Supplemental Information

In vivo formation of vacuolated multi-phase compartments lacking membranes

Hermann Broder Schmidt and Rajat Rohatgi

Contents:

Supplemental Methods

Characterization of TDP43 particle shape and size

Morphological analysis of images containing $TDP43_{RRM\text{-}GFP}$ particles was performed using custom-written Mathematica (Wolfram Research) scripts. In short, individual particles were extracted from multiple acquired images and manually inspected to (i) lay in the confocal plane, (ii) contain no direct neighbors and (iii) are not adjacent to an image border. Using the in-built component measurement functions of Mathematica, the following parameters were measured for each extracted particle: (i) particle perimeter, (ii) the perimeter of a perfectly round disk equivalent to the particle size and (iii) the radius of this equivalent disk. Using parameters i and ii, the circularity of the particles was determined as follows: circularity = particle perimeter / equivalent disk perimeter. Hence, a perfect sphere would have a circularity of 1.0 in this analysis. Given that the $TDP43_{RRM\text{-}GFP}$ particles are highly spherical according to this definition (Figure 1), the particle volumes (V) were estimated using the volume formula for a sphere: $V = 4/3\ \pi\ r^3$, where r is the determined equivalent disk radius.

Estimation of intra-particle dynamics

To visualize the intra-particle dynamics of the nucleoplasm-filled particles, time-resolved xyz-image stacks were acquired with a Leica SP8 confocal laser scanning microscope in the GFP channel after excitation with a 488 nm laser. Per particle, 25-30 z-planes in 0.33 μm steps were taken over 5 minutes. 3D reconstruction and particle sectioning were done with the Image3D function of Mathematica.

For spot-bleach fluorescence recovery after photobleaching (FRAP) experiments, IBB-MBP-mCherry in the nucleoplasm was bleached in $\approx 1.7\ \mu m^2$ areas for 10 frames (75 milliseconds each)

with a 561 nm laser and fluorescence recovery recorded every 75 milliseconds for a total of 60 frames. To bleach IBB-MBP-mCherry in the bubbles of $TDP43_{RRM\text{-}GFP}$ particles, the area had to be reduced to ≈ 0.3 μm^2 to avoid complete bleaching of the structure. Otherwise, the bleaching settings were identical. Spot-bleach experiments of $TDP43_{RRM\text{-}GFP}$ in the particles were performed in a ≈ 0.3 μm^2 area as well, however using the 488 nm laser line. A Leica SP8 confocal laser scanning microscope and the Leica FRAP software module were used for all bleaching experiments in this study. For each recorded time point (t), the fluorescence intensities within the bleached area were integrated using the Leica software and normalized to the fluorescence intensity of an unbleached control area. These normalized, time-dependent fluorescence intensities I_t were then used to calculate the fluorescence recovery (FR) according to the following formula: $FR_t = (I_t - I_{t0}) / (I_{before\ bleaching} - I_{t0})$, with $t0$ being the first time point observed after photobleaching. The data was then plotted and fitted with the Prism software (Graphpad) to derive the half-times ($\tau 1/2$) of fluorescence recovery. Diffusion coefficients (D) were estimated as described (Axelrod et al., 1976; Soumpasis, 1983). Essentially, D is proportional to $w^2 / t1/2$, where w is the width of the bleached area.

Half-bleach experiments were performed such that $TDP43_{RRM\text{-}GFP}$ particles were bleached in one hemisphere (typically ≈ 22 μm^2) for 10 frames with a 488 nm laser. Fluorescence recovery was subsequently recorded every 10 seconds for a total of 20 frames. After bleaching, we noticed a rapid equilibration of fluorescence (EF) between the bleached and unbleached particle hemispheres, which we expressed as: $EF = (Pixel_{bleached\ area}(t) + Pixel_{unbleached\ area}(t)) / Pixel_{particle}(t0)$, where $Pixel(t)$ are the pixel numbers in the indicated areas at each recorded time point (t) and $t0$ the time point before bleaching. To quantify the number of pixels, custom-written

Mathematica scripts relying on built-in morphological image analysis functions were used. The equilibration half-times and diffusion coefficients were estimated as described above.

For full-bleach experiments, xyz-images (25-30 planes of 0.33 μm) were acquired before and after photobleaching (one stack every $\approx$ 15 sec for a total of 5 min). Our setup, however, limited bleaching to a single plane only, which is why we extended the bleaching duration to 20 frames at maximum laser intensity. This was sufficient to fully bleach most particles. For analysis, the average intensity of the sectioned particle was measured in every recorded plane and plotted as fluorescence intensity heat maps against time using custom Mathematica scripts. To correct for additional bleaching during image acquisition, the same procedure was applied to unbleached control particles (recorded in parallel to the bleached ones), which were then used for normalization. To compare multiple experiments, the normalized fluorescence intensity was integrated over all recorded planes for each time point and plotted against time. Using the Prism software to fit the data, the half-times of particle fluorescence recovery were estimated.

Estimation of the critical concentration for phase separation and the final particle concentration

In order to convert measured fluorescence intensities into protein concentrations, we calibrated our setup with a titration series of purified recombinant GFP. The average image intensities of 40 μM, 30 μM, 20 μM, 15 μM, and 12 μM GFP solutions in imaging medium were measured and normalized to the average image intensity of medium lacking GFP. The data was plotted and fitted with the Prism software to obtain a linear regression function describing the relation between fluorescence intensity and protein concentration. This function was then used to estimate the concentration of $TDP43_{RRM\text{-}GFP}$ in phase-separated particles and the surrounding nucleoplasm based on the GFP signal intensities measured in the respective areas.

To form particles, $TDP43_{RRM\text{-}GFP}$ constructs were transfected into HEK293T cells and 24 hours later imaged live without fixation using a temperature- and gas-controlled Leica SP8 confocal laser scanning microscope as described in the methods. Detector settings were adjusted such that the GFP signal of the particles was well within the dynamic range. Per transfected construct, 12 images each containing > 20 particles were recorded. For quantification of the GFP signals emitted by the particles and the surrounding nucleoplasm, the recorded images were first cropped around individual particles. To do so, the morphology detection functions of Mathematica were used to automatically detect particles and determine the image coordinates of rectangular boxes bounding them. To include regions of nucleoplasm, the bounding boxes were expanded by 20 pixels. Next, all isolated particles were manually inspected to (i) lay in the confocal plane, (ii) contain no neighboring particles and (iii) are not adjacent to an image border. Finally, the fluorescence intensities in the particles (excluding bubbles) and the surrounding nucleoplasm were measured using the morphology detection and component measurements functions of Mathematica. To correct for background fluorescence, the obtained values were normalized to the average nuclear fluorescence signal of untransfected HEK293T cells.

