Organellomics: AI-driven deep organellar phenotyping reveals novel ALS mechanisms in human neurons

Sagy Krispin[1,2*], Welmoed van Zuiden[1,2*], Yehuda M. Danino[1,2*], Lena Molitor[1,2], Noam Rudberg[1,2], Chen Bar[1,2], Alyssa Coyne[3,4], Thea Meimoun[1,2], Fergal M Waldron[5,6], Jenna M Gregory[5,6], Tal Fisher[1], Aharon Nachshon[1], Noam Stern-Ginossar[1], Nancy Yacovzada[1,2#], Eran Hornstein[1,2$]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Israel.

[2]Department of Molecular Neuroscience, Weizmann Institute of Science, Israel.

[3]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[4]Brain Science Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[5]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK.

[6]Department of Pathology, NHS Grampian Tissue Biorepository, Aberdeen, UK.

* These authors contributed equally

Corresponding Author for machine learning: Nancyy@weizmann.ac.il

$ Corresponding Author for all other facets of the work: Eran.Hornstein@weizmann.ac.il

Abstract

Systematic assessment of organelle architectures, termed the organellome, offers valuable insights into cellular states and pathomechanisms, but remains largely uncharted. Here, we present a pipeline for deep phenotypic learning using vision transformers, resulting in the Neuronal Organellomics Vision Atlas (NOVA) model that studies 3 million confocal images of 25 distinct membrane-bound and membraneless organelles in human neurons. Our organellomics approach evaluates cellular phenotypes by quantifying changes in the localization and morphology of multiple organelles, and by consolidating information across organelles it yields an integrated depiction of cellular state. We reveal significant interactions between cytoplasmic mislocalized TDP-43 — a hallmark of ALS — and processing bodies (P-bodies), membraneless organelles that regulate mRNA stability, which is confirmed through patient-derived neurons and human neuropathology. Furthermore, organellomics delineates phenotypic changes in neurons carrying ALS-associated mutations and demonstrates diagnostic potential for patient-derived neurons. Together, organellomics offer a novel approach to studying the neuro-cellular biology of diseases.

Keywords: AI, computer vision, deep learning, machine learning, phenotypic learning, vision transformers, perturbations learning, contrastive learning, neurons, neurodegeneration, subcellular architecture, organelles, biomolecular condensates.

Introduction

Organellomics, the systematic study of subcellular organellar architecture (the organellome), may be a useful strategy to directly phenotype cells by their subcellular compartments. This is conceivable since the quantification of RNAs, lipids, metabolites and proteins that compose organelles cannot be directly translated into complex organelle properties. Furthermore, organelles are interconnected by functional crosstalk and inter-organelle interdependency [1,2], further supporting the value of organellomics-level insights for cell biology.

The analysis of multiple organelles [3] poses substantial challenges because of the variety of localizations and morphologies, which need to be integrated and evaluated in parallel. Systematic quantification and global evaluation of subcellular architecture has been developed using mass spectrometry-based approaches [4–8]. These do not provide insight into organellar morphology and depend on tedious organelle fractionation, which is limited to only a subset of the organelles. In addition, imaging-based approaches were developed [9–19]. However, these are largely in non-polarized cultured cells.

The study of cellular organization has been traditionally focused on membrane-restricted organelles, such as the nucleus and mitochondria. However, in recent years, certain biomolecular condensates have been identified as membraneless organelles with specialized functions [20–24]. Therefore, the study of both membrane-restricted and membraneless organelles is important for gaining a comprehensive view of cellular biology.

The dysfunction of organelles is a central driver of disease. Disorders of membrane-bound organelles such as mitochondria, lysosomes, or autophagosomes have been comprehensively characterized and are associated with distinctive clinical pictures [25–30]. In addition, compelling evidence suggests a role for membraneless organellar dysfunction in neurodegeneration [24,31,32].

We hypothesized that comprehensively capturing the localization and morphological patterns across the entire organellome would allow unbiased mapping of organellar responses to perturbations using a unified new scale. To test this hypothesis, we employed an image representation learning approach, particularly based on Vision Transformers (ViT [33,34]) perturbation learning [19,35–41], to capture high-level abstraction of morphological features from fluorescent confocal images. Deep representation learning enables automatic feature extraction, which is unbiased and does not require a pre-defined object or task. Hence it describes any localization pattern in the image, surpassing widely used approaches to extract objects from images based on a-priori assumptions [11,42,43].

ViTs have emerged as powerful models for analyzing microscopy images. Unlike traditional convolutional neural networks (CNNs), ViTs offer enhanced flexibility in capturing global relationships within images by processing the input as a sequence of image patches, leveraging self-attention mechanisms [33]. This capability is crucial for identifying small, yet meaningful, alterations in neuronal structure and organization that can arise from specific perturbations. This characteristic makes ViTs

well-suited for complex tasks such as deep phenotypic learning in diverse cellular environments, where quantification of spatial relationships of subtle organellar variations is critical.

Here, we introduce Neuronal Organellomics Vision Atlas (NOVA), a systematic and unbiased organellomics perturbation learning approach that is based on ViTs and contrastive learning [40]. The contrastive learning framework employed minimizes natural variation, forcing the model to focus on genuine perturbation effects through learning transitive relationships between samples. NOVA takes as input immunofluorescence images of a panel of cellular organelles in cultured cells and quantifies perturbational changes. NOVA was first pre-trained on massive amounts of OpenCell data [44] to learn to encode for subcellular localization patterns of 1,311 proteins. Then, the pre-trained model was fine-tuned for a perturbation learning task on data from highly polarized human neuronal cells with different chemical or disease-relevant genetic perturbations. NOVA does not require detection of the cellular borders (segmentation), overcoming a limitation in neuron image analysis, and is sensitive enough to depict organellome changes associated with single nucleotide genetic variants. NOVA enables quantitative mapping of organellar responses to perturbation, and its integrated view of the cell as a superposition of its constituent organelles represents a new phenotyping approach. Our organellomics approach offers a novel perspective on neuro-cellular disease biology by capturing cellular phenotypes through organelle localization and morphology. We reveal critical interactions between mislocalized cytoplasmic TDP-43—a hallmark of ALS—and P-bodies, which regulate mRNA stability, validated in patient-derived neurons. This approach offers foundation for future diagnostic advancements in neurobiology and proposes new biomarkers for human pathology.

Results

Organellomics: the study of 25 organelles in neurons

The complexity of the organellome necessitates a strategy that integrates unbiased, quantitative data about multiple organelles. Here, we devised a set of 25 key membrane-bound and membraneless organelles (**Figure 1A** and **Table 1**) and generated a comprehensive dataset composed of ~2.3 million images of human iPSC-derived neurons [45] under different experimental conditions (i.e., perturbations; **Table 2** (**Methods**) and **Supplementary Figure 1**). The image dataset of 25 organelles associated with the study was created in six batches, with two technical repeats per condition in each of the study batches. Images of neurons with specific dyes or immune-staining were captured by a spinning disc confocal microscope and subjected to QC and preprocessing steps, resulting in images showing 1.5 neurons on average (**Supplementary Figure 2** and **Supplementary Materials**). Data used for model training (0.5M (half a million) images), validation (0.12M images), and test (0.12M images) originated from two different batches. For robustness, inference was conducted solely on four additional held-out batches, each with two biological repeats (1.5M images, **Methods**).

ViT for encoding organelle localization patterns in perturbed human neurons

Vision Transformers (ViTs), with their self-attention mechanism, excel at modeling long-range dependencies across image patches, making them particularly effective for capturing fine-grained cellular details in microscopy images [33]. Their ability to accurately recognize intricate features is crucial for tasks involving high-resolution imaging and detailed structural analysis.

To examine organellar changes in neuronal cells upon disease-associated perturbations, we developed NOVA, a sequential fine-tuned model based on ViTs (**Figure 1B** and **Supplementary Figure 3**). NOVA is specifically designed to address perturbation learning tasks in the following way. First, we trained a ViT model with a multi-class classification head and cross-entropy loss on a publicly available dataset of 1,311 OpenCell proteins in human embryonic kidney cells that were cultured in basal growth conditions (HEK293T, 1,100,253 image tiles[44], **Supplementary Figure 4A**). By learning to predict the protein identity in the image, the model learns to encode a wide variety of organellar morphologies and localizations (referred here as topographies), in simple cells (**Supplementary Figure 4B)**.

We next extended the model to perform a perturbation learning task by continuing the training on images of neurons with perturbations (**Supplementary Figures 5, 6** and **Methods**). To preserve the knowledge learned from the localization encoding task, a partial fine-tuning step was performed using the angle-metric [46] approach on the model's layers (**Supplementary Figure 5** and **Methods**). The perturbation learning head and loss function were devised to perform contrastive loss, using Information Noise Contrastive Estimation (InfoNCE, **Figure 1B** and **Supplementary Figure 3**). InfoNCE loss encourages the model to implicitly learn to focus on similarities or dissimilarities that correspond to true perturbation effects. Hence, impactful perturbations will be pushed apart, diverging their feature representations away from the baseline, control condition. In parallel,

perturbations without effect will generate similar representations, ensuring that the model doesn't overfit irrelevant differences. Together NOVA performs two tasks: (1) the encoding of neuron images for feature representations, and (2) the detection of perturbations with actual organellar effect.

As an initial look into the neuronal organellome, images of unperturbed neurons were encoded by NOVA and the representations were used to create a Uniform Manifold Approximation and Projection (UMAP)[47] of the full landscape of neuronal organelles (**Figure 1C**, Adjusted Rand Index (ARI), 0.38). This map describes topographies, thus, clustering in proximity organelles with similarity in either localization or morphology, and is reproducible across batches (**Supplementary Figure 7**, ARI = 0.394).

Notably, some cytoplasmic organelles are in proximity. Mitochondrial outer membrane (MOM) and mitotracker, which accumulate in active mitochondria, are thus reassuringly observed together. The proximity of TDP-43 granules and paraspeckles was previously reported [48,49] and TDP-43 granules proximity to ANXA11 granules in the organellar map might align with the recent discovery of their co-aggregation in ALS [50]. NEMO (NF-κB essential modulator) granules are cytoplasmic membraneless organelles that are critical for NF-κB signaling, and their clustering at postsynapses suggests a potential association between NF-κB signaling and postsynaptic markers. Taken together, the projection of all organelles on the same manifold reveals distinctive organellar topographies, providing insights into neuronal subcellular organization.

We further benchmarked NOVA's organellar map (ARI=0.38, **Figure 1C**, and ARI=0.442, **Supplementary Figure 8A**) to organellar maps that were generated by extracting features using the Cytoself model [9] (ARI=0.31) or CellProfiler (ARI=0.246) [11,43] (**Supplementary Figure 8B, C**). The more discernible clusters demonstrate that NOVA appears to be well-suited for organellomics.

Stress granules (SGs) are cytoplasmic biomolecular condensates that are composed of RNA-binding proteins and RNAs. They are formed in response to stress conditions and are readily identifiable through specific immunostaining of the protein G3BP1. To evaluate the ability of NOVA to interpret the organellome to describe a specific cellular state (i.e., the extent of a stress response), we introduced sodium arsenite (0.25 mM, 30 minutes)[51,52] into human iPSC-derived neurons, which led to the assembly of visible SGs. Accordingly, NOVA correctly clustered neurons by the response to stress, using G3BP1 as an SG marker and FMRP, which is localized in neurons into RNA-associated granules ("FMRP granules"), including SGs under stress [53] (**Figure 1D**, ARI = 0.603, 0.875, respectively).

NOVA further revealed clear segregation of human osteosarcoma (U2OS) cells at basal growth conditions from U2OS with induced oxidative stress (0.4 mM sodium arsenite, 30 minutes) by images of stress granules (G3BP1) and also by P-bodies or actin cytoskeleton (**Supplementary Figure 9**), consistent with published works [54,55]. These images were taken with a Zeiss LSM 900 confocal. Therefore, NOVA can be effectively utilized on highly polarized neurons, but also beyond its initial training, on a previously unseen cell type (U2OS), captured by a different microscope.

Figure 1.

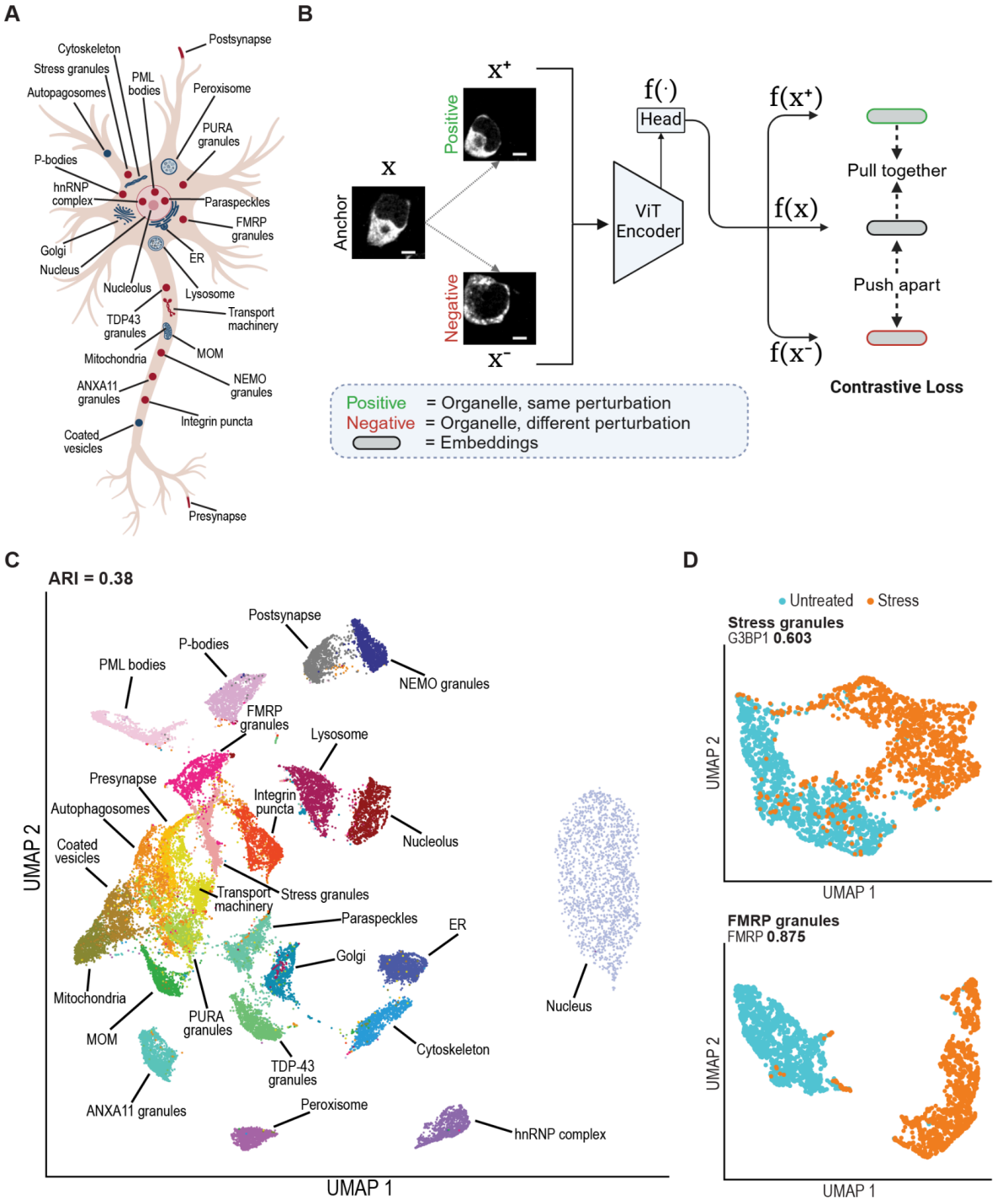

A
Cytoskeleton
Stress granules
Autopagosomes
PML bodies
Postsynapse
Peroxisome
PURA granules
P-bodies
hnRNP complex
Paraspeckles
FMRP granules
Golgi
Nucleus
ER
Nucleolus
Lysosome
TDP43 granules
Transport machinery
Mitochondria
MOM
ANXA11 granules
NEMO granules
Integrin puncta
Coated vesicles
Presynapse
B
Anchor
x
Positive
x⁺
Negative
x⁻
ViT Encoder
f(·)
Head
f(x⁺)
f(x)
f(x⁻)
Pull together
Push apart
Contrastive Loss
Positive = Organelle, same perturbation
Negative = Organelle, different perturbation
= Embeddings
C
ARI = 0.38
UMAP 2
UMAP 1
Postsynapse
P-bodies
PML bodies
NEMO granules
FMRP granules
Lysosome
Presynapse
Integrin puncta
Autophagosomes
Nucleolus
Coated vesicles
Transport machinery
Stress granules
Paraspeckles
ER
Golgi
Mitochondria
PURA granules
MOM
Cytoskeleton
TDP-43 granules
ANXA11 granules
Peroxisome
hnRNP complex
Nucleus
D
Untreated
Stress
Stress granules
G3BP1 0.603
UMAP 2
UMAP 1
FMRP granules
FMRP 0.875
UMAP 2
UMAP 1

Figure 1. Vision transformer-based perturbation learning model enables organellomic studies in neurons. (A) Illustration of the 25 organelles used for the organellomics study. **(B)** Illustration of the perturbation learning workflow that incorporates contrastive learning, whereby an input image (i.e., anchor) is encoded by 128 features in the latent space (i.e., embeddings). Each anchor image is randomly matched to other fields of view of the same organelle under the same perturbation (i.e., "positive", similar perturbational effects) and to five other fields of view of the same organelle but under different perturbations (i.e., "negatives", dissimilar or no perturbational effect). During training, the model converges to establish meaningful representations through contrastive loss: clustering positive pairs together in the feature space while pushing apart negative pairs based on the distinct differences learned from the data. Perturbations with no apparent effect result in feature representations similar to unperturbed cells. See **Supplementary Figure 3** and the **Methods** for details. **(C)** The full organellar landscape of day 8 human iPSC-derived neurons, depicting 25 organelles with distinct topographic patterns, under basal growth conditions. ARI = 0.38. **(D)** UMAP clustering of neuron cell topographies, in response to sodium arsenite stress, marked by stress granules (G3BP1, ARI=0.603), or FMRP granules (FMRP, ARI=0.875). ARI: Adjusted Rand Index.

Table 1. The 25 key organelles analyzed in this study, respective abbreviations and markers

Organelle	Abbr.	Marker	Relevant Literature
Nucleus		Hoechst 33342	[56]
Mitochondria		Mitotracker	[57]
Mitochondria Outer Membrane	MOM	TOMM20	[58]
Golgi Apparatus	Golgi	GM130	[59]
Endoplasmic Reticulum	ER	Calreticulin	[60]
Actin Cytoskeleton	Cytoskeleton	Phalloidin	[61]
Autophagosomes		SQSTM1	[62]
Lysosomes		LAMP1	[63]
Peroxisomes		PEX14	[64]
Coated Vesicles		CLTC	[65]
Integrin puncta		CD41	[66,67]
Nucleolus		NCL	[68]
Paraspeckles		NONO	[69,70]
TDP-43 granules		TDP-43	[71]
Heterogeneous Nuclear Ribonucleoprotein Complex	hnRNP complex	FUS	[72,73]
Promyelocytic Leukaemia Nuclear Bodies	PML bodies	PML	[74]
P-Bodies		DCP1A	[75]
Stress granules		G3BP1	[76]
ANXA11 granules		ANXA11	[77]
PURA granules		PURA	[78,79]
NEMO granules		NEMO	[80]
Microtubule-Associated Transport Machinery	Transport machinery	KIF5A	[81]
Presynaptic Terminals	Presynapse	SNCA	[82,83]
Post-Synaptic Subcompartments	Postsynapse	PSD95	[84]
FMRP granules		FMRP	[53]

A ranking system for quantification of the organellar response to perturbations

A comprehensive organellome analysis revealed the broad organellar topographies in response to sodium arsenite stress **(Figure 2A)**. Significant stress-related changes in human induced neurons were observed in SGs (G3BP1) (ARI = 0.603, **Figure 1D**) and FMRP granules (ARI = 0.875, **Figure 1D**) and also in TDP-43 granules (TDP-43) (ARI = 0.732, **Figure 2A**). In addition, PURA granules (PURA) effectively discriminated stressed from unstressed neurons (ARI = 0.618, **Figure 2A**), consistent with their association with SGs [78,79]. The topographies of mitochondria and its outer membrane (Mitotracker and TOMM20), and cytoskeleton (Phalloidin) were also changed substantially in response to stress (ARI = 0.645, 0.495 and 0.429, **Figure 2A**).

We sought means to quantify the organellomics changes across several batches and replicates. Therefore, we developed an orthogonal ranking technique, dubbed organelle scoring. Organelle scoring ranks organelles by their ability to discriminate between perturbations, computing the ARI of the full embeddings from different batches (**Supplementary Figure 10** and **Methods**). Organelle scoring of human neurons in response to stress revealed substantial alterations in the topography of stress granules (G3BP1, FMRP, and PURA) and other biomolecular condensates (TDP-43 granules and PML bodies) (p-value < 0.0001, **Figure 2B**). Reproducible stress-associated changes were also observed in the mitochondria, mitochondrial outer membrane, ANXA11 granules and integrin puncta (p-value < 0.0001); P-bodies and paraspeckles (p-value < 0.001); autophagosomes, peroxisomes and nucleus (p-value < 0.01), and the cytoskeleton (p-value < 0.05), across 6 experimental batches (**Figure 2B**).

Additionally, we compared organelle scoring by NOVA versus objects from CellProfiler. Analysis based on either NOVA or CellProfiler revealed changes to the mitochondrial outer membrane, mitochondria, PURA granules, cytoskeleton, peroxisomes, ANXA11 granules, FMRP granules, and P-bodies in neurons in response to stress. However, CellProfiler did not rank stress granules, marked by G3BP1, as significantly changing upon cellular stress (**Supplementary Figure 11**).

Altogether, we comprehensively characterize the broad changes that neuronal organelles undergo under stress using a single-scale system, thus reassuringly detecting known markers and suggesting new organelles affected by the stress response. Organelle scoring is a unified organellar comparison system that overcomes size heterogeneity and different biochemical properties or morphologies.

Figure 2

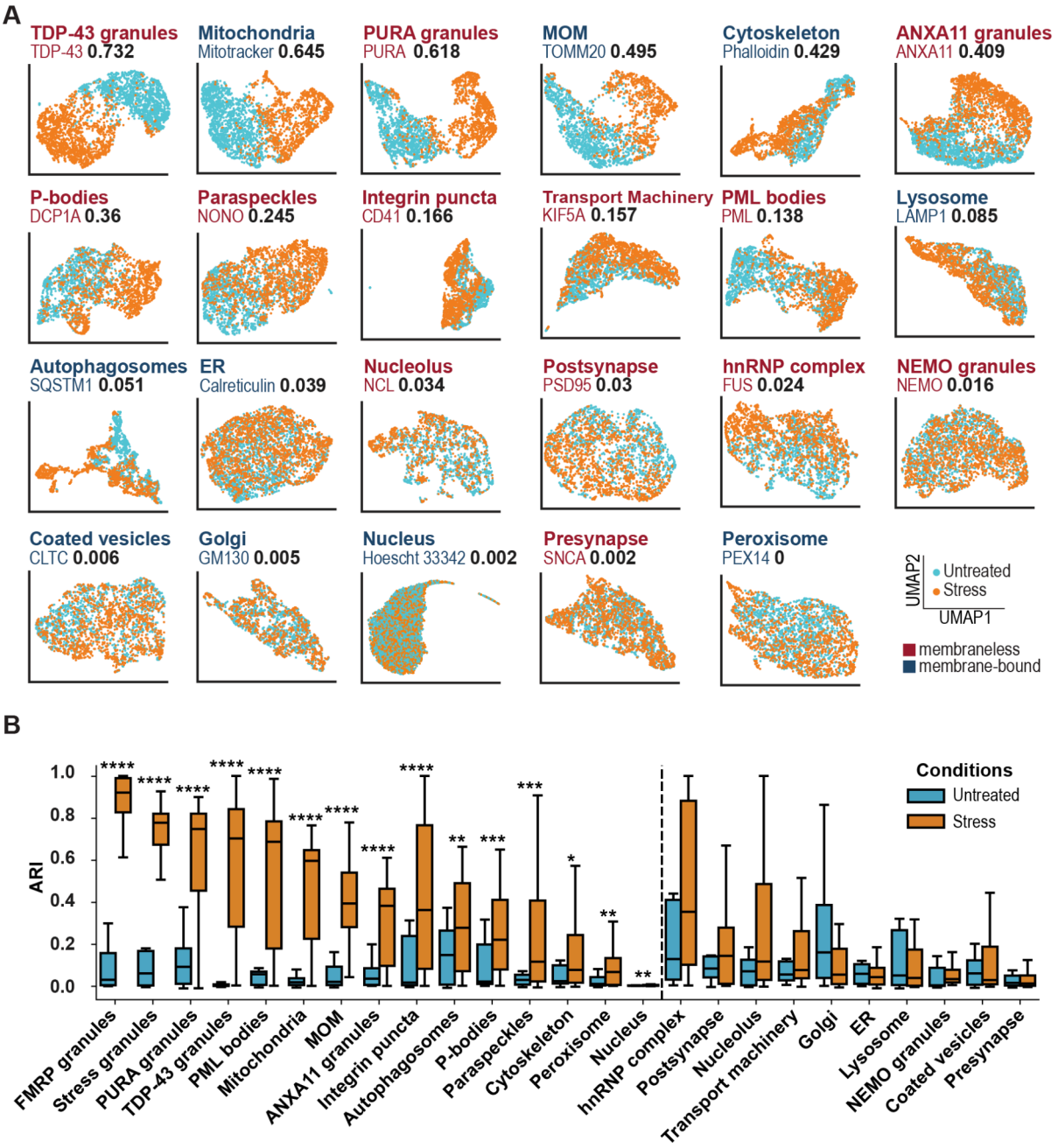


Figure 2. Neuronal organellar response to stress. (A) UMAPs of 23 organelles in human neurons in response to stress. Untreated (light blue) or sodium arsenite treated ("Stress"; orange). Membrane-bound (blue) or membraneless (red) organelles and ARI (black). Data for G3BP1 and FMRP organelles is in **Figure 1D**. Data from a representative batch. **(B)** Organelle scoring of the human neuronal response to stress (orange) vs. untreated (blue). ARI of organellar full latent feature space across 6 batches, ordered by the median of ARI values. Boxes indicate the interquartile range (IQR) and whiskers are drawn to the farthest datapoint within 1.5*IQR from the nearest hinge. Black

horizontal lines represent median values. P-values ****< 0.0001, ***< 0.001, ** < 0.01, *< 0.05; unpaired one sided t-test.

Cytoplasmic TDP-43 affects neuronal organellar dynamics and interacts with P-bodies

To investigate whether organellomics can generate novel mechanistic insights into disease states, we next focused on TDP-43 cytoplasmic malpositioning, which is a hallmark of human neuropathology in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) [85]. For this purpose, we studied human iPSC-derived neurons expressing doxycycline-inducible TDP-43 (TARDBP), which lacks a functional nuclear localization signal (NLS; TDP-43^{deltaNLS}) and is forced to the cytoplasm (**Figure 3A, B**). This construct is also fused with an engineered ascorbate peroxidase (APEX) [86].

We created a dataset composed of three imaging batches of ~300,000 image tiles post QC of TDP-43^{deltaNLS}-overexpressing neurons, stained for 25 organelles (**Table 1** and **Supplementary Materials**). Noteworthy, these batches were not included in the model training of NOVA and were used for inference only. As expected, we observed alterations in TDP-43 granule subcellular distribution (ARI = 0.497, **Figure 3C** and **Supplementary Figure 12A**). Organelle scoring of TDP-43 $\pm$ doxycycline (DOX) reassuringly revealed that the topography of TDP-43 was affected by the cytoplasmic mislocalization of TDP-43 granules (p-value < 0.0001, **Figure 3D**). Furthermore, P-bodies, ER, PML nuclear bodies, ANXA11 granules, peroxisomes, the presynapse, and MOM have also displayed significant topographic changes (p-value < 0.001, **Figure 3D** and **Supplementary Figure 12B**). In contrast, no discernible changes were identified in several other organelles, including stress granules (ranked 17th, 21st, and 23rd, out of 25, G3BP1, FMRP, and PURA, respectively). These findings suggest that cytoplasmic TDP-43 broadly affects the neuronal organellome and is largely independent of stress granules, aligning with [87], who observed that cytosolic TDP-43 assemblies operate separately from conventional stress granules.

