Compartment-specific proteostasis

Distinct protein quality control pathways in the nucleus and cytosol

Master Research Project 1

Biomedical Sciences

University of Groningen

Author: Robin Kalsbeek

Date: July 26th, 2021

Research group: Dr. M.S. Hipp

Supervisor: Dr. C.L. Klaips

Acknowledgement

I would like to express my gratitude to Prof Dr. Harrie Kampinga for giving me the opportunity to do my internship and project in his lab, and for providing the facilities that were required.

I would like to kindly thank Dr. Mark Hipp for giving me an opportunity to work in his young, international team, and for the fresh ideas during the course of the work.

I would like to sincerely thank Dr. Courtney Klaips, for her professional guidance and support, for teaching me a lot about scientific research, for the words of encouragement, and for a great time in general.

I would like to thank Joost Hendriks, for additional supervision and support at the lab. I wish you the best of luck on your Ph.D. journey.

I would like to thank Klaas Sjollema for his expertise and assistance in microscopy.

I would like to thank Mirjam Baanstra for her assistance with the flow cytometry experiments.

I would like to thank Maria van Waarde-Verhagen for the expert technical help with the luciferase activity assay.

I would like to thank Jeanette Brunsting for the everyday support at the lab.

I would like to thank, in no particular order, Eduardo, Joris, Maiara, Wouter, Els, Sophie, Danara, Marleen, and Suzanne, for the fun times inside and outside the lab.

Abstract

Protein quality control (PQC) is of vital importance of the cell. Since the various cellular compartments are confronted with distinct challenges, it is natural that the different compartments have evolved distinct PQC pathways. While much work has focused on the PQC pathways of the cytosol and ER, the PQC networks inside the nucleus are to this date the more elusive. The thermolabile protein firefly luciferase offers many benefits to study protein quality control and is therefore a very popular tool. However, the studies that have used firefly luciferase in the past were often lacking in comparing their findings to standardized control proteins. As such, in the present study we have put together a standardized, modularized library of plasmids that express proteins with varying thermodynamic properties, that localize to different compartments, and are fluorescently tagged. We show that nuclear localization signal (NLS)-Firefly luciferase-GFP localizes to the nucleolus upon heat stress, but that NLS-Renilla luciferase-GFP forms peri-nucleolar foci upon heat stress. We show that the use of a nuclear export signal (NES) improves cytoplasmic localization, but also causes accumulation at the aggresome. We have generated plasmids that express glutathione S-transferase (GST) fusion proteins that are low expressed and stable, and double fluorophore proteins that are highly expressed and stable. Lastly, we have made plasmids that express targeted versions of poly-glutamine (poly-Q) disease-related model proteins, an inert huntingtin exon 1 (Htt) 25Q-GFP and an aggregation prone Htt97Q-GFP. We conclude that the differences between misfolded proteins in the nucleus are the result of two distinct nuclear PQC pathways, and that these may play a role in the context of disease.

Table of Contents

Introduction

Proteins are biological macromolecules that play many critical roles in the cell. The correct functioning of proteins depends largely on their three-dimensional conformation or fold. The native fold of proteins is encoded in their amino acid sequences (Anfinsen, 1973). However, to ensure proteins are folded correctly in the crowded environment of the cell, additional protein quality control (PQC) factors including molecular chaperones assist in the folding of newly synthesized proteins (Frydman, 2001; Hartl et al., 2011). During the life cycle of a protein, its conformation has to be maintained by this same chaperone network.

The cell can experience a multitude of proteotoxic stress factors, such as pH, heat, and oxidative stress, that can cause proteins to become partially unfolded. As the protein misfolds, hydrophobic amino acid residues that are normally turned inward (intramolecular contacts) tend to be exposed. These hydrophobic regions can coagulate with other proteins (intermolecular contacts) and form aggregates that can be toxic to the cell. First, since the protein has lost its native conformation, it can no longer serve its purpose within the cell (loss of function). Second, the newly exposed regions of the protein can interact with other proteins (gain of function) and sequester them into the forming aggregate.

Some proteins, such as those with intrinsically disordered regions (IDRs), or low-complexity domains (LCDs), are highly sensitive to proteotoxic stress conditions, and become readily misfolded at increased temperatures (Dunker et al., 2008). Other proteins are destabilized by a genetic mutation, such as a poly-glutamine (poly-Q) trinucleotide repeat expansion in the *huntingtin* gene (Htt) causing Q-length-dependent aggregation of huntingtin (MacDonald et al., 1993). In all cases, the aggregates that are formed when proteins misfold can give rise to serious diseases that are collectively called proteinopathies.

To prevent formation of these toxic aggregates, terminally misfolded proteins can be recognized and degraded by autophagy or the ubiquitin proteasome system (UPS)(Tsuchiya et al., 2017). However, proteasomal degradation requires substrate solubility (Korolchuk et al., 2010; K. Wang et al., 2015). To this end, chaperones cooperate with the UPS to maintain solubility of the substrate, and help facilitate its ubiquitylation and subsequent proteasomal degradation. To sum up, molecular chaperones are involved in all parts of the protein life cycle, and are vital for proteostasis.

Cellular quality control compartments

The eukaryotic cell is compartmentalized by different specialized organelles. Inside these compartments are vastly distinct environments that are presented with unique challenges, and that therefore have individual needs with respect to protein quality control. The cell is adaptive to proteotoxic stress, in that it upregulates PQC factors through different pathways, increasing refolding and degradation capacities in the various compartments.

The cytoplasm is a reducing, highly crowded environment where the majority of proteins are synthesized. During nascent folding, hydrophobic regions are exposed that can undergo harmful interactions with other proteins. Specialized ribosome-associated chaperone complexes interact co-translationally with these vulnerable sequences to prevent premature misfolding (Preissler & Deuerling, 2012). When proteins do misfold in the cytoplasm due to intrinsic or extrinsic stressors, the heat shock response (HSR) is activated. The expression of various chaperones is upregulated through the HSR pathway, most

prominently under the transcription factor heat shock factor 1 (HSF1), increasing refolding and degradation capacities.

One-third of all synthesized proteins, mostly membrane proteins and secretory proteins, are folded and assembled inside the endoplasmic reticulum (ER), an oxidizing environment (Thibault & Ng, 2012). Inside the ER lumen, HSPA5, an ER-specific chaperone of the Hsp70 family, recognizes hydrophobic regions on proteins and folds them under stimulation by J proteins and nucleotide exchange factors. When the burden of unfolded proteins inside the ER exceeds a certain threshold, the unfolded protein response (UPR) is triggered by HSPA5. Unfolded and misfolded proteins are then targeted for degradation in the cytoplasm through a specialized UPS pathway: ER-associated protein degradation (J. Wang et al., 2017).

The nuclear envelope, nuclear pore complexes, and tightly regulated nuclear transport mechanisms together maintain the segregation of nucleoplasm and cytoplasm (D'Angelo et al., 2009; Hutten & Dormann, 2019). Ensuring this segregation is vital, since the integrity of genomic DNA inside the nucleus is of utmost importance to the cell. Although no protein synthesis occurs inside the nucleus, it contains many RNA-binding proteins that can only function through their disordered regions (Järvelin et al., 2016). As such, these proteins are vulnerable to pathological interactions with other unfolded or misfolded proteins (Chen-Plotkin et al., 2010; Mackenzie et al., 2010). Chaperones of the Hsp70 family are present in the nucleus to prevent the formation of aggregates, but under acute proteotoxic stress they rely on other PQC components to be imported from the ER or the cytoplasm. It is known that, upon heat stress, ATP-bound Hsp70 chaperones are actively imported by the protein Hikeshi, however, much about these nuclear PQC pathways is still unknown (Kose et al., 2012; Shibata & Morimoto, 2014).

The nucleoplasm itself is also further compartmentalized. The nucleolus is the most prominent subcompartment: a membraneless organelle, phase separated by disordered regions in fibrillarin (FBL) and nucleophosmin (NPM) and their interactions with rRNA. The nucleolus is formed around ribosomal DNA, and best known for its involvement in the assembly of ribosomal subunits (Latonen, 2019; Spector, 2001). In addition, the nucleolus serves an important purpose as a stress response organelle where stress-responsive proteins are sequestered upon various stress signals, including upon loss of proteostasis. In conclusion, the different cellular compartments have evolved distinct pathways to cope with proteotoxic stress. Although many aspects of cytoplasmic PQC and ER PQC have been unraveled, much is still unknown about nuclear proteostasis networks.

The same protein in different compartments

The differences in PQC pathways between cellular compartments can be explained as a consequence of the distinct environments, or as an adaptation to properties of compartment-specific protein species. Differentially targeted model proteins have proven valuable in studying the differences in PQC pathways between the different compartments.

Although protein aggregation in disease occurs both in the cytoplasm and in the nucleus, it has been shown that certain artificial aggregation-prone proteins have reduced toxicity when targeted to the nucleus, compared to the cytoplasm (Woerner et al., 2016). These artificial proteins serve no biological function and were designed specifically to form cross-β fibrils. When targeted to the cytoplasm, these β-sheet proteins formed toxic aggregates that sequestered nuclear transport factors, causing nucleocytoplasmic transport defects. Aggregates were still formed when targeted to the nucleus, however,

these inclusions colocalized with the nucleoli and showed coimmunoprecipitation with NPM, alluding to the protective role of the nucleolus upon proteotoxic stress.

The same protective role of the nucleolus has been shown to be true for a model protein that aggregates during stress, firefly luciferase. Upon heat stress, firefly luciferase misfolds readily and localizes to the nucleolus, a process that was shown to be driven by Hsp70. Moreover, as temperatures were decreased to normal growth temperatures, it was shown that this localization was reversible. At 37 ºC, firefly luciferase was extracted from nucleoli and refolded, again dependent on Hsp70 chaperones, suggesting that PQC pathways exist within the nucleus (Frottin et al., 2019; Nollen et al., 2001).

Firefly luciferase is a popular model protein to study proteostasis pathways, for various reasons. For one, it is a thermolabile protein, that is relatively stable at a normal growth temperature of 37 ºC, but easily misfolds when the temperature is increased by a few degrees, mimicking a non-lethal heat stress to the cell. Secondly, since it is not an endogenous protein, firefly luciferase expression does not generally interfere with cellular pathways. Moreover, as firefly luciferase is a bioluminescent protein, its enzymatic activity – and therefore 'foldedness' – under various conditions can be analyzed relatively easily.

In addition to its intrinsic instability, the thermolabile properties of firefly luciferase have been additionally modified by characterizing point mutations that destabilize the protein even further. The strongest effects on thermodynamic stability were achieved with the point mutation R188Q and further exacerbated with the additional point mutation R261Q. These variants were defined as firefly luciferase single mutant (SM) and double mutant (DM), respectively (Gupta et al., 2011).

Controls needed to make direct comparisons

Luciferase has been an important tool in dissecting the various protein quality control pathways at work in the cell. However, the reliance on one type of protein, particularly when making comparisons between different compartments across decades makes it difficult to conclude how much of these pathways are truly compartment-specific and how much of these pathways would be conserved broadly for more heterogeneous, endogenous substrates. In the present study we are trying to answer how much of these findings are specific to firefly luciferase, and how much of these findings are specific to the different compartments. As such, we have put together a standardized, modularized library of plasmids that express proteins with varying thermodynamic properties to study the different pathways in depth.

Methods

Cloning of plasmids

Plasmids were generated through restriction-ligation cloning using various endo-nucleases, Alkaline Phosphatase Calf Intestinal (CIP), and T4 ligase (all: New England Biolabs). Polymerase chain reaction (PCR) was performed using Q5 DNA Polymerase (New England Biolabs). For detailed information on cloning, please refer to Table S1 in Supplementary information. Cloning was confirmed by either restriction digestion and gel electrophoresis, or Sanger sequencing (Eurofins Genomics). DNA sequences were aligned with MUSCLE (Madeira et al., 2019). Plasmid maps were assembled from confirmed aligned sequences using SnapGene software (Insightful Science).

Cell culture

HEK293T cells were obtained from ATCC and were cultured at 37 ºC and 5% CO_2 in Dulbecco's modified Eagle medium (DMEM, Gibco), supplemented with 10% fetal bovine serum, and 1% penicillin-streptomycin. Cells were regularly checked for mycoplasma infection. Heat shock experiments were conducted in a separate dry incubator at 43 ºC and 5% CO_2.

Transfection

Transient transfections were performed by combining 1.0 µg plasmid DNA with 6.0 µL polyethylenimine (PEI) in unsupplemented DMEM (Gibco) and adding this to the media. Plasmids were combined at a 0.5 µg : 0.5 µg ratio with either pCDNA5 FRT TO (empty vector, EV) or another plasmid of interest. After 3 h incubation, media was changed to culturing medium as described above.

Widefield microscopy

Widefield fluorescent microscopy was performed on a Leica (Wetzlar, Germany) DM6B-Z microscope equipped with a Leica HC PL APO 40x/0.95 CORR dry objective and a Leica DFC 7000T camera. Micrographs were captured using Leica Application Suite X (Leica) software and analyzed in Fiji (Schindelin et al., 2012).

Confocal laser scanning microscopy

Confocal laser scanning microscopy was performed at the UMCG Imaging and Microscopy Center (Groningen, the Netherlands), on a Leica (Wetzlar, Germany) TCS SP8 microscope equipped with a Leica HC PL APO CS2 63x/1.4 oil immersion objective. The Hoechst 33342 staining was detected using an excitation wavelength of 405 nm and emission of 415-485 nm. The GFP fluorophore was detected using an excitation wavelength of 488 nm and emission of 500-550 nm. The mScarlet fluorophore was detected using an excitation wavelength of 552 nm and emission of 590-700 nm (Bindels et al., 2017). The Alexa Fluor 633 staining was detected using an excitation wavelength of 638 nm and emission of 650-750 nm. Micrographs were captured using Leica Application Suite X software and analyzed in Fiji (Schindelin et al., 2012).

Fixating and immunocytochemistry and fluorescent dye labeling

Cells were grown on PLL-coated coverslips, fixated in 3.7% formaldehyde in PBS for 15 mins and permeabilized in 0.1% Triton X-100 in PBS for 5 mins. For immunocytochemistry, coverslips were blocked in PBS+ (0.5% BSA, 0.3% glycine in PBS) and incubated

with primary antibodies overnight at 4 °C. After washing with PBS+, coverslips were incubated with secondary antibodies and washed again. For a list of antibodies used, please refer to Table S3 in Supplementary information. Nuclei were counterstained with Hoechst 33342 (10mg/ml in H_2O) for 5 minutes, before mounting slides with Citifluor (Electron Microscopy Sciences, Hatfield, PA) AF1 mountant solution.

Luciferase activity assay

Cells were lysed in ice-cold 1:1 diluted BLUC lysis buffer (25 mM Tris H_3PO_4 pH 7.8, 10 mM $MgCL_2$, 1% Triton X-100, 15% glycerol, 1 mM EDTA). Equal volumes of cell lysate were analyzed in triplicate using a Berthold (Bad Wildbad, Germany) Sirius luminometer while ensuring that all measurements were within linear detection range. RLU/s measurements were taken with optimized instrument parameters: measurement delay time 2.0 s, measurement duration 10.0 s, injector delay 2.0 s, injected BRLUC (1.25 mM ATP, 3.5 mg/ml D-luciferin, in BLUC) volume 100 µl.

