ARTICLE

Phase separation and toxicity of C9orf72 poly(PR) depends on alternate distribution of arginine

Chen Chen[1], Yoshiaki Yamanaka[2], Koji Ueda[3], Peiying Li[4], Tamami Miyagi[2], Yuichiro Harada[2], Sayaka Tezuka[1], Satoshi Narumi[5], Masahiro Sugimoto[6], Masahiko Kuroda[2], Yuhei Hayamizu[1], and Kohsuke Kanekura[2]

Arg (R)-rich dipeptide repeat proteins (DPRs; poly(PR): Pro-Arg and poly(GR): Gly-Arg), encoded by a hexanucleotide expansion in the *C9ORF72* gene, induce neurodegeneration in amyotrophic lateral sclerosis (ALS). Although R-rich DPRs undergo liquid–liquid phase separation (LLPS), which affects multiple biological processes, mechanisms underlying LLPS of DPRs remain elusive. Here, using in silico, in vitro, and in cellulo methods, we determined that the distribution of charged Arg residues regulates the complex coacervation with anionic peptides and nucleic acids. Proteomic analyses revealed that alternate Arg distribution in poly(PR) facilitates entrapment of proteins with acidic motifs via LLPS. Transcription, translation, and diffusion of nucleolar nucleophosmin (NPM1) were impaired by poly(PR) with an alternate charge distribution but not by poly(PR) variants with a consecutive charge distribution. We propose that the pathogenicity of R-rich DPRs is mediated by disturbance of proteins through entrapment in the phase-separated droplets via sequence-controlled multivalent protein–protein interactions.

Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal motor neuron disease affecting both upper and lower motor neurons. Frontotemporal dementia (FTD) is a type of brain disorder with degeneration of the frontal and temporal lobes of cerebrum. It has been shown that these neurodegenerative diseases share some genetic and clinical features, and defective *C9ORF72* is the leading cause of familial cases of ALS (C9-ALS) and FTD (Renton et al., 2011; DeJesus-Hernandez et al., 2011). C9-ALS patients carry abnormally expanded GGGGCC hexanucleotide repeats in intron 1 of *C9ORF72*, and the unconventional repeat-associated non-ATG-initiated (RAN) translations (Zu et al., 2011) produce five dipeptide repeat proteins (DPRs), poly(GA): Gly-Ala, poly(PA): Pro-Ala, poly(GP): Gly-Pro, poly(GR): Gly-Arg, and poly(PR): Pro-Arg, from the hexanucleotide expansion (Zu et al., 2011; Mori et al., 2013). Two Arg-rich (R-rich) DPRs, poly(PR) and poly(GR), are toxic (Wen et al., 2014; Kwon et al., 2014; Mizielinska et al., 2014; Lee et al., 2016) to a wide variety of potential targets, including (a) protein translation (Kanekura et al., 2016; Lee et al., 2016; Zhang et al., 2018; Moens et al., 2019); (b) dynamics of membrane-less organelles (MLOs; Lin et al., 2016; Lee et al., 2016; Boeynaems et al., 2017); (c)

metabolism of RNA (Kwon et al., 2014); (d) nucleocytoplasmic transport (Jovičič et al., 2015; Zhang et al., 2015) and (e) DNA damage response (Andrade et al., 2020) for both poly(PR) and poly(GR); (f) ubiquitin-proteasome system (Gupta et al., 2017) and (g) heterochromatin anomalies for poly(PR) alone (Zhang et al., 2019); and (h) mitochondrial function for poly(GR) alone (Choi et al., 2019). Thus, the processes impaired by R-rich DPRs are highly diverse and regulated by independent mechanisms, indicating that the interaction between R-rich DPRs and a variety of proteins involved in various processes generates ALS-related neurotoxicity.

Arg, a positively charged, highly polar amino acid with dipole moment, plays a pivotal role in protein–protein interactions and protein–nucleic acid interactions (Boeynaems et al., 2019). Abundant positive charges and a highly repetitive nature enable these R-rich DPRs to bind to proteins and nucleic acids at multiple sites via electrostatic forces and cation-π interactions, resulting in liquid–liquid phase separation (LLPS; Lee et al., 2016; Lin et al., 2016). LLPS contributes to MLO formation (Frottin et al., 2019). MLOs, such as nucleoli and stress granules, consist of condensed proteins and RNA. The biophysical characteristics

of proteins that undergo LLPS involve multivalent, relatively low-affinity interactions via intrinsically disordered regions and/or frequent presence of charged amino acids (Babinchak et al., 2019). Indeed, R-rich *C9ORF72* DPRs phase separate with proteins or nucleic acids to form droplets in vitro and preferentially accumulate in MLOs, such as nucleoli and stress granules in cells (Lee et al., 2016; Boeynaems et al., 2017). The R-rich DPRs modulate the dynamics of phase-separated proteins, thereby impeding the functions of MLOs as well as proteins composing MLOs. In particular, poly(PR) inhibits nucleolar nucleophosmin (NPM1) diffusion (White et al., 2019).

Although many peptides undergo LLPS, not all show cytotoxicity. The primary sequence of these is important for cytotoxicity (Ukmar-Godec et al., 2019). For example, while the highly cationic Arg_{12} (R_{12}) peptide phase separates with RNA under in vitro conditions, it is located in the cytosol and has no toxic effect (Meloni et al., 2014; Meloni et al., 2015). By contrast, the $(PR)_{12}$ peptide, carrying the same number of Args, is mainly located in the nucleolus and exerts toxicity by inhibiting protein translation (Kanekura et al., 2018), indicating that the toxicity of cationic peptides is defined not only by the ability to undergo LLPS but also by yet-unknown rules. These lines of evidence prompted us to structurally decode the complex coacervation and toxicity of poly(PR).

Here, we show that the alternating structure of poly(PR) governs LLPS and the resulting toxicity, including inhibition of transcription/translation and impairment of diffusion of NPM1. The toxicity of cationic peptides, including poly(PR), is determined not only by whether it undergoes LLPS but also by the number of molecules each peptide captures in the droplets via multivalent interactions. The molecular dynamics (MD) calculation shows that alternate insertion of Pro into consecutive Args is not advantageous for binding energy, but quantitative proteomic analysis indicates that alternate distribution of Arg is favorable for forming multivalent interaction with client molecules, resulting in LLPS. Poly(PR) variants with large periodicities (e.g., PnRn) have higher binding energy, form sticky droplets with poor fluidity, trap fewer proteins, and fail to exert toxicity. Data mining from the proteomic analysis revealed that proteins harboring acidic stretches and/or abundant acidic amino acids are overrepresented in the $(PR)_{12}$ interactome compared with the R_{12} interactome. Enrichment of proteins with acidic motifs is, at least in part, mediated by LLPS. Thus, we propose that entrapment of diverse proteins with acidic motifs by poly(PR) via multivalent interactions underlies motor neuronal cell death in C9-ALS.

Results

Size of the periodicity determines the toxicity of poly(PR)

The most remarkable biochemical properties of R-rich DPRs [poly(PR) and poly(GR)] are repetitive structures and an extraordinarily highly charged nature due to abundant Arg residues. Arg is a positively charged, highly polar amino acid that interacts with negatively charged molecules, such as acidic proteins and nucleic acids. Protein–protein and protein–nucleic acid interactions are expected to confer cytotoxic properties to

R-rich DPRs. Poly-Arg peptide, often used as a drug delivery carrier due to its cell-permeability properties, is highly charged and interacts with anionic molecules in a similar manner (Kosuge et al., 2008) but exerts no cytotoxic effect (Meloni et al., 2014; Meloni et al., 2015). Therefore, we speculated that the distribution of charged Args in a peptide primary sequence determines its toxicity. To study the role of distribution of Arg in poly(PR)-mediated toxicity, we adopted two biological models, protein translation and diffusion of nucleolar NPM1, which are reportedly inhibited by poly(PR) (Kanekura et al., 2016; Hartmann et al., 2018; Moens et al., 2019; White et al., 2019).

First, we investigated the effect of Arg distribution in the inhibition of translation by poly(PR). Real-time monitoring of fluorescence from GFP produced by in vitro translation (IVT) using HeLa cell lysates revealed that $(PR)_{12}$, but not R_{12}, affects protein translation (Fig. 1 A), suggesting that alternately inserting Pro residues into a continuous R_{12} sequence confers the peptide with ability to inhibit protein translation. To investigate the effect of Arg distribution on the inhibition of protein translation, we synthesized a series of $(PR)_{12}$ variants with different periodicities (size of periodicity, t, = 2, 4, 8, 24; Fig. 1 B); proportionally different variants of $(PR)_{12}$ with equal numbers of Arg, $(P_1R_2)_6$ (Pro:Arg ratio = 1:2) and $(P_1R_3)_4$ (Pro:Arg ratio = 1:3); or variants of $(PR)_{12}$ with equal numbers of Pro, $(P_2R_1)_6$ (Pro:Arg ratio = 2:1), and $(P_3R_1)_4$ (Pro:Arg ratio = 3:1; Fig. 1 C). The IVT-HeLa assay showed that $(PR)_{12}$ effectively suppressed translation of GFP, whereas $(PR)_{12}$ variants with larger periodicities showed a much milder effect compared with $(PR)_{12}$ (Fig. 1 D). The IVT-HeLa assay also showed that an increment of Pro residues up to a 1:1 ratio conferred the peptide with the ability to inhibit protein translation, whereas $(PR)_{12}$ variants with fewer Arg residues (<50%) could not inhibit protein translation (Fig. 1 E). Therefore, periodical insertion of Pro into consecutive R_{12} ≤50% confers toxicity to the peptide. Next, we investigated whether the periodicity of poly(PR) affects protein translation in live cells by expressing a series of $(PR)_{50}$ variants (Fig. 1 F). When we measured protein translation rate in live cells by puromycin labeling, GFP-$(PR)_{50}$ inhibited protein translation most, and GFP-$(PR)_{50}$ variants with large periodicity failed to inhibit (Fig. 1, G and H). Intriguingly, we found that GFP-$(PR)_{50}$ accumulated almost exclusively in the nucleolus; however, GFP-$(PR)_{50}$ variants with large periodicity localized to both the nucleolus and cytosol, and GFP-$(P_{16}R_{16})_3$ failed to localize to the nucleolus and localized almost exclusively to the cytosol (Fig. 1 G, lower panel, and Fig. 1 I).

Next, we tested whether the size of periodicity of $(PR)_{50}$ variants affects the diffusion of NPM1 (White et al., 2019). FRAP of nucleolar GFP-$(PR)_{50}$ variants showed that GFP-$(PR)_{50}$ had higher fluidity than GFP-$(P_8R_8)_6$ (Fig. 1 J), indicating that the interaction between $(PR)_{50}$ and surrounding molecules in the nucleolar matrix was relatively loose compared with that of $(P_8R_8)_6$. Next, we tested the effect of $(PR)_{50}$ variants on the diffusion of GFP-NPM1. As reported previously, $(PR)_{50}$ suppressed the mobility of GFP-NPM1; however, $(P_4R_4)_{12}$ had a modest effect and $(P_8R_8)_6$ had little effect on it (Fig. 1, K and L). To quantitatively evaluate FRAP results, we used a simplified equation to extract diffusion coefficients from FRAP data using half recovery time (Materials and methods). Half recovery time
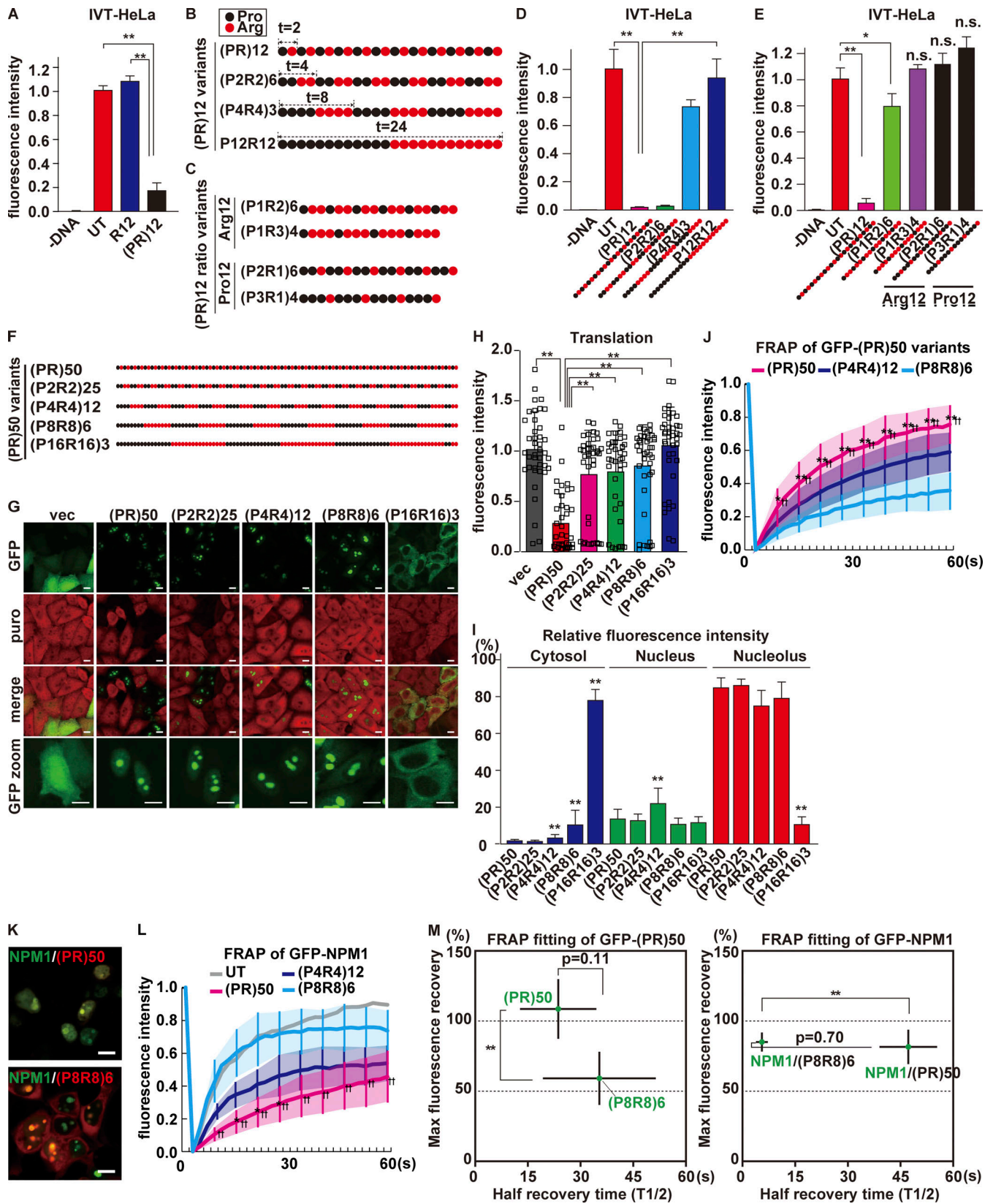
Chen et al.
Structural decoding of C9orf72 poly(PR)