We hypothesized that the significant change in P-body topography might be due to a direct interaction with TDP-43. Hence, we co-stained P-bodies and TDP-43 in TDP-43^{deltaNLS}-expressing neurons, and identified a significant increase in TDP-43 signal inside DCP1A-positive P-bodies, relative to the surrounding cytoplasm (**Figure 3E, F**). This observation was confirmed in a more disease-relevant context. We differentiated iPSCs from three C9orf72-ALS (C9ALS) patients, which exhibit increased cytoplasmic TDP-43, and six control subjects into day 60 motor neurons (**Supplementary Figure 13**). These 60-day-old patient iPSC-derived neurons showed a similar enrichment of TDP-43 signal in P-bodies compared to the healthy controls (p-value < 0.01, unpaired t-test, **Figure 3G**).

To test if known P-body proteins are predicted to physically interact with TDP-43, we used SpeedPPI, a machine-learning algorithm that predicts protein-protein interactions based on structure [88–90]. We analyzed the TDP-43 dimer, and the potential interaction of TDP-43 with 10 core P-body proteins (DCP1A, DCP2, LSM14A, DDX6, AGO1, AGO2, AGO3, YTHDC2, SMG7, EIF4ENIF1) and with 5 housekeeping genes, which are not enriched in P-bodies (VDAC1, RPAB1, GDIA, RL10A, TBA1A). Six of the P-body proteins showed hundreds of predicted contact sites with TDP-43 with high confidence

for protein-protein complex interactions (pDockQ scores above 0.23, [90]), of which three (AGO1, AGO2, AGO3) scored even higher than the predicted score for the TDP-43 dimer itself (1091 contact sites, pDockQ 0.44; **Supplementary Figure 14**). In contrast, all but one housekeeping protein had fewer contact sites compared to P-body proteins with low confidence (pDockQ scores below 0.23).

Furthermore, we directly tested if P-bodies physically interact with cytoplasmic TDP-43 using proximity labeling proteomics, which enables the characterization of cellular compartments at a nanometric radius. APEX facilitates the conjugation of radicalized biotin onto neighboring biomolecules upon exposure to hydrogen peroxide (H2O2) and can reveal weak or transient interactions that are often lost by other methods (e.g., immunoprecipitation), including within non-membrane-enclosed biomolecular condensates [86,91]. We first tested the specificity of the APEX fused to TDP-43^{deltaNLS} and confirmed by immunostaining that biotinylation was specifically observed only when both H2O2 and biotin phenol were present in the media, and that the biotinylation pattern overlapped with the construct localization (**Supplementary Figure 15**). Then, following proximity labeling, we pulled down the biotinylated proteins with streptavidin beads and performed targeted mass spectrometry (MS). We quantified the enrichment of 10 P-body proteins and five housekeeping proteins in the vicinity of TDP-43, relative to nuclear export signal (NES) that served as a non-specific cytoplasmic APEX labeling control. All P-body proteins were significantly enriched in TDP-43^{deltaNLS} samples, whereas the five housekeeping proteins were not significantly enriched (unpaired two-sided t-test; **Figure 3H**).

Finally, to verify the potential relevance of TDP-43-P-body interaction in ALS pathology, we observed colocalization between P-body markers DCP1A and LSM14A and cytoplasmic TDP-43 aggregates in human C9orf72-ALS (C9ALS) post-mortem motor cortex samples (white arrowheads, **Figure 3I**).

Taken together, these results demonstrate a physical interaction between mislocalized cytoplasmic TDP-43 and P-bodies that may be responsible for the altered P-body topography, suggesting a role for P-bodies in ALS and FTD that was not shown previously.

Figure 3

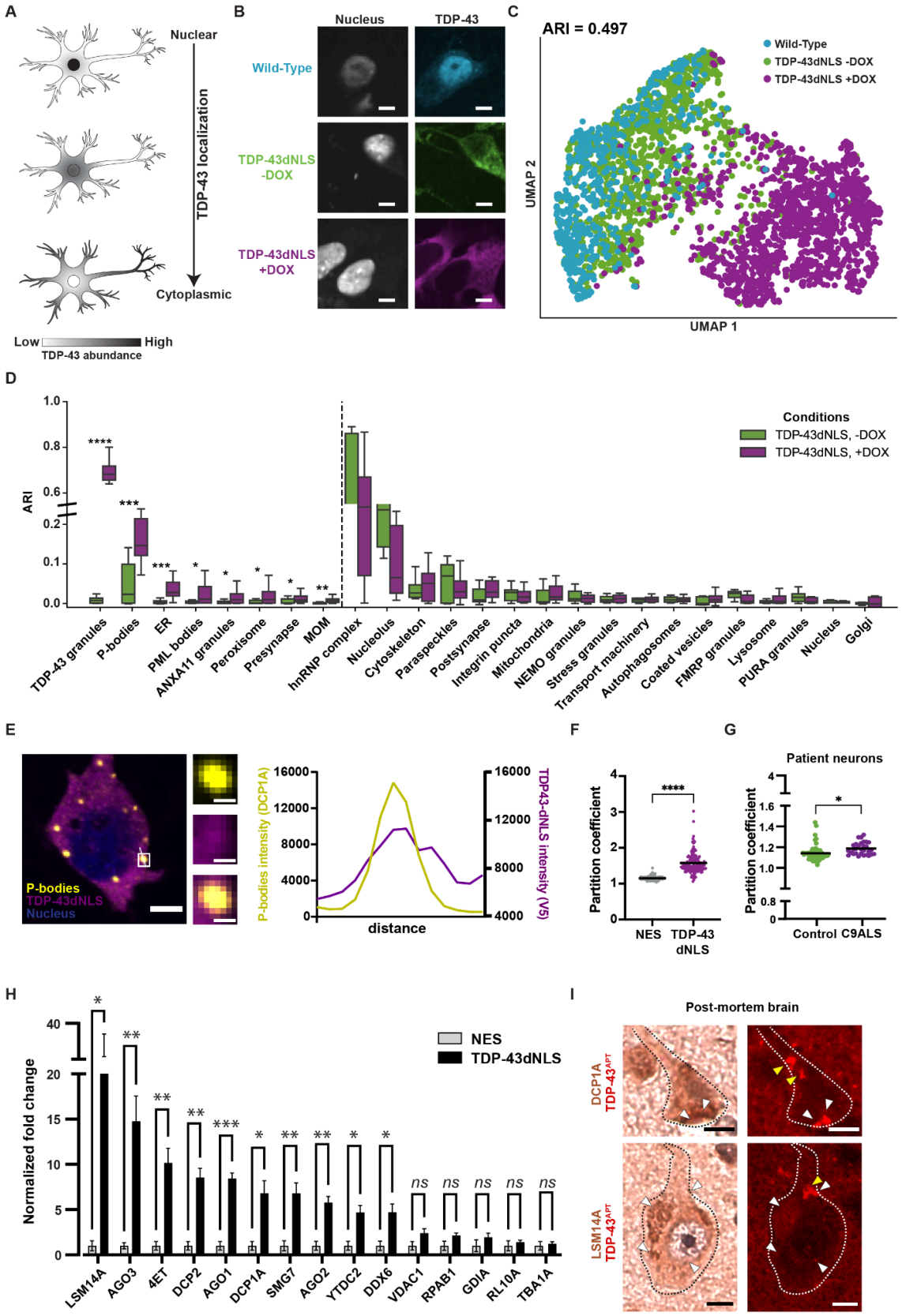

A
Nuclear
TDP-43 localization
Cytoplasmic
Low
High
TDP-43 abundance
B
Nucleus
TDP-43
Wild-Type
TDP-43dNLS -DOX
TDP-43dNLS +DOX
C
ARI = 0.497
Wild-Type
TDP-43dNLS -DOX
TDP-43dNLS +DOX
UMAP 2
UMAP 1
D
ARI
Conditions
TDP-43dNLS, -DOX
TDP-43dNLS, +DOX
TDP-43 granules
P-bodies
ER
PML bodies
ANXA11 granules
Peroxisome
Presynapse
MOM
hnRNP complex
Nucleolus
Cytoskeleton
Paraspeckles
Postsynapse
Integrin puncta
Mitochondria
NEMO granules
Stress granules
Transport machinery
Autophagosomes
Coated vesicles
FMRP granules
Lysosome
PURA granules
Nucleus
Golgi
E
P-bodies
TDP-43dNLS
Nucleus
P-bodies intensity (DCP1A)
TDP43-dNLS intensity (V5)
distance
F
Partition coefficient
NES
TDP-43 dNLS
G
Patient neurons
Partition coefficient
Control
C9ALS
H
Normalized fold change
NES
TDP-43dNLS
LSM14A
AGO3
4ET
DCP2
AGO1
DCP1A
SMG7
AGO2
YTDC2
DDX6
VDAC1
RPAB1
GDIA
RL10A
TBA1A
I
Post-mortem brain
DCP1A
TDP-43APT
LSM14A
TDP-43APT

Figure 3. The neuronal organellome reveals a physical interaction between cytoplasmic TDP-43 and P-bodies. (A) A diagram of cytoplasmic TDP-43 induction in human iPSC-derived neurons. **(B)** Representative confocal spinning-disk images of iPSC-derived neurons expressing endogenous TDP-43 (wild-type parental line), and a line expressing a doxycycline-inducible TDP-43 with a mutated nuclear localization signal (TDP-43^{deltaNLS}, uninduced -DOX; induced +DOX). Lens x63; Scale bar - 5 µm. **(C)** UMAP of TDP-43 granules in wild-type, and uninduced or induced TDP-43^{deltaNLS}. **(D)** Organelle scoring by ARI using the full latent space. The uninduced TDP-43^{deltaNLS} (-DOX) was considered the control. Three independent batches, two repeats each. Unpaired one-sided t-test. Black horizontal lines represent median values. **(E)** A representative image of the colocalization of induced cytoplasmic TDP-43^{deltaNLS} (V5 tag; purple) and P-bodies (DCP1A; yellow) in iPSC-derived neurons; lens x63; Scale bar - 5 µm. Insets - merged channels and channel separation of TDP-43^{deltaNLS} and a single P-body (bottom to top, respectively); scale bar - 1 µm, and intensity profile of TDP-43^{deltaNLS} (purple) and P-bodies (yellow). Partition coefficient analysis of **(F)** induced cytoplasmic TDP-43 signal within P-bodies, relative to nuclear export signal (NES) as control in 8-day-old iPSC-derived cortical neurons (n=130-180 images per group) or **(G)** endogenous TDP-43 signal within P-bodies in 60-day-old iPSC-derived motor neurons from C9orf72-ALS patients (C9ALS) expressing cytoplasmic TDP-43 relative to healthy control group (n=60-90 images per group). For **F** and **G**, data was analyzed by CellProfiler, unpaired two-sided t-test. Values are averages per image; Black horizontal lines represent mean values. **(H)** Normalized enrichment of 10 P-body proteins and five housekeeping proteins (left to right) in the vicinity of TDP-43^{deltaNLS}, relative to cytoplasmic control (NES). APEX-TDP-43^{deltaNLS} proximity labeling and targeted mass spectrometry. Unpaired two-sided t-test. P-values * < 0.05, ** < 0.01, *** < 0.001, **** <0.0001, ns = non-significant. **(I)** Representative images of human neurons from C9ALS cortex. TDP-43 detected by an aptamer (red) and P-body markers (DCP1A or LSM14A, upper or lower panel, respectively) detected by chromogenic immunohistochemistry. Aberrant cytoplasmic TDP-43 without (yellow arrowheads) or with (white arrowheads) co-localized P-body marker are shown. Scale bar - 10 µm.

Cytoplasmic TDP-43 alters P-body biophysics and function

We further hypothesized that mislocalized cytoplasmic TDP-43 affects P-body functions, which may be relevant for neuropathology. P-bodies are dynamic, phase-separated biomolecular condensates that are in constant interplay with other membraneless organelles, and their fluidity directly impacts cellular processes [92,93]. Therefore, we first tested the biophysical properties of P-bodies using partial fluorescence recovery after photobleaching (FRAP) in U2OS cells that constitutively overexpressed YFP-DCP1A and were transfected with inducible TDP-43^{deltaNLS}. Upon induction of cytoplasmic TDP-43, the internal diffusion of YFP-DCP1A molecules within P-bodies was delayed, relative to uninduced control (**Figure 4A** and **Supplementary Figure 16A, B**). Therefore, cytoplasmic TDP-43 impacts P-body liquidity. In addition, P-body size and numbers were reduced in iPSC-derived neurons expressing TDP-43^{deltaNLS} (CellProfiler, p-value < 0.0001, unpaired t-test; **Figure 4B, C**).

We also evaluated the effect of mislocalized TDP-43 on P-bodies in different model systems. We found that cytoplasmic TDP-43 affected P-body typical size and number also in U2OS cells

expressing TDP-43^{deltaNLS} fused to clover [87] (**Supplementary Figure 16C, D**). This was also observed in iPSC-derived neurons in which the cytoplasmic mislocalization of endogenous TDP-43 was driven by osmotic stress, induced by sorbitol [94] (0.4M, 2 hr; **Supplementary Figure 16E, F**). Consistently, P-bodies were smaller in cytoplasmic TDP-43 positive C9ALS neurons compared to control (p-value < 0.01, unpaired t-test; **Figure 4D**). However, the number of P-bodies was increased (p-value < 0.001, unpaired t-test; **Figure 4E**). Finally, we report significant differences in the size and numbers of puncta that are positive for either P-bodies protein DCP1A or LSM14A (**Figure 3I, 4F**) in human C9ALS post-mortem motor cortex, relative to non-neurological controls (Mann-Whitney U test, p-value < 0.05, **Figure 4G-J**). Intriguingly, in addition we observed large intra-neuronal and extracellular LSM14A-positive bodies in C9ALS patient neuropathology that were almost completely absent from control tissues (Mann-Whitney U test, p-value < 0.05, **Figure 4K**). Thus, TDP-43 pathology in C9ALS brains is associated with changes in the distribution of markers typical of P-bodies.

Although P-body function is still highly debated, P-bodies seem to play a role in RNA storage and decay [95]. Thus, we assessed the changes to RNA half-life in response to TDP-43^{deltaNLS} expression in neurons using the incorporation of 4-thiouridine (4sU; 200 µM) into newly transcribed RNA molecules (2 or 3 hours labeling pulse), followed by SLAM-seq and GRAND-SLAM analysis [96–98]. Induction of TDP-43^{deltaNLS} caused a 10% reduction in global RNA half-life, relative to uninduced control (p-value < 0.0001, unpaired t-test; **Figure 4L**). This reduction in mRNA stability increased to 30% when analyzing mRNAs that are enriched inside P-bodies [99] (p-value < 0.0001, unpaired t-test; **Figure 4M**). In contrast, there was a milder 5% decrease in the half-life of RNAs not specifically enriched in P-bodies (p-value < 0.01, unpaired t-test; **Figure 4N**. Therefore, accelerated mRNA decay is caused by cytoplasmic TDP-43 malposition and correlates to mRNAs that reside in P-bodies. Together, cytoplasmic TDP-43 directly interacts with P-bodies, and disrupts their biophysical properties and functions (**Figure 4O**). This suggests a potential role of P-bodies in the molecular mechanisms underlying TDP-43 proteinopathy and neurodegeneration.

Figure 4.

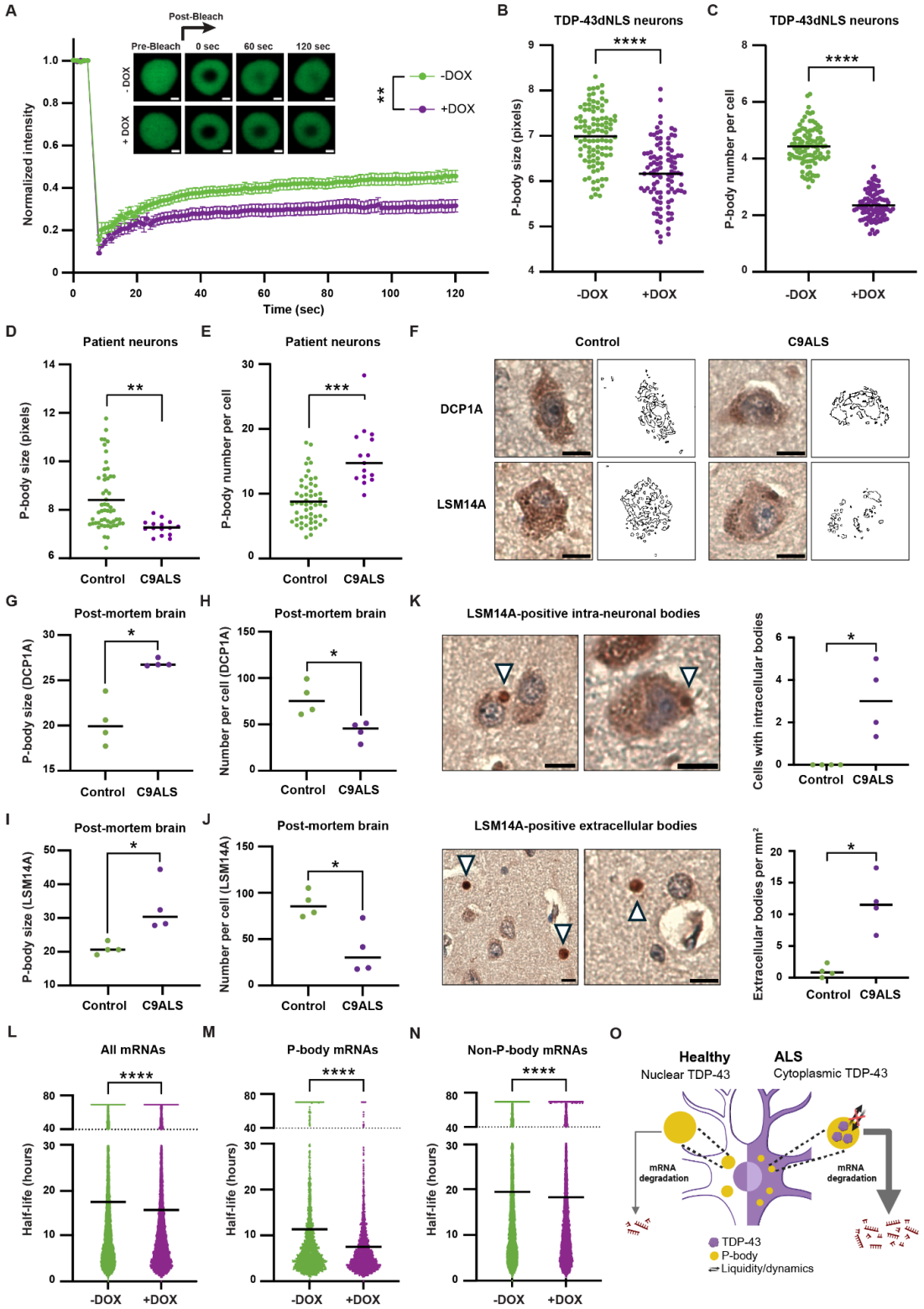

A
Post-Bleach
Pre-Bleach
0 sec
60 sec
120 sec
- DOX
+ DOX
Normalized intensity
Time (sec)
-DOX
+DOX
B
TDP-43dNLS neurons
P-body size (pixels)
C
TDP-43dNLS neurons
P-body number per cell
D
Patient neurons
P-body size (pixels)
E
Patient neurons
P-body number per cell
F
Control
C9ALS
DCP1A
LSM14A
G
Post-mortem brain
P-body size (DCP1A)
H
Post-mortem brain
Number per cell (DCP1A)
K
LSM14A-positive intra-neuronal bodies
Cells with intracellular bodies
I
Post-mortem brain
P-body size (LSM14A)
J
Post-mortem brain
Number per cell (LSM14A)
LSM14A-positive extracellular bodies
Extracellular bodies per mm²
L
All mRNAs
Half-life (hours)
M
P-body mRNAs
N
Non-P-body mRNAs
O
Healthy
Nuclear TDP-43
ALS
Cytoplasmic TDP-43
mRNA degradation
TDP-43
P-body
Liquidity/dynamics

Figure 4. Cytoplasmic TDP-43 disrupts P-body function. (A) Representative P-body image series (scale bar - 1 µm) and quantification of YFP-DCP1A fluorescence in P-body of U2OS cells upon recovery after partial photobleaching (FRAP). Uninduced (-DOX) or induced (+DOX) cytoplasmic TDP-43 (TDP-43^{deltaNLS}). Mixed-effects model with Geisser-Greenhouse correction. P-body **(B)** size or **(C)** number per cell, without or with induction of (DOX) cytoplasmic TDP-43 in 8-day-old iPSC-derived cortical neurons. CellProfiler quantification of n=100 images/condition, unpaired t-test. P-body **(D)** size or **(E)** number per cell, in 60-day-old iPSC-derived neurons from non-ALS subjects (control), versus C9orf72-ALS patients expressing cytoplasmic TDP-43 (C9ALS). P-body number is normalized to the cell number per image. P-body size measured as average pixels per image. n=50 images/group. Data points in **B-E** represent means per image. Unpaired t-test. **(F)** Representative DAB chromogenic immunohistochemistry images of P-body proteins, DCP1A or LSM14A, in C9orf72-ALS post-mortem motor cortex and an age and sex matched non-neurological control and quantification of **(G,I)**, median puncta size or **(H,J)** numbers per cell based on DCP1A or LSM14A staining respectively. Mann-Whitney U test. P-value * < 0.05. Lines represent median values. Scale bar - 10 µm. **(K)** Representative images from C9orf72-ALS post-mortem motor cortex stained with LSM14A demonstrating large intra-neuronal (top panel) and extracellular (bottom panel) LSM14A-positive bodies (white arrowheads), and corresponding quantification. Mann-Whitney U test. P-value * < 0.05. Lines represent median values. Scale bar - 10 µm. mRNA half-life in hours measured by SLAM-seq of **(L)** all mRNAs, **(M)** mRNAs enriched in P-bodies or **(N)** RNAs not enriched in P-bodies, as reported by [99]. Unpaired t-test. P-value ** <0.01; *** <0.001; **** <0.0001. Black horizontal lines in panels **C-I** represent mean values. **(O)** A proposed mechanistic model showing the interaction between cytoplasmic TDP-43 and P-bodies in ALS, and its consequences for P-body properties and function.

An integrated organellome view of human neurons

We hypothesized that an integrated organellome view that takes into account information from all 25 organelles can provide a novel and holistic map of cellular phenotypes. To address this hypothesis, we developed "synthetic superposition", an *in-silico* technique that combines NOVA's representations of all organelles, per perturbation, into a single data point (**Figure 5A** and **Methods**). The integrated organellome view is technically gained by randomly selecting an image tile for each of the 25 organelles from the same perturbation, and concatenating their NOVA embeddings.

We initially explored the effect of sodium arsenite stress or cytoplasmic mislocalization of TDP-43 on the integrated organellome, and found an ARI of 1.0, or above 0.85, respectively (**Supplementary Figure 17A, B**). UMAP of the integrated view is superior to the clustering gained by individual organelles (**Figure 2A, 3C**), suggesting that NOVA's embeddings can indeed be superpositioned to gain an integrated view of organellar responses to perturbations.

An integrated organellome of ALS-associated single-nucleotide variants in neurons

We next examined the integrated organellome of genetically-perturbed iPSC-derived neurons from the iPSC Neurodegenerative Disease Initiative (iNDI) lines [100], which are useful tools for research and allow the controlled comparison of different neurodegeneration-associated mutations

on a single isogenic background. We compared the effect of four ALS-associated mutations to the ancestral isogenic control line (Knockout Laboratory F2.1 Jackson, KOLF, wild-type). These lines include ALS alleles of two RNA-binding proteins, FUS (Fused In Sarcoma) R495* [101], TDP-43 (TAR DNA/RNA-Binding Protein 43) M337V [102,103] as well as OPTN (Optineurin) Q398E [104] and TBK1 (TANK Binding Kinase 1) E696K [105], which potentially alter the secretory pathway or autophagy. In addition, as specificity analysis, we included Parkinson's disease and Lewy body dementia-associated SNCA (Alpha-Synuclein) E46K allele [106] (**Supplementary Figure 18A**). When superimposing the embeddings from all 25 organelles, the integrated organellome of the neurodegeneration-associated lines was dissimilar to that of the wild-type. In contrast, remarkably, not all ALS-associated lines were clustered together (ARI = 0.962, **Figure 5B**), and this was reproducible in another batch (ARI = 0.65, **Supplementary Figure 18B**). Within the ALS lines, TBK1 E696K was the most different from the wild-type, whereas TDP-43 M337V and OPTN Q398E remained more proximal to the wild-type, suggesting that their organellome is less affected (**Figure 5B**). Additionally, SNCA E46K formed a distinct cluster, suggesting different alterations in the organellome in Parkinson's disease / Lewy body dementia (**Figure 5B**).

To ensure that the act of concatenating embeddings from the same group (i.e., perturbation) does not inherently differentiate groups *per se*, we applied a sensitivity analysis of the synthetic superposition by randomly concatenating organelles from a mixture of perturbations into the same data point. The resulting UMAP did not assemble any clusters (ARI = 0.01, **Supplementary Figure 18C**), supporting the notion that information accumulates via concatenation of embeddings, and that clusters authentically emerge from the perturbations they represent. In an additional specificity assessment, the SNCA mutant line was excluded and UMAP visualization of the integrated organellome resulted in similar clustering of the ALS-associated lines, suggesting that the effect (depicted in **Figure 5B**) is not driven by SNCA (ARI = 0.858 and 0.99, **Supplementary Figure 18D, E**). These data demonstrate the potential utility of the integrated organellome in exploring disease-relevant genetic perturbations.

The observation that ALS lines formed distinctive clusters led us to test whether different ALS mutations have distinct effects on organellar topographies. Thus, the cellular changes across all perturbations and organelles were quantified using a one-sided unpaired t-test and Cliff's delta effect size analysis [107] (**Methods**). This analysis showed many significant changes relative to wild-type, across six batches, despite some batch-to-batch variability (**Figure 5C**). In line with the integrated organellomic analysis (**Figure 5B**), FUS R495* and TBK1 E696K displayed alterations in many organelles. Some organelles, such as TDP-43 granules, were affected across all lines (ARI = 0.133, **Figure 5C**). This is noteworthy, as it occurs despite the absence of visible mislocalization of TDP-43 in these young neuronal cultures. In contrast, many organelles were altered only in specific lines. For example, coated vesicles were significantly affected by the TBK1 and OPTN mutations, but only minimally impacted by TDP-43 M337V (ARI = 0.447, **Figure 5C**). Furthermore, mutations in FUS, TBK1, and OPTN impaired vesicular trafficking-associated ANXA11 granules. This might be interesting to explore in the context of ALS-associated ANXA11 mutations, ANXA11-dependent co-trafficking of membrane-bound and membraneless organelles in neurons, and co-aggregation of ANXA11 with TDP-43 in type C inclusions [50,77,108] (ARI = 0.712). Interestingly, P-bodies were altered in TDP-43, TBK1,

and FUS mutant neurons (ARI = 0.201), suggesting a link between P-bodies and ALS beyond cytoplasmic TDP-43 mislocalization. Notably, autophagosomes were one of the few organelles that remained unaffected by most mutations including OPTN Q398E (ARI = 0.09).

One of the strongest effects was observed in the hnRNP complex (FUS marker) in the FUS R495* mutant line (ARI = 1.0, **Figure 5C**). When superimposing 24 organelles and excluding the hnRNP complex, the FUS mutant line clustered closer to wild-type (ARI = 0.643 and 0.946, **Supplementary Figure 19A, B**), suggesting that the hnRNP complex is a predominant organelle affected by the FUS R495* mutation.

We explored the FUS R495* in more detail, using homozygous, heterozygous, and genetically corrected (revertant) FUS R495* lines that restored the wild-type sequence. Integrated organellomic analysis of 8-day-old neurons revealed that FUS R495* mutant lines clustered closely together and far from wild-type, with the revertant line forming a separate cluster (ARI = 1, **Supplementary Figure 19C**). Similar results were obtained in 18-day-old neurons where the FUS revertant organellome converged more clearly towards wild-type (ARI = 0.87 and 0.939, **Supplementary Figure 19D, E**).

