Aneuploidy reveals insights into control of protein complex stoichiometry

by

Christopher M. Brennan

B.S. Biology
Providence College, 2013

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2019

Signature of the Author: _____
Department of Biology
July 15, 2019

Certified by: _____
Angelika Amon
Kathleen and Curtis Marble Professor of Cancer Research
Investigator, Howard Hughes Medical Institute
Thesis Supervisor

Accepted by: _____
Stephen Bell
Uncas and Helen Whitaker Professor of Biology
Investigator, Howard Hughes Medical Institute
Co-Director, Biology Graduate Committee

Aneuploidy reveals insights into control of protein complex stoichiometry

By

Christopher M. Brennan

Submitted to the Department of Biology
on July 15, 2019 in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Biology

Abstract

Aneuploidy, or an incorrect number of chromosomes, is caused by errors in chromosome segregation during cell division. Because genes are expressed in accordance with their copy number, aneuploidy simultaneously alters the gene dosage of hundreds to thousands of genes. The outcome is an imbalanced proteome, which has a negative impact on cellular physiology and places intense demand on the protein quality control system of the cell to effectively fold and/or degrade proteins. Aneuploidy further represents an ideal model for studying how cells cope with imbalances in their proteome as it allows for interrogation of the fate of hundreds to thousands of imbalanced proteins simultaneously.

Here, we identify protein complex stoichiometry imbalances as a major cause of protein aggregation in aneuploid cells. Subunits of protein complexes encoded on excess chromosomes aggregate in aneuploid cells, which is suppressed when expression of other subunits is coordinately altered. We further show that excess subunits are either degraded or aggregate, a fate that is largely mutually exclusive for individual subunits. We also demonstrate that protein aggregation is nearly as effective as protein degradation at lowering levels of excess proteins. Our study explains why proteotoxic stress is a universal feature of the aneuploid state and reveals protein aggregation as a form of dosage compensation to cope with disproportionate expression of protein complex subunits. This work informs both our comprehension of aneuploid cell physiology, and also provides a more complete understanding of how aneuploid and euploid cells cope with stoichiometric imbalances, namely that protein aggregation can function as protein quality control mechanism in this regard.

Thesis supervisor: Angelika Amon
Title: Kathleen and Curtis Marble Professor of Cancer Research; Investigator, Howard Hughes Medical Institute

This thesis is dedicated to my parents, who taught me to do what makes me happy.

Acknowledgements

First, I need to thank my advisor, Angelika, for her unparalleled enthusiasm. Thank you for always encouraging me and pushing me to be my best. I am so fortunate to have had a mentor that cares so deeply about my development as a scientist. It was an absolute honor and privilege to be able to learn from such an outstanding scientist and mentor. I also need to thank my undergraduate research advisor, Brett, whose passion for science is the reason why I decided to pursue a PhD.

The Amon lab has been a fantastic place to spend these last 5 years. To past and present members of the lab, thank you for creating a wonderful environment to do science in. Thank you for all of your time and energy spent thinking about my work. I need to highlight Ana, my rotation mentor who was instrumental in getting this project started and spent her last summer in the lab teaching me. Also, my two bay mates, Luke and Summer, both of whom I have learned so much from. Your friendship made coming to lab every day much more enjoyable.

I would also like to acknowledge my thesis committee members, Steve Bell and Gene-Wei Li. Your input throughout my grad career was invaluable for guiding the direction of my project and career. Also, thank you to Dan Finley for graciously agreeing to serve as the outside member of my thesis committee.

I have been fortunate to have great friends both in and out of the biology department. Thank you to my Biograd2013 classmates for always keeping things fun and light-hearted. To my non-MIT friends, you have been the most welcomed distraction from work. Thank you to Rushmore, the North End trivia crew, Pin and Tonic, and ENDR for drinking beers with me, bowling, playing video games, and generally just having fun. Thank you to my roommates Corey, Eric, and Erik who always at least tried to understand what happened in my day. You made living in Boston a great time in my life. I need to thank my oldest friend, Brianna, for putting up with me since we were born. I can't believe that 28 years later, our dogs are best friends.

I consider myself extremely fortunate to have a family that are also my best friends and greatest supporters. Thank you to my grandparents Mamie, Pa, Nana, and Pop for your constant love. Family gatherings with my aunts, uncles, and cousins, both Brennan's and Roberts' have been a consistent source of happiness throughout my life. And thank you to the Fletcher's for welcoming me into their family. I've greatly enjoyed all of my trips to Washington and your visits to Boston.

I could not have done this without the support of my Mom and Dad. Mom, thank you for fostering my love of science. Dad, thank you for encouraging me to find what I love to do, then figure out how to make a career out of it. To my brother and best friend JR, thank you for reminding me not to take things (especially myself) too seriously. To my dog Bradley, thank you for attacking me with same affection and enthusiasm when I come home regardless of whether my experiment failed or succeeded. Lastly, thank you to my wife and life partner, Marissa, for being my greatest source of support and inspiration since we met on the first day of graduate school. Meeting you was the best thing that ever happened to me.

Table of Contents

Chapter 1: Introduction

Aneuploidy, or an incorrect number of chromosomes, is caused by errors in chromosome segregation during cell division. Every time a cell divides, it must accurately segregate equal numbers of chromosomes between the mother and daughter cells. While mechanisms exist to ensure faithful chromosome segregation, errors occur at a rate of $5x10^{-4}$ in yeast and 10^{-4} to 10^{-5} in mammals (Hartwell and Smith, 1985; Rosenstraus and Chasin, 1978). Because genes are expressed in accordance with their copy number, aneuploidy simultaneously alters the gene dosage of hundreds to thousands of genes. The result is an imbalanced proteome, which has a negative impact on cellular physiology and places intense demand on the protein quality control system of the cell to effectively fold and/or degrade proteins.

In this introduction, I will first discuss the effects of aneuploidy on cellular and organismal fitness. I will then provide an overview of how cells utilize protein quality control mechanisms to maintain a healthy proteome, including the role of protein aggregation in protein homeostasis. Finally, I will discuss the effects of aneuploidy on the proteome and our current understanding of how cells cope with the problem of an imbalanced proteome.

ANEUPLOIDY DECREASES FITNESS

Causes of aneuploidy

Aneuploidy is caused by errors in chromosome segregation during cell division. In a normal mitotic division, replicated sister chromatids are linked by the protein cohesin during S phase (Nasmyth and Haering, 2009). In prophase, these sisters are then attached to opposite poles of the mitotic spindle by the interaction between spindle microtubules and the kinetochore, a large protein assembly that forms at the centromeres of chromosomes. Attached chromosomes align during metaphase and are bioriented, i.e. with sister kinetochores facing in opposite

directions. At this point, the spindle assembly checkpoint (SAC) ensures that sister chromatids are properly attached to the spindle and bioriented such that chromosome segregation will result in partitioning of one sister chromatid into each daughter (Fig. 1, Musacchio and Salmon, 2007). If sister chromatids are correctly attached and bioriented, the pulling force in opposite directions generated by the spindle results in tension across the sister kinetochores. This tension satisfies the checkpoint, resulting in the activation of the ubiquitin ligase known as the anaphase-promoting complex or cyclosome (APC/C) and its activating subunit Cdc20. APC/C bound to Cdc20 leads to cleavage of cohesin by degrading Securin, the inhibitor of the protease called Separase. Cleavage of cohesin by Separase begins anaphase and results in the separation of sister chromatids to opposite poles.

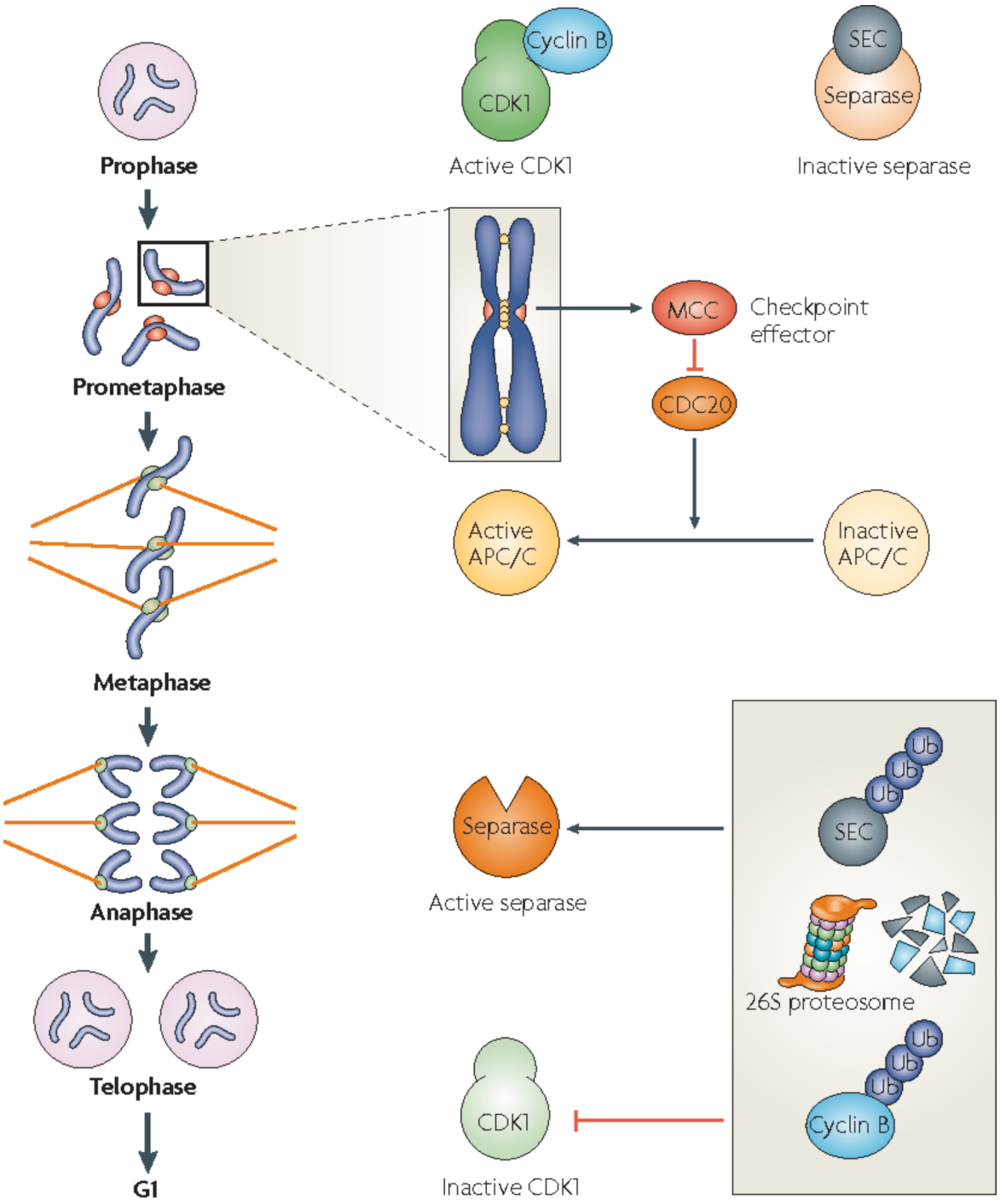

Figure 1. The spindle assembly checkpoint (SAC) ensures faithful chromosome segregation (adapted from (Musacchio and Salmon, 2007))

Unattached kinetochores and sister chromatids that lack tension lead to the generation of the mitotic checkpoint complex (MCC). The MCC keeps the APC/C inactive by inhibiting its

activating subunit Cdc20. Once all chromosomes have been correctly attached the spindle and tension is generated, the SAC is inactivated relieving the inhibition of Cdc20. APC/C-Cdc20 mediates ubiquitination and proteolysis of Securin, the inhibitor of Separase. Once activated, Separase cleaves cohesin molecules that hold the bioriented sister chromatids together causing them to be pulled to opposite poles. APC/C-Cdc20 also facilitates the degradation of mitotic cyclins, thus lowering CDK activity and preparing cells to exit from mitosis.

One source of chromosome mis-segregation is the premature loss of sister-chromatid cohesion. This can be caused by incorrect establishment of cohesion, hyperactivation of Separase, or by aberrantly low Securin activity. In these cases, sister chromatids are segregated as soon as they attach to the mitotic spindle, so any incorrect attachments result in aneuploid daughter cells (Nasmyth and Haering, 2009).

Incorrect attachments of chromosomes to the spindle and faulty inactivation of the SAC are another source of chromosome-segregation errors. Amphitelic attachments describe properly attached sister chromatids, with each sister attached to microtubules emanating from opposite polls. Monotelic (only one kinetochore is attached to the spindle) and syntelic (both sister kinetochores are attached to the same pole) do not generate tension and thus activate the SAC (Fig. 2, (Knouse et al., 2017)). If the SAC fails to arrest cells in metaphase due to mutations in checkpoint components, these types of attachments lead to chromosome mis-segregation and aneuploidy. Merotelic attachments occur when one kinetochore is attached to both poles and are thought to be poorly sensed by the SAC since merotely can still result in tension across the sister kinetochores. In these cases, chromosomes lag behind the rest of the migrating chromosomes and

can be mis-segregated or incorporated into micronuclei (Cimini et al., 2001; Thompson and Compton, 2011).

Figure 2. Types of kinetochore-microtubule attachments (adapted from Knouse et al., 2017)

Kinetochore-microtubule attachments that result in properly bioriented sister chromatids attached to opposite poles of the spindle are termed amphitelic. Amphitelic attachments satisfy the spindle assembly checkpoint (SAC) and result in equal segregation of chromosomes. Monotelic attachments occur when only one kinetochore attaches to one pole, while syntelic attachments occur when both kinetochores are attached to the same pole. Both cause activation of the SAC and, under normal circumstances, delay anaphase until the an amphitelic attachment is achieved. Merotelic attachments occur when one kinetochore is attached to both poles. Depending on the strength of the attachments to each pole, the SAC can be satisfied by merotelic attachments and anaphase proceeds. The chromatid that is attached to both poles lags behind the rest of the migrating chromosomes resulting in mis-segregation or the formation of a micronucleus. A micronucleus forms when the lagging chromosome is surrounded by membrane that is separate from the main nucleus causing the isolation of the lagging chromosome.

If chromosome mis-segregation occurs in a mitotic division early in development, this can result in a large percentage of cells in the organism being aneuploid. Chromosome

segregation errors can also occur during either division of meiosis, producing aneuploid gametes that give rise to whole organisms with constitutional aneuploidy.

Models of aneuploidy used in this thesis

In this thesis, I have employed two models of aneuploidy in budding yeast and one model in human cells to study the molecular nature of protein aggregates in aneuploid cells. In general, aneuploid cells can be generated such that they have specific, known karyotypes or random karyotypes. Defined aneuploidies have the advantage of enabling assessment of chromosome-specific effects of aneuploidy. Cells with randomly generated karyotypes are useful for studying general effects of aneuploidy and are also capable of resulting in more complex combinations of chromosome gains and losses. However, they are often karyotypically unstable. The primary model used in this work are haploid budding yeast containing a single extra copy of one of their 16 chromosomes (n+1) called disomes (Fig. 3A). Each copy of the duplicated chromosome is marked with a unique selectable marker, allowing for stable maintenance of the aneuploid karyotype. Disomes were constructed by marking the same locus with different markers in separate haploid strains. One of the parental strains contains a mutation in the karyogamy gene, *KAR1*, which when mutated in one mating partner prevents nuclear fusion from occurring (Conde and Fink, 1976). These strains were then crossed to one another, resulting in an abortive mating. Occasionally entire chromosomes are transferred from one nucleus to the other, resulting in aneuploid progeny (Torres et al., 2007). Chromosome transfer events can then be selected for by growing progeny in conditions that require the selectable marker on both chromosomes.

To study higher levels of chromosomal imbalance, I have utilized a strain harboring a temperature sensitive allele of the kinetochore component *NDC10* (Goh and Kilmartin, 1993).

When grown at the semi-permissive temperature, *ndc10-1* mutants mis-segregate chromosomes at a very high rate, resulting in a heterogeneous population of aneuploid cells with a range of karyotypes (Oromendia et al., 2012).

To assess whether principals uncovered in yeast are generalizable to higher eukaryotes, I have also utilized a series of human cell lines with stable, defined karyotypes. Retinal pigmented epithelial (RPE-1) cells immortalized by the expression of telomerase reverse transcriptase are nearly diploid, allowing for an adequate euploid control in experiments. These cells were made aneuploid by microcell mediated chromosome transfer (Fournier and Ruddle, 1977; Stingele et al., 2012). By this method, micronuclei containing whole chromosomes are generated by treating cells with a microtubule poison, then purified and fused with acceptor cells, resulting in whole chromosome gains (Fig. 3B).

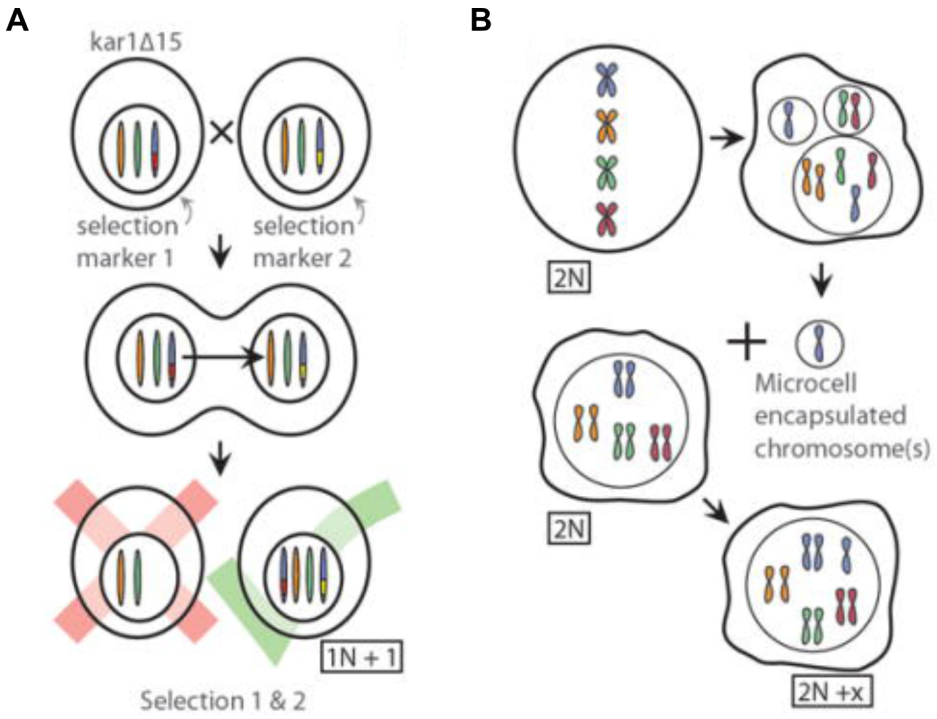

Figure 3. Models of aneuploidy used in this thesis (adapted from (Siegel and Amon, 2012))

(A) Haploid yeast containing a single extra copy of one chromosome (disomes) were created using an abortive mating scheme. One parental strain harbored a mutation in the *KAR1* gene to prevent nuclear fusion upon mating. The two parental strains harbored distinct selectable markers at the same locus allowing for selection of progeny that had inherited both copies of the chromosome due to transfer of the whole chromosome.
(B) Human cells with trisomies were generated by microcell mediated chromosome transfer. Micronuclei containing whole chromosomes can be created by treating cells with a microtubule poison. Chromosome-containing micronuclei can be purified and fused with a cell, resulting in the incorporation of the extra chromosome into nucleus of the acceptor cell.

Effects of aneuploidy on organismal physiology

Over 100 years ago, Marcella O'Grady and Theodor Boveri noted the detrimental effects that aneuploidy had on organismal fitness by observing the failed development of sea urchin embryos that had undergone chromosome mis-segregation (Boveri, 1902). Today we understand that aneuploidy that arises during gamete formation is the leading cause of miscarriage in humans because, with rare exceptions, aneuploid human embryos fail to survive to term (Alberman and Creasy, 1977; Hassold and Jacobs, 1984). Humans with trisomy of chromosome 13 (Patau syndrome) and trisomy 18 (Edwards syndrome) can survive to birth, but only ~10% of these individuals live to be one year of age (Rasmussen et al., 2003). Trisomy of chromosome 21 causes Down syndrome, which is characterized by developmental defects including impaired growth and brain development (Siegel and Amon, 2012). In all other organisms examined, aneuploid progeny either fail to survive to term or have developmental defects reminiscent of those observed in humans.

Aneuploidy can also arise spontaneously in adult organisms due to mitotic chromosome mis-segregation. In most adult organisms, aneuploidy is quite rare, occurring in only ~2-5% of cells (Knouse et al., 2014). Mosaic variegated aneuploidy (MVA) is caused by mutations in the mitotic machinery or SAC components that ensure chromosomes are properly attached to the mitotic spindle, resulting in individuals with upwards of 25% of their cells being aneuploid (Hanks et al., 2004; Snape et al., 2011). Like individuals with constitutional aneuploidies, humans with MVA exhibit growth and developmental deficiencies. Mice harboring a hypomorphic mutation in the SAC were found to have higher numbers of aneuploid cells in non-regenerating adult tissues such as the brain, while highly proliferative tissues like the intestine were almost entirely euploid, indicating that aneuploid cells may be selected against *in vivo* (Pfau et al., 2016). This could in part be due to euploid cells outcompeting slow growing aneuploid cells, but recent evidence suggests that aneuploid cells can be recognized by the immune system (Andriani et al., 2016; Santaguida et al., 2017; Sullivan et al., 2016). At least *in vitro*, aneuploid cells that have arrested in the cell cycle with complex karyotypes secrete pro-inflammatory signals that facilitates their clearance by natural killer (NK) cells (Santaguida et al., 2017). Combined with cell autonomous methods to ensure faithful chromosome segregation, immune-mediated clearance of aneuploid cells may also protect organisms against the harmful effects of aneuploidy.

Aneuploid fitness defects are caused by changes in gene expression

To understand why aneuploidy is universally detrimental to organismal development, it is important to explore how an imbalanced karyotype affects individual cells. The most important question to address is whether the extra chromosome is expressed. Using disomic yeast described

above, it was established that chromosome gain leads to a corresponding increase in RNA and protein levels for most genes on the gained chromosome (Torres et al., 2007; 2010). Consistent with developmental defects observed in aneuploid multicellular organisms, disomic yeast have impaired proliferation. This phenotype is caused by the expression of the extra chromosome, as euploid yeast containing a yeast artificial chromosome of up to 1.6 Mb (13% of the genome) encoding no yeast genes have no proliferation delay (Torres et al., 2007).

Fundamentally, aneuploidy is a problem of gene dosage. Phenotypes could be caused by changing the dosage of just one to a few genes, affecting the related cellular process, or they could be caused by mass action of simultaneously altering the dosage of many genes. There is evidence that aneuploid phenotypes are caused by both of these mechanisms.

Single gene effects of aneuploidy could be caused by haploinsufficiency (chromosome loss in a diploid) and sensitivity to increased copy number (chromosome gain). Haploinsufficiency describes when a phenotype in a diploid organism is caused by a heterozygous loss of function mutation. In yeast grown in rich medium, about 3% of the genome is haploinsufficient, and most of these genes are involved in protein translation, particularly ribosomal subunits (Deutschbauer et al., 2005). With the exception of chromosome I, all yeast chromosomes contain at least one ribosomal subunit, which could in part explain the fitness defects caused by chromosome loss.