We note the following limitations of our measurements, which may explain why our estimates are lower than expected from recent approximations of the concentration of *in vitro* FUS liquid droplets (≈ 120 mg/ml) (Burke et al., 2015) or membrane-less organelles such as Cajal bodies (≈ 140 mg/ml) or nuclear speckles (≈ 160 mg/ml) (Handwerger et al., 2005). First, a large spectrum of fluorescence intensities, ranging from the nuclear GFP signal to the recombinant GFP titration series to the highly enriched GFP signals in the particles, has to be covered with a limited dynamic range of the detector. Second, it is difficult to achieve concentrations of purified recombinant GFP that yield fluorescence signals as high as those detected in the $TDP43_{RRM\text{-}GFP}$ particles to cover the

entire signal range with a standard curve. Third, our approach relies on the assumption that GFP fluorescence remains unaltered when incorporated into TDP43 and condensed in liquid droplets. Barring these technical limitations, examples from the literature suggest that protein concentrations in liquid assemblies stably residing inside living cells may indeed not be much higher than the 50 mg/ml that we estimated (Table S1). The intrinsically-disordered FG domain of the yeast nucleoporin Nsp1, for instance, readily forms hydrogels – a much more condensed and less dynamic state compared to liquid droplets – at concentrations as low as 20 mg/ml *in vitro* (Frey and Görlich, 2007). Similarly, a fusion of the FUS IDR with GFP already started to assemble *in vitro* into solid hydrogels at a concentration of 60 mg/ml (Kato et al., 2012). Indeed, also the *in vitro* formed liquid FUS assemblies (with concentrations of 120 mg/ml) rapidly converted into aggregates, a behavior not seen in cells (Lin et al., 2015; Patel et al., 2015).

Replacement of the conserved diagnostic motif in the IDR of TDP43

Sequence alignments revealed that the IDR of TDP43 contains a highly conserved, 28 amino acid (aa) stretch, which varies notably in its sequence composition from the rest of the IDR. This finding is especially striking given that (i) the IDR itself is neither well conserved in sequence composition nor sequence length, which is expected for such low-complexity regions (Denning and Rexach, 2007), whereas (ii) the 28-aa stretch in question is indeed so conserved that it can serve as a diagnostic motif for vertebrate TDP43. We hence hypothesized that this conserved motif may play an important role on the phase separation abilities of TDP43. However, to test this experimentally, we did not simply want to delete this stretch, effectively shortening the IDR. Rather, we decided to replace it by an equally-long sequence that matches the IDR in its composition. To derive such a sequence, we computed the percentage occurrence of each amino acid in the IDR (here: residues

274-319 and 365-414) and used this to predict how many residues of each amino acid would be expected to occur in a 28-aa sequence. From the resulting list, we randomly generated one-thousand 28-aa long sequences and randomly chose one that contained a "GGFGN" motif, a consensus typically found in the IDR of TDP43. All sequence analysis was done with custom-written Mathematica scripts. The amino acid sequence was then back-translated into a non-repetitive nucleotide sequence using the GeneDesigner2 software (DNA2.0), commercially synthesized (IDT) and sub-cloned as indicated in Table S2 using Gibson Assembly.

Supplemental Tables

Table S1. Estimation of key phase descriptors for $TDP43_{RRM\text{-}GFP}$ particles. *Related to Figures 2, 4 and S4.*

$TDP43_{RRM\text{-}GFP}$ reporter	Critical concentration[1]	Particle concentration[1]	Intra-particle diffusion constant[2]	Particle viscosity[4]
WT	≈ 5.0 ± 2.5 μM	≈ 1.0 ± 0.2 mM	(≈ 0.1-0.5 μm^2/sec)[3] ≈ 0.1-1.5 μm^2/sec	≈ 0.075- 1.15 Pa · sec
M337V	≈ 5.5 ± 1.5 μM	≈ 1.0 ± 0.2 mM	≈ 0.01-0.1 μm^2/sec	≈ 1.15-11.5 Pa · sec
N345K	≈ 4.5 ± 1.0 μM	≈ 0.9 ± 0.1 mM	≈ 0.08-1.0 μm^2/sec	≈ 0.1-1.5 Pa · sec
A382T	≈ 5.5 ± 2.0 μM	≈ 0.9 ± 0.2 mM	≈ 0.07-0.9 μm^2/sec	≈ 0.1-1.5 Pa · sec

[1]Estimated as described in Fig. S4 and the supplementary methods section.
[2]Estimations based on half-bleach experiments from Figs. 2 and 4.
[3]Estimated based on spot-bleach experiments from Fig. 2.
[4]Estimated based on the crude assumptions that (i) the respective $TDP43_{RRM\text{-}GFP}$ phases behave as Newtonian fluids and (ii) an average spherical protein of 2nm radius has a similar diffusion constant as the respective $TDP43_{RRM\text{-}GFP}$ construct.
See methods for details.

Table S2. Bacterial and mammalian expression constructs used in this study. *Related to Materials and Methods.*

Construct name	Plasmid	Construct description	Promoter	Used in figures
GFP	pHBS797	His_{18}-mEGFP	LacO	S4
GFP-TDP43	pHBS861	Full-length TDP43 with an N-terminal GFP tag	CMV-TetO	S1
GFP-TDP43	pHBS837	Full-length TDP43 with an N-terminal GFP tag	CMV	S1
GFP-TDP43 4F-L	pHBS917	GFP-tagged full-length TDP43 with Phe147, Phe149, Phe229 and Phe231 mutated to Leu	CMV	S1
GFP-TDP43 ΔRRM	pHBS916	N-terminally GFP-tagged TDP43 lacking its RRM domains	CMV	S1
$TDP43_{RRM\text{-}GFP}$	pHBS838	TDP43 with its RRM domains replaced with GFP	CMV	1, 2, 3, 4, S1, S3, S4
$TDP43_{RRM\text{-}GFP}$	pHBS857	TDP43 with its RRM domains replaced with GFP	CMV-TetO	S3
$TDP43_{RRM\text{-}GFP}$ $NLS1_{mut}$	pHBS865	$TDP43_{RRM\text{-}GFP}$ with a mutated nuclear localization signal (NLS)	CMV	S3
$TDP43_{RRM\text{-}GFP}$ M337V	pHBS850	$TDP43_{RRM\text{-}GFP}$ with M337V mutation	CMV	4
$TDP43_{RRM\text{-}GFP}$ N345K	pHBS851	$TDP43_{RRM\text{-}GFP}$ with N345K mutation	CMV	4
$TDP43_{RRM\text{-}GFP}$ A382T	pHBS852	$TDP43_{RRM\text{-}GFP}$ with A382T mutation	CMV	4
$TDP43_{RRM\text{-}mCherry}$	pHBS853	TDP43 with its RRM domains replaced with mCherry	CMV	3, S3
$TDP43_{RRM\text{-}mCherry}$ ΔCR	pHBS909	$TDP43_{RRM\text{-}mCherry}$ with deleted diagnostic motif	CMV	3
$TDP43_{RRM\text{-}mCherry}$ CR_{mut}	pHBS863	$TDP43_{RRM\text{-}mCherry}$ with replaced diagnostic motif	CMV	3
$TDP43_{RRM\text{-}UnaG}$	pHBS800	TDP43 with its RRM domains replaced with UnaG	CMV	S3
mCherry		Untagged mCherry	CMV	1, S3
IBB-MBP-mCherry	pHBS864	Fusion of an Importinβ-binding domain (IBB), the maltose binding protein (MBP) and mCherry	CMV	1, 2