We demonstrated many organellar changes in the FUS R495* homozygous and heterozygous lines compared to wild-type, with the homozygous line displaying the strongest phenotype (**Supplementary Figure 19F, G**). Unexpectedly, many significant changes remained present in the revertant, especially in 8-day-old neurons, although the change of the hnRNP complex itself was diminished compared to the wild-type line (marked by FUS). This suggests that the genetic correction of the mutation results in an incomplete recovery at the cellular phenotypic level.

In summary, by computing the contribution of all organelles to a phenotype, an integrated view of neuron cellular biology that can identify and define disease phenotypes can be produced. Using this method, we have mapped novel organellar changes across ALS-associated variants, in premature and mature isogenic neurons.

An integrated organellome of patient-derived C9orf72 and sporadic ALS neurons

TDP-43 pathology is present in approximately 97% of ALS cases. It can be modeled in some of the patient-derived iPSC-induced neurons after several weeks in culture including in some sporadic cases in which the underlying disease-causing genes are unknown, and in cells that carry risk-associated hexanucleotide repeat expansions in the C9orf72 gene [109].

We tested NOVA's utility on a nuanced and clinically relevant application of identifying neuronal phenotypes as a proxy for disease diagnosis (ALS vs. healthy). This study further allowed us to examine the performance of NOVA on patient-derived neurons that were not previously seen by the model. Thus, we analyzed images of 60-day-old iPSC-derived motor neurons (**Supplementary Figure 13** and **Supplementary Figure 20A**) from healthy controls (n=6, 3 males, 3 females), C9orf72 repeat expansion carriers (C9ALS, n=3, 2 males, 1 female) and sporadic ALS (sALS) patient iPSC lines that either express nuclear TDP-43 (n=2, 1 male, 1 female) or pathological cytoplasmic TDP-43 (n=10, 6 males, 4 females; **Supplementary Figure 20A**). Neurons were stained for TDP-43 granules (TDP-43),

P-bodies (DCP1A), nucleus (Hoechst 33342), and the microtubule cytoskeleton (MAP2, marker previously unseen by the model). Integrated study via superpositioning of these four markers resulted in clustering per patient group (ARI = 0.35) that was significantly superior to the clustering gained by individual organelles (**Figure 5D** and **Supplementary Figure 20B**). The decision to focus on these organelles was made to also enable investigation of the aberrant TDP-43-P-body crosstalk in ALS (**Figure 3, 4**). Inspection of each patient group individually revealed consistent organellar differences that surpassed patient-to-patient variability and the impact of gender effect that was a limiting factor in transcriptomics-based analysis [110] (**Figure 5D**). This indicates that disease-related organellar changes are more prominent than individual variation. The analysis confirmed significant changes in organelles and each of the patient groups compared to controls, with C9ALS exhibiting the strongest phenotype (**Supplementary Figure 20C**).

Therefore, the NOVA model is capable of performing robustly on mature human neurons, with a variety of genetic backgrounds, even when the image capture setup differs significantly from the original training conditions. This includes variations such as using a non-spinning disc confocal microscope, independent experimental handling, day-60 sporadic neurons differentiated through a small-molecule protocol, and by introduction of organelle markers not previously seen by the model. Therefore, by detecting patient-derived diseased neurons that come from a different source and were not included in NOVA's training data, we demonstrate out-of-the-distribution task performance, substantiate the generalizability, and offer potential clinical value of our approach in delineating cellular states. Altogether, our data demonstrate that NOVA is broadly applicable as a tool for image analysis in neuroscience and can facilitate neurodegeneration-focused organellomic studies in human neurons.

Figure 5.

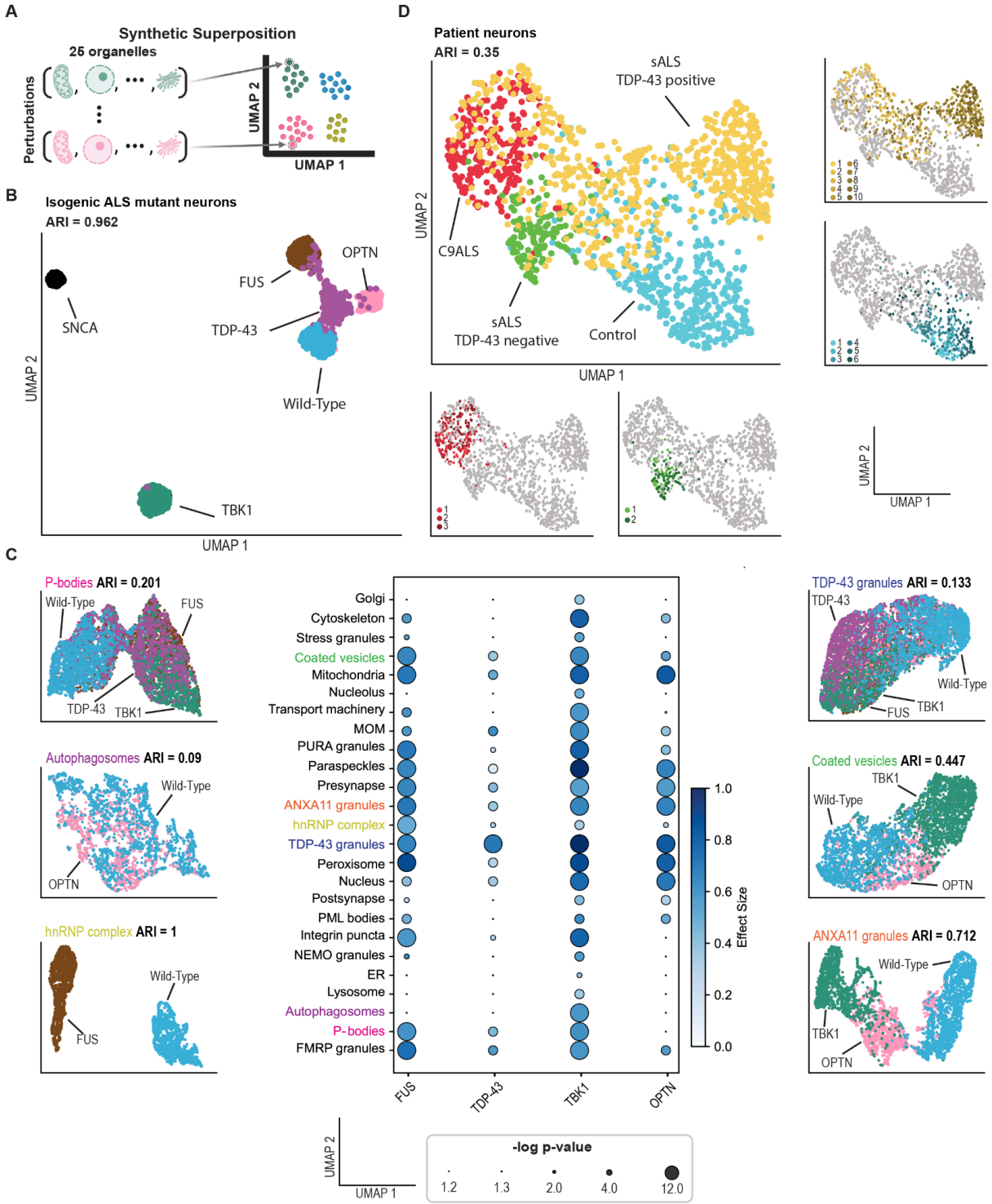


Figure 5. Integrated organellome characterization of human ALS neurons. (A) Illustration of the integrated organellome gained by synthetic superposition of 25 organelles. **(B)** UMAP of integrated organellome in wild-type (KOLF) and isogenic lines with ALS-associated mutations, FUS R495*, TDP-43 M337V, OPTN Q398E, TBK1 E696K, and a Parkinson's disease model line, SNCA E46K (ARI =

0.962). **(C)** Bubble plot of organellar changes, relative to wild-type (rows), across the ALS lines (columns). Data of 6 different batches. Bubble size - statistical significance (-$\log_{10}$ p-value; one-sided unpaired t-test; values of 1.3 or lower are not significant). Bubble color-effect size (Cliff's Delta, **Methods**) and selected UMAPs of specific organellar changes across genetic variants from representative batches with ARI values. **(D)** UMAP of synthetic superposition of 4 organelles (Nucleus (Hoechst 44432), TDP-43 granules (TDP-43), P-bodies (DCP1A), and Cytoskeleton (MAP2)) in integrated organellome view of 60-day-old iPSC-derived healthy controls motoneurons (blue), C9orf72 ALS (red), and sporadic ALS (sALS) without (green) or with (yellow) cytoplasmic TDP-43 mislocalization. ARI = 0.35. Color hues visualization of individual patients by genotype or phenotype groups.

Discussion

In this work, we present NOVA, a deep learning-driven approach for organellomic studies in neurons. We developed a ViT architecture with supervised contrastive loss to systematically encode perturbational subcellular localization and morphology patterns from immunofluorescent images of 25 membrane-bound and membraneless organelles (**Figure 1**). NOVA allows for characterization of the organellar response to perturbations on a single metric scale including in response to oxidative stress (**Figure 2**) and reveals that cytoplasmic TDP-43, the main hallmark of ALS, alters P-body biophysics and function (**Figure 3, 4**). Furthermore, combining the information from all organelles provides an integrated organellome view of human neurons under various perturbations (**Figure 5**). Utilizing ALS-associated and ALS patient neurons we pinpointed numerous specific organellar changes, thereby demonstrating both unique and shared alterations across different ALS genetics. Organellomics also enabled successful diagnostic separation of ALS patient-derived and control neurons, surpassing inter-patient heterogeneity and sex differences [110].

Organellomics is a new omics approach for high-content comparison of any number of organelles with different subcellular topographies. The relevance of organellomic studies is rooted in the need to understand the complex regulation of cellular functions via the network of membrane-bound and membraneless organelles [77,111–118]. Our image-based organellomics approach provides a compelling alternative to traditional biochemical methods of organellar fractionation [4–8], addressing the inherent difficulties of organelle fractionation, particularly in attempting the isolation of liquid-like membraneless organelles [119]. Whereas the dysfunction of membraneless organelles [24,31] and membrane-bound organelles [120–122] is suggested to contribute to neurodegeneration, our approach provides unprecedented depth to organellar studies and an integrative view of cell biology.

NOVA's ability to study perturbed cultured human neurons offers several advantages. First, the model was pre-trained on OpenCell data of 1,311 proteins (non perturbed), learning to encode localization from a publicly available dataset. This pre-training step optimized the model to predict organelle identity from image embeddings, promoting generalizability and reducing training time for the main task of perturbational learning. Next, we extended the model through partial fine-tuning using the angle metric, selectively freezing certain layers. During training on approximately 820,000 images of perturbed cultured human neurons, the loss function was replaced with a contrastive loss that included negative samples, specifically tailored for the perturbation learning task.

Furthermore, we developed organelle scoring, which ranks the response to perturbations, enables direct comparisons across all studied organelles, highlights inter-batch differences, and provides an effect size to prioritize further studies. Additionally, quantifying the perturbational changes across all organelles reveals the complete cellular response in an unprecedented way, paving the way for evaluating organelle coordination and potential interdependencies.

Another key component introduced is the integrated organellome view, which offers a unique approach to representing the cellular state as a vector composed of all studied cellular compartments. This is computed by concatenating randomly matched embeddings (each containing 128 features) of multiple organelles under the same perturbation. This process is performed *in silico*

and is not constrained by the limited number of channels available on standard light microscopes. Instead, NOVA utilizes simple confocal images, offering a cost-effective alternative to more complex multiplexed imaging technologies, such as PhenoCycler or 4i [123,124]. Notably, we demonstrate that this technique can be applied retroactively to existing microscopy datasets without any limitation on the number of fluorescent channels.

Finally, methods that generate self-explainable manual engineered features (intensity, shape, granularity, size for a given object), such as CellPainting and CellProfiler, are widely used [42,43]. However, some markers are not easily demarcated by defined objects and computer vision studies have demonstrated that multi-scale features automatically learned by a deep learning model can capture and generalize better than hand-crafted features [125–128]. Accordingly, NOVA's deep representation learning does not require pre-defined objects or tasks and can describe any localization pattern in the image.

NOVA's generalizability is demonstrated by its ability to analyze data from various cell lines including those not part of the training set (e.g., U2OS, TDP-43^{deltaNLS}, patient-derived motor neurons), across different maturation stages (day 8, day 18, and day 60 human neurons), and using data from different microscopes and experimentalists from different labs. Additionally, we were able to perform inference on a marker not included in the training set (MAP2, Microtubule-Associated Protein 2, part of the microtubule cytoskeleton). NOVA was also tested on a clinically relevant out-of-distribution task: identifying neuronal phenotypes for disease diagnosis (ALS vs. healthy controls). The analysis revealed consistent organellar differences within different patient groups, with disease-related changes outweighing individual variability. Therefore, we propose that NOVA provides a solid foundation for further AI studies in neurobiology. Moreover, our extensive in-house dataset, consisting of thousands of microscopic images of neurons under different perturbations, represents a valuable resource for future investigations.

We mapped the cellular stress response across the entire organellome for the first time. This revealed that oxidative stress induces not only stress granule formation, but also topographic changes in additional organelles including nuclear structures like PML bodies and paraspeckles, and membrane-bound organelles like mitochondria. While changes in individual organelles have been reported [129,130], our results highlight the involvement of diverse organelles in the stress response simultaneously, with the largest topographic changes occurring primarily in membraneless organelles.

A significant alteration in P-body topography was one of the effects mostly associated with cytoplasmic localization of TDP-43, in line with previous work in rat hippocampal neurons [131]. Our results suggest that TDP-43-P-body interactions may be largely independent of stress granules, which did not change in response to the expression of cytoplasmic TDP-43. This is intriguing given the reported interactions between P-bodies and SGs [114] and the suggested role of SGs in ALS [132]. TDP-43-P-body interactions were orthogonally supported by proximity labeling and mass spectrometry, which demonstrated a direct enrichment of several established P-body proteins in the vicinity of cytoplasmic TDP-43. This interaction of TDP-43 granules and P-bodies was previously unreported and might be an important facet of TDP-43 proteinopathies, including ALS.

Additionally, cytoplasmic TDP-43 seems to impact P-body size and numbers including in patient-derived day 60 motor neurons with TDP-43 pathology and human C9orf72-ALS motor cortex. Consistently, cytoplasmic TDP-43 modified P-body's liquidity, potentially impairing their ability to compartmentalize and regulate RNA processing [21,133,134]. Accordingly, we discovered that cytoplasmic TDP-43 increased the decay of mRNAs that are specifically enriched in P-bodies. Since we demonstrate that P-bodies are altered in ALS genetic models and patient-derived neurons, P-bodies may emerge as a potential therapeutic target.

The organellomic landscape of ALS-affected human cortical and motor neurons is reported here for the first time, providing a variety of novel hypotheses regarding organellar dysfunction in ALS. For example, TBK1 E696K exhibited a very broad and unexpected organellar phenotype. In contrast, TDP-43 M337V drove only modest changes, consistent with available literature that ALS-associated TDP-43 mutations do not disrupt its localization in premature iPSC-derived early culture (day 8) neurons [135,136].

In addition, the altered ANXA11 granule topography, shown in FUS R495*, OPTN Q398E, and TBK1 E696K lines, is interesting in light of ALS genetics [108,137] and its recently discovered co-aggregation with TDP-43 in type C inclusions [50]. Furthermore, P-bodies were altered in several lines, which speaks to our novel insights into their crosstalk with TDP-43 and potential insufficiency in disease. Unexpectedly, the autophagosome, ER, Golgi, and lysosome were not among the most affected organelles when mutant lines harbored variants in OPTN, which has been reported to alter the secretory pathways, cause Golgi fragmentation, dysregulate autophagosomes and enhance ER stress [138,139]. Likewise, mutations in TBK1 did not affect the autophagosome, as could be expected from past studies [140]. These discrepancies might be due to the premature age of neurons or specific growth conditions and hence could generate new hypotheses and showcase the strength of analyzing all organelles simultaneously.

FUS R495* is deficient in the nuclear localization signal and is clinically associated with an aggressive, juvenile form of ALS in which FUS is depleted from the nucleus and undergoes aberrant phase separation [101,141–143]. We analyzed the FUS R495* cortical neurons and a revertant line, in which the arginine at position 495 is restored through a second cycle of genome editing. The FUS revertant organellome was not akin to the wild-type in day 8 neurons but converged towards the wild-type in day 18 neurons. The partial organellar dysregulation in the FUS revertant line might be due to passage in culture or CRISPR off-target effects.

Finally, the Parkinson’s disease SNCA E46K line diverged from both wild-type and ALS-associated lines, demonstrating NOVA’s broader utility and revealing a distinct organellar complexity across neurodegenerative states.

We acknowledge that while the lack of segmentation makes NOVA very efficient, it is limited to studies in homogeneous cultures. Further development of segmentation in NOVA and improved preprocessing may enable the study of co-cultures, multiple genotypes, or even tissue sections. Additionally, the challenge of model explainability limits the translation of the model's observations

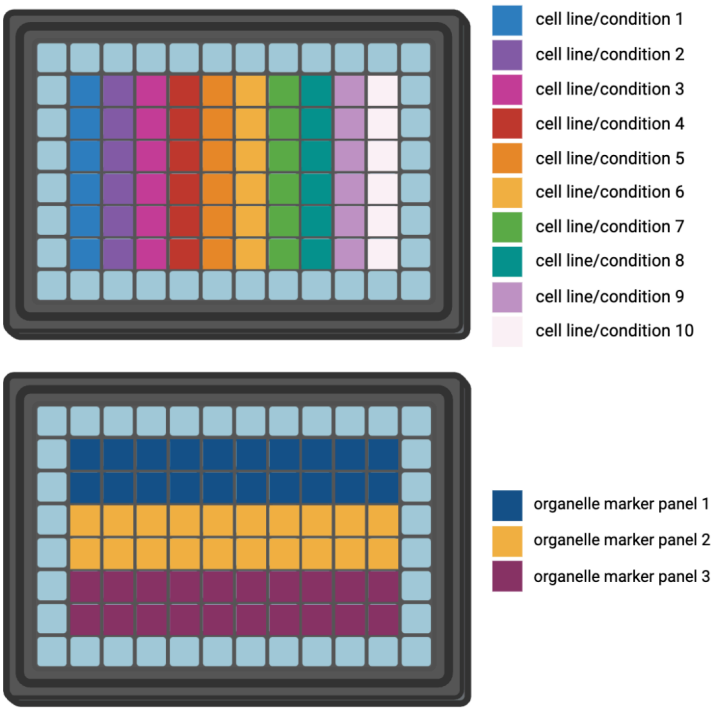
into meaningful interpretations regarding changes in topography, an aspect that can be improved in future work.

Moreover, optimizing the combination of marker panels and the utility of antibodies in 384-well plates will allow for simultaneous studies of an even higher number of organellar markers under a larger number of perturbations. Future studies could examine other organelle-associated phenotypes, such as organellar contact sites and functions. Such potential studies could provide further insights into organellomics.

Finally, the model is currently applicable to iPSC-derived cortical and motor neurons, but we expect it can be generalized to other disease-relevant cell types, such as microglia or astrocytes.

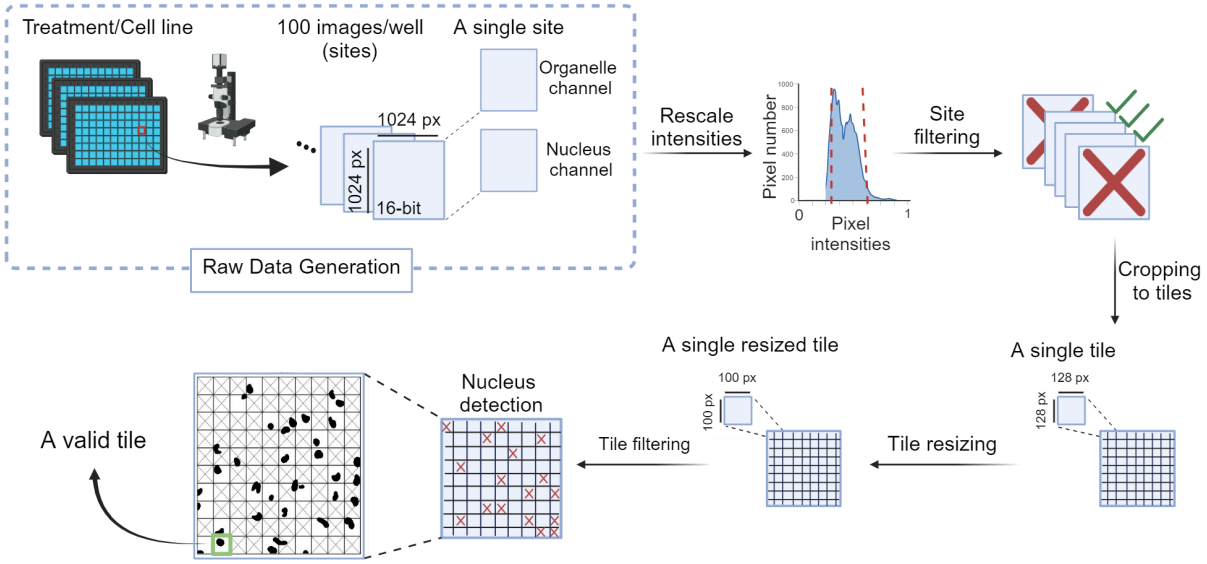
In summary, organellomics is a novel approach to understanding cellular biology. It can be extended to additional disease phenotypes, other types of patient-derived cells, and existing imaging data. Organellomics has the potential to generate novel hypotheses in cellular and disease biology research and may foster diagnosis and precision medicine for neurological diseases.

Supplementary Figures



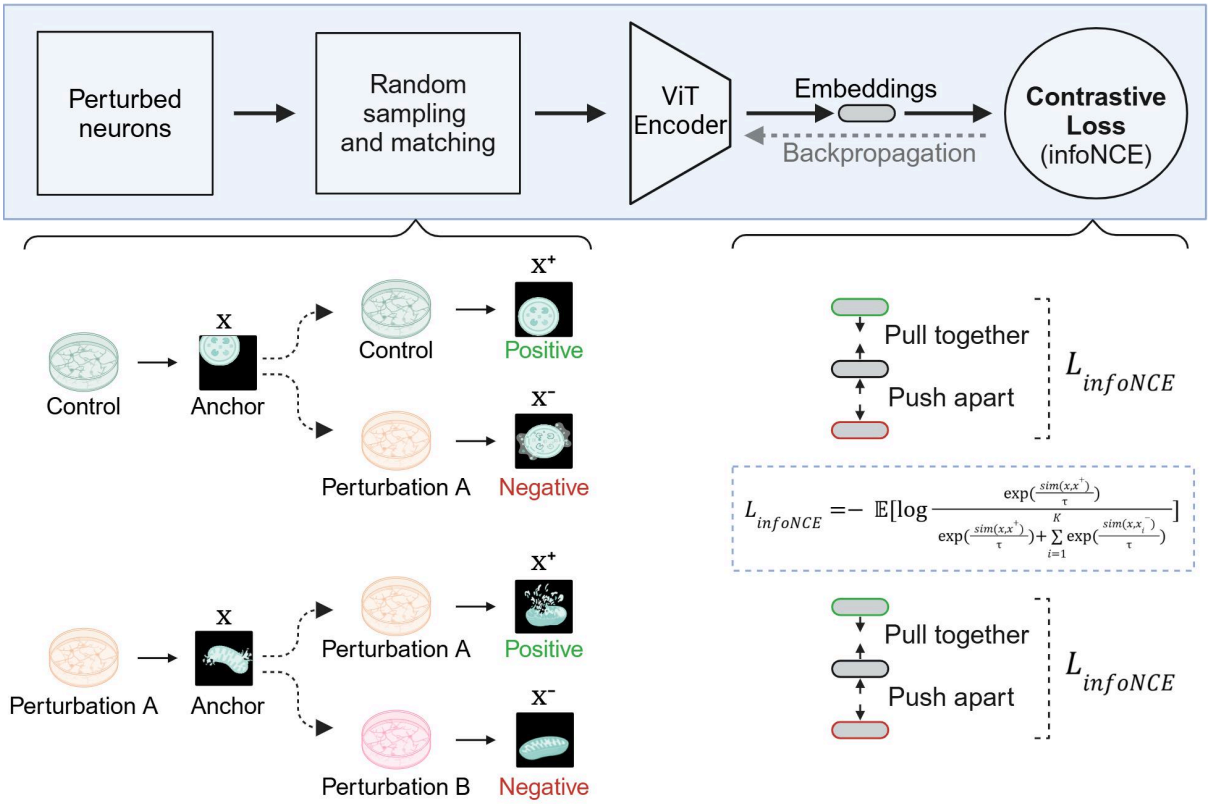
Supplementary Figure 1. Experimental setup of iPSC cortical neurons used for model training and inference.

In our experimental design, each 96-well plate included 10 cell lines/conditions and 3 marker panels. A total of 60 wells are used per plate. A full experimental batch for studying 25 organelles is generated using 11 marker panels across 4 different plates (see **Table 1**). Each well contained a distinct cell line or experimental condition (top panel). In each well, a specific marker panel with a set of antibodies targeting 2-4 organelles is used. Per batch, we image two wells (replicates) per marker panel (bottom panel). An exemplary plate design from a batch is depicted; another batch would be from the same cell differentiation but on a different plate.



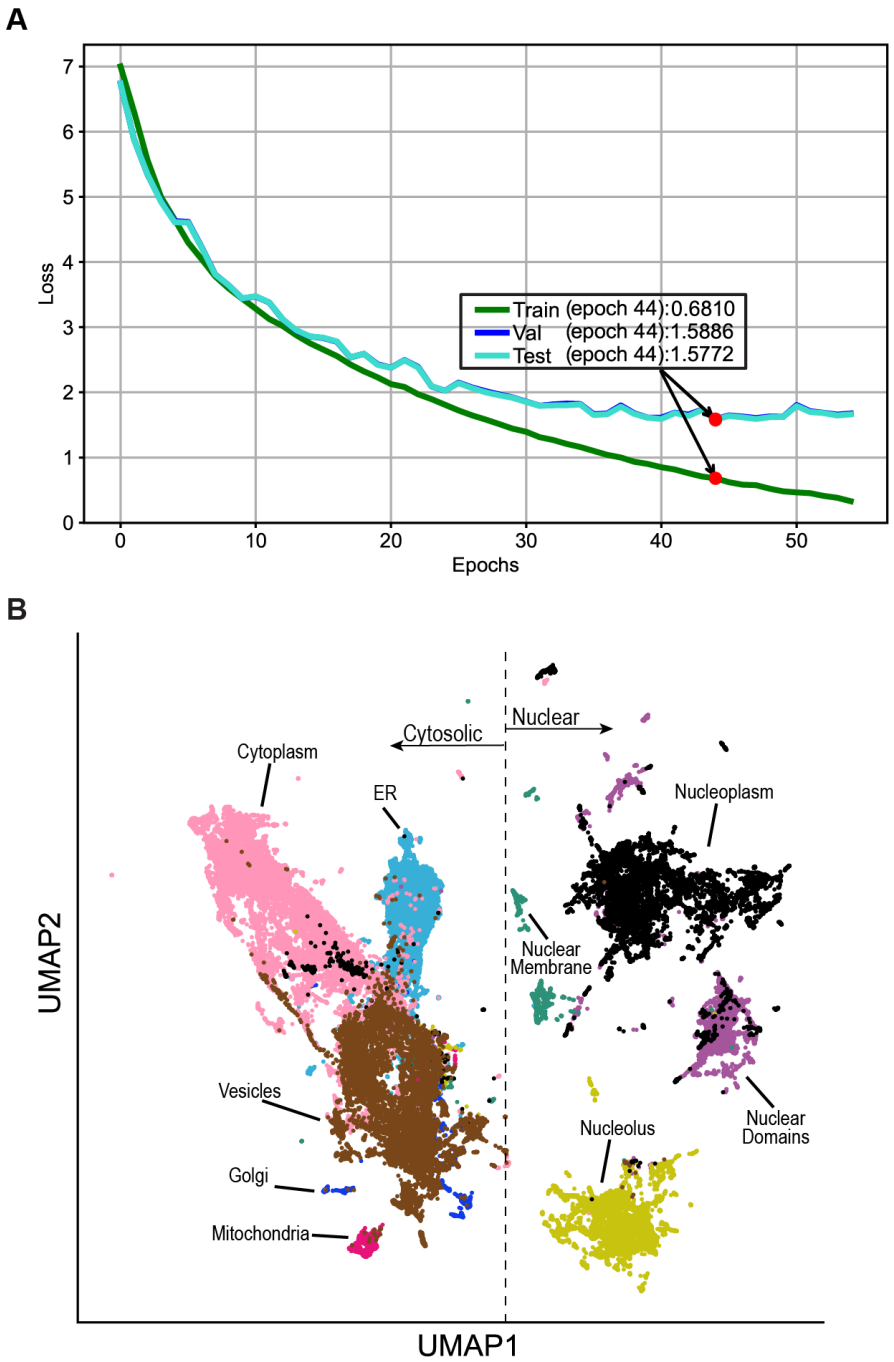
Supplementary Figure 2. Image pre-processing.

