Inclusion bodies formed by polyglutamine and poly(glycine-alanine) are enriched with distinct proteomes but converge in proteins that are risk factors for disease and involved in protein degradation

Mona Radwan[1], Jordan D. Lilley[1], Ching-Seng Ang[2], Gavin E. Reid[1,3], Danny M. Hatters[1*]

[1]*Department of Biochemistry and Molecular Biology; and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, VIC 3010. Australia*

[2]*Bio21 Mass Spectrometry and Proteomics Facility, The University of Melbourne, Parkville, Victoria, Australia*

[3]*School of Chemistry, The University of Melbourne, VIC 3010. Australia*

**Correspondence: dhatters@unimelb.edu (D.M. Hatters). Twitter: @DannyHatters*

ABSTRACT

Poly(glycine-alanine) (polyGA) is one of the dipolypeptides expressed in Motor Neuron Disease caused by C9ORF72 mutations and accumulates as inclusion bodies in the brain of patients. Superficially these inclusions are similar to those formed by polyglutamine (polyQ) in Huntington's disease and both have been reported to form an amyloid-like structure suggesting they might aggregate via similar mechanisms to confer cellular dysfunction similarly. Here we investigated which endogenous proteins were enriched in these inclusions and whether aggregation-prone lengths of polyQ (Q_{97}), in context of Huntingtin exon 1, shared similar patterns to aggregation-prone lengths of polyGA (101_{GA}). When co-expressed in the same cell, $polyGA_{101}$ and $HttQ_{97}$ inclusions adopted distinct phases with no overlap suggesting different endogenous proteins would be enriched. Proteomic analyses indeed yielded distinct sets of endogenous proteins recruited into the inclusion types. The proteosome, microtubules, TriC chaperones, and translational machinery were enriched in polyGA aggregates, whereas Dnaj chaperones, nuclear envelope and RNA splicing proteins were enriched in polyQ aggregates. Both structures revealed a synergy of degradation

machinery including proteins in the polyQ aggregates that are risk factors for other neurodegenerative diseases involving protein aggregation when mutated, which suggests a convergence point in the pathomechanisms of these diseases.

INTRODUCTION

The formation of protein inclusions is a hallmark of many neurodegenerative diseases. In Huntington's disease, amino-terminal fragments of mutant Huntingtin (Htt) protein aggregate into intraneuronal inclusions [1]. The aggregation of mutant Htt is triggered by an abnormally expanded polyglutamine (polyQ) sequence encoded in exon 1 that arises by CAG trinucleotide repeat expansions [2, 3]. Long polyglutamine sequences form cytoplasmic or nuclear inclusions in animal and mouse models and is associated with a pathological cascade of events (reviewed in [4]).

In motor neuron disease caused by *C9ORF72* GGGGCC hexanucleotide repeat expansion mutations, protein inclusions arise from the aggregation of dipolydipeptides (DPRs) expressed abnormally from the expanded GGGGCC hexanucleotide repeat sequence. 5 different DPRs are expressed, namely dipeptide polymers of glycine-alanine (polyGA), proline-arginine (polyPR), glycine-arginine (polyGR), proline-alanine (polyPA), and proline-glycine (PolyPG). Of these polyPR and polyGR are profoundly toxic when expressed in cell culture and animal models targeting mechanisms in ribosome biogenesis, translation, actin cytoskeleton among others [5-12]. PolyGA appears less toxic than the others although has been reported in some models to confer toxicity [13-21]. We previously reported polyGA to be mildly toxic to cultured Neuro2a cells and to induce a distinct network of proteome changes that occur compared to the arg-rich DPRs [12]. We also noted a distinction to the other DPRs in forming large inclusions that morphologically are similar to the inclusions formed by polyQ. Furthermore, it has been reported that polyGA inclusions are, like polyQ, SDS-insoluble and amyloid-like [22, 23].

PolyGA is also more widespread in MND-patient brain tissue compared to the other DPRs [24].

Here we investigated the proteinaceous composition of polyGA inclusions and compared the profile quantitatively to inclusions of polyQ using a Huntingtin exon 1 model ($Httex1Q_{97}$) in a mouse neuroblastoma cell culture model using a novel proteomics-based approach to enrich the inclusions from cells under mild lysis conditions. We find distinct recruitment patterns to each inclusion type and also some similarities in biological mechanisms pertaining to degradation and a convergent pathomechanism for neurodegenerative diseases involving inappropriate protein aggregation.

METHODS

Plasmids. A pEGFP-based construct expressing polyGA dipeptide repeat length of 101 dipeptides ($polyGA_{101}$) was generated as described previously [12]. This construct expresses a GFP fusion tag at N-terminus of the polyGA. pT-REx vector expressing exon 1 of Htt (Httex1) with polyQ sequence length of 97 and C-terminal mCherry or GFP fluorescent tags and pT-REx-mCherry were prepared as previously described [25, 26].

Cell lines. Neuro-2a cells, obtained originally from the American Type Culture Collection (ATCC), were maintained in Opti-MEM (Life Technologies). The medium was supplemented with 10% v/v fetal calf serum, 1 mM glutamine, and 100 Unit mL^{-1} penicillin and 100 µg mL^{-1} streptomycin, and cells were kept in a humidified incubator with 5% v/v atmospheric CO_2 at 37 °C.

Transfections. Neuro2a cells were transiently transfected with the vectors using Lipofectamine 2000 reagent (Life Technologies). Specific transfection conditions for the different culture vessel types at densities of 9×10^4 (Ibidi 8-well µ-chamber) or 6×10^6 (T75

flasks). The following day cells (confluency of 80 – 90 %) were transiently transfected with 1.25 or 60 µL Lipofectamine 2000 and 0.5 or 24 µg vector DNA, respectively, as per the manufacturer's instructions (Life Technologies). The next day, the medium was changed to Opti-MEM, and for the time course the medium was refreshed daily.

Confocal imaging. Cells co-transfected with EGFPC2-GA_{101} and Httex1Q_{97}-mCherry were fixed 24 h after transfection in 4% w/v paraformaldehyde for 15 min at room temperature. Nuclei were counterstained with Hoechst 33342 at 1:200 dilution (Thermo Fisher Scientific, San Jose, CA) for 30 min then washed twice in phosphate buffered saline (PBS). Fixed cells were imaged on a Leica SP5 confocal microscope using HCX PL APO CS 40× or 63× oil-immersion objective lens (NA 1.4) at room temperature. Laser used: 405 nm excitation, 445–500 nm emission– Hoechst 33342; 488 nm excitation, 520–570 nm emission–GFP; 561 nm excitation, 590 nm emission– mCherry. Single colour controls were used to establish and adjust to remove bleed through of the emission filter bandwidths. FIJI version of ImageJ [27] and Inkscape software were used for image processing.

Longitudinal live-cell imaging. Neuro2a cells were co-transfected with pT-REx-mCherry and either EGFPC2-GA_{101} or pT-REx-Httex1Q_{97} -GFP. Medium was refreshed 24 hr post transfection and cells were imaged longitudinally with a JuLI-stage fluorescence microscope (NanoEnTek) at 15 min intervals for 96 hours. Channels used: GFP for EGFP (466/40 nm excitation, 525/50 nm emission), RFP for mCherry (525/50 nm excitation, 580 nm emission). Measurement of time of inclusion formation were extracted from files generated with automated imaging using the FIJI version of ImageJ [28]. Image processing was performed with the FIJI version of ImageJ [27] and visual inspection. Differences in inclusion formation rates were assessed by survival curve analysis in GraphPad Prism 7.05 (Graphpad Software Inc., San Diego, CA).

Purification of PolyGA and polyQ Aggregates. Neuro-2a cells expressing either GFP-tagged 101xGA or Httex1Q_{97} in 3 replicates were harvested by pelleting (200 *g*; 5 min; 24 °C) 24 h post transfection. Cell pellets were resuspended in lysis buffer (20 mM Tris, pH 8.0; 2 mM $MgCl_2$; 150 mM NaCl; 1% (w/v) Triton X-100; 20 Units/mL Benzonase, Novagen; 1× complete mini-protease cocktail; Roche) and then incubated for 30 min on ice. Lysates were diluted 2 times with PBS supplemented with protease inhibitor and aggregates were pelleted at 1000 *g* for 6 minutes. The aggregates were washed twice with 1 mL PBS, then resuspended in 1 ml PBS and subjected to fluorescence-activated cell sorting (FACS) on a BD FACS Aria III instrument with an outlet nozzle of 100 µm in diameter. The flow rate was adjusted to ~500 events/min, and EGFP fluorescence was monitored for sorting. Sorted aggregates were pelleted (12,000 *g*; 5 min; 4 °C), resuspended in PBS and washed 3 times by pelleting as above and resuspension in PBS. The final pellets were harvested by pelleting (21,000 *g*, 6 min, 4 °C) and dissolved in 10 µL neat formic acid for 30 min at 37 °C, vortexed for 20 seconds and sonicated for 1 min three times then incubated in a shaking microfuge tube incubator (30 min, 37 °C). Samples were neutralized to pH 7.0 by titration with unbuffered 3 M Tris. The protein concentration in the sample was determined by a Bradford assay using bovine serum albumin as mass standard. A total protein of 200 µg was further processed for mass spectrometry analysis.

Collection of cells by Pulse Shape Analysis. To assess the impact of polyGA aggregation on whole proteome, Neuro2a cells expressing GFP-tagged polyGA in 3 replicates were harvested 48 h post transfection by resuspension in PBS with a cell scraper. Cells were pelleted (120 *g*; 6 min) and resuspended in 2 mL PBS supplemented with 10 units/mL DNase I and filtered through 100 µm nylon mesh before analysis by flow cytometry. DAPI or Sytox (Thermo Fisher Scientific) was spiked into cell suspensions just before sorting to stain dead cells. Cells were analyzed by a FACS ARIA III cell sorter (BD Biosciences) equipped with 405-nm, 488-nm, 561-nm and 640-nm lasers. Live cells were

gated using side and forward scatter as described previously [29]. Cells were further gated into cells with $polyGA_{101}$ in the soluble form (ni) and those with $polyGA_{101}$ inclusions (i) by pulse shape analysis (PulSA) as previously described [29]. Each gate recovered between 0.8-1 × 10^6 cells which were sorted directly into PBS and then snap frozen in liquid nitrogen and stored at – 80 °C until used.