Journal of Cell Biology   2 of 17
https://doi.org/10.1083/jcb.202103160

Figure 1. Size of periodicity determines poly(PR) toxicity. (A) IVT assay using HeLa cell lysates (IVT-HeLa) with GFP cDNA. n = 3 biological replicates. **(B)** Structures of $(PR)_{12}$ variants with different periodicities (repeat size: t = 2, 4, 8, 24). **(C)** Structures of $(PR)_{12}$ ratio variants. Peptides have either the same number of Arg residues (12 Arg) or the same number of Pro resides (12 Pro). **(D)** IVT-HeLa assay with $(PR)_{12}$ variants. **(E)** IVT-HeLa assay with $(PR)_{12}$ ratio variants. **(F)** Structures of $(PR)_{50}$ variants with different periodicities (repeat size: t = 2, 4, 8, 16, 32). To equalize the numbers of Pro and Arg to $(PR)_{50}$, P_2R_2 was added to the C-terminus of $(P_4R_4)_{12}$, $(P_8R_8)_6$, and $(P_{16}R_{16})_3$. **(G)** Immunohistochemical evaluation of protein translation rate in HeLa cells expressing GFP-$(PR)_{50}$

variants by puromycin labeling. Scale bars show 10 µm. **(H)** Quantification of puromycin signal intensities in HeLa cells expressing GFP-$(PR)_{50}$ variants. $n = 40$ cells/variant. **(I)** Quantification of subcellular signal intensities of GFP-$(PR)_{50}$ variants. $n = 20$ cells/variant. **(J)** FRAP analysis of GFP-$(PR)_{50}$, GFP-$(P_4R_4)_{12}$, or GFP-$(P_8R_8)_6$ expressed in HEK293 cells. The FRAP recovery rates of GFP-$(PR)_{50}$ variants were compared. $n = 10$ cells/variant. Asterisks indicate significant differences between $(PR)_{50}$ and $(P_4R_4)_{12}$: *, $P < 0.05$; **, $P < 0.01$. Daggers indicate significant differences between $(PR)_{50}$ and $(P_8R_8)_6$: †, $P < 0.05$; ††, $P < 0.01$. **(K)** Subcellular localization of GFP-NPM1 and mCherry-$(PR)_{50}$, or mCherry-$(P_8R_8)_6$, expressed in HEK293 cells. Scale bars show 10 µm. **(L)** FRAP analysis of GFP-NPM1 coexpressed with mCherry-$(PR)_{50}$, mCherry-$(P_4R_4)_{12}$, or mCherry-$(P_8R_8)_6$. The FRAP recovery rates of GFP-NPM1–coexpressed mCherry-$(PR)_{50}$ variants were compared. $n = 10$ cells/variant. Asterisks indicate significant differences between $(PR)_{50}$ and $(P_4R_4)_{12}$: *, $P < 0.05$; **, $P < 0.01$. Daggers indicate significant differences between $(PR)_{50}$ and $(P_8R_8)_6$: †, $P < 0.05$; ††, $P < 0.01$. **(M)** Fitting of FRAP curves of GFP-$(PR)_{50}$ variants or GFP-NPM1 coexpressed with $(PR)_{50}$ variants. All data are represented as mean ± SD ($n = 3$ biological replicates otherwise mentioned). Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test; *, $P < 0.05$; **, $P < 0.01$, n.s., not significant (A, D, E, H, I, and M).

---

indicated that diffusion of NPM1 coexpressed with $(PR)_{50}$ was much slower than when coexpressed with $(P_8R_8)_6$, implying that more NPM1 molecules are affected by $(PR)_{50}$ than by $(P_8R_8)_6$ (Fig. 1 M). These data clearly indicated that toxicity of poly(PR) was dependent on alternating Arg distribution.

Alternating Arg distribution regulates the $(PR)_{12}$ interactome
In our study, alternate insertion of Pro residues into consecutive R_{12} conferred the ability to inhibit protein translation on the peptide. Conventional proteomics indicates that poly(PR) preferentially interacts with RNA-binding proteins and ribosomal proteins to inhibit protein translation (Lee et al., 2016; Hartmann et al., 2018; Moens et al., 2019). These results suggest that the alternate distribution of Arg, along with the amino acid that alternates Arg, may determine the interactome of R-rich DPRs, which concomitantly results in cytotoxicity. The $P_{12}R_{12}$ variant, which carries exactly the same components as $(PR)_{12}$, such as the number of Args, number of Pros, and length of peptide, may enhance or inhibit protein–protein interactions to diminish the harmful properties of $(PR)_{12}$. Thus, we performed quantitative liquid chromatography–mass spectrometry (LC-MS) to identify the interactomes of these peptides (Ueda et al., 2011). Immunoprecipitation (IP) with HEK293 cell lysates followed by LC-MS identified ∼2,000 interacting proteins (Fig. 2 A; Table S1). The $(PR)_{12}$ interactome is composed mostly of nuclear and cytosolic proteins, where poly(PR) mainly localizes (Fig. 2 B). To confirm whether the repeat length impacts the interactome of poly(PR), we compared the $(PR)_{12}$ interactome with that of $(PR)_{50}$ (Lee et al., 2016) and that of $(PR)_{100}$ (Lin et al., 2016) and confirmed that the interactome of $(PR)_{12}$ is quite similar to that of $(PR)_{50}$ (Fig. S2, A and B) and that of $(PR)_{100}$ (Fig. S2 C). The statistical probability of these levels of overlaps were extremely low (P = $1.05 × 10^{-126}$ for $(PR)_{50}$ and $8.57 × 10^{-123}$ for $(PR)_{100}$). Scatterplots and volcano plots showed that $(GR)_{12}$ and R_{12} showed similar enrichment levels in the interactomes ($R^2 = 0.756$; Fig. 2, C and D), whereas enrichment in the interactome of $(PR)_{12}$ was very different from that of R_{12} ($R^2 = 0.415$; Fig. 2, E and F).

When the interactome of R_{12} and that of $(PR)_{12}$ were compared, 1,503 proteins of the 1,921 identified proteins were common to the interactomes of R_{12} and $(PR)_{12}$. The interactomes of $(P_2R_2)_6$, $(P_4R_4)_3$ and $P_{12}R_{12}$ were also qualitatively similar (Fig. S2, D–F). The signals of 171 proteins (11.4%) were increased by ≥10-fold, while the signals of only 7 proteins (0.5%) were decreased to <10% in $(PR)_{12}$, compared with those of R_{12} (Fig. 2 G). Signal intensities of the enriched interactome of $(PR)_{12}$ (>10-fold compared with R_{12}) were, in most cases, reduced in the $P_{12}R_{12}$ interactome (Fig. 2, H–L), suggesting that the enrichment of the interactome was dependent on the periodicity of $(PR)_{12}$. IP followed by immunoblotting (IB) analyses indicated that the signal intensities of coprecipitated proteins were inversely correlated with the size of periodicity (Fig. 2 M). Another IP-IB analysis showed that consecutive P_{12} residues did not contribute to protein–protein interaction, while Pro residues inserted alternately into consecutive R_{12} induced a drastic increase in protein–protein interactions, showing that the role of alternating Pro is promoting multivalent interaction with client proteins (Fig. 2 N). Overall, these findings indicated that (a) peptide–protein interaction was affected by the size of $(PR)_{12}$ periodicity, and (b) the number of molecules that 12 Arg residues interacted with was increased by alternate insertion of Pro.

We obtained 196 proteins in the $(PR)_{12}$ interactome that were enriched at least fivefold, compared with those in the R_{12} interactome, and 294 proteins detected only in the $(PR)_{12}$ interactome (Fig. 2 A). To identify biological processes involving these proteins, we investigated functional protein association networks using the STRING database (Szklarczyk et al., 2019). The >5-fold enriched proteome and $(PR)_{12}$-specific interactome integrate well and form several functional nodes (Fig. 2 O). The PANTHER overrepresentation test (Mi et al., 2019) found that components of many biological processes, which were reportedly impaired by poly(PR), such as RNA splicing (gene ontology [GO]: 0008380; false discovery rate [FDR] = $3.38 × 10^{-32}$), nucleocytoplasmic transport (GO: 0006913; FDR = $1.37 × 10^{-15}$), and translation (GO: 0006412, FDR = $4.73 × 10^{-10}$), were included (Fig. 2 P and Table S2). We found previously unknown molecular targets of poly(PR), such as transcription, DNA-templated (GO: 0006351, FDR = $1.32 × 10^{-13}$; Fig. 2 P). The heatmap showed that all three classes of RNA polymerases may be targeted by $(PR)_{12}$ (Fig. S1). To examine whether RNA polymerase activity is indeed impaired by poly(PR), RNA synthesis was evaluated by incorporation of 5-ethynyl uridine (EU), followed by visualization using Alexa Fluor 594 azide via click chemistry reaction (Fig. S3, A and B). We observed nucleolar RNA staining without treatment, but the signal intensity of newly synthesized RNA in HeLa cells treated with $(PR)_{20}$ was attenuated (Fig. S3, B and C). Furthermore, overexpression of GFP-$(PR)_{50}$ suppressed transcription, and GFP-$(PR)_{50}$ variants with large periodicities failed to suppress (Fig. S3, E and F). These data show that the $(PR)_{12}$-enriched interactome contains molecular targets of poly(PR) pathophysiology.
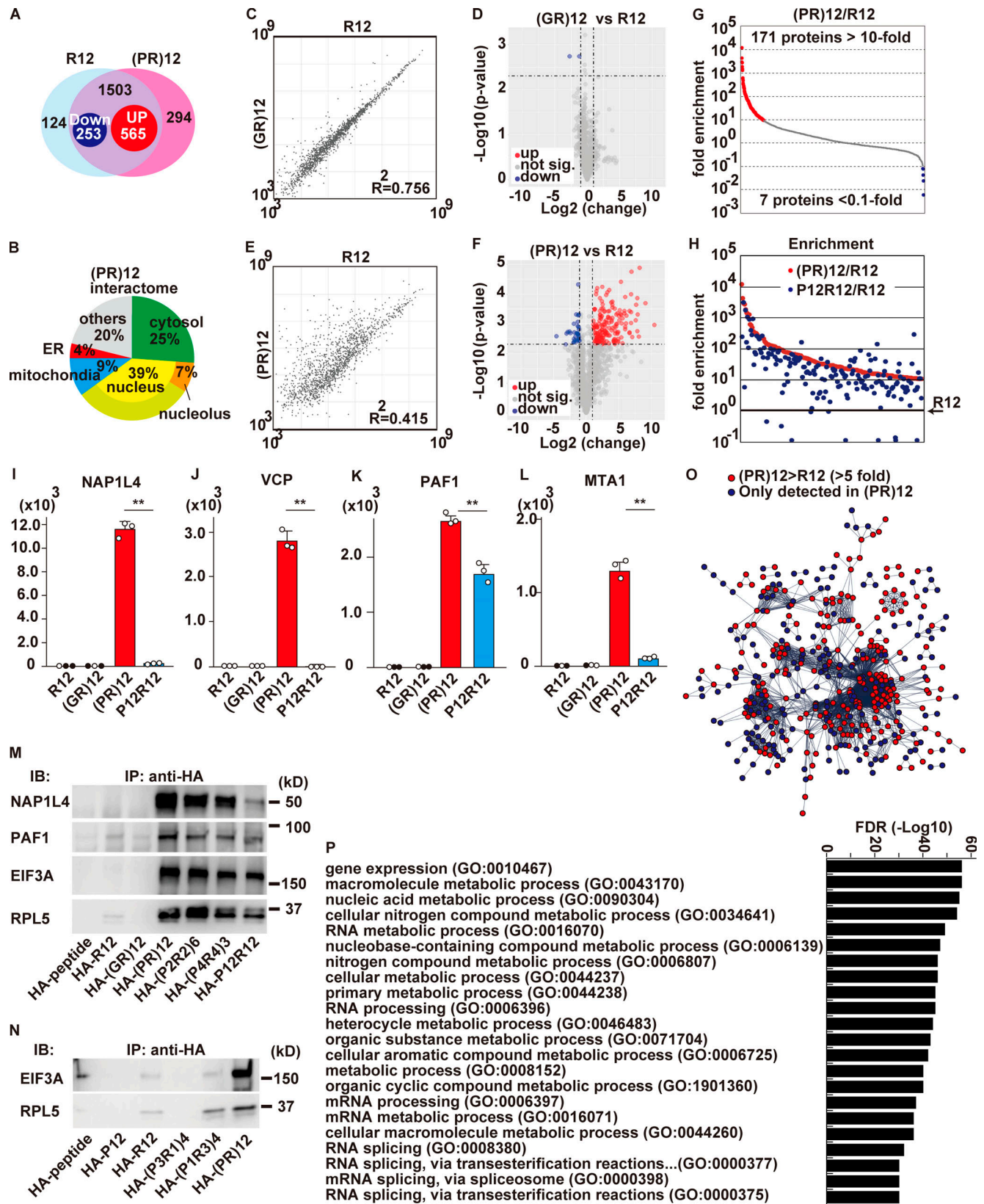
Figure 2. **Alternating Arg structure regulates the interactome of (PR)12.** **(A)** Venn diagram of identified proteins in R12 and (PR)12 interactomes. **(B)** Subcellular localization of (PR)12 interactome. **(C)** Scatterplots and correlation between interactomes of R12 and (GR)12; plots were constructed using average of three samples per condition. **(D)** Volcano plots illustrating significant differentially abundant proteins between R12 and (GR)12 Interactomes. The −log10 (Benjamini–Hochberg corrected P value) plotted against log2 fold-change. Nonaxial vertical lines denote a ±2.0 fold-change, while the nonaxial horizontal line denotes P = 0.005: red circles; up: gray circles, not significant (not sig.): blue circles; down. **(E)** Scatterplots and correlation between interactomes of

Chen et al.
Structural decoding of C9orf72 poly(PR)

Journal of Cell Biology   5 of 17
https://doi.org/10.1083/jcb.202103160

R_{12} and $(PR)_{12}$. **(F)** Volcano plots illustrating significantly differentially abundant proteins between R_{12} and $(PR)_{12}$ interactomes. **(G)** The $(PR)_{12}$ interactome enriched >10-fold (171 proteins; red circles) or enriched less than one tenth (7 proteins; blue circles) compared with R_{12} interactome. **(H)** Attenuation of enrichment of the 171 proteins $[(PR)_{12}/R_{12} > 10]$ shown in G in the $P_{12}R_{12}$ interactome. **(I–L)** Examples of signal intensities of LC-MS for identified proteins. White dots indicate the intensity of each sample; samples without signal are shown as black dots; n = 3 biological replicates. Data are represented as mean ± SD. Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test: ** = P < 0.01. **(M)** IP-IB analysis of HA-tagged $(PR)_{12}$ variants with different periodicities. **(N)** IP-IB analysis of HA-tagged $(PR)_{12}$ variants with different proportions. **(O)** Functional protein association networks of the $(PR)_{12}$-enriched proteome and $(PR)_{12}$-specific interactome visualized via STRING database. Red circles show $(PR)_{12}$-enriched proteome (more than fivefold compared with R_{12}), and blue circles show $(PR)_{12}$-specific interactome. **(P)** Examples of biochemical pathways enriched in $(PR)_{12}$ interactome.