Microscopy images are of defined fluorescent emission frequency, stained for a single organelle. Data collected at 16-bit, 1024x1024 pixels referred to as "site images", is being rescaled and tested for autofocus quality (**Methods**). All organelle images are coupled to the cognate nucleus channel. Valid site images are cropped to fixed-size tiles of 128x128 pixels and resized into 100x100 pixels. Automated cell number pre-processing is performed on the nucleus channel and tiles with at least one cell nucleus but not more than five are considered valid tiles for analyses.



Supplementary Figure 3. Detailed illustration of the perturbation learning workflow with contrastive learning.

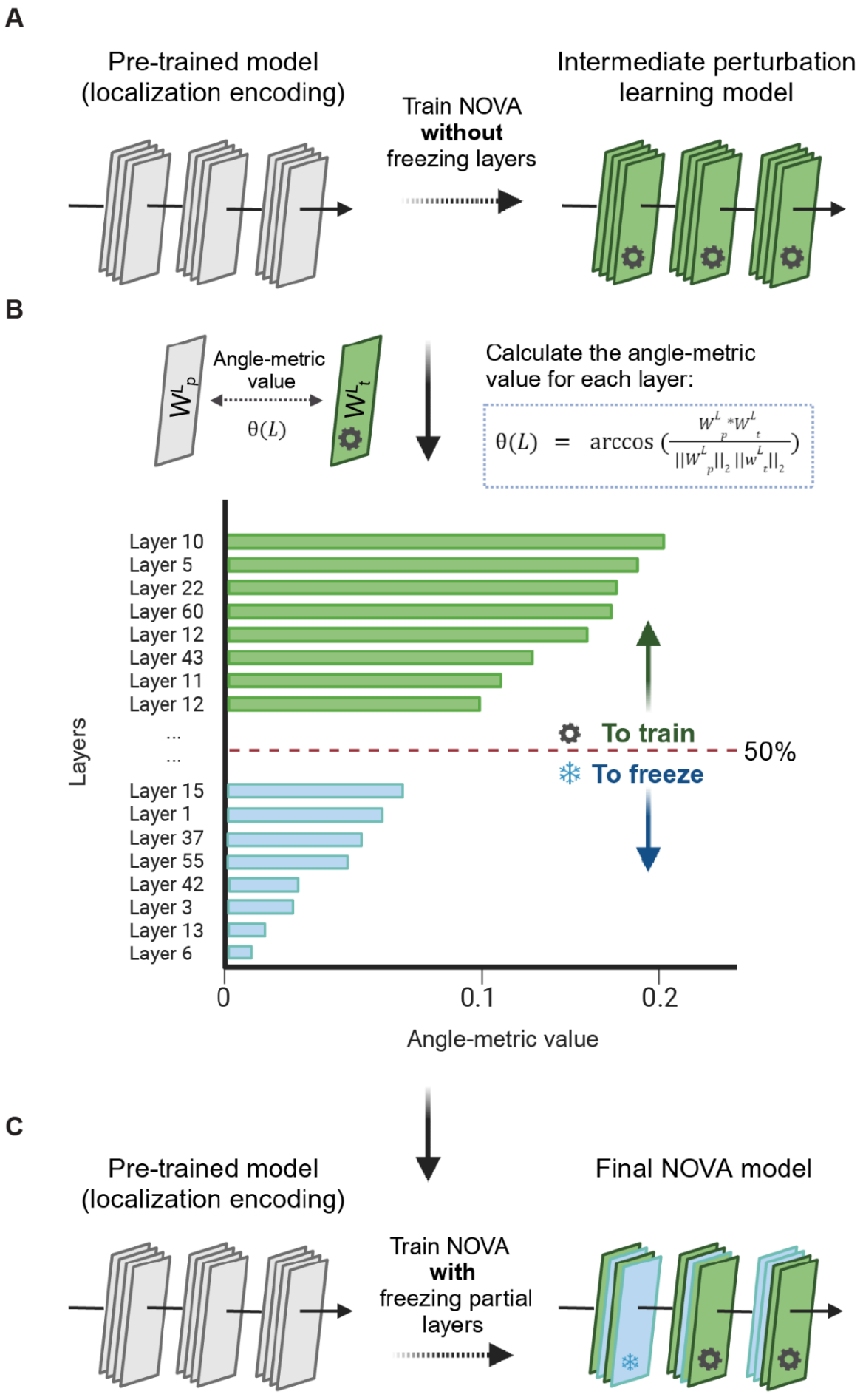
For each image in the training set, which serves as an "anchor", we randomly sample (with repetitions) six matching image tiles: one positive and five negative samples. Positives are tiles with the same organelle and perturbation as the anchor. Negatives are tiles of the same organelle, however with a different perturbation. The model then calculates the embeddings for the 7 samples and applies the infoNCE contrastive loss function. Contrastive learning allows the model to learn structural latent space and infer the relationship between the samples. It pulls together in embedding space similar images (anchor and positive) while pushing apart dissimilar images (negative). For additional information see **Methods**.



Supplementary Figure 4. ViT-based protein classification model is trained to study organelle localization in human embryonic kidney cells.

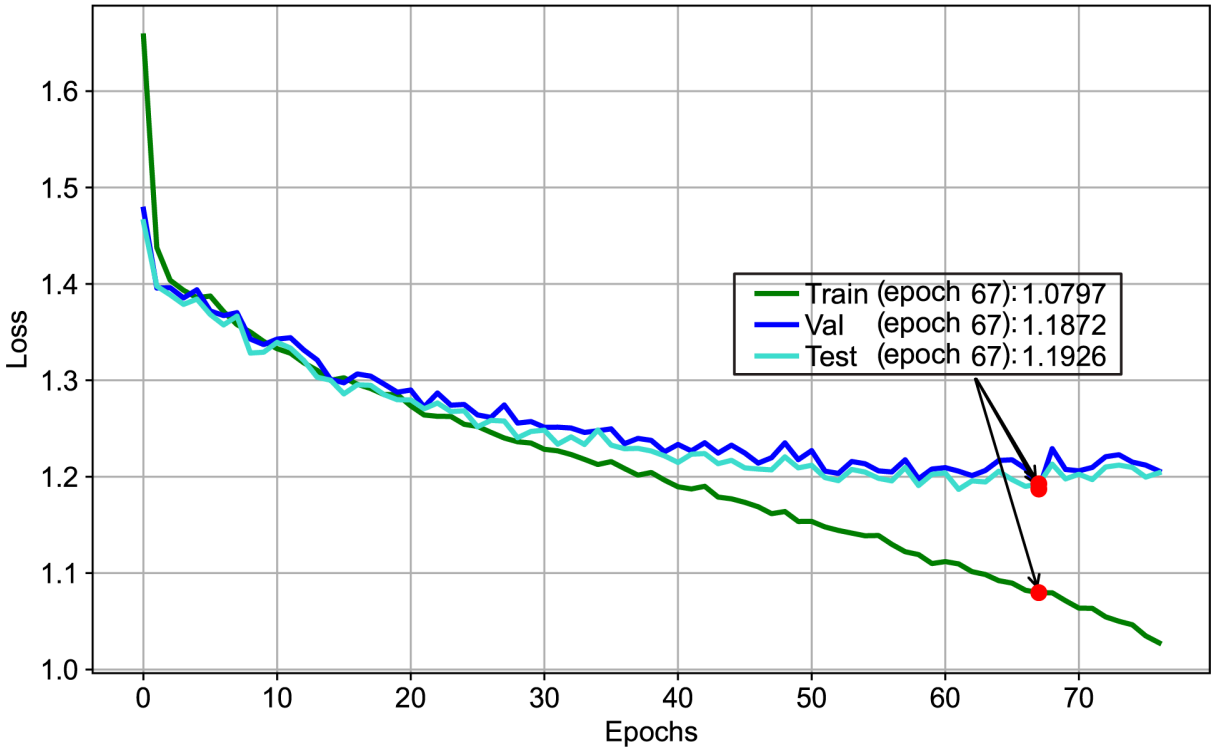
(A) The training curves of the model for encoding organelles localization patterns on the OpenCell dataset (train, val (validation), and test). Loss value of the model (y-axis) and the number of epochs (x-axis). Epoch 44, demarcated by a red dot, serves as the final model since it is the last epoch in the model's training process with an improvement in the loss function on the validation data. Black box with exact loss values. Model's classification accuracy on the test set of 59.42% (see **Methods**). **(B)**

UMAP of Opencell's test set (held-out data), colored according to protein localization categories, shows distinct organelle clustering. Only proteins with a clear and exclusive localization pattern are depicted. The list of annotated proteins is indicated in the **Supplementary Materials** (OpenCell_annotated_proteins.csv).


A
Pre-trained model
(localization encoding)
Train NOVA
without
freezing layers
Intermediate perturbation
learning model
B
Angle-metric
value
θ(L)
Calculate the angle-metric
value for each layer:
Layer 10
Layer 5
Layer 22
Layer 60
Layer 12
Layer 43
Layer 11
Layer 12
...
...
Layer 15
Layer 1
Layer 37
Layer 55
Layer 42
Layer 3
Layer 13
Layer 6
Layers
To train
50%
To freeze
0
0.1
0.2
Angle-metric value
C
Pre-trained model
(localization encoding)
Train NOVA
with
freezing partial
layers
Final NOVA model

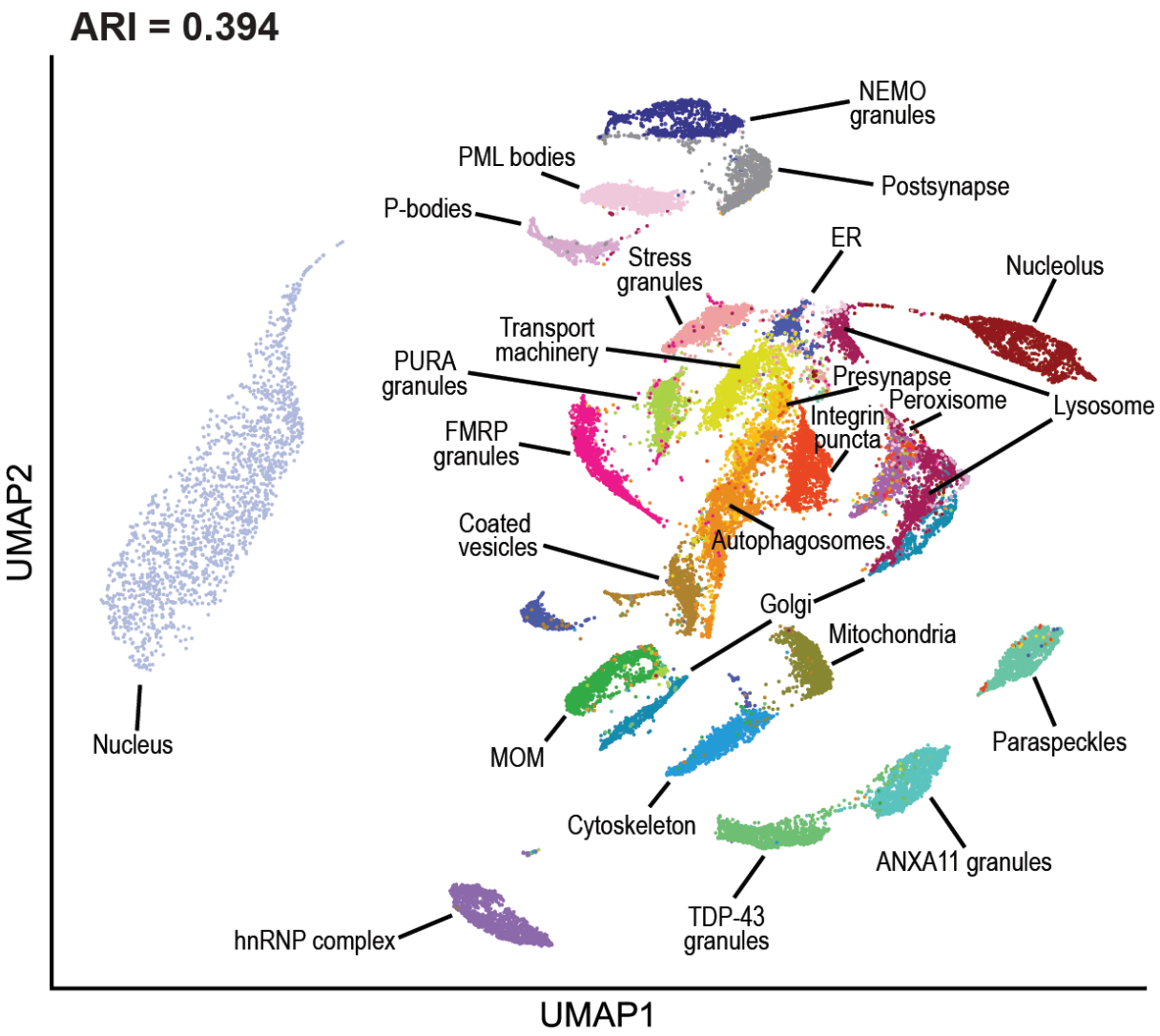
Supplementary Figure 5. Partial fine-tuning using the angle-metric approach.

Given the pre-trained model for localization encoding on the OpenCell dataset, a data-driven approach is applied to determine which layers to freeze in a fine-tuning process for perturbation learning. Freezing a layer prevents it from changing during the training. **(A)** The first step is fine-tuning the model (with a contrastive loss function, training on data of neurons with perturbations) without freezing any of the layers. **(B)** Calculate the angle metric (a measure of similarity between vectors) between pairs of corresponding layers in the pre-trained and the fine-tuned models. The layers are sorted by angle metric value and layers scored less than the median are frozen. **(C)** A final fine-tuning is applied, starting from the same pre-trained model, with selected layers frozen, resulting in the final NOVA model used in the entire work.



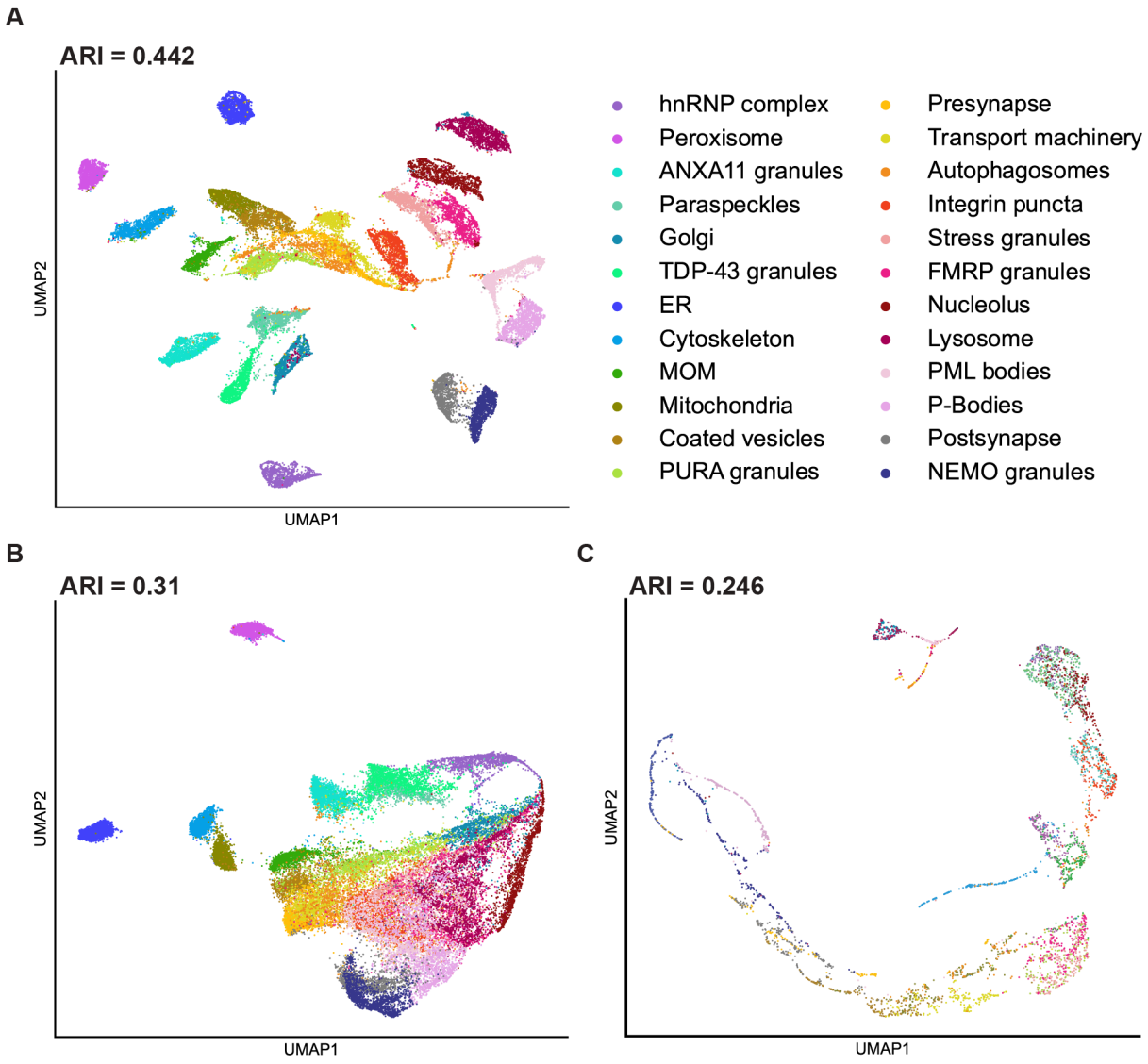
Supplementary Figure 6. Training curves for the perturbation learning ViT fine-tuned model.

The training curves of the fine-tuned model for perturbations learning on iPSCs-derived neurons (train, val (validation), and test. Loss value of the model (y-axis) and the number of epochs (x-axis). Epoch 67, demarcated by a red dot, serves as the final model since it is the last epoch in the model's training process, which showed an improved loss function on the validation data. Black box with exact loss values.



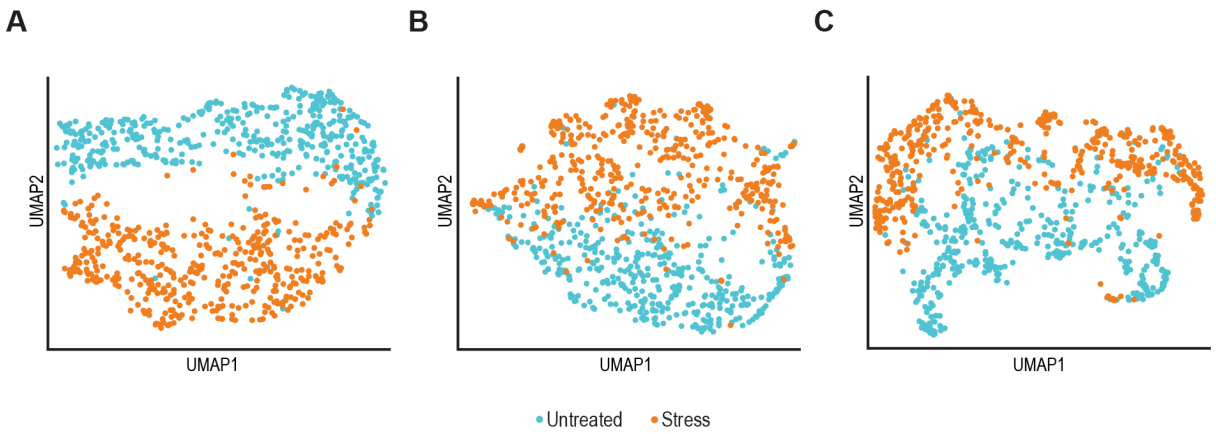
Supplementary Figure 7. Reproducibility of NOVA.

The full landscape of 25 organelles reproduced similar clusters to those presented in **Figure 1C** when applying NOVA onto a different batch (ARI=0.394).



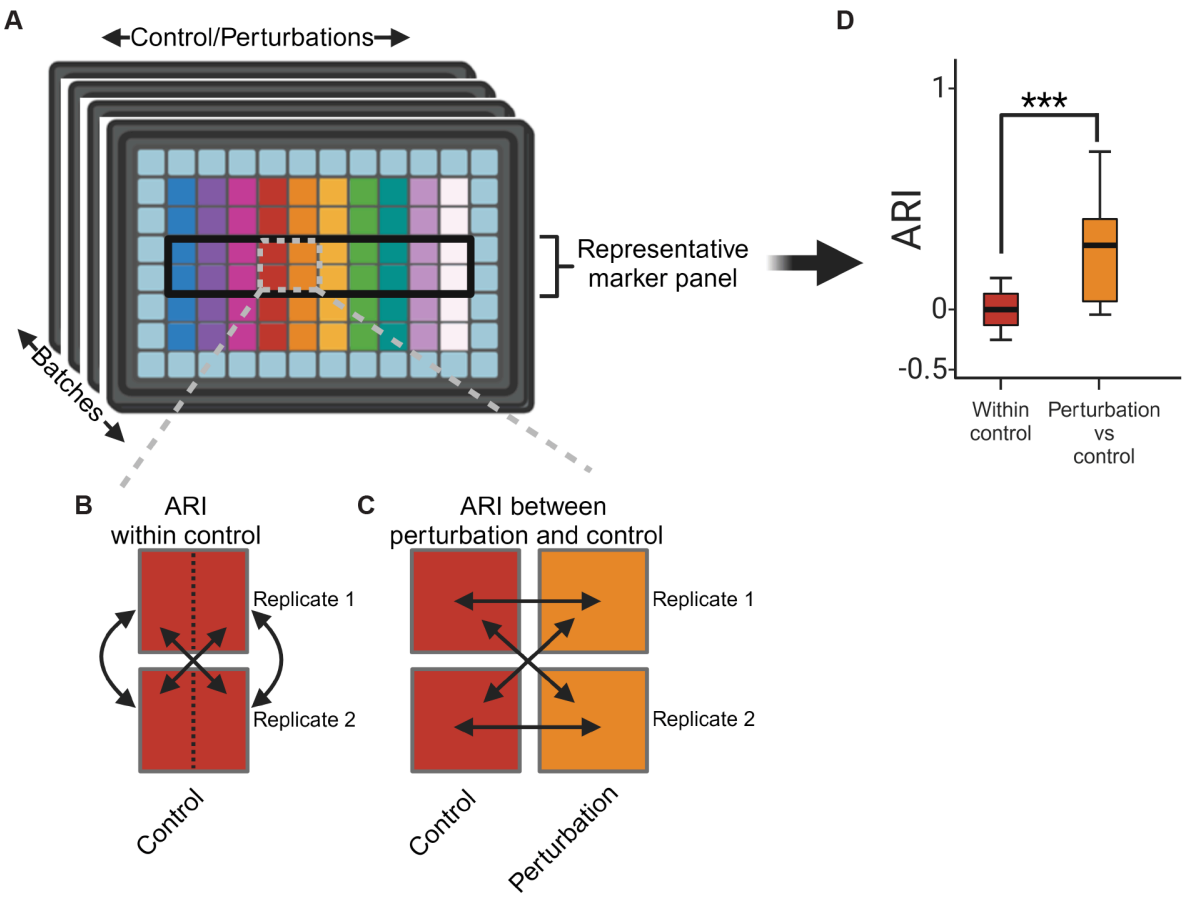
Supplementary Figure 8. NOVA compared to published methods

The full landscape of organelles (without the nucleus (Hoechst 33342)), on the data presented in **Figure 1C**, expresses better-defined clusters when based on **(A)** NOVA predicted features (ARI=0.442) in comparison to **(B)** Cytoself predicted features (ARI=0.31) or to **(C)** CellProfiler extracted features (ARI=0.246).



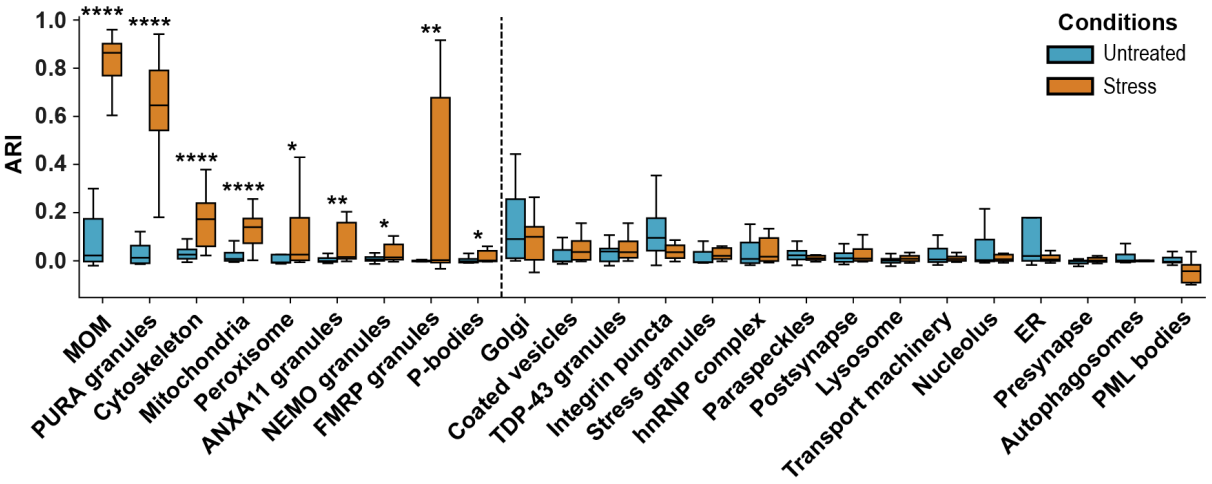
Supplementary Figure 9. NOVA effectively depicts the stress status of U2OS cells without being trained on this cell type.

UMAP reveals clustering of organellar topographies, in U2OS cells response to sodium arsenite stress (0.4 mM, 30 min), although NOVA was not trained on images of this cell line **(A)** stress granules (G3BP1), **(B)** P-bodies (DCP1A) and **(C)** Cytoskeleton (Phalloidin).



