PNAS INAUGURAL ARTICLE | BIOCHEMISTRY OPEN ACCESS

Mammalian telomeric RNA (TERRA) can be translated to produce valine–arginine and glycine–leucine dipeptide repeat proteins

Taghreed M. Al-Turki[a,b] and Jack D. Griffith[a,b,1]

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2018.
Contributed by Jack D. Griffith; received December 19, 2022; accepted January 25, 2023; reviewed by Jan Karlseder and Patricia L. Opresko

Mammalian telomeres consist of (TTAGGG)n repeats. Transcription of the C-rich strand generates a G-rich RNA, termed TERRA, containing G-quadruplex structures. Recent discoveries in several human nucleotide expansion diseases revealed that RNA transcripts containing long runs of 3 or 6 nt repeats which can form strong secondary structures can be translated in multiple frames to generate homopeptide or dipeptide repeat proteins, and multiple studies have shown them to be toxic in cells. We noted that the translation of TERRA would generate two dipeptide repeat proteins: highly charged repeating valine–arginine $(VR)_n$ and hydrophobic repeating glycine–leucine $(GL)_n$. Here, we synthesized these two dipeptide proteins and raised polyclonal antibodies to VR. The VR dipeptide repeat protein binds nucleic acids and localizes strongly to replication forks in DNA. Both VR and GL form long 8-nm filaments with amyloid properties. Using labeled antibodies to VR and laser scanning confocal microscopy, threefold to fourfold more VR was observed in the nuclei of cell lines containing elevated TERRA as contrasted to a primary fibroblast line. Induction of telomere dysfunction via knockdown of TRF2 led to higher amounts of VR, and alteration of TERRA levels using a locked nucleic acid (LNA) GapmeR led to large nuclear VR aggregates. These observations suggest that telomeres, in particular in cells undergoing telomere dysfunction, may express two dipeptide repeat proteins with potentially strong biological properties.

telomere | dipeptides | TERRA | amyloids

Significance

Over the past decade, studies have shown that eukaryotic telomeres are transcribed producing a G-rich RNA termed TERRA. It has been assumed that this RNA is purely structural and does not encode proteins. However, the repeat associated non-ATG translation mechanism (RAN) elucidated in several human diseases could, if utilized by mammalian TERRA, generate two dipeptide repeat proteins, repeating valine–arginine and repeating glycine–leucine. We present evidence for this hypothesis and suggest that these repeating dipeptide proteins are generated in greater amounts when cells undergo a telomere crisis. Their abundance could alter nucleic acid metabolism and general protein synthesis and trigger cellular inflammation responses.

Eight decades ago, work in *Drosophila* by Muller and in maize by McClintock revealed elements at chromosome ends which, in some unknown manner, protect them from fusing end-to-end. These elements were termed telomeres. In subsequent work, telomeres from most eukaryotes were found to consist of long arrays of short nucleotide repeats such as (TTAGGG)n in many mammals and other organisms (1, 2). Telomere-specific proteins were then discovered in numerous species including TRF1 and TRF2 in mammals (3). However, it was not until 6 decades later that the way in which telomeres block end-to-end fusions was revealed through the discovery of end-loop formation (t-loops) (4, 5). Throughout most of this period, it was assumed that telomeres are transcriptionally silent. However, in 2007, Azzalin et al. (6) and subsequently Schoeftner and Blasco (7) showed that not only are mammalian telomeres transcribed from subtelomeric promoters by RNA polymerase II, but the most abundant transcripts arise from the G-rich strand producing a long G-rich RNA (UUAGGG)n termed TERRA. Later, it was shown that TERRA is a ubiquitous feature of telomeres across the phyla (8).

A multitude of studies has shown that TERRA plays a strong structural role in the telomere. Mostly, it is localized to the nucleus, bound at telomeres (7) but upon serum starvation and diauxic shift, TERRA is detected in the cytoplasm (9, 10). While most TERRA molecules are relatively short, some can exceed 9,000 nt (6). Notably, a fraction of TERRA molecules shows polyadenylation at their 3′ ends with a longer half-life compared to the non-polyadenylated species (7, 11). Numerous general and telomere-specific proteins bind TERRA including TRF1, TRF2 (12) ORC, and hnRNPA1 (13). Levels of TERRA have been observed to change through the cell cycle (11), and TERRA levels are elevated in cancer cells and cells that utilize the alternative lengthening of telomeres (ALT) pathway that operates when telomerase activity is absent (14). Feretzaki et al. (15) recently showed that TERRA is expressed from many telomeres. A key structural feature of TERRA is the formation of 24-nt G-quadruplex structures due to the repeated runs of three G's (16). Due to the simple, monotonous nature of mammalian telomeric repeats and lack of canonical start codons, it has been assumed that these sequences do not encode proteins.

In 2011, studies of two human diseases, SCA8 and myotonic dystrophy type 1 (DM-1), unearthed an unexpected phenomenon in which both sense and antisense transcripts containing long runs of expanded triplet repeats which form stable hairpins are translated in all possible reading frames (excluding those that generate stop codons) which was termed

Author affiliations: [a]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295; and [b]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295

Author contributions: T.M.A. and J.D.G. designed research; performed research; analyzed data; and wrote the paper.

Reviewers: J.K., The Salk Institute for Biological Studies; and P.L.O., University of Pittsburgh.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: jdg@med.unc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221529120/-/DCSupplemental.

Published February 22, 2023.

repeat associated non-ATG translation or RAN (17). It was proposed that ribosomes are able to load and begin translation at hairpins or G-quadruplexes. However, due to the lack of an ATG signal, translation can begin at any nucleotide in the repeat, giving rise to three (triplet repeat) or six (hexanucleotide repeat) possible products in each direction (diminished in number if stop codons are encountered). In SCA8, toxic poly amino acid proteins were detected in neural cells (17, 18) and in DM-1, and polyglutamine was observed in muscle cells (17). In 2011, Renton et al. and DeJesus-Hernandez et al. (19, 20) linked familial amyotrophic lateral sclerosis and the related frontotemporal dementia (ALS/FTD) to an expansion of a C_2G_4 nucleotide block at the orf72 locus on chromosome 9 (C9orf72). RNA transcripts containing these repeats form hairpins and G-quadruplexes (21), and RAN translation (22–24) generates a series of dipeptide repeat proteins: repeating arginine–glycine (GR), arginine–proline (PR), glycine–alanine (GA), glycine–proline (GP), and proline–alanine (PA). They fall into two groups: GA, GP, and PA are hydrophobic, while PR and GR are highly charged with 50% arginine residues and would be expected to bind nucleic acids. Indeed, repeating GR and PR are toxic in cultured cells and several animal models (25, 26). Synthetic $(GR)_{20}$ and $(PR)_{20}$ taken up by cells in culture disrupt nucleoli and mRNA processing and poison RNA biogenesis (27, 28). These toxic dipeptide proteins may contribute to general genomic instability as the accumulation of double-strand DNA breaks has been noted in neurodegenerative diseases (29). Farg et al. (30) observed an upregulation of markers of the DNA damage response in motor neurons from patients with C9orf72 ALS/FTD, suggesting that this was due to the presence of the dipeptide repeat proteins, and Andrade et al. (31) found that expression of PR, GP, and GA in cells decreased the efficiency of several DNA repair pathways with PR having the strongest effect.

Of the three hydrophobic dipeptide repeat proteins expressed in C9orf72 ALS/FTD, GA has shown some toxicity in vivo (25, 26). The greatest number of studies have focused on GA. Zhang et al. (32, 33) expressed $(GA)_{50}$ in cultured cells and observed amyloid-like inclusions and inhibition of proteosomes and noted an accumulation of GA amyloids in patients with C9orf72 ALS/FTD (34). Thus, a hallmark of GA is the formation of amyloid-like structures and possibly neural toxicity.

Both mammalian TERRA and the expanded RNA in C9orf72 ALS/FTD contain a six nucleotide repeat with three and four G's, respectively, that generate G quadruplexes providing potential ribosome loading sites and this parallel did not escape us.

Inspection of the sequence of mammalian TERRA revealed that two dipeptide repeat proteins could be produced by a RAN mechanism (Fig. 1). One, repeating valine–arginine (VR), is highly charged and similar to GR and PR, would be expected to bind nucleic acids. The other, repeating glycine-leucine (GL), is hydrophobic and like GA, might form amyloid structures. While C-rich telomeric transcripts have been detected (6), they are present in very low abundance. Nevertheless, RAN translation of the C-rich transcript could generate repeating proline–asparagine and repeating threonine–leucine dipeptides by a RAN mechanism. Of note, the telomeric repeat of *Caenorhabditis elegans* is (TTAGGC)n (35). RAN translation of G-rich *C. elegans* TERRA would also produce two dipeptide repeat proteins: GL as in mammals and instead of VR, repeating leucine-arginine (LR). Telomeric RNA (UUUAGGG)n from plants generates stop codons in all frames.

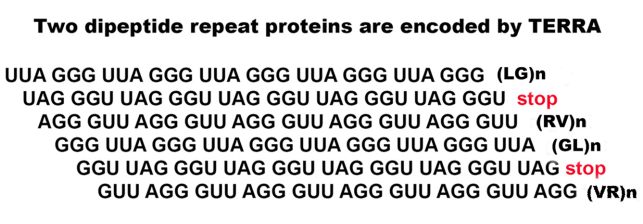

Fig. 1. Two dipeptide repeat proteins encoded by TERRA. Initiation of translation by ribosomes beginning at any of the nucleotides in the UUAGGG repeats present in TERRA will generate VR and GL dipeptide repeat proteins.

TERRA can be found in the cytoplasm (10) and is also seen as cargo in extracellular vesicles (9, 36). Very recent work from the Shadel and Karlseder groups (Nassour et al., 2023) (37) revealed that when telomeres are rendered dysfunctional, for example, through loss of TRF2, TERRA appears in the cytoplasm in larger amounts and interacts with key factors in the innate immune response pathways to activate autophagy.