SDS-PAGE and immunoblotting

Cells were lysed in ice-cold RIPA lysis buffer (25 mM Tris HCl, 150 mM NaCl, 1% Igepal CA-630, 1% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS), 1 mM $MgCl_2$) supplemented with 1x cOmplete protease inhibitor cocktail (EDTA-free, Roche) and DENARASE (c-LEcta, 50 U/ml). Protein samples were normalized using the DC Protein Assay (Bio-Rad) and suspended in sample buffer (50 mM Tris pH 6.8, 8% glycerol, 2% SDS, 10% β-mercaptoethanol, containing bromophenol blue) and heated to 99 °C for 5 minutes, and separated by electrophoresis on 10% or 12% self-made TGX FastCast acrylamide solution (Bio-Rad) gels using Laemmli running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS) at 30 mA per gel. Proteins were transferred from poly-acrylamide gels to nitrocellulose membranes using the Trans-Blot Turbo Transfer System (Bio-Rad) at 25V, 2.5A for 10 mins. Membranes were washed with PBS-T, blocked with 10% powdered milk in PBS-T and incubated with primary antibodies overnight at 4 °C. After washing with PBS-T, membranes were incubated with horseradish peroxidase-conjugated secondary antibodies and washed again. Immunoblots were visualized using chemiluminescence (ECL, Thermo Scientific) and imaged with a Bio-Rad ChemiDoc Touch imaging system. Protein densitometry was performed using Image Lab 6.1 (Bio-Rad) software. For a list of antibodies used, please refer to Table S3 in Supplementary information.

Sedimentation assay

Cells were lysed and protein levels were normalized as described as above. The homogenate was centrifuged at 15,000x g for 15 mins at 4 °C and the pellet fraction was resuspended in RIPA buffer. The resulting supernatant and pellet fraction were further processed for immunoblotting.

Flow cytometry

Cells were dissociated with trypsin-EDTA, washed with 1% BSA in PBS, and resuspended in 0.2% BSA in PBS. Cell suspensions were filtered through a 35-µm nylon cell strainer (Corning Life Sciences) and stained with 50 µg/ml propidium iodide (Roche). Flow cytometry was performed at the UMCG Flow Cytometry Unit (Groningen, the Netherlands) on a BD Biosciences (San Jose, CA) LSR II flow cytometer and measurements were taken with BD FACSDiva 8.0.1 software. Analysis was performed with FlowJo 10.7.2 (Beckton, Dickinson and Company) software.

Results

Creation of a library of compartmentalized model proteins

In order to study how the different compartments within the cell handle protein misfolding under stress, we put together a library of plasmids for expression in mammalian cells. The plasmid library was designed to be modular to facilitate future expansion. The DNA constructs consist of three modules that encode a fusion protein: a model protein of interest, tagged with an N-terminal signaling peptide, and tagged with a C-terminal fluorophore (Fig. 1). Between these modules, one or more endonuclease restriction sites were placed for interchangeability.

Our main interest lies in the thermolabile model protein firefly luciferase and different mutant variants that make it less stable (Gupta et al., 2011). For comparison, we are also studying the model protein Renilla luciferase, which, like firefly luciferase, is a bioluminescent protein, but is thought to be more stable. Furthermore, we have designed single and double fluorophore control proteins as highly abundant, stable control proteins. To create stable control proteins that are expressed at levels similar to firefly luciferase we have used glutathione S-transferase (GST) as a model protein. Lastly, to study how relevant the properties of a thermolabile protein are in the context of disease, we have made targeted versions of Htt25Q-GFP and Htt97Q-GFP, based on Huntingtin exon-1.

The N-terminal signaling peptides include the nuclear localization signal (NLS) of SV40 T antigen and a consensus nuclear export signal (NES) for cytoplasmic localization (Gupta et al., 2011). Some of the constructs were also generated without a signaling peptide. At the C-terminus all constructs were tagged with either EGFP or with mScarlet fluorophores for detection by both microscopy and immunoblotting (Bindels et al., 2017). Since all constructs exist in both a GFP-tagged form and an mScarlet-tagged form, all combinations can be made for co-expression. All plasmid sequences have been confirmed by either restriction digestion and gel electrophoresis, or Sanger sequencing. Plasmid stocks and bacteria stocks have been frozen down and catalogued. We next turned towards an initial characterization of the various constructs, focusing primarily on GFP-tagged constructs.

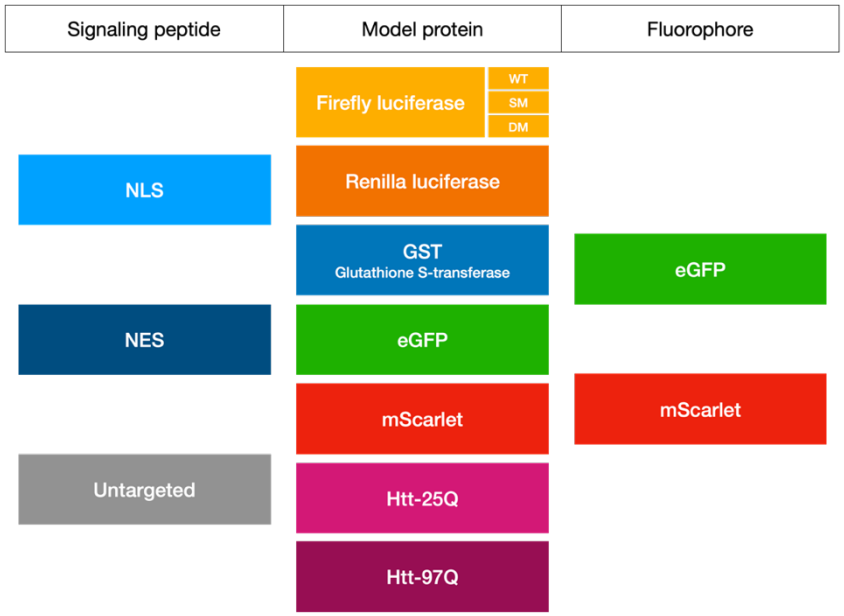

Figure 1. Schematic representation of modular plasmid library. For a complete list of the plasmids that were generated, please refer to Table S1 in Supplementary information.

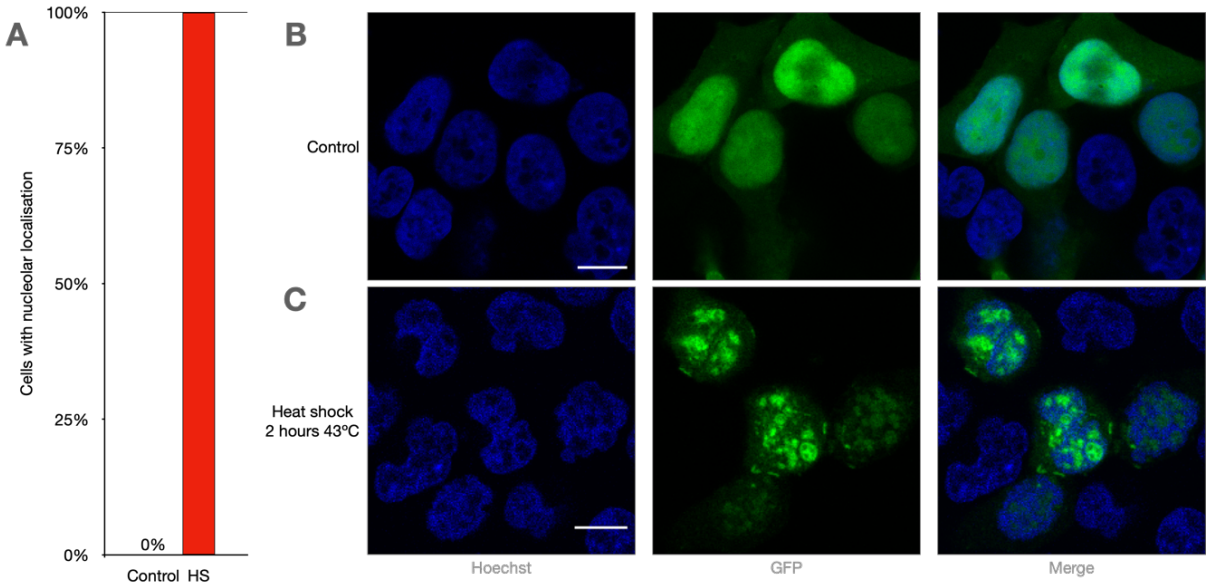

*Figure 2. NLS-LG WT was expressed for 48 h in HEK293T cells. One half was subjected to two hours heat shock (43 °C) before fixating and mounting. **(A)** Percentage of cells with foci without heat shock (blue bar, n = 138) and with heat shock (red bar, n = 240). Graphs represent mean. **(B-C)** Confocal laser scanning micrographs of NLS-LG WT expressing cells without heat shock (B) and with heat shock (C). Representative images shown. Scale bar denotes 10 µm.*

NLS-LG localizes to nucleolus upon 2 hours of heat shock

Previous studies have shown that NLS-Luciferase-GFP (NLS-LG) is a thermolabile protein that localizes to nucleoli upon misfolding, in a process driven by Hsp70 (Frottin et al., 2019; Nollen et al., 2001). To confirm that NLS-LG has the same behaviour in our hands, we expressed NLS-LG wild type (WT) in HEK293T cells for 48 h, and subjected one half of the cells to 2 hours heat stress (43 °C). Under normal growth conditions, we observed using fluorescent microscopy, that NLS-LG WT is a soluble protein (Fig. 2a), that exists primarily inside the nucleus, since it colocalizes with the Hoechst staining (Fig. 2b). Upon two hours of heat shock at 43 °C we observed in all expressing cells that NLS-LG WT has aggregated into nuclear foci with Hoechst-dim localization (Fig. 2c).

NES-LG forms cytoplasmic foci upon 2 hours of heat shock

To study how the same thermolabile protein would behave when localized to the cytoplasm, we tagged LG with an NES. We observed that under normal growth conditions, in most of the expressing cells NES-LG WT exists as a soluble protein that localizes to the cytoplasm, since it does not colocalize with the Hoechst staining. However, we found that in 5% of expressing cells, NES-LG WT had aggregated and formed cytoplasmic foci (Fig. 3a, 3b). Upon two hours of heat shock at 43 °C, NES-LG WT has aggregated in nearly all expressing cells. These aggregates showed a granular appearance throughout the cytoplasm, as well as larger protein deposits that localized to the concave part of the nucleus, which is thought to be the aggresome (Fig. 3c). The aggresome is a structure near the microtubule-organizing center (MTOC), where aggregated proteins are sequestered when the capacity of proteasomal degradation is exceeded (Johnston et al., 1998).

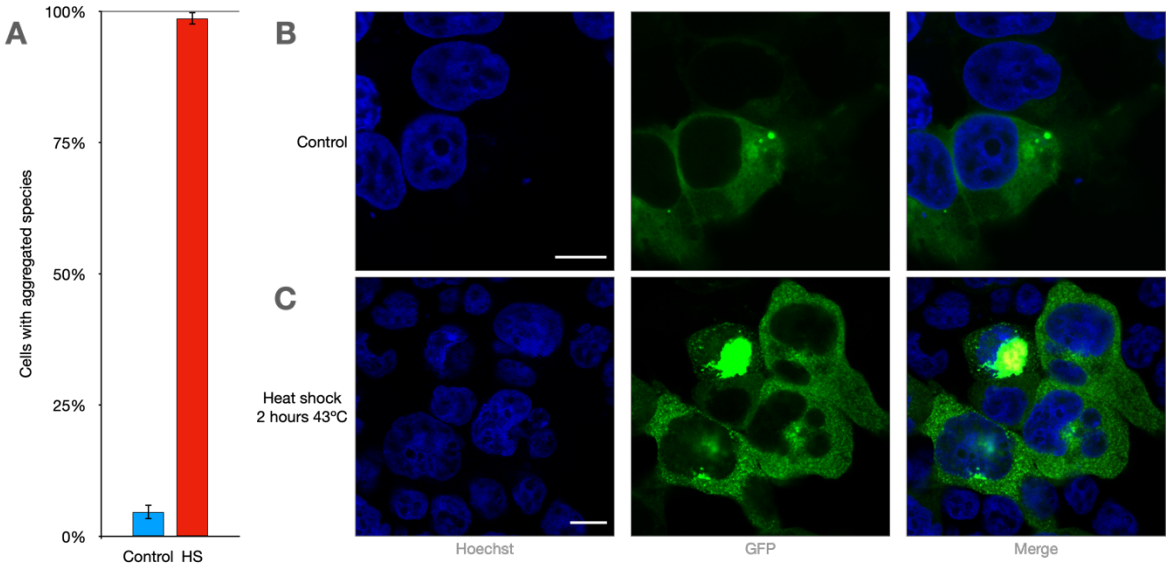

Figure 3. NES-LG WT was expressed for 48 h in HEK293T cells. One half was subjected to two hours heat shock (43 ºC) before fixating and mounting. (A) Percentage of cells with foci without heat shock (blue bar, n = 279) and with heat shock (red bar, n = 233). Graphs represent mean ± standard deviation. (B-C) Confocal laser scanning micrographs of NES-LG WT expressing cells without heat shock (B) and with heat shock (C). Representative images shown. Scale bar denotes 10 µm.

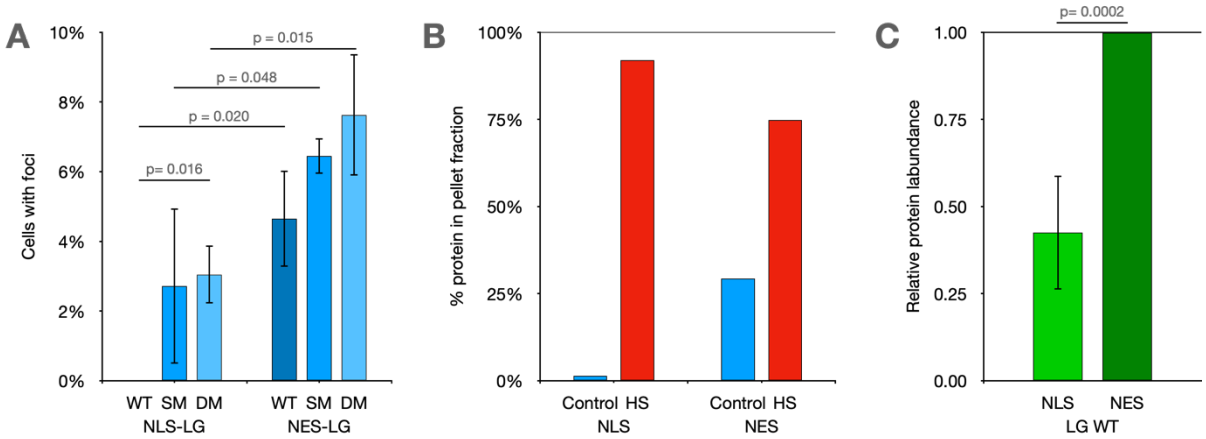

Figure 4. (A) NLS-LG WT, SM, and DM, and NES-LG WT, SM, and DM were expressed for 48 h in HEK293T cells. Percentage of cells with foci under normal growth conditions. Graphs represent mean ± standard deviation. All p-values are Student's t-test, two tailed, two-sample assuming equal variance (n = 138, 305, 231, 279, 326, 228). (B) Pelleting assay on lysates of cells expressing either NLS-LG WT or NES-LG WT, with or without 2 hours of heat shock (43 ºC). Immunoblot analysis was performed using specific antibodies against GFP and GAPDH and visualized using peroxidase-labeled secondary antibodies. Percentage protein in the pellet fraction was quantified as $GFP_{pellet}/(GFP_{supernatant} + GFP_{pellet})$. n = 1. (C) Relative abundance of NLS-LG WT and NES-LG WT under normal growth conditions after 48 h of expression. Graphs represent mean ± standard deviation. Immunoblot analysis was performed using specific antibodies against GFP and GAPDH and visualized using peroxidase-labeled secondary antibodies. GFP quantification was normalized to GAPDH, before standardizing to NES-LG WT. P-values are Student's t-test, two tailed, paired (n = 5).