Sample preparation for whole proteome analysis. Sorted cell populations were thawed and resuspended in 100 µl RIPA lysis buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% v/v NP-40, 0.1% w/v SDS, 1% w/v sodium deoxycholate, 1× complete mini-protease mixture; Roche), and incubated on ice for 30 min. The concentration of proteins was measured by the Pierce microBCA Protein Assay according to the manufacturer's instruction (Thermo Fisher Scientific). Equal amounts of protein for each sample were precipitated with six volumes of pre-chilled (–20 °C) acetone, and incubation overnight at –20°C. Samples were then pelleted (21,000 *g*, 10 min, 4 °C). Acetone was decanted without disturbing the protein pellet. The pellets were washed once with pre-chilled acetone then allowed to dry for 10 min. The protein precipitates were resuspended in 100 µl 0.1 M triethylammonium bicarbonate (TEAB) and were vortexed and then sonicated 3 times for 30 s. The samples were further processed for mass spectrometry analysis.

Protein sample preparation for mass spectrometry. Proteins were subjected to reduction with 10 mM tris(2-carboxyethyl)phosphine (TCEP), pH 8.0, and alkylation with 55 mM iodoacetamide for 45 min, followed by trypsin digestion (0.25 µg, 37 °C, overnight). The resultant peptides were adjusted to contain 1% v/v formic acid then desalted by solid-phase extraction with an SPE cartridge (Oasis HLB 1 cc Vac Cartridge, Waters Corp., Milford, MA) pre-washed with 1 ml of 80% v/v acetonitrile (ACN) containing 0.1% v/v trifluoroacetic acid (TFA) and equilibrated with 1.2 ml of 0.1% v/v TFA three times. Samples were then loaded on the cartridge and washed with 1.5 ml of 0.1% v/v TFA before being eluted with 0.8 ml of 80% v/v ACN containing 0.1% v/v TFA and

collected in 1.5 ml microcentrifuge tubes. Peptides were then lyophilized by freeze drying (Virtis, SP Scientific, Warminster, PA). The peptides were resuspended in 100 µl distilled water and quantified using microBCA assay with bovine serum albumin as the mass standard. Then, 10 µg of each sample (in a volume of 50 µl containing 100 mM TEAB) were differentially labelled by reductive dimethyl labelling using equal volumes (2 µl) of 4% light formaldehyde (CH_2O) or 4% medium formaldehyde (CD_2O, 98% D) and 0.6 M Sodium cyanoborohydride ($NaCNBH_3$). The peptide solutions were incubated on an Eppendorf Thermomixer (Eppendorf South Pacific Pty. Ltd., Macquarie Park, NSW, Australia) at room temperature for 1 h. After quenching with 8 µl of 1% v/v ammonium hydroxide followed by further quenching with 8 µl of neat formic acid, dimethyl-labelled peptides were combined in equal amounts prior to liquid chromatography-nano electrospray ionization-tandem mass spectrometry (LC-nESI-MS/MS) analysis.

NanoESI-LC-MS/MS analysis. Peptides were analyzed by LC-nESI-MS/MS using an Orbitrap Lumos mass spectrometer (Thermo Fisher Scientific) fitted with nanoflow reversed-phase-HPLC (Ultimate 3000 RSLC, Dionex, Thermo Fisher Scientific). The nano-LC system was equipped with an Acclaim Pepmap nano-trap column (Dionex - C18, 100 Å, 75 µm × 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex - C18, 100 Å, 75 µm × 50 cm, Thermo Fisher Scientific). For each LC-MS/MS experiment, 1 µg (whole proteome) or 0.135 µg (aggregate proteome) of the peptide mix was loaded onto the enrichment (trap) column at a flow of 5 µl/min in 3% CH_3CN containing 0.1% v/v formic acid for 6 min before the enrichment column was switched in-line with the analytical column. The eluents used for the LC were 5% DMSO/0.1% v/v formic acid (solvent A) and 100% CH_3CN/5% DMSO/0.1% formic acid v/v. The gradient used was 3% v/v B to 20% B for 95 min, 20% B to 40% B in 10 min, 40% B to 80% B in 5 min and maintained at 80% B for the final 5 min before equilibration for 10 min at 3% B prior to the next analysis.

The mass spectrometer was operated in positive-ionization mode with spray voltage set at 1.9 kV and source temperature at 275 °C. Lockmass of 401.92272 from DMSO was used. The mass spectrometer was operated in the data-dependent acquisition mode, with MS spectra acquired by scanning from m/z 400–1500 at 120,000 resolution with an AGC target of 5e5. For MS/MS, the "top speed" acquisition method mode (3 s cycle time) on the most intense precursor was used whereby peptide ions with charge states ≥2 were isolated with an isolation window of 1.6 m/z and fragmented with high energy collision (HCD) mode, with a stepped collision energy of 30 ± 5%. Product ion spectra were acquired in the Orbitrap at 15,000 resolution. Dynamic exclusion was activated for 30s.

Proteomic data analysis. Raw data were analyzed using Proteome Discoverer (version 2.3; Thermo Scientific) with the Mascot search engine (Matrix Science version 2.4.1). Database searches were conducted against the Swissprot *Mus musculus* database (version 2016_07; 16794 proteins) combined with common contaminant proteins. GFP sequence (UniProt ID: P42212) was also concatenated to the Httex1Q_{97} and PolyGA_{101} sequences. Search was conducted with 20 ppm MS tolerance, 0.2 Da MS/MS tolerance and 2 missed cleavages allowed. Variable modifications were used for all experiments: oxidation (M), acetylation (Protein N-term), dimethylation (K), dimethylation (N-Term), 2H(4) dimethylation: (K) and 2H(4) dimethylation (N-term). A fixed modification used for all experiments was carbamidomethyl (C). The false discovery rate (FDR) was calculated by the Percolator node in Proteome Discoverer v 2.3.0.81 and was set to 0.5 % at the peptide identification level and 1 % at the protein identification level. Proteins were filtered for those containing at least two unique peptides in 3 biological replicates. Peptide quantitation was performed in Proteome Discoverer v.2.3 using the precursor ion quantifier node. Dimethyl labelled peptide pairs were established with a 2 ppm mass precision and a signal to noise threshold of 3. The retention time tolerance of isotope pattern multiplex was set to 0.6 min. Two single peak or missing channels were allowed

for peptide identification. The protein abundance in each replicate was calculated by summation of the unique peptide abundances that were used for quantitation (light or medium derivatives). Missing quantitation values were replaced with a constant (zero-filling). The peptide group abundance and protein abundance values were normalized to account for sample loading. In brief, the total peptide abundances for each sample was calculated and the maximum sum for all files was determined. The normalization factor was the factor of the sum of the sample and the maximum sum in all files. After calculating the normalization factors, the Peptide and Protein Quantifier node normalized peptide group abundances and protein abundances by dividing abundances with the normalization factor over all samples. The normalized protein abundances were imported into Perseus software (v 1.6.5.0). Protein abundances were transformed to log2 scale. The samples were then grouped according to the replicates. For pairwise comparison of proteomes and determination of significant differences in protein abundances, paired Student's t test based on permutation-based FDR statistics was then applied (250 permutations; FDR = 0.05; S0 = 0.1). This was justified on the basis the proteomics abundance data is normally distributed.

Bioinformatics. Protein interaction networks were generated using Cytoscape 3.7.1[30] built-in STRING (v11.0) [31] using active interaction sources parameters on for Experiments, Databases, Co-expression neighborhood, Gene Fusion and Cooccurrence. The minimum required interaction score setting was 0.9 (highest confidence). The corresponding enriched GO annotation terms were determined by calculating their enrichment *P*-value, which we compute using a Hypergeometric test, as explained in [32]. The *P*-values are corrected for multiple testing using the method of Benjamini and Hochberg [33]. Selected GO terms were used to manually re-arrange nodes and were added to protein interaction network using Inkscape.

IUPred [34] were applied to predict the intrinsically unstructured/disordered regions of proteins significantly enriched in polyGA or Httex1Q_{97} aggregates. Glutamine content

was analyzed with the web-server COPid [35] (http://www.imtech.res.in/raghava/copid/). A control set of 100 random proteins (Table S1) was generated from a list of the mouse proteome obtained from the Uni-ProtKB database (http://www.uniprot.org/uniprot/?query=reviewed:yes+AND+organism:10090&random=yes). The Mann-Whitney- Wilcoxon test was employed to determine significant differences.

Statistical Analysis. The details of the tests were reported in the figure legends. All statistical analyses were performed with GraphPad Prism v 7.05 (Graphpad Software Inc., San Diego, CA). Significant results were defined on the figures for $P < 0.05$.

Data availability. The MS proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [36] partner repository with the dataset identifiers PXD018505 for aggregate proteome data and PXD018824 for whole proteome data.

RESULTS & DISCUSSION

Previously we found that polyGA$_{101}$ as a fusion to GFP formed cytosolic inclusions in Neuro2a cells when transiently transfected. To further investigate the rate that this occurs, we used live cell imaging to track cells from 24 h after transfection onwards. Of cells with detectable levels of expression by 24 h almost all of them had formed inclusions by 60 h (Fig 1A). This was faster than comparable experiments with Httex1Q$_{97}$ as a fusion to mCherry, which is well known to extensively aggregate in cell culture [37, 38] (Fig 1A). We noted that some lower-expressing cells showed detectable expression of polyGA only after 24 h (which we did not track) and that these were likely to form aggregates more slowly. When we co-expressed Httex1Q$_{97}$ as a fusion to mCherry, we found the polyGA$_{101}$ and Httex1Q$_{97}$ formed discrete inclusions in the same cell with no apparent colocalization (Fig 1B). This suggested that any concomitant co-aggregation patterns that arise with endogenous proteins may involve distinct proteins.

Figure 1. Httex1Q_{97} and polyGA_{101} rapidly form distinct inclusions in neuro2a cells. A. Kaplan-Meier curves of polyGA- or Httex1Q_{97} expressing cells that form visible inclusions as assessed by longitudinal imaging. Cells were tracked from 24 h post-transfection. *P*-values correspond to log-rank (Mantel-Cox) test. **B.** Confocal micrographs of neuro2a cells co-expressing GFP-tagged GA_{101} and mCherry-tagged Httex1Q_{97}, fixed 24 hr post-transfection and stained with Hoechst33258 (cyan) to visualize nuclei. Scale bar represents 5 µm.