In this study, initiated in 2011, we synthesized ~20 amino acid-repeating GL and VR dipeptide proteins and have examined their properties in vitro. More recently, we were successful in generating polyclonal antibodies to the repeating VR dipeptide protein and present evidence here for the presence of VR inclusions in cells, in particular ones expressing high levels of TERRA. These inclusions appeared in higher levels and in large inclusions when telomere homeostasis is disrupted. Expression of telomeric dipeptide repeat proteins when TERRA is translocated to the cytoplasm could impact cellular pathways involving nucleic acid metabolism and repair and general protein synthesis and induce genomic instability and inflammation.

Results and Discussion

GL, GA, and VR Dipeptide Repeat Proteins Form Long Filaments and Amyloids-Like Networks. A peptide consisting of nine GL repeats $(GL)_9$ and another containing seven GA repeats $(GA)_7$ were chemically synthesized (*SI Appendix, Materials and Methods* and Table S1). Both were incubated in a low salt buffer and prepared for visualization by transmission electron microscopy (TEM) using metal shadow casting (Fig. 2 *A*, *C*, and *E*), cryoEM (Fig. 2*B*), and negative staining (Fig. 2 *D* and *F*) (*SI Appendix, Materials and Methods*). Examination of fields of molecules formed by $(GL)_9$ (Fig. 2*A*) revealed long, stiff filaments, often measuring several microns in length, frequently associating into large networks. $(GA)_7$ formed nearly identical filaments and also large, dense aggregates (Fig. 2 *C* and *D*). $(GL)_9$ mean filament diameters of 7.8 ± 1.1 nm (n = 80) were obtained from negative staining and 7.9 ± 1.3 nm (n = 18) from cryoEM images. $(GA)_7$ filament diameters (negative staining) had a mean value of 7.3 ± 0.95 nm (n = 35). Efforts to obtain high-resolution structures by cryoEM are underway. Using metal shadow casting and TEM to monitor filament formation, it was observed that the GL filaments formed very rapidly in solution at 2 mg/mL of peptide. Once formed, they remained partially dispersed with some networks forming over days of incubation. GA filaments also formed quickly in solution at high peptide concentration but increased in filament length over periods of hours, and upon extended incubation, most of the filaments were in dense amyloid-like networks as shown in Fig. 2 *C* and *D*. Both the GL and GA filaments were stable to dilution to a few micrograms/milliliter in low salt buffers, but freezing led to insoluble aggregates.

Several preparations of repeating VR dipeptides were synthesized, one of which is 20 amino acids long with biotin at the N terminus ($(VR)_{10}$-bio) and another consisting of 15 VR repeats $(VR)_{15}$ (*SI Appendix*, Table S1). When these were taken up at 2 mg/mL in low salt buffer, poorly structured aggregates and short rods were observed by TEM, but they dissociated upon a 100-fold dilution. However, when the VR preparations were taken up in phosphate-buffered saline (PBS) at 2 mg/mL, long filaments and

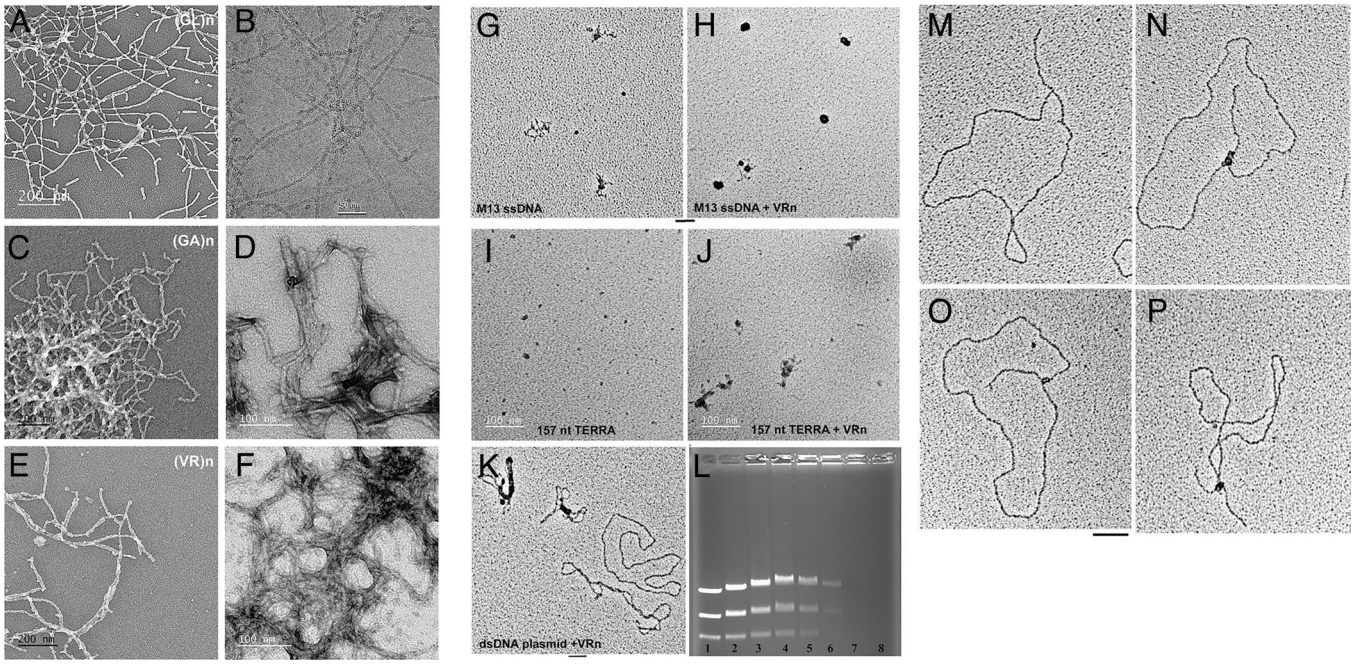

Fig. 2. Electron microscopic visualization of dipeptide filaments, and the binding of VR to nucleic acids. An 18 a.a. peptide containing 9 GL repeats $(GL)_9$ and a 14 a.a. peptide containing seven GA repeats $(GA)_7$ (*SI Appendix*, Table S1) were prepared for EM by rotary metal shadow casting (*SI Appendix, Materials and Methods*) (*A*, *C*, and *E*). (*B*) Field of $(GL)_9$ fibers frozen in vitreous ice and imaged by cryoEM (*SI Appendix, Materials and Methods*). (*D*) Negative staining of GA filaments. (*E* and *F*). VR_{10}-bio dipeptide filaments were prepared for EM by rotary metal shadow casting (*E*) or negative staining (*F*). TEM imaging was at 40 kV (*A*, *C*, and *E*), 80 kV (*D* and *F*) and 200 kV (*B*) (*SI Appendix, Materials and Methods*). *A*, *C*, and *E* shown in reverse contrast. Magnification bars (*A*–*F*) are shown for each field. (*G*) M13 ssDNA visualized by TEM in a buffer of 10 mM Hepes (pH 7.5), 50 mM NaCl. (*H*) M13 ssDNA in the same buffer incubated with a $(VR)_{10}$-bio at a 1:1 mass ratio. (*I*) Field of 157-nt TERRA molecules (small dots) visualized by TEM. (*J*) 157-nt TERRA molecules incubated with $(VR)_{10}$-bio at a 1:1 mass ratio. (*K*) A 3-kb pRST5 plasmid DNA consisting of a mixture of open circular and supertwisted forms was mixed with $(VR)_{10}$-bio at a 1:1 mass ratio and visualized by TEM. Samples shown in *G*–*K* were prepared for TEM as in *A*, *C*, and *E* at DNA or RNA concentrations of 1 μg/mL and incubations were carried out for 20 min at room temperature. Magnification bars in *G*, *H*, and *K* equal 50 nM. (*L*) 500 ng aliquots of a mixture of three pRST5 DNA fragments (1,937, 1,018, and 558 bp) were incubated with 0, 125, 250, 375, 500, 750, and 1,000 ng of $(VR)_{10}$-bio (lanes 1 to 8, respectively) and electrophoresed on an agarose gel. A 3-kb pRST5 plasmid (pGLGAP) (38) containing a 400-bp displaced arm and a 5-nt gap at the base of the fork was incubated with $(VR)_{10}$-bio and then further incubated with streptavidin as a tag for the presence of $(VR)_{10}$-bio for 20 min. This was followed by preparation for TEM (as in Fig. 2) (*SI Appendix, Materials and Methods*) (*M*–*P*). DNAs were scored sequentially as they were encountered in fields observed in the TEM. (*M*) Replication fork DNA alone. (*N*–*P*) Replication fork DNA incubated with $(VR)_{10}$-bio and streptavidin. The molecule in *P* shows a fork which had undergone slippage to generate a four-armed "chicken foot" structure. Bar equals 50 nm for *M*–*P*.

filament networks were seen both by metal shadow casting and negative staining (Fig. 2 *E* and *F*) and these were stable upon 1,000-fold dilution in PBS. The mean diameter of the VR filaments determined from negative staining was 7.5 ± 1.0 nm (n = 50). Thus, masking arginine residues in the VR dipeptide with phosphate appears to allow them to form structures similar to those formed by GL and GA dipeptides.

Flores et al. (38) showed that $(GA)_6$ will form amyloid-like structures with beta sheet properties, and the kinetics of filament formation by $(GA)_{15}$ in vitro was examined by Chang et al. (39) who described the formation of flat ribbons which slowly formed larger aggregates and exhibited amyloid properties. EM imaging in cells in which $(GA)_{175}$ was expressed from a plasmid revealed twisted filament ribbons 13 to 15 nm in diameter and 0.1 to 1 μm in length, which formed amyloid-like clusters (40). These 13- to 15-nm twisted filaments likely form from the association of several 7.3-nm primary filaments.