Mutant luciferases form more foci at 37 ºC than wild-type, and NES-LG forms more foci at 37 ºC than NLS-LG

The mutant variants LG SM and DM were engineered to be already unstable at 30-37 ºC (Gupta et al., 2011). To study how the altered thermodynamic properties of LG SM and LG DM would behave when localized to the different compartments, we expressed all 6 LG constructs (Fig. 1) in cells. Since we had seen that NES-LG WT already aggregated in cells upon heat treatment, here we compared cells only under normal growth conditions at 37 ºC. Using fluorescent microscopy, the percentage of cells with foci was

calculated by counting approximately 300 LG expressing cells across different fields, and counting how many of these cells contained foci (Fig. 4a). We observed that, contrary to wild type NLS-LG, single mutant and double mutant NLS-LG readily form aggregates at normal growth conditions. As shown previously (Fig. 3a, 3b), NES-LG WT also forms some aggregates at 37 ºC. Although not significant, there appears to be a trend that more foci are found in cells expressing LG SM and LG DM, compared to LG WT. Within all three different LG variants there is a significant increase in the number of foci when they are localized to the cytoplasm, compared to when they are localized to the nucleus.

NLS-LG pellets differently than NES-LG

In addition to studying the aggregation propensity of NLS-LG and NES-LG by microscopy, we also investigated the detergent solubility of these proteins, by performing a sedimentation assay. We observed that under normal growth conditions nearly all of NLS-LG WT is soluble. Contrastingly, we found that at 37 ºC 29% of NES-LG WT is found in the pellet fraction (Fig. 4b). These findings are in line with our previous findings using microscopy (Fig. 2a), however, this differential behaviour must be confirmed by further biological replicates. Upon heat shock we found that both NLS-LG WT and NES-LG WT have aggregated and most of the protein was found in the pellet fraction.

NLS-LG is less abundant than NES-LG

To begin to determine why NES-LG WT forms foci at 37 ºC that are found in the pellet fraction and NLS-LG WT does not, we measured their relative abundance in the cells via immunoblotting and densitometry. What stood out to us, is that at 37ºC, NLS-LG WT is much less abundant than NES-LG WT (Fig. 4c). We next asked if NLS-LG is expressed less than NES-LG, or that NLS-LG is turned over more rapidly than NES-LG.

Cycloheximide chase

To find answers to these questions, we expressed NLS-LG WT and NES-LG WT for 48 hours in HEK293T and treated these cells with either 1 mM cycloheximide to inhibit protein synthesis, or a combination of both 1 mM cycloheximide and 10 µM MG-132 to inhibit both protein synthesis and proteasomal degradation, respectively (Lee & Goldberg, 1996). Moreover, one half of the cells was subjected to a heat shock (43 ºC) for the duration of the treatment, before collecting. Unfortunately, upon analysis we observed too high variability between samples (Fig. 5). We therefore chose a single time point to study more in depth.

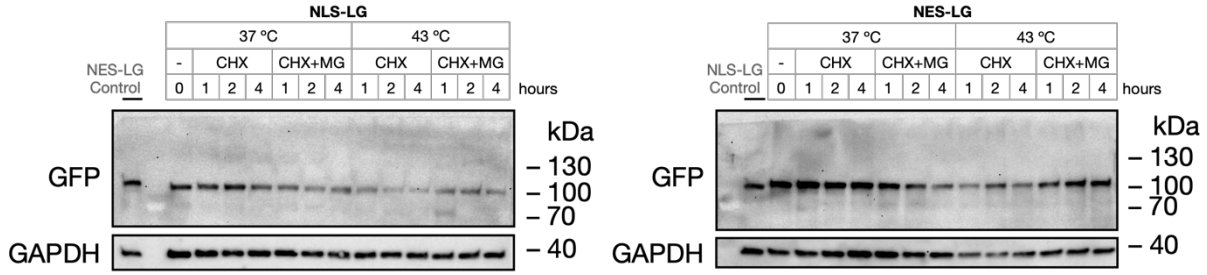

Figure 5. Immunoblot of NLS-LG WT and NES-LG WT. Lysates from transiently transfected HEK293T cells were run on 10% gels under denaturing conditions. After transfer to nitrocellulose the blotted bands were immunodetected with specific antibodies against GFP and GAPDH and subsequently visualized with peroxidase labeled secondary antibodies.

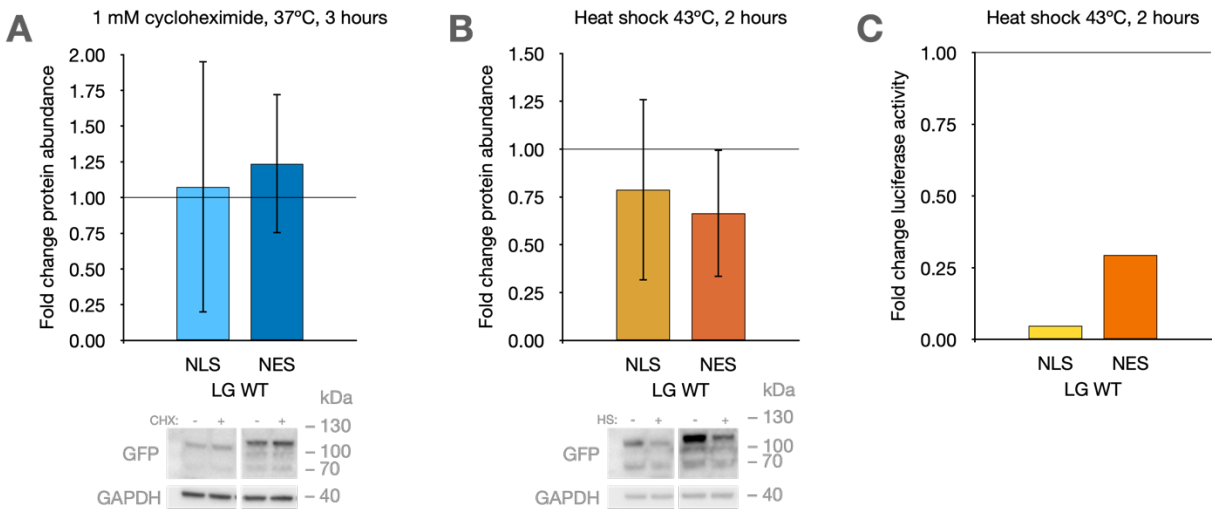

Figure 6. (A) Fold change in abundance of NLS-LG WT and NES-LG WT upon treatment with 1 mM cycloheximide for 3 h (n = 3, 4). Graphs represent mean ± standard deviation. Four images are from nonconsecutive wells on the same blot. (B) Fold change in abundance of NLS-LG WT and NES-LG WT upon 2 h of heat shock (43 ºC; n = 5, 4). Graphs represent mean ± standard deviation. Four images are nonconsecutive wells on the same blot. (C) Luciferase activity assay on lysates of cells expressing NLS-LG WT and NES-LG WT after 2 h of heat shock (43 ºC). Each lysate was sampled and measured in triplicate. Graphs represent mean fold change RLU/s over untreated cells (n = 1).

NLS-LG and NES-LG are degraded upon 2 hours of heat shock

We expressed NLS-LG WT and NES-LG WT for 48 hours in cells and treated one half of these cells with 1 mM cycloheximide to inhibit protein synthesis for 3 hours under otherwise normal growth conditions. Unfortunately, when performing immunoblot analysis, we observed a large variability between biological repeats, and could not find a significant difference between steady state degradation of NLS-LG WT and NES-LG WT (Fig. 6a).

We have previously shown that NLS-LG WT and NES-LG WT are thermolabile proteins that readily aggregate upon two hours of heat stress (43 ºC). As such, we asked how much of NLS-LG WT and NES-LG WT is degraded upon heat shock. We expressed these constructs similar to before and subjected one half of the cells to two hours of heat shock (43 ºC) before lysis. Using immunoblot analysis, we observed that both NLS-LG WT and NES-LG WT are degraded upon heat stress (Fig. 6b). Although not significant ($p = 0.23$), the data suggests that NES-LG WT is degraded marginally more than NLS-LG WT upon heat stress.

Firefly luciferase is more stable in the cytoplasm than in the nucleus

So far, we have seen that NLS-LG WT has higher detergent solubility than NES-LG WT under normal growth conditions (Fig. 4b) and that both NLS-LG WT and NES-LG WT are degraded upon heat stress (Fig. 6b). Up until this point we have only studied total protein levels. To put things into perspective, we sought to investigate what proportions of NLS-LG WT and NES-LG WT are folded correctly, both under normal growth conditions, as well as upon two hours of heat stress (43 ºC). One of the benefits of using firefly luciferase as a model protein, is that it is a bioluminescent enzyme. As such, it is very straightforward to study the 'foldedness' of NLS-LG WT and NES-LG WT, by lysing the cells, adding D-luciferin (the substrate for firefly luciferase) to the lysate, and measuring the light output.

We expressed NLS-LG WT and NES-LG WT in HEK293T for 48 h and subjected one half of the cells to 2 h heat shock (43 ºC). After lysis, we performed a bioluminescent activity assay on the different lysates. When studying the activity of NLS-LG WT and

NES-LG WT, we found that activity of both NLS-LG WT and NES-LG WT are greatly reduced upon heat stress (Fig. 6c). When comparing the two, we found that the decrease was much more pronounced in NLS-LG WT expressing cells, however, this differential behaviour must be confirmed by further biological replicates.

NLS-RG is a thermolabile protein

Next, we investigated how a different bioluminescent protein would behave, compared to firefly luciferase. Renilla luciferase is often used as a control protein when studying properties of firefly luciferase, since it is also a bioluminescent protein, but otherwise completely unrelated. In line with our previous experiments, we expressed NLS-Renilla-GFP (NLS-RG) for 48 hours in HEK293T cells and subjected one half of the cells to 2 hours of heat shock at 43 ºC. Under normal growth conditions, NLS-RG is soluble in virtually all cells (Fig. 7a) and has strong nuclear localization, as shown by its colocalization with the Hoechst staining (Fig. 7b). Upon two hours of heat shock (43 ºC) however, NLS-RG has formed nuclear foci in all cells (Fig. 7a, 7c).

NLS-RG has different localization than NLS-LG upon 2 hours of heat shock

We made the observation that, although both NLS-LG and NLS-RG form nuclear foci upon 2 hours of heat shock (43 ºC), their appearances are different. For NLS-LG we saw colocalization of foci with Hoechst-dim nuclear sub-compartments (Fig. 2c) which could indicate nucleolar localization. For NLS-RG we noticed that the foci do not intersect with Hoechst-dim nuclear sub-compartments. To verify that NLS-LG has nucleolar localization upon heat shock, and that NLS-RG does not, we repeated the previous experiment, but performed immunofluorescence against NPM1 (nucleophosmin), which is a constituent of the granular component of nucleoli. We confirmed that the NLS-RG foci that are formed upon heat shock, do not localize to nucleoli, but even more striking, we saw these foci assemble into ring-like structures that surround the nucleoli (Fig. 11).

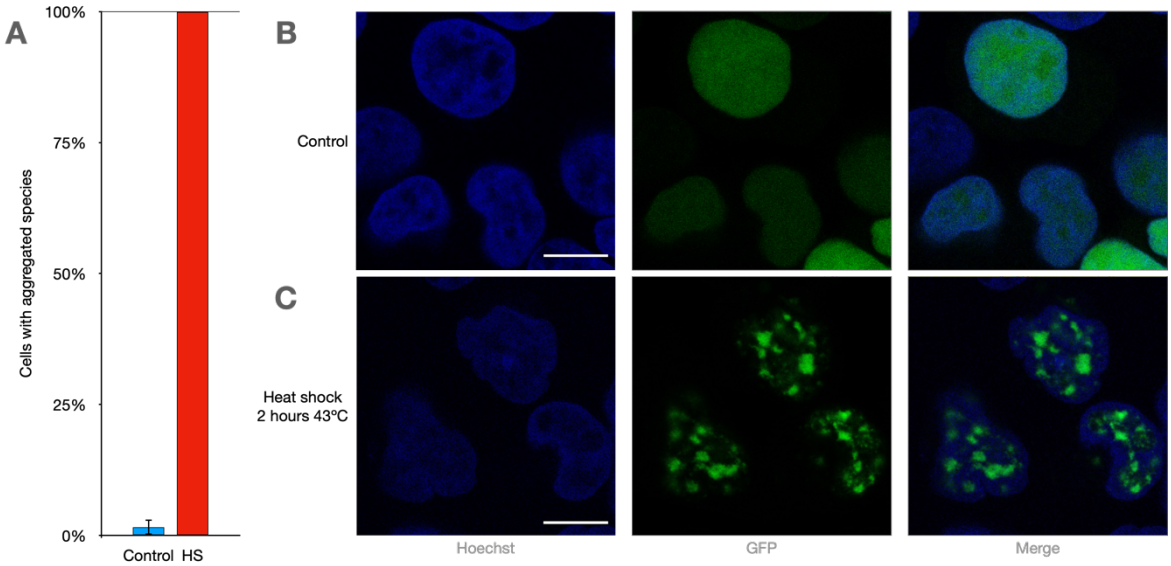

*Figure 7. NLS-RG was expressed for 48 h in HEK293T cells. One half was subjected to two hours heat shock (43 ºC) before fixating and mounting. **(A)** Percentage of cells with foci without heat shock (blue bar, n = 298) and with heat shock (red bar, n = 300). Graphs represent mean ± standard deviation. **(B-C)** Confocal laser scanning micrographs of typical NLS-RG expressing cells without heat shock (B) and with heat shock (C). Representative images shown. Scale bar denotes 10 µm.*

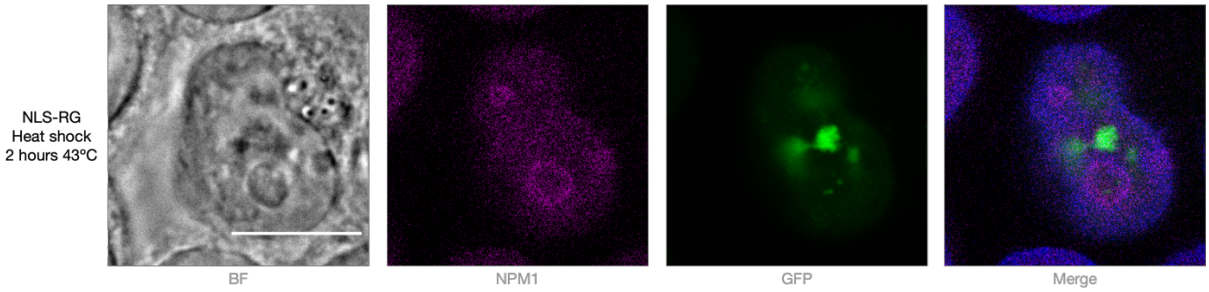

Figure 8. Confocal laser scanning micrographs of typical NLS-RG expressing cells upon 2 hours of heat shock (43 ºC). Immunocytochemistry was performed with specific antibodies to detect NPM1 (nucleophosmin) and visualized with Alexa Fluor 633 secondary antibodies. Representative images shown. Scale bar denotes 10 µm.

