)_{20}$ also inhibited translation elongation in reactions where translation initiation was blocked by lactimidomycin (**Fig. S3c**). These data, together with the presence of halfmers in $(PR)_{20}$-treated cells (**Fig. S2a**), suggested that the DPRs could be acting during and after the assembly of 80S ribosomes on mRNA. Since arginine-containing DPRs have been shown to form aggregates in the presence of RNA[22,25], we investigated if mRNA-coating by $(PR)_{20}$ could be the cause of impaired translation by the DPRs. Supporting this view, extending the length of the 5´UTR of the luciferase mRNA led to a further decrease in translation by $(PR)_{20}$ (**Fig. S3d**), due to an increased probability of DPR-binding to the 5´UTR that could block initiation. mRNA coating by the DPR can also explain why the polypeptide also blocks translation elongation, as the coding sequence of the luciferase mRNA (1600 nt) represents 90%

of its length. Of note, the interaction with $(PR)_{20}$ does not irreversibly alter RNA molecules, as mRNA recovered from $(PR)_{20}$-treated reticulocyte extracts through purification was efficiently translated in a subsequent *in vitro* translation reaction (**Fig. S3e**).

Besides translation, a number of cellular processes using RNA intermediates have been found to be altered by *C9ORF72* repeat expansion coded DPRs including splicing[15,26], rRNA biosynthesis[15] and nuclear mRNA export[27]. While these effects have been proposed to be due to the binding of the DPRs to specific factors such as the splicing regulator U2 snRNP[26], we reasoned that RNA coating by arginine-rich DPRs should affect any reaction involving RNA. Accordingly, *in vitro* reactions using RNA substrates such as reverse transcription or RNase A-mediated RNA degradation were impaired by $(PR)_{20}$ in a dose-dependent manner (**Fig. 1a,b**). Of note, the fact that $(PR)_{20}$ protects RNA from degradation can influence proteomic analyses aiming to identify DPR-interacting proteins, since these interactions may be indirectly mediated by RNA. In fact, such studies have consistently found an enrichment of RNA-binding factors as DPR interactors[22,26,28]. Regarding cellular effects, we could confirm an accumulation of poly(A) mRNA in the nucleus[27] (**Fig. 1c**), and an overall decrease in RNA biosynthesis[15] in U2OS cells exposed to $(PR)_{20}$, (**Fig. 1d**). In addition, we observed an altered nucleoplasmic distribution of Cajal Bodies, which are membrane-less nuclear assemblies that have been involved in various aspects of RNA metabolism including snRNA modifications, histone mRNA assembly or telomeric RNA processing (**Fig. 1e**) [29]. Moreover, intracellular replication of Sindibis virus, which contains a single-stranded RNA genome, in Baby Hamster Kidney (BHK-21) was also impaired by the presence of $(PR)_{20}$ (**Fig. 1f**), further supporting the notion that (PR)n DPRs have a general effect on all reactions using RNA.

In addition to RNA, oligoarginine peptides bind avidly DNA and promote its compaction[30-32]. Consistently, electrophoretic mobility assays (EMSA) revealed a similar affinity of $(PR)_{20}$ for DNA and RNA in single- or double-stranded form (**Fig. 2a, Fig. S4**). We thus explored the impact of $(PR)_{20}$ in reactions based on DNA templates. In vitro, the DPR inhibited polymerase chain reactions in a dose-dependent manner (**Fig. 2b,c**). When added to U2OS, the presence of $(PR)_{20}$ reduced DNA replication rates (**Fig. 2d**) and impaired the repair of DNA breaks generated by ionizing radiation as measured with antibodies detecting the phosphorylated form of histone H2AX (γH2AX) (**Fig. 2e**). The efficiency of gene-deletion of a stably expressed EGFP cDNA by

CRISPR/Cas9 using a sgRNA against EGFP was also significantly affected by $(PR)_{20}$ **(Fig. 2f)**, even though this effect could be influenced both by DNA or RNA binding of the DPR.

Based on the previous results, we investigated if the effects of the $(PR)_{20}$ peptide could be alleviated by the presence of DPR-scavenging non-coding oligonucleotides. In support of this, addition of a 646 nucleotide (nt) RNA lacking ATG rescued the effect of $(PR)_{20}$ on in vitro translation reactions in a dose-dependent manner (**Fig. 3a**). Furthermore, 19 or 38 nt-long ssDNA oligonucleotides rescued the toxicity of $(PR)_{20}$ in U2OS cells both in short-term viability experiments as well as in clonogenic assays (**Fig. 3b,c**). Importantly, ssRNA or ssDNA molecules did not prevent the entry of the $(PR)_{20}$ into the nucleus or its accumulation into nucleoli (**Fig. 3d, e**). On the contrary, the presence of the DPR led to the entry of ssDNA and, to a lesser extent, ssRNA, into cells, and their accumulation at nucleoli (**Fig. 3f,g**). In fact, oligoarginines have been used as non-lipidic transfection reagents for decades[33], and arginine-rich cell-penetrating proteins such as Tat from the human immunodeficiency virus (HIV1) also facilitate the delivery of nucleic acids or proteins into cells or even animals[34].

Finally, we explored if the effects of *C9ORF72* repeat-expansion encoded DPRs would be similar to those triggered by other arginine-rich peptides. To this end, we evaluated the effects of PROTAMINE, a sperm-specific small polypeptide that promotes DNA compaction and that has the highest percentage of arginine content within the animal proteome[35]. In addition, PROTAMINE has also been investigated as a DNA or RNA transfection reagent[36] and is known to be toxic for poorly understood reasons[37]. Remarkably, PROTAMINE purified from salmon sperm accumulated at nucleoli (**Fig. 4a**) and mimicked all the effects triggered by (PR) DPRs in U2OS cells including an increase in the nucleolar size (**Fig. 4b**), reduced transcription and translation rates (**Fig. 4c,d**) and alterations in splicing (**Fig. 4e**). If the cellular effects of PROTAMINE and $(PR)_{20}$ are due to a limited the access of RNA- or DNA-binding factors to their targets, we reasoned that both treatments should promote similar alterations on the chromatin-bound proteome. Indeed, a proteomic analysis of the changes in chromatin-bound factors that are triggered by PROTAMINE and $(PR)_{20}$ showed a highly statistically significant linear correlation and overlap between both conditions (**Fig. 4f,g**). In addition, and consistent with the DNA/RNA-coating model, among the proteins that showed reduced levels on the chromatin fraction upon $(PR)_{20}$ or PROTAMINE exposure there were several arginine-rich proteins such as linker histone H1 variants or High Mobility Group (HMG) proteins (**Table S2**). Since PROTAMINE has a

clinical use as antidote of the anticoagulant heparin, we tested if heparin could similarly revert the toxic effects of PROTAMINE. In fact, heparin fully rescued the viability of U2OS cells exposed to protamine or $(PR)_{20}$ in clonogenic assays (**Fig. 4h**). While the use of heparin led to a small fraction of $(PR)_{20}$ accumulating on cellular membranes, the majority of the DPR still entered into cells and accumulated at nucleoli (**Fig. S5**).