The telomeric GL dipeptide repeat protein thus appears to have properties similar to GA including filamentation and amyloid formation. The GA and GL filaments may share similarities with structures formed by human lysozyme and islet amyloid polypeptide, both of which generate filaments and amyloids, are stabilized by cross beta-sheet formation, and induce inflammation via NLRP3 and inflammasome formation (41, 42). Taken together, these observations point to the likelihood that GL and GA dipeptide protein filaments could also induce inflammatory responses in cells. The observation of filaments and amyloid-like networks formed by VR dipeptides in the presence of PBS was surprising. VR dipeptide proteins might disrupt nucleic acid pathways as do PR and GR (27, 28) but also activate inflammatory pathways typically seen in aging cells.

VR Dipeptide Repeat Protein Avidly Binds ssDNA and RNA. The highly charged nature of poly VR points to its binding nucleic acids. To examine the binding of $(VR)_{10}$-bio to nucleic acids, it was diluted in low salt buffer to disperse aggregates, and aliquots from incubations with RNA or DNA were prepared for TEM by rotary metal shadow casting as in Fig. 2. Examination of M13 ssDNA revealed previously described bush-like structures (Fig. 2*G*) and incubation with $(VR)_{10}$-bio at ratios of 0.1 to 1.0 μg of $(VR)_{10}$-bio per microgram of ssDNA resulted in progressive compaction with highly condensed structures present at a 1:1 mass ratio (Fig. 2*H*). The $(VR)_{10}$-bio molecules alone were too small to be seen by shadow casting. TERRA molecules (156 nt) were barely visible as small dots (Fig. 2*I*), but upon incubation with $(VR)_{10}$-bio at a 1:1 mass ratio, clearly visible dense aggregates were observed (Fig. 2*J*). As seen by EM, incubation of a mixture of supercoiled and relaxed plasmid species at a 1:1 mass ratio of $(VR)_{10}$-bio to DNA (Fig. 2*K*) resulted in a fraction of the supertwisted molecules appearing collapsed, but others and the relaxed species appeared unchanged. To explore this further, a plasmid digest consisting of three DNA fragments (1,937, 1,018, and 558 bp) was incubated with increasing amounts of $(VR)_{10}$-bio and electrophoresed on a 0.7% agarose gel (Fig. 2*L*). As the ratio of $(VR)_{10}$-bio to DNA increased (lanes 2 to 8), the fragments showed a shift to slower migrating species and the appearance of increasing amounts of

DNA at the top of the gel such that at the highest ratio of peptide all of the DNA was present at the top.

These observations demonstrate that the VR dipeptide repeat protein has a strong affinity for nucleic acids with preference for ssDNA and RNA over duplex DNA. This may suggest that the greater flexibility of ssDNA and RNA over dsDNA facilitates binding of the arginine and phosphate residues. Further studies with RNAs of different sizes and base composition will be required to determine if the VR peptide has any higher preference for TERRA over other RNAs.

VR Dipeptide Repeat Protein Localizes to DNA Replication Forks and Holliday Junctions. The high affinity of $(VR)_{10}$-bio for ssDNA suggested that it might localize to ssDNA gaps or unpaired structures in duplex DNA. To test this, we used a circular 3-kb plasmid containing a single replication fork with a 400-bp displaced duplex tail at a defined site (43). The DNAs contain gaps of either 1, 5, or 15 nt at the base of the fork on the displaced arm. The DNA was sequentially incubated with $(VR)_{10}$-bio and then streptavidin to identify its location and prepared for TEM (*SI Appendix*, *Materials and Methods*). DNAs with a 5-nt gap at the fork are shown in Fig. 2 *M*–*P*. The results described below with distributions are summarized in *SI Appendix*, Table S2.

In the absence of $(VR)_{10}$-bio, scoring 100 DNAs with a replication fork, none had a streptavidin particle bound at the fork or elsewhere on the DNA. In the presence of $(VR)_{10}$-bio, streptavidin tagging showed a high preference for $(VR)_{10}$-bio at the fork with similar affinity for DNAs containing 1-, 5-, or 15-nt gaps. For DNA with a 1-nt gap, 82% had a single streptavidin particle bound, and of these, 84% were localized to the fork, while the remaining 16% were either at the end of the displaced arm or along the plasmid circle. With a 5-nt gap, 91% of the streptavidin particles were localized to the fork, and with a 15-nt gap, 85% were at the fork. The ends of the displaced arm may have some ss character resulting in localization at that site. This represents a very high preference for the replication fork junction.

The fork in the molecule shown in Fig. 2*P* had regressed to generate a four-stranded chicken foot structure in which there is no ss gap at the junction. Such structures were previously observed in identical DNA preparations (44). In the current experiments, they were frequently bound by $(VR)_{10}$-bio as seen by streptavidin tagging. To examine this further, a pure Holliday junction DNA (J-12 junction with 200 bp arms) (45) was incubated with $(VR)_{10}$-bio and streptavidin. Of 83 Holliday junction DNAs scored, 46% had a streptavidin particle bound, and of those, 82% were at the junction as contrasted to being along or at the ends of the arms (*SI Appendix*, Table S2). Thus, $(VR)_{10}$-bio not only has a strong affinity for ssDNA or RNA but can target small unpaired regions in DNA such as those present at a Holliday junction.

To ask whether the nucleic acid binding properties of VR dipeptide repeat protein are unique to this sequence or rather due to the presence of the charged arginine residues, a biotin-tagged 18 amino acid poly arginine peptide (R_{18}-bio) was synthesized (*SI Appendix*, Table 1), along with a biotin-tagged 17 amino acid repeating dipeptide in which the arginine residues were replaced with lysine [$(VK)_9$-bio]. These were employed in experiments with the replication fork DNA containing a 15-nt gap at the fork and at the same ratios of the peptides to DNA as above. Results with $(VK)_9$-bio showed only slightly less specificity for the fork than $(VR)_{10}$-bio with 78% of the streptavidin particles localized to the fork as contrasted to elsewhere on the DNA (*SI Appendix*, Table S2). Scoring 112 replication fork DNAs incubated with R_{18}-bio, half (49%) showed one or more streptavidin particles bound along the circle or on the displaced arm and the other half (51%) at the junction (18% of the DNAs were not tagged by streptavidin) (*SI Appendix*, Table S2). Overall, this peptide showed significantly lower specificity for the fork junction.

These experiments revealed that the family of dipeptide repeat proteins that includes VR, and most likely PR and GR may be "tuned" to bind tightly not only to ssDNA and RNA but also to perturbations in duplex DNA such as ss gaps, replication forks, and Holliday junctions. Pure polyarginine or polylysine peptides may bind so indiscriminately to nucleic acids that any preference for perturbations in duplex DNA may be mostly lost and hence in vivo would bind primarily along the bulk of the nucleic acids. The alternation of charged and neutral amino acids may help in arranging the charged amino acids in their interaction with the phosphate groups along the nucleic acid backbone generating a higher specificity for gaps and junctions in DNA. Clearly, in vivo effects of these dipeptide repeat proteins would become more pronounced as their length increases.

Generation of an Antibody Specific for the VR Dipeptide Repeat Protein. A rabbit polyclonal antibody was raised against the VR dipeptide repeat protein, and its specificity was confirmed by dot blot analysis which showed specific staining with increasing amounts of VR_{15} but no signal against $(GL)_9$ (Fig. 3*A*). Efforts to generate a parallel antibody to the GL dipeptide repeat protein have not yet been successful.

To evaluate the VR antibody specificity in human cells, a DNA construct containing the cytomegalovirus (CMV) promoter and a 3X Flag tag followed by 60 in-frame VR repeats was synthesized (Fig. 3*B*) and transfected into U2OS cells (*SI Appendix*, *Materials and Methods*) which exhibit an ALT phenotype and were shown to have elevated levels of TERRA due to the presence of hypomethylated subtelomeric regions (47, 48). Thus, U2OS cells may exhibit a higher level of endogenous VR dipeptide proteins. Thirty-six hours after transfection, the cells were fixed and costained with VR (red) and Flag (green) antisera and counterstained with DAPI. Laser scanning confocal microscopy revealed distinct Flag and VR foci. Both the Flag epitope (Fig. 3 *C*, *Top*) and the VR staining material (Fig. 3 *C*, *Middle*) were present in punctate foci ranging from small spots to larger bodies. The VR antibody also stained large bodies in the nuclei suggesting that the VR dipeptide protein is prone to aggregation (Fig. 3*C*). This is consistent with previous cell culture-based studies as well as studies of C9orf72 ALS/FTD patients, where similar PR and GR aggregates were observed in cells (24, 27, 49–51).

Colocalization of the signals from the VR antibody and the Flag tag in 50 cells was used to determine whether the foci detected by the VR antibody represent VR dipeptides expressed from the plasmid or off-target proteins in U2OS cells. Yellow signals indicate foci where both VR and Flag antigens are in close proximity and the presence of frequent yellow signals in the merged image (Fig. 3 *C*, *Bottom*) demonstrates positive spatial correlation between images in the two channels. Interestingly, some separate red signals were detected by the VR antibody that did not colocalize with Flag (Fig. 3 *C*, *Bottom*, white arrow) suggesting the detection of endogenous VR dipeptide proteins.

To translate the yellow overlap signals into quantitative information reflecting the correlation of the two targets, we performed a colocalization coefficient analysis (46) (*SI Appendix*, *Materials and Methods*). As shown in Fig. 3*D*, 98% of the red pixels overlapped with the green pixels and 81% of the green pixels overlapped with the red pixels. This high degree of overlap suggests that the VR antibody is specific and binds its target (VR) in this case, fused to Flag.