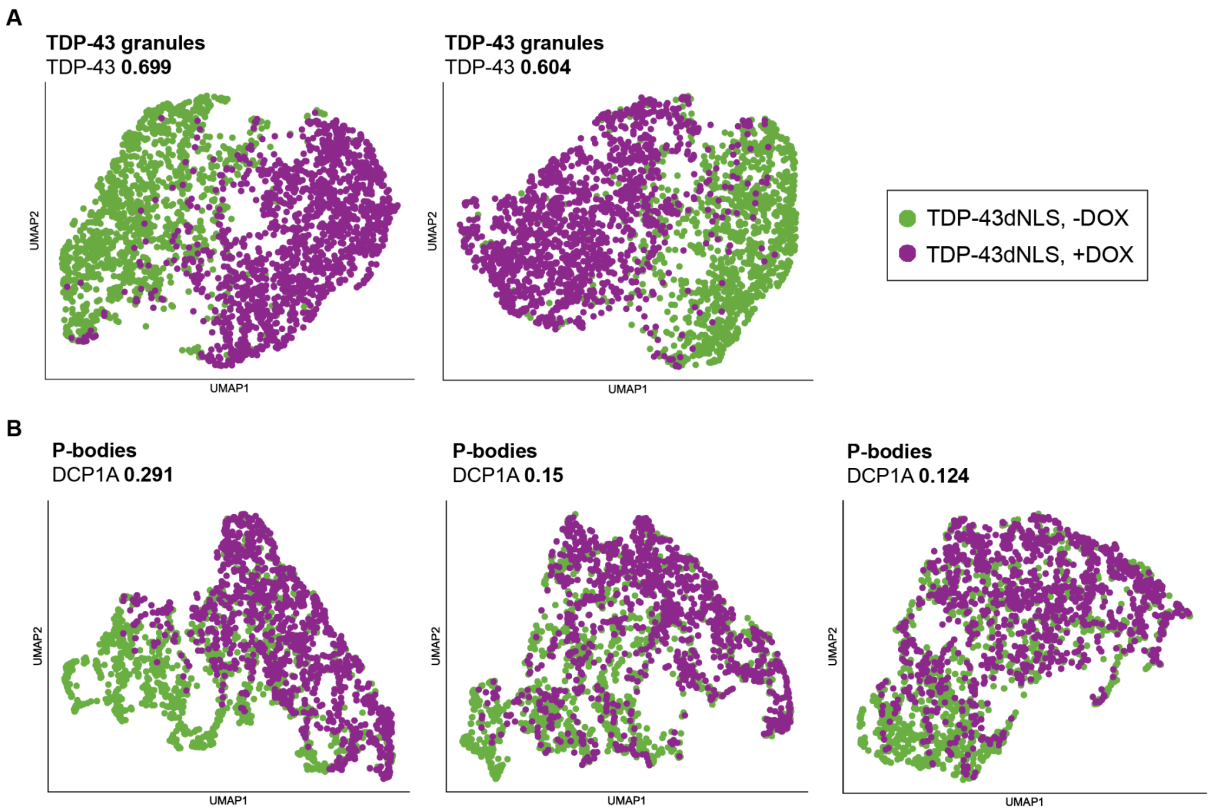
Supplementary Figure 10. Organelle scoring of perturbation effects across replicates and batches.

(A) An overview of the experimental design in a plate containing perturbations (e.g., cell lines; columns) and representative marker panels (black squares; rows) as described in **Supplementary Figure 1**. ARI scores were calculated in the model's latent space for **(B)** random splits of the control samples across different replicates, representing the noise, **(C)** and for perturbations and control across different replicates, reflecting their degree of separation. **(D)** Using aggregated ARI scores from multiple batches, the significance of a perturbation's effect relative to the noise was assessed with an unpaired one-sided t-test.



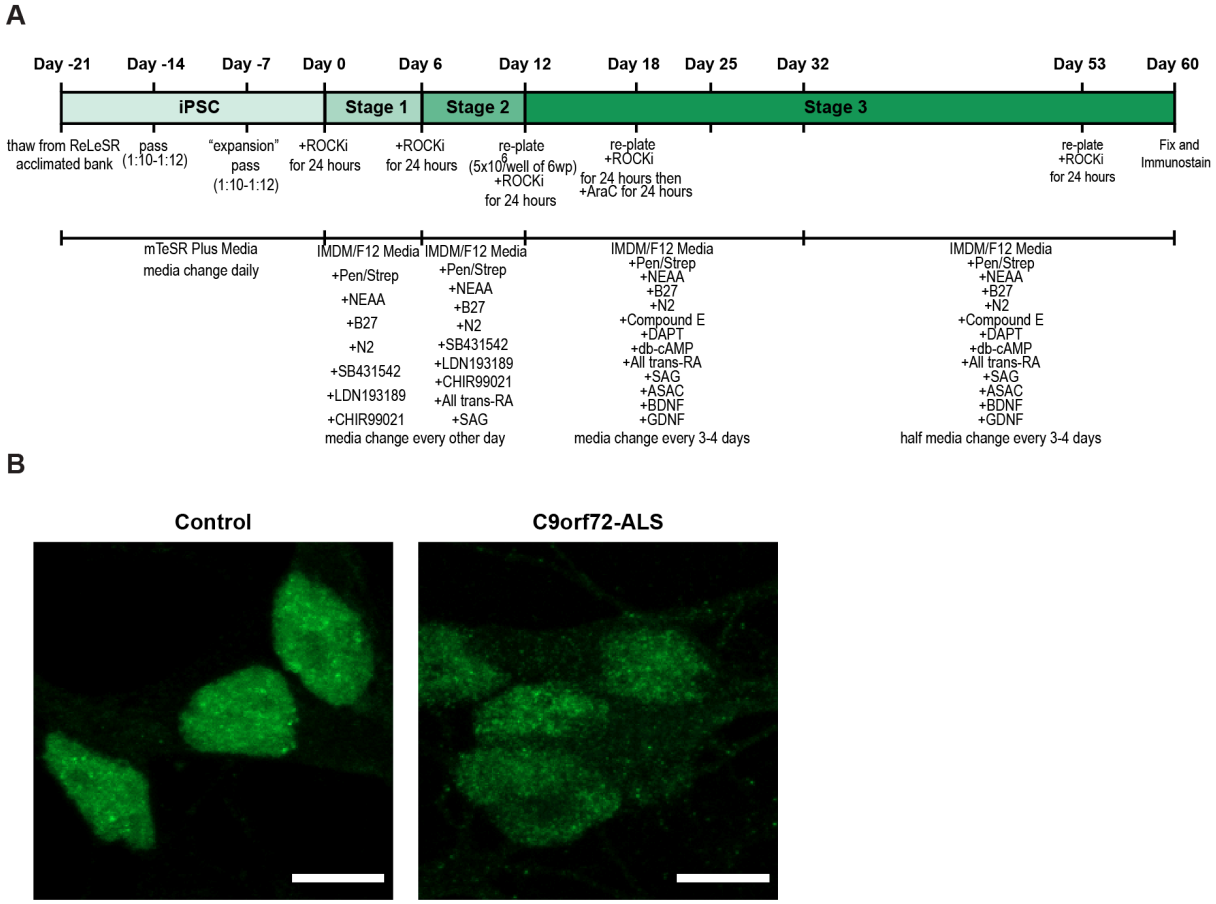
Supplementary Figure 11. The organellar stress response studied by CellProfiler features.

Organelle scoring (y-axis) without (blue) or with stress (orange), using 347 CellProfiler features across 4 batches (see **Methods**). Organelles are ordered by the median ARI values (left to right). Left of the dash line - organelles with significant topographic changes in stress. Boxes indicate the interquartile range (IQR) and whiskers are drawn to the farthest data point within 1.5*IQR from the nearest hinge. Lines represent median values. Unpaired one-sided t-test p-values * < 0.05, ** < 0.01, ****< 0.0001.



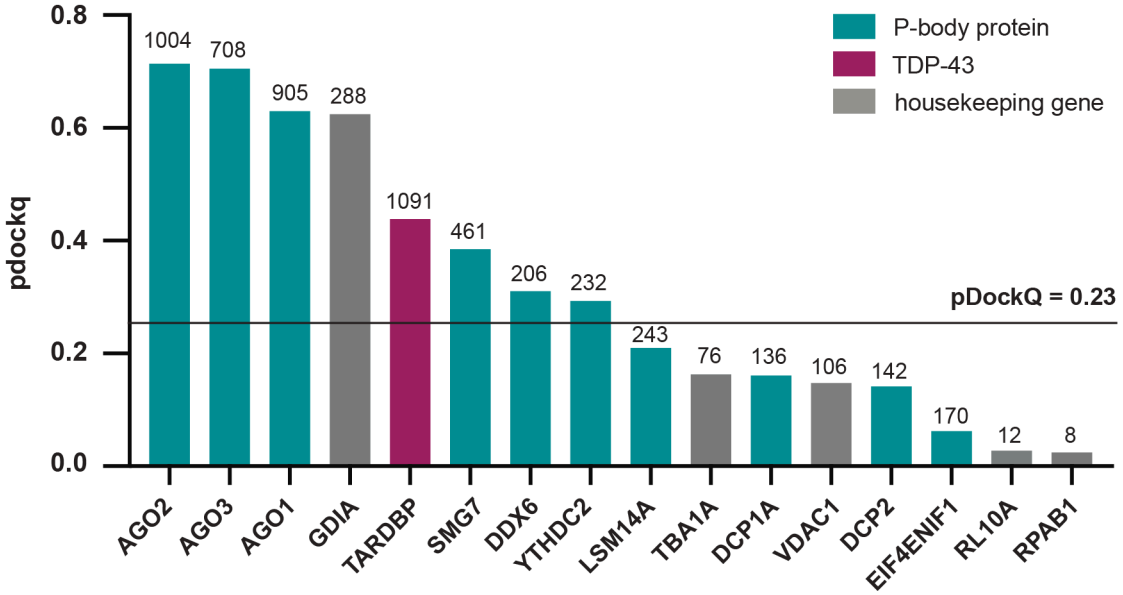
Supplementary Figure 12. Topographical changes of TDP-43 granules and P-bodies in response to cytoplasmic TDP-43.

UMAPs of **(A)** TDP-43 granules and **(B)** P-bodies, without (-DOX) or with (+DOX) the induction of cytoplasmic TDP-43. Two or three batches respectively, in addition to the one shown in **Figure 3C**. ARI scores in black.



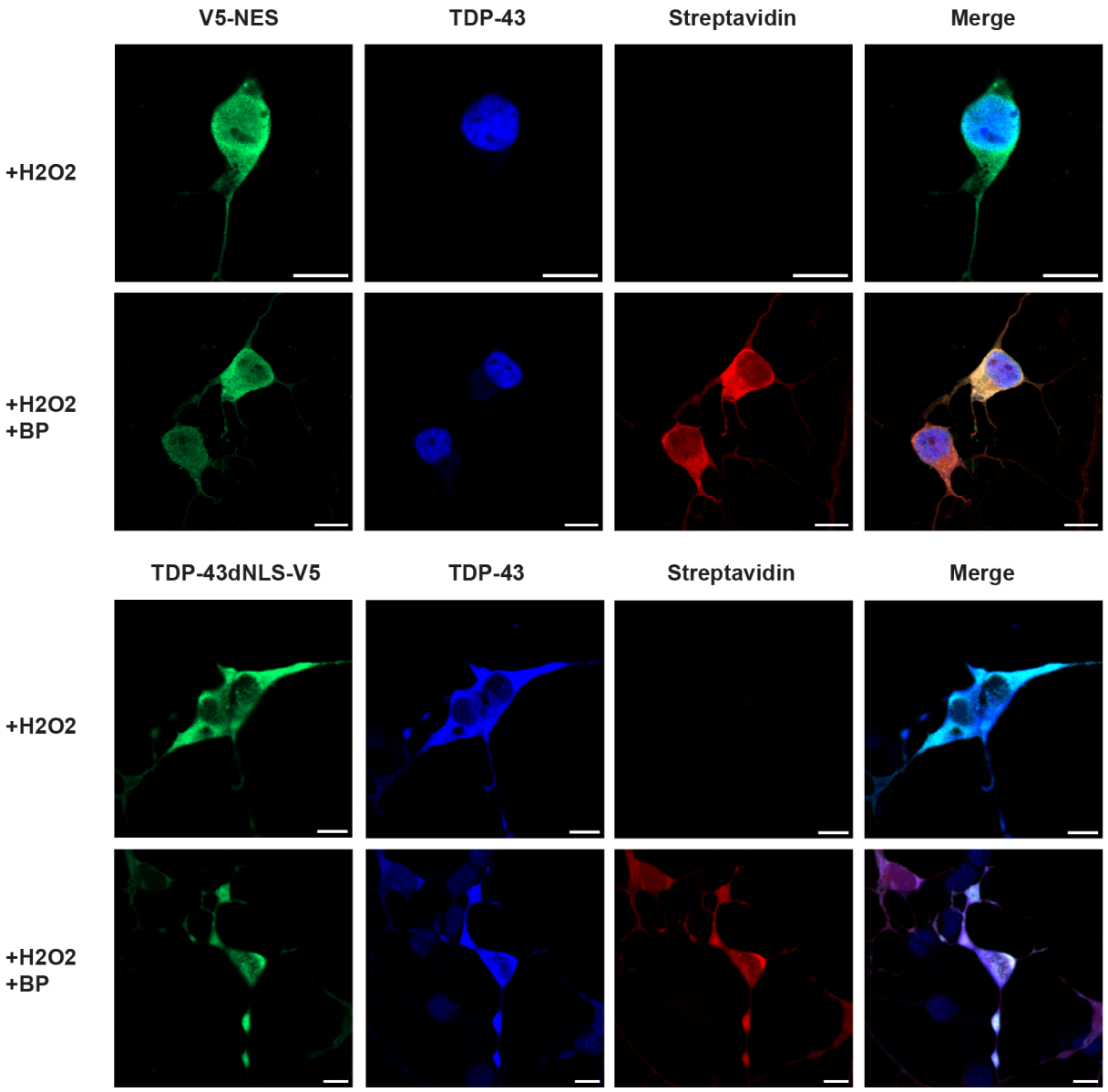
Supplementary Figure 13. Differentiation of patient-derived iPSCs into day 60 neurons.

(A) Timeline for induction and maturation of iPSC-derived motor neurons over 60 days in culture. Days from iPSC induction (Day 0). Technical steps are mentioned in more detail in the **Methods**. **(B)** Representative images of TDP-43 granules (green) in day 60 neurons, derived from a healthy (non-neuronal disease) control subject (left panel) or a familial ALS patient with *c9orf72* hexanucleotide repeat expansion (C9orf72-ALS) expressing cytoplasmic TDP-43 (right panel). Scale bar - 10 µm. Lens - X63.



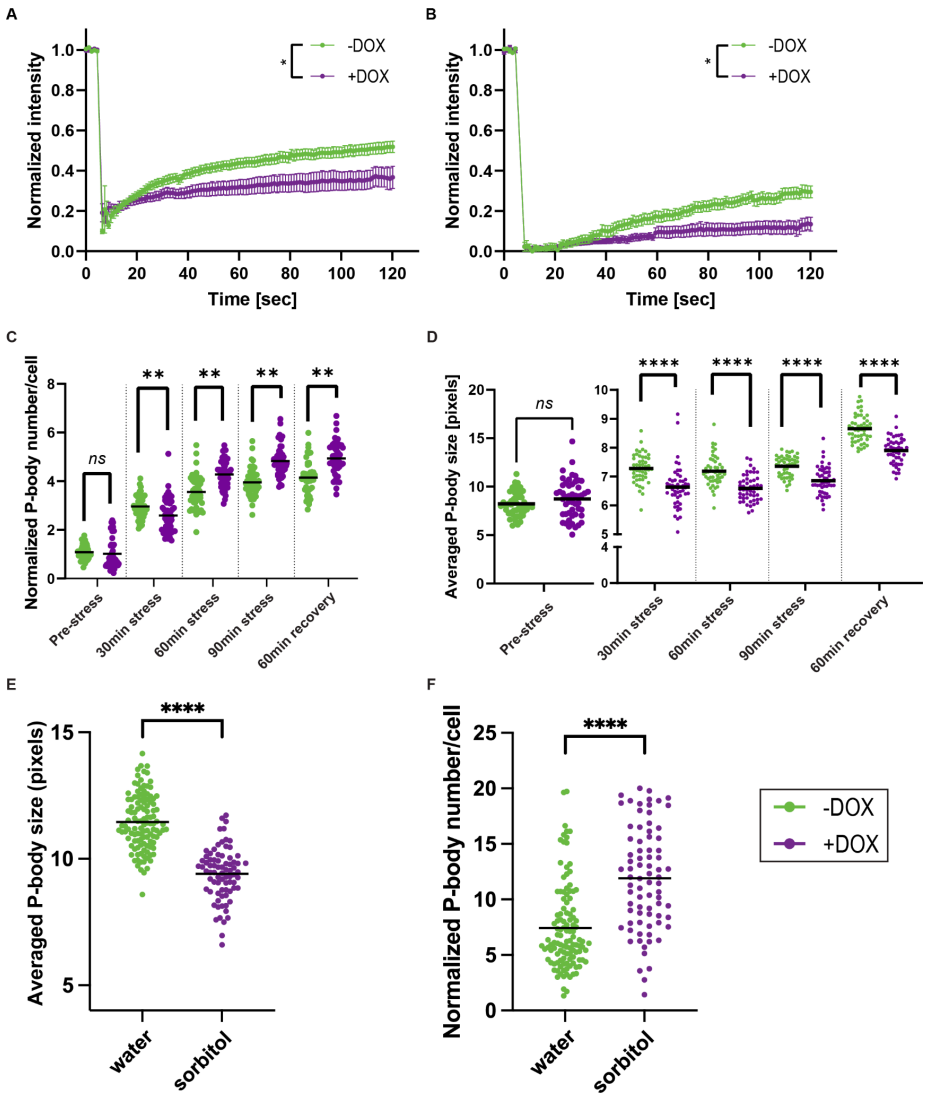
Supplementary Figure 14. Predicted protein-protein interactions between P-body resident proteins and TDP-43, as predicted by SpeedPPI [144].

The numbers above the bars indicate predicted contact sites between each protein and TDP-43. Y-axis - pDockQ scores. If pDockQ > 0.23 we call the interaction between the proteins a significant one.



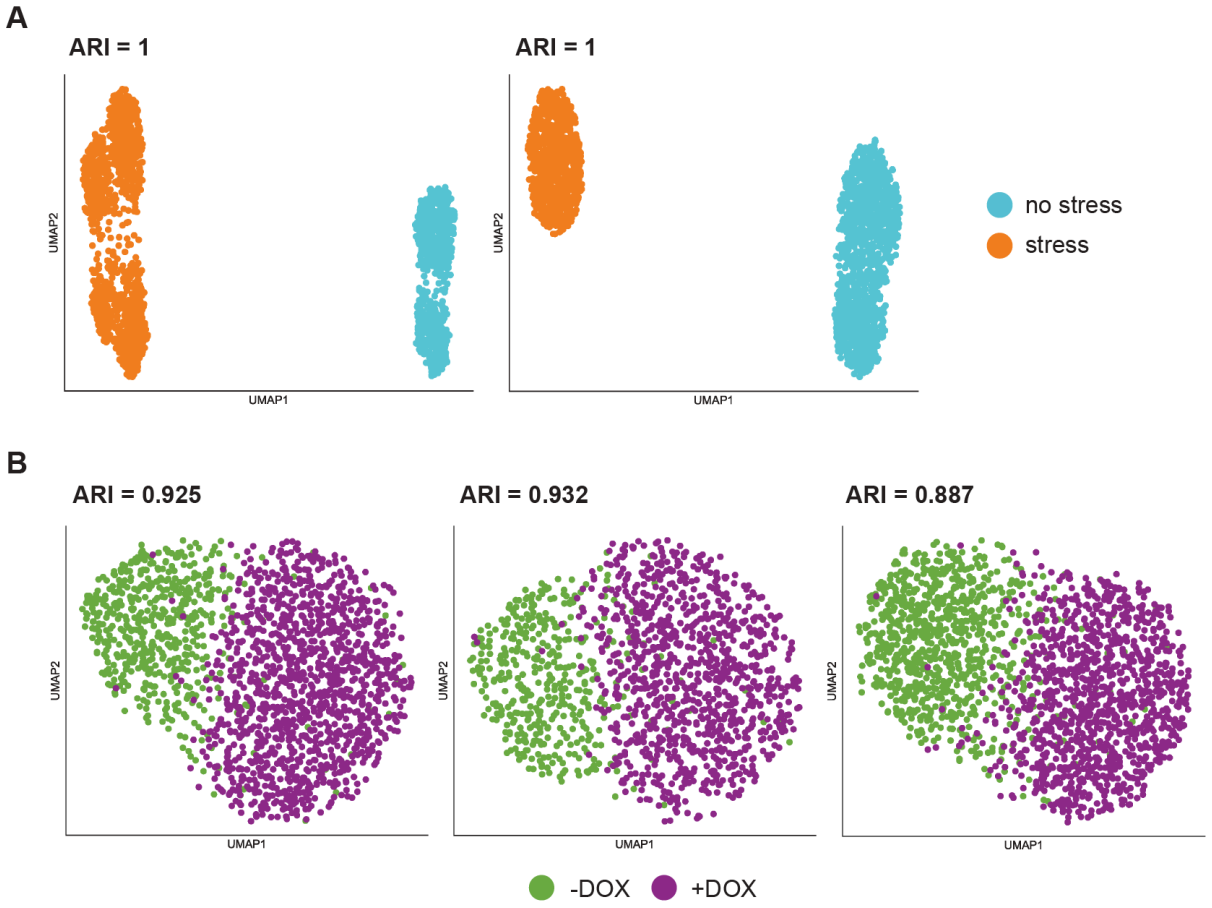
Supplementary Figure 15. Validation of APEX construct activity in iPSC-derived neurons.

Immunostaining of neurons overexpressing APEX-V5-NES (+DOX; top) or TDP-43^{deltaNLS}-V5-APEX (+DOX; bottom) without or with biotin phenol (BP) in the culture media. V5-tagged overexpressed construct (green), endogenous TDP-43 (blue), and biotinylated proteins (Streptavidin; red). Scale bar - 10 µm.



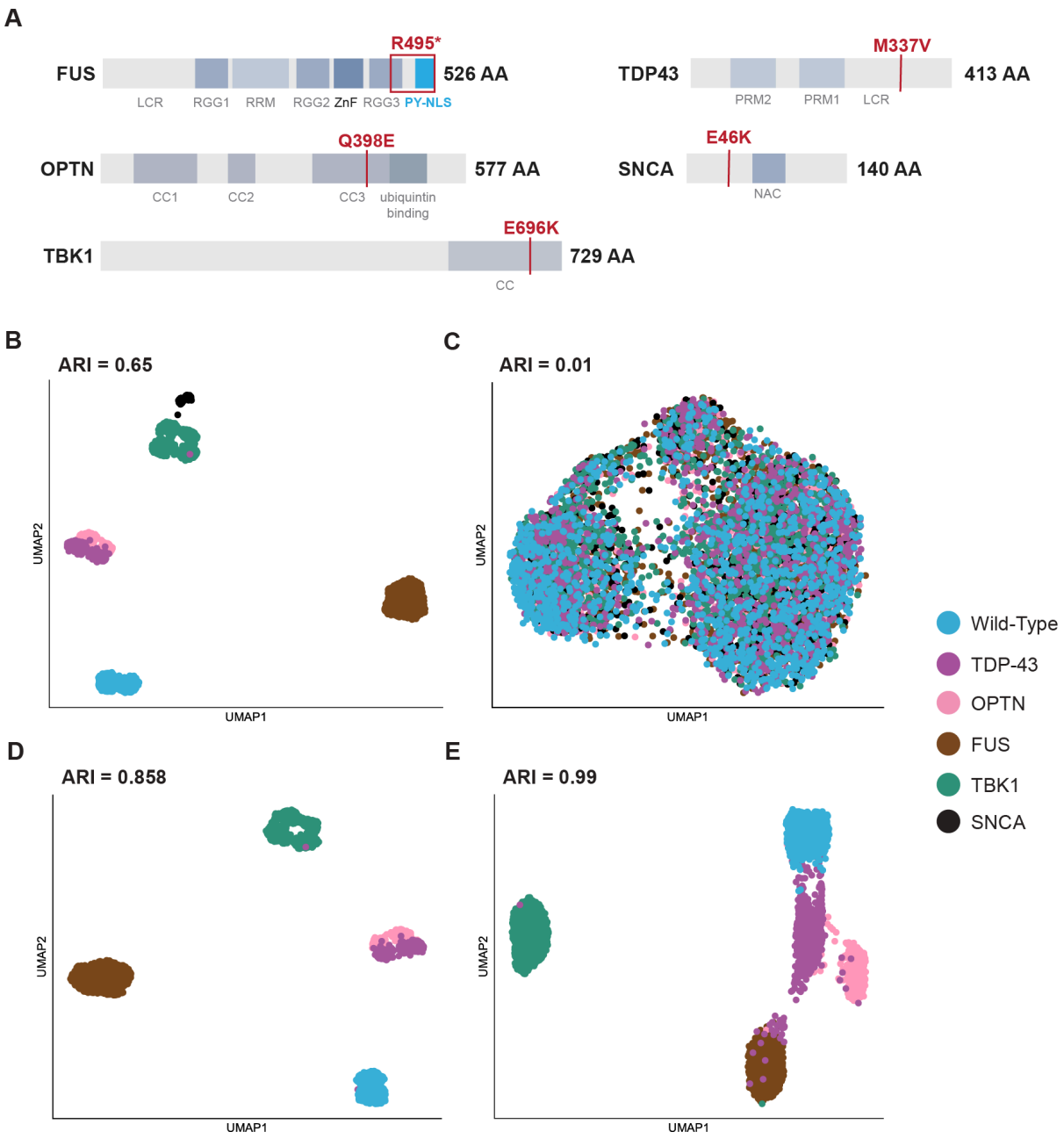
Supplementary Figure 16. Cytoplasmic TDP-43 disrupts P-body biophysics and functions under stress conditions in different cellular models.

(A, B) FRAP analysis related to **Figure 4A, B**. P-bodies depicted by the expression of YFP-DCP1A in U2OS cells partially photobleached. Fluorescence normalized to basal state and recovery monitored over time. Two independent repeats. Mixed-effects model with Geisser-Greenhouse correction; p-value < 0.01. CellProfiler analysis of **(C)** P-body number and **(D)** size in the presence or absence of induced cytoplasmic TDP-43 (-DOX or +DOX, respectively), under sodium arsenite stress (0.25 mM, 90 min) and 60 min recovery, in U2OS cells expressing YFP-DCP1A. n = 50 image fields/condition. CellProfiler analysis of **(E)** P-body size or **(F)** number per cell**,** in day 8 iPSC-derived neurons (iW11; parental line of the TDP-43^{deltaNLS} iPSC line, see **Methods**) treated with carrier (H2O, control) or sorbitol (osmotic stress; 0.4M, 2 hr) to induce mislocalization of endogenous TDP-43 into the cytoplasm. P-body number is normalized to the cell number per image. P-body size measured as average pixels per image. n = 50 images/group. Unpaired t-test p-value** < 0.01; **** < 0.0001; ns = non-significant. For panels **C-F,** black horizontal lines represent mean values.



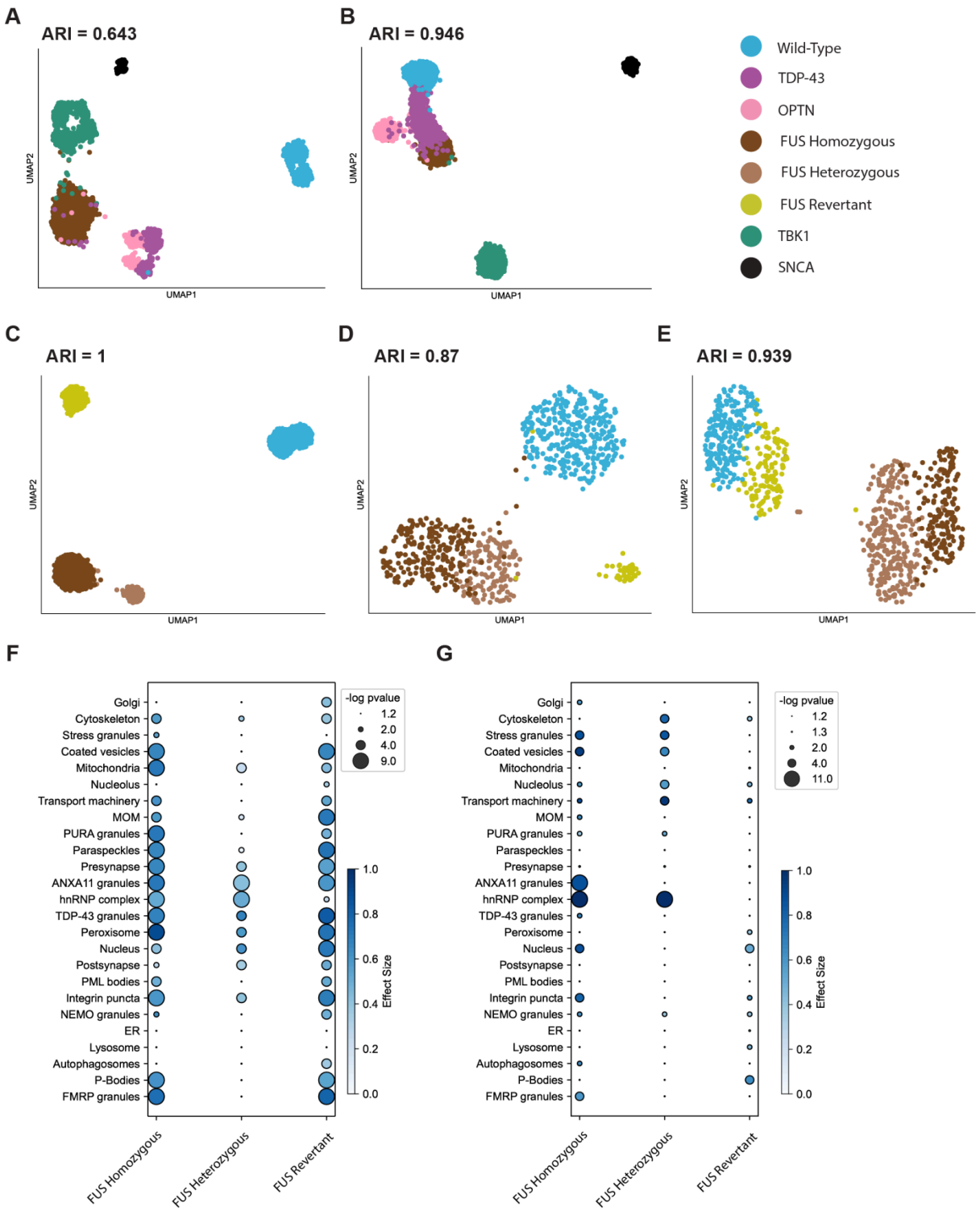
Supplementary Figure 17. Reproducibility of synthetic superposition of the organellome.