Our work provides a mechanism to explain the toxic effects of cell-penetrating oligoarginine peptides, including those translated from *C9ORF72* intronic repeat expansions found in ALS/FTD patients. We propose that the high affinity of oligoarginines for nucleic acids leads to a generalized coating of DNA and RNA in cells, impairing any cellular reaction that uses nucleic acid templates. These effects can be alleviated by heparin or DPR-scavenging non-coding oligonucleotides, providing proof-of-principle examples of how this knowledge could be exploited. These strategies are particularly relevant in the context of paracrine effects, which have been reported not only for *C9ORF72* derived DPRs[38,39] but also for other cell-penetrating peptides such as the Tat protein from the HIV-1 virus, which has been linked to AIDS-related neurodegeneration[40].

Methods

Cell lines

U2OS, BHK-21, Hela-SBP and Hela RPS9^{SBP} cells stably expressing RPS9-Flag-TEV-SBP were cultivated in standard DMEM medium supplemented with 10% FBS, 2 mM L-glutamine and 1% penicillin/streptomycin. The previously described mouse ES^{Cas9} cells[41] were grown on gelatin in DMEM (high glucose) supplemented with 15% knockout serum replacement (Invitrogen), LIF (1000 U/ml), 0.1 mM non-essential amino acids, 1% glutamax and 55 mM β-mercaptoethanol.

Treatments

Peptides containing 20 PR dipeptide repeats and a C-terminal HA epitope tag were synthesized at Genscript. Hydroxyurea (H8627), heparin sodium salt (H3149) and salmon PROTAMINE (P4005) were obtained from Sigma, cycloheximide (239764) from Calbiochem and lactimidomycin (56295) from EMD Millipore. Fluorophore-labelled oligonucleotides were synthesized by Sigma, with the following sequences: Cy3-5´-DNA (CCACTGCACCGCTGCTAGG); Cy5-5´-DNA (CCTAGCAGCGGTTGCAGTGG); Cy3-5´-RNA (CCACUGCACCGCUGCUAGG) and Cy5-RNA (CCUAGCAGCGGUUGCAGUGG).

Antibodies

The following antibodies were used in this study: rat anti-HA (Roche; 11867423001), rabbit anti-UBF1 (Santa Cruz; sc-13125), mouse anti-COILIN (204.10 from Dr. Lamond´s Lab) and mouse anti-γH2AX (Millipore; 05-636).

Immunofluorescence and High Throughput Microscopy

For immunofluorescence using HA, coilin and UBF antibodies, cells were fixed with 4% PFA prepared in PHEM buffer (60 mM Pipes, 25 mM Hepes, 10 mM EGTA, 2 mM $MgCl_2$ pH 6.9) containing 0.2% of Triton X-100, and permeabilized with 0.5% Triton X-100 after fixation. For HTM, cells were grown on μCLEAR bottom 96-well plates (Greiner Bio-One) and immunofluorescence of γH2AX was performed using standard procedures. Analysis of DNA replication by EdU, transcription by EU and translation by OPP incorporation were done using Click-It kits (Invitrogen) following manufacturer's instructions. In all cases, images were automatically acquired from each well using an Opera High-Content Screening System (Perkin Elmer). A 20x or 40x magnification lens was used and images were taken at non-saturating conditions. Images were segmented using DAPI signals to generate masks matching cell nuclei from which the mean signals for the rest of the stainings were calculated.

Cell viability

4,000 cells were seeded per well in a 96-well tissue culture plate and treated with the indicated concentrations of $(PR)_{20}$ alone or together with 4 uM of 19 or 38 nt-long Cy5-RNA or Cy5-DNA oligonucleotides. 36 hours after the treatment, cell viability was measured using a luminescent system (CellTiter-Glo, Promega), according to the manufacturer's protocol. Viability is plotted as percentages compared to untreated controls.

Clonogenic cell survival assays

2,000 U2OS cells were plated in 6-well tissue culture plates in culture medium. The following day, cells were incubated with 10 μM $(PR)_{20}$ alone or in combination with 2 μM of 19

(CCACTGCACCGCTGCTAGG) or 38 (CCACTGCACCGCTGCTAGGATCGCCTGAAATCGTTGGC) nucleotide-length ssDNA molecules. After 2 days the medium was changed and cells were then grown for 8 additional days in untreated medium. At the end of the experiments, cells were fixed and stained with 0.4% methylene blue in methanol for 30 min.

Peptide Labelling

The Cy®3 Mono 5-pack kit (Sigma) was used for the labeling of PROTAMINE. Briefly, 0.150 mM salmon PROTAMINE was incubated with 0.850 mM of Cy3 NHS Ester in 20 mM Hepes and 150 mM NaCl buffer overnight at 4°C, followed by the addition of 1 M Tris pH7.5 in order to quench the reaction. Cy3-protamine was purified using 3K Amicon Ultra 0.5 mL centrifugal filters (Sigma).

Immunoprecipitation

$10x10^6$ Hela-SBP and Hela-$RPS9^{SBP}$ cells were lysed in cold RNA-IP lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM $MgCl_2$, 2 % NP-40, 10% glycerol and freshly added Complete protease inhibitors) and incubated with 30 µL of Strep-Tactin beads (IBA) for 1 h at 4°C with constant shaking. Beads were washed 6 times in NET2 buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.5% Triton X-100).