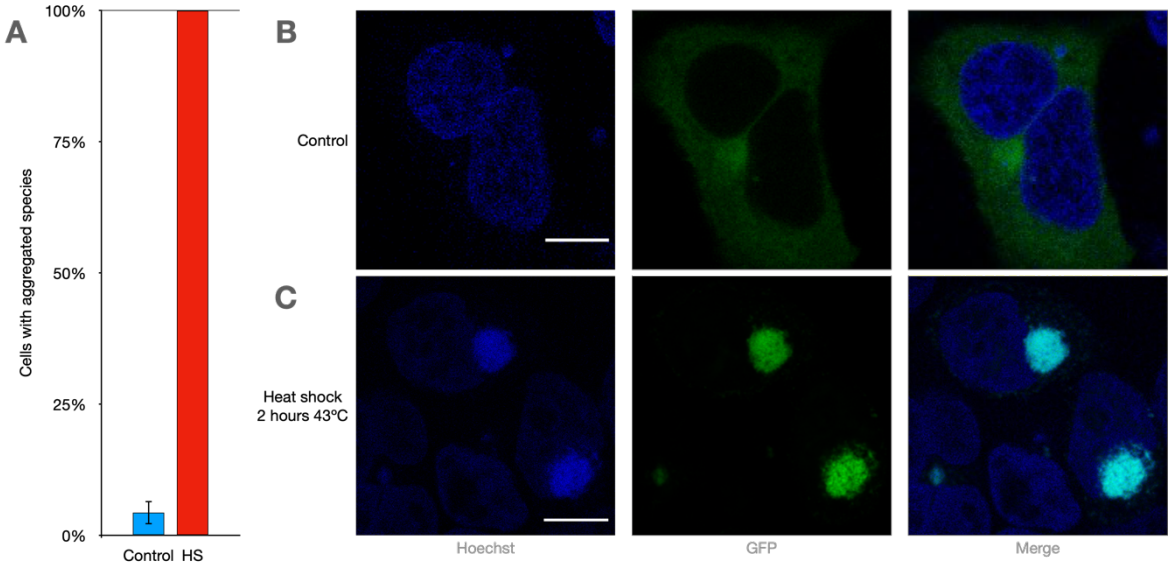

Figure 9. NES-RG was expressed for 48 h in HEK293T cells. One half was subjected to two hours heat shock (43 ºC) before fixating and mounting. (A) Percentage of cells with foci without heat shock (blue bar, n = 316) and with heat shock (red bar, n = 300). Graphs represent mean ± standard deviation. (B-C) Confocal laser scanning micrographs of typical NES-RG expressing cells without heat shock (B) and with heat shock (C). Representative images shown. Scale bar denotes 10 µm.

NES-RG localizes to the aggresome upon 2 hours of heat shock

Since we made the observation that Renilla luciferase is also a thermolabile protein in the nucleus, we next explored how stable Renilla luciferase would be when localized to the cytoplasm. We observed that at 37 ºC, in most of the expressing cells NES-RG exists as a soluble protein with strong cytoplasmic localization. We found that in 4.3% of expressing cells, NES-RG had aggregated and formed cytoplasmic foci. We also saw in higher expressing cells that NES-RG is more concentrated near the concave part of the nucleus, which is thought to be the aggresome (Fig. 9a, 9b). Upon two hours of heat shock at 43 ºC, NES-RG has aggregated in all expressing cells. In some cells these aggregates showed a granular appearance throughout the cytoplasm, but more often we saw larger protein deposits that localized to the aggresome (Fig. 9c).

Renilla luciferase is more abundant than firefly luciferase

To study whether Renilla luciferase was degraded similarly as firefly luciferase, we expressed NLS-RG and NES-RG for 48 hours and treated one half of these cells with 1 mM cycloheximide to inhibit protein synthesis for 3 hours under otherwise normal growth conditions. Immunoblot analysis showed that after 3 hours treatment with 1 mM

cycloheximide, NLS-RG and NES-RG are marginally degraded (Fig. 10a). In conclusion, the turnover rates of NLS-RG and NES-RG at 37 ºC are comparable ($p = 0.69$).

Upon two hours of heat shock (43 ºC), we observed with immunoblot analysis that NLS-RG and NES-RG are marginally degraded (Fig. 10b). Moreover, when studying steady state abundance, we found that Renilla luciferase is much more abundant than firefly luciferase at 37 ºC (Fig. 10c).

Single fluorophore control proteins do not localize as strongly as the model proteins

We next explored whether our previous findings hold true for any protein when subjected to heat shock, or that they are specific to thermolabile proteins. To this end we generated NLS-GFP and NES-GFP plasmids, as well as NLS-mScarlet and NES-mScarlet, since GFP and mScarlet are relatively thermostable proteins.

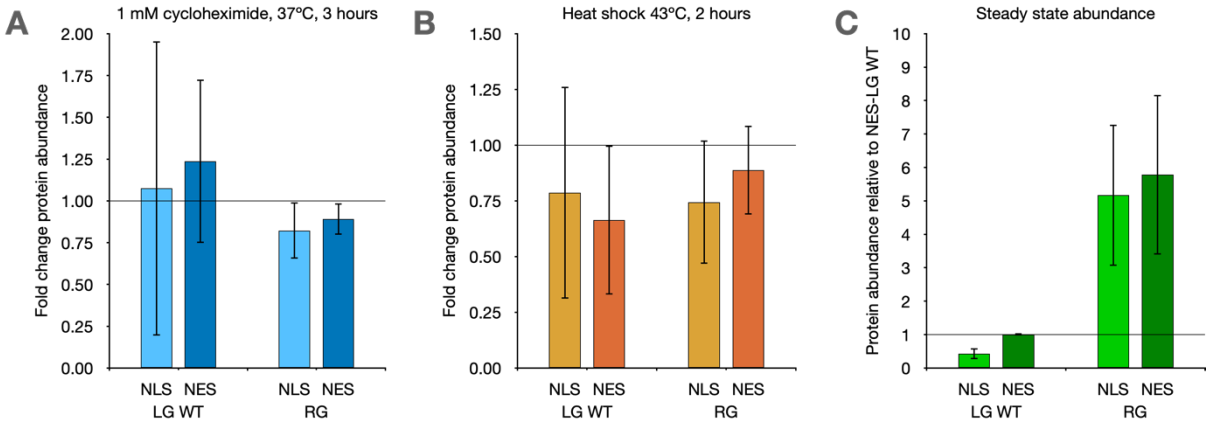

Figure 10. (A) Fold change in abundance of NLS-LG WT, NES-LG WT, NLS-RG, and NES-RG upon treatment with 1 mM cycloheximide for 3 h (n = 3, 4, 3, 3). Graphs represent mean ± standard deviation. (B) Fold change in abundance of NLS-LG WT, NES-LG WT, NLS-RG, and NES-RG upon 2 h of heat shock (43 ºC; n = 5, 4, 3, 3). Graphs represent mean ± standard deviation. (C) Relative abundance of NLS-LG WT, NES-LG WT, NLS-RG, and NES-RG under normal growth conditions after 48 h of expression (n = 6, 6, 3, 3). Graphs represent mean ± standard deviation. Immunoblot analysis was performed using specific antibodies against GFP and GAPDH and visualized using peroxidase-labeled secondary antibodies. GFP quantification was normalized to GAPDH, before standardizing to NES-LG WT.

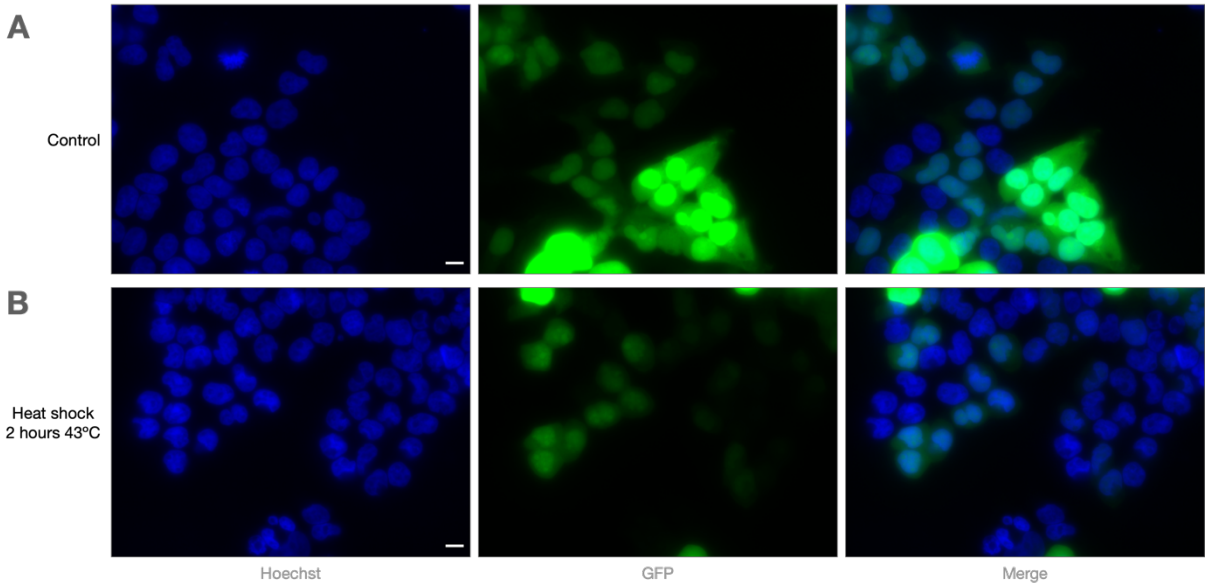

Figure 11. Widefield fluorescent microscopy, NLS-GFP was expressed 48 h in HEK293T cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

We expressed NLS-GFP, NES-GFP, NLS-mScarlet, and NES-mScarlet in HEK293T cells for 48 h, subjected one half of the cells to two hours of heat shock (43 ºC) and performed fluorescent microscopy on fixated cells. To our surprise, we saw that NLS-GFP and NLS-mScarlet leaked into the cytoplasm (Fig. 11), and that NES-GFP and NES-mScarlet leaked into the nucleus (not shown). Moreover, we saw areas where NLS-GFP and NLS-mScarlet appeared to be enriched, which colocalize with Hoechst-dim parts of the nucleus. The accumulation in these areas appeared to increase upon heat shock Fig. 12).

All in all, the single fluorophore control proteins we generated did not localize properly. We hypothesized that the proteins are too small (29.0 – 29.4 kDa) to be obstructed by nuclear pore complexes (NPCs) and could passively diffuse between compartments. To test this hypothesis, and in an effort to improve the localization of our control proteins, we increased their size by doubling the fluorophores.

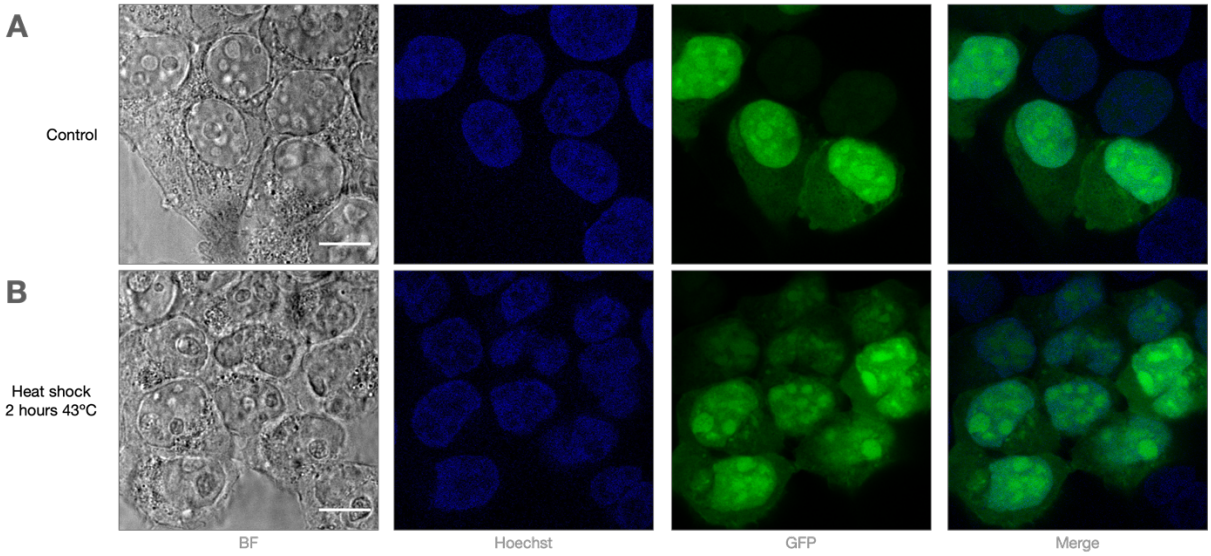

Figure 12. Confocal laser scanning micrographs of typical NLS-GFP expressing cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

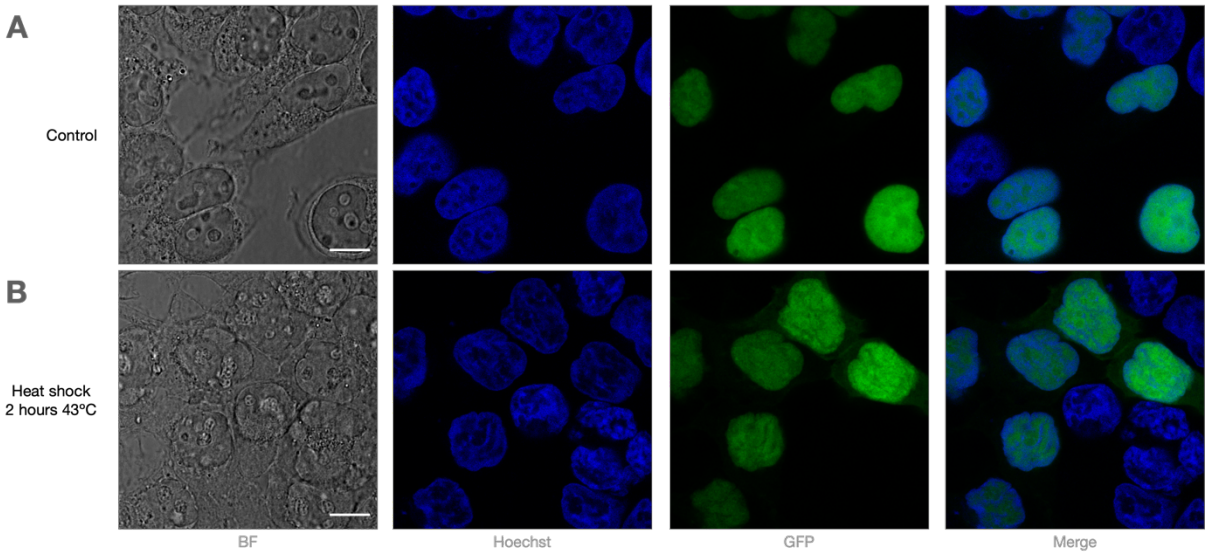

Figure 13. Confocal laser scanning micrographs of typical NLS-GFP-GFP expressing cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

NLS-GFP-GFP localizes to the nucleus and is thermostable

We generated NLS-GFP-GFP, NES-GFP-GFP, as well as NLS-mScarlet-mScarlet, NES-mScarlet-mScarlet and untargeted mScarlet-mScarlet plasmids (53.7 – 57.3 kDa). We expressed the four targeted plasmids in HEK293T cells for 48 h, subjected one half of the cells to two hours of heat shock (43 ºC) and performed fluorescent microscopy on fixated cells. We observed that NLS-GFP-GFP had proper nuclear localization, as shown by the colocalization with the Hoechst staining (Fig. 13a). Moreover, we saw that NLS-GFP-GFP remained soluble and retained nuclear localization upon two hours of heat shock (43 ºC; Fig. 13b). The same holds true for NLS-mScarlet-mScarlet (not shown).