Alternate Pro residues enhance multivalent interactions with acidic motifs

Shared features in the enriched interactome of $(PR)_{12}$ were analyzed to clarify the affinity of $(PR)_{12}$ for these proteins. Sequences of the top 20 enriched proteins tended to have long acidic stretches consisting of consecutive D/E or repetitive clusters of D/E (Fig. 3 A and Fig. S4, A and B). Thus, we screened consecutive acidic stretches in the sequence of proteins with at least a fivefold increase in $(PR)_{12}$ or a decrease of >50% in $(PR)_{12}$, compared with R_{12}, and found that $(PR)_{12}$ preferred proteins with longer acidic stretches (Fig. 3 B). To detect repetitive small clusters of D/E, we defined a weighted scoring system for D/E using 6-aa peptides (D/E score; Fig. 3 C; Materials and methods) and calculated the score from start codon to stop codon of each protein. The total D/E scores of proteins with a greater affinity for $(PR)_{12}$ were significantly higher than those of proteins with a greater affinity for R_{12} (P < 0.0001; Fig. 3, D and E). Next, we calculated the overall amino acid occurrence in the interactome. The whole interactomes of $(PR)_{12}$ and R_{12} showed no remarkable differences, but occurrence of charged residues (D, E, K, and R) were higher than those in the human proteome (Fig. 3 F; Kozlowski, 2017). When we calculated amino acid occurrence in the 4 subgroups (>500-, >200-, >5-fold, or <0.5-fold enriched in $(PR)_{12}$ compared with R_{12}, respectively), we found that acidic amino acids were overrepresented in the $(PR)_{12}$ interactome compared with R_{12} (Fig. 3 G). Aromatic acids were not enriched, suggesting that electrostatic forces rather than cation-π interactions play a central role in determining the $(PR)_{12}$ interactome (Fig. 3 G).

When compared with the interactome of R_{12}, the signal intensities of some $(PR)_{12}$ interactors were drastically increased, >10,000-fold (Fig. 2 G). Even if 1 $(PR)_{12}$ molecule interacts with 12 client molecules via each R and R_{12} interacts with only 1 molecule, the maximum difference can be 12-fold, so that such an unusually high condensation process may be explained by LLPS formation. Overrepresentation of charged amino acids in the $(PR)_{12}$-enriched interactome substantiates this postulate. To confirm this, we mixed recombinant proteins with R_{12}, $(PR)_{12}$, or $P_{12}R_{12}$. Recombinant human Importin-7 (IPO7) and recombinant human ARF GTPase activating protein 1 (GIT1) phase separated with $(PR)_{12}$ but not with R_{12} or $P_{12}R_{12}$, indicating that alternating Arg distribution facilitates LLPS (Fig. 3, H–K). We investigated whether 1,6-hexanediol, an aliphatic alcohol that disrupts LLPS, affects the IP assay (Lin et al., 2016). When the concentration of 1,6-hexanediol in the lysate was increased, the immunoprecipitated NAP1L4 signal, which is the most enriched interactor of $(PR)_{12}$, faded accordingly (Fig. 3 L). These data revealed that enrichment of the interactome may be explained, at least partially, by LLPS.

Periodicity of $(PR)_{12}$ regulates phase separation

LLPS formation with recombinant IPO7 and GIT1 indicates that the periodicity of $(PR)_{12}$ regulates phase separation. Studies indicate that the distribution of charged amino acids affects electrostatic force and phase separation, and especially blocky sequences of charged residues exhibit strong charge correlations due to the sequence alignment of two nearby chains, which favors binding (Chang et al., 2017). To understand how periodicity of $(PR)_{12}$ regulates LLPS, we tested LLPS formation with poly-rA as a model for RNA or with poly-E as a model for acidic stretches of proteins. Phase-separated droplets of $(PR)_{12}$ variants and RNA wetted the hydrophobic glass surface similarly (Fig. 4 A). However, when glass coated with a hydrophilic polymer was used, most droplets floated, and fewer droplets adhered to the surface (Fig. 4 B). Droplets of $(PR)_{12}$ variants with larger periodicities lost their spherical structure under this condition, and $P_{12}R_{12}$ formed sticky nonspherical condensates that failed to minimize surface tension, a typical characteristic of liquids (Fig. 4 B). FRAP analysis of tetramethylrhodamine (TAMRA)-labeled RNA mixed with each peptide showed that RNA mixed with $(PR)_{12}$ had higher fluidity than RNA mixed with $(P_4R_4)_3$ or $P_{12}R_{12}$ (Fig. 4 C). When mixed with poly-E, $(PR)_{12}$ and $(P_2R_2)_6$ formed phase-separated droplets (Fig. 4 D). $(P_4R_4)_3$ also formed droplets, but they contained aggregates of poly-E. Importantly, $P_{12}R_{12}$ did not form droplets.

Considering that fluidity of $(PR)_{12}$ droplets is higher than that of $(PR)_{12}$ variants with large periodicities (Fig. 4 C), Pro residues may expectedly act as looseners for the Arg–RNA interaction and Arg–poly-E interaction. Further substantiating this, whole droplet bleaching showed that RNA in the $(PR)_{12}$ droplet fraction was more capable of entering and exiting than RNA in R_{12} droplets (Fig. 4 E). To further prove this, we tested the effects of $(PR)_{12}$ variants with different proportions on LLPS. When mixed with RNA, $(PR)_{12}$, $(P_1R_2)_6$, and $(P_1R_3)_4$ formed droplets (Fig. 4 F), whereas $(P_2R_1)_6$ and $(P_3R_1)_4$ did not (data not shown). FRAP analysis indicated that fluidity of the droplets increased with the ratio of Pro residues (Fig. 4 G). We also observed the same tendency in phase separation and FRAP analysis when mixed with poly-E peptide, indicating that alternate Pro residues loosened protein–protein interactions and protein–RNA interactions. (Fig. 4, G and I). Visible droplets of $(P_2R_1)_6$ and $(P_3R_1)_4$ were not observed when mixed with poly-E (data not shown). To confirm if the effect of the distribution of Arg on FRAP recovery rate can be seen in the physiologically relevant size of poly(PR), we performed FRAP analysis with recombinant $(PR)_{50}$ and $(P_{16}R_{16})_3$. As expected, the FRAP recovery rate of $(P_{16}R_{16})_3$ was slower than that of $(PR)_{50}$ when mixed with RNA or poly-E (Fig. S5, A and B).
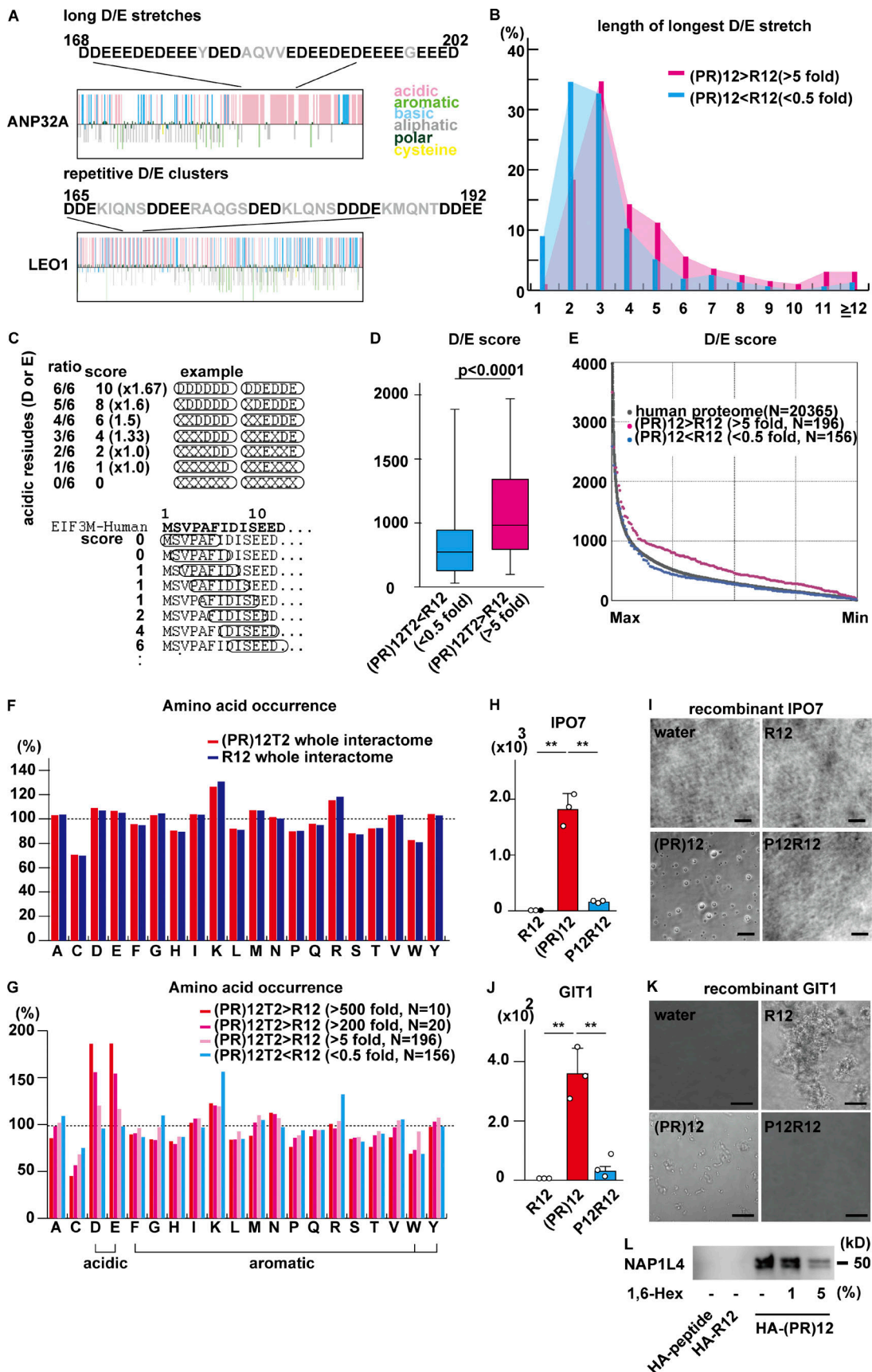
Chen et al.
Structural decoding of C9orf72 poly(PR)

Journal of Cell Biology 6 of 17
https://doi.org/10.1083/jcb.202103160

Figure 3. **Alternate Pro residues enhance multivalent interaction with acidic motifs. (A)** Amino acid compositions of ANP32A and LEO1 as visualized via PepCalc.com (https://pepcalc.com). ANP32A exemplifies proteins with long acidic stretches and LEO1 exemplifies proteins with repetitive D/E clusters.

(B) Histogram of lengths for longest D/E stretches in the enriched interactomes. **(C)** Scheme of D/E scoring weighing continuity. Six consecutive amino acids were serially scored and summed. **(D)** Boxplot of D/E scores of enriched interactome of R_{12} and $(PR)_{12}$. Whiskers are drawn down to 5th and 95th percentiles. Differences were analyzed with Mann–Whitney U test. **(E)** D/E score of enriched interactome of $(PR)_{12} < R_{12}$ (<0.5-fold; blue circles), interactome of $(PR)_{12} > R_{12}$ (>5-fold; red circles), and human proteome. 13 proteins whose D/E scores are higher than 4,000 were omitted from human proteome for graphical clarity. **(F)** Amino acid occurrence in the whole interactome of R_{12} and $(PR)_{12}$, normalized to human proteome. **(G)** Amino acid occurrence in the enriched interactome of $(PR)_{12}$ compared with that of R_{12} (>500-fold, >200-fold, >5-fold, or <0.5-fold), normalized to human proteome. **(H)** Signal intensities of LC-MS for IPO7. **(I)** Phase separation of IPO7 mixed with R_{12}, $(PR)_{12}$, or $P_{12}R_{12}$ observed by confocal microscopy. **(J)** Signal intensities of LC-MS for GIT1. **(K)** Phase separation of GIT1 mixed with R_{12}, $(PR)_{12}$, or $P_{12}R_{12}$, observed by confocal microscopy. The scale bar shows 10 μm. **(L)** IP-IB assay of HA-tagged R_{12} or $(PR)_{12}$ mixed with HEK293 cell lysates containing indicated concentrations of 1,6-hexanediol. Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test; **, P < 0.01.

We further investigated the effects of $(PR)_{12}$ variants on LLPS with proteins harboring acidic stretches. NPM1 phase separates with poly(PR) via its acidic tract-3 domain consisting of 20 consecutive acidic amino acids (White et al., 2019), and periodicity of $(PR)_{50}$ regulates the diffusion of NPM1 in cells (Fig. 1 K–M). Therefore, we speculated that periodicity of $(PR)_{12}$ may regulate the phase separation of NPM1 and that modulation of LLPS underlies impaired diffusion of NPM1 in cells. When recombinant human NPM1 was mixed with $(PR)_{12}$ variants, NPM1 phase-separated with $(PR)_{12}$ variants and droplets showed wetting properties on the glass surface, except those of $P_{12}R_{12}$, which did not phase-separate with NPM1 (Fig. 4 J). NPM1 was shown to phase-separate with ribosomal RNA (rRNA; White et al., 2019). We tested whether $(PR)_{12}$ variants affect the phase separation of NPM1 and rRNA. Recombinant human NPM1 phase separated with human rRNA, and when $(PR)_{12}$ was added, the signal intensities of NPM1 and rRNA in the droplets were increased, indicating that more NPM1 molecules were captured in the phase-separated droplets (Fig. 4, K and L). When mixed with $(P_4R_4)_3$, NPM1 formed large droplets, but rRNA formed fibrillar structures without liquid properties (Fig. 4 K). $P_{12}R_{12}$ failed to form droplets with NPM1, resulting in rRNA fibrils (Fig. 4 K). This indicated that $(PR)_{12}$ addition resulted in increased NPM1 and rRNA involvement in LLPS droplets, causing droplets to become condensed, whereas addition of $(PR)_{12}$ variants with large periodicities did not. Based on the results of in vitro and in cellulo experiments, we speculated that the rate of dynamic exchange of NPM1 in phase equilibrium was disrupted by complex coacervation with $(PR)n$. $(PR)n$ condenses NPM1 and rRNA in phase-separate droplets via multivalent interaction, thereby affecting the mobility of NPM1 in cells, whereas PnRn failed to condense NPM1 and accumulated to rRNA. Thus, PnRn did not significantly affect NPM1 mobility (Fig. 1 M and Fig. 4, K and L).