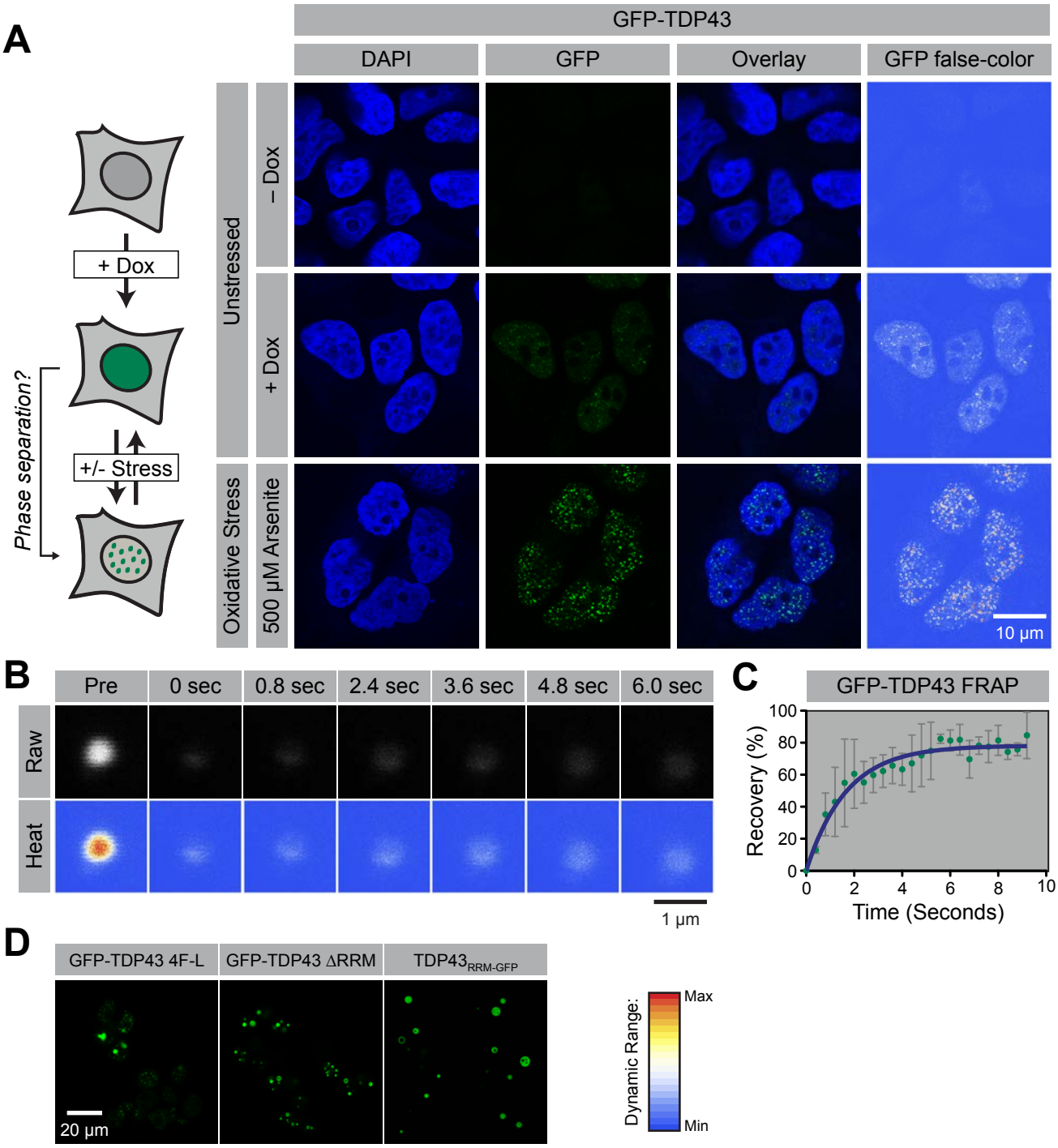
A
GFP-TDP43
DAPI
GFP
Overlay
GFP false-color
Unstressed
– Dox
+ Dox
Oxidative Stress
500 µM Arsenite
+ Dox
+/- Stress
Phase separation?
10 µm
B
Pre
0 sec
0.8 sec
2.4 sec
3.6 sec
4.8 sec
6.0 sec
Raw
Heat
1 µm
C
GFP-TDP43 FRAP
Recovery (%)
Time (Seconds)
D
GFP-TDP43 4F-L
GFP-TDP43 ΔRRM
TDP43RRM-GFP
20 µm
Dynamic Range:
Max
Min

Figure S1

Figure S1. Stress-induced TDP43 assemblies have liquid-like properties in vivo. *Related to the Results.*

(**A**) Full-length TDP43 with a GFP-tag at its N-terminus shows a mainly diffuse nuclear localization in a stable cell line after induction with 1 μg/ml doxycycline (Dox) for 48 hours. Upon treatment with 500 μM sodium arsenate for 1 hour, TDP43 assembles into spherical nuclear granules. (**B**) Representative images of the fluorescence signal recovery after photo-bleaching a stress-induced TDP43 particle in one hemisphere. (**C**) Combined analysis of the fluorescence recovery kinetics of four half-bleached TDP43 particles suggests that they have liquid-like properties. (**D**) Impairment of RNA binding, either by mutating four key phenylalanine (F) residues to leucines (L) in the RNA binding RRM domains (GFP-TDP43 4F-L; (Buratti and Baralle, 2001) or by deleting them altogether (GFP-TDP43 ΔRRM and $TDP43_{RRM\text{-}GFP}$), causes TDP43 to form particles independently of stress. Shown are representative raw images with independently adjusted detector settings.