To investigate these potential differences, pellets recovered from lysates of neuro2a cells expressing GFP-tagged Httex1Q_{97} or GFP-tagged polyGA_{101} were sorted to purify the aggregates using flow cytometry via monitoring the GFP fluorescence. Quantitative proteomics, by way of dimethyl isotope labelling, was used to define the proteins enriched in each aggregate class after normalization to total mass of protein. We observed 737 proteins. Of these 63 were significantly enriched in polyGA inclusions (3 replicates, a permutation-based FDR cut-off of 5% and S0 of 0.1) and 48 were enriched in Httex1Q_{97} (Table 1, Table S2 and Fig 2A).

Figure 2. Proteome recruitment patterns to polyGA_{101} and Httex1Q_{97} inclusions. A. Volcano plot of proteins identified in the inclusions. *P*-values were calculated by a two-sided one samples t-test with null hypothesis that abundances were unchanged and the Log_2 ratio was equal to 0. Proteins meeting stringency thresholds (hyperbolic curves, FDR≤0.05, S_0=0.1) are shown as colored circles. **B.** STRING interaction maps (v.11) determined in Cytoscape (v3.7) for proteins significantly enriched in the inclusions (the full list of proteins are in Table S2). The analysis was done at the highest confidence setting. Each protein was represented by a colored circle sized proportionally to -log_{10} (*P*-value). The color scale represents logarithm of fold change. Selected significantly

enriched GO terms (GOCC, GOPB, and UniProt keywords) are displayed (Full terms are shown in Table S3). **C.** Analysis of enriched proteomes for low-complexity regions (IUPred-L) and high glutamine content. Significance of difference was assessed against a control dataset of random mouse proteins (Table S1) with the Mann–Whitney–Wilcoxon test. Whiskers extend from 10 to 90%.

We observed notable features in this dataset consistent with known pathological markers of polyGA inclusions. Namely in C9ORF72 mediated MND, a subset of inclusions is non-reactive to TDP-43 [39]. In most other forms of MND, TDP-43-reactive inclusions are a key pathological signature of neurons in disease [40]. These TDP-43 negative inclusions were previously found to be immunoreactive for polyGA, suggesting they form by polyGA aggregation [41, 42]. We observed a lack of TDP43 in the polyGA inclusions by virtue of an enrichment in the Httex1Q_{97} inclusions (Table 1). In addition, the TDP43-negative inclusions seen in vivo are immunoreactive to p62 [43] and lack immunoreactivity to FUS, optineurin, alpha-internexin and neurofilament [44, 45]. In our data p62 (also called sequestesome 1) is one of the most enriched proteins in polyGA inclusions, whereas Fus appeared excluded by virtue of its enrichment in Httex1Q_{97} inclusions, which has been observed previously in cell models of polyQ aggregation and human pathology [25, 46-48]. Hence these data point to the cell model of polyGA inclusions mimicking the process of aggregation and recruitment seen in vivo and also providing specificity of co-recruitment relative to Httex1Q_{97}.

Analysis of the differences is shown visually in Fig 2B by a (STRING) protein-protein interaction map and annotation to functional networks. Overall both inclusions yielded an enrichment for gene ontology and KEGG networks of microtubule cytoskeleton, proteasome complex, chaperones, RNA splicing and nuclear envelope (Fig 2C; Table S3). These findings are in accordance with prior findings that protein aggregation impacts

these biological processes and in particular an involvement in machinery for their clearance and degradation [49-52].

In addition, the data points more directly to proteins and genes implicated in MND phenotype and mechanisms. Phlda1 was one of the proteins enriched in polyGA inclusions. Previously it was found that Phlda1 was upregulated in Fus-mutant motor neurons and that this was an adaptive response to protect against apoptosis [53]. Phlda1 was also observed upregulated in sporadic MND fibroblasts treated to stress compared to controls [54]. Nudt5 was also found enriched in inclusions, and expression of this gene was significantly increased in motor neurons derived from induced plutripotent cells from MND patients over controls [55]. Another protein of note enriched in the polyGA aggregates was Dpysl3. A missense mutation that has been linked to MND risk in French population and in culture expression of the mutation leads to shorted neuronal survival [56]. Hence it remains plausible that co-aggregation of these proteins into polyGA inclusions sequesters their activity and renders cells less resilient to stress triggers.

The Httex1Q_{97}-enriched proteome also yielded noteworthy findings. Previously it was found that polyQ can preferentially co-recruit proteins containing intrinsically disordered domains and proteins enriched in glutamine (IDRs) [25, 47]. These patterns were also observed in our data (Fig 2C). However, polyGA did not show these enrichment patterns, indicative of specificity for polyQ in recruiting IDRs and Q-rich proteins.

To assess whether the changes in polyGA inclusion formation had other effects on proteome abundance, we expressed polyGA$_{101}$ and at 48 h after transfection sorted live cells into those with visible aggregates from those without by flow cytometry sorting method and pulse shape analysis [57] (Fig 3A). We found cells with inclusions were more reactive to Sytox (Fig 3A inset), which is indicative of dying and dead cells, than cells

without inclusions so we excluded these cells from analysis. 35% of the remaining live cells expressing polyGA had inclusions (Fig 3A). This was a lower yield of aggregates than we measured by live cell imaging at 48h in Fig 1A (about 90%), which we attribute to this experiment being more inclusive of the lower-expressing cells and other differences in the experimental conditions that affect aggregation rates such as phototoxicity of live cell imaging. Nonetheless, the yield obtained made the experiment amenable for comparing cells with versus without inclusions. Out of 2420 proteins identified, we observed 56 proteins that significantly changed abundance in these sorted cell populations (Fig 3B; Table S4). There was no overlap in the proteins seen enriched in polyGA inclusions with proteins that changed expression due to polyGA aggregation. This provides firmer confidence that the enrichment seen in the polyGA aggregates arises from co-aggregation rather than changes in gene expression. Of the genes that changed expression, protein interaction networks yielded significant enrichment in networks including nuclear speck (GO: 0016607), ribosome biogenesis (GO: 0042254), chromosome (GO:0005694), mitochondrion (GO:0005739) and Golgi-to-ER-traffic (MMU-6811442) (Table S5). These pathways would be anticipated to be activated by stress responses incurred by protein aggregation, however, we did not note any striking changes that pertained to novel mechanisms other than that from this data.

Figure 3. Cellular protein abundance changes arising from polyGA$_{101}$ aggregation. A. Schematic of flow cytometry method of pulse shape analysis (PulSA) to sort cells enriched with inclusions (i) from those without inclusions (ni). Cells with inclusions display shorter width (W) fluorescence values versus cells with soluble protein, and typically higher height values (H) arising from the condense foci of fluorescence inside the cells. Cells were sorted to exclude dead cells by Sytox reactivity. Inset shows percentage of transfected cells reactive to Sytox by time after transfection. *n*=4, means ± SD shown. **B.** Volcano plots of proteins that changed their abundance upon polyGA

aggregation. The data for all proteins are plotted as log2-fold change versus the –log10 of the *P*-value. The dotted line indicates significance cut-off (hyperbolic curves, FDR≤0.05, S_0=0.1) and proteins meeting stringency thresholds are shown as colored circles. **C.** Protein-protein interaction network (STRING v11) of proteins significantly changed in abundance upon polyGA aggregation (the full list of proteins are in Table S4). The analysis was done at the highest confidence setting. Each protein was represented by a colored circle sized proportionally to -$\log_{10}$ (*P*-value). The color scale represents logarithm of fold change. Selected significantly enriched GO terms (GOCC, GOPB, and UniProt keywords) are displayed (Table S5).

Lastly we investigated the overlap of proteins enriched in Httex1Q_{97} inclusions with our previously reported changes in solubility of whole cell proteome before versus after inclusions had formed [58]. In that dataset we observed 17 proteins that significantly decreased in solubility as cells expressing Httex1Q_{97} shifted from a dispersed unaggregated state to forming inclusions [58]. Of these, 5 proteins were found in our list of 48 proteins significantly enriched in Httex1Q97 inclusions (Picalm, Hgs, Clint1, Ubqln2 and Dnajb1). Four of these proteins (all except Dnajb1) form a robust protein-protein interaction network with a significant gene ontology enrichment for clathrin coat assembly (GO:0048268; FDR of 0.0031) suggestive that this mechanism is involved in polyQ aggregation. Clint1 and Ubqln2 were previously shown to colocalize to polyQ inclusions, supporting this conclusion [47, 59]. An interesting note with respect to mechanism is that UBQLN2 targets ubiquitinated substrates for degradation in ERAD and autophagy [60]. Furthermore mutations in UBQLN2 cause MND, and appear to operate by impairing protein degradation of ubiquitinated proteins [61]. Further supporting an important role linking protein aggregation, degradation more broadly to these neurodegenerative diseases is the enrichment of Picalm in the polyQ inclusions. Picalm is an phosphatidylinositol-binding clathrin assembly protein and has been shown via

GWAS as a top ten risk for Alzheimer's disease [62, 63]. It has been reported to modulate intracellular APP processing and plaque pathogenesis [64], modulate autophagy and alter tau clearance [65].

Collectively the data here reports proteins that co-aggregate into two very different neurodegenerative disease proteinaceous deposits. The findings provide specificity of proteins to the aggregation type that provide useful perspective to that reported by others. Moreover, the mechanisms of protein clearance mechanism appear relevant to both aggregation types and notably of a number of proteins in the Httex1Q_{97} aggregates that when mutated are modifiers of MND risk. Therefore, the findings identify a synergy of biological mechanisms involved in protein degradation that appear central to at least two different neurodegenerative diseases, and possibly more applicable to the other neurodegenerative diseases involving inappropriate protein aggregation.

ACKNOWLEDGMENTS

We thank the Bio21 Melbourne Mass Spectrometry and Proteomics facility. This work was funded by grants to DMH (National Health and Medical Research Council APP1161803 and Motor Neuron Disease Research Institute, Australia small grant) and to DMH and GER (Australian Research Council DP170103093). MR acknowledges support from an Australian Government Research Training Program (RTP) Scholarship and an Egyptian Ministry of Higher Education and Scientific Research PhD scholarship.

REFERENCES

1. DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science. 1997;277(5334):1990-3. PubMed PMID: 9302293.

2. MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, et al. A Novel Gene Containing a Trinucleotide Repeat that is Expanded and Unstable on Huntington's Disease Chromosomes. Cell. 1993;72(6):971-83.

3. Scherzinger E, Sittler A, Schweiger K, Heiser V, Lurz R, Hasenbank R, et al. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. Proc Natl Acad Sci U S A. 1999;96(8):4604-9. Epub 1999/04/14. doi: 10.1073/pnas.96.8.4604. PubMed PMID: 10200309; PubMed Central PMCID: PMCPMC16379.