Increasing gene expression of a single gene by just two-fold can also have fitness consequences. Perhaps the most dramatic example comes from studies of tubulin in yeast. Just a 1.4-fold overexpression of β-tubulin causes a decrease in cell viability (Katz et al., 1990). Other instances of changes in fitness due to a single extra copy of a gene have been harder to discover. In a study of dosage sensitive genes (defined as causing fitness penalties when encoded in

greater than five copies) only one gene was found to have a proliferation delay when expressed at one extra copy (Bonney et al., 2015). A recent study using a pooled growth competition of strains harboring barcoded genes on single copy plasmids found that 15% of genes significantly delayed proliferation (Morrill and Amon, 2019). Thus, particularly for chromosome gain, it is uncommon for single genes to account for the entirety of proliferation defects observed in aneuploid cells. Karyotype specific aneuploid phenotypes are often observed, arguing that a two-fold increase in expression of one or a small group of genes can have specific phenotypic consequences. For example, disome XVI has defects in protein transport not observed in other disomes as evidenced by its unique sensitivity to gene deletions in protein trafficking components and drugs targeting protein trafficking (Dodgson et al., 2016a). What gene or group of genes on chromosome XVI causes this effect is still elusive.

General phenotypic consequences of aneuploidy have been studied more readily. These phenotypes have been observed in aneuploid cells with distinct karyotypes even across species. Apart from decreased proliferation, aneuploidy commonly causes metabolic stress and genomic instability (Blank et al., 2015; Hwang et al., 2017; Passerini et al., 2016; Sheltzer et al., 2011; Tang et al., 2017; Torres et al., 2007; Williams et al., 2008). These various stresses along with slow growth of aneuploid cells also causes the activation of a transcriptional response known as the environmental stress response (ESR) (Sheltzer et al., 2012). The ESR was found to downregulate genes involved in macromolecule biosynthesis and up regulate stress response genes to help cells cope with multiple different environmental stresses (Gasch et al., 2000). Recently it was discovered that multiple distinct karyotypes in yeast and mice cause an increase in variability, or “noise”, in biological pathways (Beach et al., 2017). Finally, as a direct result of altering the proteome, aneuploid cells have defects in protein folding and turnover, a condition

known as proteotoxic stress. The causes and consequences of proteotoxic stress in aneuploid cells will be discussed in depth later in this introduction and in Chapter 2.

Cancer and the aneuploidy paradox

Despite the numerous penalties of aneuploidy described above, aneuploidy is observed in 77% of tumors in The Cancer Genome Atlas (Knouse et al., 2017). This striking feature of cancer cells prompted Theodor Boveri to propose that aneuploidy causes cancer (Boveri, 1914). This presents a paradox: in most contexts, aneuploidy impairs cellular proliferation, yet it is frequently associated with cancer, a disease defined by cells with increased proliferation. This begs the question of whether aneuploidy is a cause or simply a consequence of transformation. Theoretically, aneuploidy could promote tumorigenesis by altering the copy number of tumor suppressors and oncogenes. It is also possible that dysregulation of the cell cycle observed in cancer leads to an increase in chromosome segregation errors, and aneuploidy is an unwanted side effect of rapid proliferation. Evidence for both of these scenarios exist, and the complexity of the role of aneuploidy in cancer is perhaps best exemplified by the fact that people with Down syndrome have a decreased risk of forming solid tumors and an increased risk of developing leukemia (Hasle et al., 2000; Yang et al., 2002).

Analysis of many cancer genomes revealed that gains of oncogenes and loss of tumor suppressors do impact the karyotype of tumors (Davoli et al., 2013). Direct evidence that aneuploidy can drive tumorigenesis comes from studies of chromosomal instability (CIN), a condition which causes aneuploidy due to increased rates of chromosome mis-segregation. In one study, CIN was able to accelerate tumor progression by promoting loss of heterozygosity of key tumor suppressors such as p53 (Baker et al., 2009). Indeed, individuals with mosaic

variegated aneuploidy are at an increased risk of developing childhood cancers (Hanks et al., 2004).

While genomic or chromosomal instability may be able to facilitate tumorigenesis, much evidence suggests that aneuploidy per se does not cause cancer. Aneuploid cells of multiple karyotypes transformed with oncogenes do not proliferate faster than euploid cells transformed with the same oncogene (Sheltzer et al., 2017). Additionally, levels of aneuploidy tend to be lower in early lesions, and higher in more developed tumors suggesting that aneuploidy arises later in oncogenesis (Ried et al., 1996; Ross-Innes et al., 2015). While specific chromosomal alterations are associated with some cancers, the lack of a common "pan-cancer" karyotype and the variability of karyotypes even among individuals with same type of cancer strongly argues that aneuploidy is infrequently a cancer driver (Knouse et al., 2017). Though it is rarely the cause of cancer onset, it is possible that aneuploidy plays a role in metastasis and resistance to treatment by providing genetic variability within the tumor population, thus allowing for adaptation. This is supported by the fact that high levels of aneuploidy are associated with poor patient prognosis (Byrd et al., 2002; Emdin et al., 1987). Establishing the role of aneuploidy in later stages of cancer development will be important in fully understanding the impact of karyotype on disease progression.

MAINTENANCE OF PROTEIN HOMEOSTASIS

Molecular chaperones assist protein folding

From bacteria to humans, the final step needed for expression of a gene to result in a functional protein is the acquisition of its correct three-dimensional structure. Proteins can adapt many different energetically favorable conformations, but these misfolded states do not result in

a functional protein (Balchin et al., 2016) and negatively impact the fitness of the cell (Geiler-Samerotte et al., 2011). Although the information needed for a protein to fold correctly is encoded in its amino acid sequence (Anfinsen, 1973), most proteins require assistance from molecular chaperones to achieve a properly folded state. The major cellular chaperones are HSP70s, HSP90s, and chaperonins.

HSP70 binds short stretches of hydrophobic amino acids on nascent peptides and newly synthesized proteins (Rüdiger et al., 1997). Through ATP hydrolysis cycles, HSP70 binds and releases substrates, giving them an opportunity to fold (Fig. 4). ATP-bound HSP70 has low affinity for substrates, allowing substrates to associate and dissociate freely (Mayer, 2010). HSP40, a co-chaperone for HSP70, can deliver substrates to HSP70 and stimulates its ATPase activity, causing HSP70 to transition into a high affinity, ADP-bound state (Kampinga and Craig, 2010). Nucleotide exchange factors then facilitate the exchange of ADP for ATP, causing the substrate to be released into the cytoplasm. If the substrate is able to fold, the hydrophobic residues become buried and are no longer recognized by HSP70 (Hartl et al., 2011).

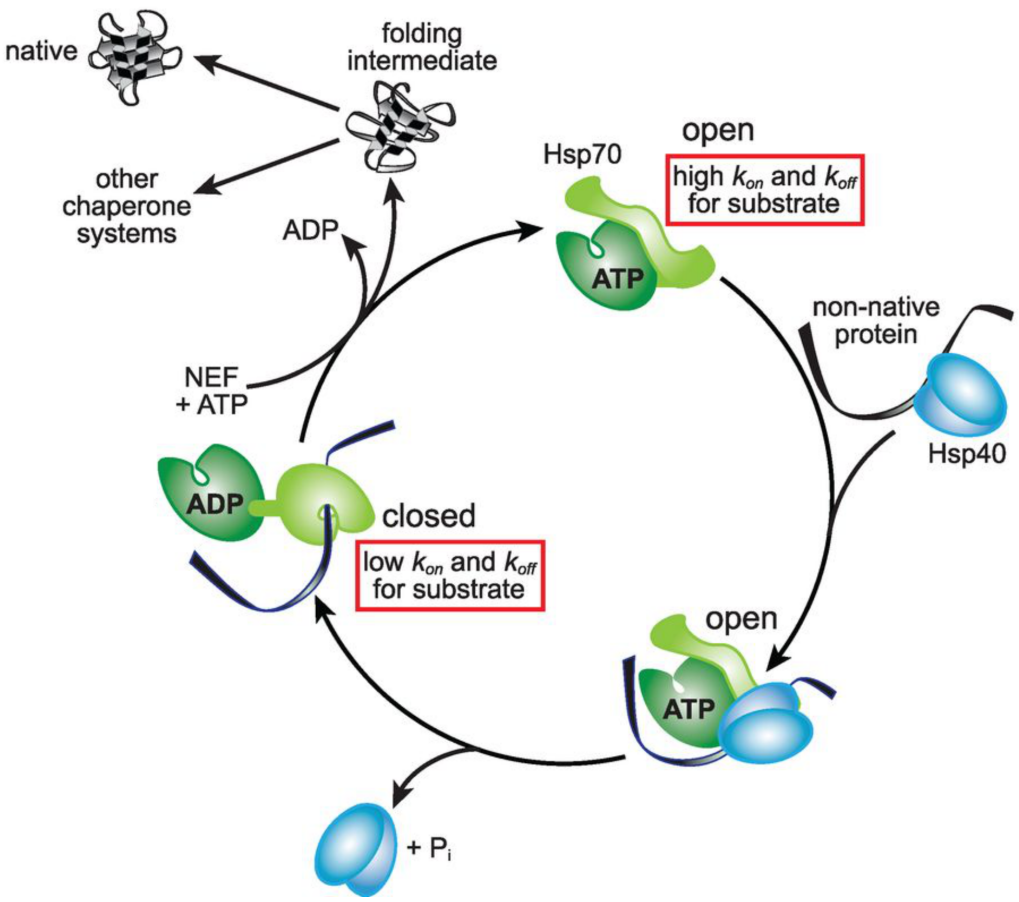

Figure 4. HSP70 conformation cycles are driven by ATP hydrolysis (adapted from (Balchin et al., 2016))

HSP70 in its ATP bound state has fast on and off rates for substrate binding and release (open conformation). The HSP70 co-chaperone, HSP40, delivers non-native proteins to ATP bound HSP70. HSP40 also stimulates the ATPase activity of HSP70, causing it to hydrolyze ATP. ADP bound HSP70 (closed conformation) tightly binds the substrate that was delivered by HSP40. HSP70 nucleotide exchange factors (NEFs) then facilitate exchange of ADP for ATP, causing HSP70 to transition back to its open conformation, releasing its substrate. This substrate can either be re-captured by HSP70 if it did not fold, be delivered to another chaperone system, or reach its fully native form with no further assistance.

HSP90 functions downstream of HSP70, assisting in the folding of a large number of client proteins that include protein kinases and steroid receptors (Taipale et al., 2010). Like HSP70, HSP90 works via an ATP hydrolysis cycle, however it functions as a dimer and features many more co-chaperones that facilitate client selection and modulate ATP hydrolysis (Shiau et al., 2006).

Chaperonins are barrel shaped protein complexes that encapsulate a single substrate protein allowing it to fold in isolation thereby preventing inappropriate interactions with other proteins. The chaperonin of the eukaryotic cytoplasm, known as TRiC (or CCT), is composed of two sets of 8 subunits that form back-to-back rings (Muñoz et al., 2011). It is most notably responsible for aiding in the folding of actin and tubulin. HSP70 can deliver substrates to TRiC, which enter the barrel and are entrapped by a lid domain which opens and closes in an ATP-dependent manner (Douglas et al., 2011). The inner walls of chaperonins are composed of hydrophilic, negatively-charged residues that encourage the enclosed substrate to fold by burying hydrophobic residues (Horwich and Fenton, 2009). Upon completion of the ATP hydrolysis cycle, the lid domain opens and the substrate is released. If one round of ATP hydrolysis was not enough time for the substrate to attain its folded conformation, the substrate can reenter the barrel for subsequent rounds of entrapment and release (Hartl et al., 2011).

Protein complex formation is the final step in protein folding

In addition to achieving a correct three-dimensional structure, many proteins rely on stable interactions with other proteins to function. Homomeric protein complexes are composed of multiple subunits of the same protein while heteromeric protein complexes are composed of at least two unique subunits. In principal, this final step in protein folding involves two or more

proteins diffusing through a crowded cytoplasm then colliding with one another in an orientation that results in their proper association (Ellis, 2001; Williams and Dichtl, 2018). This presents a large problem, as protein-protein interfaces are "sticky" and have the potential to form inappropriate interactions, especially with unfolded nascent polypeptides, when they are not buried within the structure of the mature protein complex (Levy et al., 2012; Pechmann et al., 2009). In some cases, the cell copes with this problem by employing specific chaperones to aid in a multi-step, ordered assembly process, e.g. the 20S proteasome and yeast vacuolar ATPase (Le Tallec et al., 2007; Smardon et al., 2002).

In theory, this problem could also be solved by assembling complexes as the translated proteins emerge from the ribosome. For homomeric complexes, this can happen because a single mRNA is typically translated by multiple ribosomes, resulting in spatially proximal subunits (Kiho and Rich, 1964; Zipser and Perrin, 1963). In prokaryotes, where subunits of the same complex are often encoded by the same polycistronic message, even heteromeric complexes can be assembled co-translationally (Shieh et al., 2015). In fact, subunit-encoding genes are ordered within operons to optimize complex assembly (Wells et al., 2016). Operons are rare in eukaryotes, so co-translational complex assembly would require co-localization of mRNAs or localization of one mature subunit to the site of translation of another subunit. Indeed, mRNAs encoding separate subunits of the actin nucleation complex, Arp2/3, are co-localized to the leading edge protrusions of migrating fibroblasts (Mingle et al., 2005). In fission yeast, it was estimated that ~40% of proteins (without RNA binding domains) associated with mRNAs of binding partners (Duncan and Mata, 2011). More recent evidence confirms that co-translational assembly of complexes is common in budding yeast. Shiber and colleagues performed immunopurification of single complex subunits and observed interaction of that subunit with

nascent polypeptides of their binding partners using ribosome profiling (Shiber et al., 2018). This was the case for nine out of twelve tested complexes. The three that did not show evidence of co-translational assembly have dedicated chaperones to assist in assembly as described above. Interestingly, for six of the nine complexes, assembly is directed, i.e. fully translated subunit A interacts with nascent subunit B, but not vice versa (Shiber et al., 2018). Using a similar approach, co-translational assembly of three nuclear complexes was also observed in mammalian cells (Kamenova et al., 2019). Thus, co-translational assembly of heteromeric complexes appears to be the norm and not the exception. However, it is still unclear how one subunit is targeted to the translation site of other subunits.

The ubiquitin proteasome system degrades misfolded proteins

When proteins fail to achieve their native conformation or are damaged, the ubiquitin-proteasome system (UPS) can degrade them. Misfolded or damaged proteins are recognized by ubiquitin ligases, which aid in the covalent attachment of a charged ubiquitin molecule from a ubiquitin-conjugating enzyme to a lysine residue on the substrate by binding both the ubiquitin-conjugating enzyme and the substrate (Finley et al., 2012). Thus, ubiquitin ligases confer substrate selectivity to the ubiquitination reaction. The two classes of ubiquitin ligases are RING domain and HECT domain ubiquitin ligases that differ in the mechanism by which they facilitate the ubiquitination reaction. RING domain ubiquitin ligases position the ubiquitin-conjugating enzyme to effectively transfer the ubiquitin molecule directly to the substrate (Deshaies and Joazeiro, 2009). For HECT domain ubiquitin ligases, the ubiquitin-conjugating enzyme first transfers the ubiquitin molecule to a cysteine residue on the ubiquitin ligase before it is covalently bound to the substrate (Scheffner et al., 1995).

There are numerous ubiquitin ligases that control many cellular processes, including a set that are dedicated to recognizing misfolded and damaged proteins in various cellular compartments. Some recognize misfolded proteins by binding to exposed hydrophobic residues (Fredrickson et al., 2011), whereas others bind to chaperones in order to find misfolded proteins (Heck et al., 2010). Multiple rounds of ubiquitin conjugation on a single substrate results in the creation of a polyubiquitin chain. Ubiquitin itself contains seven lysines, all of which can be ubiquitinated (Komander and Rape, 2012). All types of chains except lysine 63 (K63) linkages are competent to signal for degradation by the proteasome with K48 linkages being most common (Xu et al., 2009). A chain of at least four ubiquitin molecules is required for efficient recognition by the proteasome (Thrower et al., 2000). Polyubiquitinated substrates can be delivered to the proteasome by proteins that act as shuttles. These shuttle proteins contain a ubiquitin-like (UBL) domain that is bound by the proteasome and a ubiquitin binding (UBA) domain that binds the polyubiquitin chain on the substrate thereby physically linking the ubiquitinated protein to the proteasome (Elsasser and Finley, 2005).

The proteasome is a holoenzyme composed of two multi-subunit particles: the 19S regulatory particle recognizes substrates and the 20S core particle degrades them. The common cytosolic form of the proteasome is the 26S proteasome, which is composed of the barrel-shaped 20S core particle and two 19S regulatory particles situated on opposite ends of the barrel that serve as lids (Finley, 2009). Ubiquitin chains conjugated to substrates are recognized by two lid subunits that bind ubiquitin. Along with deubiquitinating enzymes, the lid then removes ubiquitin chains from proteins prior to their degradation so that ubiquitin can be recycled (Leggett et al., 2002; Verma et al., 2000; 2002; Yao and Cohen, 2002). The lid also performs the important function of unfolding proteins and threading them into the core particle by ATP

hydrolysis (Sauer and Baker, 2011). The barrel-shaped core particle is composed of four stacked rings, with the two inner rings harboring subunits that have trypsin-like, chymotrypsin-like, and caspase-like protease activity (Groll et al., 1997). Once inside the core particle, peptide bonds of the substrate are cleaved, resulting in short polypeptides that are typically between 3 and 22 amino acids long (Kisselev et al., 1999).

Protein aggregation as a strategy for protein quality control

When proteins do not fold properly and fail to be degraded, aggregation can occur as a result of inappropriate interactions between exposed hydrophobic residues that would be buried within the folded structure of the protein (Fig. 5, (Tyedmers et al., 2010)). Types of protein aggregates were historically defined by electron microscopy as either amyloid or amorphous. Amyloid aggregates are highly ordered, characterized by cross-β sheet secondary structure instead of the typical folded state of the protein (Dobson, 2003; Fändrich and Dobson, 2002). The amyloid form of some proteins can perform functions separate from the soluble form of the protein such as translational control in yeast meiosis and long-term memory persistence in Drosophila (Berchowitz et al., 2015; Majumdar et al., 2012). However, amyloid proteins are perhaps most known for their role in disease phenotypes including Alzheimer's, Parkinson's, and prion diseases (Knowles et al., 2014).

Amorphous aggregates have no defined structure and it is becoming apparent that they vary widely in their nature. The types of aggregates formed varies based on both the identity of the aggregating proteins and the conditions that cause aggregation (Stathopulos et al., 2003; Wang et al., 2010). Whether amorphous aggregates are toxic to cells will be discussed below.

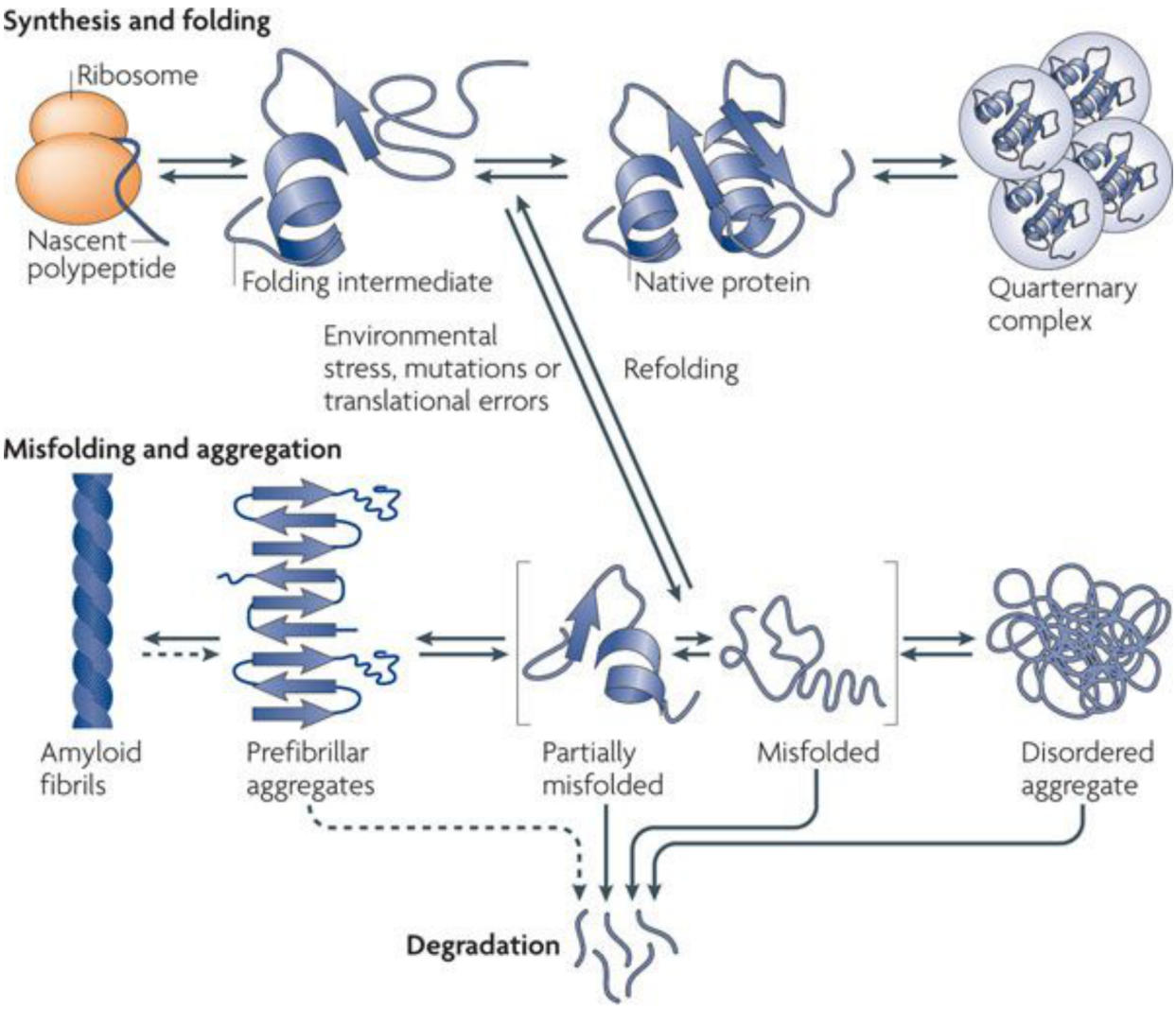

Figure 5. Protein folding and aggregation (adapted from (Tyedmers et al., 2010))

Protein folding begins at the ribosome as nascent chains adopt three dimensional structures, resulting in the formation of native proteins. Nascent chains, partially folded intermediates, and orphan complex subunits are particularly vulnerable to misfolding and aggregation because they contain hydrophobic stretches that become buried in the core of the fully folded protein or protein complex. Molecular chaperones aid in the process of protein folding and prevent aggregation by binding these exposed hydrophobic residues. When aggregation does occur,

proteins can form cross-β sheets resulting in the formation of amyloids or disordered aggregates that lack any defining structure. Proteins in aggregates can then be disaggregated and either refolded or degraded. Entire aggregates can also be degraded by selective autophagy.