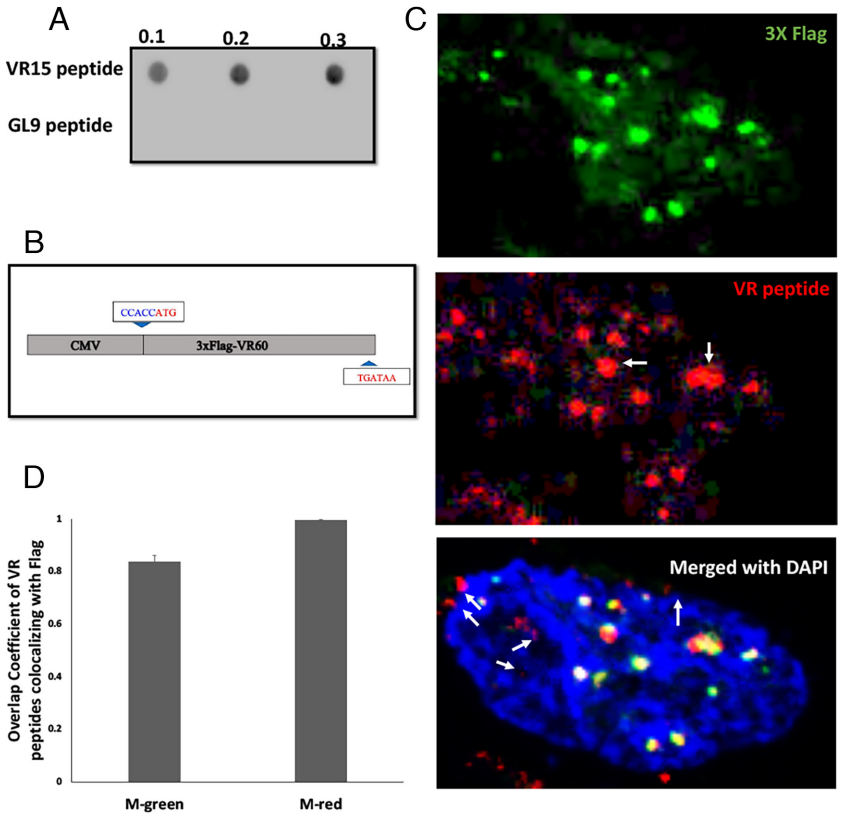

Fig. 3. Validation of VR-specific antibodies by immunoblotting and direct expression of VR in cells. (*A*) A peptide consisting of four VR repeats and a short linker (*SI Appendix*, Table S2, peptide #2) was used to raise polyclonal rabbit antibodies and was affinity-purified on a column containing $(VR)_{15}$ (*SI Appendix*, Table S2, peptide #3) which lacks the linker. Increasing amounts of $(VR)_{15}$, and $(GL)_9$ (negative control) were blotted onto nitrocellulose membranes and incubated with the VR antibody for Western Blot analysis. The experiment was repeated three times. (*B*) Schematic of a DNA construct containing the CMV promoter followed by a 3X Flag tag and 60 repeats of the VR dipeptide and terminated in two stop codons. This construct was inserted into a pcDNA3.1(+) vector (GenScript Inc.) for transfection (*SI Appendix, Materials and Methods*). (*C*) Representative confocal images of U2OS cells overexpressing the 3X Flag-VR_{60} construct. Cells were fixed, coimmunostained with Flag and VR antibodies. The white arrows indicate dot-like aggregates. The merged image (*Bottom*) shows the colocalization of Flag and VR (yellow signals) in the nucleus (DAPI-blue). White arrows indicate the VR signals (red) that do not overlap with Flag (green). (*D*) Graph of the percentage of Flag colocalized with VR. Percentage colocalization was determined by calculating Mander's Colocalization Coefficient (MCC) using ImageJ and the JACoP plugin within the Region of Interest (ROI) for individual cells. Colocalization was measured by applying Mander's Overlap Coefficient (MOC) method (46). The value of MOC can range from 0 to 1, where 0 represents no overlap and 1 represents maximum overlap. Fifty cells were quantified. Error bars indicate standard error. All images are single confocal plane images.

SDS-PAGE and Western Blot Analysis Validate the Specificity of the VR Antibody. The results from laser scanning confocal microscopy pointed to cell-specific staining with the VR antibody (Fig. 3 *C*, arrows). Thus, VR dipeptide proteins should be detected in a Western analysis using the VR antibody. It was also important to determine if any general cellular proteins contribute a background of staining. If so, they should be seen as a specific band or bands upon probing the gels. Arguments for the specificity of the VR antibody which was generated using a $(VR)_4$ peptide and affinity purified using a $(VR)_{15}$ peptide came from a Blast search of the database of human proteins. This failed to uncover any protein with a run of four VR repeats and only one (OS-9 isoform precursor) with three, arguing that this particular amino acid arrangement may have been selected against, possibly for structural or other reasons.

Previous studies of the behavior of the dipeptide repeat proteins GA, GP, and GR (24) on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gels revealed that even though they were of relatively low molecular weight, none entered the gel proper but were detected at the boundary of the stacking gel, and similar results were reported upon transfecting HEK293 cells with constructs expressing GA, GP, and GR peptides, indicating that those dipeptides are also in SDS-insoluble aggregates (23, 52). They also noted that the properties of these proteins appear to partially inhibit or lower their electrotransfer efficiency to PVDF membranes. Thus, they and other groups have concluded that dot blot analysis provides a more accurate and reliable measure of this class of proteins than Western analysis. In parallel with what these authors had reported, when the $(VR)_{15}$ dipeptide was boiled in 10% SDS and electrophoresed on an 8 to 16% polyacrylamide gel, Western analysis using the VR antibody (*SI Appendix*, *Materials and Methods*) revealed that the peptide also remained at the boundary of the stacking gel (Fig. 4*A*) paralleling the results with the ALS/FTD dipeptide proteins.

To probe for any cellular proteins that might be contributing a background to the signals observed in the light microscopy studies (Fig. 3*C*), extracts were prepared from 80% confluent U2OS cells, U2OS cells overexpressing $(VR)_{60}$, and cells from a primary human foreskin fibroblast cell line (FSK) established in this laboratory. Staining with Ponceau-S (Fig. 4*B*) revealed that comparable amounts of protein were present on the membrane for all three samples. Probing the gel with the VR antibody (Fig. 4*C*) revealed distinct bands at the boundary of the gel and stacking gel for the U2OS, U2OS-RV_{60}, cells, and the synthetic VR_{15} dipeptide protein, and a very dim signal was present from the FSK cells. The lack of any bands within the 8 to 16% gel for any of the three cell extracts provided evidence that the VR antibody used in our study is not broadly detecting other cellular species. This analysis also indicated that there are higher levels of VR dipeptide protein aggregates present in the ALT line U2OS as contrasted to the primary foreskin line FSK.

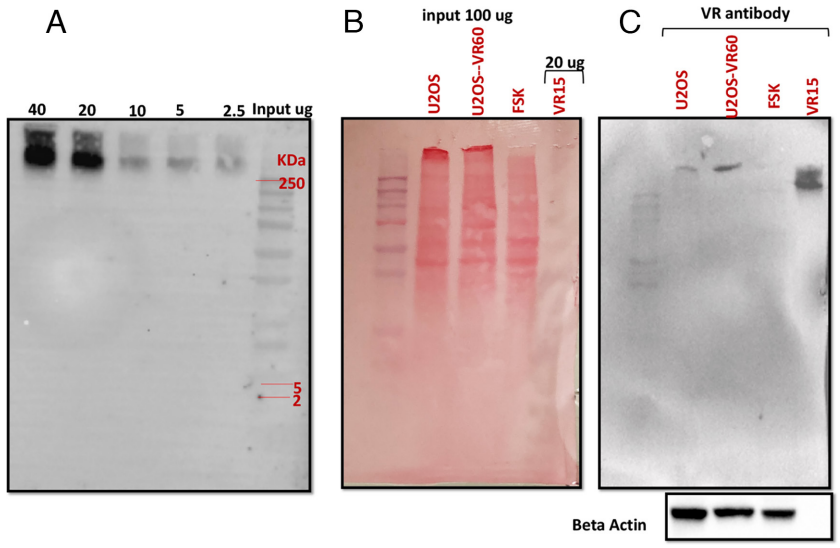

Fig. 4. Western analysis on SDS-PAGE gels reveals specific staining of the VR dipeptide protein. (*A*) Western Blot of increasing amounts of VR_{15} dipeptide protein. Three independent experiments were performed. (*B*) Ponceau-S staining of U2OS, U2OS overexpressing $(VR)_{60}$, and FSK cell lines revealed that total protein from the three samples, including the expected band at the top of the gel, was successfully transferred. Protein molecular weight markers are in the left lane. (*C*). The destained PVDF membrane in *B*; it was subjected to Western Blot analysis. Beta Actin was detected and was used as a loading control. Three independent experiments were performed.

VR Dipeptide Proteins Are Expressed in U2OS, Immunodeficiency, Centromeric Region Instability, Facial Anomalies Syndrome (ICF), and Primary Human Cells. The presence of discrete red foci (VR signals lacking a Flag tag) in the experiments with the U2OS cells (Fig. 3*C*) and the presence of specific bands in the Western analysis showing the SDS-insoluble aggregates of VR encouraged us to examine VR staining foci in several different cell lines. As a baseline, we employed the FSK cells and compared the results to those derived from two cell lines known to have high endogenous levels of TERRA: U2OS, and the telomerase-positive fibroblast line (GM08747) derived from a female patient afflicted with the Immunodeficiency, Centromeric region instability, Facial anomalies syndrome (ICF) in which high levels of TERRA have been reported also due to hypomethylation of subtelomeric regions (48, 53, 54). Using high-resolution confocal microscopy, we explored the presence of VR dipeptide proteins in the three cell lines. Leaving out the primary VR antibody was included as a negative control. A goat anti-rabbit Alexa 488 secondary antibody conjugate was used to detect VR signals in the cells. Initial imaging was carried out to optimize the specificity of staining and quality (contrast) of the images. To do so, each individual cell was identified by detecting DAPI-stained nuclei, followed by laser scanning for VR signals (any signal with pixel values above the background level). This was done for cells selected in each of the three cell lines. The U2OS cells with the best signal-to-noise ratio were used to optimize the acquisition parameters for each fluorescence channel for the three cell lines. Endogenous VR signals were detected and appeared as punctate spots and discrete foci which varied in size and intensity in the three cell lines (Fig. 5 *A*, *Left Top*, *Middle*, and *Bottom*). No VR signals were detected in the negative control (no primary antibody), confirming that the VR primary antibody binding is specific.