UMAPs of synthetic superposition of 25 organelles in cortical neurons **(A)** ±stress conditions (two batches, data from **Figure 2**) or **(B)** induced and uninduced cytoplasmic TDP-43 conditions (three batches, data from **Figure 3**).



Supplementary Figure 18. Reproducibility and validation of synthetic superposition of the organellome in iNDI neurons.

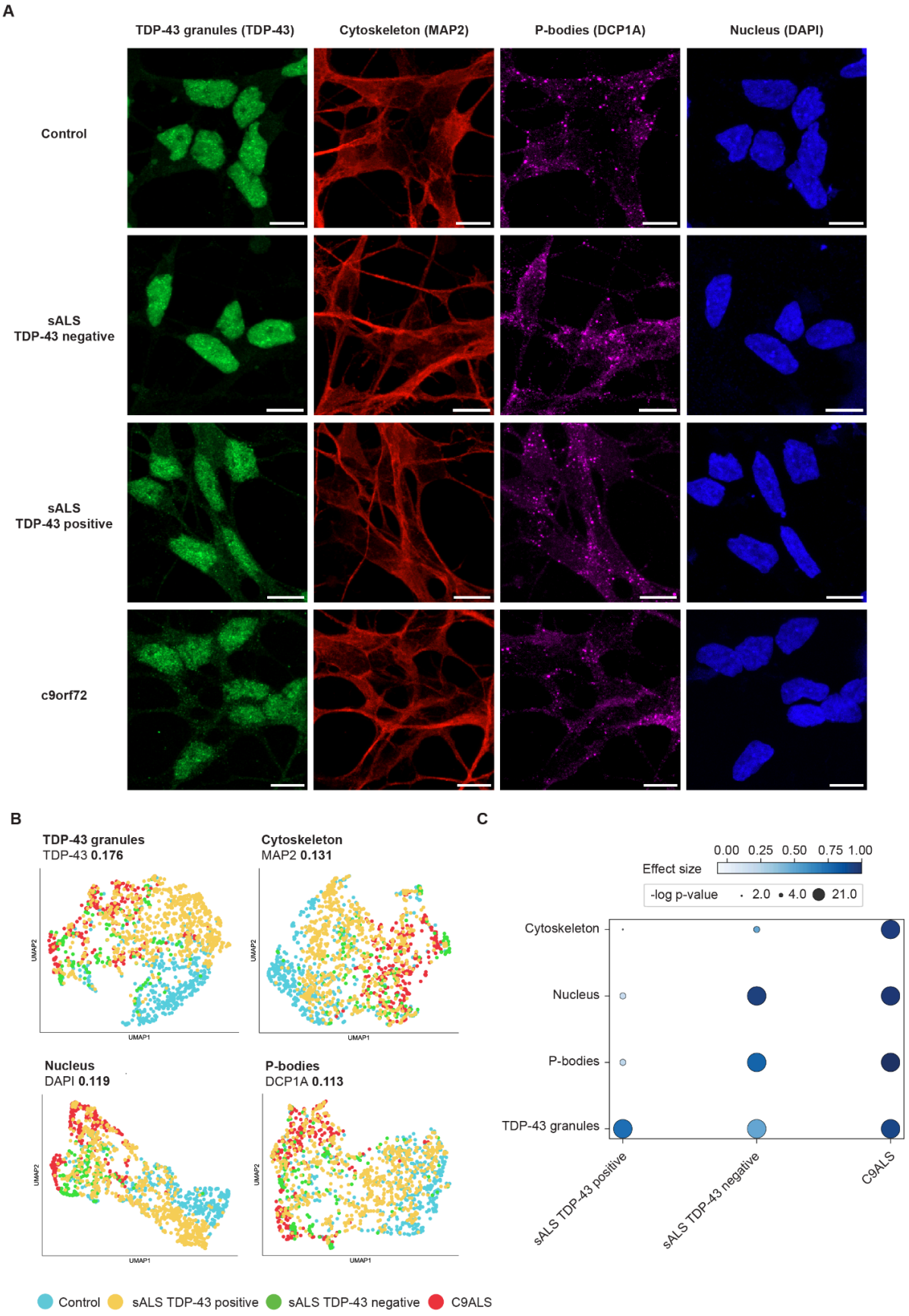
(A) Diagram of genes with ALS- and PD-associated mutation cell lines, from the iPSC Neurodegenerative Disease Initiative (iNDI) used in this study. UMAP of synthetic superposition of **(B)** 25 organelles in cortical iNDI neurons in a second batch (data from **Figure 5**) or **(C)** after cell line label shuffling control. This demonstrates cell line-dependent information content. **(D, E)** A similar analysis, excluding SNCA E46K mutant line.



Supplementary Figure 19. FUS-ALS organellomic analysis in day 8 and day 18 iNDI neurons.

(A, B) UMAPs of synthetic superposition of 24 out of 25 organelles excluding the hnRNP complex (marked by FUS) reveals hnRNP complex contribution to the clustering obtained in **Figure 5D**. Two different experimental batches. UMAPs of synthetic superposition of all 25 organelles, in **(C)** day 8 or **(D, E)** day 18 FUS R495* homozygous, heterozygous and revertant lines. Bubble plot demonstrating

organellome changes relative to wild-type control. Study of FUS 495* homozygous, heterozygous and revertant lines in **(F)** day 8 or **(G)** day 18 neurons. Bubble size - significance (-$\log_{10}$ p-value; one-sided unpaired t-test; values of 1.3 or lower are not significant). Bubble color - effect size (Cliff's Delta).


A
TDP-43 granules (TDP-43)
Cytoskeleton (MAP2)
P-bodies (DCP1A)
Nucleus (DAPI)
Control
sALS TDP-43 negative
sALS TDP-43 positive
c9orf72
B
TDP-43 granules
TDP-43 0.176
Cytoskeleton
MAP2 0.131
Nucleus
DAPI 0.119
P-bodies
DCP1A 0.113
UMAP1
UMAP2
Control
sALS TDP-43 positive
sALS TDP-43 negative
C9ALS
C
0.00 0.25 0.50 0.75 1.00
Effect size
-log p-value
2.0
4.0
21.0
Cytoskeleton
Nucleus
P-bodies
TDP-43 granules
sALS TDP-43 positive
sALS TDP-43 negative
C9ALS

Supplementary Figure 20. Organellomics of patient-derived day 60 motor neurons.

(A) Representative images of TDP-43 granules (green), Cytoskeleton (red), P-bodies (magenta), and Nucleus (blue) in day 60 motor neurons. Scale bar - 10 µm. **(B)** UMAPs of 4 organelles in day 60 motor neurons. Studied groups: healthy control (blue), sALS cytoplasmic TDP-43 positive (yellow) sporadic ALS (sALS) cytoplasmic TDP-43 negative (green) or C9ALS (red). **(C)** Bubble plot displaying organellar changes in patient-derived day 60 motor neurons relative to healthy non-neuronal disease control lines. Bubble size - significance (-$\log_{10}$ p-value; one-sided unpaired t-test; values of 1.3 or lower are not significant). Bubble color - effect size (Cliff's Delta).

Material & Methods

Cell lines

1. **Cancer cell lines and iNDI lines**

#	Cell line	Reference / gift
1	KOLF2.1J human iPSC	[145]
2	TDP-43^{deltaNLS}-V5-APEX NGN2-iPSC	This study
3	NES-V5-APEX NGN2-iPSC	This study
4	U2OS stably expressing YFP-DCP1A	A gift from Nancy Kedersha Pavel Ivanov and Paul Anderson, Harvard Medical School
5	U2OS stably expressing UBAP2L-ruby and /inducibly expressing TDP-43^{deltaNLS}-clover	[87]A gift from Don W Cleveland, University of California, San Diego
6	TARDBP M337V SNV/SNV	[100]
7	OPTN Q398E SNV/SNV	[100] Gift from Michael Ward, National Institutes of Health
8	TBK1 E696K SNV/SNV	[100]
9	FUS R495* SNV/SNV	[100]
10	FUS R495* SNV/WT	[100]
11	FUS R495* Rev/WT	[100]
12	SNCA E46K SNV/SNV	[100]

2. **Patient-derived iPSC lines including demographic Information**

#	iPSC Line Name	Source	Clinical Diagnosis	Age at Time of Collection	Sex	Origin

1	CS0002	Cedars-Sinai	Non-neurologic control	51	Male	PBMC
2	CS0201	Cedars-Sinai	Non-neurologic control	56	Female	PBMC
3	CS1ATZ	Cedars-Sinai	Non-neurologic control	60	Male	PBMC
4	CS2AE8	Cedars-Sinai	Non-neurologic control	50	Female	PBMC
5	CS8PAA	Cedars-Sinai	Non-neurologic control	58	Female	PBMC
6	CS9XH7	Cedars-Sinai	Non-neurologic control	53	Male	PBMC
7	CS0JGZ	Cedars-Sinai	sALS	56	Male	PBMC
8	CS3EPR	Cedars-Sinai	sALS	60	Female	PBMC
9	CS3XLK	Cedars-Sinai	sALS	55	Female	PBMC
10	CS4PFR	Cedars-Sinai	sALS	72	Female	PBMC
11	CS4ZCD	Cedars-Sinai	sALS	53	Male	PBMC
12	CS5JPF	Cedars-Sinai	sALS	55	Female	PBMC
13	CS5ZHY	Cedars-Sinai	sALS	64	Male	PBMC
14	CS6PYD	Cedars-Sinai	sALS	69	Female	PBMC
15	CS8JGP	Cedars-Sinai	sALS	61	Male	PBMC
16	CS9GXD	Cedars-Sinai	sALS	68	Male	PBMC
17	CS0JPP	Cedars-Sinai	sALS	61	Male	PBMC
18	CS4MLT	Cedars-Sinai	sALS	54	Male	PBMC
19	CS0BUU	Cedars-Sinai	C9orf72	63	Female	PBMC
20	CS2YNL	Cedars-Sinai	C9orf72	60	Male	PBMC
21	CS7VCZ	Cedars-Sinai	C9orf72	64	Male	PBMC

Human post-mortem brain tissue samples

Case Number	GUID	Source	Clinical Diagnosis	Age at Death	Sex
GBB-19-14	NEUNC276RND	Target ALS post-mortem core	Control	67	Male
GBB-16-10	NEUCG551GV1	Target ALS post-mortem core	Control	57	Female
GBB-18-13	NEUGV102NYT	Target ALS post-mortem core	Control	63	Female
GBB-17-22	NEUGJ583BEE	Target ALS post-mortem core	Control	70	Male
GBB-18-07 (Georgetown Case)	NEUGU437VNF	Target ALS post-mortem core	C9orf72	66	Female
GWF15-09	NEUTP019MY9	Target ALS post-mortem core	C9orf72	61	Female
GWF18-31	NEUKN179PH8	Target ALS post-mortem core	C9orf72	62	Female
GBB-18-17 (Georgetown case)	NEUBE992ZD1	Target ALS post-mortem core	C9orf72	68	Male

Antibodies used in this study

1. **Primary antibodies**

#	Target	Antibody #	Species	Company
1	ANXA11	HPA027545	rabbit	Atlas antibodies
2	Calreticulin	ab2907	rabbit	abcam
3	CD41	ab33660	rat	abcam
4	CLTC	ab21679	rabbit	abcam
5	DCP1A	PA5-82478	rabbit	Invitrogen
6	DCP1A	H00055802-M06	mouse	Novus Biologicals
7	DCP1A	ab47811	rabbit	abcam
8	FMRP	ab17722	rabbit	abcam
9	FUS	sc-47711	mouse	Santa Cruz
10	G3BP1	sc-365338	mouse	Santa Cruz
11	GM130	610822	mouse	BD
12	Hoechst 33342	62249		Thermo Fisher
13	KIF5A	ab5628	rabbit	abcam
14	LAMP1	ab25630	mouse	abcam
15	LSM14A	18336-1-AP	rabbit	Proteintech
16	MAP2	188 004	guinea pig	Synaptic Systems
17	Mitotracker	M22426		Invitrogen
18	NCL	14574S	rabbit	Cell Signaling
19	NEMO/Ikkgamma	611306	mouse	BD
20	NONO/p54nrb	GTX19407	mouse	GeneTex
21	PEX14	abc142	rabbit	EMD Millipore
22	Phalloidin	A12379		Thermo Fisher
23	PML	sc-966	mouse	Santa Cruz
24	PSD95 (K28/43 / DLG4)	75-028	mouse	Neuromab
25	PURA	12D11	rat	Gift from Dierk Niessing, Helmholtz

				Munich
26	SNCA	ab1903	mouse	abcam
27	SQSTM1	ab56416	mouse	abcam
28	TDP-43	10782-2-AP	rabbit	Proteintech
29	TDP-43	60019-2-Ig	mouse	Proteintech
30	TOMM20	ab78547	rabbit	abcam

2. Secondary antibodies

#	Name	Antibody #	Species (against)	Company
1	Anti-rat-Cy2	712-545-153	rat	Jackson ImmunoResearch
2	Anti-rabbit-Cy5	711-175-152	rabbit	Jackson ImmunoResearch
3	Anti-rabbit-Cy3	711-165-152	rabbit	Jackson ImmunoResearch
4	Anti-rabbit-Cy2	711-225-152	rabbit	Jackson ImmunoResearch
5	Anti-rabbit-Alexa Fluor 488	A11034	rabbit	Thermo Fisher Scientific
6	Anti-mouse-Alexa Fluor 647	A21236	mouse	Thermo Fisher Scientific
7	Anti-mouse-Cy5	715-175-150	mouse	Jackson ImmunoResearch
8	Anti-mouse-Cy3	715-165-150	mouse	Jackson ImmunoResearch
9	Anti-mouse-Cy2	715-225-150	mouse	Jackson ImmunoResearch
10	Anti-guinea pig-Alexa Fluor 568	A11075	Guinea pig	Thermo Fisher Scientific

Experimental methods

Generation of polyclonal NGN2 iNDI lines

Prior to transfection, 0.4-0.5 million iPSCs were seeded into a well of a 6-well plate, allowing for 12-24 hours for cell attachment. PG-TO-hNGN2 plasmid (Gift from Michael Ward, Addgene plasmid #172115) was transfected with TransfeX Transfection Reagent (ATCC® ACS-4005™) along with piggyBac™ transposase vector. Plasmid DNA and Opti-MEM I Reduced-Serum Medium (Gibco 31985-047) were equilibrated to room temperature and were gently mixed to avoid foaming or damaging the reagents. A total of 2µg of DNA was introduced per well, with a donor-to-transposase ratio of 2:1. Next, 100ul of Opti-MEM I Reduced-Serum Medium was dispensed into a sterile microcentrifuge tube, along with the donor and transposase plasmids. Then, 3ul of TransfeX Transfection Reagent ACS-4005 was incorporated into the DNA mixture and was mixed thoroughly using gentle pipetting. A mini centrifuge was used briefly to consolidate the tube's contents. The resultant lipo:DNA complexes were allowed to incubate at room temperature for 15-30 minutes, promoting the formation of transfection-ready complexes. The complexes were carefully added drop-wise to the iPSCs, ensuring diverse coverage within the well for optimal transfection efficiency. The cells were then incubated at 37°C with 5% CO2 for 24 hours. At 24 hours post-transfection, a complete media change was made and 0.4 ug/ml of puromycin (InvivoGen ant-pr-1) was added to the media. Media supplemented with 0.4 ug/ml of puromycin was replaced every day for 2-3 days until only BFP-expressing cells remained in the culture.

Generation of inducible TDP-43^{deltaNLS} iPS cell line

We generated iPS cell lines based on iW11-NGN2 (gift from Michael Ward [45]) featuring a doxycycline-inducible APEX construct [146] fused to either NLS-mutated TDP-43 (TDP-43^{deltaNLS}) (Addgene plasmid #229081) or nuclear export signal (NES) (Addgene plasmid #229082) as cytoplasmic control. A V5 tag was included to enable visualization by microscopy. After cloning these constructs into the PG-TO-hNGN2 plasmid in the place of NGN2, the cassette was integrated into the genome via the piggyBac transposon system described above. After introducing the plasmid, we carefully selected clones and cultivated them according to established protocols. The differentiation process followed the described methodology, with the induction of cytoplasmic TDP-43^{deltaNLS} achieved by adding doxycycline (2 ug/ml) for 24 hours.

Human iPSC culture (NGN2 lines)

iPSCs were grown at 37°C with 5% CO2 on tissue culture dishes coated with Geltrex (Gibco, A1413302) for at least 1 hour at 37°C. Cell culture media was exchanged daily with mTeSR1 medium (STEMCELL Technologies, 85850). For passaging, iPSCs were dissociated with accutase (ThermoFisher Scientific, A1110501) and seeded into mTeSR1 supplemented with ROCK inhibitor Y-27632 (Tocris Bioscience, 1254). Cells were frozen in mTeSR1 supplemented with 30% KnockOut Serum Replacement (ThermoFisher Scientific, 1082802) and 10% DMSO.

Differentiation of iPSCs to day 8 neurons (NGN2 lines)

Differentiation was performed as in [45] with minor adjustments. Briefly, after thawing, cells were transferred to tubes with DPBS, centrifuged, and resuspended in culture medium before seeding onto one Geltrex (Gibco, A1413302) coated well of a 6-well plate. Media was supplemented with 10 µM ROCK inhibitor (Y-27632, Tocris Bioscience 1254) for cell survival. During expansion, cells were cultured with daily media replacements of mTeSR1 medium (STEMCELL Technologies, 85850). Upon reaching 70%-80% confluency, cells were transferred to larger plates in preparation for induction. On the day of differentiation induction, cells were seeded with ROCK inhibitor (Tocris Bioscience, 1254, 10 µM) supplemented to induction medium comprising DMEM/F12, HEPES (Gibco, 11330032), N2 supplement (Gibco, 17502048), NEAA (Gibco, 11140050), L-glutamine (Gibco, 25030081), and freshly prepared doxycycline solution (2mg/ml stock, Sigma D9891). Based on plate size, either 1.5 million or 2.5 million cells were seeded, targeting 5 million and 15 million neurons, respectively. On induction days 1-3 induction media was replaced daily. For maturation starting from day 3 onwards, overnight PDL-coated (sigma P7405, 0.04mg/ml) 96-well plates (Brooks, MGB096-1-2-LG-L) were prepared, and approximately 60,000 day 3 neurons were seeded per well in maturation medium containing BrainPhys (STEMCELL Technologies, 05790), B27 supplement (Gibco, 17504044), BDNF (PeproTech, 450-02), NT3 (PeproTech, 450-03), and Laminin (Gibco, 23017-015). On the first maturation day, 40mM BrdU (Sigma, B9285) was added to the maturation medium. Neurons matured with half-medium replacements every 3 days. For staining mitochondria, cells were incubated with mitotracker (M22426, 1:20,000 dilution) for 30 min prior to fixation. For stress experiments, cells were incubated with sodium arsenite (250 µM, 30 minutes, Sigma Aldrich, S7400) or D-sorbitol (sorbitol, 0.4 M, 2 hr, Sigma Aldrich, S7547) prior to fixation.

Differentiation of iPSCs to day 18 neurons (iNDI lines)

Differentiation of iPSCs to day 18 neurons was performed similarly as described above for day 8 neurons, with a few minor changes as described in [147]. Neuronal cultures were seeded on day 3 of differentiation on Poly-L-Ornithine (PLO, Merck P4957). PLO was introduced to PhenoPlate™ 96-well microplates (Revvity, 6055300) in a concentration of 0.01% and were incubated overnight at 37°C. After incubation, the plates were washed 3 times with ddH2O, and were left to dry completely. After drying, 40,000 neurons were seeded into each well of a 96-well plate with a total media volume of 300 ul per well. Neurons were matured in maturation medium containing BrainPhys (STEMCELL Technologies, 05790), NCS21 50x serum free (Diagnovum, D621-10ml), GDNF 10 ng/mL (PeproTech, 450-10), BDNF 10 ng/mL (PeproTech, 450-02), NT3 10 ng/mL (PeproTech, 450-03), Laminin 1 µg/mL (Gibco, 23017-015), doxycycline 2 µg/mL (Sigma-Aldrich, D9891-10G), Uridine 1 µM (Sigma-Aldrich, U3003), 5-Fluoro-2′-deoxyuridine 1 µM (Sigma-Aldrich, F0503), with half-medium replacements every 3-4 days.

Differentiation of iPSCs into 60-day spinal motor neurons

Sporadic ALS (sALS), C9orf72, and non-neurological control patient iPSC lines were acquired from the Answer ALS repository at Cedars Sinai (see Materials list). All iPSCs were maintained in 6 well plates coated with growth factor reduced Matrigel (Corning) and mTeSR Plus Media (StemCell

Technologies). iPSC lines were passaged at a ratio of 1:10 with ReLeSR (StemCell Technologies). As shown in Supplemental Figure 15, spinal motor neuron cultures were generated by differentiating iPSCs with a modified version of the direct induced motor neuron (diMNs) protocol that has been recently described [148,149]. Briefly, iPSCs were cultured and expanded for 3 weeks to facilitate acclimation and stabilization post-thaw. On day 0, iPSCs were dissociated with ReLeSR and plated in 6 well plates (coated with growth factor reduced Matrigel) at a density of 500,000 cells/well in stage 1 media (47.5% IMDM (Gibco), 47.5% F12 (Gibco), 1% NEAA (Gibco), 1% Pen/Strep (Gibco), 2% B27 (Gibco), 1% N2 (Gibco), 0.2 mM LDN193189 (Stemgent), 10 mM SB431542 (StemCell Technologies), and 3 mM CHIR99021 (Sigma Aldrich)) supplemented with 20 mM ROCKi (Y-27632, Millipore). The next day, media was exchanged and replaced with stage 1 media without ROCKi. A full volume of stage 1 media was subsequently exchanged on days 3 and 5 of differentiation. On day 6 of differentiation, cells were dissociated with accutase and plated in T25 flasks (coated with growth factor reduced Matrigel) at a density of $2x10^6$ cells/flask in stage 2 media (47.5% IMDM, 47.5% F12, 1% NEAA, 1% Pen/Strep, 2% B27, 1% N2, 0.2 mM LDN193189, 10 mM SB431542, 3 mM CHIR99021, 0.1 mM all-trans RA (Sigma Aldrich), and 1 mM SAG (Cayman Chemicals)) supplemented with 20 mM ROCKi. The next day, media was exchanged and replaced with stage 2 media without ROCKi. A full volume of stage 2 media was subsequently exchanged on days 9 and 11 of differentiation. On day 12 of differentiation, cells were dissociated with Trypsin and plated in 6 well plates (coated with growth factor reduced Matrigel) at a density of $5x10^6$ cells/well in stage 3 media (47.5% IMDM, 47.5% F12, 1% NEAA, 1% Pen/Strep, 2% B27, 1% N2, 0.1 mM Compound E (Millipore), 2.5 mM DAPT (Sigma Aldrich), 0.1 mM db-cAMP (Millipore), 0.5 mM all-trans RA, 0.1 mM SAG, 200 ng/mL Ascorbic Acid (Sigma Aldrich), 10 ng/mL BDNF (PeproTech), and 10 ng/mL GDNF (PeproTech) supplemented with 20 mM ROCKi. The next day, media was exchanged and replaced with stage 3 media without ROCKi. A full volume of stage 3 media was exchanged on day 16 of differentiation. On day 18 of differentiation, iPSNs were dissociated with accutase and plated in 6 well plates (coated with growth factor reduced Matrigel) at a density of $5x10^6$ cells/well in stage 3 media supplemented with 20 mM ROCKi. The next day, media was exchanged and replaced with stage 3 media supplemented with 20 mM AraC. The next day, media was exchanged and replaced with stage 3 media without AraC. A full volume of stage 3 media was subsequently exchanged every 3-4 days until day 32 of differentiation. Beginning on day 32 of differentiation, a half volume of stage 3 media was manually exchanged with a P1000 pipette every 3-4 days until day 53 of differentiation. On day 53 of differentiation, iPSNs were dissociated with accutase and plated in glass bottom 24 well plates (coated with growth factor reduced Matrigel) at a density of 350,000 cells per well in stage 3 media supplemented with 20 mM ROCKi. The next day, media was exchanged and replaced with stage 3 media without ROCKi. A half volume of stage 3 media was subsequently exchanged on day 57 of differentiation. All iPSC lines used in this study were differentiated in a single batch. All iPSCs and iPSNs were maintained at 37°C with 5% CO_2 and routinely tested negative for mycoplasma.

Immunofluorescence

Cells in 96-well plates were treated with 4% PFA in 1X PBS++ for a fixed duration of 15 minutes. Post-fixation, cells were rinsed with PBS and were then exposed to 0.1% TritonX in PBS for a permeabilization step for 10-15 minutes. Blocking was achieved by incubating the cells with CasBlock

reagent for 10 minutes at RT. Primary antibodies (listed in Materials) in CasBlock were prepared, and cells were submerged in 75 µl of this mix overnight at 4°C. The following day, cells were washed twice in 100 µl PBS for 5 minutes. Secondary antibodies (listed in Materials), diluted in PBS, were added to the cells, ensuring a 75 µl coverage for an hour while minimizing light exposure. Hoechst, diluted in PBS at a 1:8000 ratio, served as a DNA stain, and cells were treated with 75-100 µl of this mix for 10-30 minutes at room temperature. Before proceeding with imaging, the Hoechst solution was replaced with PBS. It was crucial to ensure the cells remained submerged in PBS, particularly during extended imaging sequences to prevent desiccation.

U2OS immunofluorescence experiments were performed on coverslips but using the same protocol as described for neurons. We used G3BP1-GFP stably expressing U2OS cells and stained for Phalloidin, DCP1A using antibodies (see Materials) and DAPI was applied with the coverslip mounting media. For stress conditions, cells were incubated with sodium arsenite (0.4 mM) for 30 minutes prior to fixation.