Alternate distribution of Arg is not advantageous with respect to binding energy

As $(PR)_{12}$ variants have the same length and density of Arg, the only difference is in the distribution of Arg. To determine the manner in which Arg distribution affects molecular interaction, we performed MD simulation to calculate the binding energy between RNA and $(PR)_4$ variants with different distributions of Arg residues (sizes of periodicity t = 2, 4, 8; Fig. 5 A). This calculation focused on one-by-one interactions between a peptide and RNA in water (Fig. 5 B). We analyzed the root mean square deviations (RMSDs) for each peptide by comparing it with the

reference structure. $(PR)_4$ and P_4R_4 had relatively small and constant values >30 ns, while $(P_2R_2)_2$ had a larger RMSD (Fig. 5 C), suggesting that proline plays an important role in the rigidity of the peptide during its interaction with RNA. $(PR)_4$ had the smallest RMSD among the peptides, because the prolines were inserted evenly in the sequence. On the other hand, $(P_2R_2)_2$ showed the largest RMSD, indicating that RR was more flexible than PR. P_4R_4 had the highest binding energy, while $(PR)_4$ had the lowest (Fig. 5 E). We examined each binding energy component according to the equation $Eb = Evdw + Eele + Epol + Enonpol$, where Evdw is van der Waals energy, Eele is the electrostatic energy, Epol is polar solvation energy and Enonpol is nonpolar solvation energy. The first two represent molecular interactions, while the last two are related to solvation energy. Both van der Waals forces and electrostatic forces showed negative values, and absolute values increased in the order of $(PR)_4$, $(P_2R_2)_2$ and P_4R_4 (Fig. 5 E). Interestingly, Epol was positive for all peptides, while $(PR)_4$ displayed a significantly smaller value than the others, indicating that solvation of water was significantly modulated by the presence and spatial distribution of Pro. Decomposed energies for each residue in the peptide sequence and RNA (poly-A) are shown (Fig. 5, F–I). While Arg residues contributed to binding energy, the Epol of each Arg depended on the peptide sequence. $(PR)_4$ displayed a relatively suppressed Epol, due to the presence of water molecules between $(PR)_4$ and RNA. However, P_4R_4 showed a relatively large Epol, suggesting that water molecules surrounding the peptide were excluded from the interaction between peptides and RNA. We also analyzed the radius of gyration (Rg), which approximately estimates the compactness of molecules. We plotted the free energy of the system with respect to Rg and RMSD (Fig. 5, J–M). The energy minima of $(PR)_4$ had a larger Rg value and a smaller RMSD, whereas R_8 was distributed at a smaller Rg and a larger RMSD, indicating that Pro residues conferred rigidity to the structure of peptides. We propose that such structural stability with relatively extended conformation intervenes in the interaction between cationic amino acids and anionic protein/nucleic acids. Our proposal would also be supported by relevant work demonstrating the folding of PR into a helical structure (Edun et al., 2021).

Next, we verified MD simulation results by other experiments. The strength of binding was determined using critical salt concentration (Alberti et al., 2019). Increasing salt concentration interferes with ionic interaction and hinders LLPS. When the concentration of NaCl in the mixture containing phase-separated droplets consisting of $(PR)_{12}$ variants and RNA was increased, the $P_{12}R_{12}$ droplets survived up to 800 mM, whereas
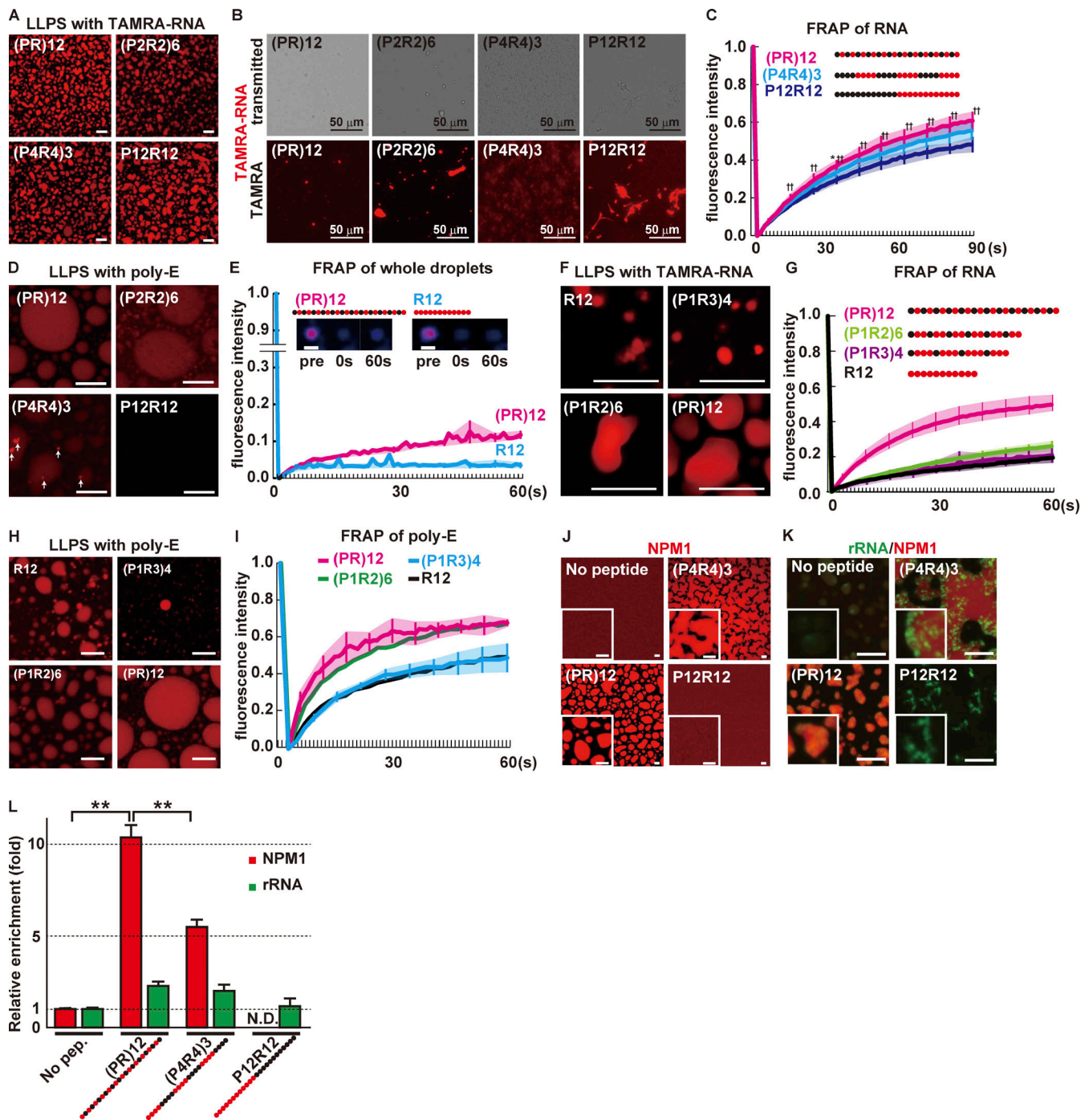
Chen et al.
Structural decoding of C9orf72 poly(PR)

Journal of Cell Biology   8 of 17
https://doi.org/10.1083/jcb.202103160

Figure 4. **Periodicity of (PR)$_{12}$ regulates phase separation. (A)** LLPS of (PR)$_{12}$ variants with poly-rA containing TAMRA-labeled rA$_{15}$ on a glass surface. **(B)** LLPS of (PR)$_{12}$ variants with poly-rA containing TAMRA-labeled rA$_{15}$ on a hydrophilic MAS-coated glass surface. The scale bars show 50 μm. **(C)** FRAP analysis of TAMRA-RNA phase separated with (PR)$_{12}$ variants. Lines are average of five droplets. Asterisks indicate significant differences between (PR)$_{12}$ and (P$_4$R$_4$)$_3$: *, P < 0.05; **, P < 0.01. Daggers indicate significant differences between (PR)$_{12}$ and P$_{12}$R$_{12}$: †, P < 0.05; ††, P < 0.01. **(D)** (PR)$_{12}$ variants with different periodicity phase separated with HiLyte555-labeled poly-E. White arrows show aggregates. **(E)** Whole-droplet FRAP analysis of R$_{12}$ or (PR)$_{12}$ mixed with RNA. Lines show average of six droplets/condition. The scale bars show 1 μm. **(F)** Phase-separated droplets of (PR)$_{12}$ ratio variants mixed with poly-rA containing TAMRA-labeled rA$_{15}$. **(G)** FRAP analysis of TAMRA-RNA phase separated with (PR)$_{12}$ ratio variants. Lines show average of 8–10 droplets/condition. **(H)** (PR)$_{12}$ ratio variants phase separated with poly-E labeled with HyLite555. **(I)** FRAP analysis of poly-E–HyLite555 mixed with (PR)$_{12}$ ratio variants. Lines show the average of eight droplets/condition. Error bars of (P$_1$R$_2$)$_6$ and (P$_1$R$_3$)$_4$ were omitted for graphical clarity. **(J)** Recombinant human NPM1 labeled with HyLite555 mixed with (PR)$_{12}$ variants. **(K)** Recombinant human NPM1 mixed with (PR)$_{12}$ variants and human rRNA. **(L)** Signal enrichment of NPM1 and rRNA in the phase-separated droplets; n = 10 per condition. The scale bars show 10 μm unless otherwise mentioned. All data are represented as mean ± SD. Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test. **, P < 0.01; N.D., not done; n.s., not significant.
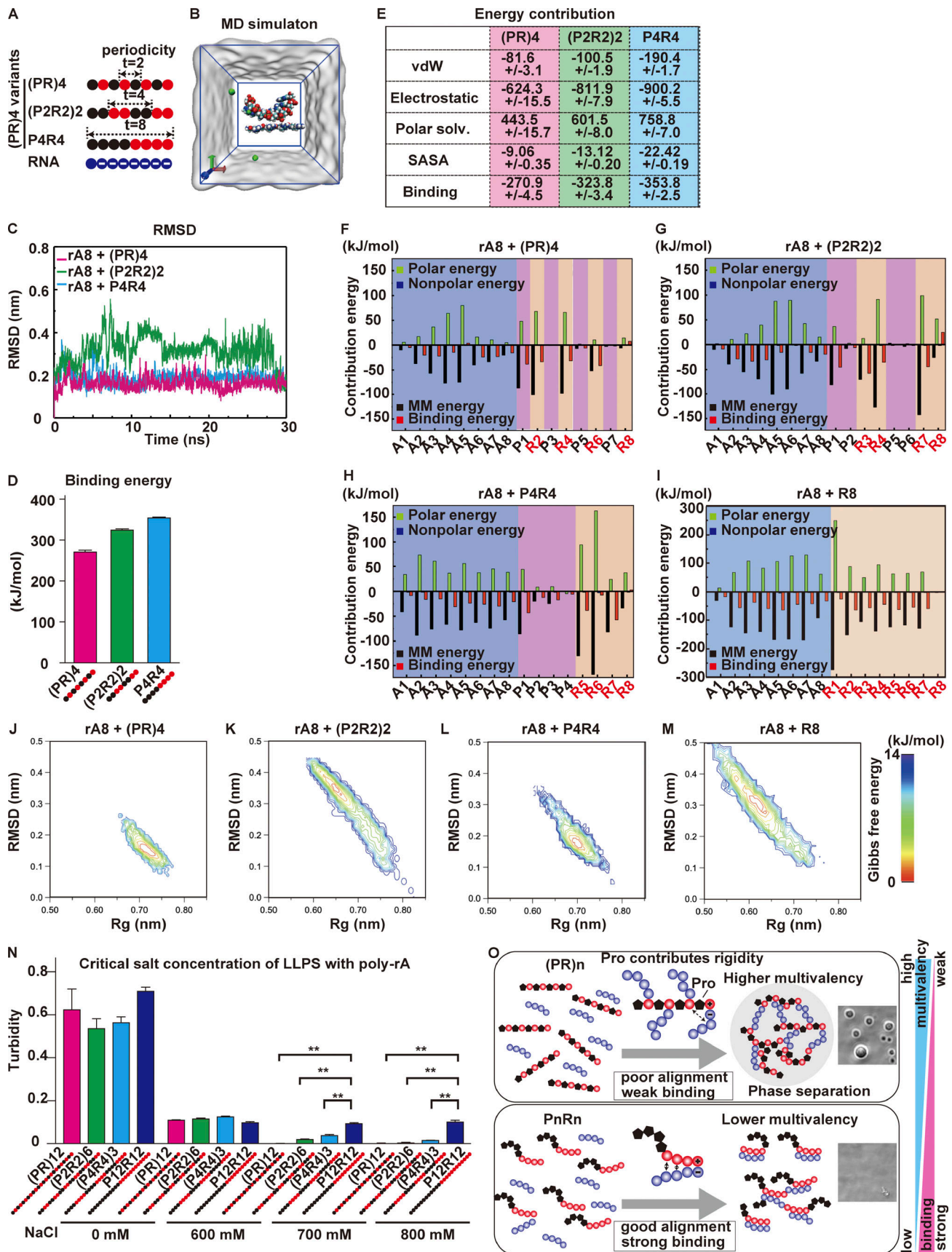
Figure 5. **Alternate distribution of Arg is not advantageous with respect to binding energy. (A)** Structures of (PR)$_4$ variants with different periodicities (t = 2, 4, 8) and rA$_8$ RNA used for MD simulation. Red circles indicate positions of Arg. **(B)** Scheme of MD simulation including one (PR)$_4$ variant, one rA$_8$, and

Chen et al.
Structural decoding of C9orf72 poly(PR)

Journal of Cell Biology   10 of 17
https://doi.org/10.1083/jcb.202103160

Na+ counter ions to equilibrate net charge. **(C)** RMSD during the 30-ns simulations. **(D)** Binding energy between each (PR)4 variant and rA8 RNA determined by MD simulation. n = 3 simulations/each condition. **(E)** Energy contribution of each component revealed that van der Waals force (vdW) and electrostatic force enhanced the binding energy of (PR)4 variants. solv., solvation. **(F–I)** Dissection of energetic contribution into each component of binding energy. MM energy, molecular mechanic energy. **(J–M)** Energy landscape of rA8 and (PR)4 variants and R8. **(N)** Critical salt concentration of (PR)12 variants diminishing phase separation with RNA. Turbidity was determined by OD600; n = 3 biological replicates. All data are represented as mean ± SD. Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test. **, $P < 0.01$. **(O)** Scheme of multivalent interaction and phase separation of (PR)n and PnRn.

(PR)12 droplets remained by 600 mM and dissolved by 700 mM, indicating that a stronger interaction was induced by the blocky charge sequence (Fig. 5, N and O).

Our data indicated that poly(PR) entraps proteins with acidic motifs via multivalent interaction LLPS both in vitro and in cellulo. We propose that the highly multivalent interactions enabled by the alternating distribution of Arg and Pro in poly(PR) contributes to the disturbance of biochemical reactions and the pathophysiology of C9-ALS.

Discussion

ALS is a lethal motor neuron disease affecting >200,000 patients worldwide (Logroscino et al., 2018). To date, almost 30 ALS-causative genes have been identified, and *C9ORF72* has been named as the most prevalent causative gene of familial ALS (Maurel et al., 2018). Here, we shed light on the molecular mechanism underlying poly(PR) toxicity that contributes to pathogenesis of C9-ALS.