Once an aggregate has formed, cells generally have two mechanisms for removing them: disaggregation and autophagy. In yeast, disaggregation is carried out by Hsp70 (Ssa1), Hsp40 (Ydj1), and the AAA+ ATPase Hsp104 (Glover and Lindquist, 1998). Hsp104 functions as a homohexameric ring, with each subunit capable of hydrolyzing ATP. Hsp70 and Hsp40 transfer proteins from the aggregates to the central pore of the Hsp104 ring (Haslberger et al., 2007; 2008). Hsp104 then threads the substrate through its central pore by hydrolyzing ATP, thus disentangling it from the aggregate and allowing for refolding or degradation (Mogk et al., 2018). Metazoans notably lack Hsp104, however they still effectively eliminate protein aggregates (Pinto et al., 1991). Instead, separate classes of metazoan HSP40s form transient complexes that are capable of conferring disaggregation activity to HSP70 (Nillegoda et al., 2015). Precisely how this system is capable of removing proteins from aggregates is still unknown.

Rather than individually disaggregate proteins, cells can also degrade aggregates entirely utilizing the cellular recycling system known as autophagy. Autophagy involves the formation of a double membrane structure within the cytoplasm of the cell around the substrate that is to be degraded. Eventually, this autophagosome fuses with the lysosome (or vacuole in yeast) where the contents are deposited for degradation (Reggiori and Klionsky, 2013). This process can be used to turn over bulk cytoplasm or specific cargo such as entire organelles. In mammals, p62 functions as the cargo receptor for protein aggregates, tagging them for turnover by autophagy.

p62 interacts with ubiquitinated protein aggregates via its ubiquitin binding (UBA) domain and with the autophagosome membrane bound receptor, LC3, through its LC3-interacting region (Bjørkøy et al., 2005; Pankiv et al., 2007). In yeast, Cue5 has been identified as the link between ubiquitination and autophagy. Like, p62, Cue5 contains a ubiquitin-binding domain and was demonstrated to bind Atg8, the yeast homolog of LC3 (Lu et al., 2014). Cue5 is unrelated to p62, as it binds ubiquitin via its CUE domain rather than a UBA domain. Interestingly, the identification of Cue5 in yeast led to the discovery of CUE domain containing autophagy adaptors in mammals, suggesting that these proteins may be the ancestral autophagy adaptors for ubiquitin conjugated proteins (Lu et al., 2014). In yeast, disaggregation and refolding of aggregated proteins appears to be favored over clearance either by disaggregation followed by proteolysis or aggregate autophagy (Wallace et al., 2015). Since disaggregation is a highly ATP demanding process, it is possible that under starvation conditions, autophagy may become the dominant pathway for aggregate clearance (Miller et al., 2015b).

Increasing evidence suggests that rather than being an uncontrolled cellular catastrophe, protein aggregation may be an additional branch of protein quality control (Chen et al., 2011). Aggregating misfolded proteins instead of degrading or attempting to fold them may relieve components of the protein quality control system. Further, if misfolded proteins can be sequestered, their potential toxicity caused by their ability to form inappropriate interactions with native or nascent proteins could be mitigated (Mogk et al., 2018). Indeed, yeast and mammals contain specific aggregate deposition sites. In yeast, the main sites for protein aggregation are called IPOD (insoluble protein deposit), INQ (intra nuclear quality control), and CytoQ (cytosolic quality control compartment) (Fig. 6A, (Kaganovich et al., 2008; Miller et al., 2015b)). The IPOD exists near the vacuole and contains mainly amyloidogenic proteins such as

the yeast prions (Kaganovich et al., 2008; Kumar et al., 2016). The INQ forms inside the nucleus, adjacent to the nucleolus and contains misfolded nuclear and cytosolic proteins (Miller et al., 2015a). The CytoQ describes the cytoplasmic aggregation site, which begins as many distinct aggregates upon stress and eventually coalesces into a few or even just one aggregate (Specht et al., 2011). Interestingly, the formation of the INQ and CytoQ depend on the proteins Btn2, and Hsp42, respectively (Malinovska et al., 2012; Specht et al., 2011). Btn2 contains a nuclear localization sequence and upon induction by heat, colocalizes with the INQ. Hsp42 is a constitutively expressed, cytosolic small heat shock protein (sHSP) (Haslbeck et al., 2004). sHSPs are ATP-independent chaperones that function as "holdases" in that they bind partially unfolded substrates, rendering them competent for refolding upon alleviation of the stress (Fig. 6B, (Jakob et al., 1993; Ungelenk et al., 2016)). sHSPs frequently exist as dimers but can form oligomers of 48 or greater, facilitating their ability to bind many misfolded proteins simultaneously (Basha et al., 2012). Because of their capacity to directly promote aggregation as well as fuse many small aggregates into larger aggregates, Hsp42 and other sHSPs have been called "aggregases" (Mogk et al., 2018). Mammalian cells contain a specialized aggregation site near their microtubule organizing center (MTOC) that forms upon stress called the aggresome (Johnston et al., 1998). Misfolded proteins are marked with K63 ubiquitin chains by the ubiquitin ligase, Parkin (Olzmann et al., 2007). Histone deacetylase 6 (HDAC6) simultaneously binds ubiquitin chains and the microtubule motor dynein, thereby allowing transport of the misfolded protein to the MTOC (Kawaguchi et al., 2003).

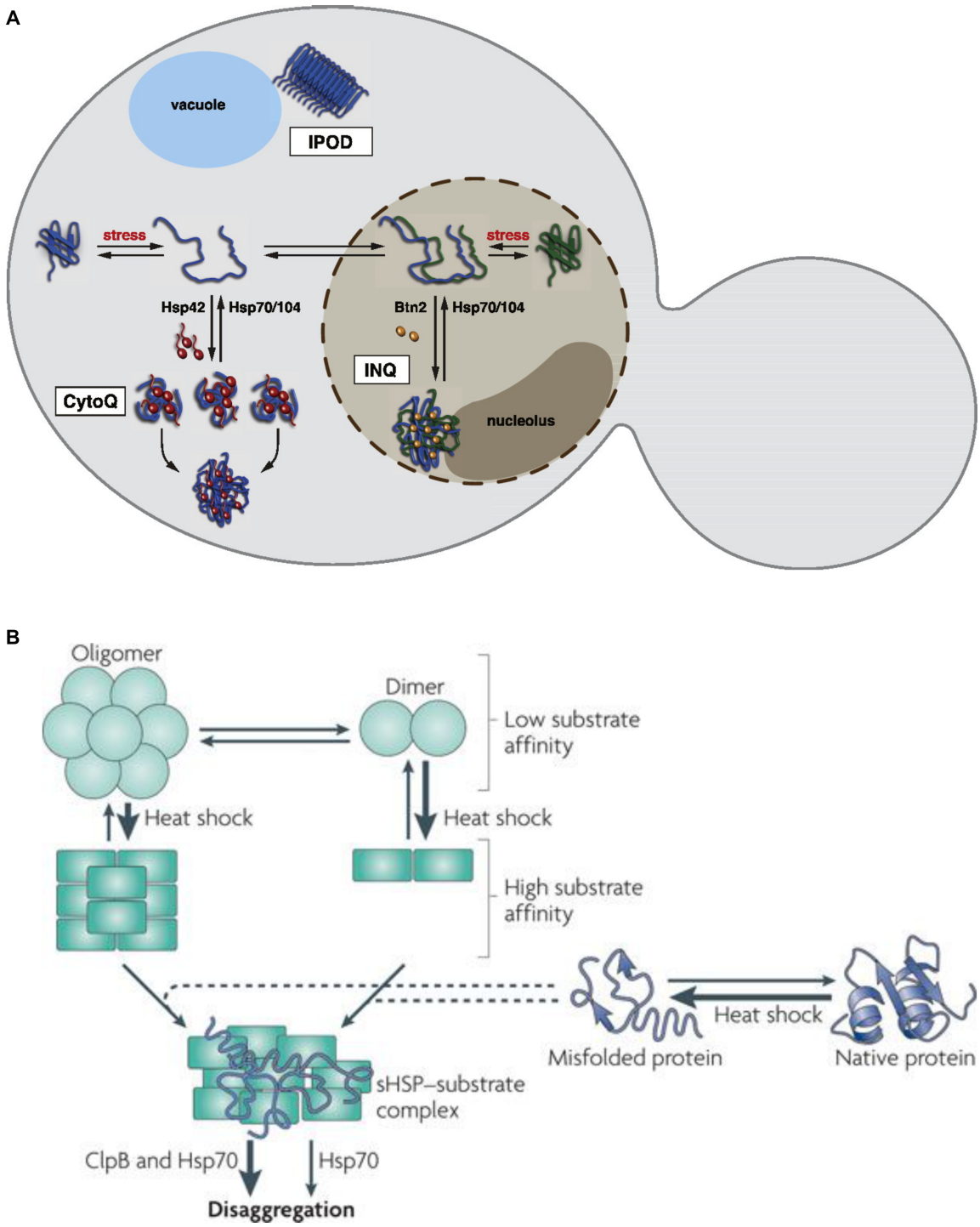

Figure 6. Small heat shock proteins (sHSPs) facilitate aggregation at deposition sites (adapted from (Miller et al., 2015a) (A) and (Tyedmers et al., 2010) (B))

(A) Yeast contain three identified aggregate deposition sites. The insoluble protein deposit (IPOD) resides near the vacuole which contains mainly amyloid proteins such as yeast prions.

The intra nuclear quality control (INQ) compartment contains misfolded proteins in the nucleus and requires Btn2 for its formation. The cytosolic quality control compartment (CytoQ) contains cytoplasmic misfolded proteins and requires the sHSP Hsp42. This compartmentalization has been proposed to help cells by allowing the concentration of dissagregation, refolding, and degradation factors at only a few sites of protein aggregation.

(B) sHSPs are ATP-independent molecular chaperones that hold misfolded proteins during stress. They can exist as dimers or high molecular weight oligomers. Some sHSPs undergo a conformational change at increased temperatures that make them become high affinity binders of misfolded proteins. sHSPs are thought to interact with proteins before they become fully unfolded, thereby keeping them in a folding competent conformation for refolding after dissagregation by Hsp104 or Hsp70.

The existence of such cellular mechanisms to promote aggregation argues in favor of the hypothesis that aggregation is cytoprotective. In addition to response to protein-folding stress, increasing evidence suggests that the toxic form of amyloid proteins are actually smaller, soluble oligomers, and that the insoluble amyloid form of the protein is in fact cytoprotective (Caughey and Lansbury, 2003). Furthermore, amyloid forming proteins were found to only be toxic when present in the cytoplasm, but were harmless when targeted to the nucleus (Woerner et al., 2016). In *C. elegans*, it has been demonstrated that an increase in protein aggregation mediated by sHSPs correlates with longevity, indicating that sHSP induced protein aggregation can aid in protein quality control within aged cells (Walther et al., 2015). The emerging picture is that, although protein misfolding is undesirable, controlled aggregation can be an effective mechanism to cope with the problem.

EFFECTS OF AN ANEUPLOID PROTEOME

Chromosome gain and loss cause protein aggregation

Aneuploidy presents a unique challenge to the protein quality control machinery. In the case of a gain of a single chromosome in a diploid cell, the cell must cope with a 1.5-fold increase in the amount of gene products for hundreds to thousands of otherwise normal proteins. This is a rather different scenario than contexts in which protein homeostasis has previously been studied, namely where heat or chemical stress causes misfolding of the majority of the proteome or expression of a single toxic protein that misfolds. Aneuploid cells must fold or degrade this excess protein or else they are at risk of containing large amounts of potentially harmful misfolded protein. Disomic yeast harbor twice as many protein aggregates compared to euploid cells when visualized by Hsp104 foci, regardless of which chromosome is amplified (Oromendia et al., 2012). In disomes, it appears that aggregation occurs as a result of the chaperone systems and the UPS becoming overwhelmed, as disomic cells have reduced capacity to fold a model Hsp90 substrate and are sensitive to chemical and genetic inhibition of chaperones and the UPS (Dodgson et al., 2016b; Oromendia et al., 2012; Torres et al., 2007). Likewise, mammalian cells trisomic or tetrasomic for single chromosomes are deficient in protein folding, increase autophagy in an attempt to cope with misfolded protein, and are sensitive to drugs that inhibit chaperones and autophagy (Donnelly et al., 2014; Tang et al., 2011).

Although it is clear that increasing the amount of protein in the cell by gaining a chromosome causes proteotoxic stress, it is less intuitive that having less protein caused by chromosome loss would have the same effect. Heterogeneous populations of aneuploid cells generated by random chromosome mis-segregation show increased protein aggregation and lysosomal stress caused by altered autophagic flux respectively (Oromendia et al., 2012;

Santaguida et al., 2015; Stingele et al., 2013; Tang et al., 2011). The most compelling piece of evidence that chromosome loss causes proteotoxic stress comes from the generation of yeast monosomic for specific chromosomes (Beach et al., 2017). Like disomic and trisomic yeast, monosomic yeast form protein aggregates immediately following chromosome mis-segregation. In this system, the degree of protein aggregation is well correlated with the number of protein coding genes on the aneuploid chromosome(s) for chromosome gain and loss (Beach, 2016). This suggests that the source of proteotoxic stress in aneuploid cells originates from an imbalance in the proteome, not simply excess protein. It has been proposed that this is caused by stoichiometric imbalance of protein complexes, where proteins that are expressed at levels higher than their binding partners due to aneuploidy are prone to misfold (Fig. 3, (Torres et al., 2008)). This idea will be explored in detail in Chapter 2.

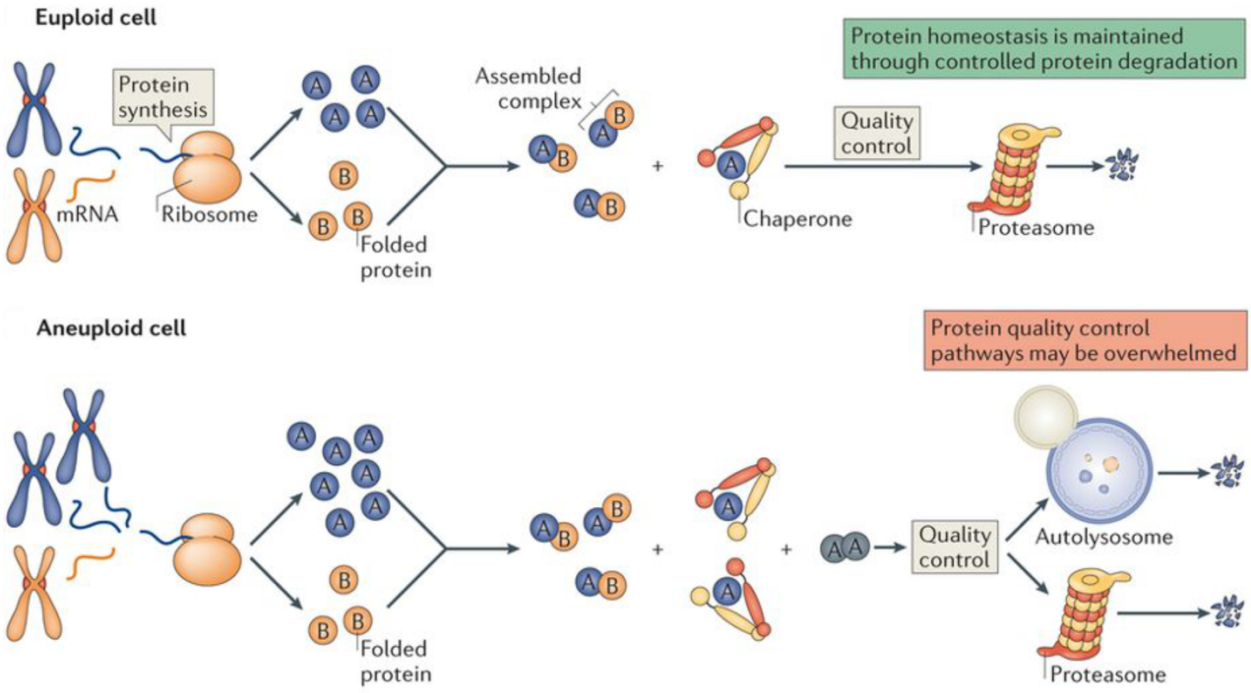

Figure 7. An imbalanced proteome causes proteotoxic stress (adapted from (Santaguida and Amon, 2015))

In euploid cells (top) synthesis of protein complex subunits is proportional to their stoichiometry in the complex. Any slight deviations in synthesis results in excess subunits that can be dealt with by cellular protein quality control (PQC) mechanisms. In aneuploid cells (bottom), proteins are synthesized in accordance with their genetic copy number, resulting in stoichiometric imbalance. Any protein complex with a subunit encoded on the aneuploid chromosome(s) then has excess subunits that need to be handled by PQC machinery. This has been proposed to overwhelm the PQC system, leading to the observed proteotoxic stress that is a universal consequence of aneuploidy.

Protein quality control defects underlie the fitness disadvantage of aneuploid cells

Although aneuploid cells exhibit multiple different stresses, evidence suggests that coping with an imbalanced proteome underlies the reduced proliferation that is common among aneuploid cells. First, as described above, aneuploid cells are exquisitely sensitive to additional perturbations to the protein quality control network. Second, in disomic yeast, the magnitude of each strains proliferation disadvantage is well correlated with the number of protein coding genes on that chromosome (Torres et al., 2007). Additionally, yeast artificial chromosomes that contain no yeast protein-coding genes do not slow proliferation. Finally, increasing the cells capacity to fold and/or degrade proteins is one of the few mechanisms shown to rescue the proliferation defects of multiple aneuploid lines with distinct karyotypes. In a screen of disomic yeast for mutations that confer growth advantages to aneuploid cells, a loss of function mutation in the deubiquitinase, *UBP6*, was one of just three genetic alterations to rescue the proliferation of more than one disomic strain (Torres et al., 2010). Ubp6 functions at the proteasome to recycle ubiquitin from degraded substrates, however this activity has the potential to allow

substrates to escape degradation effectively antagonizing the UPS (Hanna et al., 2006; Leggett et al., 2002). As such, deletion of UBP6 causes increased degradation of some proteins. In disomic cells harboring a deletion in *ubp6*, it was shown that many proteins encoded by the extra chromosome were more effectively degraded, thereby attenuating the imbalance caused by aneuploidy at the protein level (Torres et al., 2010). In human cells, increased expression of HSF1, the master transcription factor of the heat shock response, rescues sensitivity to autophagy inhibition by increasing expression of HSP90 and restoring the folding capacity of cells (Donnelly et al., 2014). This corresponds with an increase in the proliferative capacity for trisomic cells (Donnelly et al., 2014). Thus, proteotoxic stress can account for at least part of the proliferative defects that are characteristic of aneuploidy.

Control of gene dosage by protein degradation

Because imbalance in the proteome is at the heart of many deleterious aneuploid phenotypes, aneuploid cells would benefit by decreasing expression of proteins from excess chromosomes, or increasing expression of proteins on lost chromosomes. Some mechanism of dosage compensation would thus restore the balance of the proteome to a euploid state. As discussed above, most genes are expressed in proportion to their genetic copy number at both the RNA and protein level, indicating that widespread dosage compensation on autosomes does not exist. However, it was initially reported that most genes encoded in extra copy fail to accumulate more protein, because only a few proteins were examined and most of them were subunits of protein complexes (Torres et al., 2007). Subsequent analyses utilizing mass spectrometry (ms) demonstrated that it is more common for protein levels to reflect genetic copy number (Pavelka et al., 2010; Torres et al., 2010). Later work generated a comprehensive picture of the proteome for individual gains of most yeast chromosomes utilizing mass spec, and identified many

proteins that fail to accumulate in aneuploid cells despite having twice as much mRNA (Dephoure et al., 2014). These attenuated proteins are downregulated primarily by degradation because inhibition of proteasomal and autophagic degradation causes them to accumulate to levels predicted by their copy number. Further, disomic yeast do not utilize translational control as a dosage compensation strategy (Taggart and Li, 2018). These proteins that are degraded when encoded in excess are highly enriched for subunits of protein complexes, indicating that aneuploid cells utilize protein degradation to control stoichiometry of imbalanced complexes (Dephoure et al., 2014). The finding that aneuploid cells rely on protein degradation to perform dosage compensation is likely one reason why they are so sensitive to perturbations in the UPS.

Concluding remarks

Aneuploidy results in an imbalanced proteome that has numerous fitness penalties to both cells and organisms. These phenotypes occasionally derive from change in dosage of one or a few genes, but more commonly are caused by the simultaneous change in dosage of many genes. This large-scale change in the proteome of the cell results in proteotoxic stress, a universal feature of aneuploid cells that is at the heart of their proliferation defects. The molecular basis for aneuploid-associated proteotoxic stress is unknown, though it has been proposed that stoichiometric imbalance of protein complexes could account for the increased burden on the protein quality control system. Additionally, aneuploidy represents an ideal model to study how cells cope with stoichiometric imbalance. Since eukaryotes rarely employ operons to ensure that stoichiometric ratios of complex subunits are properly maintained, how eukaryotic cells deal with imbalances in their proteome is a fundamental question.

In this thesis, I have developed an assay to purify protein aggregates from aneuploid yeast and human cells to identify and quantify proteins within those aggregates using mass spectrometry. Uncovering the identity of proteins in aggregates revealed that stoichiometric imbalance of protein complexes causes protein aggregation in both aneuploid and euploid cells. Further, by combining this dataset with previously published data, I was able to track the fate of unassembled subunits of protein complexes genome-wide, revealing protein degradation and aggregation as mutually exclusive mechanisms to neutralize proteins that lack their binding partners. Remarkably, this work identified protein aggregation as a mechanism of dosage compensation, capable of eliminating excess protein by rendering it insoluble.

In light of recent findings that aggregation functions as an additional branch of protein quality control rather than a source of cellular toxicity, I propose that dosage compensation by protein aggregation may function as a cytoprotective mechanism in aneuploid and euploid cells. This is may be of particular importance to cancer cells. Cancer cells are frequently highly aneuploid, yet exhibit high proliferative potential relative to untransformed cells. How cancer cells avoid the negative fitness penalties caused by aneuploidy-induced stoichiometric imbalance could reveal therapeutic strategies. It will be important to study whether cancer cells employ aggregation or other dosage compensation mechanisms to maintain a balanced proteome.

References

Alberman, E.D., and Creasy, M.R. (1977). Frequency of chromosomal abnormalities in miscarriages and perinatal deaths. Journal of Medical Genetics *14*, 313–315.

Andriani, G.A., Almeida, V.P., Faggioli, F., Mauro, M., Tsai, W.L., Santambrogio, L., Maslov, A., Gadina, M., Campisi, J., Vijg, J., et al. (2016). Whole Chromosome Instability induces senescence and promotes SASP. Nature Publishing Group *6*, 35218.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230.

Baker, D.J., Jin, F., Jeganathan, K.B., and van Deursen, J.M. (2009). Whole Chromosome Instability Caused by Bub1 Insufficiency Drives Tumorigenesis through Tumor Suppressor Gene Loss of Heterozygosity. Cancer Cell *16*, 475–486.

Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2016). In vivo aspects of protein folding and quality control. Science *353*, aac4354.

Basha, E., O'Neill, H., and Vierling, E. (2012). Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. Trends Biochem. Sci. *37*, 106–117.

Beach, R.R. (2016). Insights into the consequences of chromosome gains and losses in *S. cerevisiae*. Massachusetts Institute of Technology.