For 60-day motor neurons, iPSNs were fixed and immunostained on day 60, as previously described [149,150]. Briefly, iPSNs were fixed in 4% PFA for 15 mins and washed 3X 5 mins with 1X PBS prior to permeabilization with 0.1% Triton X-100 in PBS for 15 mins. iPSNs were then incubated in 10% goat serum/1X PBS blocking solution for 30 mins. Primary antibodies (1:500 Rabbit Anti-TDP-43 (ProteinTech, 10782-2-AP), 1:500 Mouse Anti-DCP1A (Novus Biologicals, H00055802-M06), and 1:1000 Guinea Pig Anti-Map2 (Synaptic Systems, 188 004) were diluted in block and iPSNs were incubated with resulting primary antibody solution for 2 hours at room temperature. iPSNs were then washed 3X 10 mins with 1X PBS and incubated with secondary antibody solution (1:1000 Goat Anti-Rabbit Alexa Fluor 488, 1:1000 Goat Anti-Guinea Pig Alexa Fluor 568, 1:1000 Goat Anti-Mouse Alexa Fluor 647; Thermo Fisher) diluted in block for 1 hour at room temperature. iPSNs were washed 3X 10 mins with 1X PBS, incubated with 1:1000 Hoechst/1X PBS solution for 5 mins at room temperature, and washed 2X 5 mins with 1X PBS prior to mounting with Prolong Gold and sealing of wells with a 15 mm glass coverslip. Agitation was not used at any step in the immunostaining protocol.

Microscopy

U2OS cells were imaged on a Zeiss LSM 900 confocal laser scanning microscope with 63x oil objective using standard settings. Imaging was performed for 100 tiles using Zeiss system autofocus and tiles were stitched together using 10% image overlap by the Zeiss Software.

iPSC-derived neurons (iNDI lines, day 8) were assessed using the Visitron Systems laser and laser control, complemented by the spinning disk confocal scanner unit from Yokogawa. The Olympus IX83 served as the primary microscope in this configuration, images were acquired at X60 using oil immersion. We used a filter wheel system featuring EF460/50m, E525/50m, FF01-609/54, and ET700/75m filters, along with the DM405/488/561/640 dichroic mirror. To optimize image acquisition, we employed specific parameters, including binning at 2 for pixel averaging and heightened sensitivity, an exposure time of 300ms, and dynamic adjustments to laser power (range: 0.1-100%) for optimal imaging of distinct fluorophores. Our high-throughput imaging strategy

encompassed the capture of 10x10 images per well within a 96-well plate, resulting in a total of 100 images per well, without overlap between adjacent images. This systematic approach was applied to all cell lines in a consecutive run. Each image has 2-4 channels of a panel of 2-4 antibodies. The entire process was controlled through VisiView Software, ensuring seamless microscope control and efficient image acquisition.

60-day motor neurons were imaged with a 63X objective with 1X digital zoom on a Zeiss LSM980 confocal in 24 well plates. Identical imaging parameters (e.g. laser power, gain) were used to acquire images. Ten images per well, 4 antibodies (see above), without overlap between adjacent images were taken.

Table 2: Experimental batches of isogenic iPSC-derived neurons

The table below depicts the overview of experimental batches generated as shown in **Supplementary Figure 1**.

Cell lines	**Conditions**	**N biological repeats**	**N batches (replicates) total**	**N image tiles**
Wild-type	+/- Sodium arsenite (SA)	2	6 (12)	627,301
TDP-43^{deltaNLS}	+/- Doxycycline (DOX)	1	3 (6)	305,407
Wild-type, FUS R495* homozygous, FUS R495* heterozygous, FUS R495* revertant, TDP-43 M337V, OPTN Q398E, TBK1 E696K, SNCA E46K	Untreated, day 8	2	6 (12)	2,010,906 (Wild-type images are from the sodium arsenite experiment)
Wild-type, FUS R495* homozygous, FUS R495* heterozygous, FUS R495* revertant	Untreated, day 18	1	2 (4)	218,540

Fluorescence recovery after photobleaching (FRAP)

U2OS cells overexpressing DCP1A-YFP (yellow fluorescent protein, a gift from Dr. Nancy Kadersha and Prof. Paul Tylor) were maintained as described above (section of: U2OS cells). For FRAP of P-bodies under uninduced or induced cytoplasmic TDP-43 expression, ~250K U2OS cells were seeded in

individual 6-wells with coverslips. Next, the cells were co-transfected with two different plasmids: piggyBac transposase plasmid and PG-TO-hNGN2 that carries doxycycline-inducible TDP-43^{deltaNLS}-V5-APEX (see above, expresses BFP; Blue Fluorescent Protein) in the place of NGN2, using Transfex reagent (ATCC, ACS-4005). The TDP-43^{deltaNLS}-V5-APEX construct was integrated into the genome via the piggyBac transposon system. Twenty-four hours post-transfection, the cells were incubated for twenty-four hours with doxycycline (DOX; 1ug/ml) for cytoplasmic TDP-43 induction or only with media (as control; -DOX; uninduced) and with sodium arsenite (250 uM, Sigma, S7400) 60-90 min before the FRAP measurement. FRAP image sequences were obtained on a Zeiss LSM 900 inverted scanning confocal microscope equipped with a heated chamber and a Plan-Apochromat 40× water objective (Carl Zeiss, Jena, Germany). First, we looked for cells expressing BFP to verify the successful uptake of TDP-43^{deltaNLS}-V5-APEX. Those cells were scanned using a 488 laser for the detection of YFP-labeled P-bodies. For analysis of the fluorescence recovery, an internal area within a few P-bodies in a field of view was bleached and the fluorescence recovery was measured every ~1 sec for 2 min. A single P-body and background area without bleaching were also measured per field of view (a site in a well). For each condition (uninduced or induced), 3-4 sites were analyzed in two independent 6-wells. In each site, between two to nine P-bodies were measured. FRAP data were normalized and calculated by subtracting the background signal and then normalized to the P-body that was not bleached. Measured P-bodies that were out-of-focus or exhibiting strange behaviors were excluded. The average of the normalized data for an experiment was synchronized by the average time point value. For each biological replicate, the statistical analysis was a mixed-effects model with Geisser-Greenhouse correction.

APEX proximity labeling

Monoclonal iPSC lines carrying TDP-43^{deltaNLS} and NES APEX constructs were differentiated into neurons as described previously. After 5 days of maturation, construct expression was induced for 24 and 4 hours respectively to achieve comparable levels of expression. APEX activity was induced by supplementing biotin-phenol (BP, 500 µM, Iris Biotech GmbH, LS-3500) for 90 minutes and H2O2 (1 mM, J.T.Baker 7722-84-1) for 1 min. Labeling activity was halted with quenching solution (QS: sodium azide (10mM, Mallinckrodt, 1953-57), sodium ascorbate (10mM, Sigma-Aldrich, A7631) and Trolox (5mM, Sigma-Aldrich, 238813) in PBS. Then, cells were scraped directly in RIPA lysis buffer supplemented with Complete Protease Inhibitor Cocktail (Roche, 4693116001) and PhosphoSTOP (Roche, 4906837001). Lysates were centrifuged at 15,000 × g for 10 min at 4°C after which protein concentration was quantified with Bio-Rad Protein Assay Dye Reagent (Bio-Rad, 500-0006). Streptavidin-coated magnetic beads (Pierce Streptavidin Magnetic Beads, Thermo-Fisher, 88816) were incubated at a ratio of 50 µl beads per 130 µg of sample with rotation overnight at 4°C. The next day, bead samples were washed with a series of buffers, 1 ml per wash, for 3 minutes at 1100 rpm at room temp to remove non-specific binders: twice with RIPA lysis buffer, once with 1 M KCl (MERCK, 104936), once with 0.1 M Na2CO3 (Sigma-Aldrich, S7795), and once with 2 M Urea solution containing 10 mM Tris-HCl, pH 8.0 (Invitrogen, 15568–025). Finally, the washed biotinylated proteins were transferred to on-bead digestion process by trypsinization.

Liquid Chromatography and Mass spectrometry

Samples were subjected to on-bead tryptic digestion followed by a desalting step. The resulting peptides were analyzed using nanoflow liquid chromatography (nanoAcquity) coupled to high resolution, high mass accuracy mass spectrometry (Q-Exactive HF). The samples were analyzed by targeted analysis while monitoring relevant peptides. The targeted analysis data were analyzed using the Skyline algorithm and manually curated for confident identifications and accurate quantification. The data were also searched against the human protein database using the Proteome Discoverer software to increase confidence in the assignment of the relevant peaks. Peptide intensities were normalized against the Total Ion Current Area per peptide for each protein. These values were averaged to get a single value per protein. Then, data from both groups (TDP-43^{deltaNLS} and NES) were normalized to the average of the NES group per protein before comparing between conditions to get fold changes.

Human tissue staining and quantification

Motor cortical tissues from four C9orf72 cases and 4 age-and sex-matched controls were from the Target ALS post-mortem core. Cases were included in triplicate in a tissue microarray (TMA) format (TMA available through the Target ALS Post-mortem core). For P-body immunohistochemistry analysis, sections were cut at 4 µm thickness onto superfrost charged slides and deparaffinized. Sections then underwent heat mediated antigen retrieval in citric acid (pH=6) for 15 minutes in a pressure cooker. Following this antibodies were incubated for 90 minutes at the following concentrations: DCP1A 1:100 (Abcam ab47811) and LSM14A 1:100 (Proteintech 18336-1-AP) for 1 hr at room temperature. Sections were stained using the Novolink Polymer detection system and 3,3'-Diaminobenzidine (DAB) chromogen was used and counterstaining was performed with haematoxylin, according to standard operating procedures. TDP-43 aptamer dual chromogenic stains were performed using our standard operating protocol freely available online [151] using the following concentrations: DCP1A 1:100, LSM14A 1:100, secondary anti-biotin Alkaline phosphatase antibody at 1:100 (30 minutes incubation). Digital analysis was performed using ImageJ analyze particle function. Images were batch processed as follows: Image>8-bit. Process>Sharpen. Image>Adjust>Brightness & Contrast>minimum = 53, maximum = 144. Set measurements> area and limit to threshold. Image>Adjust>Threshold>48-103. Analyze>Analyze Particles>Size=5-Infinity, Circularity = 0-1, Show = Bare Outlines, Display results, Add to Manager.

SLAM-seq and mRNA stability analysis

For metabolic RNA labeling of TDP-43^{deltaNLS}-expressing iPSC-derived neurons, the maturation medium was replaced with a medium containing 4sU (200 µM, T4509, Sigma). After 2 or 3 hours of labeling, cells were harvested with Tri-reagent (Sigma) followed by RNA extraction under reducing conditions and treatment with iodoacetamide (A3221, Sigma) as previously described [96]. RNA-seq libraries were prepared at the Crown Genomics Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science. A bulk adaptation of the MARS-Seq protocol

[152,153] was used to generate RNA-Seq libraries for SLAM-seq. Briefly, 30 ng of input RNA from each sample was barcoded during reverse transcription and pooled. Following Agencourt Ampure XP beads cleanup (Beckman Coulter), the pooled samples underwent second-strand synthesis and were linearly amplified by T7 in vitro transcription. The resulting RNA was fragmented and converted into a sequencing-ready library by tagging the samples with Illumina sequences during ligation, RT, and PCR. Libraries were quantified by Qubit and TapeStation as well as by qPCR for Human GAPDH gene as previously described [152,153]. Sequencing was done on a Nextseq 500, using Nextseq High Output 75 cycles kit mode, allocating 400M reads in total (Illumina). RNA-seq reads were aligned to the human genome (version GRCh38.99) with STAR (version 2.7.3a) [154] using parameters "--outFilterMismatchNmax 20 --outFilterScoreMinOverLread 0.4 --outFilterMatchNminOverLread 0.4 --outSAMattributes nM MD NH --alignEndsType Extend5pOfRead1 --quantMode GeneCounts". Bam files from STAR were processed with GRAND-SLAM (version 2.0.7b) [97] with trimming of 20 nucleotides in the 5' end due to higher mismatch rates across all nucleotides in this region. The error rate correction linear model was set to AT instead of the default TA, because the latter showed a higher error rate and variability across samples. Unlabeled samples without 4sU were used for the model estimation. The main output table of GRAND-SLAM ({prefix}.tsv) was used as input for analysis with grandR (version 0.2.5) [98] for half-life estimations as previously explained [155]. For analysis of P-body-enriched RNAs, we used a list of P-body-enriched RNAs (FDR < 0.05, log2FC > 1) from [99]. Final statistical testing was done with Prism (version 10).

Computational methods

Preprocessing of U2OS or neuron microscopic images

Day 8 and TDP-43^{deltaNLS} iPSC-derived neurons: Immunofluorescence confocal images were taken using a Spinning disc confocal microscope (see **Methods**). Each image (1,024x1,024 pixels, also referred as "site image") is dual-channel, holding the target marker as the first channel and the nucleus (Hoechst 33342) as the second. Each raw image contained on average ~17 individual cells, as detected using Cellpose[156] (V2.0.3, parameters: model type='nuclei', diameter=60) on the nucleus channel convolved with [[-1,-1,-1], [-1,25,-1], [-1,-1,-1]] for edges enhancement. First, raw pixels were scaled to a range of 0 to 1 using the rescale_intensity function from skimage python package (V0.21.0), with the in_range parameter sets to 0.5 percentile and 99.9 percentile of the pixels' intensity distribution and out_range sets to numpy.float32. Corrupted/out-of-focus sites were filtered (see **Methods**). Site image was then cropped to 64 equal-sized tiles (128x128 pixels). Each tile was resized to 100x100 pixels (using the resize function from skimage package V0.19.3, with anti_aliasing=True). Since images of neurons are highly sparse, all tiles were subjected to QC process, where we kept only valid tiles. See also **Supplementary Figure 2**.

U2OS: immunofluorescence confocal images (13,414x13,414 pixels) were taken using Zeiss LSM 900 confocal microscope (Zeiss). The images were downsampled by a factor of 2 (block_size=2) using the skimage.measure.block_reduce function (v0.19.3) with func=numpy.mean, followed by a min-max normalization to transform the pixel range between 0 to 1. Then, a stacking was made, along the channel axis, for each target marker and its corresponding nucleus (Hoechst 33342). Site images were then cropped into equal-sized tiles of 300x300 pixels each, and downsampled by a factor of 3 (using block_reduce, block_size=3, func=numpy.mean) to reach the size expected by the model, 2x100x100. The quality control applied here used the same criteria for a valid tile, using cell nuclei detection by employing Cellpose.

Immunofluorescent images of **day 18 iPSC-derived neurons** (1,080x1,080 pixels) were taken using the Opera Phenix Plus microscope (Revvity). Their frame was cropped by 28 pixels from each margin to match 1,024x1,024 pixels. The rest of the preprocessing was made as described for the day 8 neurons data.

Immunofluorescent confocal images of **sALS patient-derived motor neurons (day 60)** (1,024x1,024 pixels) were taken using Zeiss LSM 980 confocal microscope (Zeiss), with a larger zoom magnitude (X63) than the rest of the data (X60). Hence, each image was cropped to 49 equal-sized tiles (146x146 pixels) and resized to 100x100 pixels as described before. Neither intensity rescaling nor filtering of unfocused images were needed.

Quality control reports of all experimental and imaging data can be found here.
Site images that contained too much debris or were out of focus were filtered out using marker-dependent self-curated thresholds for Brenner's Gradient score [157], an approach for focus evaluation. Here we quantified the Brenner's Gradient score by summing its values computed on

both the original image and its transpose, enabling a comprehensive evaluation of the image in both horizontal and vertical dimensions. Site images that passed the Brenner's Gradient score filtering for both the target and the nucleus channels were then cropped to tiles. A valid tile is considered a tile with at least a single whole nucleus with fewer than 6 nuclei. To filter for non-valid tiles, we performed cell nuclei detection by employing Cellpose (with the same parameters as above).

To determine whether a nucleus appears fully within a tile, we first computed its polygon using the Python package Shapely (version 2.0.1). We also defined a polygon for the entire tile boundary, applying a padding of 2 units on each side. If the nucleus polygon was completely contained within the tile's polygon, we classified it as a whole nucleus (**Supplementary Figure 2**). The neuron count in our data, therefore, reflects the number of whole nuclei captured within each valid tile.

In our day 8 and TDP-43^{deltaNLS} neurons data, 7 tiles out of 64 (on average, per site) passed the QC, with 1.5 nuclei on average. In total, 2,622,157 tiles passed the QC, depicting ~916,705 neurons. In the day 18 neurons data, one tile on average passed QC (6 neurons on a site on average). In total, 218,540 valid tiles for ~80,859 neurons.
In the day 60 neurons data, 5,888 valid tiles, depicting ~1983 neurons, passed the QC.

Localization encoding ViT-based model

A ViT model with linear head and cross-entropy loss (for full architecture see **Supplementary Materials**) was trained from scratch on the publicly available OpenCell dataset (1,100,253 cropped images [44]). The data was split into 70%, 15%, 15% for training, validation, and testing data, respectively. Model parameters: embed_dim=192, patch_size=14, in_chans=2, output_dim=1311. The input for the model are dual-channel tiles of 100x100 (a 3D vector sized (2,100,100)). Model training converged after 44 epochs (**Supplementary Figure 4**). In training, the following augmentations were applied with a probability of 0.5 for each: horizontal flip, vertical flip, and rotation of 90, 180, or 270 degrees. Train parameters: we used the AdamW optimizer with adjustable weight_decay for non bias nor Norm parameters. A cosine scheduler was used for controlling the weight decay (base value=0.04, final value=0.4, number of epochs=300) and the learning rate (base value=0.0008*batch_size/256, final value= 0.000001, number of epochs=300, number of warmup epochs=5). The batch size was set to 350 site images, each containing 7 tiles on average. Early stopping was applied after ten epochs in which the validation loss showed no improvement. A grad scaler (torch.cuda.amp.GradScaler) was used for calculating the gradients. Unscaled gradients' norms were clipped (torch.nn.utils.clip_grad_norm_) with max_norm=3.
Model's Accuracy was calculated on the test set using sklearn.metrics.accuracy_score function. The true labels given to the function were the proteins' identification. Predicted labels were acquired by applying softmax (torch.nn.functional.softmax) to the model's output and selecting the label that had the highest probability (using torch.argmax) per sample.

Perturbation learning ViT-based model using InfoNCE contrastive loss

For perturbation learning tasks, we first initialized the weights of a ViT model with the layers' weights from the pre-trained model (that was trained for localization encoding), while partially freezing layers

from being updated during fine-tuning (see **Methods**, **Supplementary Figure 5**). The model was trained using the contrastive loss function termed infoNCE[40] in a supervised manner [41], which learns to simultaneously minimize the distance in the latent feature space between samples with similar perturbation effect and maximize the distance between dissimilar ones. Positive/negative pairs, definition, and matching: In the supervised approach, positive pairs are tiles (different fields of view of the same organelle) with similar perturbation effects (treated as different augmentations of the same perturbation), while negative pairs underwent a different perturbation that might present similar, dissimilar, or no effects. Given a tile image x, which serves as an anchor, additional tile images that are considered positive $x^{+}$and negatives x^{-} are matched to it. This matching is made by randomly sampling (with repetitions) one positive and five negative tiles of the same organelle. The positive sample is an image tile of the same organelle and under the same perturbation (as the anchor). The negative samples are image tiles of the same organelle, but under different perturbations than the anchor. Notably, the five negative tiles are not restricted to sharing the same perturbation.

The seven matched tiles are fed during training to the ViT and their predicted embeddings are sent as an input to the InfoNCE loss function (**Figure 1B**, **Supplementary Figure 3**). The InfoNCE loss function is calculated as follows:

$$L_{infoNCE} = -\ \mathbb{E}[\log \frac{f(x,x^{+})}{f(x,x^{+}) + \sum_{i=1}^{K} f(x,x^{-}{}_{i})}],$$

$$where\ f(x,x') = \exp(\frac{sim(x,x')}{\tau}),\ sim(x,x') = \frac{x \cdot x'}{||x|| \cdot ||x'||},\ \tau = 0.1, K = 5$$

Pulling together the anchor x and the positive x^{+} reduces natural variation in the dataset, such as batch/well effect. Pushing away the negative samples x^{-} aims to refine the representation of a sample to hold real perturbational effects on the organelle. Notably, each anchor could serve as a positive or negative sample for other anchors, which touches on a powerful mechanism within contrastive loss frameworks; the triangle relationship or transitivity of sample comparisons. This dynamic is beneficial for learning real perturbation effects since, when an anchor is pulled closer to its positive and pushed further from its negatives, this relationship also influences other triplets where the same anchor appears, which may refine their representations to be more accurate, consistent, and robust. This means that the model is not just learning pairwise relationships between the anchor-positive and anchor-negative, but also indirectly learning relationships between other samples, through the process of comparison. When anchors are shared between multiple samples, it also prevents the model from overfitting to a small set of pairwise relationships and therefore the model generalizes better.

Although the labels, that are used for matching an anchor with positive and negative samples, are indications of the experimental conditions performed on the cells, they are considered as “weak labels”. Since not all perturbations have a consistent effect on the organelle topographies, one may think that differing samples solely for being labeled differently might lead to false relationships

between perturbations. Yet, a contrastive framework with negatives mitigates this issue. When a negative sample shares a similar perturbation effect with its anchor, it also shares it with its positive sample. Hence, pushing away the negative sample means pushing the positive sample away from the anchor as well. This loss function prioritizes maintaining the similarity between an anchor and its positive sample rather than differing an anchor from its negative. Therefore, pushing away a negative sample from its anchor would be mitigated if they both shared a similar perturbation effect, despite being labeled differently. This approach allows the model to seek real differences in perturbation effects, while acknowledging that there might be perturbations with similar effects or with no effect at all.

Training procedure and parameters: Training parameters as the pre-trained model, except for batch_size=750 files and the dimension of the output was changed to 128. Model parameters: embed_dim=192, patch_size=14, in_chans=2, output_dim=128. The training was performed on 2 batches (820,972 tiles) of our in-house curated dataset of perturbed neurons (day 8) and converged after 67 epochs (**Supplementary Figure 6**). Each input is a dual-channel image (100x100 pixels). The data was split into train (70%), validation (15%), and test (15%) sets. The inference and data presented in the results figures were not part of model development, and/or on the test set. The training dataset includes 25 organelles in 9 different cell lines/conditions. The labels contain the name of the organelle stained, the cell line, and whether it's untreated or stressed.

Partial fine-tuning: the angle-metric approach for freezing layers in perturbation learning

Fine tuning is the adaptationof a pre-trained model, initially trained on massive amounts of labeled data, to perform well on a different but related problem. This involves selectively updating the model's parameters to capture the nuances of the new task. [158]. The "localization encoding model" is extended to perform "perturbation learning" in neurons (**Figure 1B**) using partial fine-tuning (**Supplementary Figure 5**). All weights of the transformer encoder are copied to the new model, except for the head, but not all model layers were re-trained with data of neurons. The training process is done by utilizing a data-drive approach for selecting which layers to freeze and which to fine-tune with our perturbed neurons. Frozen layers are not being updated during the fine-tuning process. For selecting which layers to freeze we first trained a perturbation learning model (see **Methods**) without freezing any layer. Then, we calculated for each layer how much it deviated from the pre-trained model weights using the angle metric [46]:

$$\theta(L) = \arccos\left(\frac{W^L_p * W^L_t}{||W^L_p||_2 \, ||w^L_t||_2}\right)$$

where W^L_p are the weights of layer L in the pre-trained model, and W^L_t are the weights of the same layer L in the model after no-freezing fine tuning.

Finally, since the tasks are different, yet share basic commonalities, we picked the least deviated layers (50% of the layers) to be the ones we freeze.

Image Encoding

While our training pipeline runs the training consecutively (the localization encoding model and then the perturbation learning model), the inference step is done only with the perturbation learning model. A 100x100 pixels image (target and nucleus channel) is fed to the model to extract a numerical representation of 128 values that capture the phenotype of the organelle presented in the image.

UMAP, dimensionality reduction of latent space and 2D clustering

To cluster images of organelles from different experimental perturbations, the image representations are subjected to dimensionality reduction using the Uniform Manifold Approximation and Projection (UMAP) algorithm (umap-learn v0.5.3, random_state=1). The first 2 UMAP components were plotted in a scatter plot to visualize clusters and colored according to their experimental perturbation. Our work presents different types of UMAPs based on the model output: single organelle plots (**Figure 1D**, **Figure 2A**, **Figure 3C**, **Figure 5C**, **Supplementary Figure 9**, **Supplementary Figure 12**, and **Supplementary Figure 20B**), multiple organelles plots (**Figure 1C** and **Supplementary Figure 7**,**8**) and plots presenting synthetic superimposed data of multiple organelles under the same condition (**Figure 5B**,**D**, **Supplementary Figure 17**, **Supplementary Figure 18B-E**, and **Supplementary Figure 19A-E**).

Clustering scoring: Adjusted Rand Index

To evaluate the clustering quality within the latent space, as presented in the UMAPs, we used the supervised measure called Adjusted Rand Index (ARI). The ARI was calculated using the adjusted_rand_score function (provided by sklearn.metrics, version=1.2.2) on the true labels together with the constrained K-Means' predicted labels. The ARI is bounded between -0.5 to 1, whereby a value of 1 indicates a perfect clustering, a value close to 0 indicates random labeling (independently of the number of clusters and samples), and -0.5 for especially discordant clusterings. The ARI presented is rounded up to 3 digits after the decimal point.

Since ARI is a supervised measure, we needed to infer the predicted label from the closest neighbors in the latent space. For that purpose, we clustered the embeddings using the constrained KMeans algorithm [159] (provided by k-means-constrained package, version=0.7.3) with the following parameters: n_clusters= number of true unique labels, size_min=10% of the input data, n_init=10, random_state=42. Except for the whole organellome landscape, for which we set size_min to be 2% of the input data since a high number of clusters is expected. We used the constrained version of KMeans due to its ability to overcome outliers, which confused the clustering process of the original KMeans algorithm (provided by sklearn.cluster, version=1.2.2), via the additional minimum cluster size parameter it provides.

Synthetic Superposition

Assuming an experiment with K perturbations (e.g., conditions/cell lines). Images of m organelles are obtained under each perturbation. Let $x_{j,k} \in \mathbb{N}^{128}$ be latent representation of a given image of

organelle j and perturbation k, where $j \in [1, m]$ and $k \in [1, K]$. Superpositioning of m organelles from the same perturbation into a single data point as follows:

Synthetic Superposition

Input: Set of perturbations $\{k\}_{k=1}^{K}$; Number of organelles m; Collection of *NOVA* representations where each $x_{j,k} \in N^{128}$ encodes an image of organelle j under perturbation k

Output: Group of superimposed representations $\{x'\}$, with each $x' \in N^{m \cdot 128}$

Let: $O_k = \{x_{j,k}\}_{j=1}^{m}$, representations of m distinct organelles under the same perturbation k

```
for each k ∈ {1, ..., K} do
    while at least one O_k remains do
        Randomly sample (without replacement) an O_k
        # Concatenate the representations in the selected O_k
        x' = (x_{1,k}, x_{2,k}, ..., x_{m,k}) ∈ N^{m·128}
        Add x' to the output group {x'}
    end while
end for
return {x'}
```

For visualization purposes, we reduce the dimensionality of each x' to a vector of size 2 using the UMAP algorithm.

Evaluating the significance of a perturbation's effect via ARI on the latent space

For each perturbation, we aimed to statistically test which organelles have changed when compared to its control. Specifically, we sought to compare the separation between perturbed and unperturbed representations against the natural variability (or "noise") in our data, where a separation was quantified using ARI scores calculated on the embeddings (see **Methods**).

First, we calculated ARI scores for each organelle with and without a perturbation, across two replicates from the same batch, yielding four ARI scores per perturbation (see **Supplementary Figure 10**). These ARI scores quantify the extent of separation between the perturbed and control in the model's latent space.

To establish a baseline measure of inherent data variability, we next used only the control samples to compute ARI scores. For each organelle in the control condition, we randomly split the samples from the same replicate into two artificial sub-groups, and calculated the ARI scores between these sub-groups across replicates (also yielding four ARI scores). These scores reflect the inherent variation in the data, that can stem from using two different physical replicates, and serve as the measure of noise in the system.

This procedure was repeated on multiple batches. For each organelle, we then applied an unpaired one-sided t-test on the ARI scores from all batches to measure the significance of the perturbation effect when comparing it to the noise (using scipy.stats, version 1.12.0). Significant results indicate that the higher observed differences are likely due to the perturbation rather than inherent variability.

Organelle ranking barplot

The organellome can be ranked based on the level of differences detected between two perturbations measured by the ARI score (see **Methods**). Here, for each organelle we show its ARI scores between the perturbation and the control, between the control to itself (the noise), and the level of significance, if the difference was found significant. The organelles are ordered by the median ARI score of the perturbation vs the control, higher means better separation between the perturbation and the control for the organelle. Then the organelles are split into two groups, significant and non-significant effects.