4. Cox D, Raeburn C, Sui X, Hatters DM. Protein aggregation in cell biology: An aggregomics perspective of health and disease. Semin Cell Dev Biol. 2018. Epub 2018/05/14. doi: 10.1016/j.semcdb.2018.05.003. PubMed PMID: 29753879.

5. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science. 2014;345(6201):1139-45. Epub 2014/08/02. doi: 10.1126/science.1254917. PubMed PMID: 25081482; PubMed Central PMCID: PMCPMC4459787.

6. Tao Z, Wang H, Xia Q, Li K, Li K, Jiang X, et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum Mol Genet. 2015;24(9):2426-41. Epub 2015/01/13. doi: 10.1093/hmg/ddv005. PubMed PMID: 25575510.

7. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature. 2015;525(7567):56-61. Epub 2015/08/27. doi: 10.1038/nature14973. PubMed PMID: 26308891; PubMed Central PMCID: PMCPMC4800742.

8. Kanekura K, Yagi T, Cammack AJ, Mahadevan J, Kuroda M, Harms MB, et al. Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. Hum Mol Genet. 2016;25(9):1803-13. Epub 2016/03/05. doi: 10.1093/hmg/ddw052. PubMed PMID: 26931465; PubMed Central PMCID: PMCPMC4986334.

9. Lee KH, Zhang P, Kim HJ, Mitrea DM, Sarkar M, Freibaum BD, et al. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell. 2016;167(3):774-88 e17. Epub 2016/10/22. doi: 10.1016/j.cell.2016.10.002. PubMed PMID: 27768896; PubMed Central PMCID: PMCPMC5079111.

10. Freibaum BD, Taylor JP. The Role of Dipeptide Repeats in C9ORF72-Related ALS-FTD. Front Mol Neurosci. 2017;10(35):35. Epub 2017/03/01. doi: 10.3389/fnmol.2017.00035. PubMed PMID: 28243191; PubMed Central PMCID: PMCPMC5303742.

11. Zhang YJ, Gendron TF, Ebbert MTW, O'Raw AD, Yue M, Jansen-West K, et al. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat Med. 2018;24(8):1136-42. Epub 2018/06/27. doi: 10.1038/s41591-018-0071-1. PubMed PMID: 29942091; PubMed Central PMCID: PMCPMC6520050.

12. Radwan M, Ang CS, Ormsby AR, Cox D, Daly JC, Reid GE, et al. Arginine in C9ORF72 dipolypeptides mediates promiscuous proteome binding and multiple modes of toxicity. Mol Cell Proteomics. 2020. Epub 2020/02/23. doi: 10.1074/mcp.RA119.001888. PubMed PMID: 32086375.

13. Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, et al. p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta Neuropathol. 2011;122(6):691-702. Epub 2011/11/22. doi: 10.1007/s00401-011-0911-2. PubMed PMID: 22101323.

14. Bigio EH, Weintraub S, Rademakers R, Baker M, Ahmadian SS, Rademaker A, et al. Frontotemporal lobar degeneration with TDP-43 proteinopathy and chromosome 9p repeat expansion in C9ORF72: clinicopathologic correlation. Neuropathology.

2013;33(2):122-33. Epub 2012/06/19. doi: 10.1111/j.1440-1789.2012.01332.x. PubMed PMID: 22702520; PubMed Central PMCID: PMCPMC3449045.

15. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013;77(4):639-46. Epub 2013/02/19. doi: 10.1016/j.neuron.2013.02.004. PubMed PMID: 23415312; PubMed Central PMCID: PMCPMC3593233.

16. Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. 2013;126(6):829-44. Epub 2013/10/17. doi: 10.1007/s00401-013-1192-8. PubMed PMID: 24129584; PubMed Central PMCID: PMCPMC3830741.

17. Mann DM, Rollinson S, Robinson A, Bennion Callister J, Thompson JC, Snowden JS, et al. Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol Commun. 2013;1(1):68. Epub 2013/11/21. doi: 10.1186/2051-5960-1-68. PubMed PMID: 24252525; PubMed Central PMCID: PMCPMC3893586.

18. Mori K, Arzberger T, Grasser FA, Gijselinck I, May S, Rentzsch K, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. 2013;126(6):881-93. Epub 2013/10/18. doi: 10.1007/s00401-013-1189-3. PubMed PMID: 24132570.

19. Zu T, Liu Y, Banez-Coronel M, Reid T, Pletnikova O, Lewis J, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A. 2013;110(51):E4968-77. Epub 2013/11/20. doi:

10.1073/pnas.1315438110. PubMed PMID: 24248382; PubMed Central PMCID: PMCPMC3870665.

20. May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathologica. 2014;128(4):485-503. doi: 10.1007/s00401-014-1329-4. PubMed PMID: PMC4159571.

21. Mizielinska S, Isaacs AM. C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia: gain or loss of function? Curr Opin Neurol. 2014;27(5):515-23. doi: 10.1097/WCO.0000000000000130. PubMed PMID: 25188012.

22. May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. 2014;128(4):485-503. Epub 2014/08/15. doi: 10.1007/s00401-014-1329-4. PubMed PMID: 25120191; PubMed Central PMCID: PMCPMC4159571.

23. Chang YJ, Jeng US, Chiang YL, Hwang IS, Chen YR. The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. J Biol Chem. 2016;291(10):4903-11. Epub 2016/01/16. doi: 10.1074/jbc.M115.694273. PubMed PMID: 26769963; PubMed Central PMCID: PMCPMC4777828.

24. Mackenzie IR, Frick P, Grasser FA, Gendron TF, Petrucelli L, Cashman NR, et al. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol. 2015;130(6):845-61. Epub 2015/09/17. doi: 10.1007/s00401-015-1476-2. PubMed PMID: 26374446.

25. Ramdzan YM, Trubetskov MM, Ormsby AR, Newcombe EA, Sui X, Tobin MJ, et al. Huntingtin Inclusions Trigger Cellular Quiescence, Deactivate Apoptosis, and

Lead to Delayed Necrosis. Cell Rep. 2017;19(5):919-27. Epub 2nd May 2017. doi: 10.1016/j.celrep.2017.04.029. PubMed PMID: 28467905.

26. Tsvetkov AS, Miller J, Arrasate M, Wong JS, Pleiss MA, Finkbeiner S. A small-molecule scaffold induces autophagy in primary neurons and protects against toxicity in a Huntington disease model. Proc Natl Acad Sc USA. 2010;107(39):16982-7. doi: 10.1073/pnas.1004498107.

27. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature Methods. 2012;9(7):676-82. doi: 10.1038/nmeth.2019.

28. Abramoff MD, Magalhaes PJ, Ram SJ. Image Processing with ImageJ. Biophotonics International. 2004;11(7):36-42.

29. Ramdzan YM, Polling S, Chia CPZ, Ng IHW, Ormsby AR, Croft NP, et al. Tracking protein aggregation and mislocalization in cells with flow cytometry. Nature Methods. 2012;9:467. doi: 10.1038/nmeth.1930

https://www.nature.com/articles/nmeth.1930#supplementary-information.

30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003;13(11):2498-504.

31. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research. 2016;45(D1):D362-D8. doi: 10.1093/nar/gkw937.

32. Rivals I, Personnaz L, Taing L, Potier M-C. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics. 2006;23(4):401-7. doi: 10.1093/bioinformatics/btl633.

33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society:

Series B (Methodological). 1995;57(1):289-300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

34. Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005;21(16):3433-4. doi: 10.1093/bioinformatics/bti541.

35. Kumar M, Thakur V, Raghava GPS. COPid: Composition Based Protein Identification. In Silico Biology. 2008;8(2):121-8.

36. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research. 2015;44(D1):D447-D56. doi: 10.1093/nar/gkv1145.

37. Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature. 2004;431(7010):805-10. doi: 10.1038/nature02998. PubMed PMID: 15483602.

38. Ormsby AR, Ramdzan YM, Mok YF, Jovanoski KD, Hatters DM. A platform to view huntingtin exon 1 aggregation flux in the cell reveals divergent influences from chaperones hsp40 and hsp70. J Biol Chem. 2013;288(52):37192-203. doi: 10.1074/jbc.M113.486944. PubMed PMID: 24196953; PubMed Central PMCID: PMCPMC3873573.

39. Cruts M, Gijselinck I, Van Langenhove T, van der Zee J, Van Broeckhoven C. Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci. 2013;36(8):450-9. Epub 2013/06/12. doi: 10.1016/j.tins.2013.04.010. PubMed PMID: 23746459.

40. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006;314(5796):130-3. Epub 2006/10/07. doi: 10.1126/science.1134108. PubMed PMID: 17023659.

573 41. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford
574 NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of
575 C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011;72(2):245-56.
576 Epub 2011/09/29. doi: 10.1016/j.neuron.2011.09.011. PubMed PMID: 21944778;
577 PubMed Central PMCID: PMCPMC3202986.
578 42. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, et al. A
579 hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-
580 linked ALS-FTD. Neuron. 2011;72(2):257-68. Epub 2011/09/29. doi:
581 10.1016/j.neuron.2011.09.010. PubMed PMID: 21944779; PubMed Central PMCID:
582 PMCPMC3200438.
583 43. Pikkarainen M, Hartikainen P, Alafuzoff I. Ubiquitinated p62-positive, TDP-43-
584 negative inclusions in cerebellum in frontotemporal lobar degeneration with TAR
585 DNA binding protein 43. Neuropathology. 2010;30(2):197-9. Epub 2009/07/23. doi:
586 10.1111/j.1440-1789.2009.01043.x. PubMed PMID: 19622109.
587 44. King A, Al-Sarraj S, Shaw C. Frontotemporal lobar degeneration with ubiquitinated
588 tau-negative inclusions and additional alpha-synuclein pathology but also unusual
589 cerebellar ubiquitinated p62-positive, TDP-43-negative inclusions.
590 Neuropathology. 2009;29(4):466-71. Epub 2008/08/22. doi: 10.1111/j.1440-
591 1789.2008.00966.x. PubMed PMID: 18715271.
592 45. King A, Maekawa S, Bodi I, Troakes C, Al-Sarraj S. Ubiquitinated, p62
593 immunopositive cerebellar cortical neuronal inclusions are evident across the
594 spectrum of TDP-43 proteinopathies but are only rarely additionally
595 immunopositive for phosphorylation-dependent TDP-43. Neuropathology.
596 2011;31(3):239-49. Epub 2010/12/02. doi: 10.1111/j.1440-1789.2010.01171.x. PubMed
597 PMID: 21118398.
598 46. Doi H, Okamura K, Bauer PO, Furukawa Y, Shimizu H, Kurosawa M, et al. RNA-
599 binding protein TLS is a major nuclear aggregate-interacting protein in huntingtin

exon 1 with expanded polyglutamine-expressing cells. J Biol Chem. 2008;283(10):6489-500. Epub 2008/01/03. doi: 10.1074/jbc.M705306200. PubMed PMID: 18167354.