Beach, R.R., Ricci-Tam, C., Brennan, C.M., Moomau, C.A., Hsu, P.-H., Hua, B., Silberman, R.E., Springer, M., and Amon, A. (2017). Aneuploidy Causes Non-genetic Individuality. Cell *169*, 229–242.e21.

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell *163*, 406–418.

Bjørkøy, G., Lamark, T., Brech, A., Outzen, H., Perander, M., Øvervatn, A., Stenmark, H., and Johansen, T. (2005). p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. J. Cell Biol. *171*, 603–614.

Blank, H.M., Sheltzer, J.M., Meehl, C.M., and Amon, A. (2015). Mitotic entry in the presence of DNA damage is a widespread property of aneuploidy in yeast. Mol. Biol. Cell *26*, 1440–1451.

Bonney, M.E., Moriya, H., and Amon, A. (2015). Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. Genes & Development *29*, 898–903.

Boveri, T. (1902). Uber mehrpolige Mitosen als Mittle zur Analyse des Zellkerns. Verhandl Phys-Med Ges Wulzburg NF *35*, 67–90.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren (Jena, Germany: Gustav Fischer Verlag).

Byrd, J.C., Mrózek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S., et al. (2002). Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood *100*, 4325–4336.

Caughey, B., and Lansbury, P.T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu. Rev. Neurosci. *26*, 267–298.

Chen, B., Retzlaff, M., Roos, T., and Frydman, J. (2011). Cellular strategies of protein quality control. Cold Spring Harb Perspect Biol *3*, a004374.

Cimini, D., Howell, B., Maddox, P., Khodjakov, A., Degrassi, F., and Salmon, E.D. (2001). Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. J. Cell Biol. *153*, 517–527.

Conde, J., and Fink, G.R. (1976). A mutant of Saccharomyces cerevisiae defective for nuclear fusion. Proc Natl Acad Sci USA *73*, 3651–3655.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948–962.

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Elife *3*, e03023.

Deshaies, R.J., and Joazeiro, C.A.P. (2009). RING domain E3 ubiquitin ligases. Annu. Rev. Biochem. *78*, 399–434.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics *169*, 1915–1925.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Dodgson, S.E., Kim, S., Costanzo, M., Baryshnikova, A., Morse, D.L., Kaiser, C.A., Boone, C., and Amon, A. (2016a). Chromosome-Specific and Global Effects of Aneuploidy in Saccharomyces cerevisiae. Genetics *202*, 1395–1409.

Dodgson, S.E., Santaguida, S., Kim, S., Sheltzer, J., and Amon, A. (2016b). The pleiotropic deubiquitinase Ubp3 confers aneuploidy tolerance. Genes & Development *30*, 2259–2271.

Donnelly, N., Passerini, V., Dürrbaum, M., Stingele, S., and Storchová, Z. (2014). HSF1 deficiency and impaired HSP90-dependent protein folding are hallmarks of aneuploid human cells. The EMBO Journal *33*, 2374–2387.

Douglas, N.R., Reissmann, S., Zhang, J., Chen, B., Jakana, J., Kumar, R., Chiu, W., and Frydman, J. (2011). Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. Cell *144*, 240–252.

Duncan, C.D.S., and Mata, J. (2011). Widespread cotranslational formation of protein complexes. PLoS Genet *7*, e1002398.

Ellis, R.J. (2001). Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. *26*, 597–604.

Elsasser, S., and Finley, D. (2005). Delivery of ubiquitinated substrates to protein-unfolding machines. Nat Cell Biol *7*, 742–749.

Emdin, S.O., Stenling, R., and Roos, G. (1987). Prognostic value of DNA content in colorectal carcinoma. A flow cytometric study with some methodologic aspects. Cancer *60*, 1282–1287.

Fändrich, M., and Dobson, C.M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. The EMBO Journal *21*, 5682–5690.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annu. Rev. Biochem. *78*, 477–513.

Finley, D., Ulrich, H.D., Sommer, T., and Kaiser, P. (2012). The ubiquitin-proteasome system of Saccharomyces cerevisiae. Genetics *192*, 319–360.

Fournier, R.E., and Ruddle, F.H. (1977). Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. Proc Natl Acad Sci USA *74*, 319–323.

Fredrickson, E.K., Rosenbaum, J.C., Locke, M.N., Milac, T.I., and Gardner, R.G. (2011). Exposed hydrophobicity is a key determinant of nuclear quality control degradation. Mol. Biol. Cell *22*, 2384–2395.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol. Biol. Cell *11*, 4241–4257.

Geiler-Samerotte, K.A., Dion, M.F., Budnik, B.A., Wang, S.M., Hartl, D.L., and Drummond, D.A. (2011). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc. Natl. Acad. Sci. U.S.a. *108*, 680–685.

Glover, J.R., and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell *94*, 73–82.

Goh, P.Y., and Kilmartin, J.V. (1993). NDC10: a gene involved in chromosome segregation in Saccharomyces cerevisiae. J. Cell Biol. *121*, 503–512.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 A resolution. Nature *386*, 463–471.

Hanks, S., Coleman, K., Reid, S., Plaja, A., Firth, H., FitzPatrick, D., Kidd, A., Méhes, K., Nash, R., Robin, N., et al. (2004). Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat. Genet. *36*, 1159–1161.

Hanna, J., Hathaway, N.A., Tone, Y., Crosas, B., Elsasser, S., Kirkpatrick, D.S., Leggett, D.S., Gygi, S.P., King, R.W., and Finley, D. (2006). Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. Cell *127*, 99–111.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324–332.

Hartwell, L.H., and Smith, D. (1985). Altered fidelity of mitotic chromosome transmission in cell cycle mutants of S. cerevisiae. Genetics *110*, 381–395.

Haslbeck, M., Braun, N., Stromer, T., Richter, B., Model, N., Weinkauf, S., and Buchner, J. (2004). Hsp42 is the general small heat shock protein in the cytosol of Saccharomyces cerevisiae. The EMBO Journal *23*, 638–649.

Haslberger, T., Weibezahn, J., Zahn, R., Lee, S., Tsai, F.T.F., Bukau, B., and Mogk, A. (2007). M domains couple the ClpB threading motor with the DnaK chaperone activity. Molecular Cell *25*, 247–260.

Haslberger, T., Zdanowicz, A., Brand, I., Kirstein, J., Turgay, K., Mogk, A., and Bukau, B. (2008). Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments. Nat. Struct. Mol. Biol. *15*, 641–650.

Hasle, H., Clemmensen, I.H., and Mikkelsen, M. (2000). Risks of leukaemia and solid tumours in individuals with Down's syndrome. The Lancet *355*, 165–169.

Hassold, T.J., and Jacobs, P.A. (1984). Trisomy in man. Annu. Rev. Genet. *18*, 69–97.

Heck, J.W., Cheung, S.K., and Hampton, R.Y. (2010). Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. Proc. Natl. Acad. Sci. U.S.a. *107*, 1106–1111.

Horwich, A.L., and Fenton, W.A. (2009). Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q. Rev. Biophys. *42*, 83–116.

Hwang, S., Gustafsson, H.T., O'Sullivan, C., Bisceglia, G., Huang, X., Klose, C., Schevchenko, A., Dickson, R.C., Cavaliere, P., Dephoure, N., et al. (2017). Serine-Dependent Sphingolipid Synthesis Is a Metabolic Liability of Aneuploid Cells. Cell Rep *21*, 3807–3818.

Jakob, U., Gaestel, M., Engel, K., and Buchner, J. (1993). Small heat shock proteins are molecular chaperones. J. Biol. Chem. *268*, 1517–1520.

Johnston, J.A., Ward, C.L., and Kopito, R.R. (1998). Aggresomes: a cellular response to misfolded proteins. J. Cell Biol. *143*, 1883–1898.

Kaganovich, D., Kopito, R., and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. Nature *454*, 1088–1095.

Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.-M., Dembele, D., Capponi, S., Timmers, H.T.M., et al. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. Nat Commun *10*, 1740.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nature Publishing Group *11*, 579–592.

Katz, W., Weinstein, B., and Solomon, F. (1990). Regulation of tubulin levels and microtubule assembly in Saccharomyces cerevisiae: consequences of altered tubulin gene copy number. Mol. Cell. Biol. *10*, 5286–5294.

Kawaguchi, Y., Kovacs, J.J., McLaurin, A., Vance, J.M., Ito, A., and Yao, T.P. (2003). The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. Cell *115*, 727–738.

Kiho, Y., and Rich, A. (1964). Induced enzyme formed on bacterial polyribosomes. Proc Natl Acad Sci USA *51*, 111–118.

Kisselev, A.F., Akopian, T.N., Woo, K.M., and Goldberg, A.L. (1999). The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. J. Biol. Chem. *274*, 3363–3371.

Knouse, K.A., Davoli, T., Elledge, S.J., and Amon, A. (2017). Aneuploidy in Cancer: Seq-ing Answers to Old Questions. Annu. Rev. Cancer Biol. *1*, 335–354.

Knouse, K.A., Wu, J., Whittaker, C.A., and Amon, A. (2014). Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. Proc. Natl. Acad. Sci. U.S.a. *111*, 13409–13414.

Knowles, T.P.J., Vendruscolo, M., and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. Nature Publishing Group *15*, 384–396.

Komander, D., and Rape, M. (2012). The ubiquitin code. Annu. Rev. Biochem. *81*, 203–229.

Kumar, R., Nawroth, P.P., and Tyedmers, J. (2016). Prion Aggregates Are Recruited to the Insoluble Protein Deposit (IPOD) via Myosin 2-Based Vesicular Transport. PLoS Genet *12*, e1006324.

Le Tallec, B., Barrault, M.-B., Courbeyrette, R., Guérois, R., Marsolier-Kergoat, M.-C., and Peyroche, A. (2007). 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. Molecular Cell *27*, 660–674.

Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H., and Finley, D. (2002). Multiple associated proteins regulate proteasome structure and function. Molecular Cell *10*, 495–507.

Levy, E.D., De, S., and Teichmann, S.A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc. Natl. Acad. Sci. U.S.a. *109*, 20461–20466.

Lu, K., Psakhye, I., and Jentsch, S. (2014). Autophagic clearance of polyQ proteins mediated by ubiquitin-Atg8 adaptors of the conserved CUET protein family. Cell *158*, 549–563.

Majumdar, A., Cesario, W.C., White-Grindley, E., Jiang, H., Ren, F., Khan, M.R., Li, L., Choi, E.M.-L., Kannan, K., Guo, F., et al. (2012). Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. Cell *148*, 515–529.

Malinovska, L., Kroschwald, S., Munder, M.C., Richter, D., and Alberti, S. (2012). Molecular chaperones and stress-inducible protein-sorting factors coordinate the spatiotemporal distribution of protein aggregates. Mol. Biol. Cell *23*, 3041–3056.

Mayer, M.P. (2010). Gymnastics of molecular chaperones. Molecular Cell *39*, 321–331.

Miller, S.B.M., Ho, C.-T., Winkler, J., Khokhrina, M., Neuner, A., Mohamed, M.Y.H., Guilbride, D.L., Richter, K., Lisby, M., Schiebel, E., et al. (2015a). Compartment-specific aggregases direct distinct nuclear and cytoplasmic aggregate deposition. The EMBO Journal *34*, 778–797.

Miller, S.B.M., Mogk, A., and Bukau, B. (2015b). Spatially Organized Aggregation of Misfolded Proteins as Cellular Stress Defense Strategy. Journal of Molecular Biology 1–11.

Mingle, L.A., Okuhama, N.N., Shi, J., Singer, R.H., Condeelis, J., and Liu, G. (2005). Localization of all seven messenger RNAs for the actin-polymerization nucleator Arp2/3 complex in the protrusions of fibroblasts. J. Cell. Sci. *118*, 2425–2433.

Mogk, A., Bukau, B., and Kampinga, H.H. (2018). Cellular Handling of Protein Aggregates by Disaggregation Machines. Molecular Cell *69*, 214–226.

Morrill, S.A., and Amon, A. (2019). Why haploinsufficiency persists. Proc Natl Acad Sci USA *10*, 201900437–6.

Muñoz, I.G., Yébenes, H., Zhou, M., Mesa, P., Serna, M., Park, A.Y., Bragado-Nilsson, E., Beloso, A., de Cárcer, G., Malumbres, M., et al. (2011). Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. Nature Publishing Group *18*, 14–19.

Musacchio, A., and Salmon, E.D. (2007). The spindle-assembly checkpoint in space and time. Nat. Rev. Mol. Cell Biol. *8*, 379–393.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annu. Rev. Genet. *43*, 525–558.

Nillegoda, N.B., Kirstein, J., Szlachcic, A., Berynskyy, M., Stank, A., Stengel, F., Arnsburg, K., Gao, X., Scior, A., Aebersold, R., et al. (2015). Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation. Nature 1–22.

Olzmann, J.A., Li, L., Chudaev, M.V., Chen, J., Perez, F.A., Palmiter, R.D., and Chin, L.-S. (2007). Parkin-mediated K63-linked polyubiquitination targets misfolded DJ-1 to aggresomes via binding to HDAC6. J. Cell Biol. *178*, 1025–1038.

Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. Genes & Development *26*, 2696–2708.

Pankiv, S., Clausen, T.H., Lamark, T., Brech, A., Bruun, J.-A., Outzen, H., Øvervatn, A., Bjørkøy, G., and Johansen, T. (2007). p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. J. Biol. Chem. *282*, 24131–24145.

Passerini, V., Ozeri-Galai, E., de Pagter, M.S., Donnelly, N., Schmalbrock, S., Kloosterman, W.P., Kerem, B., and Storchová, Z. (2016). The presence of extra chromosomes leads to genomic instability. Nat Commun *7*, 10754.

Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., and Li, R. (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature *468*, 321–325.

Pechmann, S., Levy, E.D., Tartaglia, G.G., and Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc. Natl. Acad. Sci. U.S.a. *106*, 10159–10164.

Pfau, S.J., Silberman, R.E., Knouse, K.A., and Amon, A. (2016). Aneuploidy impairs hematopoietic stem cell fitness and is selected against in regenerating tissues in vivo. Genes & Development *30*, 1395–1408.

Pinto, M., Morange, M., and Bensaude, O. (1991). Denaturation of proteins during heat shock. In vivo recovery of solubility and activity of reporter enzymes. J. Biol. Chem. *266*, 13941–13946.

Rasmussen, S.A., Wong, L.-Y.C., Yang, Q., May, K.M., and Friedman, J.M. (2003). Population-based analyses of mortality in trisomy 13 and trisomy 18. Pediatrics *111*, 777–784.

Reggiori, F., and Klionsky, D.J. (2013). Autophagic Processes in Yeast: Mechanism, Machinery and Regulation. Genetics *194*, 341–361.

Ried, T., Knutzen, R., Steinbeck, R., Blegen, H., Schröck, E., Heselmeyer, K., Manoir, du, S., and Auer, G. (1996). Comparative genomic hybridization reveals a specific pattern of

chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer *15*, 234–245.

Rosenstraus, M.J., and Chasin, L.A. (1978). Separation of linked markers in Chinese hamster cell hybrids: mitotic recombination is not involved. Genetics *90*, 735–760.

Ross-Innes, C.S., Becq, J., Warren, A., Cheetham, R.K., Northen, H., O'Donovan, M., Malhotra, S., di Pietro, M., Ivakhno, S., He, M., et al. (2015). Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. Nat. Genet. *47*, 1038–1046.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. The EMBO Journal *16*, 1501–1507.

Santaguida, S., and Amon, A. (2015). Short- and long-term effects of chromosome mis-segregation and aneuploidy. Nature Publishing Group *16*, 473–485.

Santaguida, S., Richardson, A., Iyer, D.R., M'Saad, O., Zasadil, L., Knouse, K.A., Wong, Y.L., Rhind, N., Desai, A., and Amon, A. (2017). Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. Dev. Cell *41*, 638–651.e5.

Santaguida, S., Vasile, E., White, E., and Amon, A. (2015). Aneuploidy-induced cellular stresses limit autophagic degradation. Genes & Development *29*, 2010–2021.

Sauer, R.T., and Baker, T.A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. Annu. Rev. Biochem. *80*, 587–612.

Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature *373*, 81–83.

Sheltzer, J.M., Blank, H.M., Pfau, S.J., Tange, Y., George, B.M., Humpton, T.J., Brito, I.L., Hiraoka, Y., Niwa, O., and Amon, A. (2011). Aneuploidy drives genomic instability in yeast. Science *333*, 1026–1030.

Sheltzer, J.M., Ko, J.H., Replogle, J.M., Habibe Burgos, N.C., Chung, E.S., Meehl, C.M., Sayles, N.M., Passerini, V., Storchová, Z., and Amon, A. (2017). Single-chromosome Gains Commonly Function as Tumor Suppressors. Cancer Cell *31*, 240–255.

Sheltzer, J.M., Torres, E.M., Dunham, M.J., and Amon, A. (2012). Transcriptional consequences of aneuploidy. Proc. Natl. Acad. Sci. U.S.a. *109*, 12644–12649.

Shiau, A.K., Harris, S.F., Southworth, D.R., and Agard, D.A. (2006). Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. Cell *127*, 329–340.

Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., and Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. Nature *561*, 268–272.

Shieh, Y.-W., Minguez, P., Bork, P., Auburger, J.J., Guilbride, D.L., Kramer, G., and Bukau, B. (2015). Operon structure and cotranslational subunit association direct protein assembly in bacteria. Science *350*, 678–680.

Siegel, J.J., and Amon, A. (2012). New insights into the troubles of aneuploidy. Annu. Rev. Cell Dev. Biol. *28*, 189–214.

Smardon, A.M., Tarsio, M., and Kane, P.M. (2002). The RAVE complex is essential for stable assembly of the yeast V-ATPase. J. Biol. Chem. *277*, 13831–13839.

Snape, K., Hanks, S., Ruark, E., Barros-Núñez, P., Elliott, A., Murray, A., Lane, A.H., Shannon, N., Callier, P., Chitayat, D., et al. (2011). Mutations in CEP57 cause mosaic variegated aneuploidy syndrome. Nature Publishing Group *43*, 527–529.

Specht, S., Miller, S.B.M., Mogk, A., and Bukau, B. (2011). Hsp42 is required for sequestration of protein aggregates into deposition sites in Saccharomyces cerevisiae. J. Cell Biol. *195*, 617–629.

Stathopulos, P.B., Rumfeldt, J.A.O., Scholz, G.A., Irani, R.A., Frey, H.E., Hallewell, R.A., Lepock, J.R., and Meiering, E.M. (2003). Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. Proc Natl Acad Sci USA *100*, 7021–7026.

Stingele, S., Stoehr, G., and Storchová, Z. (2013). Activation of autophagy in cells with abnormal karyotype. Autophagy *9*, 246–248.

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchová, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Molecular Systems Biology *8*, 608.

Sullivan, K.D., Lewis, H.C., Hill, A.A., Pandey, A., Jackson, L.P., Cabral, J.M., Smith, K.P., Liggett, L.A., Gomez, E.B., Galbraith, M.D., et al. (2016). Trisomy 21 consistently activates the interferon response. Elife *5*, 1709.

Taggart, J.C., and Li, G.-W. (2018). Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst *7*, 580–589.e584.

Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nature Publishing Group *11*, 515–528.

Tang, Y.-C., Williams, B.R., Siegel, J.J., and Amon, A. (2011). Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell *144*, 499–512.

Tang, Y.-C., Yuwen, H., Wang, K., Bruno, P.M., Bullock, K., Deik, A., Santaguida, S., Trakala, M., Pfau, S.J., Zhong, N., et al. (2017). Aneuploid Cell Survival Relies upon Sphingolipid Homeostasis. Cancer Res. *77*, 5272–5286.

Thompson, S.L., and Compton, D.A. (2011). Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors. Proc. Natl. Acad. Sci. U.S.a. *108*, 17974–17978.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. The EMBO Journal *19*, 94–102.

Torres, E.M., Dephoure, N., Panneerselvam, A., Tucker, C.M., Whittaker, C.A., Gygi, S.P., Dunham, M.J., and Amon, A. (2010). Identification of aneuploidy-tolerating mutations. Cell *143*, 71–83.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916–924.

Torres, E.M., Williams, B.R., and Amon, A. (2008). Aneuploidy: cells losing their balance. Genetics *179*, 737–746.

Tyedmers, J., Mogk, A., and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. Nature Publishing Group *11*, 777–788.

Ungelenk, S., Moayed, F., Ho, C.-T., Grousl, T., Scharf, A., Mashaghi, A., Tans, S., Mayer, M.P., Mogk, A., and Bukau, B. (2016). Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. Nat Commun *7*, 13673.

Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., and Deshaies, R.J. (2000). Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. Mol. Biol. Cell *11*, 3425–3439.

Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., Koonin, E.V., and Deshaies, R.J. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. Science *298*, 611–615.

Wallace, E.W.J., Kear-Scott, J.L., Pilipenko, E.V., Schwartz, M.H., Laskowski, P.R., Rojek, A.E., Katanski, C.D., Riback, J.A., Dion, M.F., Franks, A.M., et al. (2015). Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. Cell *162*, 1286–1298.

Walther, D.M., Kasturi, P., Zheng, M., Pinkert, S., Vecchi, G., Ciryam, P., Morimoto, R.I., Dobson, C.M., Vendruscolo, M., Mann, M., et al. (2015). Widespread Proteome Remodeling and Aggregation in Aging C. elegans. Cell *161*, 919–932.

Wang, L., Schubert, D., Sawaya, M.R., Eisenberg, D., and Riek, R. (2010). Multidimensional structure-activity relationship of a protein in its aggregated states. Angew. Chem. Int. Ed. Engl. *49*, 3904–3908.

Wells, J.N., Bergendahl, L.T., and Marsh, J.A. (2016). Operon Gene Order Is Optimized for Ordered Protein Complex Assembly. Cell Rep *14*, 679–685.

Williams, B.R., Prabhu, V.R., Hunter, K.E., Glazier, C.M., Whittaker, C.A., Housman, D.E., and Amon, A. (2008). Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science *322*, 703–709.

Williams, N.K., and Dichtl, B. (2018). Co-translational control of protein complex formation: a fundamental pathway of cellular organization? Biochem. Soc. Trans. *46*, 197–206.

Woerner, A.C., Frottin, F., Hornburg, D., Feng, L.R., Meissner, F., Patra, M., Tatzelt, J., Mann, M., Winklhofer, K.F., Hartl, F.U., et al. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. Science *351*, 173–176.

Xu, P., Duong, D.M., Seyfried, N.T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., and Peng, J. (2009). Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell *137*, 133–145.

Yang, Q., Rasmussen, S.A., and Friedman, J.M. (2002). Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. The Lancet *359*, 1019–1025.

Yao, T., and Cohen, R.E. (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. Nature *419*, 403–407.

Zipser, D., and Perrin, D. (1963). Complementation on Ribosomes. Cold Spring Harb Symp Quant Biol *28*, 533–537.

Chapter 2: Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells

Reproduced from Genes and Development:

Brennan, C.M., Vaites, L.P., Wells, J.N., Santaguida, S., Paulo, J.A., Storchová, Z., Harper, J.W., Marsh, J.A., and Amon, A. (2019). Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. Genes & Development.

The experiments in Figures 1C; 2B,C; 5; 8; 12A,B were performed by CMB, LPV, and JAP.

The experiments in Figures 6A,D were performed by SS, LPV, CMB, and JAP.

The experiments in Figure 11A were performed by JNW and CMB.