Bubble plot

For showing multiple perturbations in comparison to a control perturbation (e.g., wild-type), all at once, we used the bubble plot visualization. For each organelle and perturbation, we calculated the ARI score (see **Methods**) compared to a different perturbation that serves as the baseline, followed by the same statistical test as described above for the ranking barplot. In addition, for evaluating the effect size, we utilize the Cliff's delta metric [107], a non-parametric measure that quantifies the difference between two groups' ARI. A value of 1 means large differences, while 0 suggests no differences.

CellProfiler analysis

For organellome-wide analysis, CellProfiler (version 4.2.1; jdk version 14.0.1; python version 3.8) [43] objects were identified in images in three different size ranges to cover all possible organellar stainings; in addition, 'secondary object' identification was used for more diffuse cellular stains. Entire site images (1,024x1,024) were used as input images. All main object-related CellProfiler features relating to area, intensity, distribution, neighbors and texture were then extracted. Object measurements were averaged per image and concatenated across object types (using pandas, version 1.4.4). For directed analysis of P-bodies, a specific pipeline was applied that identifies P-bodies in the correct size range before extracting object features. In case of TDP-43 enrichment analysis, P-body objects were expanded by 2 pixels and subtracted from the expanded object to get mean intensity values from the surrounding cytoplasm. Partition coefficient was then calculated by dividing mean intensity in P-bodies by mean intensity in surrounding cytoplasm. Consequent statistical testing was done with Prism (version 10). After removing constant features and features containing missing values, the remaining 347 features calculated with Cellprofiler as described above were used as features for ARI calculations, statistical testing, boxplot ranking and UMAP plotting as described above.

Prediction of protein-protein interactions

SpeedPPI [144] was implemented to predict interaction strength between TDP-43 and P-body proteins. Briefly, the pipeline generates multiple sequence alignments (MSAs) for TDP-43 and its potential interactors using HHlibs. It offers multiple operational modes, including "all-vs-all" and "some-vs-some"; we utilized the "some-vs-some" mode to compare TDP-43 against a defined list of target proteins. The MSAs are then subjected to structure prediction and validation using AlphaFold2

[89], providing high-confidence structural models for each protein. These validated structures are fed into the SpeedPPI predictor, a machine learning-based model that integrates several features. The final output is a structural modeling prediction that illustrates the potential interaction interfaces between TDP-43 and its interactors. In addition, a pDockQ score is generated to estimate the confidence of the predicted interaction.

Software and hardware

All deep-learning architectures were implemented in Pytorch version 2.2.2[160], cuda version 11.8 on Python v3.9.19. The training was performed on an NVIDIA A40 (48GB GPU) card.

Code availability

The code can be found on our Github page: https://github.com/Sagykri/NOVA

Data availability

Proximity labeling proteomics data are available in ProteomeXchange (PXD056534).
SLAM-seq RNA-seq data are available in GEO (GSE279980).
The imaging dataset can be downloaded via Amazon S3 API. Bucket name: organellomics, region: eu-west-1. A script for that can be found on our Github repository.

Figure design

We placed and organized all the figures by using Adobe Illustrator software, which was also used for generating **Figures 1A**, **Supplementary Figures 16A** and **18A**. The graphs were generated either by Prism (version 5.9.1 or 10) or via python (v3.9.19; matplotlib (v3.8.4) and seaborn (v0.13.2)). Representative confocal images were generated by FiJi software v. 1.52p. or ZEN software. We generated the graphical abstract, **Figures 1B**, **4O**, **5A**, and **Supplementary Figures 1-3**, **5**, and **10** by Biorender.com.

Acknowledgments

We wish to thank Leeat Keren (WIS), Maya Schuldiner (WIS), Eran Segal (WIS) Sasha Devore (WIS), Ernest Fraenkel (MIT), Hirofumi Kobayashi (Chan Zuckerberg Biohub), Assaf Zaritzky (BGU), Tomer Lapidot (UPENN), Ophir Shalem (UPENN) and Johnathan Cooper Knock (U Sheffield) for advice during project development and on the manuscript draft. We thank Michael Ward (NINDS, NIH) and Bill Skarnes (Jackson labs) for early access to iNDI lines Nancy Kadersha Paul Anderson and Pavel Ivanov (HMS) for G3BP1 and YFP-DCP1A U2OS cells. Shai Bagon (WIS) and Ron Rotkofp (WIS) for advice in statistical and computational sides. Alon Savidor (WIS) and Yishai Levin (WIS) for Mass spectrometry. For the RNA-seq procedure for SLAM-seq, we wish to thank Inbal Bolocan Nachman (WIS). We thank Chen Eitan, Aviad Siany, Jazz Lubliner, Yahel Cohen, Emmanuel Amzalleg, and Lior Lin for preliminary experiments that are not in the scope of the final manuscript.

EH is the Mondry Family Professorial Chair and Head of the Andrea L. and Lawrence A. Wolfe Family Center for Research on Neuroimmunology and Neuromodulation. Funding for research: Binational Science Foundation (BSF); Association Francaise Contre les Myopathies (AFM); Amyotrophic Lateral Sclerosis Association (ALSA); Target ALS; Israel Science Foundation (ISF 3497/21, 424/22, 494/24); ALS Canada; Minna-James-Heineman Stiftung through Minerva, Minerva Foundation, with funding from the Federal German Ministry for Education and Research; Robert Packard Center for ALS Research at Johns Hopkins; McGill University; EU - ERA-Net; Radala Foundation for ALS Research; Additional support generously provided by the Kekst Family Institute for Medical Genetics. Weizmann SABRA - Yeda-Sela - WRC Program, the Estate of Emile Mimran, and The Maurice and Vivienne Wohl Biology Endowment. Nella and Leon Benoziyo Center for Neurological Diseases. Goldhirsh-Yellin Foundation. Dr. Sydney Brenner and friends. Weizmann - Center for Research on Neurodegeneration. Redhill Foundation – Sam and Jean Rothberg Charitable Trust Dr. Dvora and Haim Teitelbaum Endowment Fund. LM is funded by a Minerva Postdoctoral Fellowship. YDM is funded by the Weizmann - CNRS (Centre National de la Recherche Scientifique) Collaboration Program.

References

1. Gottschling, D. E. & Nyström, T. The Upsides and Downsides of Organelle Interconnectivity. *Cell* **169**, 24–34 (2017).
2. Zung, N. & Schuldiner, M. New horizons in mitochondrial contact site research. *Biol. Chem.* **401**, 793–809 (2020).
3. Gu, J. *et al.* Mapping multimodal phenotypes to perturbations in cells and tissue with CRISPRmap. *Nat. Biotechnol.* 1–15 (2024).
4. Itzhak, D. N. *et al.* A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. *Cell Rep.* **20**, 2706–2718 (2017).
5. Borner, G. H. H. Organellar Maps Through Proteomic Profiling - A Conceptual Guide. *Mol. Cell. Proteomics* **19**, 1076–1087 (2020).
6. Schessner, J. P., Albrecht, V., Davies, A. K., Sinitcyn, P. & Borner, G. H. H. Deep and fast label-free Dynamic Organellar Mapping. *Nat. Commun.* **14**, 5252 (2023).
7. Itzhak, D. N., Schessner, J. P. & Borner, G. H. H. Dynamic organellar maps for spatial proteomics. *Curr. Protoc. Cell Biol.* **83**, e81 (2019).
8. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, (2017).
9. Kobayashi, H., Cheveralls, K. C., Leonetti, M. D. & Royer, L. A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **19**, 995–1003 (2022).
10. Spitzer, H., Berry, S., Donoghoe, M., Pelkmans, L. & Theis, F. J. Learning consistent subcellular landmarks to quantify changes in multiplexed protein maps. *Nat. Methods* **20**, 1058–1069 (2023).
11. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

12. Heinrich, L. *et al.* Whole-cell organelle segmentation in volume electron microscopy. *Nature* **599**, 141–146 (2021).

13. Lu, A. X., Kraus, O. Z., Cooper, S. & Moses, A. M. Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.* **15**, e1007348 (2019).

14. Husain, S. S. *et al.* Single-cell subcellular protein localisation using novel ensembles of diverse deep architectures. *Commun Biol* **6**, 489 (2023).

15. Long, W., Yang, Y. & Shen, H.-B. ImPLoc: a multi-instance deep learning model for the prediction of protein subcellular localization based on immunohistochemistry images. *Bioinformatics* **36**, 2244–2250 (2020).

16. Kim, K.-M., Son, K. & Palmore, G. T. R. Neuron Image Analyzer: Automated and Accurate Extraction of Neuronal Data from Low Quality Images. *Sci. Rep.* **5**, 17062 (2015).

17. Ascoli, G. A., Donohue, D. E. & Halavi, M. NeuroMorpho.Org: a central resource for neuronal morphologies. *J. Neurosci.* **27**, 9247–9251 (2007).

18. Oberlaender, M., Bruno, R. M., Sakmann, B. & Broser, P. J. Transmitted light brightfield mosaic microscopy for three-dimensional tracing of single neuron morphology. *J. Biomed. Opt.* **12**, 064029 (2007).

19. Moshkov, N. *et al.* Learning representations for image-based profiling of perturbations. *Nat. Commun.* **15**, 1594 (2024).

20. Hirose, T., Ninomiya, K., Nakagawa, S. & Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* (2022) doi:10.1038/s41580-022-00558-8.

21. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

22. Brangwynne, C. P. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.*

203, 875–881 (2013).

23. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
24. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 196–213 (2021).
25. Platt, F. M., d'Azzo, A., Davidson, B. L., Neufeld, E. F. & Tifft, C. J. Lysosomal storage diseases. *Nat Rev Dis Primers* **4**, 27 (2018).
26. Cox, T. M. & Cachón-González, M. B. The cellular pathology of lysosomal diseases. *J. Pathol.* **226**, 241–254 (2012).
27. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat Rev Dis Primers* **2**, 16080 (2016).
28. Schapira, A. H. V. Mitochondrial disease. *Lancet* **368**, 70–82 (2006).
29. Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
30. Mizushima, N. & Levine, B. Autophagy in Human Diseases. *N. Engl. J. Med.* **383**, 1564–1576 (2020).
31. Nedelsky, N. B. & Taylor, J. P. Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 272–286 (2019).
32. Alberti, S. & Dormann, D. Liquid-Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
33. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv [cs.CV]* (2020).
34. Kim, V., Adaloglou, N., Osterland, M., Morelli, F. & Zapata, P. A. M. Self-supervision advances morphological profiling by unlocking powerful image representations. (2023) doi:10.1101/2023.04.28.538691.
35. Doron, M. *et al.* Unbiased single-cell morphology with self-supervised vision transformers. *bioRxiv* (2023).

36. Kraus, O. *et al.* Masked autoencoders for microscopy are scalable learners of cellular biology. *arXiv [cs.CV]* (2024).

37. Atmaramani, R. *et al.* Deep Learning Analysis on Images of iPSC-derived Motor Neurons Carrying fALS-genetics Reveals Disease-Relevant Phenotypes. *bioRxiv* (2024).

38. Natekar, P., Wang, Z., Arora, M., Hakozaki, H. & Schöneberg, J. Self-supervised deep learning uncovers the semantic landscape of drug-induced latent mitochondrial phenotypes. *bioRxiv* (2023) doi:10.1101/2023.09.13.557636.

39. Perakis, A. *et al.* Contrastive learning of single-cell phenotypic representations for treatment classification. in *Machine Learning in Medical Imaging* 565–575 (Springer International Publishing, Cham, 2021).

40. van den Oord, A., Li, Y. & Vinyals, O. Representation learning with Contrastive Predictive Coding. *arXiv [cs.LG]* (2018).

41. Khosla, P. *et al.* Supervised Contrastive Learning. *arXiv [cs.LG]* (2020).

42. Bray, M.-A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).

43. Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).

44. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).

45. Fernandopulle, M. S. *et al.* Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. *Curr. Protoc. Cell Biol.* **79**, e51 (2018).

46. Ye, P. *et al.* Partial Fine-tuning: A successor to full fine-tuning for vision transformers. *arXiv [cs.CV]* (2023).

47. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

48. West, J. A. *et al.* Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817–830 (2016).

49. Naganuma, T. *et al.* Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034 (2012).

50. Arseni, D. *et al.* Heteromeric amyloid filaments of ANXA11 and TDP-43 in FTLD-TDP Type C. *Nature* (2024) doi:10.1038/s41586-024-08024-5.

51. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26**, 668–679 (2016).

52. Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **10**, 430–436 (2009).

53. Lai, A., Valdez-Sinon, A. N. & Bassell, G. J. Regulation of RNA granules by FMRP and implications for neurological diseases. *Traffic* **21**, 454–462 (2020).

54. Li, W. & Chou, I. N. Effects of sodium arsenite on the cytoskeleton and cellular glutathione levels in cultured cells. *Toxicol. Appl. Pharmacol.* **114**, 132–139 (1992).

55. Kedersha, N. *et al.* Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* **169**, 871–884 (2005).

56. Chazotte, B. Labeling nuclear DNA with hoechst 33342. *Cold Spring Harb. Protoc.* **2011**, db.prot5557 (2011).

57. Chazotte, B. Labeling mitochondria with MitoTracker dyes. *Cold Spring Harb. Protoc.* **2011**, 990–992 (2011).

58. Wurm, C. A. *et al.* Nanoscale distribution of mitochondrial import receptor Tom20 is adjusted to cellular conditions and exhibits an inner-cellular gradient. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13546–13551 (2011).

59. Turner, D. L., Korneev, D. V., Purdy, J. G., de Marco, A. & Mathias, R. A. The host exosome pathway underpins biogenesis of the human cytomegalovirus virion. *Elife* **9**, e58288 (2020).

60. Nava, C. *et al.* Hypomorphic variants of cationic amino acid transporter 3 in males with autism spectrum disorders. *Amino Acids* **47**, 2647–2658 (2015).

61. Chazotte, B. Labeling cytoskeletal F-actin with rhodamine phalloidin or fluorescein phalloidin for imaging. *Cold Spring Harb. Protoc.* **2010**, db.prot4947 (2010).

62. Kageyama, S. *et al.* p62/SQSTM1-droplet serves as a platform for autophagosome formation and anti-oxidative stress response. *Nat. Commun.* **12**, 16 (2021).

63. Gowrishankar, S., Wu, Y. & Ferguson, S. M. Impaired JIP3-dependent axonal lysosome transport promotes amyloid plaque pathology. *J. Cell Biol.* **216**, 3291–3305 (2017).

64. Koyano, F. *et al.* Parkin-mediated ubiquitylation redistributes MITOL/March5 from mitochondria to peroxisomes. *EMBO Rep.* **20**, e47728 (2019).

65. Tsygankova, O. M. & Keen, J. H. A unique role for clathrin light chain A in cell spreading and migration. *J. Cell Sci.* **132**, (2019).

66. Xiang, Q., Ji, S.-D., Zhang, Z., Zhao, X. & Cui, Y.-M. Identification of ITGA2B and ITGB3 Single-Nucleotide Polymorphisms and Their Influences on the Platelet Function. *Biomed Res. Int.* **2016**, 5675084 (2016).

67. Mobarrez, F. *et al.* Microparticles in the blood of patients with systemic lupus erythematosus (SLE): phenotypic characterization and clinical associations. *Sci. Rep.* **6**, 36025 (2016).

68. Srivastava, M., Fleming, P. J., Pollard, H. B. & Burns, A. L. Cloning and sequencing of the human nucleolin cDNA. *FEBS Lett.* **250**, 99–105 (1989).

69. Fox, A. H., Bond, C. S. & Lamond, A. I. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell* **16**, 5304–5315 (2005).

70. Fox, A. H. *et al.* Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12**, 13–25 (2002).

71. Tsuji, H. *et al.* Epitope mapping of antibodies against TDP-43 and detection of protease-resistant fragments of pathological TDP-43 in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Biochem. Biophys. Res. Commun.* **417**, 116–121 (2012).

72. Calvio, C., Neubauer, G., Mann, M. & Lamond, A. I. Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *RNA* **1**, 724–733 (1995).

73. Yang, L., Gal, J., Chen, J. & Zhu, H. Self-assembled FUS binds active chromatin and regulates gene transcription. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17809–17814 (2014).

74. Weis, K. *et al.* Retinoic acid regulates aberrant nuclear localization of PML-RAR alpha in acute promyelocytic leukemia cells. *Cell* **76**, 345–356 (1994).

75. Eystathioy, T. *et al.* The GW182 protein colocalizes with mRNA degradation associated proteins hDcp1 and hLSm4 in cytoplasmic GW bodies. *RNA* **9**, 1171–1173 (2003).

76. Tourrière, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell Biol.* **160**, 823–831 (2003).

77. Liao, Y.-C. *et al.* RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether. *Cell* **179**, 147–164.e20 (2019).

78. Daigle, J. G. *et al.* Pur-alpha regulates cytoplasmic stress granule dynamics and ameliorates FUS toxicity. *Acta Neuropathol.* **131**, 605–620 (2016).

79. Molitor, L. *et al.* Depletion of the RNA-binding protein PURA triggers changes in posttranscriptional gene regulation and loss of P-bodies. *Nucleic Acids Res.* **51**, 1297–1316 (2023).

80. Du, M., Ea, C.-K., Fang, Y. & Chen, Z. J. Liquid phase separation of NEMO induced by polyubiquitin chains activates NF-κB. *Mol. Cell* **82**, 2415–2426.e5 (2022).

81. Liu, M. *et al.* KIF5A-dependent axonal transport deficiency disrupts autophagic flux in trimethyltin chloride-induced neurotoxicity. *Autophagy* **17**, 903–924 (2021).

82. Ray, S. *et al.* α-Synuclein aggregation nucleates through liquid-liquid phase separation. *Nat. Chem.* **12**, 705–716 (2020).

83. Bayer, T. A. *et al.* Neural expression profile of alpha-synuclein in developing human cortex. *Neuroreport* **10**, 2799–2803 (1999).

84. Sampedro, M. N., Bussineau, C. M. & Cotman, C. W. Postsynaptic density antigens: preparation and characterization of an antiserum against postsynaptic densities. *J. Cell Biol.* **90**, 675–686 (1981).

85. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).

86. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (1 2014).

87. Gasset-Rosa, F. *et al.* Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. *Neuron* **102**, 339–357.e7 (2019).

88. Bryant, P. & Noé, F. Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. *PLoS Comput. Biol.* **20**, e1012253 (2024).

89. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).

90. Burke, D. F. *et al.* Towards a structurally resolved human protein interaction network. *Nat. Struct. Mol. Biol.* **30**, 216–225 (2 2023).

91. Marmor-Kollet, H. *et al.* Spatiotemporal Proteomic Analysis of Stress Granule Disassembly Using APEX Reveals Regulation by SUMOylation and Links to ALS Pathogenesis. *Mol. Cell* **80**, 876–891.e6 (2020).

92. Standart, N. & Weil, D. P-bodies: Cytosolic droplets for coordinated mRNA storage. *Trends Genet.* **34**, 612–626 (2018).

93. Jain, S. & Parker, R. The discovery and analysis of P Bodies. *Adv. Exp. Med. Biol.* **768**, 23–43 (2013).

94. Dewey, C. M. *et al.* TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Mol. Cell. Biol.* **31**, 1098–1108 (2011).

95. Luo, Y., Na, Z. & Slavoff, S. A. P-Bodies: Composition, Properties, and Functions. *Biochemistry* **57**, 2424–2431 (5 2018).

96. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).

97. Jürges, C., Dölken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218–i226 (2018).

98. Rummel, T., Sakellaridi, L. & Erhard, F. grandR: a comprehensive package for nucleotide conversion RNA-seq data analysis. *Nat. Commun.* **14**, 3559 (2023).

99. Hubstenberger, A. *et al.* P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol. Cell* **68**, 144–157.e5 (10 2017).

100. Ramos, D. M., Skarnes, W. C., Singleton, A. B., Cookson, M. R. & Ward, M. E. Tackling neurodegenerative diseases with genomic engineering: A new stem cell initiative from the NIH. *Neuron* **109**, 1080–1083 (2021).

101. Waibel, S., Neumann, M., Rabe, M., Meyer, T. & Ludolph, A. C. Novel missense and truncating mutations in FUS/TLS in familial ALS. *Neurology* **75**, 815–817 (2010).

102. Rutherford, N. J. *et al.* Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS Genet.* **4**, e1000193 (2008).

103. Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672 (2008).

104. Kenna, K. P. *et al.* Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *J. Med. Genet.* **50**, 776–783 (2013).

105. Freischmidt, A. *et al.* Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat. Neurosci.* **18**, 631–636 (2015).

106. Zarranz, J. J. *et al.* The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **55**, 164–173 (2004).

107. Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.* **114**, 494–509 (1993).

108. Smith, B. N. *et al.* Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. *Sci. Transl. Med.* **9**, (2017).

109. Taylor, J. P., Brown, R. H., Jr & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539**, 197–206 (2016).

110. Workman, M. J. *et al.* Large-scale differentiation of iPSC-derived motor neurons from ALS and control subjects. *Neuron* **111**, 1191–1204.e5 (2023).

111. Bussi, C. *et al.* Stress granules plug and stabilize damaged endolysosomal membranes. *Nature* **623**, 1062–1069 (2023).

112. Amen, T. & Kaganovich, D. Stress granules inhibit fatty acid oxidation by modulating mitochondrial permeability. *Cell Rep.* **35**, 109237 (2021).

113. Qin, W. *et al.* Dynamic mapping of proteome trafficking within and between living cells by TransitID. *Cell* **186**, 3307–3324.e30 (2023).

114. Stoecklin, G. & Kedersha, N. Relationship of GW/P-bodies with stress granules. *Adv. Exp. Med. Biol.* **768**, 197–211 (2013).

115. Cohen, S., Valm, A. M. & Lippincott-Schwartz, J. Interacting organelles. *Curr. Opin. Cell Biol.* **53**, 84–91 (2018).

116. Donahue, E. K. F., Ruark, E. M. & Burkewitz, K. Fundamental roles for inter-organelle communication in aging. *Biochem. Soc. Trans.* **50**, 1389–1402 (2022).

117. Rossini, M., Pizzo, P. & Filadi, R. Better to keep in touch: investigating inter-organelle cross-talk. *FEBS J.* **288**, 740–755 (2021).

118. Schrader, M., Kamoshita, M. & Islinger, M. Organelle interplay-peroxisome interactions in health and disease. *J. Inherit. Metab. Dis.* **43**, 71–89 (2020).

119. Mitrea, D. M. *et al.* Methods for Physical Characterization of Phase-Separated Bodies and

Membrane-less Organelles. *J. Mol. Biol.* **430**, 4773–4805 (2018).

120. Fleming, A. *et al.* The different autophagy degradation pathways and neurodegeneration. *Neuron* **110**, 935–966 (3 2022).
121. Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787–795 (10 2006).
122. Root, J., Merino, P., Nuckols, A., Johnson, M. & Kukar, T. Lysosome dysfunction as a cause of neurodegenerative diseases: Lessons from frontotemporal dementia and amyotrophic lateral sclerosis. *Neurobiol. Dis.* **154**, (7 2021).
123. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
124. Gut, G., Herrmann & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, (2018).
125. Singh, S. P. *et al.* 3D Deep Learning on Medical Images: A Review. *Sensors* **20**, (2020).
126. Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
127. Sun, J., Tárnok, A. & Su, X. Deep learning-based single-cell optical image studies. *Cytometry A* **97**, 226–240 (2020).
128. Lavitt, F., Rijlaarsdam, D. J., van der Linden, D., Weglarz-Tomczak, E. & Tomczak, J. M. Deep learning and transfer learning for automatic cell counting in microscope images of human cancer cell lines. *Appl. Sci.* **11**, 4912 (2021).
129. Sahin, U. *et al.* Oxidative stress-induced assembly of PML nuclear bodies controls sumoylation of partner proteins. *J Cell Biol* **204**, 931–945 (2014).
130. Chen, J. *et al.* Oxidative stress disrupts the cytoskeleton of spinal motor neurons. *Brain Behav* **13**, e2870 (2023).
131. Lee, Y.-B. *et al.* Cytoplasmic TDP-43 is involved in cell fate during stress recovery. *Hum. Mol.*

Genet. **31**, 166–175 (2021).

132. Cao, X., Jin, X. & Liu, B. The involvement of stress granules in aging and aging-associated diseases. *Aging Cell* **19**, e13136 (2020).

133. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).

134. Ramaswami, M., Taylor, J. P. & Parker, R. Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell* **154**, 727–736 (2013).

135. Mutihac, R. *et al.* TARDBP pathogenic mutations increase cytoplasmic translocation of TDP-43 and cause reduction of endoplasmic reticulum Ca^{2+} signaling in motor neurons. *Neurobiol. Dis.* **75**, 64–77 (2015).

136. Hulme, A. J., Maksour, S., St-Clair Glover, M., Miellet, S. & Dottori, M. Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Reports* **17**, 14–34 (2022).

137. Wang, Y. *et al.* ANXA11 mutations are associated with amyotrophic lateral sclerosis-frontotemporal dementia. *Front. Neurol.* **13**, 886887 (2022).

138. Feneberg, E. *et al.* An ALS-linked mutation in TDP-43 disrupts normal protein interactions in the motor neuron response to oxidative stress. *Neurobiol. Dis.* **144**, 105050 (2020).

139. Qiu, Y. *et al.* Emerging views of OPTN (optineurin) function in the autophagic process associated with disease. *Autophagy* **18**, 73–85 (2022).

140. Li, F. *et al.* Structural insights into the interaction and disease mechanism of neurodegenerative disease-associated optineurin and TBK1 proteins. *Nat. Commun.* **7**, 12708 (2016).

141. Hofweber, M. *et al.* Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706–719.e13 (2018).

142. Yoshizawa, T. *et al.* Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell* **173**, 693–705.e22 (2018).

143. Dormann, D. *et al.* ALS-associated fused in sarcoma (FUS) mutations disrupt

Transportin-mediated nuclear import. *EMBO J.* **29**, 2841–2857 (2010).

144. Bryant, P. & Noé, F. Rapid protein-protein interaction network creation from multiple sequence alignments with Deep Learning. *bioRxiv* 2023.04.15.536993 (4 2023).

145. Pantazis, C. B. *et al.* A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* **29**, 1685–1702.e22 (2022).

146. Lee, S.-Y. *et al.* APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell Rep.* **15**, 1837–1847 (2016).

147. Lara Flores, E., Qi, A., Reilly, L., Santiana, M. & Cookson, M. iNDI Transcription Factor-NGN2 differentiation of human iPSC into cortical neurons Version 2 v1. (2023) doi:10.17504/protocols.io.n2bvj3owblk5/v1.

148. Baskerville, V., Rapuri, S., Mehlhop, E. & Coyne, A. N. SUN1 facilitates CHMP7 nuclear influx and injury cascades in sporadic amyotrophic lateral sclerosis. *Brain* **147**, 109–121 (2024).

149. Rothstein, J. D., Warlick, C. & Coyne, A. N. Highly variable molecular signatures of TDP-43 loss of function are associated with nuclear pore complex injury in a population study of sporadic ALS patient iPSNs. *bioRxiv* (2023) doi:10.1101/2023.12.12.571299.

150. Coyne, A. N. *et al.* Nuclear accumulation of CHMP7 initiates nuclear pore complex injury and subsequent TDP-43 dysfunction in sporadic and familial ALS. *Sci. Transl. Med.* **13**, (2021).

151. Waldron, F. M., Rifai, O. & Gregory, J. Antibody and TDP-43 RNA aptamer dual staining to detect patterns of co-pathology in FFPE-preserved human t. (2024).

152. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).

153. Keren-Shaul, H. *et al.* MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat. Protoc.* **14**, 1841–1862 (2019).

154. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

155. Finkel, Y. *et al.* SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis.

Nature **594**, 240–245 (2021).

156. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
157. Brenner, J. F. *et al.* An automated microscope for cytologic research a preliminary evaluation. *J. Histochem. Cytochem.* **24**, 100–111 (1976).
158. Bozinovski, S. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica (Ljubl.)* **44**, (2020).
159. Bradley, P. S., Bennett, K. P. & Demiriz, A. Constrained K-Means clustering. http://machinelearning102.pbworks.com/f/ConstrainedKMeanstr-2000-65.pdf (2000).
160. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, (2019).