Mass spectrometry

IP samples were eluted in urea, FASP-digested with Lys-C/trypsin and analyzed by LC-MS/MS. Label-free quantification was performed using MaxQuant. Chromatin and whole cell extract samples were trypsin-digested using S-traps, isobaric-labelled with iTRAQ 8-plex (chromatin) or TMT 11-plex (whole cell) reagents and pre-fractionated with RP-HPLC at high pH. Fractions were then analyzed by LC-MS/MS and raw data was processed with MaxQuant. Statistical analyses were performed with Perseus and ProStar. Mass spectrometry proteomics data associated to this work have been deposited to the ProteomeXchange Consortium via the PRIDE[42] partner repository with the dataset identifier PXD010555 (Username: reviewer92012@ebi.ac.uk; Password: ZnXyUNzE).

Polysome analysis

Hela-$RPS9^{SBP}$ cells were treated with either water as control or 10 µM of $(PR)_{20}$ during 16 hours. Ribosomes were stalled by addition of 100 µg/mL cycloheximide (CHX) for 5 min, and cells were lysed in polysome lysis buffer (15 mM Tris-HCl pH 7.4, 15 mM $MgCl_2$, 300 mM NaCl, 1% Triton-X-100, 0.1% β-mercaptoethanol, 200 U/mL RNAsin (Promega), 1 complete Mini Protease Inhibitor Tablet (Roche) per 10 mL). Nuclei were removed by centrifugation (9300×G, 4°C, 10 min) and the cytoplasmic lysate was loaded onto a sucrose density gradient (17.5–50% in 15 mM Tris-HCl pH 7.4, 15 mM $MgCl_2$, 300 mM NaCl and, for fractionation from BMDM, 200 U/ml Recombinant RNAsin Ribonuclease Inhibitor, Promega). After ultracentrifugation (2.5 h, 35,000 rpm at 4°C in a SW60Ti rotor), gradients were eluted with a Teledyne Isco Foxy Jr. system into 16 fractions of similar volume.

Ribosome assembly

Ribosomal subunits were purified from BHK21 cells using previously described procedures[43]. Briefly, the ribosomal fraction (P100) was treated with 1 mM puromycin for 15 min on ice and

then for 15 min at 37ºC. The suspension was adjusted to 0.5 M KCl and loaded onto 10 to 30% sucrose density gradients which were centrifuged at 22,000 rpm for 16 h at 4ºC in a SW40 rotor. The peaks corresponding to 40S and 60S subunits were identified by optical density at 260 nm. Finally, fractions were concentrated using YM-100 centricons. 1 pmol of 40S and 60S were incubated in assembly buffer (25 mM of TrisHCl pH7.5, 100 mM KCl, 5 mM $MgCl_2$ and 3mM of DTT) for 45 minutes at 30 ºC. To control for non-assembly, the concentration of $MgCl_2$ in the buffer was reduced to 1 mM.

Electron microscopy

For negative staining 3 µl of purified 40S+60S ribosome fractions or alternatively 3 µl of 40S+60S ribosome fractions in the presence of $(PR)_{20}$ (1:10 molar ratio) were deposited onto freshly glow-discharged carbon-coated 400 mesh copper electron microscopy (EM) grids (Electron Microscopy Sciences) and retained on the grid for 1 min. Afterwards, grids were washed with three distinct 50 µL drops of MilliQ water. The excess of water was removed with filter paper and the grids were placed on the top of three different 40 µL drops of 2% uranyl acetate and stained on the last drop for 1 min. The grids were then gently stripped for 4 s and air dried. Finally, grids were visualized on a Tecnai 12 transmission electron microscope (Thermo Fisher Scientific Inc.) with a lanthanum hexaboride cathode operated at 120 keV. Images were recorded at 61320 nominal magnification on a 4kx4k TemCam-F416 CMOS camera (TVIPS).

In vitro translation

In vitro translation was performed in nuclease-treated rabbit reticulocyte lysates (RRL) (Promega cat #L4960) using a luciferase mRNAs in the presence of the indicated concentrations of $(PR)_{20}$. Reactions were incubated for 60 mins at 30ºC and a 5 µl alicuot was used to measure luciferase activity in 100 µl of reaction buffer (15mM K2HPO4, 15mM MgSO4, 4mM EGTA pH 8, 4mM DTT, 2mM ATP and 0.1mM luciferin) in a Berthold Luminometer. For radioactive measurements, in vitro translation was performed in nuclease-treated RRL (Promega) using 150 ng of luciferase mRNA in the presence of 15 µCi of [35S]-Met for 60 min at 30°C. Samples were denatured in sample buffer and analyzed by SDS-PAGE and autoradiography.

In situ hybridization

In situ hybridization was carried out as previously described[44]. Briefly, U2OS cells were fixed with 4% PFA prepared in PHEM buffer. An oligo dT $_{(50)}$-mer, 5′-end labeled with biotin (MWG-Biotech, Germany) was used as a probe for in situ hybridization to poly(A) RNA. After hybridization, cells were washed and the hybridization signal was detected with FITC-avidin for 30 min. All samples were mounted with the ProLong anti-fading medium (Invitrogen).

RNase A digestion

RNA was extracted from mouse embryonic stem cells using the Absolutely RNA microprep kit (Agilent). 1 µg of RNA was pre-incubated with 4.5 µg of $(PR)_{20}$ or water for 10 min at room temperature, and subsequently treated with serially diluted RNase A (Qiagen) at 37°C for 15 min, at the end of which the reaction was halted by RNasin ribonuclaease inhibitor (Promega) and 1% SDS. The resulting mixture was column-purified using the RNA Clean & Concentrator-5 kit (Zymo Research), according to manufacturer's instructions. The recovered RNA was analyzed by agarose gel electrophoresis, and quantified by image analysis (ImageJ, NIH, USA).

Reverse transcription

500 ng of RNA isolated from mESC was pre-incubated on ice for 5 minutes with increasing concentrations of $(PR)_{20}$. The mixture was then used as template for cDNA synthesis using the SuperScript III First Strand Synthesis System (ThermoFisher Scientific) according to manufacturer's instructions. The reaction mixture was subsequently treated with RNase A (15 min at 37°C), solubilized with SDS 1%, and finally purified using PCR cleanup columns (Qiagen). The yield of cDNA was visualized by agarose gel electrophoresis, and quantified by fluorometry using the Qubit kit (ThermoFisher Scientific), according to manufacturer's instructions.