All other experiments and analyses were performed by CMB

ABSTRACT

Aneuploidy, a condition characterized by chromosome gains and losses, causes reduced fitness and numerous cellular stresses, including increased protein aggregation. Here, we identify protein complex stoichiometry imbalances as a major cause of protein aggregation in aneuploid cells. Subunits of protein complexes encoded on excess chromosomes aggregate in aneuploid cells, which is suppressed when expression of other subunits is coordinately altered. We further show that excess subunits are either degraded or aggregate and that protein aggregation is nearly as effective as protein degradation at lowering levels of excess proteins. Our study explains why proteotoxic stress is a universal feature of the aneuploid state and reveals protein aggregation as a form of dosage compensation to cope with disproportionate expression of protein complex subunits.

INTRODUCTION

Eukaryotes have a problem - subunits of protein complexes are not encoded in operons. Although eukaryotes have evolved to coordinate expression of subunits of the same complex (Li et al., 2014; Taggart and Li, 2018), changes in gene dosage of a subset of subunits of a protein complex, transient gene copy number imbalances during DNA replication, or fluctuations in gene expression can disrupt this coordinate expression, leading to the production of complex subunits that lack their binding partners. These orphan subunits have the potential to mis-fold and cause proteotoxic stress. How eukaryotic cells deal with stoichiometric imbalances is a fundamental, yet largely unexplored question.

Aneuploidy represents an especially dramatic case of gene dosage alteration because changes in autosome copy number generally lead to a corresponding change in the levels of RNAs and proteins produced by genes located on aneuploid chromosomes (Dephoure et al., 2014; Pavelka et al., 2010; Torres et al., 2010). Not surprisingly these dramatic alterations in cellular protein composition significantly impact cellular physiology, causing cell proliferation defects, metabolic alterations, and oxidative stress (reviewed in (Santaguida and Amon, 2015)). Protein homeostasis defects are especially prevalent in aneuploid cells. In budding yeast, many different aneuploidies harbor more protein aggregates, display decreased chaperone activity, and exhibit sensitivity to conditions that interfere with proteasomal degradation (Oromendia et al., 2012; Torres et al., 2007). In mammals, aneuploidy also disrupts protein homeostasis leading to altered autophagy, abnormal protein folding, and accumulation of protein aggregates (Donnelly and Storchová, 2014; Santaguida et al., 2015; Stingele et al., 2012; Tang et al., 2011). Why protein aggregation is so wide-spread in aneuploid cell lines was not understood.

We previously generated a series of haploid yeast strains each harboring an extra copy of one of the 16 yeast chromosomes (henceforth disomes; (Torres et al., 2007)). Analysis of the localization of Hsp104, a disaggregase that associates with protein aggregates (Liu et al., 2010), showed that gain of any of the 16 yeast chromosomes causes an increase in the number of cells harboring Hsp104 foci, providing evidence for increased protein aggregation (Oromendia et al., 2012). Here we determine the molecular basis of this phenotype. We find that subunits of protein complexes encoded on excess chromosomes aggregate in aneuploid cells. Our analyses also provide insights into how cells respond to protein stoichiometry imbalances. Protein aggregation can quantitatively deplete excess protein from the cytosol. We conclude that protein aggregation commonly performs functional dosage compensation.

RESULTS

Identification of proteins that aggregate in aneuploid yeast cells

Previous studies of proteotoxic stress in aneuploidy had shown that cells with defined chromosome gains or randomly generated, unknown karyotypes harbored increased levels of protein aggregates (Oromendia et al., 2012; Santaguida et al., 2015; Stingele et al., 2012). To gain insight into why protein aggregation is so common in aneuploid cells, we determined the composition of protein aggregates in disomic yeast strains. We isolated protein aggregates from disomic yeasts strains by a stringent differential centrifugation method (Koplin et al., 2010). Hsp104 was enriched in aggregate fractions (Fig. 1A), however we note that aggregates isolated in this manner may not contain all Hsp104 decorated aggregates, and also may contain aggregates not recognized by Hsp104. Analysis of protein aggregates by SDS-PAGE revealed

that protein aggregation is increased in aneuploid cells. Cells that mis-segregate chromosomes at a high frequency because they carry a temperature sensitive allele in the kinetochore component encoding gene *NDC10* harbored high levels of protein aggregates (Fig. 1B). Increased amounts of aggregated proteins were also observed in haploid cells disomic for chromosome V (Fig. 1B).

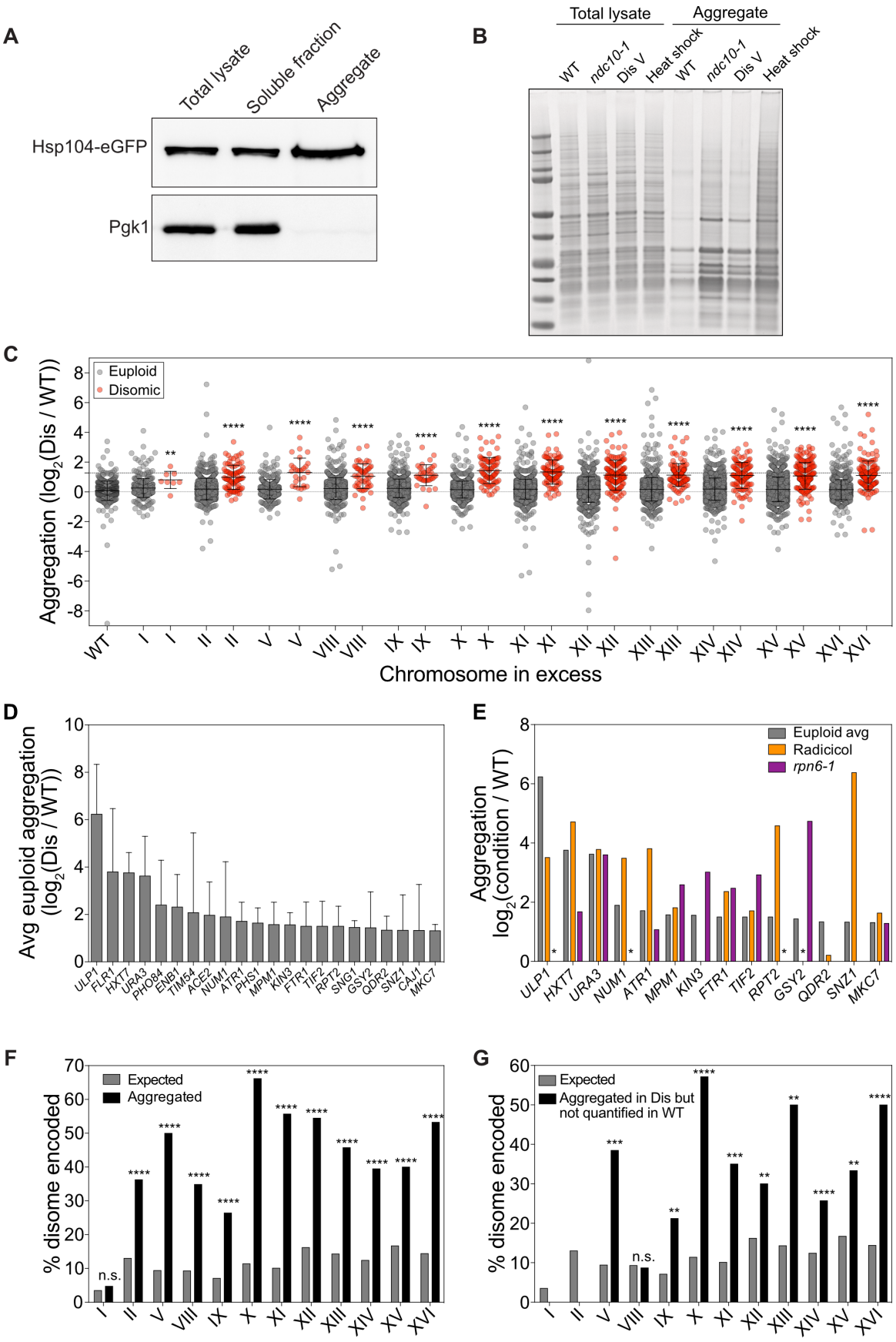
A
Total lysate
Soluble fraction
Aggregate
Hsp104-eGFP
Pgk1
B
Total lysate
Aggregate
WT
ndc10-1
Dis V
Heat shock
WT
ndc10-1
Dis V
Heat shock
C
Euploid
Disomic
Aggregation (log2(Dis / WT))
Chromosome in excess
D
Avg euploid aggregation (log2(Dis / WT))
ULP1 FLR1 HXT7 URA3 PHO84 ENB1 TIM54 ACE2 NUM1 ATR1 PHS1 MPM1 KIN3 FTR1 TIF2 RPT2 SNG1 GSY2 QDR2 SNZ1 CAJ1 MKC7
E
Euploid avg
Radicicol
rpn6-1
Aggregation log2(condition / WT)
ULP1 HXT7 URA3 NUM1 ATR1 MPM1 KIN3 FTR1 TIF2 RPT2 GSY2 QDR2 SNZ1 MKC7
F
Expected
Aggregated
% disome encoded
n.s.
G
Expected
Aggregated in Dis but not quantified in WT
% disome encoded
n.s.

Figure 1. Identification of proteins that aggregate in aneuploid yeast cells.

(A) Total lysate, aggregates and soluble fractions obtained from exponentially growing cells expressing Hsp104-eGFP (A31392) were analyzed for Hsp104 and Pgk1 abundance.

(B) Protein aggregates and total lysates were prepared from euploid cells (WT, A35797), *ndc10-1* (*ndc10-1*, A13413) cells grown at 30°C for 4 hours, disome V cells (Dis V, A28265), and euploid (A2587) cells after an 8 minute heat shock at 42°C (heat shock). Total lysates and aggregate fractions were subjected to SDS-PAGE and stained with coomassie.

(C) WT and disome cells were grown to exponential phase in SC medium containing heavy lysine and light lysine, respectively. Aggregated proteins are separated into two dot plots with red dots indicating proteins encoded on the duplicated chromosome and gray dots indicating proteins encoded on euploid chromosomes. The first column represents aggregates purified from a mixed sample of heavy lysine-labeled WT and light lysine-labeled WT. Lines represent mean and standard deviation. Upper dashed line at log2 1.27 shows the cut-off used to define aggregating proteins. ** $p < 0.01$, **** $p < 0.0001$; Mann-Whitney test.

(D) The average aggregate enrichment of proteins encoded on euploid chromosomes that were identified in aggregates of at least 3/12 disomes. Only proteins with an average enrichment of $\geq$log2 1.27 as measured in Fig. 1C are shown. Error bars indicate standard deviation.

(E) The enrichment of proteins from (D) was compared to their enrichment in aggregates purified from cells treated with radicicol (orange) or cells harboring the *rpn6-1* allele (purple) from (Fig. 5). * indicates proteins that were not quantified in either the radicicol or *rpn6-1* experiments because they did not pass the detection threshold in aggregates purified from the reference strain but were readily detected in aggregates isolated from radicicol treated or *rpn6-1* cells.

(F) The percentage of proteins encoded by the duplicated chromosome that were enriched at a level greater than the aggregation threshold of log2 1.27 (black bars) and the percentage of proteins encoded by the duplicated chromosome as a fraction of the whole proteome (gray bars). n.s. not significant, **** $p<0.0001$; cumulative distribution function (CDF) for a hypergeometric distribution.

(G) The percentage of proteins encoded by the duplicated chromosome that were not quantified by SILAC ms because the heavy-labeled (WT) peptides did not pass the detection threshold (black bars) and the percentage of proteins encoded by the duplicated chromosome as a fraction of the whole proteome (gray bars). n.s. not significant, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$; CDF for a hypergeometric distribution.

Abbreviations: WT, wild-type; Dis, disome; avg, average.

Having established that aneuploidy causes an increase in protein aggregates that can be isolated by differential centrifugation, we used SILAC mass spectrometry (ms) to identify proteins that preferentially aggregate in 12 different disomic yeast strains (Ong et al., 2002; Shevchenko et al., 2006) (Fig. 2A; Fig. 1C; Supplemental Data S1). Reproducibility was high between individual experiments: 70% of proteins were identified in repeats of individual experiments (Fig. 2B, C). Although biological replicates were well correlated, the mean of the SILAC ratios for all proteins combined in aggregates varied between replicates of the same disome (e.g. for disome II, the means were 0.59, 0.69, and 0.30). To account for this variability and to be able to conduct analyses on the aggregate data set as a whole, we mean centered all experiments such that the mean relative enrichment was equal across experiments (Fig. 1C). Each experiment was mean-centered to 0 by subtracting the mean of all SILAC ratios in that experiment from all data points.

To return the normalized values to a baseline that more closely resembles the increase in protein aggregation in disomic strains observed in the raw data, a constant (log2 0.27) was added to all normalized data points. This constant is the mean log2 ratio of all euploid encoded proteins in the dataset prior to normalization. Of note, we also identified proteins that were enriched in aggregates isolated from euploid strains compared to disome strains. However, in triplicate experiments for disome II, only 4 proteins (1.4%) were enriched more than two-fold in aggregates from euploid cells and their enrichment across replicate experiments was highly variable (Fig. 2D, E).

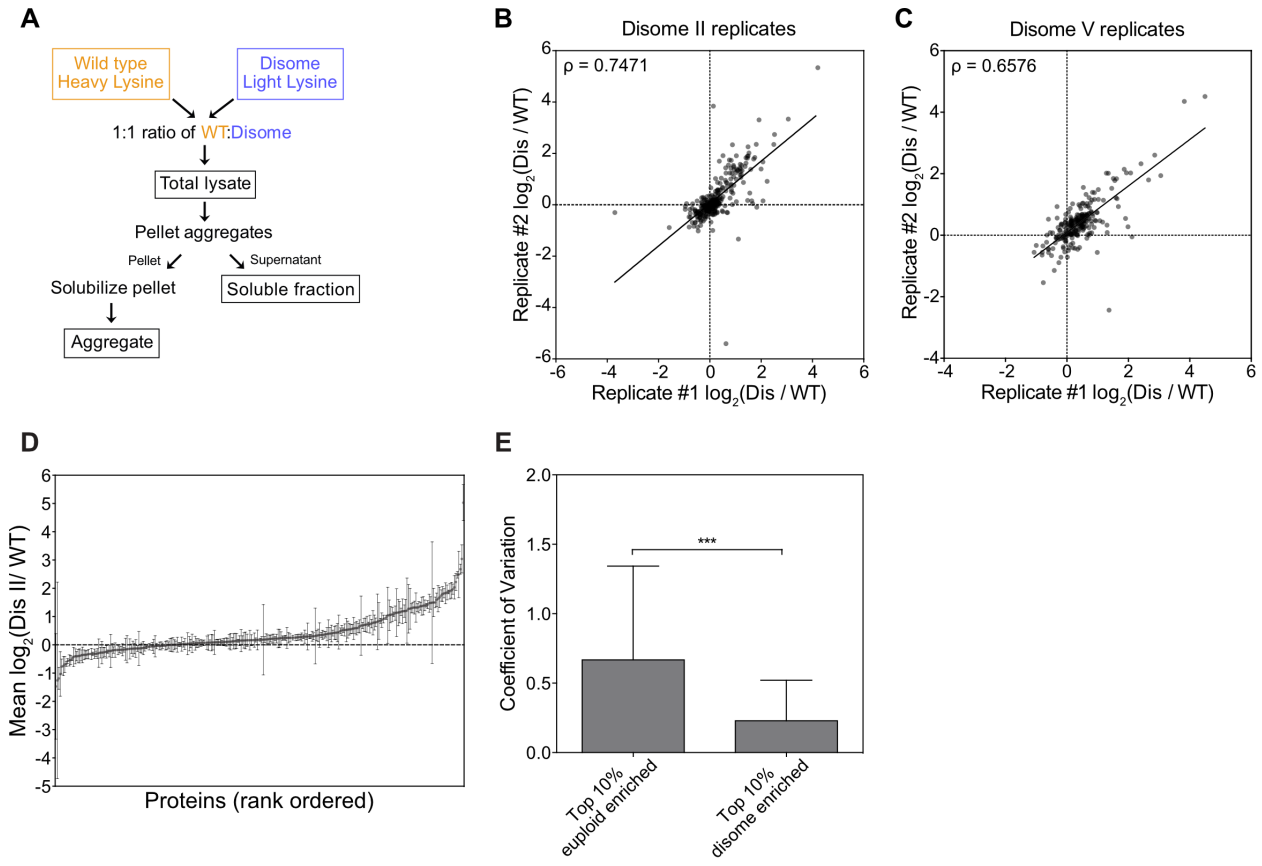

Figure 2. Characterization of aggregate purification method.

(A) Schematic for cell labeling and aggregate purification for SILAC ms experiments described in Fig. 1C and Fig. 12A, B.

(B, C) Biological replicate SILAC ms experiments were performed for disome II (B) and disome V (C). Proteins identified in both replicates were plotted as a function of their log2 enrichment in disome aggregates. Solid line represents a linear regression. Spearman correlation was 0.7471 for disome II replicates (B) and 0.6576 for disome V replicates (C).

(D) Three biological replicates of protein aggregate identification were performed for disome II cells. Proteins identified in all 3 experiments were plotted as a function of their mean enrichment in disome II aggregates. Bars represent standard deviation.

(E) The coefficient of variation for the top 10% of proteins enriched in aggregates from euploid cells and for the top 10% of proteins enriched in aggregates from disome II cells was calculated from data shown in (D). *** indicates $p < 0.0001$, Mann-Whitney test.

Abbreviations: WT, wild-type; Dis, Disome.

Which proteins aggregate in disomic yeast strains? The similar banding patterns of WT and aneuploid aggregates on SDS-PAGE gels (Fig. 1B) indicated that aggregates were composed of the same proteins, but that they aggregate more in aneuploid strains than euploid strains. Comparison of the banding pattern of protein aggregates on SDS-PAGE with the banding pattern of purified ribosomes further suggested that protein aggregates of both euploid and disomic yeast strains were enriched for ribosomes (Fig. 3A). To estimate the contribution of ribosomes to protein aggregates in disomic yeast strains we first determined the abundance of proteins in aggregates in each strain relative to its euploid reference by summing the raw total intensity of all heavy-labeled peptides and all light-labeled peptides and then calculating a ratio of the two

(Fig. 3B). Nine out twelve disomic strains contained more aggregated protein than euploid controls by this estimate. We then calculated the signal of each ribosomal protein as a percentage of the total signal for all aggregated proteins and determined that 75% of aggregated proteins were ribosomal proteins. Interestingly, the disomic strains with fewer ribosomes aggregating were the same strains that showed lower levels of total aggregate burden (compare Fig. 3B and C) confirming that ribosomes make up the majority of aggregating proteins in disomic yeast strains. Two lines of evidence indicate that it is assembled ribosomes rather than individual subunits that accumulate in aggregates. First, almost all excess ribosomal subunits are quantitatively degraded in disomic yeast strains (Dephoure et al., 2014). Second, the coomassie staining pattern of protein aggregates on SDS-PAGE resembles the pattern of purified intact ribosomes (Fig. 3A). We conclude that ribosomes are abundant in aggregates purified by our method in both euploid and aneuploid yeast strains, but that they aggregate more in disomes.

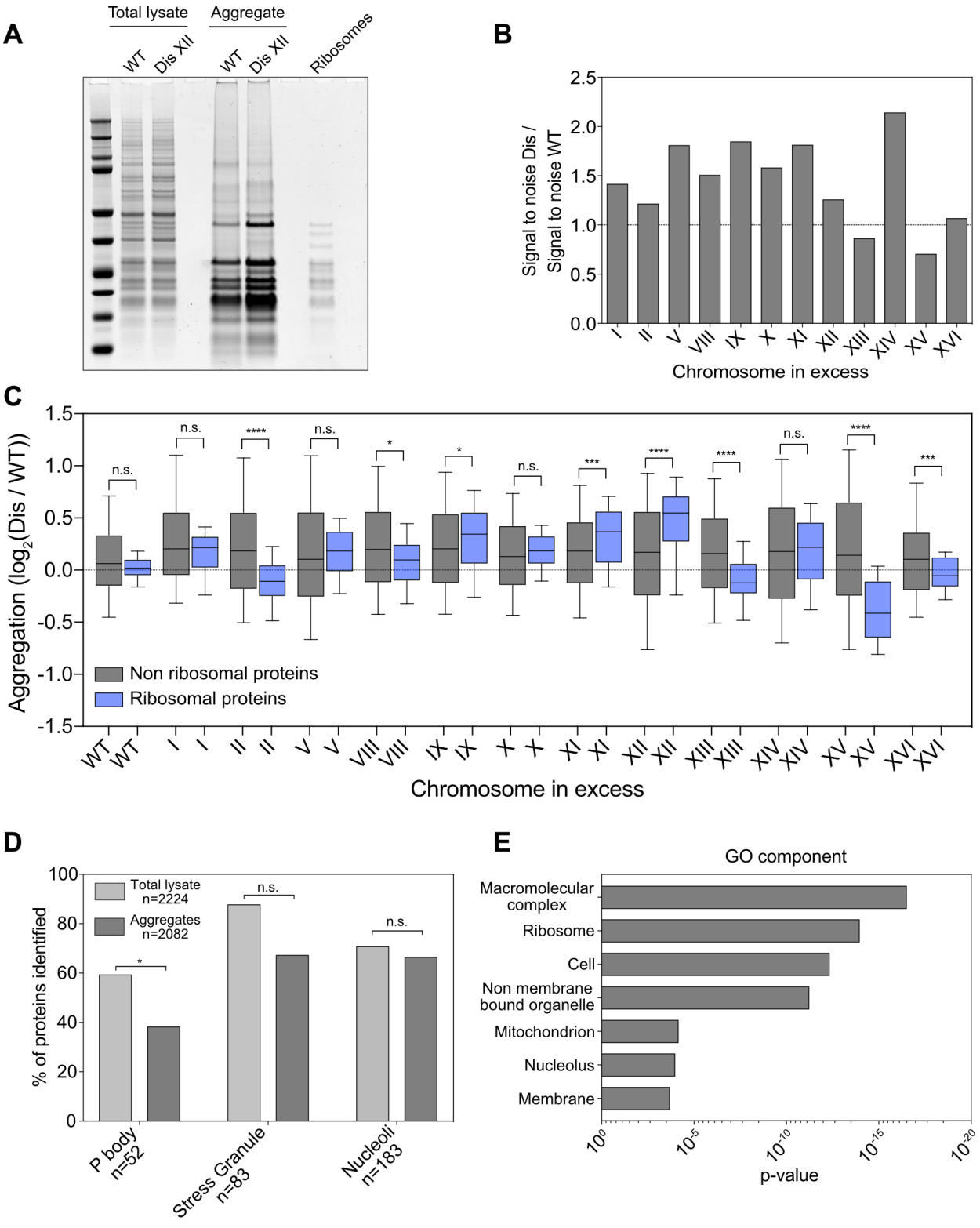

Figure 3. Ribosomal proteins are major constituents of purified protein aggregates.