47. Wear MP, Kryndushkin D, O'Meally R, Sonnenberg JL, Cole RN, Shewmaker FP. Proteins with Intrinsically Disordered Domains Are Preferentially Recruited to Polyglutamine Aggregates. PLoS One. 2015;10(8):e0136362. Epub 2015/09/01. doi: 10.1371/journal.pone.0136362. PubMed PMID: 26317359; PubMed Central PMCID: PMCPMC4552826.

48. Doi H, Koyano S, Suzuki Y, Nukina N, Kuroiwa Y. The RNA-binding protein FUS/TLS is a common aggregate-interacting protein in polyglutamine diseases. Neurosci Res. 2010;66(1):131-3. Epub 2009/10/17. doi: 10.1016/j.neures.2009.10.004. PubMed PMID: 19833157.

49. Ross CA, Poirier MA. What is the role of protein aggregation in neurodegeneration? Nat Rev Mol Cell Biol. 2005;6(11):891-8.

50. Woerner AC, Frottin F, Hornburg D, Feng LR, Meissner F, Patra M, et al. Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. Science. 2016;351(6269):173-6. Epub 2015/12/05. doi: 10.1126/science.aad2033. PubMed PMID: 26634439.

51. Boland B, Yu WH, Corti O, Mollereau B, Henriques A, Bezard E, et al. Promoting the clearance of neurotoxic proteins in neurodegenerative disorders of ageing. Nat Rev Drug Discov. 2018;17(9):660-88. Epub 2018/08/18. doi: 10.1038/nrd.2018.109. PubMed PMID: 30116051; PubMed Central PMCID: PMCPMC6456907.

52. Nussbacher JK, Tabet R, Yeo GW, Lagier-Tourenne C. Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. Neuron. 2019;102(2):294-320. Epub 2019/04/19. doi: 10.1016/j.neuron.2019.03.014. PubMed PMID: 30998900; PubMed Central PMCID: PMCPMC6545120.

626 53. De Santis R, Santini L, Colantoni A, Peruzzi G, de Turris V, Alfano V, et al. FUS
627 Mutant Human Motoneurons Display Altered Transcriptome and microRNA
628 Pathways with Implications for ALS Pathogenesis. Stem Cell Reports.
629 2017;9(5):1450-62. Epub 2017/10/11. doi: 10.1016/j.stemcr.2017.09.004. PubMed PMID:
630 28988989; PubMed Central PMCID: PMCPMC5830977.
631 54. Raman R, Allen SP, Goodall EF, Kramer S, Ponger LL, Heath PR, et al. Gene
632 expression signatures in motor neurone disease fibroblasts reveal dysregulation
633 of metabolism, hypoxia-response and RNA processing functions. Neuropathol
634 Appl Neurobiol. 2015;41(2):201-26. Epub 2014/04/23. doi: 10.1111/nan.12147. PubMed
635 PMID: 24750211; PubMed Central PMCID: PMCPMC4329387.
636 55. Egawa N, Kitaoka S, Tsukita K, Naitoh M, Takahashi K, Yamamoto T, et al. Drug
637 screening for ALS using patient-specific induced pluripotent stem cells. Sci Transl
638 Med. 2012;4(145):145ra04. Epub 2012/08/03. doi: 10.1126/scitranslmed.3004052.
639 PubMed PMID: 22855461.
640 56. Blasco H, Bernard-Marissal N, Vourc'h P, Guettard YO, Sunyach C, Augereau O, et
641 al. A rare motor neuron deleterious missense mutation in the DPYSL3 (CRMP4)
642 gene is associated with ALS. Hum Mutat. 2013;34(7):953-60. Epub 2013/04/10. doi:
643 10.1002/humu.22329. PubMed PMID: 23568759.
644 57. Ramdzan YM, Polling S, Chia CP, Ng IH, Ormsby AR, Croft NP, et al. Tracking
645 protein aggregation and mislocalization in cells with flow cytometry. Nat Methods.
646 2012;9(5):467-70. Epub 2012/03/20. doi: 10.1038/nmeth.1930. PubMed PMID:
647 22426490.
648 58. Sui X, Pires DEV, Ormsby AR, Cox D, Nie S, Vecchi G, et al. Widespread
649 remodeling of proteome solubility in response to different protein homeostasis
650 stresses. Proc Natl Acad Sci U S A. 2020. Epub 2020/01/23. doi:
651 10.1073/pnas.1912897117. PubMed PMID: 31964829.

652 59. Yang H, Yue HW, He WT, Hong JY, Jiang LL, Hu HY. PolyQ-expanded huntingtin
653 and ataxin-3 sequester ubiquitin adaptors hHR23B and UBQLN2 into aggregates
654 via conjugated ubiquitin. FASEB J. 2018;32(6):2923-33. Epub 2018/02/07. doi:
655 10.1096/fj.201700801RR. PubMed PMID: 29401586.
656 60. Renaud L, Picher-Martel V, Codron P, Julien JP. Key role of UBQLN2 in
657 pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia. Acta
658 Neuropathol Commun. 2019;7(1):103. Epub 2019/07/20. doi: 10.1186/s40478-019-0758-
659 7. PubMed PMID: 31319884; PubMed Central PMCID: PMCPMC6889556.
660 61. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, et al. Mutations
661 in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and
662 ALS/dementia. Nature. 2011;477(7363):211-5. Epub 2011/08/23. doi:
663 10.1038/nature10353. PubMed PMID: 21857683; PubMed Central PMCID:
664 PMCPMC3169705.
665 62. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al.
666 Genome-wide association study identifies variants at CLU and PICALM associated
667 with Alzheimer's disease. Nat Genet. 2009;41(10):1088-93. Epub 2009/09/08. doi:
668 10.1038/ng.440. PubMed PMID: 19734902; PubMed Central PMCID:
669 PMCPMC2845877.
670 63. Elias-Sonnenschein LS, Bertram L, Visser PJ. Relationship between genetic risk
671 factors and markers for Alzheimer's disease pathology. Biomark Med.
672 2012;6(4):477-95. Epub 2012/08/25. doi: 10.2217/bmm.12.56. PubMed PMID: 22917148.
673 64. Xiao Q, Gil SC, Yan P, Wang Y, Han S, Gonzales E, et al. Role of
674 phosphatidylinositol clathrin assembly lymphoid-myeloid leukemia (PICALM) in
675 intracellular amyloid precursor protein (APP) processing and amyloid plaque
676 pathogenesis. J Biol Chem. 2012;287(25):21279-89. Epub 2012/04/28. doi:
677 10.1074/jbc.M111.338376. PubMed PMID: 22539346; PubMed Central PMCID:
678 PMCPMC3375549.

679 65. Moreau K, Fleming A, Imarisio S, Lopez Ramirez A, Mercer JL, Jimenez-Sanchez M,
680 et al. PICALM modulates autophagy activity and tau accumulation. Nat Commun.
681 2014;5:4998. Epub 2014/09/23. doi: 10.1038/ncomms5998. PubMed PMID: 25241929;
682 PubMed Central PMCID: PMCPMC4199285.

683

684

SUPPORTING INFORMATION CAPTIONS

Table S1. List of random proteins from mouse Uniprot database.

Table S2. Proteins enriched in inclusions of polyGA$_{101}$ and Httex1Q$_{97}$. Relates to Table 1 & Fig 2.

Table S3. Gene Ontology terms enriched among proteins identified in polyGA$_{101}$ or Httex1Q$_{97}$ inclusions. Relates to Fig 2.

Table S4. Cellular abundances of proteins caused by polyGA$_{101}$ aggregation. Relates to Fig 3.

Table S5. Gene Ontology terms enriched among proteins that changed abundance upon polyGA$_{101}$ aggregation. Relates to Fig 3.

Table 1: Proteins enriched in inclusions of polyGA$_{101}$ and Httex1Q$_{97}$*

Enriched in polyGA$_{101}$			**Enriched in Httex1Q$_{97}$**		
Description	*Gene ID*	*log2 enrichment (mean SD)*	*Description*	*Gene ID*	*log2 enrichment (mean SD)*
Pleckstrin homology-like domain family A member 1	Phlda1	3.91 ±1.18	Hsc70-interacting protein	St13	4.78 ±0.36
DNA replication licensing factor MCM3	Mcm3	3.73 ±10.23	Histone H3.1	Hist1h3a	4.3 ±1.28
Eukaryotic translation initiation factor 2 subunit 1	Eif2s1	3.65 ±5.69	*Clathrin interactor 1*	*Clint1*	*4.28 0.91*
Proteasome subunit beta type-4	Psmb4	2.6 ±2.39	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2	Chchd2	4.02 ±0.73
Sequestosome-1 (p62)	Sqstm1	2.47 ±0.11	**RNA-binding protein FUS**	**Fus**	**3.26 ±1.23**
Interferon-inducible double-stranded RNA-dependent protein kinase activator A	Prkra	2.4 ±2.63	Tight junction protein ZO-1	Tjp1	3.25 ±0.93
Sorting nexin-3	Snx3	2.36 ±1.05	Ubiquitin-associated protein 2	Ubap2	3.24 ±0.36