NES-GFP-GFP localizes to the cytoplasm and is thermostable

We also studied NES-GFP-GFP and NES-mScarlet-mScarlet localization using fluorescent microscopy. We found that NES-GFP-GFP localized properly to the cytoplasm, since it does not colocalize with the Hoechst staining (Fig. 14a). Although NES-GFP-GFP was mostly soluble, we did notice some cytoplasmic foci that localized to the aggresome. However, there was no difference in the number or appearance of these foci, upon two hours of heat shock (43 ºC, Fig. 14b). In conclusion, we have shown that NLS-GFP-GFP and NES-GFP-GFP have improved localization over the single fluorophore control proteins and are thermostable.

Double fluorophore control proteins have a low turn-over rate and are thermostable

Next, we investigated at what rate NLS-GFP-GFP and NES-GFP-GFP are degraded at steady-state. We expressed NLS-GFP-GFP and NES-GFP-GFP in HEK293T cells for 48 h. One half of these cells was treated with 1 mM cycloheximide to inhibit protein synthesis for 3 hours under otherwise normal growth conditions. Immunoblot analysis showed that after 3 hours treatment with 1 mM cycloheximide, NLS-GFP-GFP and NES-GFP-GFP were not degraded (Fig. 15a).

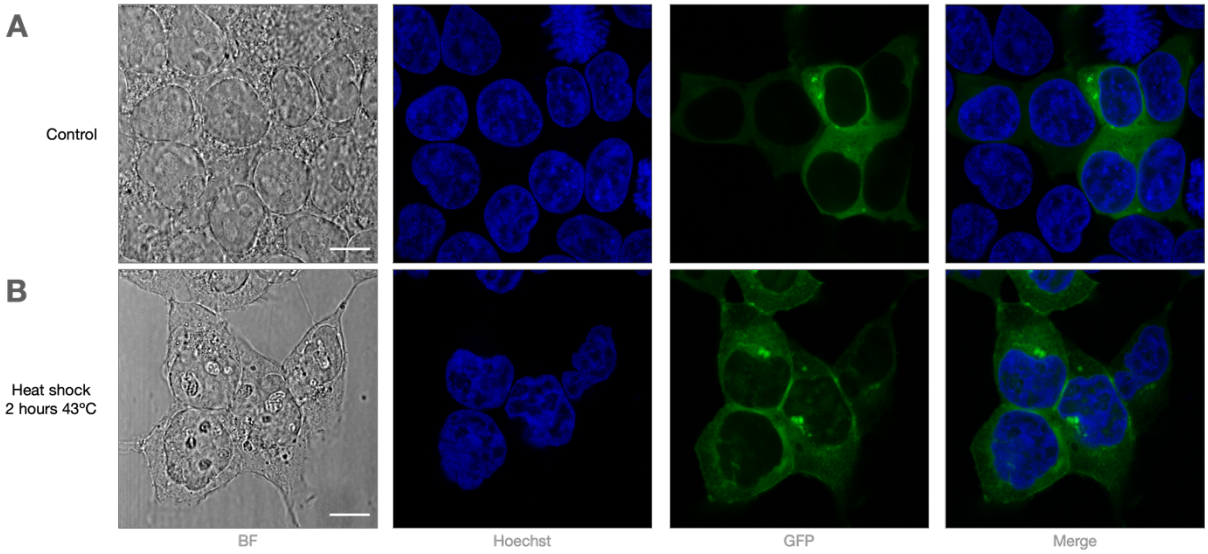

*Figure 14. Confocal laser scanning micrographs of typical NES-GFP-GFP expressing cells without heat shock **(A)** and with heat shock **(B)**. Representative images shown. Scale bar denotes 10 µm.*

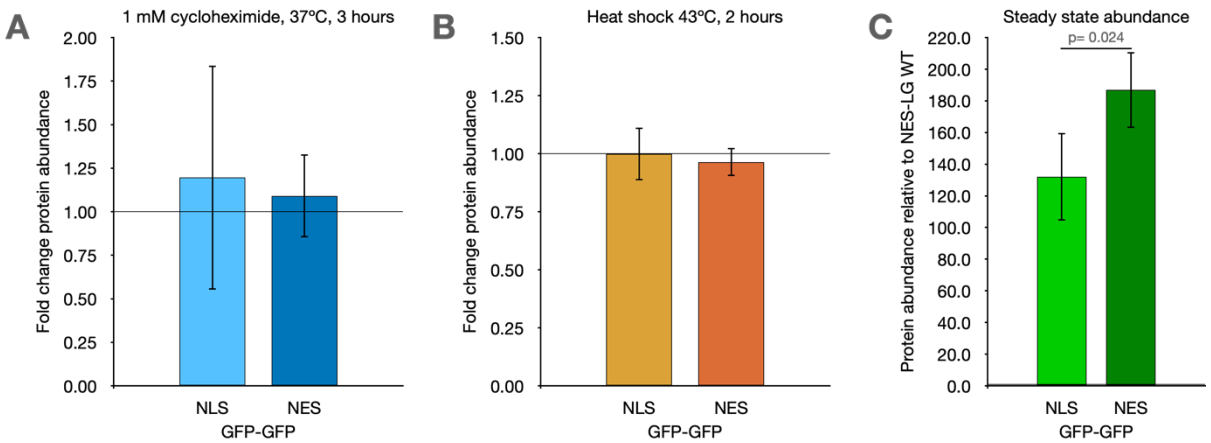

Figure 15. (A) Fold change in abundance of NLS-GFP-GFP and NES-GFP-GFP upon treatment with 1 mM cycloheximide for 3 h (n = 3, 3). Graphs represent mean ± standard deviation. (B) Fold change in abundance of NLS-GFP-GFP and NES-GFP-GFP upon 2 h of heat shock (43 ºC; n = 3, 3). Graphs represent mean ± standard deviation. (C) Relative abundance of NLS-GFP-GFP and NES-GFP-GFP after 48 h of expression (n = 3, 3). Graphs represent mean ± standard deviation. Cell lysates of NLS-GFP-GFP and NES-GFP-GFP were diluted 1:80 in 2x sample buffer before loading. Immunoblot analysis was performed using specific antibodies against GFP and GAPDH and visualized using peroxidase-labeled secondary antibodies. GFP quantification was normalized to GAPDH, before standardizing to NES-LG WT.

To study how thermostable NLS-GFP-GFP and NES-GFP-GFP are, we expressed these constructs again and this time we subjected one half of the cells to 2 hours of heat shock (43 ºC) and performed immunoblot analysis. We found that NLS-GFP-GFP and NES-GFP-GFP were not degraded upon heat shock (Fig. 15b). Taken together, we conclude that NLS-GFP-GFP and NES-GFP-GFP are thermostable proteins.

Double fluorophore proteins are highly abundant

While performing these immunoblot analyses, we ran into issues with quantification and also had noticed that the primary antibodies we used to detect GFP were consumed at a much faster rate than normal. We concluded that NLS-GFP-GFP and NES-GFP-GFP were much more abundant than other proteins of interest we had previously analyzed using immunoblotting. After analyzing series dilutions of NLS-GFP-GFP and NES-GFP-GFP cell lysates (not shown), we discovered that a 1:80 dilution was required to overcome antibody saturation issues. As a result, we calculated from these 1:80 dilutions that NLS-GFP-GFP was 132 times more abundant than NES-LG WT (the GFP protein we chose as a standard), and that NES-GFP-GFP was 187 times more abundant than NES-LG WT (Fig. 15c). These values were corrected for the 1:80 dilution and for the double GFP epitopes present on the protein of interest.

Thermostable proteins with such high abundance could be a useful tool to study, for example, the limits of nucleocytoplasmic transport when the cell is subjected to stress. However, we preferred to also include thermostable control proteins in our library that are abundant on the same order of magnitude as firefly and Renilla luciferases. Since we had learned that the size of the double fluorophore proteins (56.1 – 57.3 kDa) was sufficient to facilitate proper localization, we set out to create another set of constructs with similar size. In our modular system of constructs [signaling peptide – model protein – fluorophore], we chose to use glutathione S-transferase (GST) as a model protein. As such, we created NLS-GST-GFP, NES-GST-GFP, NLS-GST-mScarlet, NES-GST-Scarlet and untargeted GST-Scarlet following the same design philosophy.

NLS-GST-GFP localizes to the nucleus and is thermostable

To verify that NLS-GST-GFP and NLS-GST-mScarlet are thermostable, and that they localize properly to the nucleus, we expressed these constructs for 48 h in HEK293T cells, and subjected one half of the cells to two hours of heat stress (43 ºC) before fixating and performing microscopy. We observed that these proteins localized to the nucleus, as shown by the colocalization with the Hoechst staining (Fig. 16). Although NLS-GST-GFP appeared to be enriched in some areas within the nucleus, it did not form foci in either group of cells. Moreover, these enrichments did not overlap with Hoechst-dim areas, instead, NLS-GST-GFP appeared to be excluded from these areas that could be nucleolar. Since NLS-GST-GFP and NLS-GST-mScarlet did not form foci upon two hours of heat stress, we conclude that they are relatively thermostable proteins.

NLS-GST-mScarlet does not aggregate in the presence of misfolded NLS-LG WT

Next, we explored whether the GST control proteins would remain stable in the presence of a heat-denatured protein. To this end, we co-expressed NLS-LG WT and NLS-GST-Scarlet in HEK293T cells for 48 h and divided these into two groups. One group was subjected to two hours of heat stress (43 ºC), before fixating both and performing microscopy. In line with our previous findings, we saw that under normal growth conditions both NLS-LG WT and NLS-GST-Scarlet localize to the nucleus, as shown by the colocalization with the Hoechst staining, and are soluble proteins (Fig. 17a). Upon two hours of heat shock, however, we observed that NLS-LG formed nuclear foci that colocalize with Hoechst-dim nuclear areas, whereas NLS-GST-mScarlet remained soluble and localized to the nucleus (Fig. 17b). From this we conclude that NLS-GST-mScarlet is a thermostable protein that is not affected by the presence of a misfolded protein in the same compartment.

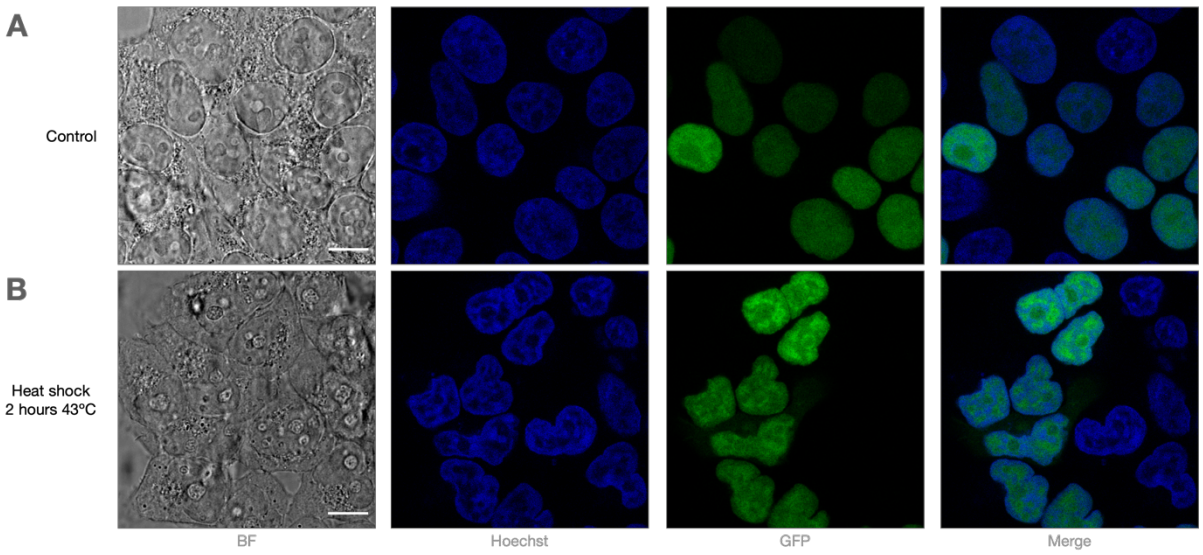

Figure 16. Confocal laser scanning micrographs of typical NLS-GST-GFP expressing cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

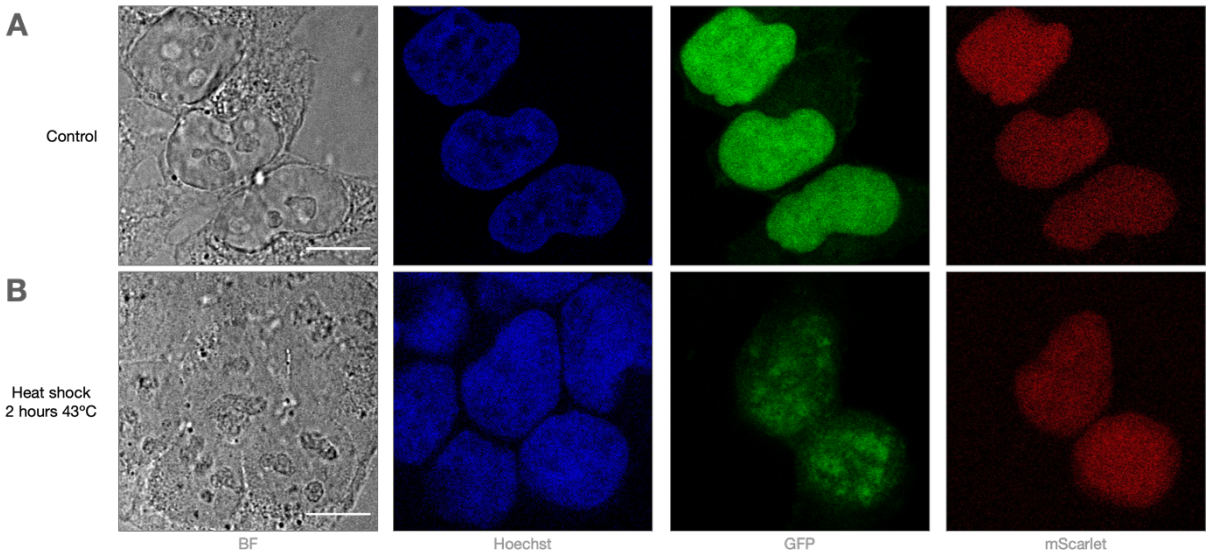

Figure 17. Confocal laser scanning micrographs of typical NLS-LG WT + NLS-GST-mScarlet expressing cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

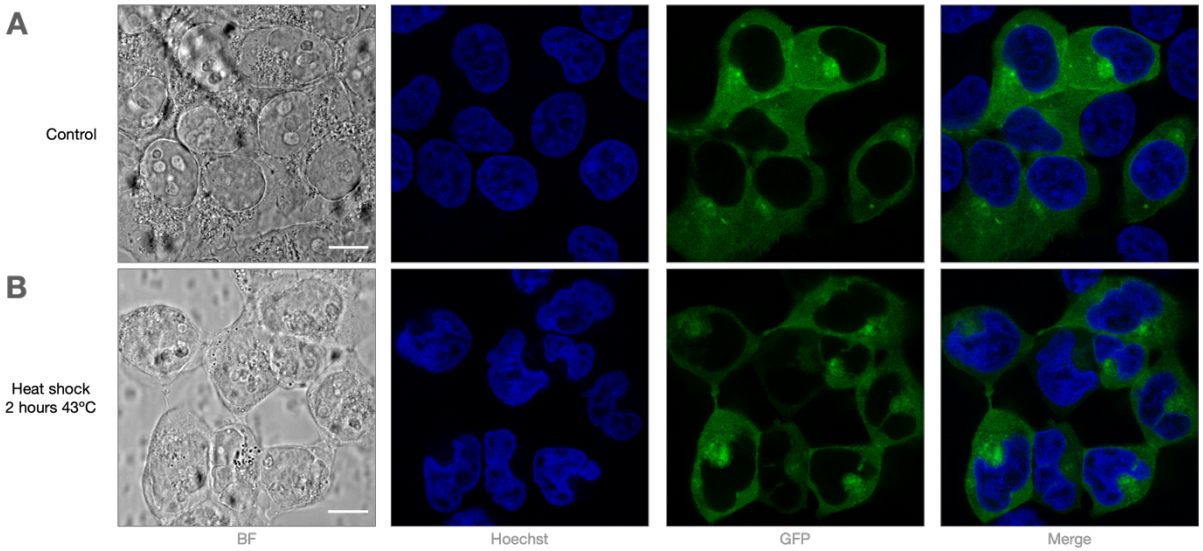

Figure 18. Confocal laser scanning micrographs of typical NES-GST-GFP expressing cells without heat shock (A) and with heat shock (B). Representative images shown. Scale bar denotes 10 µm.