(A) Total lysates and protein aggregates from euploid and disome XII cells were visualized by SDS-PAGE and coomassie staining. Purified ribosomes were included for comparison.
(B) In each SILAC ms experiment, the signal to noise ratio for all heavy-labeled (WT) peptides and light-labeled (disome) peptides were summed. The ratio of these sums was plotted to compare the amount of protein present in aggregates from each strain relative to the reference.
(C) Data were acquired as in Fig. 1C. Euploid chromosome -encoded aggregated proteins are separated into two box plots with gray boxes representing non ribosomal proteins and blue boxes representing ribosomal proteins. Error bars indicate the 10^{th} and 90^{th} percentiles. n.s. not significant, * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$; Mann-Whitney test
(D) The percentage of P body, stress granule, and nucleolar proteins identified in total lysates and protein aggregates was calculated. * indicates $p<0.05$, n.s. not significant; cumulative distribution function (CDF) for a binomial distribution.
(E) GO component analysis was performed on all proteins identified in protein aggregates. Proteins identified in total lysates served as the control data set. Corrected p-values for selected GO categories are plotted.
Abbreviations: WT, wild-type; Dis, Disome.

To determine which proteins other than ribosomes are found in aggregates purified by our differential centrifugation method, we assessed the presence of known phase-separated structures such as P bodies, stress granules, or nucleoli in purified aggregates (Banani et al., 2017) using the annotations by (Jain et al., 2016). P body proteins were significantly underrepresented in aggregates compared to total lysates (Fig. 3D). We also identified fewer stress granule proteins and nucleolar proteins in aggregates compared to total lysates, however this difference was not

significant (Fig. 3D). We conclude that components of phase-separated structures are not significantly enriched in our aggregate preparations.

Next, we conducted an unbiased Gene Ontology (GO) analysis (Fig. 3E) to determine which proteins were in fact enriched in aggregates isolated from aneuploid strains. Macromolecular complex was the GO term most significantly associated with proteins in aggregates, suggesting that complex subunits may be predisposed to aggregation. Ribosomes were the second most significant GO term, confirming our observation that the organelle is enriched in our aggregate preparation. The term "non-membrane bound organelle" (fourth most significant GO term) includes ribosomes which likely drives the significance of this GO term. The third most significant GO term was "cell", which we interpret to mean that all GO terms with p-values larger than this generic GO term, which includes the GO terms mitochondrion, nucleolus, and membrane, are not likely to be meaningful. This conclusion is supported by our observation that proteins known to be components of the nucleolus are not enriched in aggregates isolated from disomic yeast strains.

Next, we examined the physical and chemical properties of aggregating proteins. For this analysis, we compared aggregating proteins to the whole yeast proteome and to proteins identified in lysates from which the aggregates were purified. Proteins in both aggregates and total lysates had more disordered regions than the whole proteome (Fig. 4A). Hydrophobicity did not affect aggregation propensity (Fig. 4B). Amino acid composition appeared to affect protein aggregation in aneuploid yeast strains. Aggregated proteins were more basic and contained slightly more aromatic residues than proteins identified in lysates (Fig. 4C-E). Whether these

features influence aggregation propensity remains to be determined. Based on our knowledge of the effects of aneuploidy on cell physiology and the fact that “macromolecular complex” was the most significant GO term describing proteins that aggregate in disomic yeast strains, we can, however, envision two classes of proteins that aggregate in disomic yeast strains:

(1) Proteins that rely extensively on protein folding pathways to achieve their native conformation. Such proteins could aggregate because protein quality control is compromised in aneuploid yeast strains (Oromendia et al., 2012; Torres et al., 2010). Aggregation of these proteins ought to be independent of the identity of the aneuploid chromosomes and should occur in multiple different disomic yeast strains.

(2) Proteins encoded by the duplicated chromosome. Such proteins could aggregate because they are in excess.

We first focused on proteins in class 1.

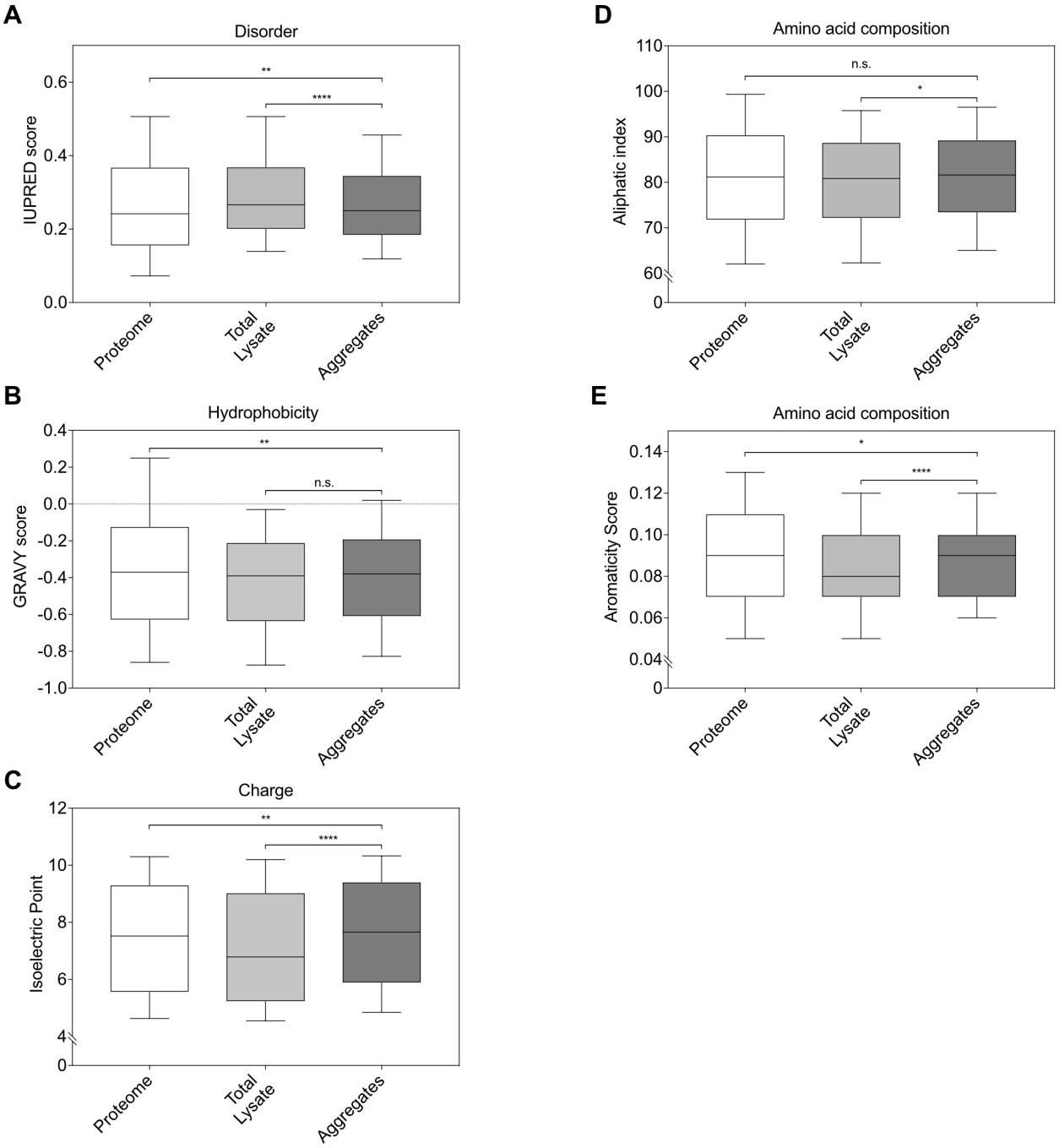

Figure 4. Properties of proteins in aggregates isolated from disomic yeast strains.

Proteins in the yeast proteome (proteome, white), proteins identified in total lysates (total lysate, light gray), and proteins identified in purified protein aggregates (aggregates, dark gray) from experiments shown in Fig. 1C were analyzed.

(A) Disorder predictions for protein sequences were calculated using IUPRED on the 'long' setting. For each sequence, per-residue disorder scores were averaged across the entire protein. (B-E) Grand average hydropathy (GRAVY) scores (B), isoelectric points (C), aliphatic indices (D), and aromaticity scores (E) were calculated using the YeastMine tool in the *Saccharomyces* Genome Database (SGD). **** indicates $p<0.0001$, ** $p<0.01$, * $p<0.05$, n.s. not significant, Mann-Whitney test.

Proteins that aggregate in multiple disomes also aggregate in cells with compromised protein quality control

Aneuploid cells experience proteotoxic stress (Oromendia et al., 2012; Santaguida et al., 2015; Stingele et al., 2012; Tang et al., 2011). Proteins that rely extensively on protein folding pathways to achieve their native conformation could thus aggregate in aneuploid cells more so than in euploid cells. To identify proteins that exhibit an increased aggregation due to the aneuploid state, we identified proteins that (1) were encoded on euploid chromosomes, (2) aggregated in at least 3 different disomic yeast strains, and (3) had an average enrichment of at least 2.4-fold [log2 1.27] in disomic aggregates. We arrived at this cutoff for enrichment of proteins in disome aggregates using a false discovery rate of 5% (see Materials and Methods). This analysis identified 22 proteins (Fig. 1D; Supplemental Data S2).

If proteins aggregate in multiple different disomic strains because proteostasis is compromised they should also aggregate in euploid cells in which protein quality control pathways are inhibited. To test this, we analyzed protein aggregates of cells harboring a temperature-sensitive mutation in the proteasome subunit encoding gene *RPN6* (Isono et al., 2005) and of cells treated

with the Hsp90 inhibitor radicicol (Fig. 5; Supplemental Data S2). For the 22 proteins identified as aggregating in multiple disomic strains, we obtained quantitative information for 14 in aggregates of either *rpn6-1* or radicicol-treated cells, or both. Of these, 13 (94%) were enriched 2.4-fold [log2 1.27] in aggregates of *rpn6-1* or radicicol-treated cells, or both (Fig. 1E, $p<0.001$, hypergeometric cumulative distribution function). We conclude that proteostasis deficiency causes a subset of proteins to aggregate in aneuploid cells regardless of whether gene copy number is altered or not.

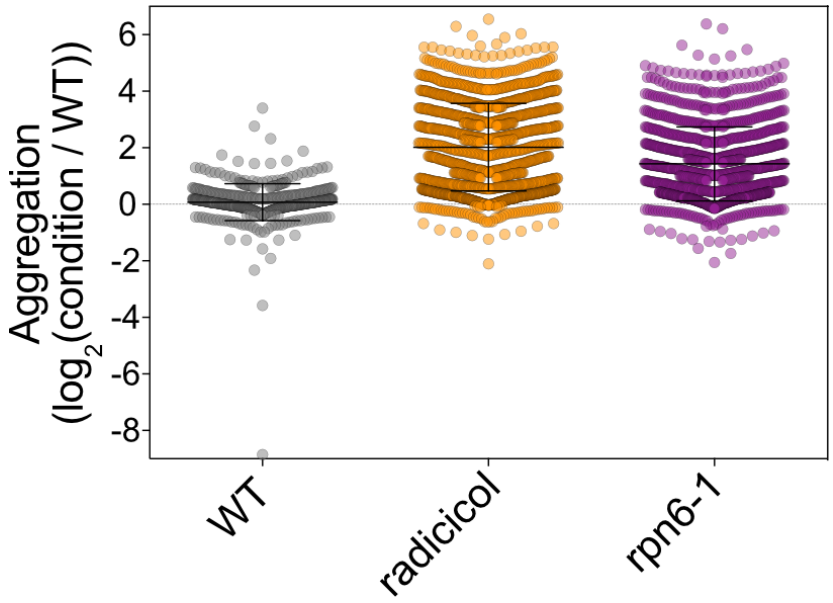

Figure 5. Aggregate analysis of cells with proteostasis defects.

SILAC ms was performed on protein aggregates purified from cells containing a temperature sensitive mutation in *RPN6* (*rpn6-1*; A38239), or cells harboring a deletion *PDR5* (A15548) and treated with 70 µM radicicol. For *rpn6-1* experiments, cells were grown in SC medium containing light lysine overnight, then diluted and grown to exponential phase at the permissive temperature. Cells were then shifted to the semi-permissive temperature (30°C) and grown for 4-6 hours. WT cells (A23504) grown in SC medium containing heavy lysine and cultured in the

same way as *rpn6-1* mutants served as the reference. For the radicicol experiment, *pdr5*Δ cells were grown to exponential phase at 30°C in SC medium containing either light or heavy lysine. 70μM radicicol was added to the light lysine-labeled culture and DMSO vehicle was added to the heavy lysine-labeled culture for 30 minutes before harvesting cells.

Abbreviations: WT, wild-type.

Duplicated proteins are highly enriched in aneuploid aggregates

We next examined proteins encoded on disomic chromosomes. These proteins were indeed significantly enriched in aggregates isolated from their respective disomic strain (Fig. 1C; red dots). To determine the number of proteins that aggregate because they are produced from two rather than one gene copy, we used a cutoff of 2.4-fold [log2 1.27] (based on a 5% FDR, described above) to define aggregate enrichment. Considering all of the disome strains, we identified 437 proteins that were enriched in aggregates in a disome-specific manner (Supplemental Data S1; Fig. 1F). For example, chromosome II encodes for 13% of the yeast proteome when duplicated, yet 36% of proteins that aggregate in strains disomic for chromosome II are encoded on chromosome II. Similarly dramatic results are observed in all disomic strains (Fig. 1F). We further note that this enrichment of disome-encoded proteins in aggregates underestimates proteins that aggregate because they are produced in excess. Proteins that fall below an established signal to noise ratio in either the heavy or light channel cannot be included because a SILAC ratio cannot be calculated. To mitigate this limitation, we examined proteins that did not pass the signal to noise threshold in one channel, but had a signal-to-noise ratio (S:N) of at least twice the threshold in the other channel (see Materials and Methods). This identified an additional 320 proteins that were quantified only in disome aggregates compared to just 72

that were only identified in aggregates of the euploid control strain (Supplemental Data S1). Importantly, 92 of the proteins quantified only in aggregates of disomic strains were encoded on the disomic chromosome, as opposed to just 2 that were quantified only in aggregates of euploid strains (Fig. 1G). We conclude that at least 529 proteins (437 proteins identified as enriched in disomes compared to wild-type aggregates + 92 identified only in disome aggregates) aggregate when their gene copy number is increased by two-fold.

Increasing gene copy number by one causes protein aggregation in human cells

Is protein aggregation a feature of aneuploidy that is conserved across eukaryotes? To address this question, we analyzed protein aggregation in near diploid human RPE-1 cells and two derivatives that were trisomic for chromosomes 12 or 21 (Stingele et al., 2012). As in yeast, proteins most enriched in aggregates were encoded on the trisomic chromosome. Within the 10^{th} percentile of proteins most enriched in trisomy 12 aggregates, 16.2% of proteins were encoded by chromosome 12 (Fig. 6A, B; Supplemental Data S3). This enrichment was highly significant (Fisher's exact test $p<0.0001$). In contrast, proteins encoded by other chromosomes had no significant enrichment, with the exception of proteins encoded by chromosome 4 which were slightly depleted (Fisher's exact test $p=0.0297$). We conclude that chromosome 12-encoded proteins are enriched in aggregates of trisomy 12 cells.

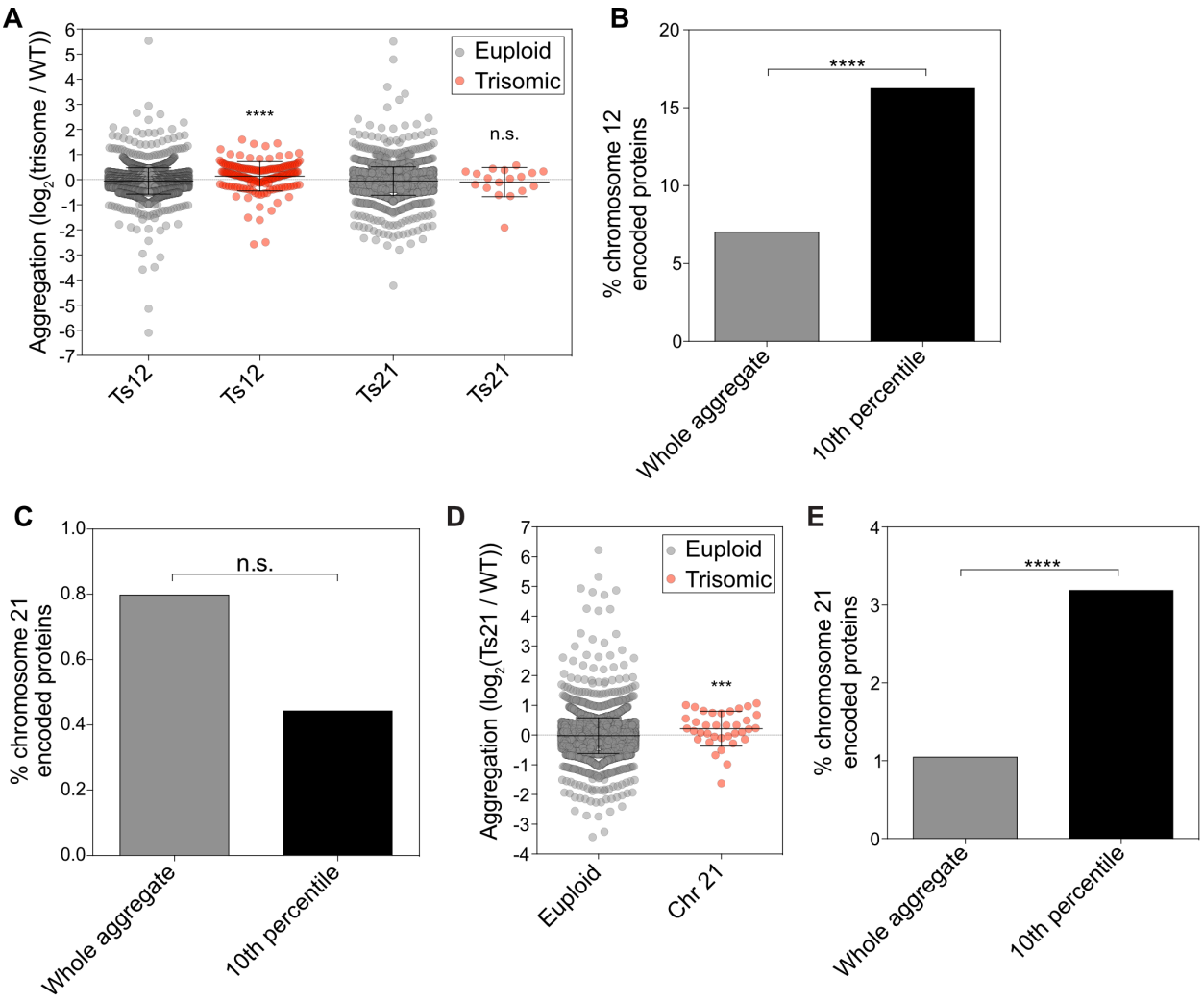

Figure 6. Aggregate analysis in trisomic human cells.

(A) RPE-1 cells and RPE-1 cells trisomic for either chromosome 12 or chromosome 21 were cultured in medium containing heavy lysine or light lysine, respectively, for 10 generations. Within each experiment, euploid encoded proteins (gray dots) were plotted separately from trisome encoded proteins (red dots). **** $p<0.0001$, n.s. not significant; Mann-Whitney test. (B, C) Enrichment of chromosome 12 (B) and 21 (C) encoded proteins in total aggregates (gray bars) and among the top 10% most highly enriched aggregated proteins (black bars). n.s. not significant, **** $p<0.0001$; cumulative distribution function (CDF) for a hypergeometric distribution.

(D) Euploid and trisomy 21 cells were treated with 25nM chloroquine and 1μM MG-132 and aggregates were plotted as in (A). *** $p<0.001$; Mann-Whitney test.

(E) Enrichment of chromosome 21 encoded proteins in total aggregates (gray bar) and among the top 10% most highly enriched aggregated proteins (black bar). **** $p<0.0001$; CDF for a hypergeometric distribution.

Abbreviations: WT, wild-type; Ts, trisome; Chr, chromosome.

Enrichment of chromosome 21-encoded proteins was not evident in aggregates purified from trisomy 21 cell lines (Fig. 6A, C; Supplemental Data S3) most likely due to the fact that chromosome 21 is the gene poorest chromosome in humans. To increase protein aggregation in trisomy 21 cells, we prevented protein degradation by inhibiting the proteasome and lysosomal degradation with MG-132 and chloroquine, respectively. In this experiment, we observed that within the 10^{th} percentile of proteins most enriched in trisomy 21 aggregates, 3.2% were encoded by chromosome 21. Overall, proteins encoded on chromosome 21 represent only 1% of proteins in aggregates (Fig. 6D, E; Supplemental Data S3). This enrichment was highly significant (Fisher's exact test $p=0.0005$). In contrast, proteins encoded on other chromosomes did not show significant enrichment. Among the chromosome 21 encoded proteins enriched in aggregates were two proteins known to contribute to disease phenotypes in Trisomy 21 - DOPEY2, a protein implicated in causing mental retardation (Rachidi et al., 2005) and APP, the precursor of the Alzheimer's disease-associated amyloid-β. We conclude that aggregation of proteins encoded on excess chromosomes is also a feature of aneuploidy in humans. We note that the enrichment of proteins encoded on excess chromosomes is less apparent in trisomic human cells than in

disomic yeast cells. This is to be expected. Gain of a copy of a chromosome in a diploid cell causes fewer protein imbalances than gain of a chromosome in a haploid cell.

Stoichiometric imbalance of protein complexes can cause protein aggregation

We next examined the mechanism whereby altering gene copy number causes protein aggregation in aneuploid cells. Protein complex analysis using the annotation by (Pu et al., 2009) showed that 44.2% of the 529 proteins, that aggregated in disomic yeast strains when encoded on an excess chromosome, were subunits of protein complexes (Fig. 7A). By comparison, only 29.2% of proteins encoded by euploid chromosomes that were enriched in aggregates, were subunits of protein complexes (Fig. 7A).