Many neurodegenerative disease–related proteins, such as TAR DNA-binding protein 43 (TDP-43), fused in sarcoma (FUS), and tau, reportedly undergo LLPS. These basically phase separate by themselves via intrinsically disordered regions, a process known as simple coacervation (Wegmann et al., 2018; Babinchak et al., 2019; Yoshizawa et al., 2018). Disease-causing mutations in these proteins disturb LLPS homeostasis and contribute to aggregate/amyloid formation. C9-DPRs as well as cationic peptides, such as R12, undergo LLPS when mixed with other molecules, a process known as complex coacervation. Complex coacervation is driven by a combination of electrostatic forces, cation–π interactions, and hydrophobic interactions. Studies indicate that poly(PR) and poly(GR) undergo LLPS in the presence of low-complexity proteins and RNA-binding proteins, leading to cytotoxicity (Lee et al., 2016; Lin et al., 2016). However, the R12 peptide, which lacks substantial toxicity, also undergoes LLPS in a similar manner. Thus, simply undergoing LLPS does not explain toxicity. We hypothesized that an alternatingly charged structure of poly(PR) may play a key role in cytotoxicity. FRAP analyses and critical salt concentration experiments using a series of (PR)12 variants indicated that the order of Pro and Arg determines the environment of phase-separated droplets, concomitantly leading to the differences seen in the inhibition of protein translation.

Based on the highly charged nature of Arg, we surmised that the interactomes of (PR)12 and R12 were qualitatively similar. To understand the pathophysiology of C9-DPRs, we used quantitative proteome analyses to compare their interactomes. Quantitative proteome analyses revealed that R12 interacts with a wide variety of proteins, including ribosomal proteins and RNA-binding proteins, which are reportedly molecular targets of poly(PR) (Fig. 3 C; Radwan et al., 2020; Moens et al., 2019).

Although R12 and (PR)12 share most of the interactomes, they were quantitatively divergent, and alternating insertion of Pro into Arg sequences resulted in multivalent protein interactions, causing an increase in the number of molecules with which one peptide can interact (Fig. 1 M and Fig. 2, F and M). These multivalent interactions were reduced when Arg was consecutively distributed, as shown by minimization of the effect exerted by P12R12 on protein translation. Multivalency of protein interaction by poly(PR) also affects the mobility of phase-separated NPM1 in cells by modulating dynamic phase equilibrium. When a major fraction of NPM1 is entrapped by poly(PR) via weak but multivalent interaction, the ratio of freely mobile NPM1 is decreased, disrupting the overall mobility of NPM1. When a minor portion of NPM1 is entrapped by poly(PR) with large periodicity via strong but less-multivalent interactions, the poly(PR) variant strongly influences the minor fraction of NPM1. However, most NPM1 remain mobile, maintaining the overall mobility of NPM1. This clarifies the absence of toxicity in consecutive polyR, or poly(PR) variants with consecutive Args, as opposed to the toxicity of poly(PR).

Quantitative proteomic analyses also revealed that poly(PR) targets proteins with acidic stretches and/or abundant acidic clusters. Cationic peptides interact with acidic proteins. However, a comparison of the interactomes of R12, (GR)12, and (PR)12 indicated that Pro insertion led to selective enrichment of proteins with acidic motifs up to 10,000-fold when compared with R12. These enriched proteins display extraordinarily high acidic amino acid contents of >30%, suggesting susceptibility to LLPS. IPO7 and GIT1 phase separate with (PR)12 but not with R12 or P12R12, suggesting that extreme enrichment is mediated, at least in part, by LLPS. This is substantiated by the attenuation of NAP1L4 IP by 1,6-hexanediol.

The next issue to be clarified is the determinant of toxicity of poly(PR). MD calculation indicated that structural rigidity promoted by Pro may negatively impact the formation of protein–protein or protein–nucleotide interactions, which may be beneficial for multivalent interactions in the LLPS droplets. This demonstrated that the specific distribution of Arg in poly(PR) plays a pivotal role in phase separation, determination of interactomes, and cytotoxic properties, including inhibition of transcription and translation and disturbed NPM1 dynamics. In addition, the difference of subcellular localization of (PR)50 variants should be noted. It remains unclear why (PR)50 localizes to the nucleolus whereas (P16R16)3 exclusively localizes to the cytosol. In addition, it also remains uncertain why poly(PR) mainly locates in the cytosol rather than the nucleolus in the postmortem tissues of C9-ALS patients, even though poly(PR) locates in the nucleolus when overexpressed in mammalian cells, including human motor neurons derived from C9orf72 patient induced pluripotent stem cells (Wen et al., 2014; Mackenzie

et al., 2015). The localization may be affected by continuous inhibition of nucleocytoplasmic transport (Zhang et al., 2015) or recruitment to cytosol by the interactors of poly(PR) (Hartmann et al., 2018), and further investigation is warranted. Another important problem to be solved is the pathophysiological roles of DPRs. Several poly(PR) transgenic mice exhibit ALS-like phenotypes including motor neuronal loss (Hao et al., 2019; Zhang et al., 2019); however, 500 repeat C9orf72 BAC mice lack motor deficits (Mordes et al., 2020). Although there are several explanations for this, such as that accumulation of RAN translation products increases with age (Cleary and Ranum, 2017) and can be exaggerated in humans due to long lifespans and potential differences in the efficacy of RAN translation in a species-dependent manner, the discrepancy also remains to be investigated. It should be noted that we mainly used $(PR)_{12}$ in this study due to the limitation of the experimental design; however, the number of repeats observed in C9-ALS patients is >50, so we may not have captured all the elements in terms of reproducing human pathology.

We originally postulated that the number of molecules that a DPR binds to is determined by the number of Arg residues. Indeed, it was increased by insertion of alternating Pro residues, which function as spacers to enable the initiation of LLPS. Our results were substantiated by a report indicating that aromatic acids (which contribute to protein–protein interaction via π–π interaction and cation–π interaction and function as "stickers," the appropriately distanced positioning of which is facilitated by "spacers") induce phase separation (Martin et al., 2020). Thus, the balance and distribution of stickers [Arg, in the case of poly(PR)] and spacers [Pro, in the case of poly(PR)] regulate phase separation and subsequent biological events.

In conclusion, we revealed the previously unknown role of charge distribution in poly(PR). Alternate charge distributions govern phase separation, multivalent protein interactions, and cytotoxicity. Thus, it is the disruption of the dynamic exchanges of proteins in the phase equilibrium, and not simply undergoing LLPS, that determines the toxicity of R-rich peptides.

Materials and methods

Reagents
1-Step Human Coupled IVT kit, poly-rA RNA, Click-iT RNA Alexa Fluor 594 imaging kit, and SYTO-RNASelect reagent were purchased from Thermo Fisher Scientific. Recombinant human IPO7 (Ag27257), recombinant human GIT1 (Ag10658), rabbit polyclonal anti-NAP1L4 antibody (16018-1-AP), rabbit polyclonal anti-PAF1 antibody (15441-1-AP), rabbit polyclonal anti-IPO7 antibody (28289-1-AP), rabbit polyclonal anti-RPL5 antibody (15430-1-AP), and rabbit polyclonal anti-EIF3A antibody (11423-1-AP) were purchased from Proteintech. Horseradish peroxidase–conjugated anti-rabbit secondary antibody (#7074S) was purchased from Cell Signaling. Chambered coverglasses (uncoated or hydrophilic polymer MAS-coated) were obtained from Matsunami Glass. 1,6-Hexanediol was purchased from Tokyo Chemical Industry.

LLPS of peptides, nucleotides, and recombinant proteins
All peptides were chemically synthesized by Genscript with purity >85%. Human rRNA was extracted from HEK293 cells

using the RNeasy kit (Qiagen). Human NPM1 cDNA was amplified from NPM1-GFP (gift of Xin Wang, Center for Cancer Research, National Cancer Institute, Bethesda, MD; Addgene plasmid #17578; RRID: Addgene_17578) by PCR and subcloned into the pET28a vector. BL21 (DE3) strain (New England Biolabs) was used for production of recombinant proteins. The expression of recombinant NPM1 was induced by 0.1 mM IPTG at 16°C overnight, and the expressed NPM1 was purified with a Ni-Sepharose column (Clontech) followed by dialysis and snap freezing. For the expression of recombinant GFP-$(PR)_{50}$ and GFP-$(P_{16}R_{16})_3$, cDNAs encoding GFP-$(PR)_{50}$ and GFP-$(P_{16}R_{16})_3$ were subcloned into pET28a vector. The expression of the proteins was induced by 1 mM IPTG at 25°C overnight, and the recombinant proteins were denatured with 6 M guanidium chloride-PBS, followed by purification with Ni-NTA agarose (Thermo Fisher Scientific) and serial dialysis. Poly-E was purchased from Fujifilm. Fluorescent labeling of recombinant NPM1 or poly-E was achieved using a HiLyte-Fluor 555 labeling kit (Dojindo), following the manufacturer's protocol. TAMRA-labeled rA_{15} was synthesized by Fasmac (Kanagawa, Japan). For observation of LLPS of $(PR)_{12}$ variants and poly-rA, each peptide (final concentration, 100 μM) and poly-rA (final concentration, 0.5 mg/ml) containing 100 nM of TAMRA-rA_{15} was mixed, and the droplets were observed on a chambered coverglass via an LSM-710 confocal microscope with a water-immersion 60× lens of NA 1.2 at room temperature (Carl Zeiss). FRAP analyses were also performed using the LSM-710 confocal microscope with a water-immersion 60× lens of NA 1.2 at room temperature, and data were analyzed with Zen software (Carl Zeiss). For observation of LLPS of $(PR)_{12}$ variants and poly-E, each peptide (final concentration, 200 μM) and poly-E (final concentration, 1 mg/ml) containing 1 μg/ml of poly E labeled with HiLyte555 was mixed. For observation of LLPS of $(PR)_{12}$ variants and human NPM1, each peptide (final concentration, 50 μM) and NPM1 (final concentration, 20 μM) containing 0.2 μM of NPM1 labeled with HiLyte555 was mixed. For observation of LLPS of $(PR)_{12}$ variants, human NPM1 and human rRNA, each peptide (final concentration, 50 μM), NPM1 (final concentration, 20 μM) containing 0.2 μM of NPM1 labeled with HiLyte555 and 50 ng/μl of human rRNA was mixed. RNA was visualized using SYTO-RNASelect (Thermo Fisher Scientific). For LLPS with IPO7 or GIT1, 0.2 μg/μL of GST-fused human recombinant IPO7 (Proteintech) or 0.1 μg/μL of His_6-fused human recombinant GIT1 (Proteintech) was mixed with 100 μM/μL of R_{12}, $(PR)_{12}$, $P_{12}R_{12}$, or water and observed by FV10i inverted confocal microscope (Olympus) with a 60× oil-immersion objective lens (NA 1.35) at room temperature.

FRAP analysis of live cells
HEK293 cells cultured in DMEM (Thermo Fisher Scientific) supplemented with 10% FBS and antibiotics were plated on a chambered coverglass (Matsunami). The cells were transiently transfected with GFP-$(PR)_{50}$ variants or GFP-NPM1 in combination with mCherry-$(PR)_{50}$ variants using Lipofectamine 2000 (Thermo Fisher Scientific). FRAP analyses were performed using the LSM-710 confocal microscope with a water-immersion 60× lens of NA 1.2 at 37°C, and data were analyzed with Zen software.

FRAP fitting

To quantitatively evaluate the FRAP results, we used a simplified equation listed below to extract diffusion coefficients from the FRAP data using the half time of recovery (Ladha et al., 1994; Yguerabide et al., 1982):

$$F(t) = \frac{F_0 + F_\infty \left(t/\tau_{1/2}\right)}{1 + \left(t/\tau_{1/2}\right)},$$

where $F(t)$ refers to the fluorescence intensity from the bleached region of interest, F_0 refers to the initial postbleach fluorescence intensity, F_∞ is the postbleach steady-state fluorescence intensity, and $\tau_{1/2}$ is half the time of recovery.

Quantitative LC-MS and IP analyses

1 nmol of HA peptide, HA-tagged $(GR)_{12}$, HA-$(PR)_{12}$, or HA-$P_{12}R_{12}$ was mixed with 200 µg of HEK293 cell lysates. IP was performed with IP buffer (150 mM NaCl, 20 mM Hepes, pH 7.4, 1 mM EDTA, 0.5% Triton X-100, and protease inhibitor cocktails) and 10 µl of anti-HA antibody beads (Fujifilm). After rotation at 4°C for 4 h, the beads were washed 4× with IP buffer and the precipitated proteins were eluted with 2× Laemmli sample buffer (Bio-Rad). The extracted samples were analyzed with LC-MS. Briefly, samples were reduced with 10 mM Tris(2-carboxyethyl)phosphine at 100°C for 10 min, alkylated with 50 mM iodoacetamide at ambient temperature for 45 min, and subjected to SDS-PAGE. Electrophoresis was terminated at a migration distance of 2 mm from the top edge of the separation gel. After Coomassie brilliant blue staining, protein bands were excised, destained, and finely cut before in-gel digestion with Trypsin/Lys-C Mix (Promega) at 37°C for 12 h. The resulting peptides were extracted from gel fragments and analyzed with an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific) combined with Ulti-Mate 3000 RSLC nano-flow HPLC system (Thermo Fisher Scientific) in higher-energy C-trap dissociation MS/MS mode. Peptides were identified and quantified using Proteome Discoverer 2.4 software (Thermo Fisher Scientific), where the MS/MS spectra were searched against the *Homo sapiens* protein database in SwissProt (https://www.uniprot.org/), with an FDR of 1% as an identification filter. The results from the HA control peptide were omitted due to the very low abundance of proteins in the extract. The total counts of peptide fragments were 5.861×10^{10} for all HA peptides.

IP and IB analyses

1 nmol of HA peptide, HA-tagged $(GR)_{12}$, or HA-$(PR)_{12}$ variant was mixed with 200 µg of HEK293 cell lysates. IP was performed as described above. The immunoprecipitated samples were then subjected to SDS-PAGE analyses, followed by IB analyses. The signals were visualized by ECL-select reagent (GE-Amersham) and detected by ChemiDoc touch imaging system (Bio-Rad).

MD simulation

All biomolecular complexes between $(PR)_4$ variants and RNA shown in Fig. 5 were modeled using AVOGADRO software and then studied at pH 7, with both N-terminal Pro residue and C-terminal Arg residue being protonated. MD simulations were performed using the open-source simulation package GROMACS v5.1.4 (Hanwell et al., 2012; Pronk et al., 2015). The amber99sb-ildn forcefield was used for all energetic parameters for inter- and intramolecular interactions (Lindorff-Larsen et al., 2010; Hornak et al., 2006). Simulations were performed in a cubic box ($50 \times 50 \times 50$ Å^3) with periodic boundary conditions applied in all directions, where all complexes were located at the center of the box. The entire system was solvated with an explicit TIP3P water model and neutralized with sodium ions (Mahoney and Jorgensen, 2000; Jorgensen et al., 1983). The initial whole simulation system is shown in Fig. 2 A. Subsequently, energy minimization with the steepest descent method was performed to reach the maximum force <1,000 kJ/mol. The systems were preequilibrated for 1 ns at a constant temperature of 300 K using the V-rescale coupling method in a canonical (number, volume, absolute temperature [NVT]) ensemble and 1 ns at a constant pressure of 1 bar using the Parrinello–Rahman coupling method (Parrinello and Rahman, 1981) in the isothermal-isobaric (number, pressure, absolute temperature [NpT]) ensemble. The coupling constants were set to 0.1 and 2.0 ps for temperature and pressure, respectively. Electrostatic interactions were calculated using the particle mesh Ewald method (Darden et al., 1993), and van der Waals interactions were treated with a smooth cutoff at a distance of 10 Å. The LINCS algorithm (Hess et al., 1997), was used to constrain the bond lengths, thus permitting use of integration time steps of 2 fs. Finally, a 30-ns MD simulation was performed for all the complexes. Simulation trajectories were analyzed using the Visual Molecular Dynamics program (Humphrey et al., 1996), and all trajectories were stored every 10 ps for further analysis. Structural deviations along the trajectories were evaluated using the RMSDs of the protein backbone in Fig. 2 C.