A total of 684 cycling cells were randomly imaged in three independent experiments. To avoid any bias, the images were acquired blindly to the VR signals. To analyze the data, we set the threshold to discriminate the positive signals from the negative signals using CellProfiler software (55). We then scored the number of cycling cells with five or more VR foci in the three cell lines. The results (Fig. 5 *A* and *C*) indicate that approximately 37% of U2OS and 33% of the ICF cells contain ≥5 VR foci. By comparison, only 11% of the primary FSK cells had ≥5 VR foci. These results reveal that U2OS and ICF cells exhibit a threefold ($P < 0.001$, $P < 0.01$), greater number of VR foci compared to the FSK cells pointing to the conclusion that cells with hypomethylated subtelomeres and elevated TERRA produce more telomeric VR dipeptide proteins.

It was of interest to examine nuclear versus cytoplasmic localization of the VR staining material. Both arginine rich-dipeptide proteins (GR and PR) from C9orf72 ALS/FTD studies showed enrichment in the nuclear compartment (56, 57). We asked if the VR dipeptide protein would parallel this trend. F-actin and nuclei were counterstained with phalloidin and DAPI, respectively (Fig. 5 *A*, *Middle* and *Right Top*, *Middle*, and *Bottom* and *B*). Scoring VR puncta was restricted to cells with ≥5 VR foci (Fig. 5*C*). Localization analysis (Fig. 5*D*) in the three cell lines demonstrated that 76%, 75%, and 69% of VR staining material showed preferential nuclear localization in U2OS, ICF, and the primary FSK line, respectively ($P < 0.01$, $P < 0.0001$, and $P < 0.0001$). This can be explained by the arginine-rich nature of the VR dipeptide proteins which resemble nuclear localization signals in the enrichment of arginine and lysine amino acids (58).

In summary, high-resolution laser scanning microscopy demonstrates increased levels of VR dipeptide staining in cells and cells with higher TERRA levels. This staining is preferentially localized to the nuclei, and in cells with higher TERRA levels, VR staining was more commonly present in larger aggregates as contrasted to the human primary line. Moreover, the two high TERRA lines contained threefold more staining material, consistent with our hypothesis that this material represents VR dipeptide protein produced by RAN translation from TERRA molecules.

Altering TERRA Levels in U2OS Cells Results in Large Solid Nuclear VR Aggregates. Ideally, one would stably reduce TERRA levels to background and ask if VR dipeptide proteins are depleted. However, achieving efficient suppression of TERRA in cells has been difficult despite multiple approaches. TERRA as a key structural component of the telomere cannot be eliminated without concomitant loss of telomere integrity. This is further confounded by the binding of RNA polymerase II to promoters

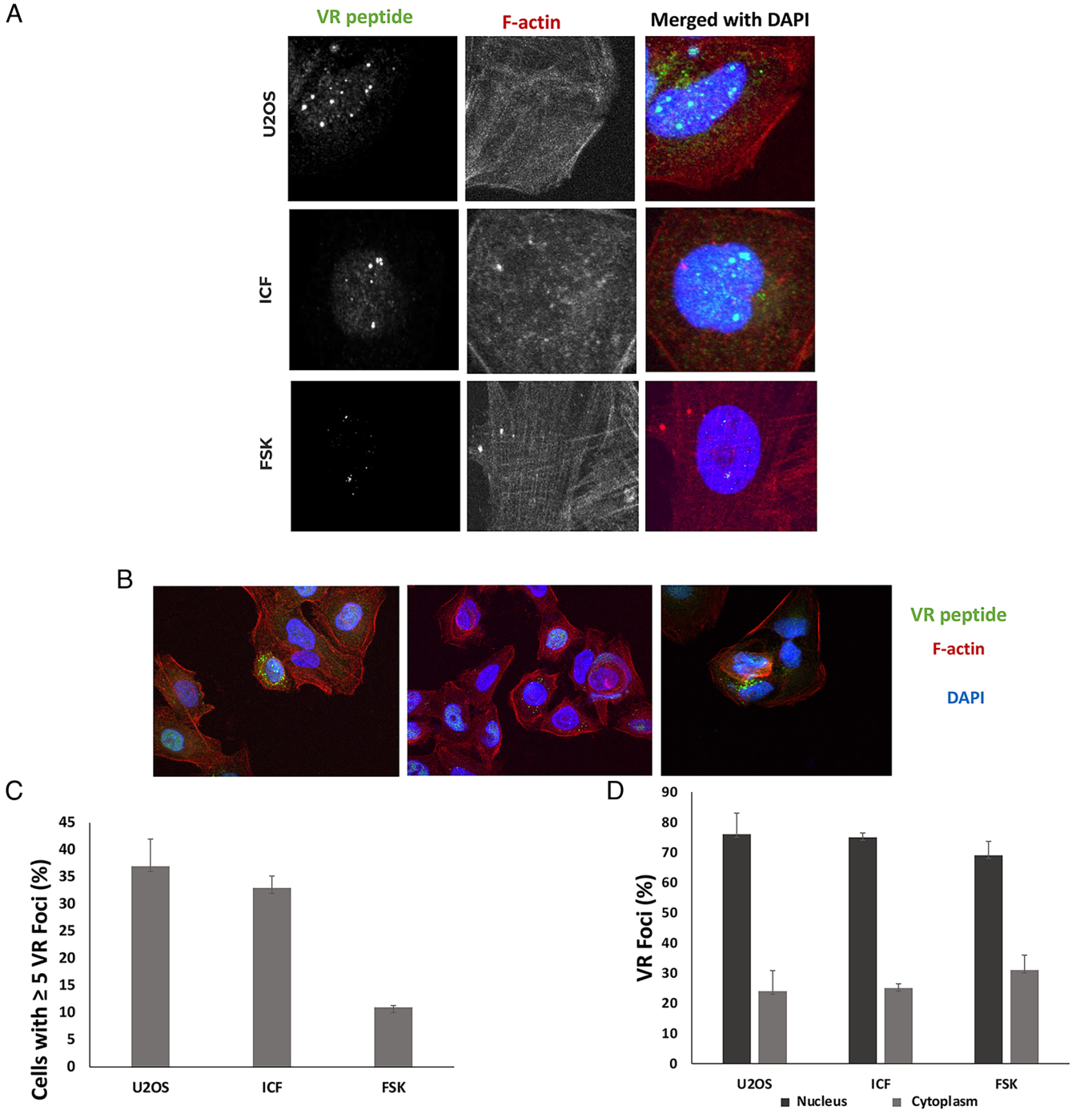

Fig. 5. Identification and characterization of VR peptides in U2OS (ALT), ICF, and primary human cells FSK. (*A*) Representative confocal images (Z-projections) showing nuclei (DAPI-blue), nuclear VR foci heterogeneous in size and intensity (green), and actin rhodamine-phalloidin (red). (*B*) Representative confocal images of nuclear and cytoplasmic VR staining sites. (*C*) Percentage of VR-positive cells (≥5 VR foci) in U2OS, ICF, and FSK. Data presented are the ±SEM of three independent experiments. Unpaired two-tailed *t* test; **P* < 0.05, ***P* < 0.01, ****P* < 0.01, *****P* < 0.0001, ****P* = 0.0003 for U2OS versus FSK, and ***P* = 0.0019 for ICF versus FSK. (*D*) Percentage of VR foci localized in the nucleus or cytoplasm. Unpaired two-tailed *t* test; **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, ***P* = 0.002 for nuclear U2OS versus cytoplasmic U2OS cells, ***P* = 0.006 for nuclear FSK versus cytoplasmic FSK cells.

within the subtelomeric sequences at multiple chromosomes ends where it initiates transcription of TERRA (6, 15, 59–61), and Feretzaki et al. (15) reported their failure to efficiently suppress TERRA levels utilizing Crispr/Cas9 technology due to TERRA being produced from multiple chromosomes.

One approach to transiently reduce TERRA levels employed by several authors involves treating cells with small oligonucleotides termed locked nucleic acid (LNA) GapmeRs. These can be tailored to bind TERRA and induce its degradation by RNaseH. Such treatment has been shown to generate Telomere-dysfunctional foci as seen by light microscopy (12, 62–64).

Studies from Chu et al. (62) and González-Vasconcellos et al. (65) employing an LNA GapmeR approach in U2OS cells reported a brief depletion of TERRA using 8 mM or 100 nM LNA GapmeR, respectively. However, TERRA levels were restored to normal levels within 24 h. Although this limitation prevented us from measuring VR levels in cells stably depleted for TERRA, we were able to ask how transient alteration of TERRA alters VR