RT-PCR

cDNA generated from mESC was pre-incubated on ice for 5 min with increasing concentrations of $(PR)_{20}$. Out of this mixture, 50 ng of cDNA were used in each reaction. RT-PCR was carried out using the 2X SYBR Select Master Mix (Life Technologies) in MicroAmp© Optical 384-Well plates (Applied Biosystems) in the QuantStudio™ 6 Flex Real-Time PCR System (Thermo Fisher) using standard protocols. The sequences of the primers used are as follows: *Gapdh*-F (TTCACCACCATGGAGAAGGC), *Gapdh*-R CCCTTTTGGCTCCACCCT, 5.8S-F (GTCGATGAAGAACGCAGCTA) and 5.8S-R (AACCGACGCTCAGACAGG).

Viral infection

BHK-21 cells were infected with Sindbis virus (Alphavirus, ssRNA genome with positive polarity) at a MOI of 20 viral PFU/cell in the absence or presence of $(PR)_{20}$. Cells were collected 4 hours after infection and total RNA was purified using RNeasy Mini Kit (Quiagen). Virus-specific primers (F:GGTACTGGAGACGGATATCGC and R:CGATCAAGTCGAGTAGTGGTTG) were used to detect viral RNA by qRT-PCR, which were normalized against cellular *GAPDH*.

Splicing

U2OS cells were plated at a density of 2.4×10^4 cells/mL on 6-well plates and treated on the following day with either 20 μM of $(PR)_{20}$ or 30 μM protamine. RT-PCR was performed as described above using previously described primers to detect alternative splicing events at *GADD45A* and *NACA* mRNAs[15].

Electrophoretic mobility shift assay (EMSA)

To analyze the binding of $(PR)_{20}$ to ssDNA and ssRNA, 0.2 μM of Cy3-RNA (CCACUGCACCGCUGCUAGG) or Cy3-DNA (CCACTGCACCGCTGCTAGG) oligonucleotides were incubated with increasing amounts of $(PR)_{20}$ for 10 min at room temperature. To evaluate the binding of $(PR)_{20}$ to dsDNA and dsRNA, Cy3-RNA and Cy3-DNA oligonucleotides were annealed with the complementary RNA (CCUAGCAGCGGUUGCAGUGG) or DNA (CCTAGCAGCGGTTGCAGTGG) molecule prior to incubation with $(PR)_{20}$. Reactions were supplemented with 4 μL of 6x loading buffer (30% glycerol, bromphenol blue, xylene cyanol) and were resolved by polyacrylamide gel electrophoresis in 4-12% TBE Gels (Invitrogen) in 1%TBE buffer at 100 V for 45 min at 4°C. The gels were scanned and Cy3 intensity was measured using Typhoon Trio (GE Health Care). Band quantification was performed using ImageJ and data fitting was performed using Graphpad Prism 7.

CRISPR/Cas9 efficiency

The previously described ES^{Cas9} mESC line[41] which carries a Doxycycline-inducible Cas9 was co-infected with lentiviral vectors expressing EGFP (pLVTHM, Addgene, 12247) and an EGFP-targeting sgRNA cloned in the pKLV-U6gRNA-PGKpuro2ABFP backbone (Addgene, 50946). Two independent clones with stable expression of all components were seeded on gelatin. After six hours, Doxycycline (1 µg/mL) and/or $(PR)_{20}$ (10 µM) were added to the medium. 72 hours later, cells were recovered and the percentage of GFP-positive cells was quantified by flow cytometry using the FlowJo software (BD).

References

1 Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nat Rev Dis Primers* **3**, 17085, doi:10.1038/nrdp.2017.85 (2017).

2 Kapeli, K., Martinez, F. J. & Yeo, G. W. Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum Genet* **136**, 1193-1214, doi:10.1007/s00439-017-1830-7 (2017).

3 Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133, doi:10.1126/science.1134108 (2006).

4 Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416-438, doi:10.1016/j.neuron.2013.07.033 (2013).

5 DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256, doi:10.1016/j.neuron.2011.09.011 (2011).

6 Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268, doi:10.1016/j.neuron.2011.09.010 (2011).

7 Farg, M. A. *et al.* C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet* **23**, 3579-3595, doi:10.1093/hmg/ddu068 (2014).

8 Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med* **24**, 313-325, doi:10.1038/nm.4490 (2018).

9 Donnelly, C. J. *et al.* RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415-428, doi:10.1016/j.neuron.2013.10.015 (2013).

10 Lee, Y. B. *et al.* Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* **5**, 1178-1186, doi:10.1016/j.celrep.2013.10.049 (2013).

11 Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200, doi:10.1038/nature13124 (2014).

12 Ash, P. E. *et al.* Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646, doi:10.1016/j.neuron.2013.02.004 (2013).

13 Mori, K. *et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338, doi:10.1126/science.1232927 (2013).

14 Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977, doi:10.1073/pnas.1315438110 (2013).

15 Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145, doi:10.1126/science.1254917 (2014).

16 Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194, doi:10.1126/science.1256800 (2014).