Electron transfer flavoprotein subunit alpha, mitochondrial	Etfa	2.24 ±2.41
Cytochrome c oxidase subunit 5B, mitochondrial	Cox5b	2.19 ±2.64
Nuclear migration protein nudC	Nudc	2.17 ±1.67
Receptor of activated protein C kinase 1	Rack1	2.17 ±0.91
40S ribosomal protein S2	Rps2	2.12 ±1.85
Nuclear fragile X mental retardation-interacting protein 2	Nufip2	2.11 ±1.41
26S proteasome non-ATPase regulatory subunit 12	Psmd12	2.09 ±2.28
Vigilin	Hdlbp	2.07 ±0.36
Insulin-like growth factor 2 mRNA-binding protein 3	Igf2bp3	2.06 ±2.16
GTP cyclohydrolase 1	Gch1	2.04 ±2.12
60S ribosomal protein L10	Rpl10	1.99 ±1.03
ATPase WRNIP1	Wrnip1	1.98 ±2.39
Protein SOGA3	Soga3	1.96 ±0.27
Small glutamine-rich tetratricopeptide repeat-containing protein alpha	Sgta	3.22 ±0.8
DnaJ homolog subfamily B member 1	*Dnajb1*	*3.11 1.11*
Chromobox protein homolog 1	Cbx1	3.1 ±0.36
Ubiquilin-2	***Ubqln2***	***3.06 ±0.77***
Phosphatidylinositol-binding clathrin assembly protein	***Picalm***	**2.81 ±0.25**
CUGBP Elav-like family member 1	Celf1	2.51 ±0.37
Transgelin-2	Tagln2	2.17 ±0.49
RNA-binding protein 25	Rbm25	2.16 ±0.51
Nucleolysin TIAR	Tial1	2.06 ±0.19
Caprin-1	Caprin1	2.04 ±0.43
Probable ATP-dependent RNA helicase DDX17	Ddx17	2.02 ±0.18
Protein PRRC2C	Prrc2c	2.01 ±0.48
RNA-binding motif, single-stranded-interacting protein	Rbms1	1.98 ±0.13

Ubiquitin fusion degradation protein 1 homolog	Ufd1l	1.95 ±1.62
Proteasome subunit alpha type-6	Psma6	1.93 ±1.54
26S proteasome non-ATPase regulatory subunit 3	Psmd3	1.91 ±2.13
40S ribosomal protein S27	Rps27	1.91 ±0.84
Adenine phosphoribosyltransferase	Aprt	1.84 ±1.62
COP9 signalosome complex subunit 7a	Cops7a	1.83 ±1.14
Cytochrome c oxidase subunit NDUFA4	Ndufa4	1.82 ±0.68
Interferon-induced protein with tetratricopeptide repeats 1	Ifit1	1.76 ±1.24
Proteasome subunit beta type-5	Psmb5	1.74 ±0.84
60S ribosomal protein L23	Rpl23	1.73 ±0.47
E3 ubiquitin-protein ligase TRIM32	Trim32	1.72 ±1
T-complex protein 1 subunit eta	Cct7	1.72 ±0.86
Ankyrin repeat domain-containing protein 17	Ankrd17	1.95 ±0.26
Pre-mRNA-processing factor 40 homolog A	Prpf40a	1.82 ±0.23
DnaJ homolog subfamily C member 9	Dnajc9	1.78 ±0.32
Hepatocyte growth factor-regulated tyrosine kinase substrate	*Hgs*	*1.73 0.25*
Ubiquitin-associated protein 2-like	Ubap2l	1.65 ±0.2
Nuclear pore complex protein Nup214	Nup214	1.65 ±0.17
Poly [ADP-ribose] polymerase 1	Parp1	1.6 ±0.47
Calponin-3	Cnn3	1.6 ±0.32
DnaJ homolog subfamily A member 2	Dnaja2	1.58 ±0.63
Serine/arginine repetitive matrix protein 2	Srrm2	1.55 ±0.6
Muscleblind-like protein 2	Mbnl2	1.54 ±0.34
Protein phosphatase 1 regulatory subunit 12A	Ppp1r12a	1.5 ±0.44

ZW10 interactor	Zwint	1.68 ±0.74	Poly(rC)-binding protein 1	Pcbp1	1.41 ±0.33
Cyclin-dependent kinase 1	Cdk1	1.59 ±0.65	**TAR DNA-binding protein 43**	**Tardbp**	**1.41 ±0.15**
ATP-dependent 6-phosphofructokinase, platelet type	Pfkp	1.57 ±0.67	Poly(rC)-binding protein 3	Pcbp3	1.35 ±0.16
Nuclear protein localization protein 4 homolog	Nploc4	1.54 ±1.38	5'-3' exoribonuclease 2	Xrn2	1.32 ±0.36
Large proline-rich protein BAG6	Bag6	1.48 ±0.66	Heterogeneous nuclear ribonucleoprotein F	Hnrnpf	1.29 ±0.43
26S proteasome non-ATPase regulatory subunit 14	Psmd14	1.48 ±0.56	Lamina-associated polypeptide 2, isoforms alpha/zeta	Tmpo	1.22 ±0.46
Malate dehydrogenase, cytoplasmic	Mdh1	1.48 ±0.43	Pumilio homolog 1	Pum1	1.22 ±0.07
ADP-sugar pyrophosphatase	Nudt5	1.47 ±0.21	Tropomodulin-3	Tmod3	1.21 ±0.37
26S protease regulatory subunit 6A	Psmc3	1.42 ±0.87	14-3-3 protein beta/alpha	Ywhab	1.15 ±0.43
Bifunctional glutamate/proline--tRNA ligase	Eprs	1.39 ±0.81	Plectin	Plec	0.97 ±0.2
Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	Aimp1	1.3 ±0.91	Small ubiquitin-related modifier 1	Sumo1	0.83 ±0.22
Ribosome-binding protein 1	Rrbp1	1.26 ±0.71	Regulator of nonsense transcripts 1	Upf1	0.82 ±0.21
40S ribosomal protein S27-like	Rps27l	1.23 ±0.68	Vimentin	Vim	0.72 ±0.13

Ras GTPase-activating protein-binding protein 1	G3bp1	1.18 ±0.66
Glyceraldehyde-3-phosphate dehydrogenase	Gapdh	1.15 ±0.48
Dihydropyrimidinase-related protein 2	Dpysl2	1.09 ±0.45
Ataxin-10	Atxn10	1.05 ±0.29
Dihydropyrimidinase-related protein 3	**Dpysl3**	**1.05 ±0.08**
ATP-binding cassette sub-family E member 1	Abce1	1.04 ±0.34
60S ribosomal protein L38	Rpl38	1.03 ±0.39
Green fluorescent protein	GFP	1.01 ±0.4
Multifunctional protein ADE2	Paics	0.98 ±0.35
Polymerase delta-interacting protein 3	Poldip3	0.97 ±0.15
Melanoma-associated antigen D1	Maged1	0.94 ±0.31
Dihydropyrimidinase-related protein 1	Crmp1	0.92 ±0.19
60S acidic ribosomal protein P0	Rplp0	0.89 ±0.35
ADP/ATP translocase 2	Slc25a5	0.89 ±0.17
T-complex protein 1 subunit beta	Cct2	0.88 ±0.22
Eukaryotic translation initiation factor 3 subunit D	Eif3d	0.87 ±0.32
Nuclear pore complex protein Nup98-Nup96	Nup98	0.68 ±0.15
Importin subunit alpha-1	Kpna2	0.58 ±0.1
Protein C10	Grcc10	0.56 ±0.01

T-complex protein 1 subunit delta	Cct4	0.84 ±0.27
Tubulin beta-5 chain	Tubb5	0.79 ±0.14
26S protease regulatory subunit 10B	Psmc6	0.76 ±0.15
Golgi-associated plant pathogenesis-related protein 1	Glipr2	0.75 ±0.22
IgE-binding protein	Iap	0.75 ±0.22
Cell division control protein 42 homolog	Cdc42	0.71 ±0.15
Non-POU domain-containing octamer-binding protein	Nono	0.67 ±0.1
Aspartate aminotransferase, cytoplasmic	Got1	0.45 ±0.05

* Only proteins that meet significance cut-off (hyperbolic curves, permutation-based FDR≤0.05, S_0=0.1). Full table of proteins are shown in Table S2; **Bold** are genes with known causes or risk factors for MND (or other neurodegenerative diseases in the case of Picalm); *Italics* are cellular proteins previously seen to become more insoluble when Httex1Q_{97} formed inclusions [58]