NES-GST-GFP localizes to the cytoplasm and is thermostable

We expressed NES-GST-GFP in HEK293T cells for 48 h. One half of the cells was subjected to two hours of heat stress (43 ºC). Next, cells were fixated and slides were mounted. Upon performing microscopy, we observed strong cytoplasmic localization for NES-GST-GFP (Fig. 18). However, similar to NES-GFP-GFP (Fig. 14), we saw small cytoplasmic foci that localized to the concave part of the nucleus. Since these foci did not increase in size or intensity upon two hours of heat stress, we conclude that NES-GST-GFP is a thermostable protein that localizes to the cytoplasm.

Additionally, we observed by eye that both NLS-GST-GFP and NES-GST-GFP had much lower fluorescent intensity than NLS-GFP-GFP and NES-GFP-GFP, meaning that they are less abundant than the double fluorophore proteins after 48 h of expression.

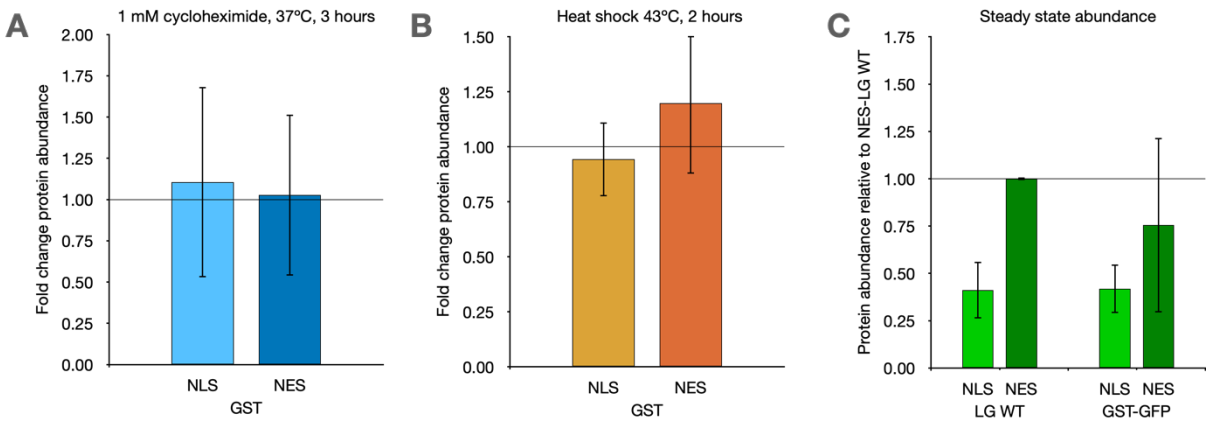

Figure 19. (A) Fold change in abundance of NLS-GST-GFP and NES-GST-GFP upon treatment with 1 mM cycloheximide for 3 h (n = 4, 4). Graphs represent mean ± standard deviation. (B) Fold change in abundance of NLS-GST-GFP and NES-GST-GFP upon 2 h of heat shock (43 ºC; n = 3, 3). Graphs represent mean ± standard deviation. (C) Relative abundance of NLS-LG WT, NES-LG WT, NLS-GST-GFP, and NES-GST-GFP under normal growth conditions after 48 h of expression (n = 6, 6, 3, 3). Graphs represent mean ± standard deviation. Immunoblot analysis was performed using specific antibodies against GFP and GAPDH and visualized using peroxidase-labeled secondary antibodies. GFP quantification was normalized to GAPDH, before standardizing to NES-LG WT.

GST fusion proteins are thermostable and their abundance is comparable to that of firefly luciferase

Having shown using microscopy that NLS-GST-GFP and NES-GST-GFP appear to be stable proteins, we next sought to verify this using biochemistry. We expressed NLS-GST-GFP and NES-GST-GFP in HEK293T cells for 48 h and treated one half of the cells with 1 mM cycloheximide to inhibit protein synthesis for 3 h. Using immunoblot analysis, we observed no decrease in NLS-GST-GFP and NES-GST-GFP abundance after three hours of inhibited protein synthesis (Fig. 19a).

After this, we studied whether NLS-GST-GFP and NES-GST-GFP are degraded upon heat stress. To this end we expressed these constructs similar to before and subjected one half of the cells to two hours of heat shock (43 ºC) before lysis. Subsequent immunoblot analysis showed that the heat stress-induced degradation is negligible for both NLS-GST-GFP and NES-GST-GFP (Fig. 19b). We therefore conclude that NLS-GST-GFP and NES-GST-GFP are thermostable proteins.

Since we noticed, when performing microscopy, that NLS-GST-GFP and NES-GST-GFP appeared much less bright than NLS-GFP-GFP and NES-GFP-GFP, we studied this using biochemistry. When comparing steady-state abundance of NLS-GST-GFP and NES-GST-GFP to other model proteins, we found that NLS-GST-GFP abundance is similar to NLS-LG WT and that NES-GST-GFP abundance is similar to NES-LG WT (Fig. 19c). From this we conclude that we succeeded in designing thermostable model proteins that are expressed comparable to firefly luciferase.

NES signaling peptide causes localization to aggresome

Looking back at the microscopy we have performed so far, we noticed that all proteins that were tagged with an NES showed some degree of aggregation, or at the very least accumulation near the concave part of the nucleus, where the aggresome is located. We asked whether this is due to having a highly expressed protein localized to the cytoplasm, but this hypothesis was dismissed since we noticed the same phenomenon for the lower expressing NES-GST-GFP protein (Fig. 18). Alternatively, we asked if the NES tag itself is causative to the accumulation of protein near the aggresome.

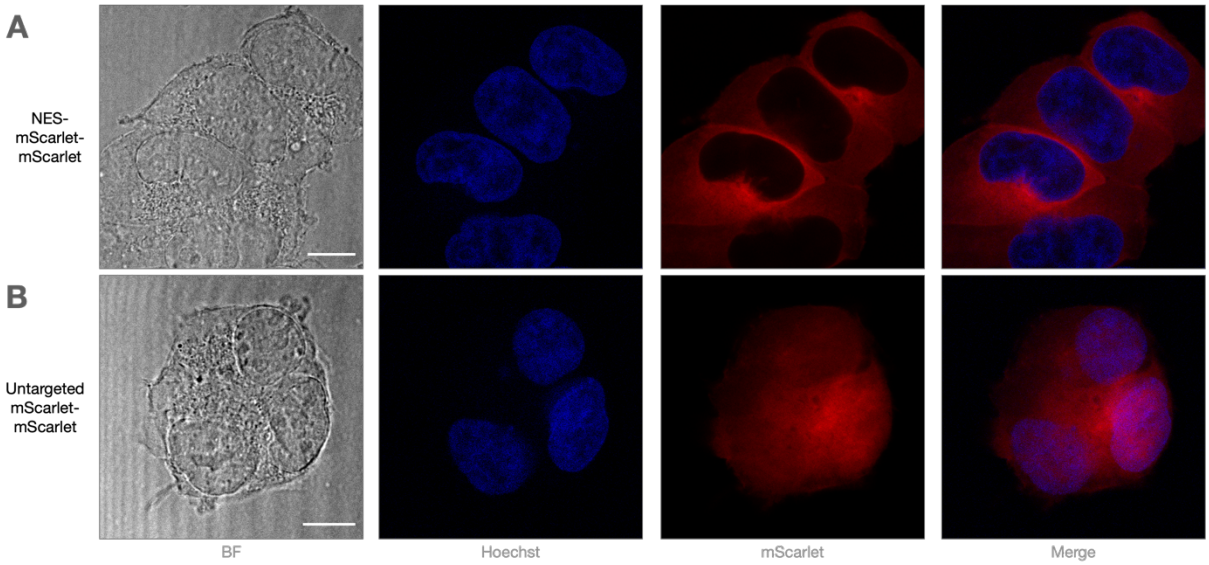

Figure 20. Confocal laser scanning micrographs of typical NES-mScarlet-mScarlet expressing cells (A) and untargeted mScarlet-mScarlet expressing cells (B) under normal growth conditions. Representative images shown. Scale bar denotes 10 µm.

We expressed untargeted GST-mScarlet, NES-GST-mScarlet, untargeted mScarlet-mScarlet, and NES-mScarlet-mScarlet in HEK293T cells for 48 h. We fixated these cells and performed microscopy. When comparing untargeted mScarlet-mScarlet to NES-mScarlet-mScarlet, we observed kidney-shaped nuclear morphology only in NES-mScarlet-mScarlet expressing cells, as indicated by the Hoechst staining. We only observed cytoplasmic accumulation near the concave part of the nucleus for NES-mScarlet-mScarlet (Fig. 20). Similar results were observed when comparing untargeted GST-mScarlet and NES-GST-mScarlet (not shown). We therefore conclude that the NES-tag does cause some sequestration to the aggresome. However, since the localization to the aggresome of NES-GST-GFP and NES-GFP-GFP does not change upon heat shock, we conclude that these NES-tagged proteins are stable, and therefore still viable controls for conducting heat shock experiments.

Context of disease-related proteins

To study how the characteristics of thermolabile proteins such as firefly luciferase and Renilla luciferase would compare to disease-related proteins, we have created NLS- and NES- targeted constructs of Htt25Q-GFP and Htt97Q-GFP.

After 48 h of expression in HEK293T cells, we observed using microscopy that untargeted 25Q-GFP is a soluble protein that is predominantly present in the cytoplasm, since there is virtually no overlap with the Hoechst staining (Fig. 21a). For NLS-tagged 25Q-GFP we saw that localized to the nucleus, since it colocalizes with the Hoechst staining, but is excluded from sub-compartments within the nucleus that could be nucleoli (Fig. 21b). When targeting 25Q-GFP to the cytoplasm, by tagging the protein with an NES signaling sequence, we see that the cytoplasmic localization has become stronger, compared to the untargeted 25Q-GFP (Fig. 21c). Additionally, in both untargeted and NES-tagged 25Q-GFP, we see some accumulation of the protein near the concave part of the nucleus, where the aggresome is located.

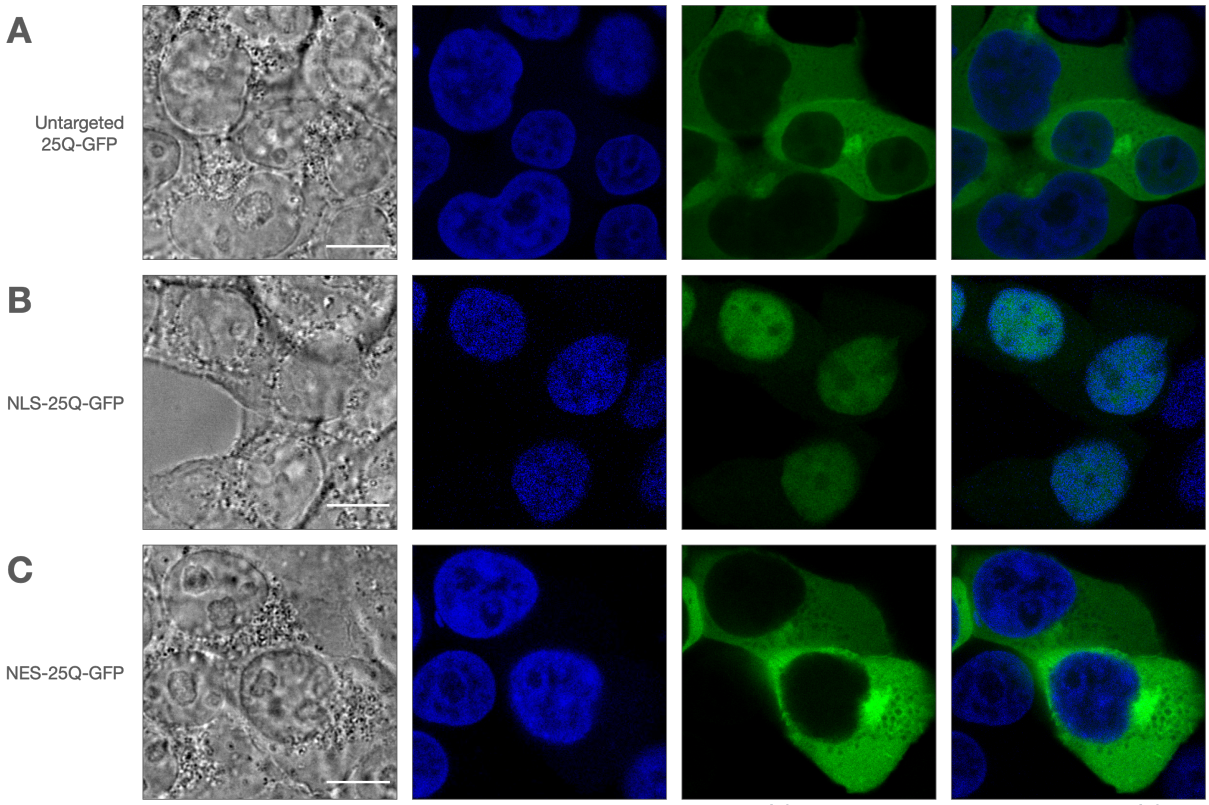

*Figure 21. Confocal laser scanning micrographs of 25Q-GFP expressing cells **(A)**, NLS-25Q-GFP expressing cells **(B)**, and NES-25Q-GFP expressing cells **(C)** under normal growth conditions. Representative images shown. Scale bar denotes 10 µm.*