17 Wen, X. *et al.* Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**, 1213-1225, doi:10.1016/j.neuron.2014.12.010 (2014).
18 Chew, J. *et al.* Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348**, 1151-1154, doi:10.1126/science.aaa9344 (2015).
19 Stopford, M. J. *et al.* C9ORF72 hexanucleotide repeat exerts toxicity in a stable, inducible motor neuronal cell model, which is rescued by partial depletion of Pten. *Hum Mol Genet* **26**, 1133-1145, doi:10.1093/hmg/ddx022 (2017).
20 Swaminathan, A. *et al.* Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model. *Hum Mol Genet* **27**, 1754-1762, doi:10.1093/hmg/ddy083 (2018).
21 Zhang, Y. J. *et al.* Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med*, doi:10.1038/s41591-018-0071-1 (2018).
22 Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum Mol Genet* **25**, 1803-1813, doi:10.1093/hmg/ddw052 (2016).
23 Lopez-Gonzalez, R. *et al.* Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron* **92**, 383-391, doi:10.1016/j.neuron.2016.09.015 (2016).
24 Chai, N. & Gitler, A. D. Yeast screen for modifiers of C9orf72 poly(glycine-arginine) dipeptide repeat toxicity. *FEMS Yeast Res* **18**, doi:10.1093/femsyr/foy024 (2018).
25 Boeynaems, S. *et al.* Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol Cell* **65**, 1044-1055 e1045, doi:10.1016/j.molcel.2017.02.013 (2017).
26 Yin, S. *et al.* Evidence that C9ORF72 Dipeptide Repeat Proteins Associate with U2 snRNP to Cause Mis-splicing in ALS/FTD Patients. *Cell Rep* **19**, 2244-2256, doi:10.1016/j.celrep.2017.05.056 (2017).
27 Rossi, S. *et al.* Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. *J Cell Sci* **128**, 1787-1799, doi:10.1242/jcs.165332 (2015).
28 Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789-802 e712, doi:10.1016/j.cell.2016.10.003 (2016).
29 Nizami, Z., Deryusheva, S. & Gall, J. G. The Cajal body and histone locus body. *Cold Spring Harb Perspect Biol* **2**, a000653, doi:10.1101/cshperspect.a000653 (2010).
30 Tan, R. & Frankel, A. D. Structural variety of arginine-rich RNA-binding peptides. *Proc Natl Acad Sci U S A* **92**, 5282-5286 (1995).
31 Mascotti, D. P. & Lohman, T. M. Thermodynamics of oligoarginines binding to RNA and DNA. *Biochemistry* **36**, 7272-7279, doi:10.1021/bi970272n (1997).
32 DeRouchey, J., Hoover, B. & Rau, D. C. A comparison of DNA compaction by arginine and lysine peptides: a physical basis for arginine rich protamines. *Biochemistry* **52**, 3000-3009, doi:10.1021/bi4001408 (2013).
33 Emi, N., Kidoaki, S., Yoshikawa, K. & Saito, H. Gene transfer mediated by polyarginine requires a formation of big carrier-complex of DNA aggregate. *Biochem Biophys Res Commun* **231**, 421-424, doi:10.1006/bbrc.1997.6125 (1997).

34 Schwarze, S. R., Ho, A., Vocero-Akbani, A. & Dowdy, S. F. In vivo protein transduction: delivery of a biologically active protein into the mouse. *Science* **285**, 1569-1572 (1999).
35 Balhorn, R. The protamine family of sperm nuclear proteins. *Genome Biol* **8**, 227, doi:10.1186/gb-2007-8-9-227 (2007).
36 Scheicher, B., Schachner-Nedherer, A. L. & Zimmer, A. Protamine-oligonucleotide-nanoparticles: Recent advances in drug delivery and drug targeting. *Eur J Pharm Sci* **75**, 54-59, doi:10.1016/j.ejps.2015.04.009 (2015).
37 Sokolowska, E., Kalaska, B., Miklosz, J. & Mogielnicki, A. The toxicology of heparin reversal with protamine: past, present and future. *Expert Opin Drug Metab Toxicol* **12**, 897-909, doi:10.1080/17425255.2016.1194395 (2016).
38 Westergard, T. *et al.* Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. *Cell Rep* **17**, 645-652, doi:10.1016/j.celrep.2016.09.032 (2016).
39 Zhou, Q. *et al.* Antibodies inhibit transmission and aggregation of C9orf72 poly-GA dipeptide repeat proteins. *EMBO Mol Med* **9**, 687-702, doi:10.15252/emmm.201607054 (2017).
40 King, J. E., Eugenin, E. A., Buckner, C. M. & Berman, J. W. HIV tat and neurotoxicity. *Microbes Infect* **8**, 1347-1357, doi:10.1016/j.micinf.2005.11.014 (2006).
41 Ruiz, S. *et al.* A Genome-wide CRISPR Screen Identifies CDC25A as a Determinant of Sensitivity to ATR Inhibitors. *Mol Cell* **62**, 307-313, doi:10.1016/j.molcel.2016.03.006 (2016).
42 Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**, D447-456, doi:10.1093/nar/gkv1145 (2016).
43 Pisarev, A. V., Unbehaun, A., Hellen, C. U. & Pestova, T. V. Assembly and analysis of eukaryotic translation initiation complexes. *Methods Enzymol* **430**, 147-177, doi:10.1016/S0076-6879(07)30007-4 (2007).
44 Palanca, A., Casafont, I., Berciano, M. T. & Lafarga, M. Reactive nucleolar and Cajal body responses to proteasome inhibition in sensory ganglion neurons. *Biochim Biophys Acta* **1842**, 848-859, doi:10.1016/j.bbadis.2013.11.016 (2014).

Acknowledgements We thank Dr. Sergio Ruiz for providing the Dox-inducible Cas9 mESC line stably expressing EGFP and an EGFP-targeting sgRNA. Research was funded by Fundación Botín, by Banco Santander through its Santander Universities Global Division and by grants from the Spanish Ministry of Economy and Competitiveness (MINECO) (SAF2014-54498-R, co-financed with European FEDER funds), Howard Hughes Medical Institute and the European Research Council (ERC-617840) to OF; DKFZ NCT3.0 Integrative Project in Cancer Research grant (NCT3.0_2015.54 DysregPT) and SFB 1036/TP07 from the Deutsche Forschungsgemeinschaft to G.S.

Author Contributions V.L. and O.S. contributed to most experiments and data analyses and to the preparation of the figures. I.D. and I.V. helped with in vitro translation and viral infection experiments. E.Z. and J.M. helped with proteomic analyses. M. H. and G. S. conducted polysome analyses and B.J. generated $RPS9^{SBP}$-expressing HeLa cells. J.B. performed electron microscopy experiments. R.F. helped with DNA and RNA binding experiments. O.F. supervised the study and wrote the MS.

Author Information Proteomic data related to this work are available via ProteomeXchange with identifier PXD010555. The authors declare no competing financial interests. Correspondence and request for materials should be addressed to O.F. (ofernandez@cnio.es).