To study the main driving force for the rA$_8$ and $(PR)_4$ variants interaction, the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) analysis was performed using the g_mmpbsa tool of GROMACS (Kumari et al., 2014; Baker et al., 2001). The MM-PBSA method for calculation of the binding energy in general can be expressed as

$$\Delta G_{binding} = G_{complex} - (G_{rA4} + G_{(PR)4}), \quad (1)$$

where $\Delta G_{binding}$ is the binding free energy; and $G_{complex}$, G_{rA4}, and $G_{(PR)4}$ are the free energies of the complex, rA4, and $(PR)4$ variants in the solvent, respectively.

$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S, \quad (2)$$

where ΔE_{MM} is the difference in intramolecular energy under vacuum, ΔG_{sol} is the solvation free energy difference, T is the absolute temperature, and ΔS is the entropy change. The ΔE_{MM} can be calculated using the MM method; ΔG_{sol} is composed of the polar solvation free energy difference and nonpolar solvation free energy difference, and the polar component is obtained by solving the finite difference Poisson–Boltzmann equation. The nonpolar component is fitted by estimating the solvent-accessible surface area; $T\Delta S$ is calculated using the normal mode method and is often omitted completely in the ranking of relative binding affinities, as its inclusion often does not improve the agreement with the experiment (Yang et al., 2011).

By using the g_mmpbsa tool, the average binding energy was calculated from the last 10 ns in the MD trajectories. The dielectric constant of the aqueous solvent was set to 80, the interior dielectric constant was set to 4, and the surface tension constant g was set to 0.022 kJ/mol. The binding energy results for all simulated systems, including the contribution of residues in terms of all components, namely MM energy, polar energy, and nonpolar energy, are shown in Fig. 5, E–I.

Free energy landscape
The free energy landscapes of the dynamic behaviors of the rA_8 and $(PR)_4$ variant complexes were estimated by an appropriate sampling method. To obtain a 2D energy landscape method, the RMSD and Rg of each complex system were chosen as two reaction coordinates. The energy landscape was calculated along these two reaction coordinates using the following equation (Papaleo et al., 2009):

$$G\alpha = -kTln\left[\frac{P(q_\alpha)}{P_{max}(q)}\right],$$

where k is the Boltzmann constant, T is the temperature of the simulation, $P(q_\alpha)$ is an estimate of the probability density function obtained from a histogram of the MD data, and $P_{max}(q)$ is the probability of the most probable state. Considering two different reaction coordinates, q_i and q_j, 2D free-energy landscapes were obtained from the joint probability distributions $P(q_i, q_j)$ of the system.

Data mining
To extract effective features from protein sequence data that contained significant discriminatory information, data mining was performed to obtain accurate predictions. In this work, we analyzed protein sequence data based on Anaconda, a data science platform, by importing the data into Spyder (Python v3.7.3) in Anaconda (https://docs.anaconda.com/anaconda/).

Step 1: Detecting consecutive acidic stretches
Take the protein (i.e., EIF 3M-Human) with part of the whole amino acid sequence. We continually performed fishing from the start codon to the stop codon, and each time the fish consisted of six consecutive amino acids. Likewise, the same fishing process was followed for all other proteins selected, so that all amino acids of each protein could be covered.

Step 2: Scoring
After fishing for all proteins, we performed scoring for all sequences. Here, we focused on the occurrence of D/E amino acids in each sequence consisting of six consecutive amino acids. In particular, if the D/E are consecutive, they are expected to have a greater effect on the electrostatic force (Fig. 1 F). Therefore, we adapted a weighted scoring system as follows. For each sequence, if no D/E amino acids, the score value was 0; if one D/E amino acid, the score value was 1; if two D/E amino acids, the score value was 2; if three D/E amino acids, the score value was 4; if four D/E amino acids, the score value was 6; if five D/E amino acids, the score value was 8; and if six D/E amino acids, the score value was 10. Consequently, all six consecutive amino acids for all proteins were serially scored and summed to obtain the final score for each protein.

IVT
An IVT assay was performed using a one-step human coupled IVT kit (Thermo Fisher Scientific) as reported previously (Kanekura et al., 2016). Briefly, 100 µM of each peptide was added to HeLa cell lysates containing substrates and accessory proteins and incubated at 37°C, and fluorescence was monitored every 1 min using a LightCycler (Roche).

Evaluation of translation rate in cells by puromycin labeling
HeLa cells transiently transfected with $GFP-(PR)_{50}$ variants were treated with 2 µg/ml of puromycin for 1 h. After fixation with 4% formalin-PBS, the cells were permeabilized with 0.5% Triton X-100, followed by immunostaining with anti-Puromycin antibody (mouse monoclonal, clone 12D10; Millipore) at 1:500 dilution and visualized by Alexa Fluor 594–labeled anti-mouse secondary antibody at 1:500 dilution (Thermo Fisher Scientific). The cells were mounted with ProLong Gold mounting reagent (Thermo Fisher Scientific). The images were obtained by FV10i inverted confocal microscope (Olympus) with a 60× oil immersion objective lens (NA 1.35) at room temperature. The signal intensities were determined by ImageJ software.

Evaluation of transcription in cells
HeLa cells treated with 10 µM $(PR)_{20}$ peptide or transiently transfected with $GFP-(PR)_{50}$ variants were incubated with 1 mM of EU for 2 h. After fixation with 4% formalin-PBS, the cells were permeabilized with 0.5% Triton X-100, followed by immunostaining with anti-GFP antibody (rabbit polyclonal; MBL) at 1:500 dilution for $GFP-(PR)_{50}$ variant–expressing cells. Then newly synthesized RNA was visualized by Click-iT RNA Alexa Fluor 594 Imaging kit (Thermo Fisher Scientific). $GFP-(PR)_{50}$ signals were visualized by Alexa Fluor 488–labeled anti-rabbit secondary antibody (Thermo Fisher Scientific). The cells were mounted with ProLong Gold mounting reagent (Thermo Fisher Scientific). The images were obtained by an FV10i inverted confocal microscope (Olympus) with a 60× oil-immersion objective lens (NA 1.35) at room temperature. The signal intensities were determined by ImageJ software.

Statistical analysis
All data are represented as mean values ± SD, and each statistical analysis is detailed in the figure legend. Data distribution was assumed to be normal, but this was not formally tested. Statistical analyses of the data were performed with SPSS software 26 (IBM). n indicates the number of biological replicates in an experiment unless otherwise mentioned.

Online supplemental material
Fig. S1 shows the heatmap for the interactome of R_{12}, $(GR)_{12}$, $(PR)_{12}$, or $P_{12}R_{12}$. Fig. S2 compares the interactome of $(PR)_{12}$ with the interactome of $(PR)_{50}$ or the interactome of $(PR)_{100}$. Fig. S3 shows that the transcription of RNA is potentially targeted by poly(PR) toxicity. Fig. S4 characterizes the 20 most enriched $(PR)_{12}$-interacting proteins. Fig. S5 shows FRAP analyses of

recombinant $(PR)_{50}$ and $(P_{16}R_{16})_3$. Table S1 contains quantitative proteomics data. Table S2 contains GO enrichment data.

Data availability

Acknowledgments

References

Alberti, S., A. Gladfelter, and T. Mittag. 2019. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell.* 176:419–434. https://doi.org/10.1016/j.cell.2018.12.035

Andrade, N.S., M. Ramic, R. Esanov, W. Liu, M.J. Rybin, G. Gaidosh, A. Abdallah, S. Del'Olio, T.C. Huff, N.T. Chee, et al. 2020. Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair in C9ORF72 ALS/FTD. *Mol. Neurodegener.* 15:13. https://doi.org/10.1186/s13024-020-00365-9

Babinchak, W.M., R. Haider, B.K. Dumm, P. Sarkar, K. Surewicz, J.-K. Choi, and W.K. Surewicz. 2019. The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* 294:6306–6317. https://doi.org/10.1074/jbc.RA118.007222

Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome.

Proc. Natl. Acad. Sci. USA. 98:10037–10041. https://doi.org/10.1073/pnas.181342398

Boeynaems, S., E. Bogaert, D. Kovacs, A. Konijnenberg, E. Timmerman, A. Volkov, M. Guharoy, M. De Decker, T. Jaspers, V.H. Ryan, et al. 2017. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell.* 65:1044–1055.e5. https://doi.org/10.1016/j.molcel.2017.02.013

Boeynaems, S., A.S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P.S. Tompa, R.V. Pappu, and A.D. Gitler. 2019. Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. USA.* 116:7889–7898. https://doi.org/10.1073/pnas.1821038116

Chang, L.-W., T.K. Lytle, M. Radhakrishna, J.J. Madinya, J. Vélez, C.E. Sing, and S.L. Perry. 2017. Sequence and entropy-based control of complex coacervates. *Nat. Commun.* 8:1273. https://doi.org/10.1038/s41467-017-01249-1

Choi, S.Y., R. Lopez-Gonzalez, G. Krishnan, H.L. Phillips, A.N. Li, W.W. Seeley, W.-D. Yao, S. Almeida, and F.-B. Gao. 2019. C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. *Nat. Neurosci.* 22:851–862. https://doi.org/10.1038/s41593-019-0397-0

Cleary, J.D., and L.P. Ranum. 2017. New developments in RAN translation: insights from multiple diseases. *Curr. Opin. Genet. Dev.* 44:125–134. https://doi.org/10.1016/j.gde.2017.03.006

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092. https://doi.org/10.1063/1.464397

DeJesus-Hernandez, M., I.R. Mackenzie, B.F. Boeve, A.L. Boxer, M. Baker, N.J. Rutherford, A.M. Nicholson, N.A. Finch, H. Flynn, J. Adamson, et al. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 72:245–256. https://doi.org/10.1016/j.neuron.2011.09.011

Edun, D.N., M.R. Flanagan, and A.L. Serrano. 2021. Does liquid-liquid phase separation drive peptide folding? *Chem. Sci. (Camb.).* 12:2474–2479. https://doi.org/10.1039/D0SC04993J

Frottin, F., F. Schueder, S. Tiwary, R. Gupta, R. Körner, T. Schlichthaerle, J. Cox, R. Jungmann, F.U. Hartl, and M.S. Hipp. 2019. The nucleolus functions as a phase-separated protein quality control compartment. *Science.* 365:342–347. https://doi.org/10.1126/science.aaw9157

Gupta, R., M. Lan, J. Mojsilovic-Petrovic, W.H. Choi, N. Safren, S. Barmada, M.J. Lee, and R. Kalb. 2017. The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. eneuro. ENEURO.0249-16.2017. https://doi.org/10.1523/ENEURO.0249-16.2017

Hanwell, M.D., D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, and G.R. Hutchison. 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 4:17. https://doi.org/10.1186/1758-2946-4-17

Hao, Z., L. Liu, Z. Tao, R. Wang, H. Ren, H. Sun, Z. Lin, Z. Zhang, C. Mu, J. Zhou, and G. Wang. 2019. Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat. Commun.* 10:2906. https://doi.org/10.1038/s41467-019-10956-w

Hartmann, H., D. Hornburg, M. Czuppa, J. Bader, M. Michaelsen, D. Farny, T. Arzberger, M. Mann, F. Meissner, and D. Edbauer. 2018. Proteomics and *C9orf72* neuropathology identify ribosomes as poly-GR/PR interactors driving toxicity. *Life Sci. Alliance.* 1:e201800070. https://doi.org/10.26508/lsa.201800070

Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725. https://doi.org/10.1002/prot.21123

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38: 27–28. https://doi.org/10.1016/0263-7855(96)00018-5

Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935. https://doi.org/10.1063/1.445869

Jovičić, A., J. Mertens, S. Boeynaems, E. Bogaert, N. Chai, S.B. Yamada, J.W. Paul III, S. Sun, J.R. Herdy, G. Bieri, et al. 2015. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* 18:1226–1229. https://doi.org/10.1038/nn.4085

Kanekura, K., T. Yagi, A.J. Cammack, J. Mahadevan, M. Kuroda, M.B. Harms, T.M. Miller, and F. Urano. 2016. Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum. Mol. Genet.* 25: 1803–1813. https://doi.org/10.1093/hmg/ddw052

Kanekura, K., Y. Harada, M. Fujimoto, T. Yagi, Y. Hayamizu, K. Nagaoka, and M. Kuroda. 2018. Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. *Sci. Rep.* 8: 12740. https://doi.org/10.1038/s41598-018-31096-z

Kosuge, M., T. Takeuchi, I. Nakase, A.T. Jones, and S. Futaki. 2008. Cellular internalization and distribution of arginine-rich peptides as a function of extracellular peptide concentration, serum, and plasma membrane associated proteoglycans. *Bioconjug. Chem.* 19:656–664. https://doi.org/10.1021/bc700289w

Kozlowski, L.P. 2017. Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* 45(D1):D1112–D1116. https://doi.org/10.1093/nar/gkw978

Kumari, R., R. Kumar, and A. Lynn. Open Source Drug Discovery Consortium. 2014. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54:1951–1962. https://doi.org/10.1021/ci500020m

Kwon, I., S. Xiang, M. Kato, L. Wu, P. Theodoropoulos, T. Wang, J. Kim, J. Yun, Y. Xie, and S.L. McKnight. 2014. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science.* 345:1139–1145. https://doi.org/10.1126/science.1254917

Ladha, S., A.R. Mackie, and D.C. Clark. 1994. Cheek cell membrane fluidity measured by fluorescence recovery after photobleaching and steady-state fluorescence anisotropy. *J. Membr. Biol.* 142:223–228. https://doi.org/10.1007/BF00234944

Lee, K.-H., P. Zhang, H.J. Kim, D.M. Mitrea, M. Sarkar, B.D. Freibaum, J. Cika, M. Coughlin, J. Messing, A. Molliex, et al. 2016. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell.* 167:774–788.e17. https://doi.org/10.1016/j.cell.2016.10.002

Lin, Y., E. Mori, M. Kato, S. Xiang, L. Wu, I. Kwon, and S.L. McKnight. 2016. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell.* 167:789–802.e12. https://doi.org/10.1016/j.cell.2016.10.003

Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 78:1950–1958. https://doi.org/10.1002/prot.22711