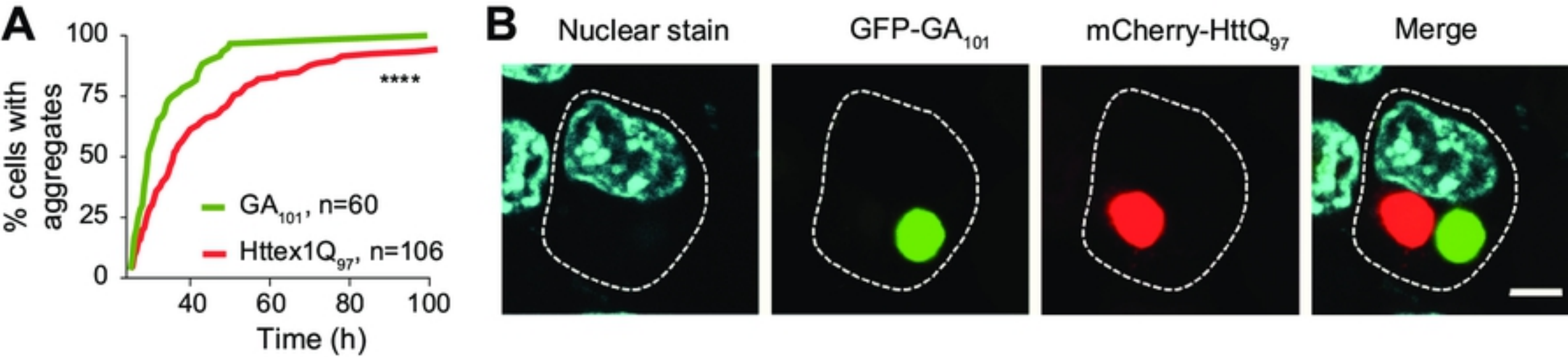
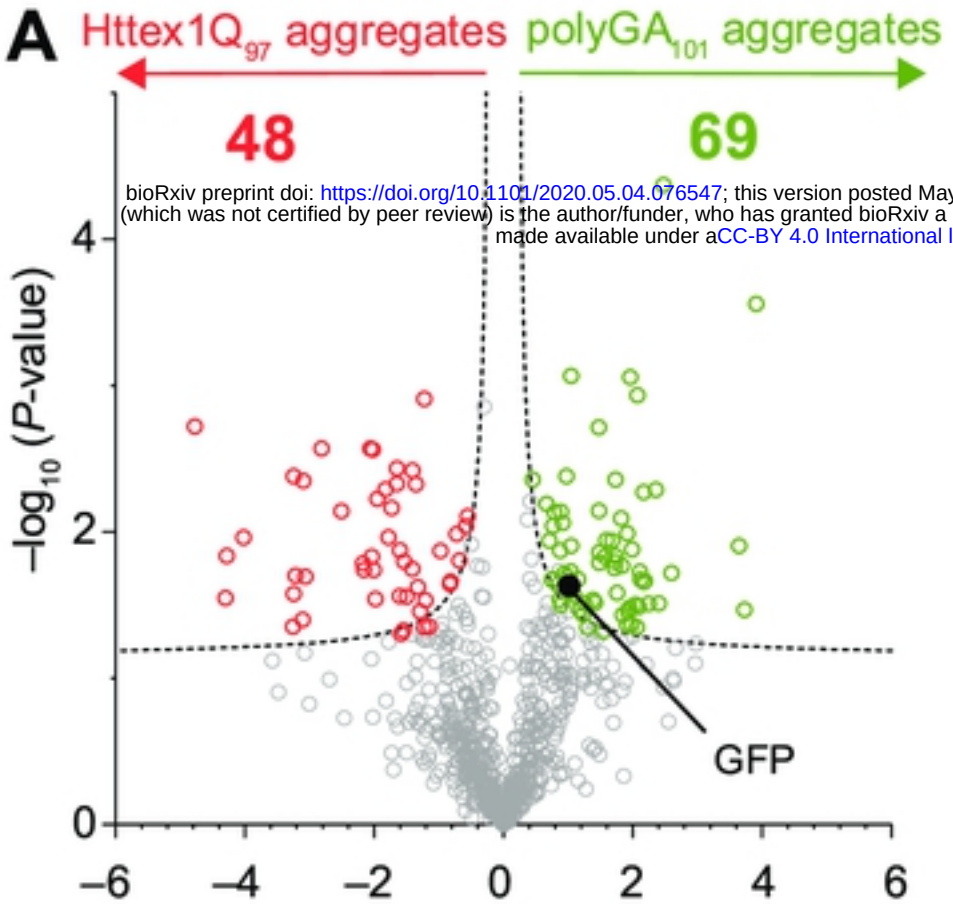
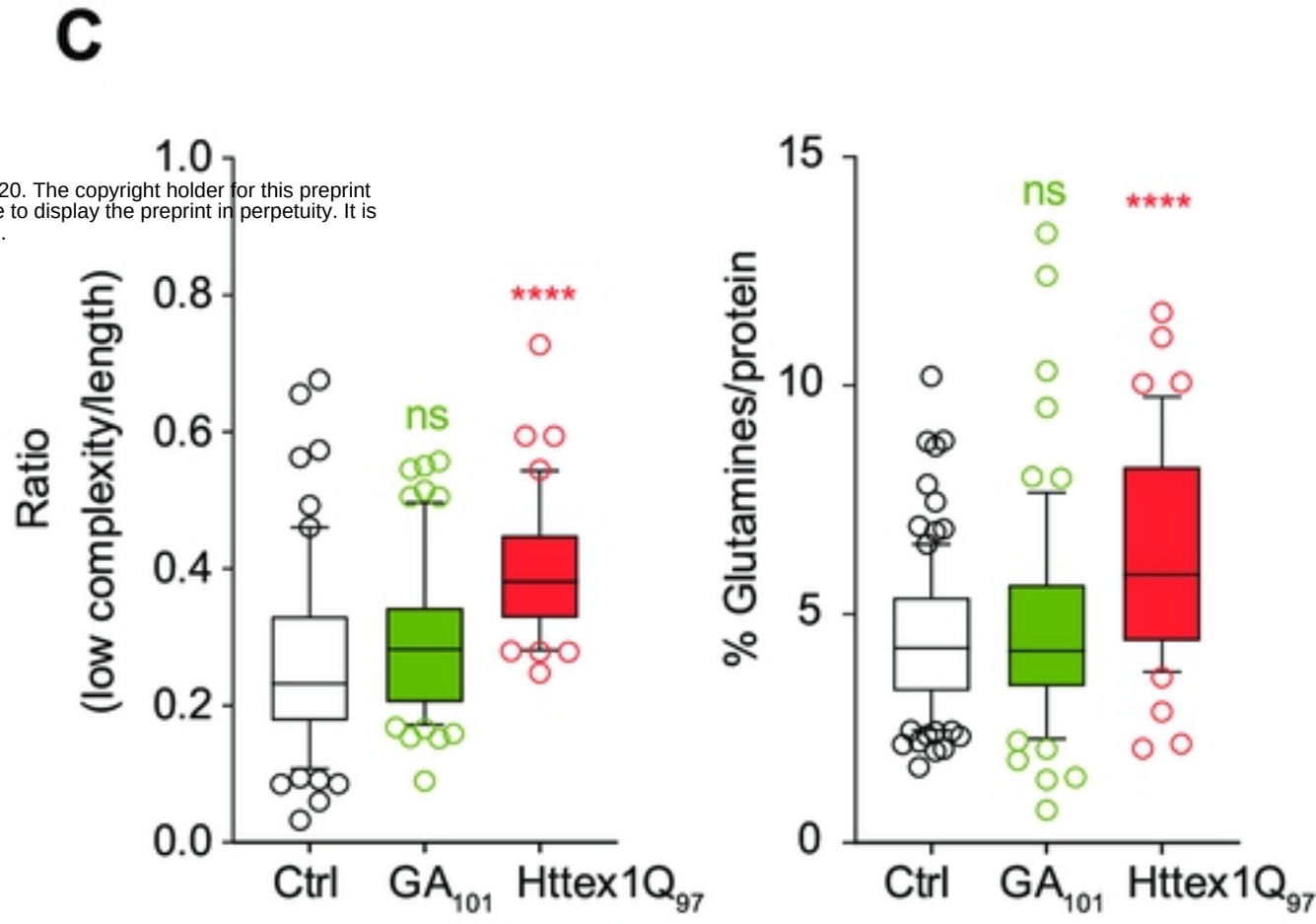

A
100
75
50
25
0
% cells with aggregates
****
GA101, n=60
Httex1Q97, n=106
40
60
80
100
Time (h)
B
Nuclear stain
GFP-GA101
mCherry-HttQ97
Merge