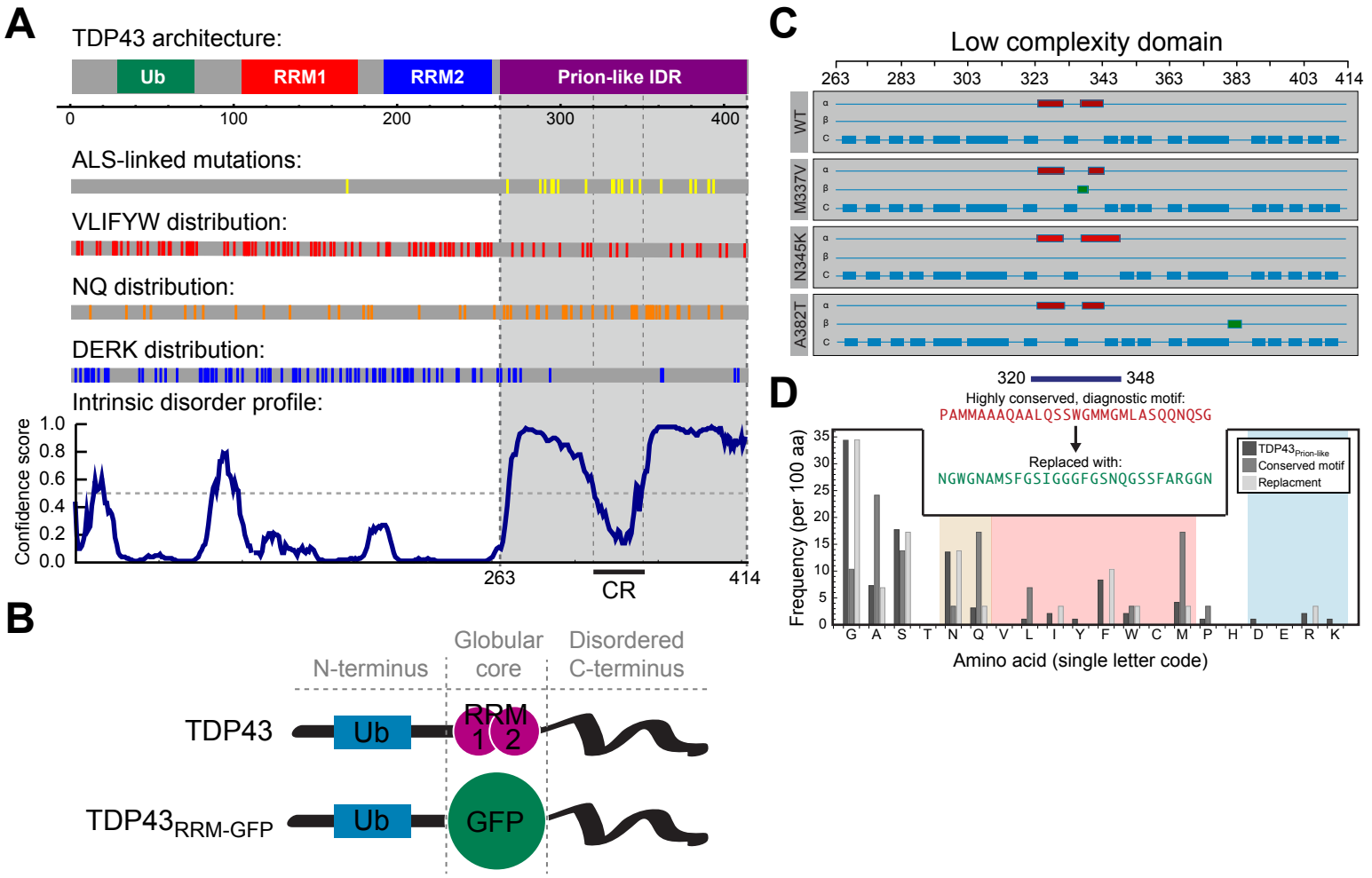
A
TDP43 architecture:
Ub
RRM1
RRM2
Prion-like IDR
0
100
200
300
400
ALS-linked mutations:
VLIFYW distribution:
NQ distribution:
DERK distribution:
Intrinsic disorder profile:
Confidence score
1.0
0.8
0.6
0.4
0.2
0.0
263
414
CR
B
N-terminus
Globular core
Disordered C-terminus
TDP43
Ub
RRM 1 2
TDP43RRM-GFP
Ub
GFP
C
Low complexity domain
263
283
303
323
343
363
383
403
414
WT
M337V
N345K
A382T
D
320
348
Highly conserved, diagnostic motif:
PAMMAAAQAALQSSWGMMGMLASQQNQSG
Replaced with:
NGWGNAMSFGSIGGGFGSNQGSSFARGGN
TDP43Prion-like
Conserved motif
Replacment
Frequency (per 100 aa)
35
30
25
20
15
10
5
0
G A S T N Q V L I Y F W C M P H D E R K
Amino acid (single letter code)

Figure S2

Figure S2. Domain and sequence overview of TDP43. *Related to Figures 1, 2, 3 and 4.*

(**A**) Domain architecture of TDP43, indicating the positions of the Ubiquitin-like domain (Ub), the RNA-binding RRM domains and the intrinsically-disordered, prion-like low-complexity (LC) domain. The ruler indicates residue number. Shown below are distributions of known ALS-linked mutations (yellow), hydrophobic amino acids (red), amine-group containing amino acids (orange) and charged residues (blue). The dark blue curve depicts the degree of intrinsic structural disorder in TDP43, as predicted by the DisoPred3 webserver (http://bioinf.cs.ucl.ac.uk/psipred/). The LC domain is enriched in asparagine (N) and glutamine (Q), but depleted of charged and hydrophobic residues. This is a characteristic sequence signature of NQ-rich prions (Alberti et al., 2009). Moreover, almost all known disease-linked mutations cluster in the intrinsically disordered region (IDR) of TDP43 (Pesiridis et al., 2009; Sreedharan et al., 2008). The stretch with the predicted low disorder in the LC domain corresponds to the conserved region (CR); see panels C and D for details. (**B**) Design of TDP43 construct used in this study. To investigate phase separation of TDP43 *in vivo*, we replaced its central RRM domains (together ≈ 18 kDa) with GFP (≈ 26 kDa). This strategy preserved the overall domain architecture of TDP43, in which a globular core of folded domains is flanked by a disordered C-terminal LC domain and a more structured N-terminal domain (see main text for details). (**C**) Secondary structure predictions of the TDP43 IDR for the wild-type, M337V, N345K and A382T variants used in this study. The ruler indicates residue numbers of full-length TDP43. Stretches of predicted α-helices are shown in red, stretches of predicted β-strands in green and putative coils (unstructured regions) in blue. Predictions were made using the Chou-Fasman algorithm implemented in the Protean program (DNASTAR). Note that the predicted secondary structure falls into the highly conserved motif (sequence in red letters; below the secondary structure predictions). Shown in green is the randomly-generated sequence

that was used to replace the conserved motif (see methods for details). (**D**) Sequence comparison of the prion-like IDR (dark gray), the conserved motif (gray) and the replacement motif (light gray), which was designed to match the amino acid composition of the IDR. See supplemental methods for details on the replacement strategy. Whereas the conserved motif is particularly AQM-rich, the IDR and replacement motifs are GNF-rich.