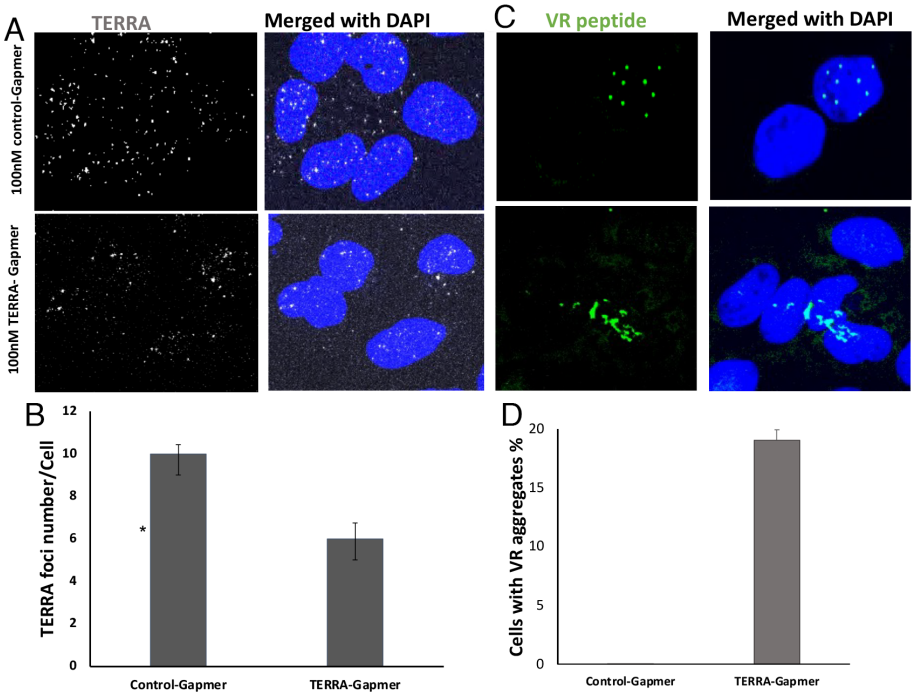

Fig. 6. Alteration of TERRA levels cause aggregation of VR dipeptides. (*A*) Representative confocal images (Z-projections) of 100 nM control and TERRA-LNA GapmeR treated U2OS 24 h after transfection. TERRA molecules were detected using the TelC-Alexa647 probe (*Upper* and *Lower Left*). Nuclei were stained with DAPI, and merged images with TERRA signals reveal nuclear and cytoplasmic TERRA (*Upper* and *Lower Right*). (*B*) Quantification of TERRA signals from *A* in U2OS cells show ∼40% depletion of TERRA. Data presented are the ±SEM of two independent experiments. Two-tailed, unpaired *t* test $*P < 0.05$, $P = 0.04$. (*C*) U2OS cells from *A* were immunolabeled with the rabbit VR primary antibody overnight at 4 °C and labeled with Alexa flour 488 secondary antibody. The large solid VR aggregates (*Left Bottom*) were distinguished from the small spherical VR signals. Nuclei were stained with DAPI and merged images with VR dipeptides show the nuclear signals. (*D*) Quantification of percentage of cells showing an accumulation of large solid nuclear aggregates. No bar is shown for the control-Gapmer as no cells with large solid aggregates were observed. Data are presented as mean ± SEM of two independent experiments.

levels and localization in U2OS cells. Using the LNA GapmeR these authors employed together with another of the same length but with a scrambled nucleic acid sequence (control-LNA GapmeR) (*SI Appendix*, *Materials and Methods*), U2OS cells were treated with 8 mM LNA GapmeRs. However, this resulted in nearly complete death of the cells within the first 24 h. Thus, we reduced treatment to 100 nM for 24 h, fixed the cells, and performed RNA-FISH to detect TERRA foci (Fig. 6*A*). TERRA foci in a total of 643 cells from three independent experiments were scored. To generate unbiased data, images from each condition were acquired randomly and blindly to TERRA signals. The threshold of the positive signals was determined (*SI Appendix*, *Materials and Methods*), and quantification was performed using CellPriofiller software (55). The results revealed a 40% reduction ($P < 0.05$) in TERRA in the LNA GapmeR-treated cells as contrasted to the control (Fig. 6*B*). This is consistent with the observations of González-Vasconcellos et al. (65) who reported a ∼50% depletion under similar treatment conditions in the same cell line.

To determine whether reduction of TERRA levels alters the expression of the VR dipeptides, we stained the LNA GapmeR and control-LNA GapmeR-treated U2OS cells with the VR primary and the Alexa Fluor 488 secondary antibodies followed by imaging by confocal microscopy (Fig. 6*C*). Strikingly, in cells with lowered TERRA and thus dysfunctional telomeres, the VR dipeptide was seen in large solid nuclear aggregates compared to the distinct small spherical signals detected in cells treated with the control-LNA GapmeR (Fig. 6*C*) or untreated U2OS cells (Fig. 5*A*). Scoring 726 U2OS cells in two independent experiments revealed that 19% of the U2OS cells displayed large solid nuclear VR aggregates (Fig. 6*D*) upon partial depletion of TERRA. This phenotype was completely absent in the scored U2OS cells treated with the scrambled GapmeR (Fig. 6*D*).

In summary, transient reduction of TERRA levels by 40% as seen by RNA-FISH resulted in the appearance of large aggregate forms of VR in the nuclei. The formation of these large solid nuclear VR aggregates was in contrast to the distinct small spherical signals detected in cells treated with the control-LNA GapmeR. Their appearance is suggestive of phase transitions in cellular proteins, in particular, droplet liquid-state to reversible amyloid cross-β fibrils, that have been described in cancer and neurodegenerative diseases (66–68). Proteins known to bind TERRA include FUS and the hnRNPs (69–71) which are RNA-binding proteins that contain prion-like domains and glycine–arginine-rich domains (72, 73). It had been shown that the reduction of the levels of noncoding RNA and the presence of arginine-rich dipeptides strengthens the electrostatic interaction between the arginines and FUS leading to the formation of solid dense aggregates (68, 74, 75). Based on these observations, we would suggest that the VR dipeptides play a key role in accelerating an aberrant phase transition via its possible direct electrostatic interaction with FUS and hnRNPs. These aggregates may contain in addition to VR and FUS, the hnRNPs. Future biochemical and or colocalization studies may shed light on their composition.

Lentivirus shRNA Knockdown of TRF2 Leads to Higher Levels of Cytoplasmic VR Dipeptide Protein. Previously Cesare et al. (64) utilized two lentivirus constructs expressing antisense RNAs, shTRF2-1488 and shTRF2-18358, to knock down TRF2 protein. Recently, using these same constructs and approach, Nassour et al. (2023) observed a significant increase in cytoplasmic TERRA upon infecting IMR90^{E6E7} cells with shTRF2-18358, indicating that accumulation of TERRA in the cytoplasm is associated with telomere dysfunction. Thus, it is possible that this would result in an increase in cytoplasmic VR dipeptides in U2OS cells. If so, this

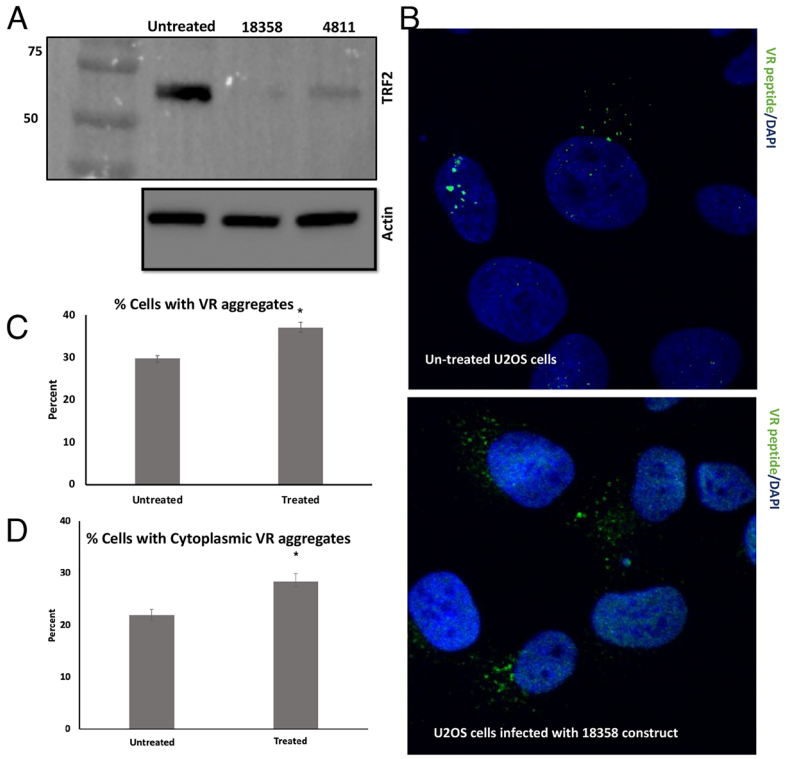

Fig. 7. TRF2 knockdown leads to higher levels of cytoplasmic VR dipeptide. (*A*) Western Blot analysis (*SI Appendix, Materials and Methods*) showing the level of TRF2 expressed in U2OS cells (untreated) or infected with two lentivirus constructs (18358, 4811) encoding TRF2 shRNAs. Actin was used as loading control. (*B*) Untreated U2OS cells and U2OS cells infected with shTRF2-18358 as shown in *A* were immunolabeled with VR primary antibody. Nuclei were stained with DAPI, and merged images with VR dipeptides show the nuclear versus cytoplasmic signals. (*C*) Quantification of the percentage of cells expressing five or more VR aggregates. Data presented are the ±SEM of two independent experiments. Two-tailed, unpaired *t* test **P* < 0.05, *P* = 0.036. (*D*) Quantification of percentage of cells expressing five or more cytoplasmic VR aggregates. Data are presented as mean ± SEM of two independent experiments. Two-tailed, unpaired *t* test **P* < 0.05, *P* = 0.04.

would provide additional evidence linking dysfunctional telomeres to the production of VR dipeptide proteins.

U2OS cells were infected with lentiviruses expressing shTRF2-1488 and shTRF2-18358 RNAs and selected for puromycin resistance (2 mg/mL for 14 d) (*SI Appendix, Materials and Methods*). Cell extracts were prepared for Western analysis using SDS-PAGE gels. As shown (Fig. 7*A*), infection with either lentivirus construct resulted in near complete reduction of TRF2 with shTRF2-18358 being the most potent. A similar pattern of reduction had been observed by Cesare et al in IMR90^{E6E7} cells (65). The pooled U2OS cells infected with shTRF2-18358 were seeded on slides, fixed, and stained with the VR antibody as described above. The percentage of cells with five or more aggregates per cell was determined by randomly scoring 657 cells in two independent experiments. Unbiased scoring was carried out as described above. Analysis revealed a significant increase (22%) in the number of cells expressing five or more VR aggregates per cell $P < 0.05$ (Fig. 7 *B* and *C*) in the population with depleted TRF2. Because TRF2 depletion in IMR90^{E6E7} cells resulted in higher levels of cytoplasmic TERRA, cells were also scored for five or more VR aggregates localized to the cytoplasm as contrasted to the nucleus (Fig. 7 *B* and *D*). This revealed an even greater increase (27%) relative to untreated cells $P < 0.05$.