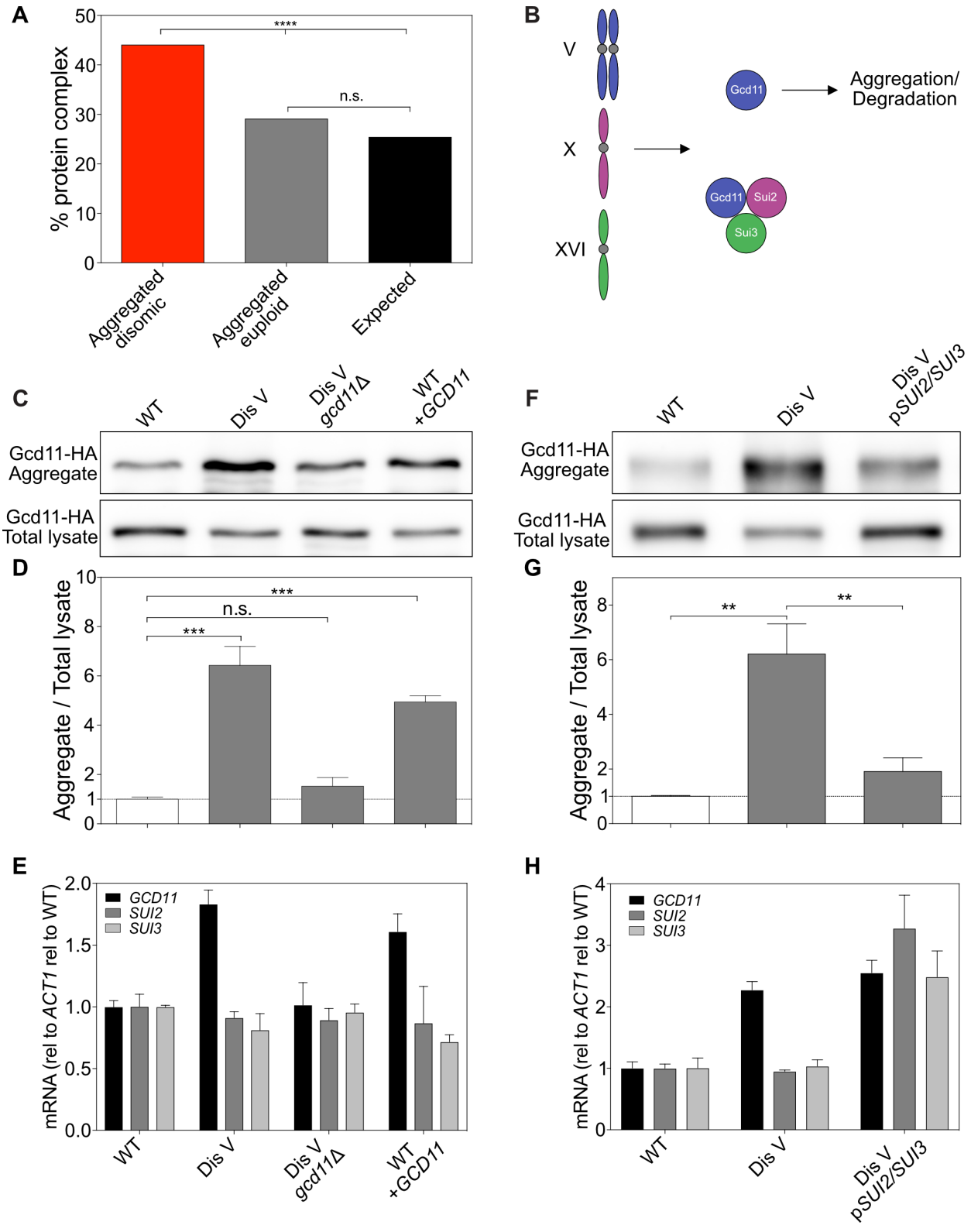

Figure 7. Stoichiometric imbalances of protein complex subunits cause protein aggregation.

(A) The percentage of proteins encoded by duplicated chromosomes enriched by log2 1.27 or more in aggregates (red), the percentage of proteins encoded by euploid chromosomes enriched by a log2 1.27 or more in aggregates (gray), and the percentage of proteins in the genome (black) that are annotated to form protein complexes by Pu et al. (2009) was calculated. **** indicates $p< 0.0001$, n.s. not significant; cumulative distribution function (CDF) for a binomial distribution.

(B) Diagram of eIF2 complex stoichiometry in disome V cells.

(C) Gcd11-HA in aggregates and total lysates in WT, (A40189), disome V (A40190), disome V GCD11-HA/gcd11Δ (A40191), and WT *URA3::GCD11-HA* (WT +*GCD11*; A40192) cells. Only one of the two *GCD11* genes in disome V cells was tagged with HA to ensure that protein levels are comparable between strains.

(D) Quantification of Western blots in (C) (n=3; SD; *** indicates $p<0.001$, T test)

(E) Cells were grown as in (C) and mRNA levels for eIF2 subunits was determined. Values were normalized to *ACT1* then to WT expression levels (n=3; SD).

(F-H) Western blot, relative aggregation quantification, and mRNA expression was determined for WT (A40193), disome V (A40194), and disome V p*SUI2/SUI3* (A40195) cells as described in (C-E; n=3; SD; ** indicates $p<0.01$, T test)

Abbreviations: WT, wild-type; Dis, disome; rel, relative.

We hypothesized that the enrichment of duplicated protein complex subunits in disome aggregates was due to protein complex subunits requiring binding to other subunits to acquire their native state. This hypothesis predicts that when gene copy number of the other complex subunits is altered in accordance with expression of the subunit produced in excess, aggregation

should be prevented. We tested this prediction by studying the eIF2 complex, which is required for translation initiation and composed of three subunits. The eIF2 γ-subunit, Gcd11, is encoded on chromosome V (Fig. 7B). Gcd11 was found in aggregates isolated from yeast strains disomic for this chromosome but returned to euploid levels in aggregates from disome V strains in which one copy of *GCD11* was deleted (Fig. 7C-E). In fact, a single extra copy of *GCD11* in an otherwise euploid strain was sufficient to cause Gcd11 aggregation (Fig. 7C-E). Thus, increased gene dosage of *GCD11* is necessary and sufficient to cause aggregation of the protein. It is worth noting that Gcd11 aggregated less when expressed in excess by itself than when overexpressed due to an extra copy of chromosome V. This observation suggests that proteotoxic stress caused by disomy of chromosome V exaggerates aggregation of Gcd11.

To test whether aggregation of Gcd11 in disome V strains can be prevented by doubling the gene copy number of the eIF2 α-subunit encoding gene *SUI2* and the eIF2 β-subunit encoding gene *SUI3*, we introduced a centromeric plasmid carrying both genes under their native promoters into disome V strains. Expression of *SUI2* and *SUI3* dramatically reduced Gcd11 aggregation in disome V cells (Fig. 7F-H). We note that expression of *SUI2* and *SUI3* also increased Gcd11 levels in lysates (Fig. 7F). A pool of Gcd11 is degraded when expressed in excess (Dephoure et al., 2014). Increased expression of Sui2 and Sui3 likely also protects Gcd11 from degradation. We conclude that stoichiometric imbalance of protein complexes can cause protein aggregation in aneuploid cells.

Excess proteins either aggregate or are degraded

Previous studies showed that many proteins that function in protein complexes are degraded when in excess in aneuploid cells, restoring their levels to euploid or near euploid levels (Dephoure et al., 2014; Gonçalves et al., 2017; Ori et al., 2016). We observed that many proteins aggregate when in excess. An important question regarding these two observations is whether excess subunits are down-regulated by both mechanisms or whether they are neutralized by one or the other. To distinguish between these possibilities, we analyzed protein degradation and aggregation propensity for all disome-encoded proteins identified in our protein aggregate analysis and the protein degradation analysis of Dephoure et al. (2014). Note that both data sets were created with the same strains. Analysis of the relative abundance of proteins encoded on disomic chromosomes confirmed that the two data sets were indeed similar to each other, despite slight differences in growth conditions (Fig. 8).

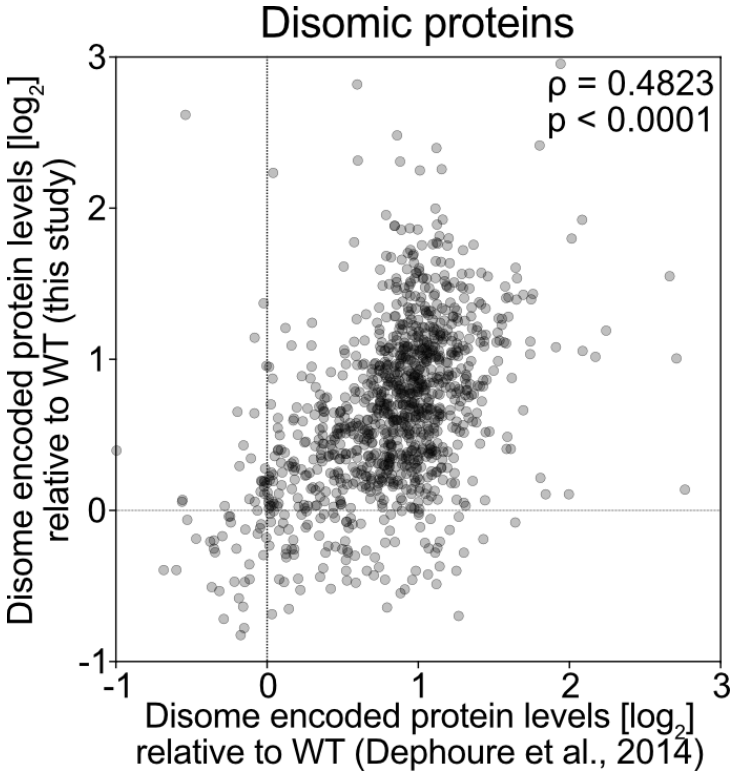

Figure 8. Comparison between our proteomic data set and that generated by Dephoure et al. (2014).

The strains used in the two studies were identical. In our analysis, proteins were isolated from cells grown in SC medium. Dephoure et al. (2014) grew cells in SC medium containing G418 and lacking histidine. Proteins were plotted as a function of their levels in strains where their corresponding genes were duplicated relative to a WT strain. The two data sets were correlated with a Spearman correlation of 0.4823 ($p < 0.0001$) (note: 20 data points fell outside the range of the axes but they were included in the correlation).

We first asked whether relative abundance of a protein in total lysates correlates with protein aggregation - or in other words, is a protein more likely to aggregate when its levels in total lysate are higher than in a euploid strain? We indeed observed a weak but significant correlation

(Fig. 9A). This observation indicates that proteins that are dosage compensated by protein degradation are less likely to be found in aggregates. Because excess cytoplasmic ribosomal subunits are almost exclusively degraded (Dephoure et al., 2014), we also asked whether the correlation was driven by ribosomal proteins. Removing cytoplasmic ribosomal proteins from the data set weakened the correlation between abundance of proteins in extracts and their propensity to aggregate, but it was still significant (Fig. 10A).

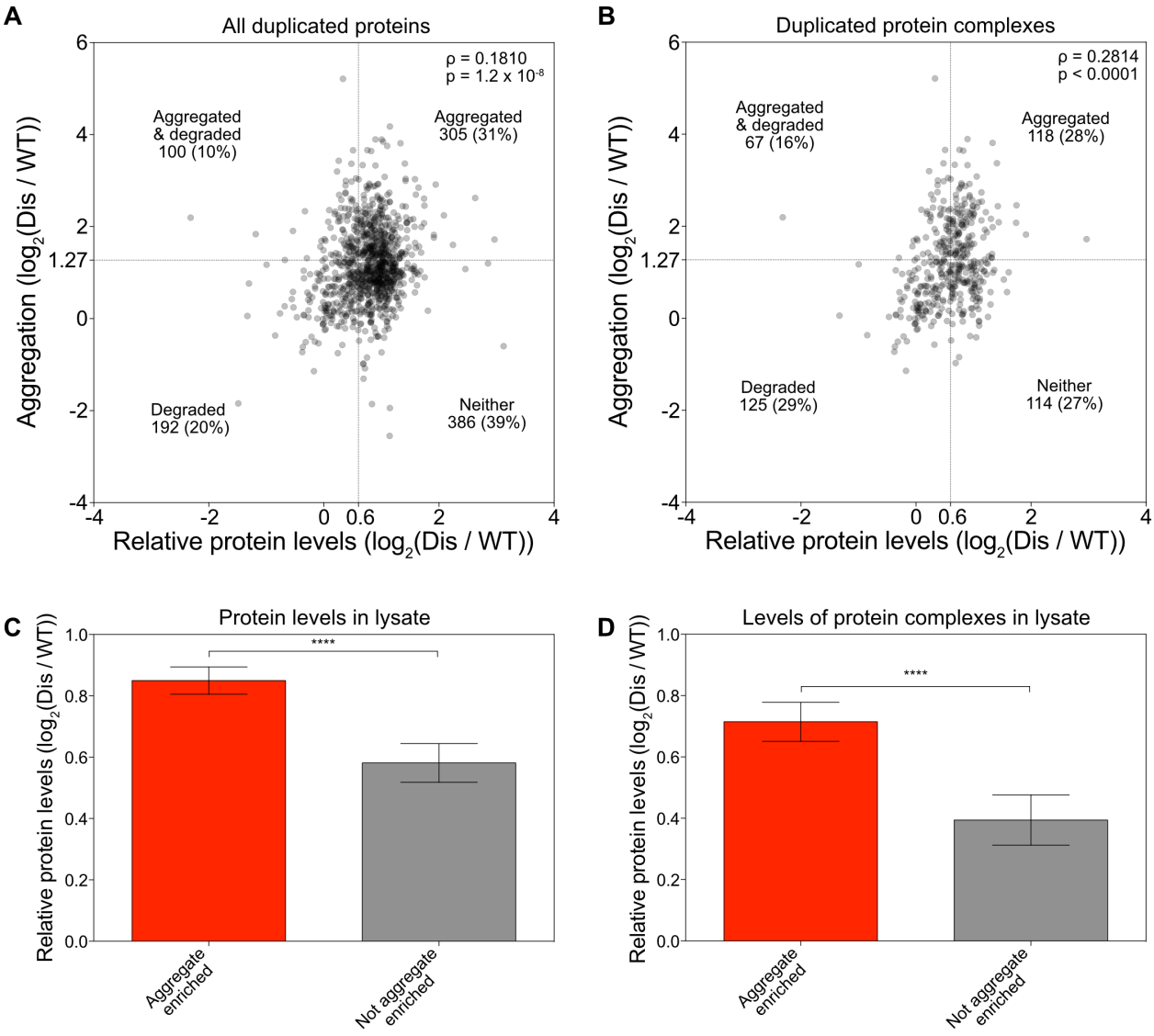

Figure 9. Excess proteins are either aggregated or degraded.

(A, B) Correlation between enrichment in protein aggregates (measured in Fig. 1C) and relative protein levels (measured in disome lysates by Dephoure et al. (2014)) was determined for all proteins encoded by the duplicated chromosome quantified in both data sets (A), and duplicated subunits of protein complexes (B). Spearman correlation of 0.1810 ($p=1.2x10^{-8}$) in (A) and 0.2814 ($p< 0.0001$) in (B). Dashed lines indicate thresholds for proteins being considered aggregated (y-axes) or degraded (x-axes). The number of proteins that fall into each quadrant is indicated.

(C, D) All duplicated proteins (C) and duplicated complex subunits (D) were separated into two bins: (1) aggregated proteins (red bars), which were defined as proteins with an enrichment of at least log2 1.27 in disomic aggregates; (2) non-aggregated proteins (gray bars) which were defined as proteins with an enrichment of log2 0.727 or less. Average relative levels in disome lysates as measured by Dephoure et al. (2014) are plotted. Error bars represent 95% confidence intervals. **** indicates $p< 0.0001$, Mann-Whitney test.

Abbreviations: WT, wild-type; Dis, disome

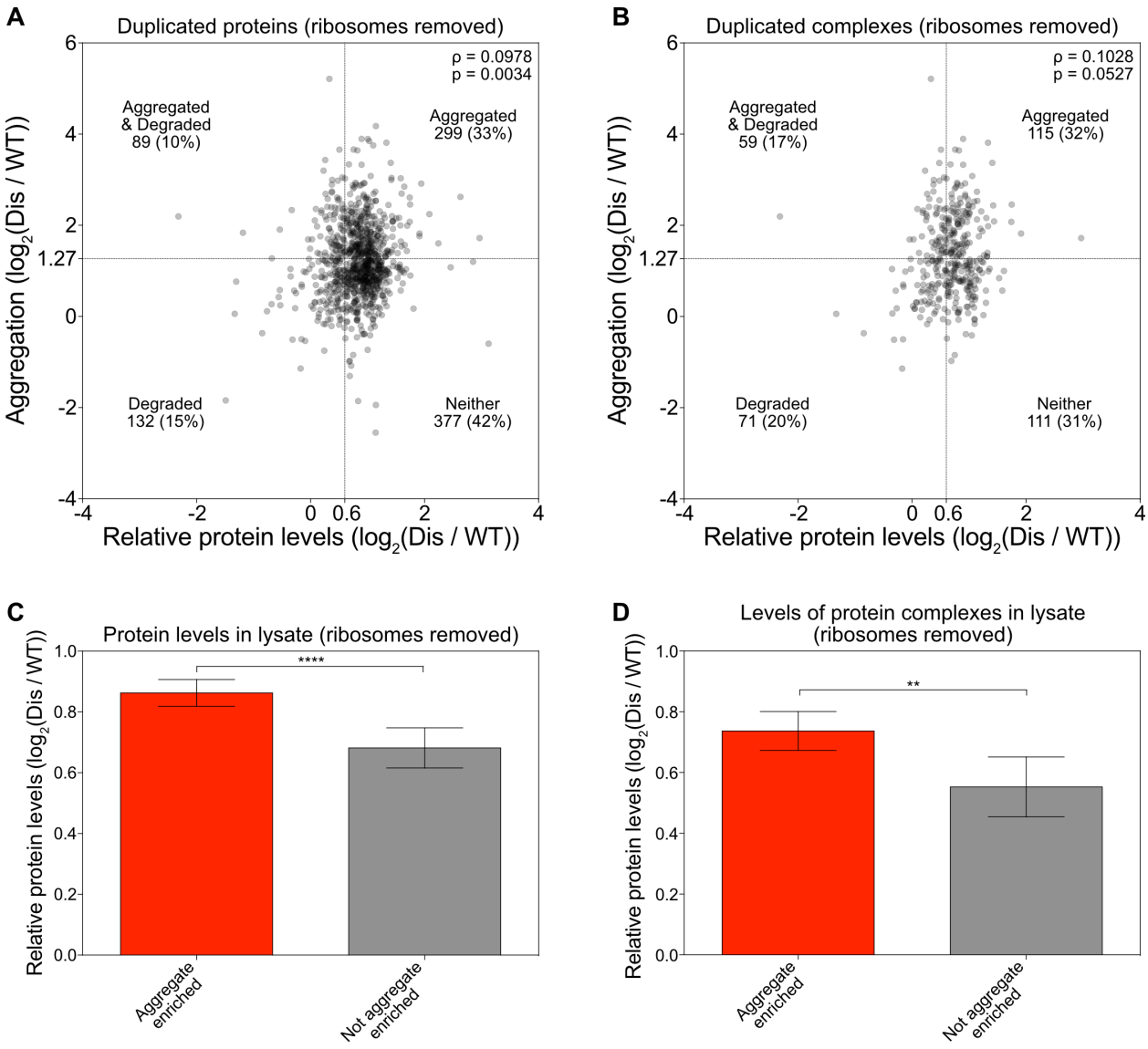

Figure 10. Excess proteins are either aggregated or degraded.

(A, B) Correlation between enrichment of a protein in protein aggregates (measured in Fig. 1C) and its relative abundance in total extracts as measured by Dephoure et al. (2014). This correlation analysis was performed for all proteins encoded by a duplicated chromosome quantified in both data sets (A), and duplicated subunits of protein complexes (B) after removing duplicated subunits of the cytoplasmic ribosome. All duplicated proteins were correlated with a Spearman correlation of 0.0978 (p=0.0034) (A) and duplicated complex subunits were correlated

with a Spearman correlation 0.1028 (p=0.0527) (B). Dashed lines indicate thresholds for proteins being considered aggregated (y-axes) or degraded (x-axes). The number of proteins that fall into each quadrant is indicated.

(C, D) After removing duplicated subunits of the cytoplasmic ribosome, proteins were separated into two bins: (1) aggregated proteins (red bars), which were defined as proteins with an enrichment of at least log2 1.27 in disomic aggregates; (2) non-aggregated proteins (gray bars) which were defined as proteins with an enrichment of log2 0.727 or less. Average relative levels in disome lysates as measured by Dephoure et al. (2014) are plotted. Error bars represent 95% confidence intervals. **** p< 0.0001, ** p< 0.01, Mann-Whitney test.

Abbreviations: WT, wild-type; Dis, disome.

To further assess whether protein aggregation and degradation are mutually exclusive we analyzed the data using cutoffs previously set to define a protein to aggregate or to be degraded when in excess. Dephoure et al. (2014) defined a disome-encoded protein as degraded when its abundance in extracts relative to a euploid reference was 1.52 [log2 0.6] instead of the expected 2 [log2 1]. We defined any disome-encoded protein aggregated when it was found enriched 2.4-fold [log2 1.27] in aggregates (Fig. 1). Based on these criteria, only 100/983 (10%) disomic proteins were considered both to aggregate and to be degraded when in excess. This is evident by the fact that the upper left quadrant in the graph in Fig. 9A is underpopulated (p= 0.0045; Fisher's exact test). This mutual exclusive behavior of disome encoded proteins was not driven by ribosomal proteins because we observed the same under population of the left upper quadrant when we removed ribosomal proteins (Fig. 10A). When we restricted our analysis to subunits of protein complexes only 67 out of 424 (16%) proteins were both enriched in disome aggregates

and degraded when in excess (p= 0.0012; Fisher's exact test; Fig. 9B), although we note that this observed lack of overlap between aggregation and degradation was, to a significant extent, driven by ribosomal proteins (Fig. 10B).

Another way to assess whether disome-encoded proteins aggregate or are degraded but not both, is to ask: Are disome-encoded proteins that are highly enriched in aggregates present in total lysates at the 2x level expected for disome-encoded proteins or are their levels lower? We found that disome-encoded proteins that are enriched in aggregates, are present at levels close to the expected level of two-fold indicating that they are not degraded (Fig. 9C). In contrast, levels of proteins that were detected in aggregates but not enriched in aggregates were lower in total lysates (Fig. 9C). Removal of ribosomal subunits from the analysis did not alter this conclusion (Fig. 10C). The results were even more striking when we focused our analysis on disome-encoded proteins that are part of protein complexes (Fig. 9D, 10D). We conclude that proteins, when present in excess, have a preferred fate of either degradation or aggregation.

Protein complex subunits that aggregate when in excess have lower turnover rates than degraded subunits

What determines whether excess protein complex subunits aggregate or are degraded? Computational analyses revealed that both classes of subunits harbor large protein-protein interfaces within their complexes compared to proteins that neither aggregate nor are degraded (Fig. 11A; Supplemental Data S4; (Marsh et al., 2013)). This is consistent with earlier work in mammalian cells, demonstrating that protein complex subunits with larger interfaces tend to be unstable and rapidly degraded when not bound to their partners (McShane et al., 2016). If

overexpressed relative to other components of the complex, these large interfaces could facilitate inappropriate protein-protein interactions, leading to aggregation instead of degradation (Levy et al., 2012).