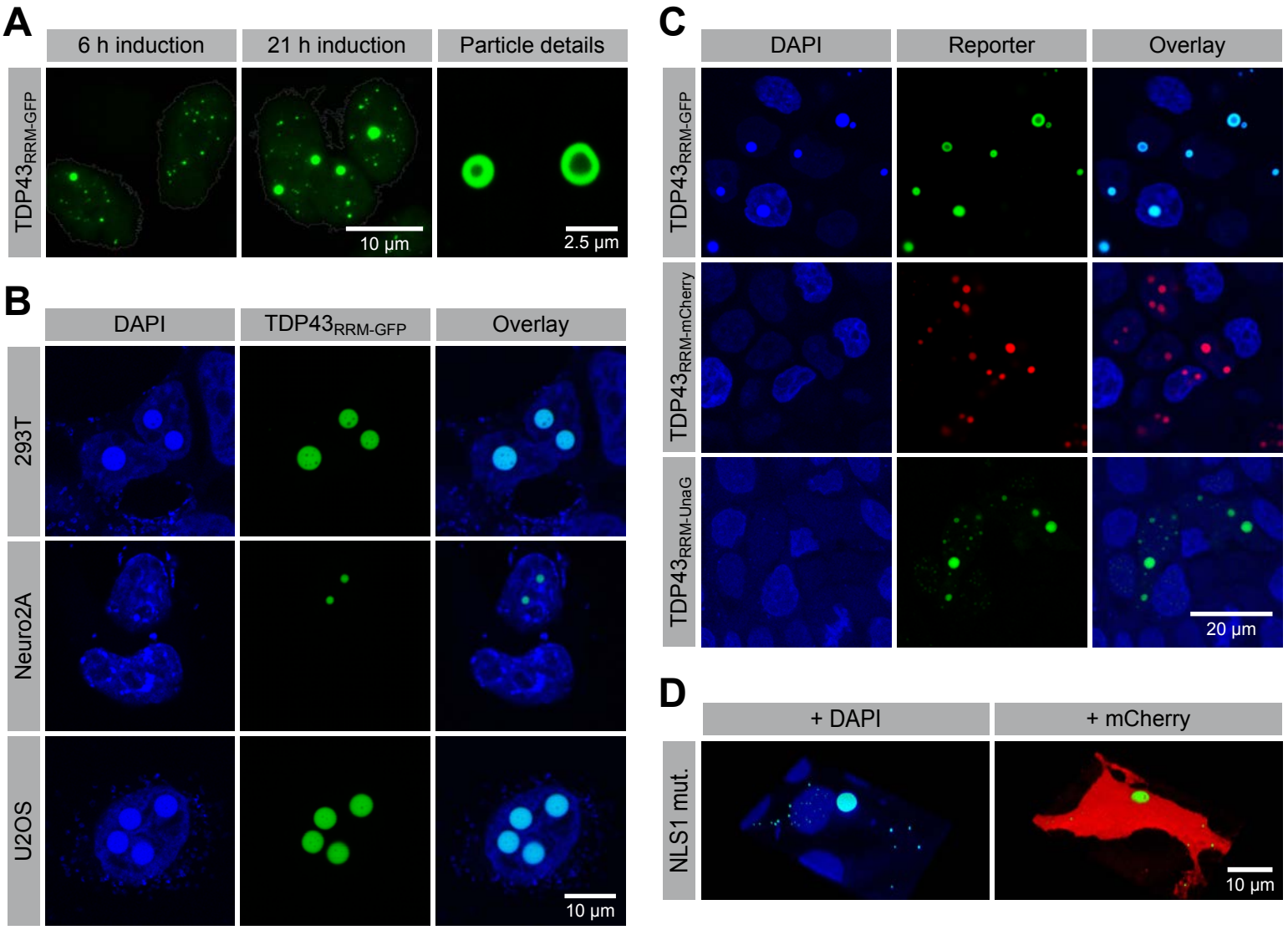
A
6 h induction
21 h induction
Particle details
TDP43RRM-GFP
10 µm
2.5 µm
B
DAPI
TDP43RRM-GFP
Overlay
293T
Neuro2A
U2OS
10 µm
C
DAPI
Reporter
Overlay
TDP43RRM-GFP
TDP43RRM-mCherry
TDP43RRM-UnaG
20 µm
D
+ DAPI
+ mCherry
NLS1 mut.
10 µm

Figure S3

Figure S3. TDP43 particles form independently of expression level, host cell type, fluorescent tag or compartment. *Related to Figure 1.*

(**A**) Stable cell lines expressing $TDP43_{RRM\text{-}GFP}$ under a doxycycline (Dox)-inducible promoter clearly contain $TDP43_{RRM\text{-}GFP}$ particles after induction with 1 μg/ml Dox for 48 hours. In contrast to transient transfections, both the transgene copy number and promoter strength are significantly reduced. Hence, $TDP43_{RRM\text{-}GFP}$ particles are not an artifact resulting from dramatic over-expression. Note that these particles, produced in stable cell lines, also contain internal bubbles. (**B**) Transiently-transfected $TDP43_{RRM\text{-}GFP}$ forms particles in HEK293T, Neuro2A and U2OS cells. (**C**) To rule out that particle formation is an artifact resulting from GFP, the TDP43 RRM domains were also exchanged for mCherry and the structurally unrelated green fluorescent protein UnaG (Kumagai et al., 2013). (**D**) Inactivation of the TPD43 nuclear localization signal (NLS) by mutation (Ayala et al., 2008; Winton et al., 2008) resulted in the formation of cytoplasmic (instead of nuclear) $TDP43_{RRM\text{-}GFP}$ particles, suggesting that the latter can form outside the nucleoplasmic environment. 3D image reconstructions from xyz-stacks are shown to demonstrate the cytoplasmic localization of the particles. Particles are in green, nuclei (DAPI) in blue, and mCherry (as a marker for the cytoplasm) in red.

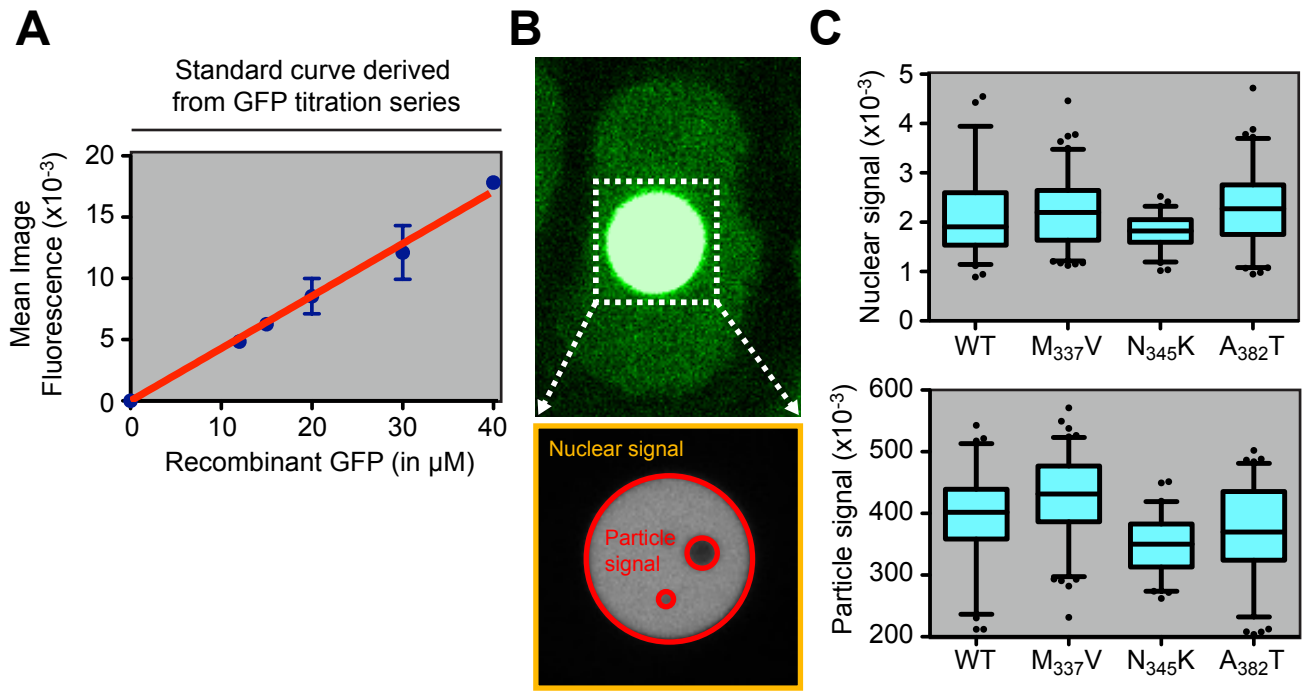
A
Standard curve derived
from GFP titration series
Mean Image
Fluorescence (x10-3)
20
15
10
5
0
0
10
20
30
40
Recombinant GFP (in µM)
B
Nuclear signal
Particle
signal
C
Nuclear signal (x10-3)
5
4
3
2
1
0
WT
M337V
N345K
A382T
Particle signal (x10-3)
600
500
400
300
200
WT
M337V
N345K
A382T