Logroscino, G., M. Piccininni, B. Marin, E. Nichols, F. Abd-Allah, A. Abdelalim, F. Alahdab, S.W. Asgedom, A. Awasthi, Y. Chaiah, et al. GBD 2016 Motor Neuron Disease Collaborators. 2018. Global, regional, and national burden of motor neuron diseases 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 17: 1083–1097. https://doi.org/10.1016/S1474-4422(18)30404-6

Mackenzie, I.R.A., P. Frick, F.A. Grässer, T.F. Gendron, L. Petrucelli, N.R. Cashman, D. Edbauer, E. Kremmer, J. Prudlo, D. Troost, and M. Neumann. 2015. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol.* 130:845–861. https://doi.org/10.1007/s00401-015-1476-2

Mahoney, M.W., and W.L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112:8910–8922. https://doi.org/10.1063/1.481505

Martin, E.W., A.S. Holehouse, I. Peran, M. Farag, J.J. Incicco, A. Bremer, C.R. Grace, A. Soranno, R.V. Pappu, and T. Mittag. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science.* 367:694–699. https://doi.org/10.1126/science.aaw8653

Maurel, C., A. Dangoumau, S. Marouillat, C. Brulard, A. Chami, R. Hergesheimer, P. Corcia, H. Blasco, C.R. Andres, and P. Vourc'h. 2018. Causative Genes in Amyotrophic Lateral Sclerosis and Protein Degradation Pathways: a Link to Neurodegeneration. *Mol. Neurobiol.* 55: 6480–6499. https://doi.org/10.1007/s12035-017-0856-0

Meloni, B.P., A.J. Craig, N. Milech, R.M. Hopkins, P.M. Watt, and N.W. Knuckey. 2014. The neuroprotective efficacy of cell-penetrating peptides TAT, penetratin, Arg-9, and Pep-1 in glutamic acid, kainic acid, and in vitro ischemia injury models using primary cortical neuronal cultures. *Cell. Mol. Neurobiol.* 34:173–181. https://doi.org/10.1007/s10571-013-9999-3

Meloni, B.P., L.M. Brookes, V.W. Clark, J.L. Cross, A.B. Edwards, R.S. Anderton, R.M. Hopkins, K. Hoffmann, and N.W. Knuckey. 2015. Poly-arginine and arginine-rich peptides are neuroprotective in stroke models. *J. Cereb. Blood Flow Metab.* 35:993–1004. https://doi.org/10.1038/jcbfm.2015.11

Mi, H., A. Muruganujan, D. Ebert, X. Huang, and P.D. Thomas. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47(D1): D419–D426. https://doi.org/10.1093/nar/gky1038

Mizielinska, S., S. Grönke, T. Niccoli, C.E. Ridler, E.L. Clayton, A. Devoy, T. Moens, F.E. Norona, I.O.C. Woollacott, J. Pietrzyk, et al. 2014. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science.* 345:1192–1194. https://doi.org/10.1126/science.1256800

Moens, T.G., T. Niccoli, K.M. Wilson, M.L. Atilano, N. Birsa, L.M. Gittings, B.V. Holbling, M.C. Dyson, A. Thoeng, J. Neeves, et al. 2019. C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A. *Acta Neuropathol.* 137:487–500. https://doi.org/10.1007/s00401-018-1946-4

Mordes, D.A., B.M. Morrison, X.H. Ament, C. Cantrell, J. Mok, P. Eggan, C. Xue, J.Y. Wang, K. Eggan, and J.D. Rothstein. 2020. Absence of Survival and Motor Deficits in 500 Repeat C9ORF72 BAC Mice. *Neuron.* 108: 775–783.e4. https://doi.org/10.1016/j.neuron.2020.08.009

Mori, K., S.-M. Weng, T. Arzberger, S. May, K. Rentzsch, E. Kremmer, B. Schmid, H.A. Kretzschmar, M. Cruts, C. Van Broeckhoven, et al. 2013. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science.* 339:1335–1338. https://doi.org/10.1126/science.1232927

Papaleo, E., P. Mereghetti, P. Fantucci, R. Grandori, and L. De Gioia. 2009. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. *J. Mol. Graph. Model.* 27: 889–899. https://doi.org/10.1016/j.jmgm.2009.01.006

Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190. https://doi.org/10.1063/1.328693

Pronk, S., I. Pouya, M. Lundborg, G. Rotskoff, B. Wesén, P.M. Kasson, and E. Lindahl. 2015. Molecular simulation workflows as parallel algorithms: the execution engine of Copernicus, a distributed high-performance computing platform. *J. Chem. Theory Comput.* 11:2600–2608. https://doi.org/10.1021/acs.jctc.5b00234

Radwan, M., C.-S. Ang, A.R. Ormsby, D. Cox, J.C. Daly, G.E. Reid, and D.M. Hatters. 2020. Arginine in C9ORF72 Dipolypeptides Mediates Promiscuous Proteome Binding and Multiple Modes of Toxicity. *Mol. Cell. Proteomics.* 19:640–654. https://doi.org/10.1074/mcp.RA119.001888

Renton, A.E., E. Majounie, A. Waite, J. Simón-Sánchez, S. Rollinson, J.R. Gibbs, J.C. Schymick, H. Laaksovirta, J.C. van Swieten, L. Myllykangas, et al. ITALSGEN Consortium. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 72:257–268. https://doi.org/10.1016/j.neuron.2011.09.010

Szklarczyk, D., A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, et al. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47(D1):D607–D613. https://doi.org/10.1093/nar/gky1131

Ueda, K., N. Saichi, S. Takami, D. Kang, A. Toyama, Y. Daigo, N. Ishikawa, N. Kohno, K. Tamura, T. Shuin, et al. 2011. A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. *PLoS One.* 6:e18567. https://doi.org/10.1371/journal.pone.0018567

Ukmar-Godec, T., S. Hutten, M.P. Grieshop, N. Rezaei-Ghaleh, M.-S. Cima-Omori, J. Biernat, E. Mandelkow, J. Söding, D. Dormann, and M. Zweckstetter. 2019. Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* 10:2909. https://doi.org/10.1038/s41467-019-10792-y

Wegmann, S., B. Eftekharzadeh, K. Tepper, K.M. Zoltowska, R.E. Bennett, S. Dujardin, P.R. Laskowski, D. MacKenzie, T. Kamath, C. Commins, et al. 2018. Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* 37:e98049. https://doi.org/10.15252/embj.201798049

Wen, X., W. Tan, T. Westergard, K. Krishnamurthy, S.S. Markandaiah, Y. Shi, S. Lin, N.A. Shneider, J. Monaghan, U.B. Pandey, et al. 2014. 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. https://doi.org/10.1016/j.neuron.2014.12.010

White, M.R., D.M. Mitrea, P. Zhang, C.B. Stanley, D.E. Cassidy, A. Nourse, A.H. Phillips, M. Tolbert, J.P. Taylor, and R.W. Kriwacki. 2019. C9orf72 Poly(PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. *Mol. Cell.* 74:713–728.e6. https://doi.org/10.1016/j.molcel.2019.03.019

Yang, T., J.C. Wu, C. Yan, Y. Wang, R. Luo, M.B. Gonzales, K.N. Dalby, and P. Ren. 2011. Virtual screening using molecular simulations. *Proteins.* 79: 1940–1951. https://doi.org/10.1002/prot.23018

Yguerabide, J., J.A. Schmidt, and E.E. Yguerabide. 1982. Lateral mobility in membranes as detected by fluorescence recovery after photobleaching. *Biophys. J.* 40:69–75. https://doi.org/10.1016/S0006-3495(82)84459-7

Yoshizawa, T., R. Ali, J. Jiou, H.Y.J. Fung, K.A. Burke, S.J. Kim, Y. Lin, W.B. Peeples, D. Saltzberg, M. Soniat, et al. 2018. Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell.* 173:693–705.e22. https://doi.org/10.1016/j.cell.2018.03.003

Zhang, K., C.J. Donnelly, A.R. Haeusler, J.C. Grima, J.B. Machamer, P. Steinwald, E.L. Daley, S.J. Miller, K.M. Cunningham, S. Vidensky, et al. 2015. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature.* 525:56–61. https://doi.org/10.1038/nature14973

Zhang, Y.-J., T.F. Gendron, M.T.W. Ebbert, A.D. O'Raw, M. Yue, K. Jansen-West, X. Zhang, M. Prudencio, J. Chew, C.N. Cook, et al. 2018. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat. Med.* 24:1136–1142. https://doi.org/10.1038/s41591-018-0071-1

Zhang, Y.-J., L. Guo, P.K. Gonzales, T.F. Gendron, Y. Wu, K. Jansen-West, A.D. O'Raw, S.R. Pickles, M. Prudencio, Y. Carlomagno, et al. 2019. Heterochromatin anomalies and double-stranded RNA accumulation underlie *C9orf72* poly(PR) toxicity. *Science.* 363:eaav2606. https://doi.org/10.1126/science.aav2606

Zu, T., B. Gibbens, N.S. Doty, M. Gomes-Pereira, A. Huguet, M.D. Stone, J. Margolis, M. Peterson, T.W. Markowski, M.A. Ingram, et al. 2011. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. USA.* 108:260–265. https://doi.org/10.1073/pnas.1013343108
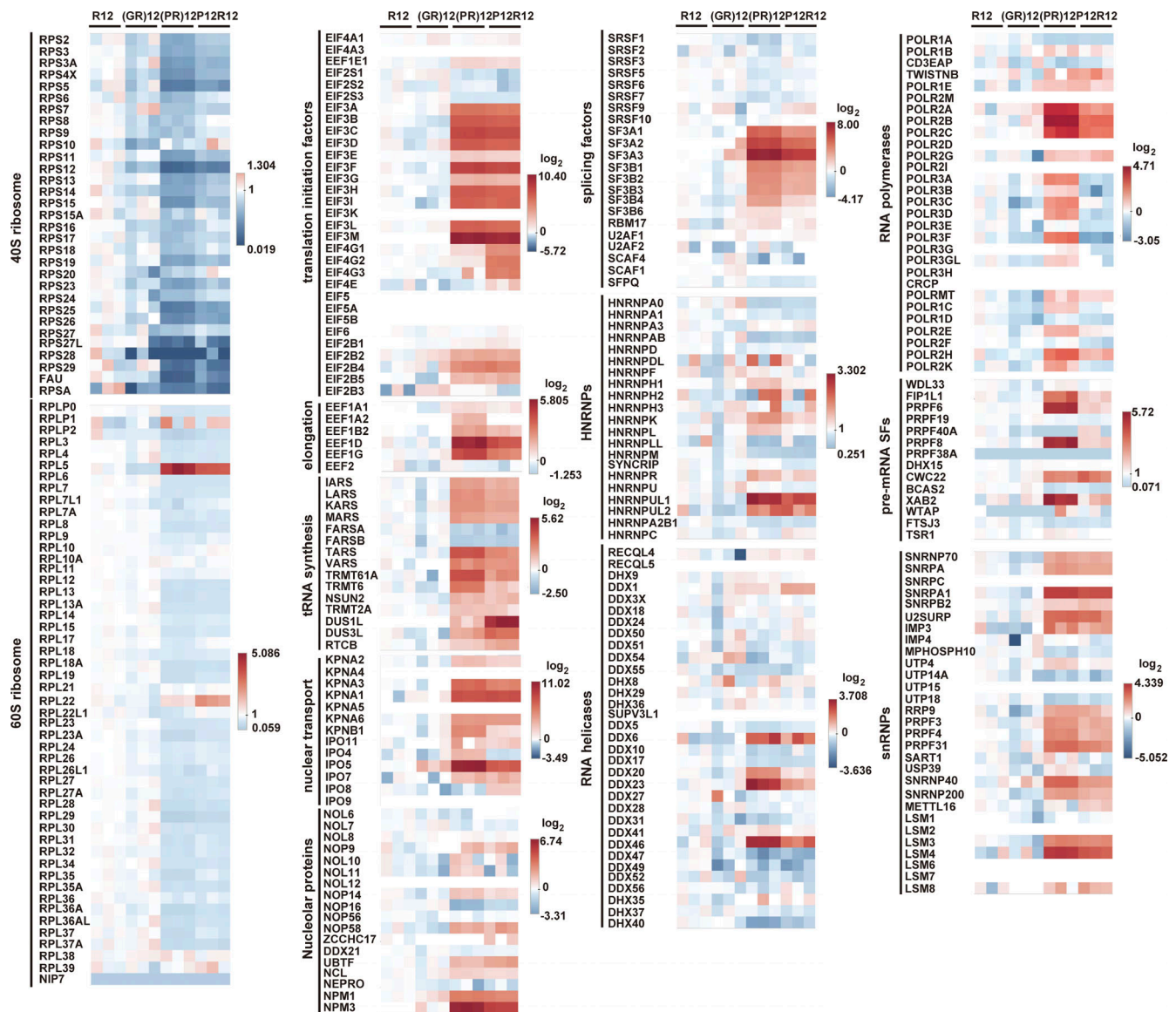
Supplemental material

Figure S1. **Heatmap for the interactome of R$_{12}$, (GR)$_{12}$, (PR)$_{12}$, or P$_{12}$R$_{12}$.**

A

(PR)12 interactome (HEK293)

(PR)50 interactome (HEK293)
(SAINT > 0.9)
total 147

p-value: 1.05e-126

12 135 (92%) 1662

D

(P2R2)6   (PR)12

122 1737 60

B

(PR)12 interactome (HEK293)

(PR)50 interactome(HEK293)
(SAINT > 0.6)
total 236

p-value: 5.05e-189

27 209 (89%) 1588

E

(P4R4)3   (PR)12

130 1726 71

C

(PR)12 interactome (HEK293)