Figure 1





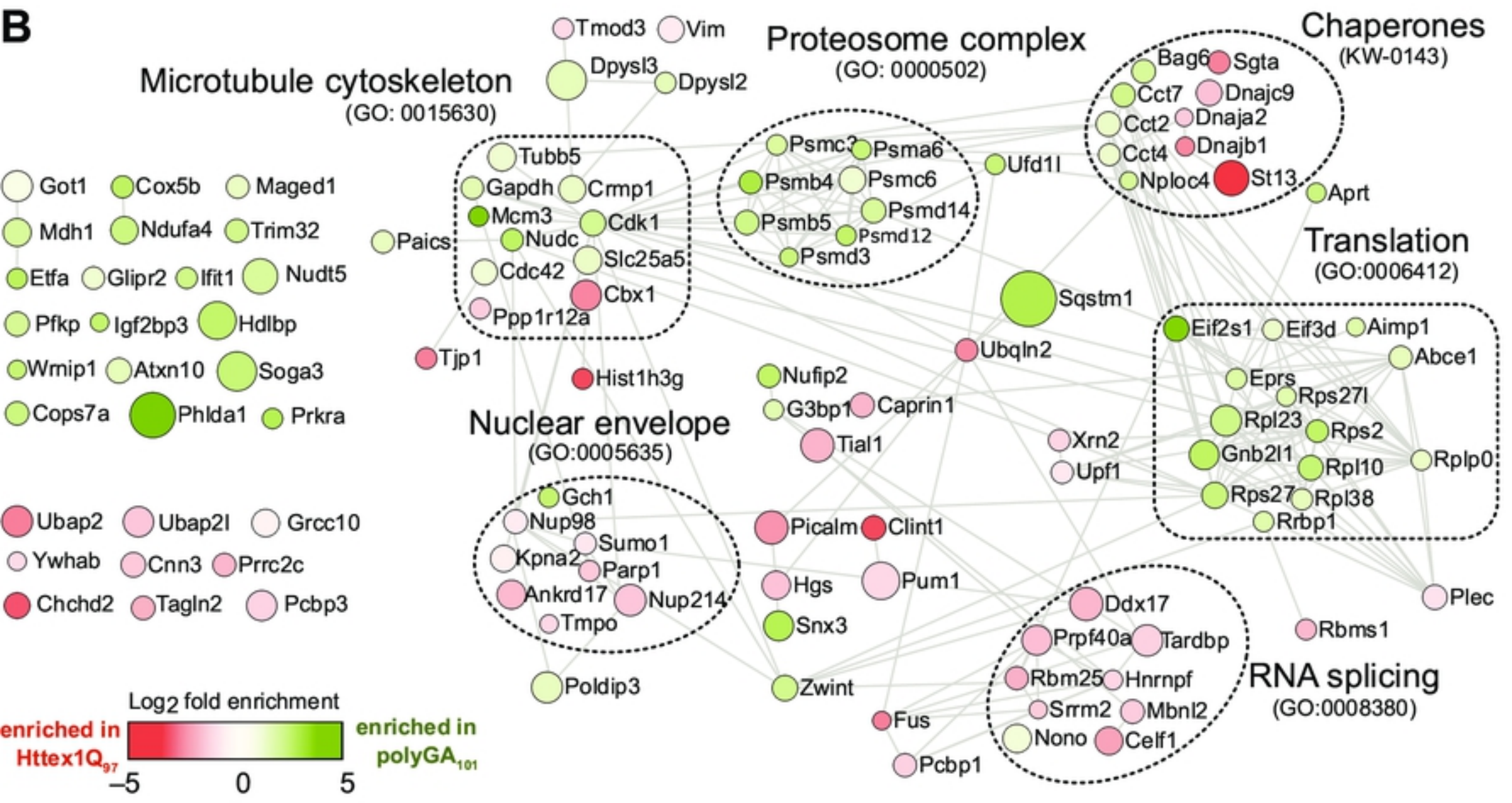
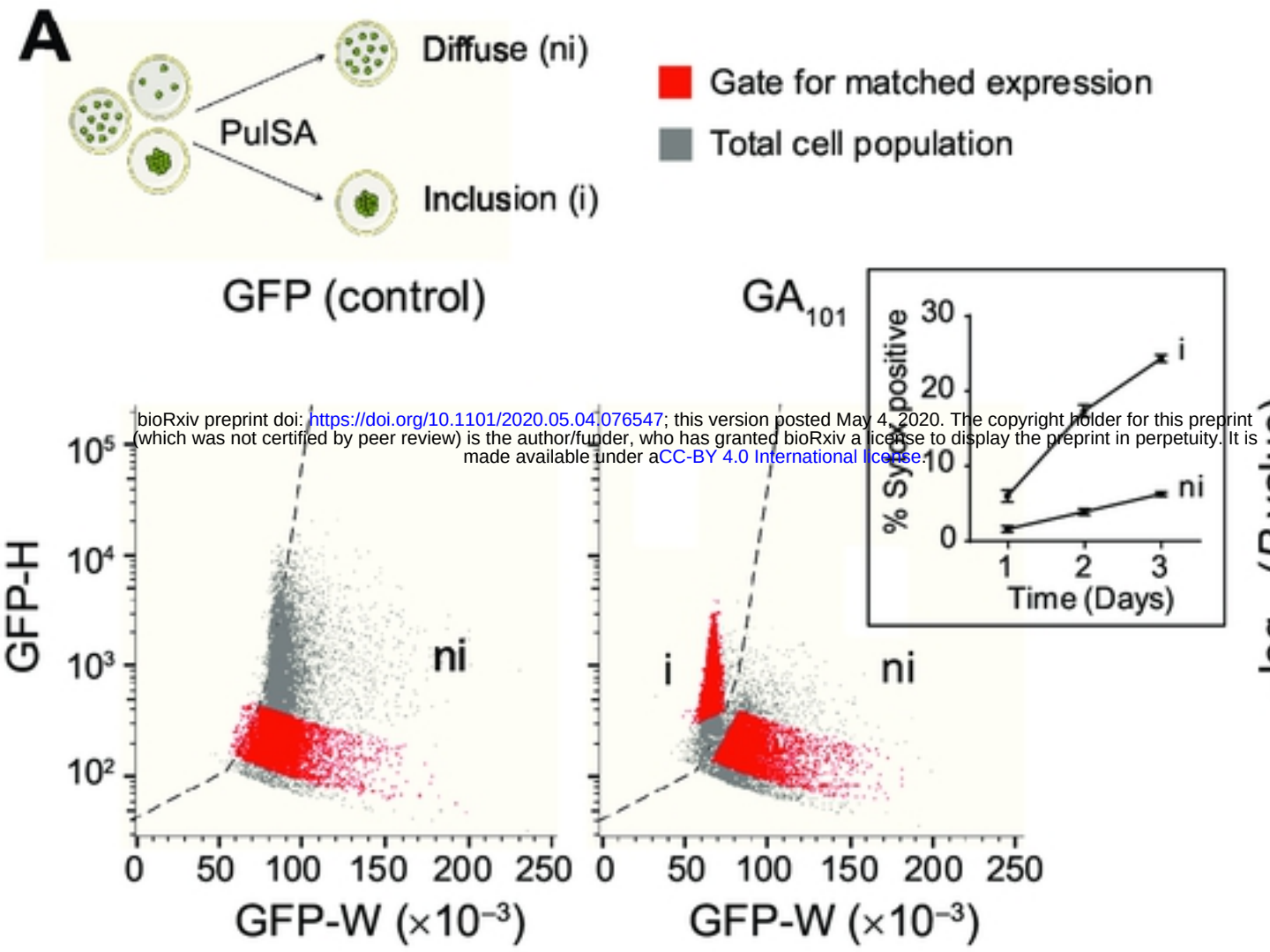
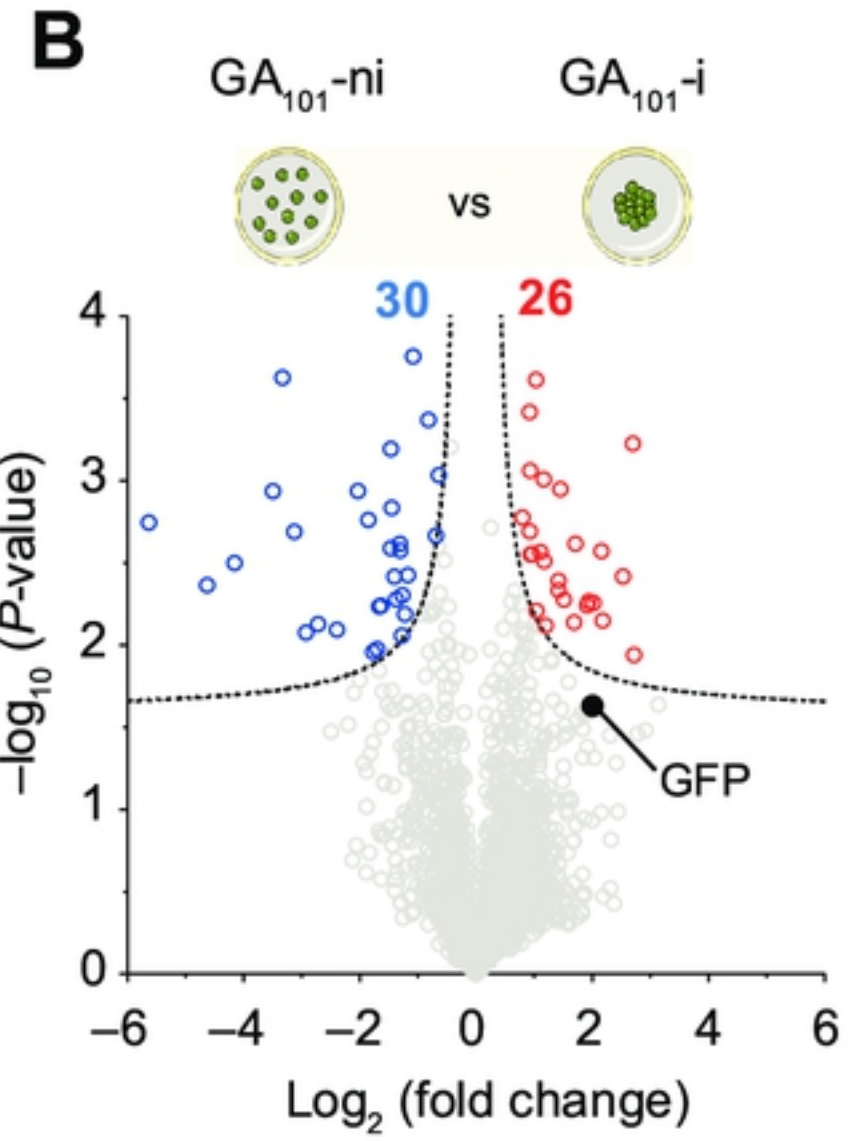


Figure 2





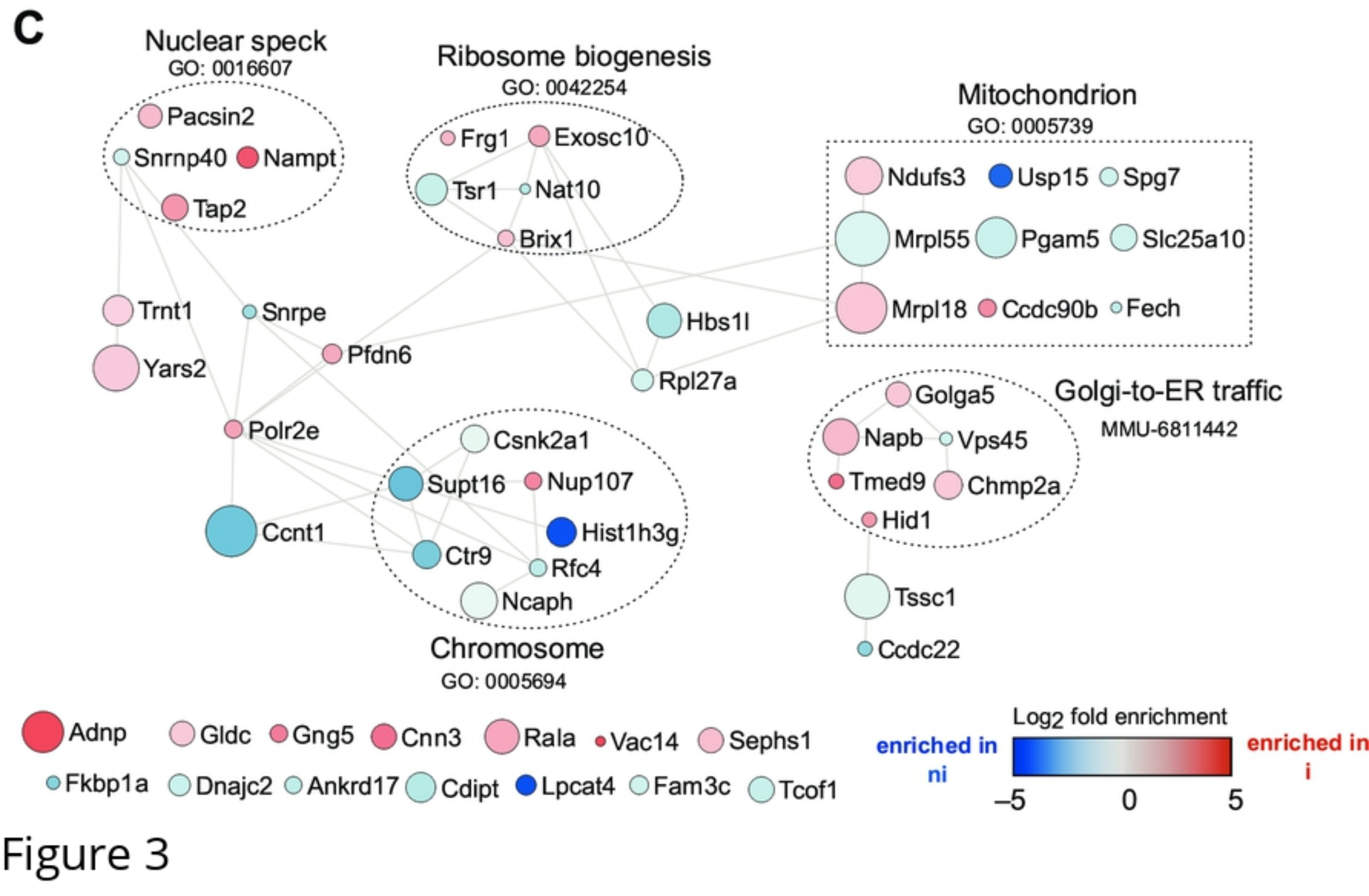


Figure 3