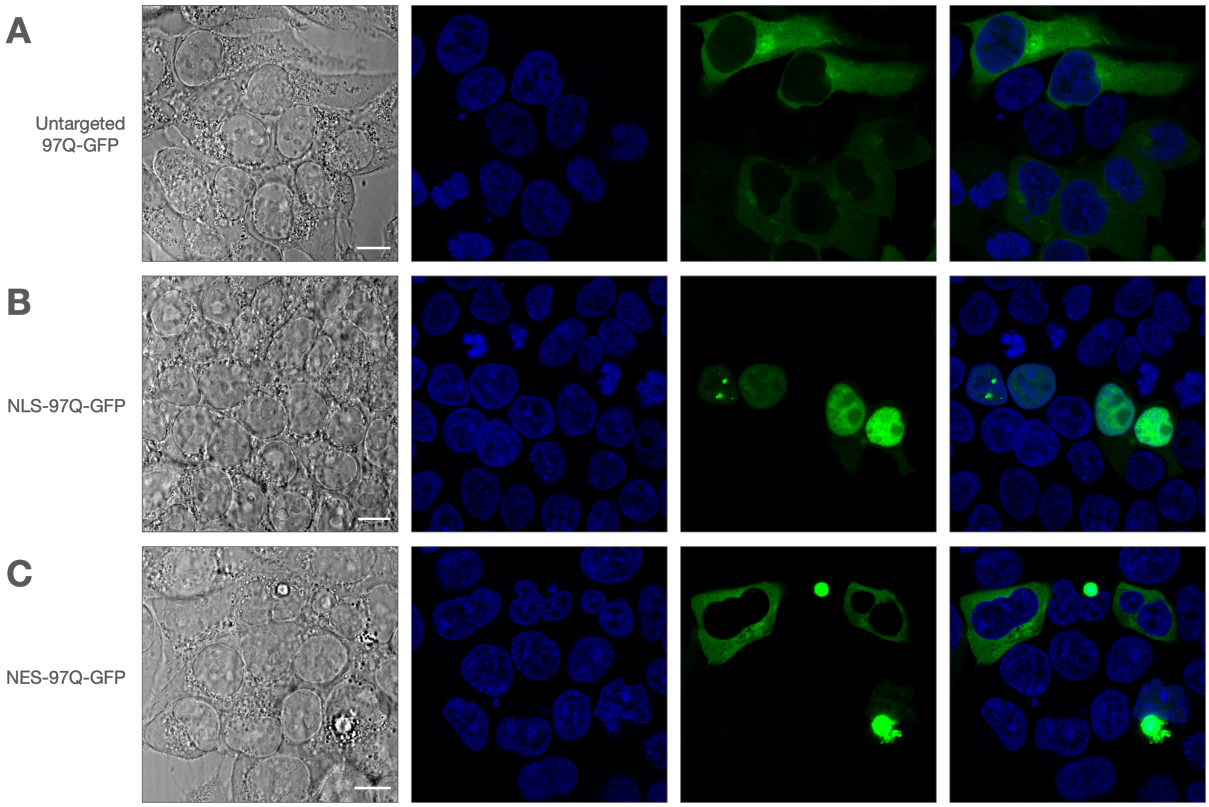

*Figure 22. Confocal laser scanning micrographs of 97Q-GFP expressing cells **(A)**, NLS-97Q-GFP expressing cells **(B)**, and NES-97Q-GFP expressing cells **(C)** under normal growth conditions. Representative images shown. Scale bar denotes 10 µm.*

After 48 h of expression in HEK293T cells, we observed using microscopy that untargeted 97Q-GFP localizes predominantly to the cytoplasm, since it does not overlap with the Hoechst staining (Fig. 22a). In a fraction of the cells, 97Q-GFP readily forms cytoplasmic aggregates. When tagged with an NES signaling sequence, we see that the cytoplasmic localization is slightly improved over untargeted 97Q-GFP (Fig. 22c). Targeting 97Q-GFP to the nucleus, by tagging the protein with an NLS signaling sequence, we see that NLS-97Q-GFP successfully localizes to the nucleus, as shown by the colocalization with the Hoechst staining (Fig. 22b). Moreover, it was seen that NLS-97Q-GFP does not aggregate as readily as untargeted 97Q-GFP and NES-97Q-GFP do (not shown). Instead, NLS-97Q-GFP appears to be soluble in most expressing cells, and is excluded from Hoechst-dim areas of the nucleus that could indicate nucleoli. In cells where NLS-97Q-GFP did form foci, these foci appeared to cluster around nucleoli (Fig. 22b).

NES-LG is more abundant than NLS-LG on a cellular level

When analyzing immunoblots of the different proteins, we noticed that all NES-tagged proteins were consistently more abundant than NLS-tagged proteins ($p = 0.001$). To verify that these differences are true on a cellular level, and not due to differences in cell populations, we turned to single-cell analysis and we chose to do this using flow cytometry. We expressed NLS-LG and NES-LG, and an empty vector control in HEK293T cells for 48 h before making single-cell suspensions and performing flow cytometry. For each sample, 50,000 events were captured and analyzed. Live single cells were selected by gating FSC and SSC, and their viability was confirmed by negative PI staining.

We observed that the live cell population expressing NLS-LG WT had a mean FITC value (a measure of GFP intensity) of 13140, compared to NES-LG with a mean FITC value of 18676. Additionally, when studying the FITC histogram we saw that the peak of NES-LG is shifted to the right, compared to NLS-LG, showing that there are more higher expressing cells for NES-LG than for NLS-LG. For NLS-LG we observed that the mean FITC value is lower and that the peak of NLS-LG expressing cells is more spread out, showing larger variability (Fig. 23a).

.

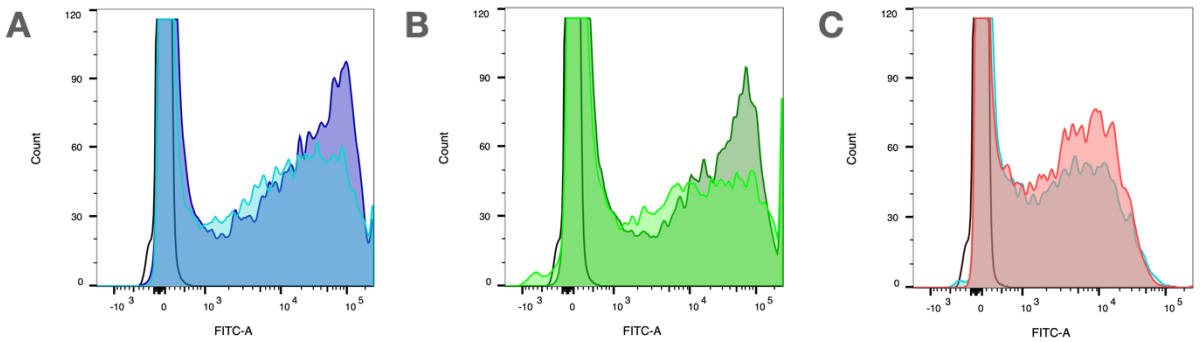

Figure 23. Histograms showing cell count vs. FITC-A intensity (=GFP intensity) on the set of live cells gated with FSC and SSC. (A) NLS-LG WT (light blue), NES-LG WT (dark blue), empty vector control (black line). (B) NLS-RG (light green), NES-RG (dark green), empty vector control (black line). (C) NLS-GST-GFP (red), NES-GST-GFP (light blue), empty vector control (black line).

NES-RG is more abundant than NLS-LG on a cellular level

Having confirmed that NES-LG is more abundant on a cellular level than NLS-LG, we next studied whether NES-RG is also more abundant than NLS-RG. To this end we repeated the experiment, expressing NLS-RG and NES-RG in HEK293T for 48 h. We calculated a mean FITC value of 16450 for live NLS-RG expressing cells and 19003 for live NES-RG expressing cells, following the same trend as when comparing NLS-LG to NES-LG expressing cells. Similarly, we found that the FITC peak of NES-RG is shifted to the right, compared to NLS-RG, and that the FITC peak of NLS-RG is more spread out, showing larger variability within the NLS-RG expressing population (Fig. 23b).

NLS-GST-GFP and NES-GST-GFP are expressed equally

To study if the same trend would hold true for a more stable control protein, we next compared NLS-GST-GFP expressing cells to NES-GST-GFP expressing cells. Interestingly, we found that the FITC peaks of NLS-GST-GFP and NES-GST-GFP expressing cells overlap, showing that on a cellular level, both are expressed equally (Fig. 23c). We did observe that the peak was higher for NLS-GST-GFP expressing cells than for NES-GST-GFP expressing cells, suggesting that NLS-GST-GFP was expressed in more cells than NES-GST-GFP.

Discussion

In the present study we have developed a comprehensive library of plasmids that express model proteins with various thermodynamic properties that localize to different cellular compartments, and we have performed an initial characterization of these proteins (summarized in Table 1).

We have shown that NLS-LG is a thermolabile protein that localizes to the nucleolus upon heat stress (Fig. 2c), which is in line with previous studies (Frottin et al., 2019; Nollen et al., 2001). The thermolabile properties of NLS-LG were further exacerbated by point mutants LG SM and LG DM. Moreover, we found that LG is more prone to form aggregates at 37 ºC when localized to the cytoplasm (Fig. 3b). Additionally, we have shown that the nucleolar localization of these proteins upon heat stress is specific to NLS-LG, since both NLS-GST-GFP and NLS-GFP-GFP remain soluble in the nucleus.

In the past it has been shown that cytoplasmic proteins can be transported to the nucleus for degradation in a Hsp70- and Hsp40-dependent pathway (Park et al., 2013). We have now shown that this pathway is inhibited when luciferase is targeted to the cytoplasm with an NES. Instead, NES-LG forms small cytoplasmic foci upon misfolding, that concentrate at the aggresome (Fig. 3c).

While studying different control protein candidates, we found that Renilla luciferase is not a thermostable protein, as it forms aggregates upon heat stress (Fig. 7c, 9c), even though it is often used as a control protein in firefly luciferase studies, or as a reporter for stress protein transcription (Gupta et al., 2011). Thus, further work will be needed to determine the 'foldedness' of these constructs during stress, by analyzing bioluminescence. When comparing NLS-LG to NLS-RG we found a striking difference in the foci that are formed upon heat stress. For NLS-LG it has been shown that, upon misfolding, it can form small nuclear foci with a granular appearance, but the majority of the protein localizes to the nucleolus (Fig. 2c). For NLS-RG we have now shown that it does not

Table 1. Summary of preliminary characterization of various proteins.

	Localization (microscopy)	Stable (biochemistry)	Abundance (biochemistry)	Aggregation (microscopy)
NLS-LG	37ºC: nuclear 43ºC: nucleolar	enzymatic activity lost upon 2 h heat shock	.4x NES-LG	nucleolar
NES-LG	cytoplasm/ aggresome	more stable than NLS-LG	1x NES-LG	aggresome/ granular foci
NLS-RG	37ºC: nuclear 43ºC: peri-nucleolar	less degraded than NLS-LG	5x NES-LG	peri-nucleolar foci
NES-RG	cytoplasm/ aggresome	less degraded than NES-LG	7x NES-LG	aggresome
NLS-GST-GFP	nuclear diffuse	not degraded	.4x NES-LG	diffuse
NES-GST-GFP	cytoplasm/ aggresome	not degraded	.8x NES-LG	aggresome (minor)
NLS-GFP-GFP	nuclear diffuse	not degraded	130x NES-LG	diffuse
NES-GFP-GFP	cytoplasm/ aggresome	not degraded	180x NES-LG	aggresome (minor)

localize to the nucleolus upon heat stress, but that it forms large peri-nucleolar foci. Similarly, we found that the disease-related aggregation-prone model protein 97Q-GFP forms the same peri-nucleolar foci when targeted to the nucleus with an NLS signaling sequence. When studying nucleolar morphology in the brightfield and Hoechst micrographs, it appears that the nucleoli are well-defined structures still and have not dissolved in these cells.

From the current data, it appears that multiple pathways exist that handle misfolded within the nucleus, and that these pathways are substrate-specific. If they are distinct pathways, what features of these proteins make NLS-LG qualify for nucleolar PQC and conversely make NLS-RG and NLS-97Q-GFP qualify for peri-nucleolar PQC? It would be interesting to study whether these two pathways can coexist in one cell, or that one pathway excludes the other. Since we have developed our plasmid library to have most proteins tagged with either GFP or mScarlet, the tools are now available. It could be of value to study how e.g. NLS-LS and NLS-RG would behave when expressed together.

Moreover, it could be interesting to study which PQC factors participate in these pathways. Although HSPA6 is a heat-inducible Hsp70 chaperone, it was not found to play a role in the refolding of luciferase when recovering from heat stress (Hageman et al., 2011). It is known that upon heat stress, Hsp70 chaperones are actively transported from the cytoplasm into the nucleus, using a specialized transport protein, Hikeshi (Kose et al., 2012). It could be investigated using knockdown of Hikeshi if either one of these pathways is dependent on cytoplasmic or nuclear Hsp70s.

Additionally, it remains to be studied what would happen to these model proteins in a different time scale, since it could also be possible that it simply takes longer for certain substrates to be taken up by the nucleolus. New live cell imaging techniques could be an excellent tool to investigate the progression of these inclusion bodies beyond the 2-hour time point we have chosen, as well as during recovery from stress.

Taken together, the data suggest that the nucleus has a higher capacity for coping with misfolded proteins. We have seen that NLS-LG does not form foci at normal growth temperatures, whereas NES-LG already forms foci in 4.6% of cells under the same conditions. We have also seen that at 37 °C, virtually all of NLS-LG is soluble, whereas 29% of NES-LG is detergent insoluble. These findings could be explained as a consequence of the lower abundance of NLS-LG, compared to NES-LG, and that the higher abundance of NES-LG exceeds the PQC capacity in the cytoplasm, leading to accumulation at the aggresome.

Unfortunately, we could not explain the difference in abundance as a consequence of faster degradation of NLS-LG than NES-LG, due to large variability between experiments when treating cells with cycloheximide. In fact, in 3 out of 5 biological repeats, more NLS-LG WT was measured in cells treated with cycloheximide for 3 h than in untreated cells. Since these cells were transfected concurrently and with the same transfection mixture, it cannot be explained as consequence of different transfection efficiency between samples. Further experimental optimization will be necessary to confirm the efficiency of our cycloheximide treatment in the inhibition of protein synthesis. Alternatively, the steady-state degradation could be studied by cloning these constructs into Tet-On or Tet-Off vectors, for doxycycline-inducible gene expression or silencing, respectively. In short, such a system allows for selective gene modulation, without interfering with other cellular processes (T. Das et al., 2016).

The finding that LG was more abundant when targeted to the cytoplasm than when targeted to the nucleus, was common to all proteins we have studied using biochemistry.

We have confirmed these findings at single-cell level for NLS-LG and NES-LG, as well as for NLS-RG and NES-RG, using flow cytometry. Interestingly, we could not observe this trend when comparing NLS-GST-GFP expressing cells to NES-GST-GFP expressing cells using the same technique. This could suggest that the difference in abundance is in fact due to increased degradation of a thermolabile protein inside the nucleus, compared to the cytoplasm.

Moreover, we have found that all NES-tagged proteins exhibit some accumulation at the aggresome. We could not correlate this accumulation with higher abundance, since we had seen this for the lowest expressed NES-GST-GFP as well as for the highest expressed NES-GFP-GFP proteins (Fig. 14a, 18a). Although localization to the aggresome is a hallmark of exceeded PQC capacity in the cytoplasm (Johnston et al., 1998), we concluded instead that this was in fact due to the NES-tag, since this accumulation was not seen for the same control proteins when they were not targeted to a specific compartment (Fig. 20). Alternatively, it could be that these proteins must be transported to the nucleus for efficient degradation, and that the presence of an NES tag inhibits this PQC pathway.