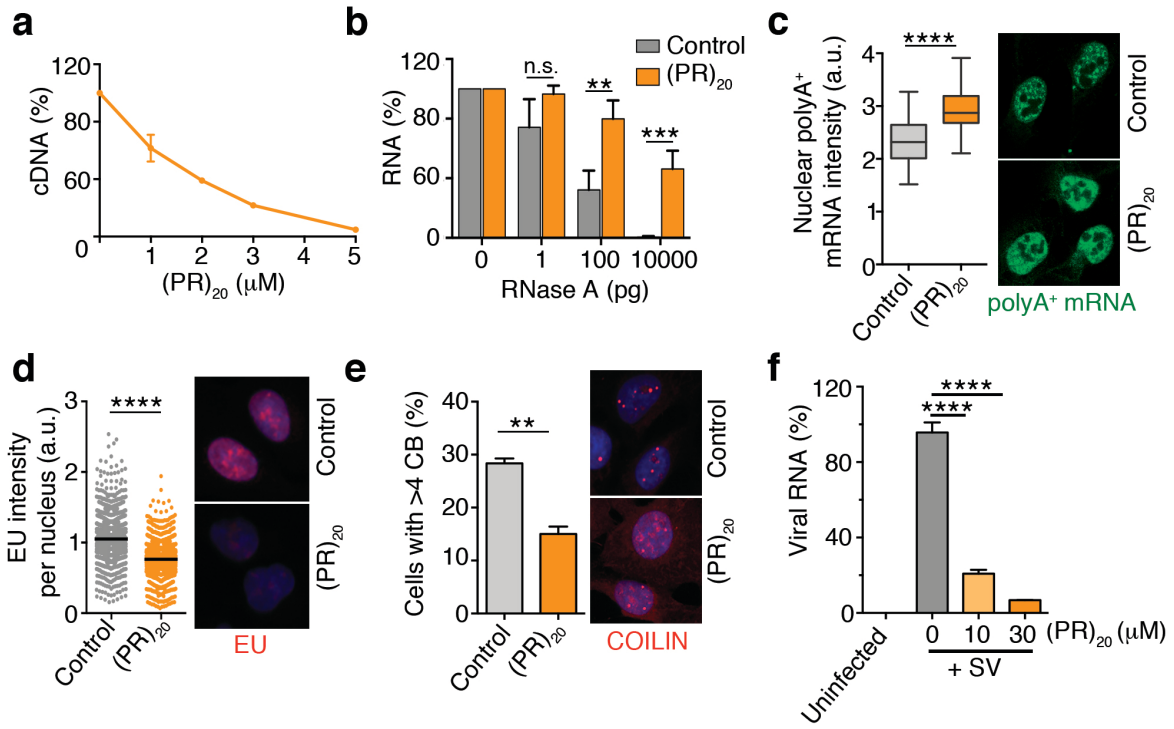

Fig. 1. RNA-based reactions are collectively impaired by (PR) DPRs. a, Reverse transcription of RNA (500 ng) with oligo-dT in the presence of increasing doses of $(PR)_{20}$. Data represent the fluorometric quantification of the resultant cDNA (n=2). **b,** Percentage of RNA (1 μg) remaining after a 15´digestion with increasing doses of RNase A in the presence or absence of $(PR)_{20}$ (5 μg) (n=3). (**c, d**) High-throughput microscopy (HTM)-mediated analysis of polyA+ mRNA (**c**) and 5-ethynyl-uridine (EU) (**d**) levels per nucleus found in U2OS cells exposed to 10 μM (PR)20 for 16 h. EU was added 30´ prior to fixation. **e,** Percentage of cells with more than 4 Cajal Bodies (CB) identified by immunofluorescence with anti-COILIN antibodies, in U2OS cells exposed to 10 μM $(PR)_{20}$ for 16 h. Note that besides its impact on the number of CBs, $(PR)_{20}$ also promotes a dispersed distribution of COILIN in the nucleoplasm. Representative images from these analyses (C-E) are provided in each case (right). **f,** Accumulation of Sindibis virus (SV) genomic RNA in BHK-21 cells in the absence or presence of increasing doses of $(PR)_{20}$. The amount of viral and cellular RNA was quantified from total RNA isolated 4 h after infection by quantitative real-time PCR (qRT-PCR) with specific primers. **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$.