Figure S4

Figure S4. Estimation of the critical concentration for $TDP43_{RRM\text{-}GFP}$ phase separation and the final $TDP43_{RRM\text{-}GFP}$ concentration in the particles. *Related to Figures 1, 4 and Table S1.*

Please see methods for experimental details. (**A**) Titration series of recombinant GFP to convert fluorescence intensities into concentrations. Dark blue points represent mean pixel fluorescence intensity values (of five measurements per concentration), dark blue error bars show the standard deviation (SD) and the red line a linear fit of the data. (**B**) *Upper*: A high-gain, saturated image of the nucleus shows soluble and phase-separated wild-type $TDP43_{RRM\text{-}GFP}$. *Lower*: representative particle extracted from a recorded image to illustrate the quantified areas. Intensities of all pixels between the yellow and red lines were integrated to calculate the nuclear signal, and all pixels in the region outlined in red to determine the particle signal (excluding bubbles). (**C**) Measured particle (*upper*) and nucleoplasmic (*lower*) fluorescence intensities for the indicated $TDP43_{RRM\text{-}GFP}$ variants. Shown are box-whiskers-plots indicating median values (horizontal line) and 5-95% percentiles (error bars); outliers are depicted as black points.

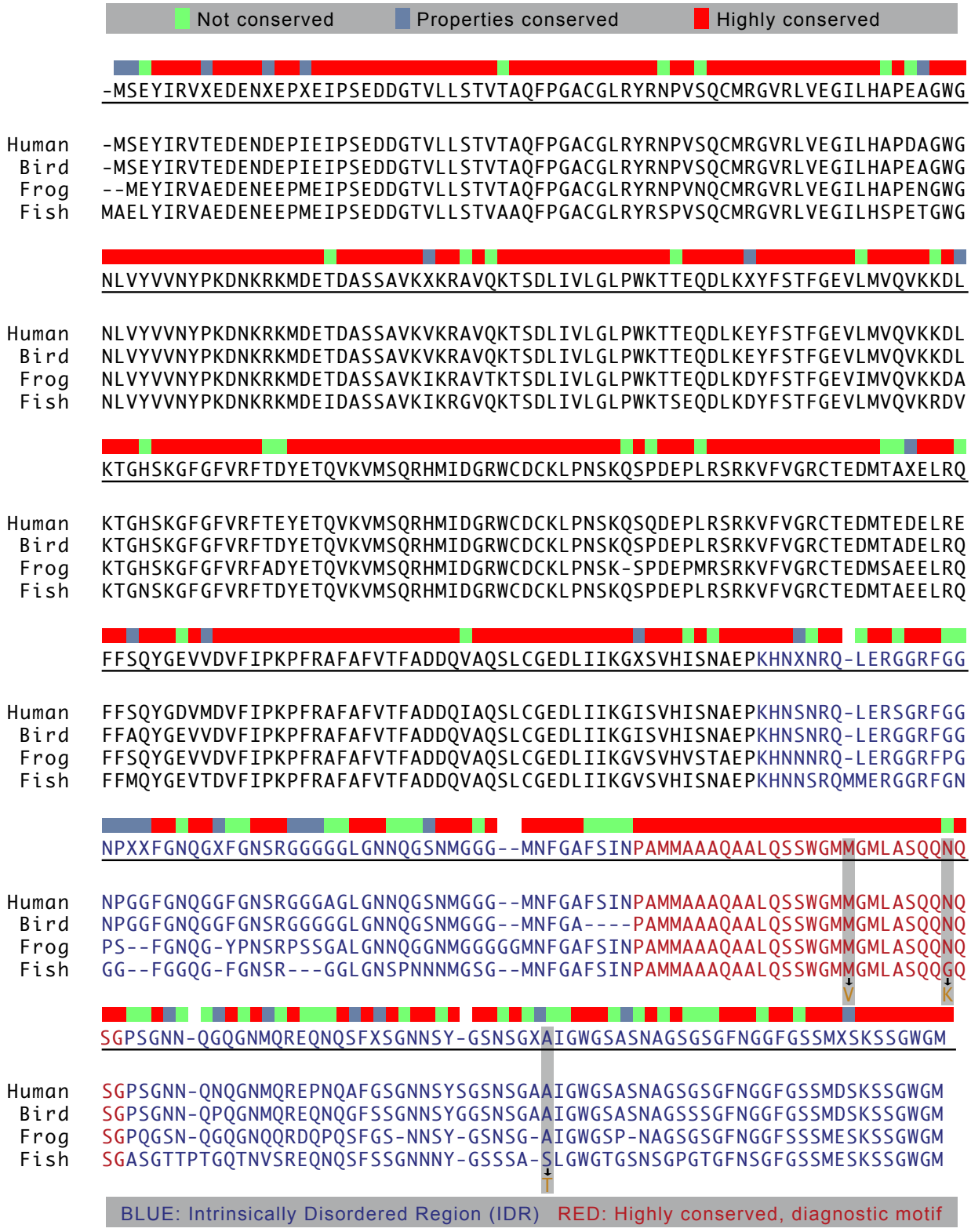
Not conserved
Properties conserved
Highly conserved
-MSEYIRVXEDENXEPXEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGILHAPEAGWG
Human -MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGILHAPDAGWG
Bird -MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGILHAPEAGWG
Frog --MEYIRVAEDENEEPMEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVNQCMRGVRLVEGILHAPENGWG
Fish MAELYIRVAEDENEEPMEIPSEDDGTVLLSTVAAQFPGACGLRYRSPVSQCMRGVRLVEGILHSPETGWG
NLVYVVNYPKDNKRKMDETDASSAVKXKRAVQKTSDLIVLGLPWKTTEQDLKXYFSTFGEVLMVQVKKDL
Human NLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDLKEYFSTFGEVLMVQVKKDL
Bird NLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDLKEYFSTFGEVLMVQVKKDL
Frog NLVYVVNYPKDNKRKMDETDASSAVKIKRAVTKTSDLIVLGLPWKTTEQDLKDYFSTFGEVIMVQVKKDA
Fish NLVYVVNYPKDNKRKMDEIDASSAVKIKRGVQKTSDLIVLGLPWKTSEQDLKDYFSTFGEVLMVQVKRDV
KTGHSKGFGFVRFTDYETQVKVMSQRHMIDGRWCDCKLPNSKQSPDEPLRSRKVFVGRCTEDMTAXELRQ
Human KTGHSKGFGFVRFTEYETQVKVMSQRHMIDGRWCDCKLPNSKQSQDEPLRSRKVFVGRCTEDMTEDELRE
Bird KTGHSKGFGFVRFTDYETQVKVMSQRHMIDGRWCDCKLPNSKQSPDEPLRSRKVFVGRCTEDMTADELRQ
Frog KTGHSKGFGFVRFADYETQVKVMSQRHMIDGRWCDCKLPNSK-SPDEPMRSRKVFVGRCTEDMSAEELRQ
Fish KTGNSKGFGFVRFTDYETQVKVMSQRHMIDGRWCDCKLPNSKQSPDEPLRSRKVFVGRCTEDMTAEELRQ
FFSQYGEVVDVFIPKPFRAFAFVTFADDQVAQSLCGEDLIIKGXSVHISNAEPKHNXNRQ-LERGGRFGG
Human FFSQYGDVMDVFIPKPFRAFAFVTFADDQIAQSLCGEDLIIKGISVHISNAEPKHNSNRQ-LERSGRFGG
Bird FFAQYGEVVDVFIPKPFRAFAFVTFADDQVAQSLCGEDLIIKGISVHISNAEPKHNSNRQ-LERGGRFGG
Frog FFSQYGEVVDVFIPKPFRAFAFVTFADDQVAQSLCGEDLIIKGVSVHVSTAEPKHNNNRQ-LERGGRFPG
Fish FFMQYGEVTDVFIPKPFRAFAFVTFADDQVAQSLCGEDLIIKGVSVHISNAEPKHNNSRQMMERGGRFGN
NPXXFGNQGXFGNSRGGGGGLGNNQGSNMGGG--MNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQ
Human NPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGG--MNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQ
Bird NPGGFGNQGGFGNSRGGGGGLGNNQGSNMGGG--MNFGA----PAMMAAAQAALQSSWGMMGMLASQQNQ
Frog PS--FGNQG-YPNSRPSSGALGNNQGGNMGGGGGMNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQ
Fish GG--FGGQG-FGNSR---GGLGNSPNNNMGSG--MNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQGQ
V
K
SGPSGNN-QGQGNMQREQNQSFXSGNNSY-GSNSGXAIGWGSASNAGSGSGFNGGFGSSMXSKSSGWGM
Human SGPSGNN-QNQGNMQREPNQAFGSGNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM
Bird SGPSGNN-QPQGNMQREQNQGFSSGNNSYGGSNSGAAIGWGSASNAGSSSGFNGGFGSSMDSKSSGWGM
Frog SGPQGSN-QGQGNQQRDQPQSFGS-NNSY-GSNSG-AIGWGSP-NAGSGSGFNGGFSSSMESKSSGWGM
Fish SGASGTTPTGQTNVSREQNQSFSSGNNNY-GSSSA-SLGWGTGSNSGPGTGFNSGFGSSMESKSSGWGM
T
BLUE: Intrinsically Disordered Region (IDR)
RED: Highly conserved, diagnostic motif