Contrary to our previous conclusion, we observed when studying disease-related aggregation-prone model proteins, that untargeted 25Q-GFP and untargeted 97Q-GFP do accumulate at the aggresome, and that 97Q-GFP can form very dense aggregates at this location. These findings can be explained by the fact that the first 17 amino acids of Htt already constitute an NES (Zheng et al., 2013). When tagged with an NLS signaling sequence, both NLS-25Q-GFP and NES-97Q-GFP do localize to the nucleus, showing that the SV40 NLS is a stronger driver of localization than the intrinsic localization signal present in Htt. Moreover, we have observed that targeting 97Q-GFP to the nucleus slightly reduces the propensity to form aggregates. This is in line with the previous finding that artificial aggregation-prone proteins are less toxic when targeted to the nucleus, compared to the cytoplasm (Woerner et al., 2016).

All in all, this would suggest that nuclear PQC is more capable in handling and degrading misfolding proteins than cytoplasmic PQC, especially if multiple PQC pathways exist within the nucleus, and even more so if cytoplasmic PQC relies in part on nuclear PQC. It could very well be that at the crossroads of these PQC pathways, lies disease.

Bibliography

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223–230. https://doi.org/10.1126/science.181.4096.223

Bindels, D. S., Haarbosch, L., van Weeren, L., Postma, M., Wiese, K. E., Mastop, M., Aumonier, S., Gotthard, G., Royant, A., Hink, M. A., & Gadella, T. W. J. (2017). mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature Methods*, *14*(1), 53–56. https://doi.org/10.1038/nmeth.4074

Chen-Plotkin, A. S., Lee, V. M. Y., & Trojanowski, J. Q. (2010). TAR DNA-binding protein 43 in neurodegenerative disease. *Nature Reviews Neurology*, *6*(4), 211–220. https://doi.org/10.1038/nrneurol.2010.18

D'Angelo, M. A., Raices, M., Panowski, S. H., & Hetzer, M. W. (2009). Age-Dependent Deterioration of Nuclear Pore Complexes Causes a Loss of Nuclear Integrity in Postmitotic Cells. *Cell*, *136*(2), 284–295. https://doi.org/10.1016/j.cell.2008.11.037

Dunker, A. K., Silman, I., Uversky, V. N., & Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, *18*(6), 756–764. https://doi.org/10.1016/j.sbi.2008.10.002

Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Körner, R., Schlichthaerle, T., Cox, J., Jungmann, R., Hartl, F. U., & Hipp, M. S. (2019). The nucleolus functions as a phase-separated protein quality control compartment - Supplementary Materials. *Science*, *365*(6451), 342–347. https://doi.org/10.1126/science.aaw9157

Frydman, J. (2001). Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annual Review of Biochemistry*, *70*(1), 603–647. https://doi.org/10.1146/annurev.biochem.70.1.603

Gupta, R., Kasturi, P., Bracher, A., Loew, C., Zheng, M., Villella, A., Garza, D., Hartl, F. U., & Raychaudhuri, S. (2011). Firefly luciferase mutants as sensors of proteome stress. *Nature Methods*, *8*(10), 879–884. https://doi.org/10.1038/nmeth.1697

Hageman, J., Van Waarde, M. A. W. H., Zylicz, A., Walerych, D., & Kampinga, H. H. (2011). The diverse members of the mammalian HSP70 machine show distinct chaperone-like activities. *Biochemical Journal*, *435*(1), 127–142. https://doi.org/10.1042/BJ20101247

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324–332. https://doi.org/10.1038/nature10317

Hutten, S., & Dormann, D. (2019). Nucleocytoplasmic transport defects in neurodegeneration — Cause or consequence? *Seminars in Cell and Developmental Biology*, *May*, 1–12. https://doi.org/10.1016/j.semcdb.2019.05.020

Järvelin, A. I., Noerenberg, M., Davis, I., & Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Communication and Signaling*, *14*(1). https://doi.org/10.1186/s12964-016-0132-3

Johnston, J. A., Ward, C. L., & Kopito, R. R. (1998). Aggresomes: A cellular response to misfolded proteins. *Journal of Cell Biology*, *143*(7), 1883–1898. https://doi.org/10.1083/jcb.143.7.1883

Korolchuk, V. I., Menzies, F. M., & Rubinsztein, D. C. (2010). Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. *FEBS Letters*, *584*(7), 1393–1398. https://doi.org/10.1016/j.febslet.2009.12.047

Kose, S., Furuta, M., & Imamoto, N. (2012). Hikeshi, a nuclear import carrier for Hsp70s, protects cells from heat shock-induced nuclear damage. *Cell*, *149*(3), 578–589. https://doi.org/10.1016/j.cell.2012.02.058

Latonen, L. (2019). Phase-to-phase with nucleoli - Stress responses, protein aggregation and novel roles of RNA. *Frontiers in Cellular Neuroscience*, *13*(April), 1–10. https://doi.org/10.3389/fncel.2019.00151

Lee, D. H., & Goldberg, A. L. (1996). Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *271*(44), 27280–27284. https://doi.org/10.1074/jbc.271.44.27280

MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., Kirby, S., … Harper, P. S. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's

disease chromosomes. *Cell*, *72*(6), 971–983. https://doi.org/10.1016/0092-8674(93)90585-E
Mackenzie, I. R. A., Rademakers, R., & Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet Neurology*, *9*(10), 995–1007. https://doi.org/10.1016/S1474-4422(10)70195-2
Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268
Nollen, E. A. A., Salomons, F. A., Brunsting, J. F., Van Der Want, J. J. L., Sibon, O. C. M., & Kampinga, H. H. (2001). Dynamic changes in the localization of thermally unfolded nuclear proteins associated with chaperone-dependent protection. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12038–12043. https://doi.org/10.1073/pnas.201112398
Park, S. H., Kukushkin, Y., Gupta, R., Chen, T., Konagai, A., Hipp, M. S., Hayer-Hartl, M., & Hartl, F. U. (2013). PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell*, *154*(1), 134–145. https://doi.org/10.1016/j.cell.2013.06.003
Preissler, S., & Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. *Trends in Biochemical Sciences*, *37*(7), 274–283. https://doi.org/10.1016/j.tibs.2012.03.002
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019
Shibata, Y., & Morimoto, R. I. (2014). How the nucleus copes with proteotoxic stress. *Current Biology*, *24*(10), R463–R474. https://doi.org/10.1016/j.cub.2014.03.033
Spector, D. L. (2001). Nuclear domains. *Journal of Cell Science*, *114*(16), 2891–2893. https://doi.org/10.1242/jcs.114.16.2891
T. Das, A., Tenenbaum, L., & Berkhout, B. (2016). Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy*, *16*(3), 156–167. https://doi.org/10.2174/1566523216666160524144041
Thibault, G., & Ng, D. T. W. (2012). The Endoplasmic Reticulum-Associated Degradation Pathways of Budding Yeast. *Cold Spring Harbor Perspectives in Biology*, *4*(a013193). https://doi.org/10.1101/cshperspect.a013193
Tsuchiya, H., Ohtake, F., Arai, N., Kaiho, A., Yasuda, S., Tanaka, K., & Saeki, Y. (2017). In Vivo Ubiquitin Linkage-type Analysis Reveals that the Cdc48-Rad23/Dsk2 Axis Contributes to K48-Linked Chain Specificity of the Proteasome. *Molecular Cell*, *66*(4), 488–502. https://doi.org/10.1016/j.molcel.2017.04.024
Wang, J., Lee, J., Liem, D., & Ping, P. (2017). HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum. *Gene*, *618*, 14–23. https://doi.org/10.1016/j.gene.2017.03.005
Wang, K., Redeker, V., Madiona, K., Melki, R., & Kabani, M. (2015). The 26S proteasome degrades the soluble but not the fibrillar form of the yeast prion Ure2p in vitro. *PLoS ONE*, *10*(6), 11–18. https://doi.org/10.1371/journal.pone.0131789
Woerner, A. C., Frottin, F., Hornburg, D., Feng, L. R., Meissner, F., Patra, M., Tatzelt, J., Mann, M., Winklhofer, K. F., Hartl, F. U., & Hipp, M. S. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science*, *351*(6269), 173–176. https://doi.org/10.1126/science.aad2033
Zheng, Z., Li, A., Holmes, B. B., Marasa, J. C., & Diamond, M. I. (2013). An N-terminal nuclear export signal regulates trafficking and aggregation of huntingtin (Htt) protein exon 1. *Journal of Biological Chemistry*, *288*(9), 6063–6071. https://doi.org/10.1074/jbc.M112.413575

Supplementary Information

Table S1. Plasmid library

All plasmids below were cloned in the pcDNA3.1(+)/myc-His A (Invitrogen) backbone.

ID	Name	Source or cloning method
1	NLS-LG (WT)	Hartl lab (6239)
2	NLS-LG (SM)	Hartl lab (6240)
3	NLS-LG (DM)	Hartl lab (6241)
4	NES-LG (WT)	Hartl lab (6236)
5	NES-LG (SM)	Hartl lab (6237)
6	NES-LG (DM)	Hartl lab (6238)
7	NLS-LS (WT)	Hartl lab (6988)
8	NLS-LS (SM)	Vector 2 and insert 7, digested with XbaI and BamHI-HF
9	NLS-LS (DM)	Vector 3 and insert 7, digested with XbaI and BamHI-HF
10	NES-LS (WT)	Vector 4 and insert 7, digested with XbaI and BamHI-HF
11	NES-LS (SM)	Vector 5 and insert 7, digested with XbaI and BamHI-HF
12	NES-LS (DM)	Vector 6 and insert 7, digested with XbaI and BamHI-HF
13	RG	Vector 2 and insert from PCR with plasmid Rluc (5256, Hartl lab) as template and primers 8 and 9, digested with KpnI and BamHI-HF
14	NLS-RG	Vector 23 and insert 13, digested with XhoI and XbaI
15	NES-RG	Vector 24 and insert 13, digested with XhoI and XbaI
16	RS	Vector 7 and insert from PCR with plasmid Rluc (5256, Hartl lab) as template and primers 8 and 9, digested with KpnI and BamHI-HF
17	NLS-RS	Vector 23 and insert 16, digested with XhoI and XbaI
18	NES-RS	Vector 24 and insert 16, digested with XhoI and XbaI
19	GFP	Vector 2 and insert from PCR with plasmid 2 as template and primers 10 and 11, digested with KpnI and XbaI
20	NLS-G	Vector 23 and insert 19, digested with XhoI and XbaI
21	NES-G	Vector 24 and insert 19, digested with XhoI and XbaI
22	Scarlet	Vector 2 and insert from PCR with plasmid 7 as template and primers 10 and 12, digested with KpnI and XbaI
23	NLS-S	Vector 22 and insert from hybridized oligos 4 and 5, digested with KpnI and XhoI
24	NES-S	Vector 22 and insert from hybridized oligos 6 and 7, digested with KpnI and XhoI
25	25Q-GFP	Hartl lab (5113)
26	NLS-25Q-GFP	Vector 25 and insert from hybridized oligos 4 and 5, digested with KpnI and XhoI
27	NES-25Q-GFP	Vector 25 and insert from hybridized oligos 6 and 7, digested with KpnI and XhoI
28	97Q-GFP	Hartl lab (5114)
29	NLS-97Q-GFP	Vector 26 and insert 28, digested with XhoI and PmeI
30	NES-97Q-GFP	Vector 27 and insert 28, digested with XhoI and PmeI

ID	Name	Source or cloning method
31..35		
36	NLS-GST-GFP	Vector 14 and insert from PCR with plasmid pGEX-5X-2 (GE Healthcare Life Sciences) as template and primers 13 and 14, digested with XhoI and BamHI-HF
37	NES-GST-GFP	Vector 15 and insert from PCR with plasmid pGEX-5X-2 (GE Healthcare Life Sciences) as template and primers 13 and 14, digested with XhoI and BamHI-HF
38	NLS-GST-Scarlet	Vector 17 and insert from PCR with plasmid pGEX-5X-2 (GE Healthcare Life Sciences) as template and primers 13 and 14, digested with XhoI and BamHI-HF
39	NES-GST-Scarlet	Vector 18 and insert from PCR with plasmid pGEX-5X-2 (GE Healthcare Life Sciences) as template and primers 13 and 14, digested with XhoI and BamHI-HF
40	NLS-GFP-GFP	Vector 14 and insert from PCR with plasmid 19 as template and primers 15 and 16, digested with XhoI and BamHI-HF
41	NES-GFP-GFP	Vector 15 and insert from PCR with plasmid 19 as template and primers 15 and 16, digested with XhoI and BamHI-HF
42	NLS-Scarlet-Scarlet	Vector 17 and insert from PCR with plasmid 22 as template and primers 15 and 16, digested with XhoI and BamHI-HF
43	NES-Scarlet-Scarlet	Vector 18 and insert from PCR with plasmid 22 as template and primers 15 and 16, digested with XhoI and BamHI-HF
44	Scarlet-Scarlet	Vector 16 and insert from PCR with plasmid 22 as template and primers 15 and 16, digested with XhoI and BamHI-HF
45	GST-Scarlet	Vector 16 and insert from PCR with plasmid pGEX-5X-2 (GE Healthcare Life Sciences) as template and primers 13 and 14, digested with XhoI and BamHI-HF

Table S2. Oligonucleotides used for cloning and sequencing

ID	Name	Sequence
1	CMV-fwd	CGCAAATGGGCGGTAGGCGTG
2	BGH-rev	TAGAAGGCACAGTCGAGG
3	EGFP-N-rev	CGTCGCCGTCCAGCTCGACCAG
4	5 KpnI NLS	CATGTCCTACCCATACGATGTTCCAGATTACGCTTTACCAAAA AAGAAGAGAAAGGTAGGC
5	3 NLS XhoI	TCGAGCCTACCTTTCTCTTCTTTTTTGGTAAAGCGTAATCTGGA ACATCGTATGGGTAGGACATGGTAC
6	5 KpnI NES	CATGTCCTACCCATACGATGTTCCAGATTACGCTTTATTGGAA CTGCTGGAAGATCTGACCCTGGGC
7	3 NES XhoI	TCGAGCCCAGGGTCAGATCTTCCAGCAGTTCCAATAAAGCGT AATCTGGAACATCGTATGGGTAGGACATGGTAC
8	5 KpnI Rluc	aaaaGGTACCGGGCCCCCCCTCGAGGTCGATGGCTTCCAAGG TGTAC
9	3 Rluc BHI	aaaaGGATCCCGGGTTTCTGCTCGTTCTTCAGC
10	5 KpnI GFP	aaaaGGTACCGGGCCCCCCCTCGAGGTCGATGGTGAGCAAG GGCG
11	3 GFP XbaI	aaaaTCTAGATTACTTGTACAGC
12	3 Scar XbaI	aaaaTCTAGACTACTTGTACAGC
13	5 XhoI GST	aaaaCTCGAGGTCGATGTCCCCTATACTAGG
14	3 GST BHI	aaaaGGATCCCGGGTTTTTTTGGAGGATGGTCGC
15	5 XhoI GFP	aaaaCTCGAGGTCGATGGTG
16	3 GFP BHI	aaaaGGATCCCCCCGGACTTGTACAGCTCGTCC

Table S3. Antibodies used for immunoblotting and immunofluorescence

Primary antibodies	**Supplier**	**Species**	**Dilution**	**Reference**
GFP	Santa Cruz	rabbit	1:1,000	sc8334
GAPDH	Fitzgerald	mouse	1:10,000	10R-G109a
NPM1	Abcam	mouse	1:1,000	ab10530
Secondary antibodies				
anti-mouse HRP	GE Healthcare	sheep	1:5,000	NXA931V
anti-rabbit HRP	GE Healthcare	donkey	1:5,000	NXA934V
anti-mouse Alexa Fluor 633	Invitrogen	goat	1:1,500	A21050