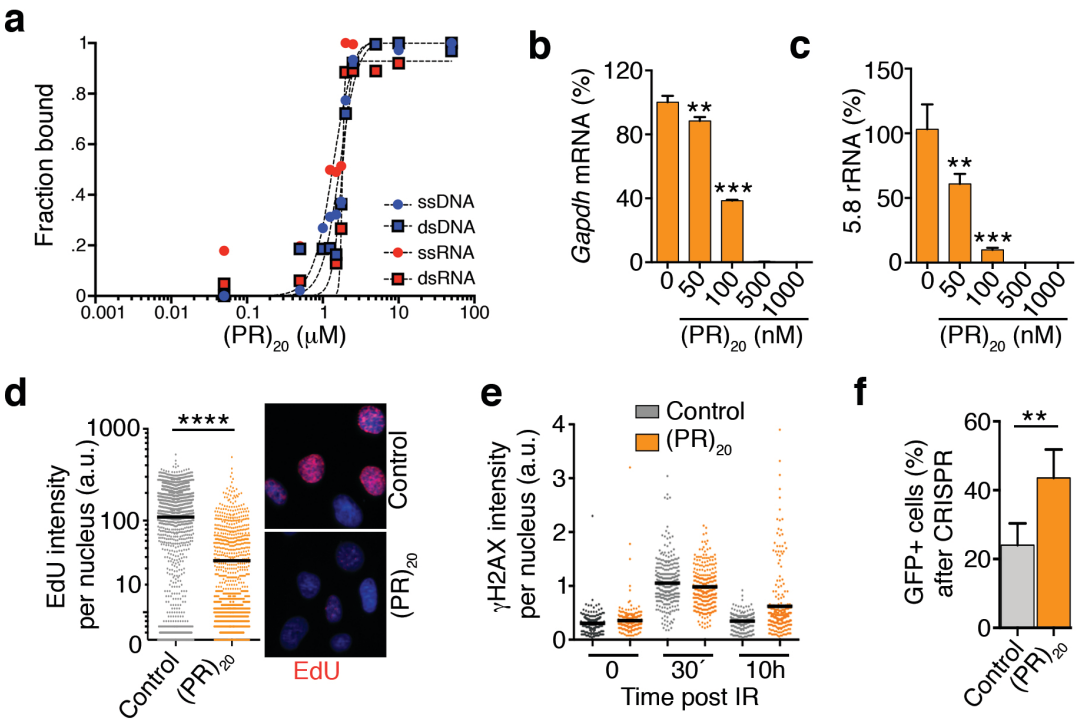

Fig. 2. Effects of $(PR)_{20}$ on DNA-based reactions. a, Quantification of EMSA assays evaluating the binding of $(PR)_{20}$ to 19 nt-long ssDNA, ssRNA, dsDNA and dsRNA molecules. Each probe (0.2 μM) was incubated with increasing concentrations of $(PR)_{20}$ for 10´. Curve-fitting was performed using non-linear regression with the Hill equation. (**b, c**) Percentage of *GAPDH* mRNA or 5.8 rRNA levels quantified by qRT-PCR in reactions containing increasing doses of $(PR)_{20}$ (n=3). **d,** HTM-mediated analysis of 5-ethynyl-deoxyuridine (EdU) levels per nucleus found in U2OS cells exposed to 10 μM $(PR)_{20}$ for 4 h. EdU was added 30´ prior to fixation. Representative images from these analyses are provided (right). **e**, HTM-mediated analysis of γH2AX levels per nucleus in U2OS cells exposed to 3Gy of IR in the presence or absence of 10 μM $(PR)_{20}$ for the indicated times. Note the accumulation of cells with persistent DNA damage (γH2AX) 10 h after IR in the presence of the DPR. **f**, Efficiency of CRISPR-mediated gene deletion quantified in mouse embryonic stem cells stably expressing EGFP, a doxycycline-inducible Cas9 and an EGFP-targeting sgRNA. The percentage of EGFP-positive cells was quantified by flow cytometry 72 h after inducing Cas9 expression with doxycycline in the presence or absence of 10 μM $(PR)_{20}$ (n=3). **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$.

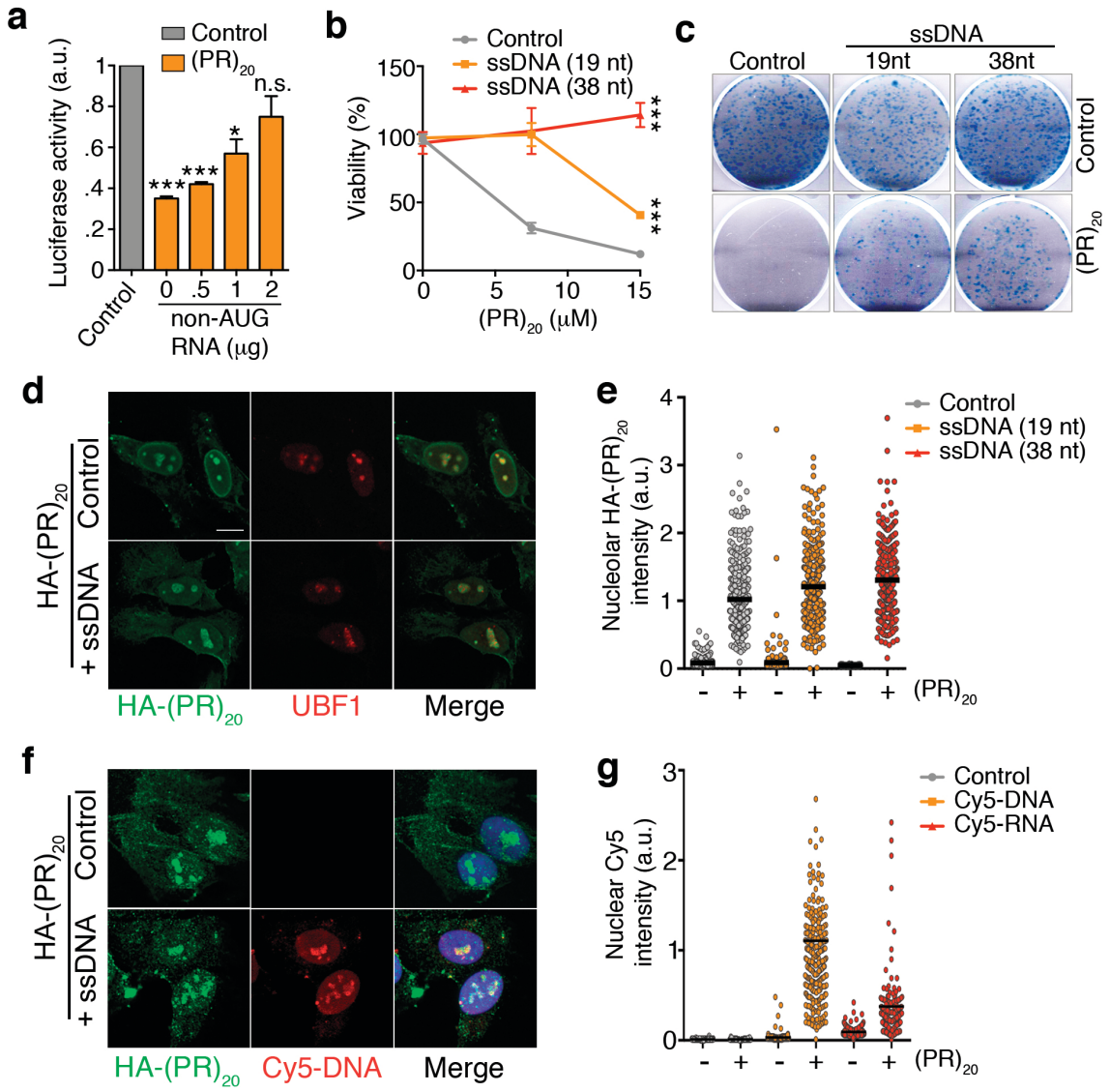

Fig. 3. Non-coding nucleic acids rescue the effects of $(PR)_{20}$ DPRs. a, *In* vitro translation of a luciferase mRNA with or without of 0.5 μM $(PR)_{20}$ and in the presence of increasing amounts of a 646 nt non-coding RNA. **b,** Percentage of viable cells as evaluated with a CellTiter-Glo luminescent assay in U2OS cells treated with increasing doses of $(PR)_{20}$ alone or together with 2 μM of ssDNA oligonucleotides (19 nt or 38 nt) (n=3). **c,** Clonogenic survival assay of U2OS cells exposed to 7.5 μM $(PR)_{20}$ alone or together with 2 μM of ssDNA oligonucleotides (19 nt or 38 nt). **d,** Immunofluorescence of HA-$(PR)_{20}$ (green) and the nucleolar factor UBF1 (red) in U2OS cells treated with 7.5 μM HA-$(PR)_{20}$ alone or together with 2 μM of a 19 nt ssDNA oligonucleotide for 8 h. **e,** HTM-mediated quantification of the nucleolar HA-$(PR)_{20}$ intensity from U2OS cells treated as in (**d**). **f,** Immunofluorescence of HA (green) and Cy5 (red) in U2OS cells treated with 7.5 μM HA-$(PR)_{20}$ alone or together with 4 μM of a Cy5-labeled 19 nt ssDNA oligonucleotide for 8 h. **g,**

HTM-mediated quantification of the nuclear Cy5 intensity from U2OS cells treated as in (**f**) with 2 μM of Cy5-labeled 19 nt ssDNA or ssRNA oligonucleotides. *, $p<0.05$; ***, $p<0.001$.