In summary, based on the finding of Nassour et al. (2023) showing that lentivirus knockdown of TRF2 resulted in an increase of TERRA in the cytoplasm, we have observed that this also results in an increase in cytoplasmic VR aggregates. Whether these VR aggregates colocalize with dysfunctional telomeres will be of interest to probe in future experiments.

Summary and Conclusions

In this paper, we present findings supporting our hypothesis that eukaryotes whose telomeres consist of (TTAGGG)n repeats have evolved this sequence to enable the G-rich RNA transcript TERRA to encode two repeating dipeptide proteins, GL and VR, by a RAN translation mechanism. This feature has been retained in the worm *C. elegans* whose telomeric repeat is (TTAGGC)n where GL and RL would be produced. We show that both GL and VR form filaments and amyloid-like structures and VR has strong nucleic acid binding properties and can be observed, in particular in cells with elevated TERRA levels. These telomeric microproteins are expected to be variable in size. High levels of either likely contribute to amyloid formation. Amyloid deposits could trigger general inflammatory responses through the activation of innate immune response (41, 76, 77). Indeed, it was recently observed that telomere dysfunction instigates inflammation in inflammatory bowel disease (78). Production of larger amounts of VR would in parallel to what has been found for PR and GR in the C9orf72 ALS/FTD studies alter RNA and ribosome biogenesis and nucleic acid metabolism. If incorporated into extracellular vesicles as cargo, they could transmit these signals to nearby cells. If our hypothesis is correct, this could point to a new way of thinking about telomeres, suggesting that they are able to generate both telomere-specific RNA and biologically active microproteins.

Given the small and presumably variable size of the VR and GL species, their resistance to trypsin digestion for mass spectrometry analysis, and their failure to enter standard SDS-PAGE gels, it is not surprising that their presence would have been overlooked. It has

not been possible to determine the length distribution of the VR species as this material remained in aggregates at the top of the SDS-PAGE gels. As discussed above, this observation parallels studies of the repeating dipeptide proteins expressed in the C9orf72 ALS/FTD disease. We have not yet been successful in raising antibodies against repeating GL dipeptides; however, efforts continue.

The dependence of biological activity on length may be different for GL and VR. GL dipeptides with as few as six repeats will assemble into stiff amyloid fibers as does $(GA)_7$. Thus, there may not be a significant size dependence for GL dipeptides in forming amyloid filaments and hence their biological activity. In contrast, repeating VR dipeptides may exist in the cell both as individual species which can interact with nucleic acids and in filaments with properties similar to GL filaments. Here, as the length of the VR species increases, the affinity of the individual species for nucleic acids and hence propensity to collapse DNA and RNA would increase.

TERRA molecules generated in ALT cells (more so than normal cells) have been found to contain frequent variations in runs of perfect UUAGGG repeats (79). As an example, TCA GGG TCA GGG would insert Ser-Gly or Gln-Gly within the VR or GL repeats. Other variations could shift the repeat pattern from VR to GL, for example, or generate a stop codon. The properties of such hybrid species would be complex and vary from one molecule to the next.

The observation of large dense nuclear VR aggregates when VR dipeptide proteins were overexpressed from a plasmid or when TERRA levels were altered by an LNA oligonucleotide suggests that the VR dipeptide protein is able to induce the formation of inclusions possibly involving protein phase transitions. This would parallel observations of aggressive toxicity phenotypes (neurodegeneration) and cell death associated with the arginine-rich dipeptide GR and PR nuclear aggregates observed in C9orf72 ALS/FTD patient cultures, cell culture, and *Drosophila* models (27, 49, 50, 70, 71). In the *C. elegans* model, the toxicity of overexpressed $(PR)_{50}$ in muscle cells and motor neurons requires an aged cellular environment suggesting that the toxicity is aging-dependent. Similarly, Tau fibrillization has been reported as a hallmark of Alzheimer's disease, and aggregates of TLS/FUS have been implicated in causing ALS (67, 68). It will be of interest to examine human cells from different aging or telomere pathologies for aggregates involving VR or GL dipeptide repeat proteins as they could be a marker of dysfunctional or deprotected telomeres. This would include telomeric genetic diseases such as idiopathic pulmonary fibrosis where short telomeres and high levels of TERRA are observed (80).

We speculate that in normal cells with functioning telomeres, the level of TERRA and thus VR and GL dipeptide proteins in the cytoplasm are maintained at a very low level. However, TERRA transport to the cytoplasm due to dysfunctional telomeres would result in the synthesis of significant amounts of the two dipeptide repeat proteins. Our observation of VR inclusions in the nuclei of cells with damaged telomeres suggests that the levels of VR in these cases may be significant and likely high enough to bind replication forks and Holliday junctions as observed in the in vitro experiments.

Until more is known about these dipeptide repeat proteins, it would be a mistake to assume their only biological activity is related to toxicity. Future studies of the synthesis, degradation, and levels of VR and GL dipeptide proteins during the cell cycle and general analysis of their toxicity will be of great value in providing clues into their function in the cell. Toxicity analysis employing overexpression from plasmids and adding synthetic repeating dipeptides directly to cells, as carried out in the C9orf72 ALS/FTD studies (reviewed above), represent in-depth studies in themselves and may vary between different cell types. For example, U2OS cells may have adapted to the presence of levels of VR and GL dipeptide protein which in other cells would be lethal.

Telomeres have been considered sensitive detectors of cellular DNA damage and recently Barnes et al. (81) demonstrated that even mild damage to telomeres in the form of 8-oxo-guanine can trigger cellular senescence even in the absence of telomere shortening. The precise details of how damaged telomeres activated senescence pathways were unclear; however, the findings presented here suggest the potential involvement of VR and or GL dipeptide proteins as signals for dysfunctional telomeres. For the future, studies of the cell cycle expression of VR and GL, their general toxicity, and their involvement in immune signaling pathways should shed insights into their function in normal and transformed cells. It is noteworthy that an unexpected discovery in a rare human expanded nucleotide repeat disease, SCA8, by Ranum et al. (17) may have helped reveal that mammalian telomeres encode two novel dipeptide repeat proteins.

Materials and Methods

See *SI Appendix* for further materials and methods.

Antibodies. The antibody to (VR) was generated in rabbits by Bio synth Inc. using peptide #2 (*SI Appendix*, Table S1).

Peptides and Proteins. The biotin-labeled $(VR)_{10}$ peptide (*SI Appendix*, Table S1) was synthesized at University of North Carolina (UNC) in a core facility directed by David Klapper. All other peptides were synthesized by Vivitide inc (now Bio synth Inc.). Streptavidin was from Invitrogen Inc.

Preparation of TERRA. TERRA was prepared as described (14). In brief, the pRST5 plasmid was linearized so that one end contains a T7 RNA polymerase promoter followed by a long TTAGGG repeat block. The DNA was transcribed with T7 RNA polymerase (MAXIscript T7 transcription Kit, Invitrogen) using conditions described by the vendor, and the RNA was purified using an RNA Clean & Concentrator kit (Zymo Research).

Image Adjustment and Manipulation. TEM images in TIFF format were adjusted for contrast and brightness using Adobe Photoshop software. In Fig. 2*B*, the micron bar was copied and moved into the area shown.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We wish to thank Ms. Smaranda Willcox for her help in amplifying and purifying the 3X Flag-RV_{60} plasmid, generating and purifying TERRA, and reviewing the manuscript. We thank Lubomir Tomaska, Laura Ranum, Maurice Swanson, and David Maranon for helpful advice on this study. In particular, we thank the two reviewers of this manuscript for insightful comments and suggestions. We thank James Bear for his help and overseeing the light microscopy facility in the Lineberger Comprehensive Cancer Center. We thank Tal Kofri in the UNC Viral Vector Core for generating the Lentivirus stock and the laboratory of Ronald Swanstrom in the Lineberger Comprehensive Cancer Center in help with the Lentivirus infections. Scott Lewis at Bio synth Inc. was most helpful in generating the peptides and antibodies. This work was supported by grants from the NIH (GM31819, ES013773, and ESO31635).

1. B. McClintock, The stability of broken ends of chromosomes in Zea mays. *Genetics* **26**, 234–282 (1941).
2. J. Meyne, R. L. Ratliff, R. K. Moyzis, Conservation of the human telomere sequence (TTAGGG)(n) among vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7049–7053 (1989).
3. T. de Lange, Shelterin-mediated telomere protection. *Annu. Rev. Genet* **52**, 223–247 (2018).
4. J. D. Griffith *et al.*, Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
5. L. Tomáška, A. J. Cesare, T. M. AlTurki, J. D. Griffith, Twenty years of t-loops: A case study for the importance of collaboration in molecular biology. *DNA Repair (Amst)* **94**, 102901 (2020).
6. C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto, J. Lingner, Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801 (2007).