Figure S5

Figure S5. Sequence conservation of TDP43 in vertebrates. *Related to Figures 3, 4 and S1.*

Sequence alignments of human (GI: 6678271), bird (*Apaloderma vittatum*; GI: 699588799), frog (*Xenopus tropicalis;* GI: 38174066) and fish (*Lepisosteus oculatus;* GI: 573908659) TDP43. The sequence of the IDR is highlighted in blue. Embedded within the loosely conserved IDR is a highly conserved motif (shown in red and denoted as "CR" hereafter). Shaded residues highlight the three disease-linked mutations studied here and the orange letters below the corresponding amino acid changes. The color code above the consensus sequence indicates the degree of conservation: red denotes identical residues, blue denotes residues with similar physiochemical properties and yellow denotes non-conserved residues. Sequence alignments were performed with ClustalW using MegAlign (DNASTAR).

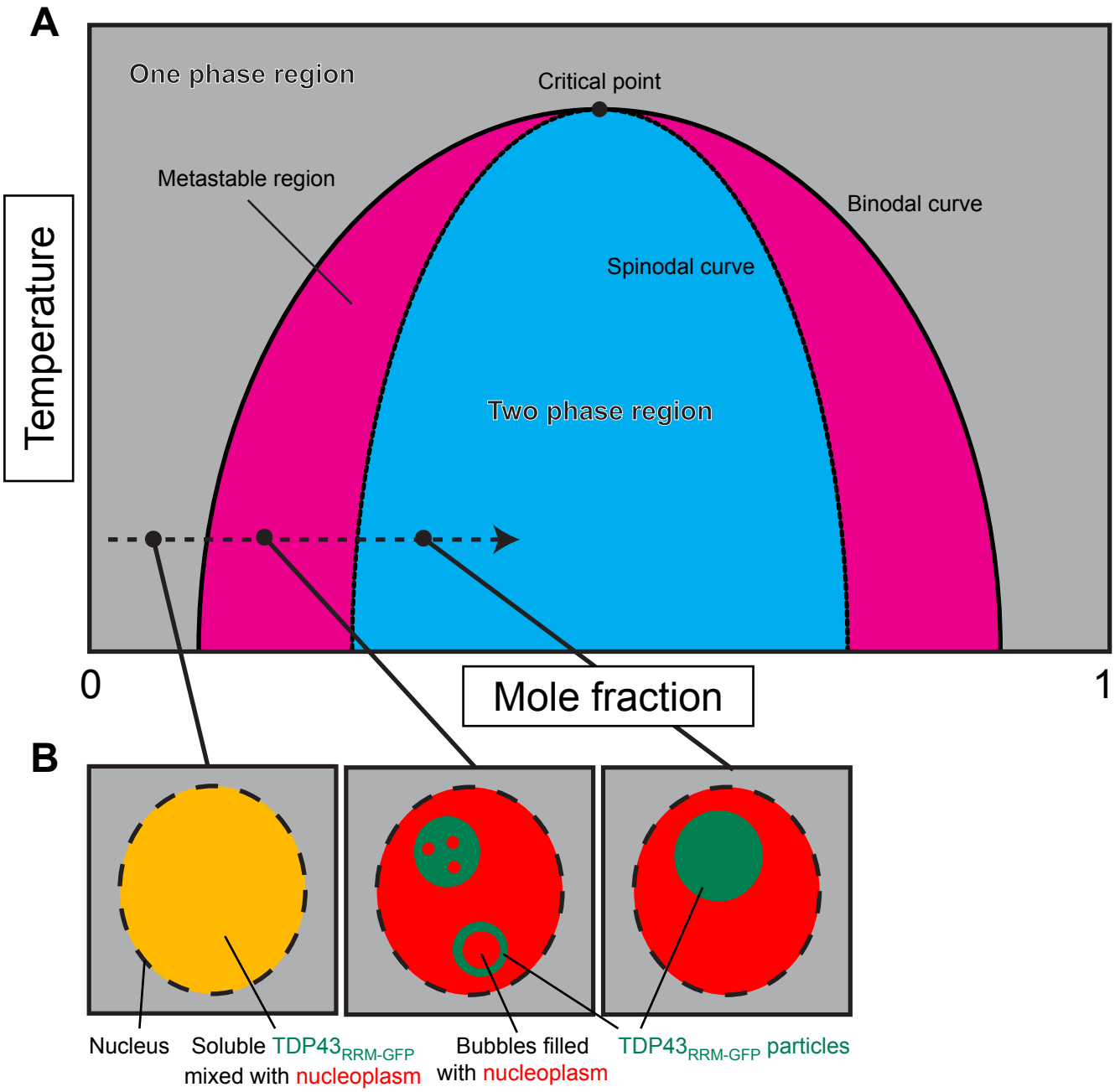
A
One phase region
Critical point
Metastable region
Binodal curve
Spinodal curve
Temperature
Two phase region
0
Mole fraction
1
B
Nucleus
Soluble TDP43RRM-GFP mixed with nucleoplasm
Bubbles filled with nucleoplasm
TDP43RRM-GFP particles