(PR)100 interactome(U2OS)
total 178

p-value: 8.57e-123

31 147 (83%) 1650

F

P12R12   (PR)12

91 1705 92
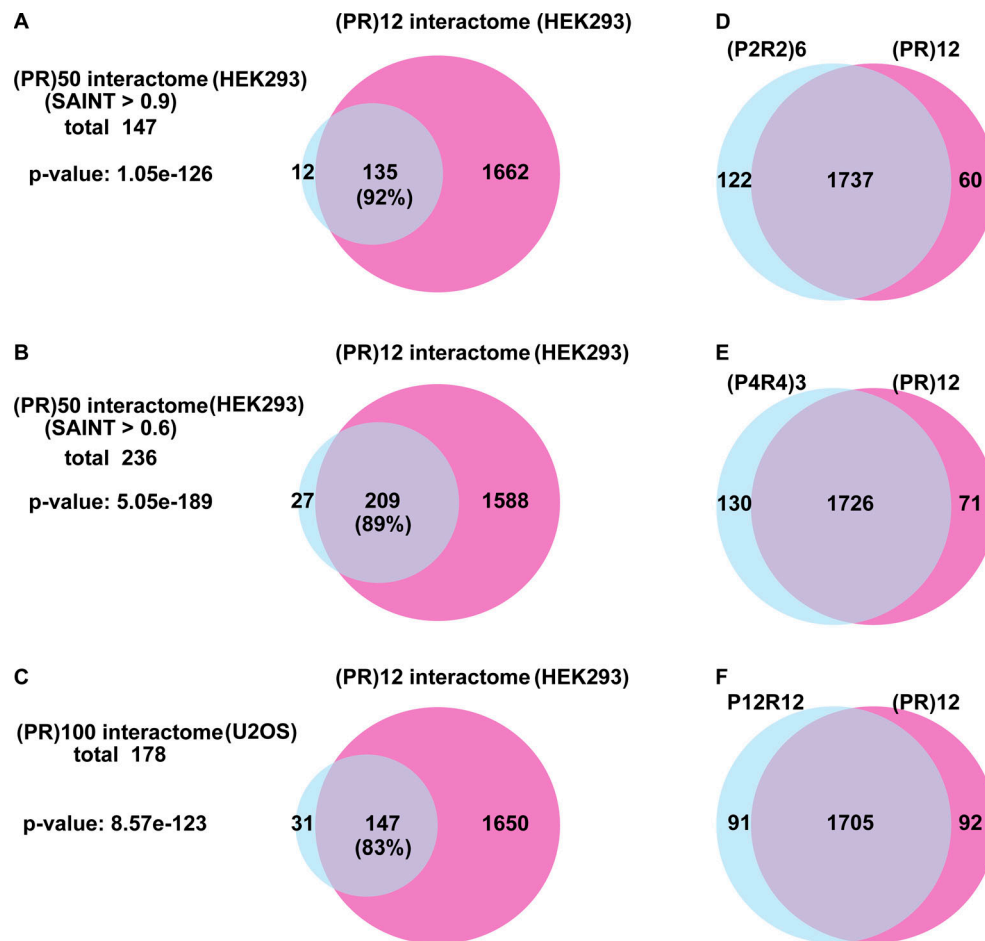
Figure S2. **Comparison between the interactome of (PR)$_{12}$ and interactome of (PR)$_{50}$. (A)** Venn diagram of (PR)$_{50}$ interactome, published before (Lee et al., 2016), and (PR)$_{12}$ interactome. The cutoff for the (PR)$_{50}$ interactome was set as Significance Analysis of Interactome (SAINT) score >0.9, which was the most stringent in the article. **(B)** Venn diagram of (PR)$_{50}$ interactome with cutoff SAINT score >0.6 and (PR)$_{12}$ interactome (Lee et al., 2016). **(C)** Venn diagram of (PR)$_{100}$ interactome (Lin et al., 2016) and (PR)$_{12}$ interactome. Hypergeometrical P values were calculated with the web server https://systems.crump.ucla.edu/ hypergeometric/index.php. **(D)** Venn diagram of (P$_2$R$_2$)$_6$ interactome and (PR)$_{12}$ interactome. **(E)** Venn diagram of (P$_4$R$_4$)$_3$ interactome and (PR)$_{12}$ interactome. **(F)** Venn diagram of P$_{12}$R$_{12}$ interactome and (PR)$_{12}$ interactome.
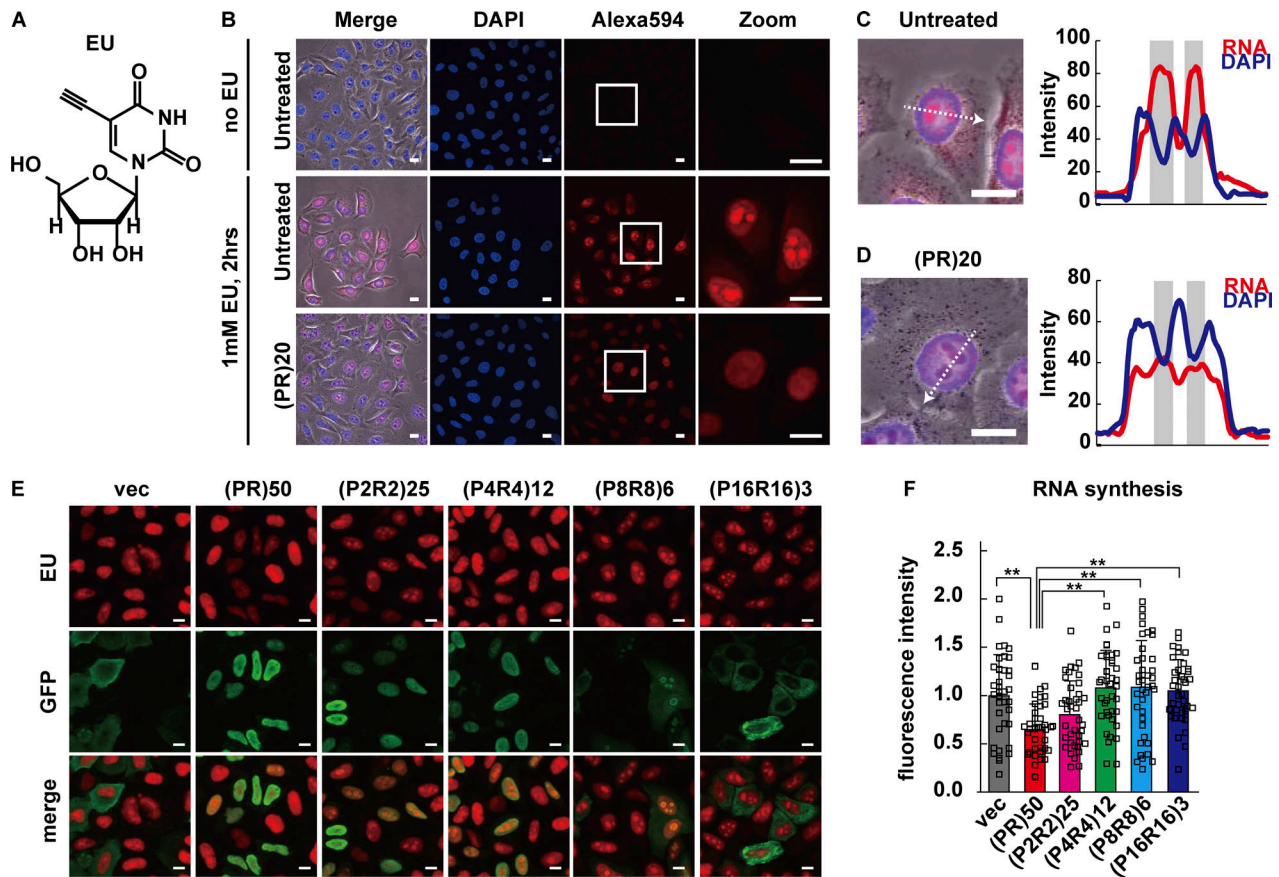
Figure S3. **Transcription of RNA is potentially targeted by poly(PR) toxicity. (A)** Structure of EU. **(B)** HeLa cells treated with or without 10 µM of $(PR)_{20}$ for 2 h in the presence or absence of 1 mM EU. Incorporated EU was treated with Alexa Fluor 594 azide via click chemistry. **(C and D)** Line profiles of nascent RNA signals in Hela cells treated with or without $(PR)_{20}$. Gray area shows nucleolus. **(E)** Transcription in HeLa cells transiently transfected with GFP-$(PR)_{50}$ variants. vec, vector. **(F)** Signal intensities of Alexa Fluor 594 were quantified. The scale bars show 10 µm. All data are represented as mean ± SD. Asterisks indicate significant differences derived via one-way ANOVA with Tukey's test: **, P < 0.01.

A

| Rank	Gene ID	Uniplot Entry	Enrichment	pI	Tyr %	Phe%	Trp%	Glu %	Asp%	aroma F+W+Y	acidic D+E	Longest D/E stretch
1	NAP1L4	Q99733	11596.36	4.60	2.1	6.4	0.8	13.6	8.6	9.3	22.2	5
2	CDK11B	P21127	4215.21	5.42	2.1	2.4	1.0	16.9	5.5	5.5	22.4	13
3	VCP	P55072	2850.76	5.14	1.6	3.9	0.4	8.4	7.7	5.9	16.1	5
4	PAF1	Q8N7H5	2649.20	4.70	3.2	3.2	0.6	13.7	9.4	7.0	23.1	7
5	IPO7	O95373	1813.91	4.39	3.8	3.8	1.3	8.9	6.8	8.9	15.7	5
6	ANP32A	P39687	1429.92	3.98	2.4	1.6	0.0	22.9	12.9	4.0	35.8	11
7	MTA1	Q13330	1294.66	9.34	4.2	2.1	1.4	6.7	4.3	7.7	11.0	3
8	EIF3M	Q7L2H7	1257.79	5.41	2.7	3.5	1.1	7.5	7.2	7.3	14.7	3
9	LEO1	Q8WVC0	587.77	4.37	1.4	1.7	0.5	14.7	14.4	3.6	29.1	9
10	SET	Q01105	559.71	4.22	2.4	4.2	1.4	18.6	11.0	8.0	29.6	11
11	AHCYL1	O43865	469.49	6.48	3.4	2.8	1.5	6.2	6.0	7.7	12.2	3
12	GIT1	Q9Y2X7	358.86	6.33	2.1	2.5	0.5	7.0	6.4	5.1	13.4	3
13	SNX9	Q9Y5X1	330.12	5.40	3.5	4.0	2.2	7.9	6.1	9.7	14.0	4
14	SUPT6H	Q7KZ85	309.66	4.81	3.8	3.9	1.1	11.4	8.2	8.8	19.6	12
15	SSRP1	Q08945	304.51	6.46	3.1	4.9	0.8	10.4	6.9	8.8	17.3	5
16	EIF3F	O00303	274.76	5.24	3.1	2.0	0.3	4.5	4.5	5.4	9.0	3
17	EIF3C	Q99613	242.46	5.48	3.4	2.6	0.8	11.2	6.9	6.8	18.1	7
18	SF3A3	Q12874	237.00	5.27	4.8	3.4	1.6	13.2	5.6	9.8	18.8	6
19	COPG1	Q9Y678	231.83	5.32	2.6	4.0	0.2	8.9	4.8	6.8	13.7	4
20	KPNA1	P52294	217.81	4.94	1.9	3.7	1.3	8.7	4.6	6.9	13.3	4
	Proteome	——	——	6.81	2.66	3.65	1.22	7.09	4.73	7.53	11.82	——

B
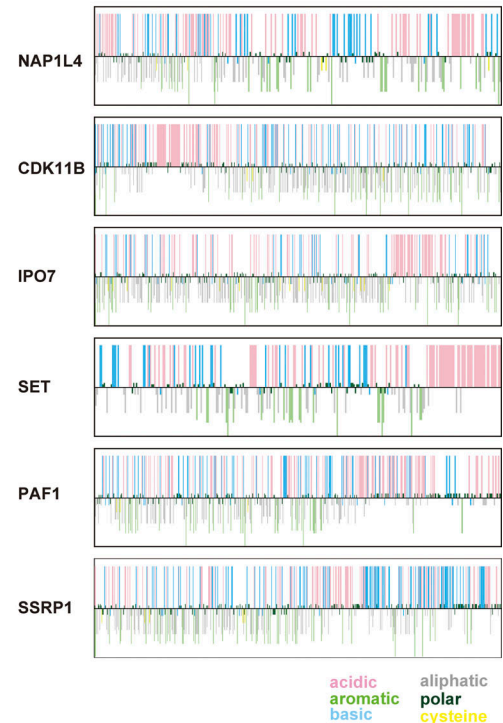
Figure S4. **Characterization of the 20 most enriched (PR)$_{12}$ interacting proteins. (A)** A list of the 20 most enriched (PR)$_{12}$ interactome. Occurrence of acidic or aromatic amino acids; isoelectric point (pI) and longest D/E stretch are indicated. **(B)** Visualization of amino acid compositions of NAP1L4, CDK11B, IPO7, SET, PAF1, and SSRP1.
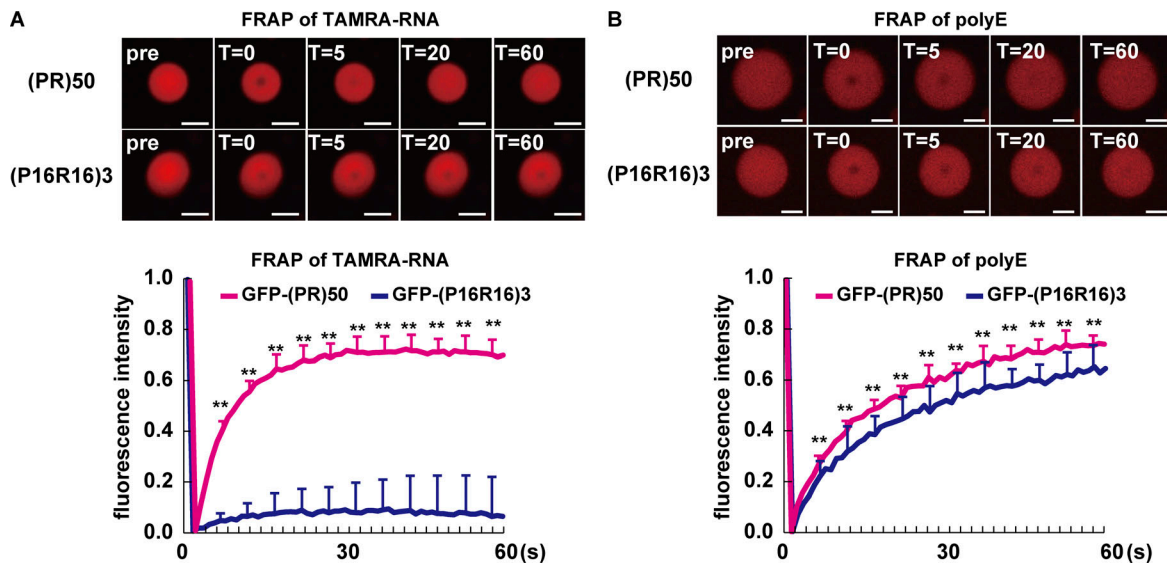

Figure S5. **FRAP analyses of recombinant (PR)$_{50}$ and (P$_{16}$R$_{16}$)$_3$. (A)** FRAP analysis of poly-rA mixed with recombinant GFP-(PR)$_{50}$ or GFP-(P$_{16}$R$_{16}$)$_3$. For observation of LLPS of (PR)$_{50}$ variants and poly-rA, recombinant protein (final concentration, 10 µM) and poly-rA (final concentration, 0.5 mg/ml) containing 100 nM of TAMRA-rA$_{15}$ were mixed. n = 16/variant. **(B)** FRAP analysis of poly-E mixed with recombinant GFP-(PR)$_{50}$ or GFP-(P$_{16}$R$_{16}$)$_3$. For observation of LLPS of (PR)$_{50}$ variants and poly-E, each peptide (final concentration, 10 µM) and poly-E (final concentration, 1 mg/ml) containing 1 µg/ml of poly-E labeled with HiLyte555 were mixed. n = 16/variant. All data are represented as mean ± SD. Asterisks indicate significant differences derived via two-tailed Student's t test. **, $P < 0.01$. The scale bars show 2 µm.

Provided online are two tables. Table S1 contains quantitative proteomics data. Table S2 contains GO enrichment data.