7. S. Schoeftner, M. A. Blasco, Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10**, 228–236 (2008).
8. C. M. Azzalin, J. Lingner, Telomere functions grounding on TERRA firma. *Trends Cell Biol.* **25**, 29–36 (2015).
9. Z. Wang *et al.*, Telomeric repeat-containing RNA (TERRA) constitutes a nucleoprotein component of extracellular inflammatory exosomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6293–E6300 (2015).
10. C. A. Perez-Romero, M. Lalonde, P. Chartrand, E. Cusanelli, Induction and relocalization of telomeric repeat-containing RNAs during diauxic shift in budding yeast. *Curr. Genet.* **64**, 1117–1127 (2018).
11. A. Porro, S. Feuerhahn, P. Reichenbach, J. Lingner, Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol. Cell Biol.* **30**, 4808–4817 (2010).
12. Z. Deng, J. Norseen, A. Wiedmer, H. Riethman, P. M. Lieberman, TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).
13. P. N. Le, D. G. Maranon, N. H. Altina, C. L. R. Battaglia, S. M. Bailey, TERRA, hnRNP A1, and DNA-PKcs interactions at human telomeres. *Front Oncol.* **3**, 91 (2013).
14. R. Arora, C. M. Azzalin, Telomere elongation chooses TERRA ALTernatives. *RNA Biol.* **12**, 938–941 (2015).
15. M. Feretzaki, P. R. Nunes, J. Lingner, Expression and differential regulation of human TERRA at several chromosome ends. *RNA* **25**, 1470–1480 (2019).
16. A. Randall, J. D. Griffith, Structure of long telomeric RNA transcripts: The G-rich RNA forms a compact repeating structure containing G-quartets. *J. Biol. Chem.* **284**, 13980–13986 (2009).
17. T. Zu *et al.*, Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 260–265 (2011).
18. F. Ayhan *et al.*, SCA 8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF 3F. *EMBO J.* **37**, e99023 (2018).
19. A. E. Renton *et al.*, A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
20. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
21. K. Reddy, B. Zamiri, S. Y. R. Stanley, R. B. Macgregor, C. E. Pearson, The disease-associated r(GGGGCC) n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* **288**, 9860–9866 (2013).
22. L. P. W. Ranum *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–E4977 (2013).
23. P. E. A. Ash *et al.*, Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639–649 (2013).
24. K. Mori *et al.*, The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1338 (2013).
25. M. Banez-Coronel, L. P. W. Ranum, Repeat-associated non-AUG (RAN) translation: Insights from pathology. *Lab. Invest.* **99**, 929–942 (2019).
26. L. Nguyen, J. D. Cleary, L. P. W. Ranum, Repeat-associated non-ATG translation: Molecular mechanisms and contribution to neurological disease. *Annu. Rev. Neurosci.* **42**, 227–247 (2019).
27. I. Kwon *et al.*, Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **1979** (345), 1139–1145 (2014).
28. Y. Lin *et al.*, Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **167**, 789–802.e12 (2016).
29. R. Lopez-Gonzalez *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 (2016).
30. M. A. Farg, A. Konopka, K. Y. Soo, D. Ito, J. D. Atkin, The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis. *Hum. Mol. Genet* **26**, 2882–2896 (2017).
31. N. S. Andrade *et al.*, Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair in C9ORF72 ALS/FTD. *Mol. Neurodegener.* **15**, 13 (2020).
32. Y. J. Zhang *et al.*, Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol.* **128**, 505–524 (2014).
33. Y. J. Zhang *et al.*, C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat. Neurosci.* **19**, 668–677 (2016).
34. M. Proudfoot *et al.*, Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. *Acta Neuropathol.* **127**, 451–458 (2014).
35. C. Wicky *et al.*, Telomeric repeats (TTAGGC)n are sufficient for chromosome capping function in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8983–8988 (1996).
36. Z. Wang, P. M. Lieberman, The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol.* **13**, 690–695 (2016).
37. J. Nassour *et al.*, ZBP1-mediated telomere to mitochondria cross talk prevents cancer initiation. *Nature*, (2023), https://doi.org/10.1038/s41586-023-05710-8.
38. B. N. Flores *et al.*, Distinct c9orf72-associated dipeptide repeat structures correlate with neuronal toxicity. *PLoS One* **11**, e0165084 (2016).
39. Y. J. Chang, U. S. Jeng, Y. L. Chiang, I. S. Hwang, Y. R. Chen, The glycine-alanine dipeptide repeat from C9 or f72 hexanucleotide expansions forms toxic amyloids possessing cell-to-cell transmission properties. *J. Biol. Chem.* **291**, 4903–4911 (2016).
40. Q. Guo *et al.*, In situ structure of neuronal C9orf72 Poly-GA aggregates reveals proteasome recruitment. *Cell* **172**, 696–696.e12 (2018).
41. A. Gustot *et al.*, Activation of innate immunity by lysozyme fibrils is critically dependent on cross-β sheet structure. *Cell Mol. Life Sci.* **70**, 2999–3012 (2013).
42. C. Westwell-Roper, H. C. Denroche, J. A. Ehses, C. B. Verchere, Differential activation of innate immune pathways by distinct islet amyloid polypeptide (IAPP) aggregates. *J. Biol. Chem.* **291**, 8908–8917 (2016).
43. D. Subramanian, J. D. Griffith, Modulation of p53 binding to Holliday junctions and 3-cytosine bulges by phosphorylation events. *Biochemistry* **44**, 2536–2544 (2005).
44. D. Subramanian, J. D. Griffith, p53 monitors replication fork regression by binding to "chickenfoot" intermediates. *J. Biol. Chem.* **280**, 42568–42572 (2005).
45. S. Lee, L. Cavallo, J. Griffith, Human p53 binds holliday junctions strongly and facilitates their cleavage. *J. Biol. Chem.* **272**, 7532–7539 (1997).
46. E. M. M. Manders, F. J. Verbeek, J. A. Aten, Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).
47. L. J. Ng, J. E. Cropley, H. A. Pickett, R. R. Reddel, C. M. Suter, Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. *Nucleic Acids Res.* **37**, 1152–1159 (2009).
48. R. Arora *et al.*, RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5**, 5220 (2014).
49. W. Xinmei *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 (2014).
50. Z. Tao *et al.*, Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum. Mol. Genet* **24**, 2426–2441 (2015).
51. C. Chen *et al.*, Phase separation and toxicity of c9orf72 poly(Pr) depends on alternate distribution of arginine. *J. Cell Biol.* **220**, e202103160 (2021).
52. S. May *et al.*, C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol.* **128**, 485–503 (2014).
53. S. Yehezkel, Y. Segev, E. Viegas-Péquignot, K. Skorecki, S. Selig, Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum. Mol. Genet* **17**, 2776–2789 (2008).
54. S. Sagie *et al.*, Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat. Commun.* **8**, 14015 (2017).
55. A. E. Carpenter *et al.*, Cell profiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
56. I. R. A. Mackenzie *et al.*, Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol.* **130**, 845–861 (2015).
57. M. H. Schludi *et al.*, Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol.* **130**, 537–555 (2015).
58. D. Kalderon, B. L. Roberts, W. D. Richardson, A. E. Smith, A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509 (1984).
59. N. Arnoult, A. van Beneden, A. Decottignies, Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. *Nat. Struct. Mol. Biol.* **19**, 948–956 (2012).
60. A. Porro *et al.*, Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* **5**, 5379 (2014).
61. J. Aguado *et al.*, Inhibition of DNA damage response at telomeres improves the detrimental phenotypes of Hutchinson-Gilford progeria syndrome. *Nat. Commun.* **10**, 4990 (2019).
62. H. P. Chu *et al.*, TERRA RNA antagonizes ATRX and protects telomeres. *Cell* **170**, 86–101.e16 (2017).
63. J. J. Montero, I. López De Silanes, O. Granā, M. A. Blasco, Telomeric RNAs are essential to maintain telomeres. *Nat. Commun.* **7**, 12534 (2016).
64. A. J. Cesare, M. T. Hayashi, L. Crabbe, J. Karlseder, The telomere deprotection response is functionally distinct from the Genomic DNA damage response. *Mol. Cell* **51**, 141–155 (2013).
65. I. González-Vasconcellos, M. A. Cobos-Fernández, M. J. Atkinson, J. Fernandez-Piqueras, J. Santos, Quantifying telomeric lncRNAs using PNA-labelled RNA-Flow FISH (RNA-Flow). *Commun. Biol.* **5**, 513 (2022).
66. A. Patel *et al.*, A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
67. S. Boyko, K. Surewicz, W. K. Surewicz, Regulatory mechanisms of tau protein fibrillation under the conditions of liquid-liquid phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31882–31890 (2020).
68. D. A. Solomon, R. Smikle, M. J. Reid, S. Mizielinska, Altered phase separation and cellular impact in C9orf72-linked ALS/FTD. *Front Cell Neurosci.* **15**, 664151 (2021).
69. S. Yehezkel *et al.*, Characterization and rescue of telomeric abnormalities in ICF syndrome type I fibroblasts. *Front Oncol.* **3**, 35 (2013).
70. D. Yang *et al.*, FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol.* **130**, 525–535 (2015).
71. B. D. Freibaum, J. P. Taylor, The role of dipeptide repeats in C9ORF72-related ALS-FTD. *Front Mol. Neurosci.* **10**, 35 (2017).
72. N. A. Lanson, U. B. Pandey, FUS-related proteinopathies: Lessons from animal models. *Brain Res.* **1462**, 44–60 (2012).
73. E. N. Guerrero *et al.*, TDP-43/FUS in motor neuron disease: Complexity and challenges. *Prog. Neurobiol.* **145–146**, 79–97 (2016).
74. N. Hamad *et al.*, Non-coding RNA suppresses FUS aggregation caused by mechanistic shear stress on pipetting in a sequence-dependent manner. *Sci. Rep.* **11**, 9523 (2021).
75. S. Maharana *et al.*, RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
76. A. Halle *et al.*, The NALP3 inflammasome is involved in the innate immune response to amyloid-β. *Nat. Immunol.* **9**, 857–865 (2008).
77. A. Salminen, J. Ojala, A. Kauppinen, K. Kaarniranta, T. Suuronen, Inflammation in Alzheimer's disease: Amyloid-β oligomers trigger innate immunity defence via pattern recognition receptors. *Prog. Neurobiol.* **87**, 181–194 (2009).
78. D. Chakravarti *et al.*, Telomere dysfunction instigates inflammation in inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024853118 (2021).
79. D. Conomos *et al.*, Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *J. Cell Biol.* **199**, 893–906 (2012).
80. Y. Gao *et al.*, Regulation of TERRA on telomeric and mitochondrial functions in IPF pathogenesis. *BMC Pulm Med.* **17**, 163 (2017).
81. R. P. Barnes *et al.*, Telomeric 8-oxo-guanine drives rapid premature senescence in the absence of telomere shortening. *Nat. Struct. Mol. Biol.* **29**, 639–652 (2022).