Figure S6

Figure S6. Schematic illustration of $TDP43_{RRM\text{-}GFP}$ phase separation and bubble formation. *Related to the Discussion.*

(**A**) General phase diagram depicting the dependence of polymer phase separation on temperature (on the ordinate) and concentration (expressed as mole fraction; on the abscissa) applied to $TDP43_{RRM\text{-}GFP}$. Nuclear $TDP43_{RRM\text{-}GFP}$ remains soluble in the one-phase region, but phase separates within the two-phase region. Note the metastable region between the binodal and spinodal curves, where phase separation occurs via nucleation. Spontaneous phase separation (spinodal decomposition) occurs only when the system is abruptly forced from the one-phase region into the boundaries of the spinodal curve, either by a sudden increase of $TDP43_{RRM\text{-}GFP}$ concentration or by super-cooling of a sufficiently concentrated $TDP43_{RRM\text{-}GFP}$ solution. See main text for details. (**B**) Schematic representation of expected phase morphologies for $TDP43_{RRM\text{-}GFP}$ particles at different points in the phase diagram assuming nucleation-based phase separation in the metastable region (see main text for detailed discussion).

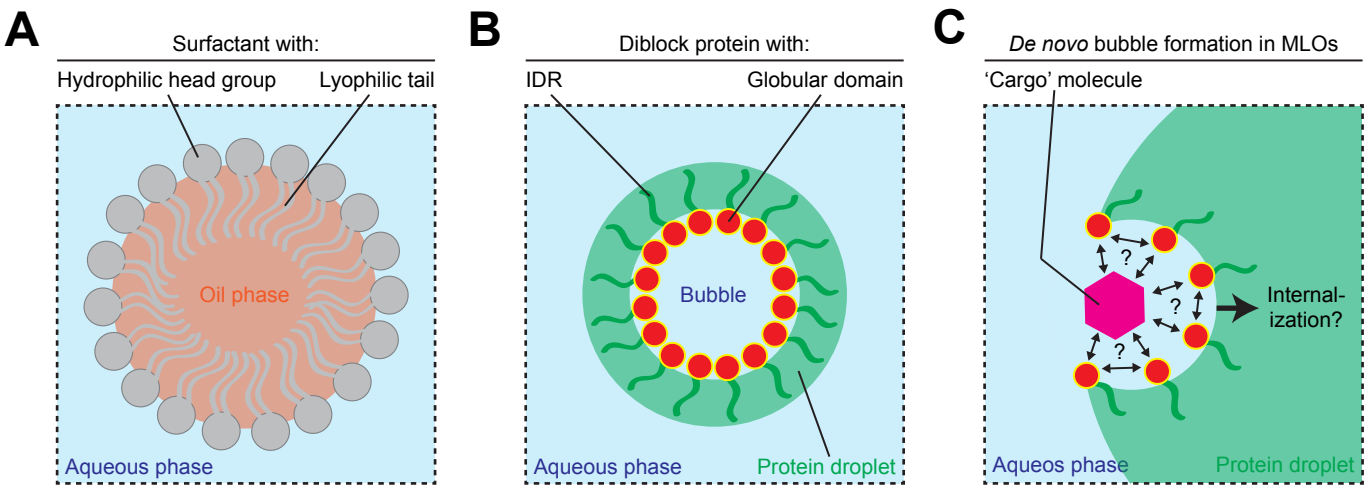
A
Surfactant with:
Hydrophilic head group
Lyophilic tail
Oil phase
Aqueous phase
B
Diblock protein with:
IDR
Globular domain
Bubble
Aqueous phase
Protein droplet
C
De novo bubble formation in MLOs
'Cargo' molecule
?
?
?
Internal-
ization?
Aqueos phase
Protein droplet

Figure S7

Figure S7. Models for the formation of bubble-containing multi-phase assemblies. *Related to the Discussion*.

(**A**) Surfactants such as detergents allow the formation of oil-in-water emulsions. (**B**) Diblock proteins comprised of intrinsically disordered and globular domains may act as analogues of surfactants and stabilize bubbles in MLOs. (**C**) Interactions between the globular domains (and cargoes) in the bubbles may contribute to drive bubble formation and stabilization. See main text for details.

Supplemental References

Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell *137*, 146-158.

Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E., and Webb, W. W. (1976). Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J *16*, 1055-1069.

Ayala, Y. M., Zago, P., D'Ambrogio, A., Xu, Y. F., Petrucelli, L., Buratti, E., and Baralle, F. E. (2008). Structural determinants of the cellular localization and shuttling of TDP-43. J Cell Sci *121*, 3778-3785.

Buratti, E., and Baralle, F. E. (2001). Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. J Biol Chem *276*, 36337-36343.

Burke, K. A., Janke, A. M., Rhine, C. L., and Fawzi, N. L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. Mol Cell *60*, 231-241.

Denning, D. P., and Rexach, M. F. (2007). Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. Mol Cell Proteomics *6*, 272-282.

Frey, S., and Görlich, D. (2007). A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. Cell *130*, 512-523.

Handwerger, K. E., Cordero, J. A., and Gall, J. G. (2005). Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure. Mol Biol Cell *16*, 202-211.

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T., and Miyawaki, A. (2013). A bilirubin-inducible fluorescent protein from eel muscle. Cell *153*, 1602-1611.

Lin, Y., Protter, D. S., Rosen, M. K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol Cell *60*, 208-219.

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Pesiridis, G. S., Lee, V. M., and Trojanowski, J. Q. (2009). Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. Hum Mol Genet *18*, R156-62.

Soumpasis, D. M. (1983). Theoretical analysis of fluorescence photobleaching recovery experiments. Biophys J *41*, 95-97.

Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J. D., Leigh, P. N., Al-Chalabi, A., Miller, C. C., Nicholson, G., and Shaw, C. E. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668-1672.

Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., and Lee, V. M. (2008). Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. J Biol Chem *283*, 13302-13309.