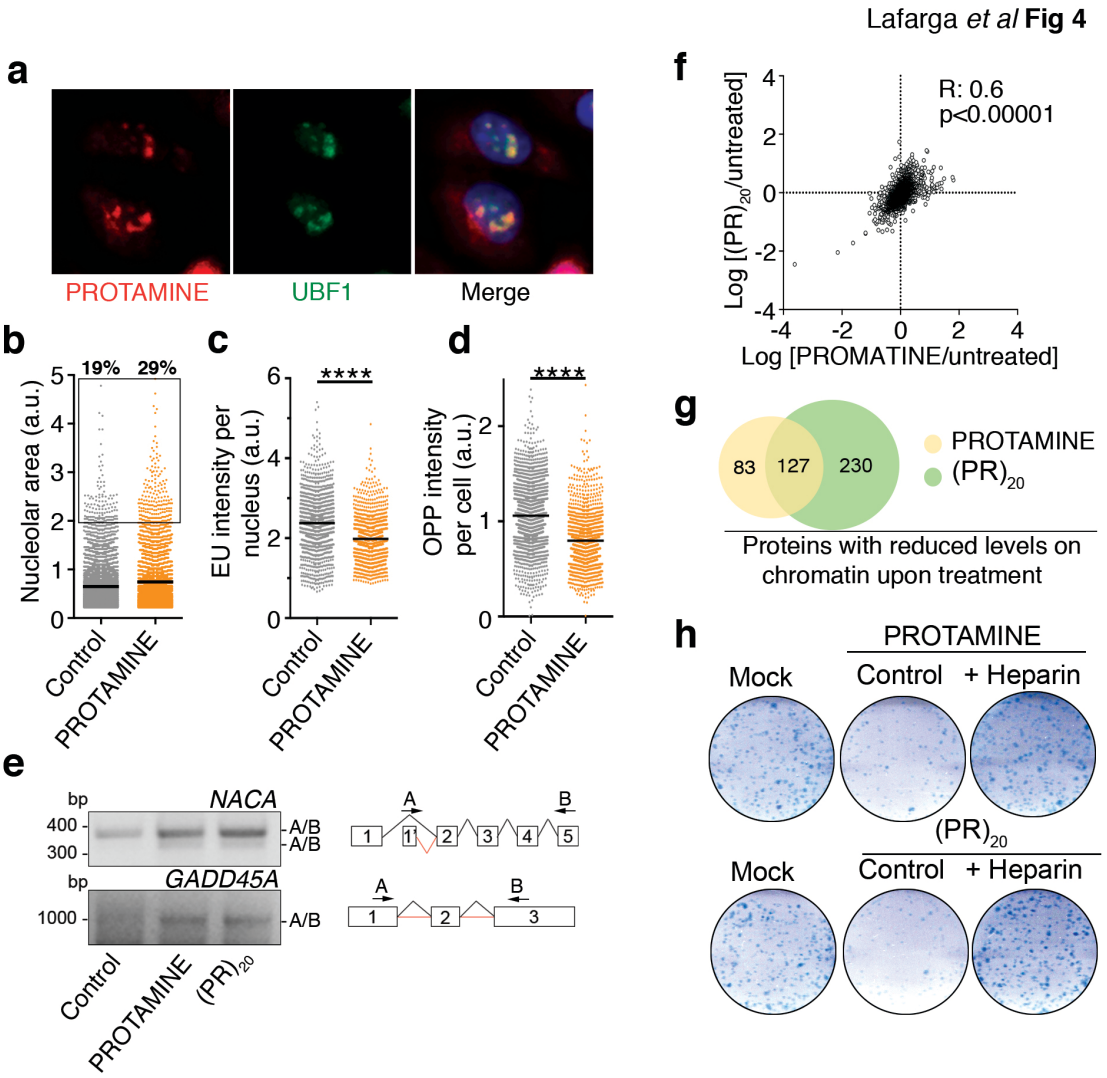

Fig. 4. PROTAMINE mimics the cellular effects of $(PR)_{20}$. a, Immunofluorescence of Cy3-labelled salmon PROTAMINE (red) and the nucleolar factor UBF1 (green) in U2OS cells treated with 30 µM Cy3-PROTAMINE for 4 h. (**b-d**) HTM-mediated quantification of the nucleolar area (**b**), EU levels per nucleus (**c**) and O-propargyl-puromycin (OPP) levels per cell (**d**) in U2OS cells exposed to 30 µM PROTAMINE for 16 (**c-d**) or 24 (**b**) h. EU and OPP were added 30´ prior to fixation. In the case of the nucleolar area, the percentage of cells with large nucleoli (defined as those with values 3 sd above the mean) is indicated. **e,** RT-PCR of *NACA* and *GADD45A* mRNAs in U2OS cells treated with PROTAMINE (30 µM) or $(PR)_{20}$ (20 µM) using primers in non-consecutive exons to monitor alternative splicing events. A scheme of the primers used is provided on the right side of the panel. **f,** Distribution of the chromatin-bound levels of all proteins identified by proteomics in U2OS cells treated with PROTAMINE (30 µM) or $(PR)_{20}$ (20 µM) for 90´. Numbers indicate the Pearson´s correlation coefficient (R) and the p-value for the linear correlation of the changes induced by both treatments. **g,** Venn diagram of the proteins that show significantly reduced levels on chromatin upon treatment of U2OS cells with PROTAMINE (30 µM) or $(PR)_{20}$

(20 μM) for 90′. **h,** Clonogenic survival assay of U2OS cells exposed to 7.5 μM $(PR)_{20}$ or PROTAMINE (20 μM) alone or together with 0.5 μM of heparin. $p<0.0001$.