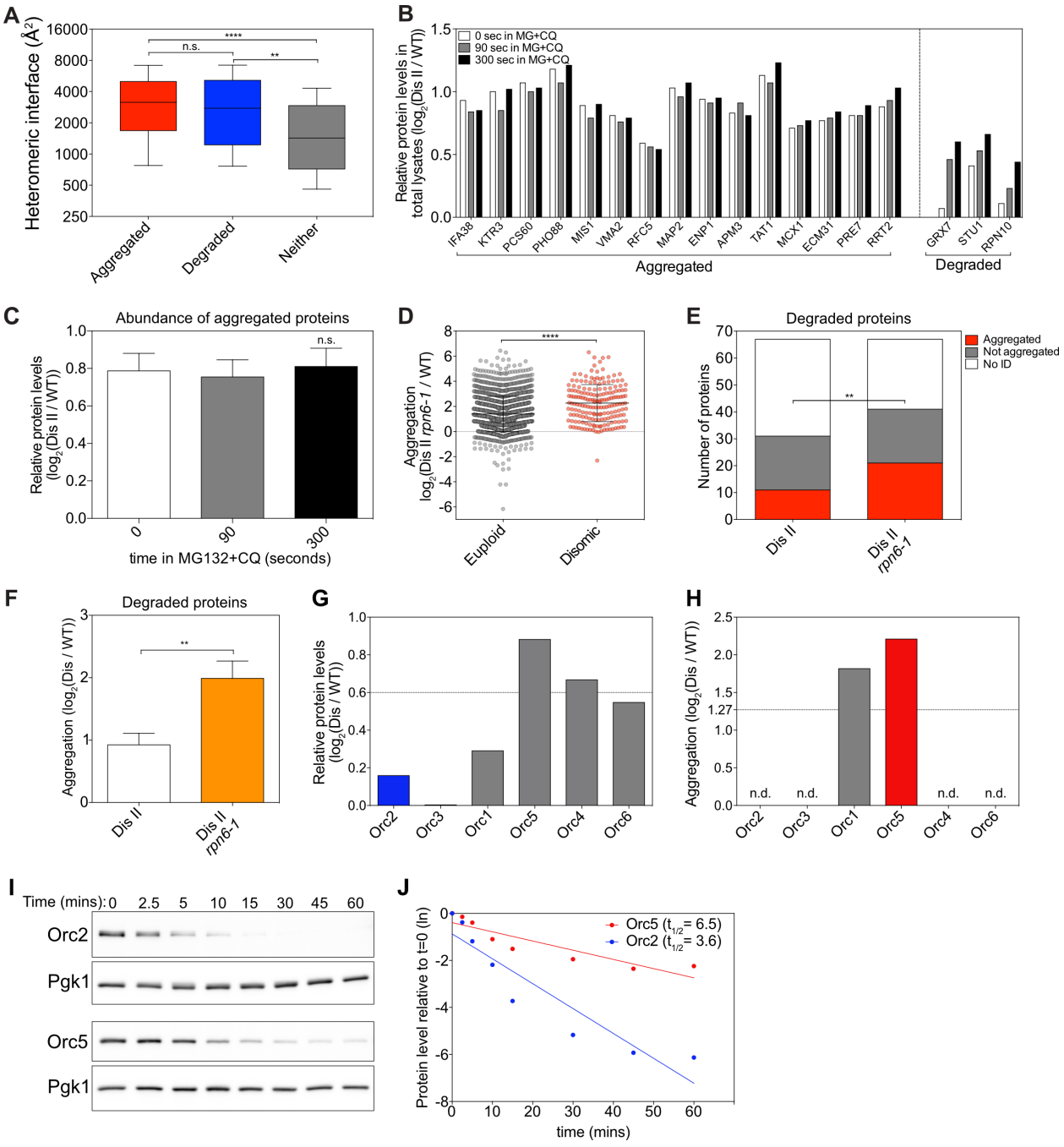

Figure 11. Protein half-life determines whether a protein aggregates or is degraded.

(A) Proteins encoded on duplicated chromosomes were separated into three categories: aggregated, degraded, and neither. Aggregated proteins (red), degraded proteins (blue) were defined as in Fig. 9A. Proteins that were identified in our analysis and that by Dephoure et al. (2014) but did not pass the threshold for aggregation or degradation were considered neither aggregated nor degraded (gray). The heteromeric interface sizes of the proteins in each category are plotted as box plots with whiskers representing the 10^{th}-90^{th} percentile. **** $p < 0.0001$, ** $p < 0.01$, n.s. not significant; Mann-Whitney test.

(B) The change in levels of proteins encoded on chromosome II in disome II in total lysates of cells treated with 100 μM MG-132 and 10 mM chloroquine relative to WT (data from Dephoure et al. (2014)). Examples of aggregating proteins as determined in Fig. 1C and of degraded proteins as determined by Dephoure et al. (2014)) are shown. White bars indicate relative levels immediately before the addition of MG-132 and chloroquine, gray bars and black bars relative levels 90 seconds and 300 seconds thereafter, respectively.

(C) Mean levels of all aggregating proteins as measured in (B) at each time point (error bars SEM; n.s. not significant; Wilcoxon matched-pairs signed rank test).

(D) Disome II *rpn6-1* (A40196) or WT (A23504) cells were grown to exponential phase at 30°C in SC medium containing light lysine and heavy lysine respectively and aggregating proteins were identified. Lines indicate mean and standard deviation. **** $p < 0.0001$, Mann-Whitney test.

(E) Proteins considered degraded when duplicated by Dephoure et al. (2014) were examined in aggregates purified from disome II cells (Fig. 1C) and from disome II *rpn6*-1 cells (D). A protein was considered to aggregate when it was enriched by more than log2 1.27 in aggregates (red), not aggregated as enriched by less than log2 1.27 (gray). no ID indicates proteins that were not

identified in aggregates (white). ** $p < 0.01$, cumulative distribution function for a binomial distribution.

(F) Degree of aggregation was determined for all proteins in (E) in disome II aggregates and disome II *rpn6-1* aggregates. Bars represent SD; ** $p<0.01$, Mann-Whitney test.

(G, H) Relative protein levels as determined by Dephoure et al. (2014) (G) and relative aggregation as measured in Fig. 1C (H) for ORC subunits when encoded by disomic chromosomes (n.d. - not detected in aggregates).

(I, J) Cells were grown to exponential phase at 30°C in YEP medium containing 2% raffinose. Expression of HA tagged *ORC2* (A40197) and *ORC5* (A40198) was induced with 2% galactose for 20 minutes. Then protein synthesis was halted by the addition of 2% glucose and 0.5 mg/mL cycloheximide (t=0). Protein levels were determined. Pgk1 was used as a loading control (I). Protein levels were quantified relative to the loading control and normalized to the 0 minute time point (J).

Abbreviations: WT, wild-type; Dis, disome; MG, MG-132; CQ, chloroquine; sec, second; ln, natural logarithm; ID, identification; mins, minutes.

Given that excess protein complex subunits that are degraded and that aggregate both harbor large protein binding interfaces, we hypothesized that their fate could be determined by differences in recognition by the ubiquitin/proteasome or lysosomal degradation systems. Previous work showed that degradation of proteins that are encoded on a disomic chromosome, can be prevented by treating disomic cells with the proteasome inhibitor MG-132 and the lysosomal-degradation inhibitor chloroquine ((Dephoure et al., 2014), Fig. 11B; right part of the graph). Importantly, we find that preventing protein degradation did not affect overall levels of

individual proteins that aggregate (Fig. 11B, left part of the graph) and average levels of all proteins that aggregate (Fig. 11C; Supplemental Data S4). Therefore, proteins found in aggregates generally have low turnover rates.

To further assess whether protein stability determines whether or not a protein aggregates, we analyzed the effects of proteasome inhibition on the fate of proteins that are normally degraded when in excess. Partial inhibition of the proteasome using a temperature sensitive *rpn6-1* mutant grown at the semi-permissive temperature (30°C) caused a general increase in protein aggregation for all proteins in disome II cells (Fig. 11D; Supplemental Data S4). 67 disomic proteins are normally degraded in cells disomic for chromosome II (Dephoure et al., 2014). We identified 31 of these in our analysis. 11 out of 31 (35.5%) aggregated in disome II cells with a functional proteasome. In disome II *rpn6-1* cells we identified 41 out of these 67 proteins. 21 (51.2%) of these were enriched in aggregates (Fig. 11E). We then asked whether the 67 proteins that were typically degraded when duplicated in disome II cells were enriched in aggregates when proteasome function is compromised. In cells with normal proteasome function, the 67 proteins had a mean enrichment of 1.9 [log2 0.92] in disome II aggregates. In disome II *rpn6-1* cells, their mean enrichment increased to 4.0 [log2 2.0] indicating that many of these usually degraded proteins are now deposited in protein aggregates (Fig. 11F).

We next determined the effects of protein stability on the choice between degradation or aggregation by analyzing the origin recognition complex (ORC) in which some subunits are degraded when in excess while others are not (Dephoure et al., 2014). The origin recognition

complex (ORC) is a six subunit complex essential for the initiation of DNA replication (Bell et al., 1993). All six subunits are encoded on different chromosomes in yeast.

Orc2 levels are mostly attenuated (reduced to 1.1 [log2 0.16] relative to WT) by degradation when the *ORC2* gene is present in two copies instead of one ((Dephoure et al., 2014), Fig. 11G). In contrast, Orc5 is not attenuated (present at 1.8 [log2 0.88] relative to WT) by degradation but is highly enriched in aggregates when encoded by two gene copies (Fig. 11G, H). To mimic stoichiometric imbalance, we transiently overexpressed Orc2 and Orc5 individually from the galactose-inducible *GAL1-10* promoter. Half-life measurements showed that overproduced Orc5 was twice as stable as overproduced Orc2 (Fig. 11I, J). We conclude that the stability of a protein can determine whether it aggregates or is degraded when produced in excess. Our data further indicate that surprisingly subtle half-life differences can determine a protein's dosage compensation fate. What minimal half-life is required for a protein to be eliminated by degradation when in excess remains to be determined.

Dosage compensation by protein aggregation

Degradation of excess subunits of protein complexes serves as a form of dosage compensation (Dephoure et al., 2014). Can protein aggregation serve the same purpose by sequestering excess subunits in an inaccessible aggregate compartment? To address this question, we quantified the relative abundance of proteins that remained in the *soluble fraction* following aggregate isolation in yeast strains disomic for chromosome II or XII and compared it to their abundance in extracts prior to aggregate removal (henceforth *total lysate*) (Fig. 2A). We observed a small but nevertheless significant difference in relative abundance of proteins encoded on euploid

chromosomes between soluble fractions and total lysates (Fig. 12A, B; Supplemental Data S5), which is consistent with the observation that proteins encoded on euploid chromosomes are also found enriched in aggregates isolated from aneuploid cells (Fig. 1C). Depletion of proteins encoded on disomic chromosomes from the soluble fraction was, however, much more dramatic. Their mean relative abundance in the soluble fraction was noticeably decreased to 1.59-fold [log2 0.67] compared to 1.66-fold [log2 0.74] in total lysates for disome II and 1.69-fold [log2 0.76] compared to 1.77-fold [log2 0.82] for disome XII (Fig. 12A, B; Supplemental Data S5).

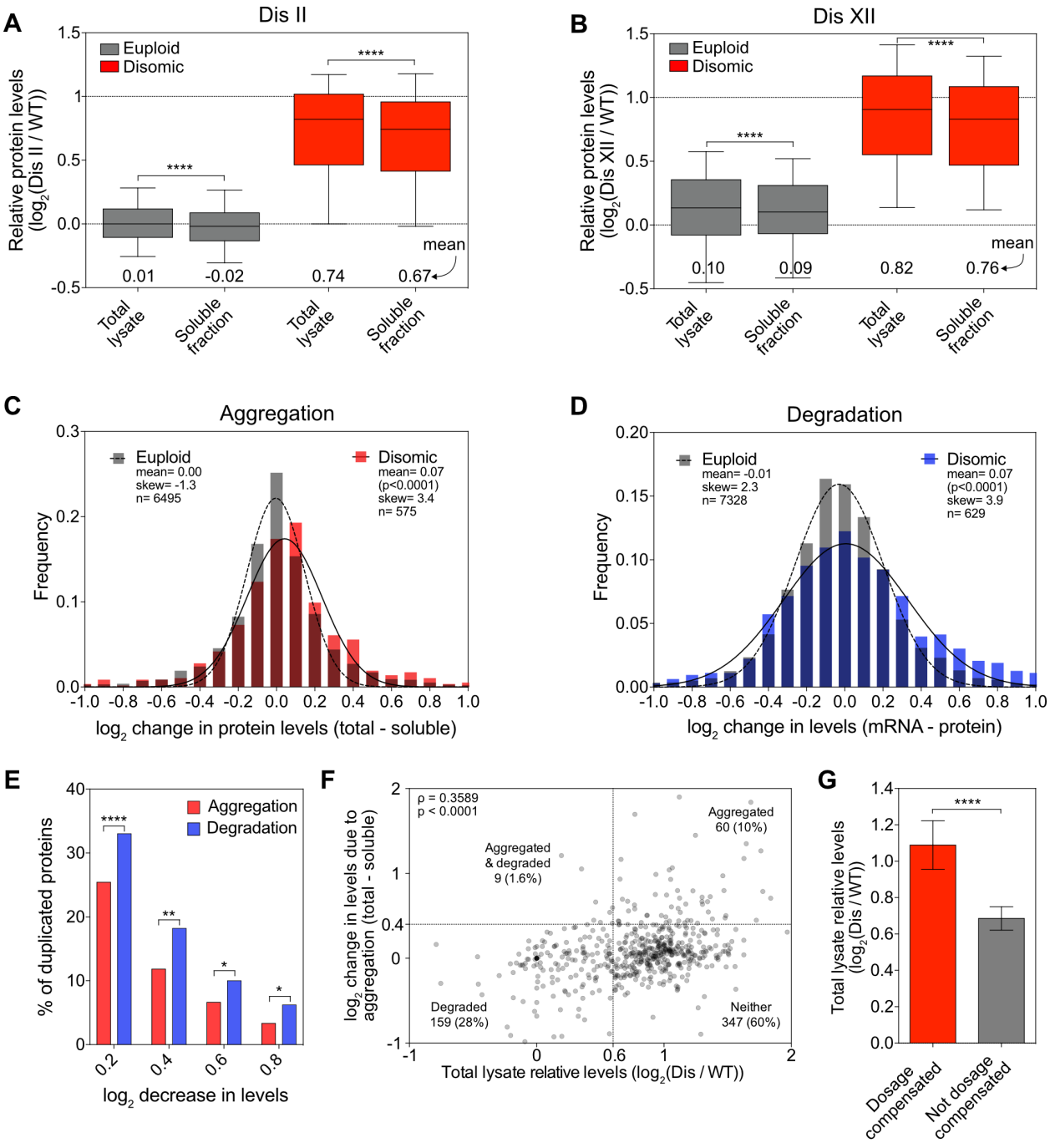

Figure 12. Dosage compensation by protein aggregation.

(A, B) Total lysate and soluble fractions, obtained as described in Fig. 1A, were analyzed for disome II (A) and disome XII (B). Proteins encoded by euploid chromosomes (gray), proteins encoded by the duplicated chromosome (red). Data are represented as box plots with whiskers extending to 10th and 90th percentiles. Means are indicated below. Upper dashed line represents

the theoretical mean for proteins encoded by the duplicated chromosome. **** p<0.0001; Mann-Whitney test.

(C) Changes in protein levels due to aggregation for proteins quantified in (A, B) were calculated for every protein by subtracting its log2 ratio in the soluble fraction from its log2 ratio in the total lysate. Disome II and disome XII data were pooled and separated into two subsets: euploid chromosome encoded proteins (gray) and disomic chromosome encoded proteins (red). Frequency distributions of each subset were then generated using a bin size of 0.1. Distributions were fit to Gaussian curves for euploid (dashed line) and disome (solid line). p-value shows that mean change of disome encoded proteins is significantly different from the mean change of euploid encoded proteins (Mann-Whitney test).

(D) Changes in protein levels due to degradation for disome II and disome XII were calculated by subtracting the relative protein level from the relative mRNA level for each gene as measured by Dephoure et al. (2014). Frequency distributions were generated and curve fitting was performed on the pooled data as in (C). p-value shows that mean change of disome encoded proteins is significantly different from the mean change of euploid encoded proteins (Mann-Whitney test).

(E) The percentage of proteins encoded by the duplicated chromosome that decreased in levels by at least a log2 of 0.2, 0.4, 0.6, and 0.8 due to aggregation (red bars) was calculated using the data described in (C) and due to degradation (blue bars) using the data in (D). **** p<0.0001, ** p<0.01, * p<0.05; Chi-squared test.

(F) Correlation between protein levels in total lysate and reduction in levels due to aggregation was determined for proteins encoded on duplicated chromosomes from the pooled data set of disome II (A) and disome XII (B). Spearman correlation of 0.3589 (p< 0.0001). Dashed lines

indicate thresholds for proteins being considered dosage compensated by aggregation (y-axes) or degradation (x-axes). The number of proteins that fall into each quadrant is indicated (note: 19 data points fell outside the range of the axes but were included in the calculations).
(G) Proteins encoded on duplicated chromosomes were separated into two categories: dosage compensated proteins (red) were defined as their levels being reduced by at least log2 0.4 in the soluble fraction relative to the total lysate. Not dosage compensated proteins (gray) were defined as their levels being reduced by less than log2 0.4. Mean levels in total lysates are plotted; error bars indicate 95% confidence intervals. **** $p<0.0001$; Mann-Whitney test.
Abbreviations: WT, wild-type; Dis, disome.

The decrease of disome-encoded proteins in the soluble fraction could be due to many proteins experiencing small amounts of aggregation or a few proteins aggregating to a large degree. To distinguish between these possibilities, we first calculated the change in levels for each protein by subtracting protein levels in the soluble fraction from those in the total lysate. We then pooled the data from the disome II and disome XII analysis and generated two bins – proteins that were encoded on disomic chromosomes and proteins encoded on euploid chromosomes. As expected, the changes in levels of the 6495 proteins encoded by euploid chromosomes were evenly distributed (Fig. 12C; gray bars); changes in levels of the 575 proteins encoded by disomic chromosomes were not. Both the mean and skew significantly deviated from the expected distribution (Fig. 12C). 147 of the 575 (26%) disome-encoded proteins were depleted by 1.15-fold [log2 0.2] in the soluble fraction compared to total lysate (Fig. 12E), 12% by 1.32-fold [log2 0.4]. Full attenuation, as defined as a 1.74-fold [log2 0.8] decrease occurred for 3% of proteins (Fig. 12E). This indicates that reduction in protein levels for duplicated proteins in the soluble

fraction is largely due to many proteins decreasing by a small degree. However, full attenuation by aggregation also occurs.

How does dosage compensation by aggregation compare to dosage compensation by protein degradation? Previous studies showed that 21% of proteins decrease by 1.52-fold [log2 0.6] either due to downregulation of gene expression or protein degradation (Dephoure et al., 2014). To assess what fraction of the proteome is attenuated solely by protein degradation we subtracted protein levels from RNA levels in the data set published by Dephoure et al. (2014). We then pooled data for disome II and disome XII, allowing us to examine 7328 measurements for proteins encoded on euploid chromosomes and 629 measurements for disome II and XII-encoded proteins. As expected, the changes in levels of the 7328 proteins encoded by euploid chromosomes were evenly distributed (Fig. 12D, gray bars). The 629 proteins encoded by disomic chromosomes behaved differently. Both the mean and skew were larger than observed in the expected distribution (Fig. 12D). 209 of the 629 (33.2%) disome-encoded proteins were depleted by 1.15-fold [log2 0.2] relative to RNA levels (Fig. 12E); 18.4% by 1.32-fold [log2 0.4]. Full attenuation, (1.74-fold [log2 0.8]) occurred for 6.4% of proteins (Fig. 12E). We conclude that small degrees of attenuation of disome-encoded proteins occurs by both aggregation and degradation. Reduction to levels approaching those seen in euploid cells occurs predominantly through protein degradation.

Finally, we asked whether dosage compensation of a particular protein was mediated by both protein degradation and aggregation, or one or the other in our disome II and XII data set. First, we determined whether a correlation existed between relative abundance of a disome-encoded

protein in the total lysate and degree of dosage compensation by aggregation. This was the case, indicating that proteins with higher relative expression levels are more likely to be dosage compensated by aggregation (Fig. 12F). To further assess whether protein aggregation and degradation are mutually exclusive we analyzed the data using cutoffs previously defined (Dephoure et al., 2014). Disome-encoded proteins present in the total lysate at levels below 1.52 [log2 0.6] (instead of the expected 2 [log2 1]) were considered dosage compensated by degradation. Disome-encoded proteins that were depleted in the soluble fraction by at least 1.52 [log2 0.6] compared to the total lysate were considered dosage compensated by aggregation. Based on these criteria, only 9 out of 575 (1.6%) disomic proteins were dosage compensated by both aggregation and degradation, as evident by the fact that the upper left quadrant in the graph in Fig. 12F is underpopulated ($p= 0.0011$; Fisher's exact test). We also asked whether disome-encoded proteins that are dosage compensated by aggregation are present in total lysates at the 2x level expected for disome-encoded proteins. This was the case (Fig. 12G), demonstrating that they are not degraded. In contrast, levels of proteins considered not dosage compensated by aggregation were lower in total extracts. Our data indicate that excess proteins are dosage compensated by aggregation or degradation but rarely both.

DISCUSSION

Our analysis of protein aggregation in disomic yeast strains provided insight into why protein aggregation is a universal feature of the aneuploid state and revealed protein aggregation as a means of dosage compensation. This dosage compensation is not only relevant in aneuploid cells but could very well contribute to stoichiometry control in euploid cells that encounter stoichiometric imbalances due to transient imbalances caused by variability in gene expression.

Remarkably, aggregation is nearly as effective as protein degradation at lowering levels of excess proteins. Whether protein aggregation serves a cytoprotective role in situations where high levels of unassembled complex subunits are present in cells (i.e. in aneuploid cells) is an important question that remains to be answered.

Which proteins aggregate in aneuploid cells?

Our analysis of protein aggregation in a series of disomic yeast strains provided insight into why protein aggregation is a universal feature of the aneuploid state. We identified proteins encoded on the disomic chromosomes as well as proteins encoded on euploid chromosomes to be enriched in aggregates.

Among the proteins found in aggregates of multiple different aneuploid strains, ribosomes were the most prominent. Ribosomes comprise ~75% of aggregates isolated from both euploid cells and aneuploid cells, but ribosomes aggregate more readily in aneuploid cells as judged by the fact that they harbor more aggregates. Why ribosomes are enriched in aggregates of disomic yeast strains is not clear. Perhaps a higher fraction of ribosomes is defective in disomic yeast strains causing them to aggregate.

Ribosomes were not the only proteins commonly found in aggregates of aneuploid yeast strains. Interestingly, the proteins that aggregated in multiple different aneuploid cell lines also aggregated when protein quality control pathways were impaired. This observation raises the interesting possibility that proteins that are especially dependent on protein quality control pathways to maintain their solubility are “canary in the coal mine” proteins of the state of

cellular protein quality control. It will be interesting to determine which properties make the canary proteins so sensitive to perturbations in protein homeostasis.

We also analyzed proteins that aggregate because their encoding genes were duplicated due to disomy. These proteins were strongly enriched for subunits of protein complexes leading us to hypothesize that subunits of heteromeric protein complexes present at levels twice their binding partners are prone to aggregation. We went on to demonstrate that this was true in the case of the eIF2 complex. Our analysis of human trisomic cell lines further revealed that aggregation of proteins that are in excess is surprisingly sensitive to alterations in gene expression. Changing gene expression by as little as 50% causes aggregation of many proteins.

Cellular response to excess subunits of protein complexes

Our findings have important implications for euploid cells. Although expression of subunits of the same complex is tightly coordinated in eukaryotes (Li et al., 2014; Taggart and Li, 2018), it is likely that even euploid cells encounter transient stoichiometric imbalances. Variability in expression of individual subunits can lead to stoichiometric imbalances where individual subunits lack binding partners or where protein complexes are only partially assembled.

Our study shows that 61.5% of proteins either aggregate or are degraded when produced in excess. Among proteins that function in complexes, 73% of proteins either aggregate or are degraded. Interestingly, the 27% that are neither aggregated nor degraded have significantly smaller heteromeric protein binding interfaces than protein complex subunits that aggregate or are degraded. We propose that excess proteins that neither aggregate nor are degraded are simply

less aggregation prone and lack the signals that target them for degradation. Whether there are features that distinguish proteins that are degraded from proteins that are aggregated remains to be determined.

Our data also indicate that aggregation and degradation of proteins encoded on disomic chromosomes is largely mutually exclusive. Intuitively this makes sense - if excess proteins are degraded there is no excess protein to aggregate. What determines whether excess proteins are degraded or aggregate is their half-life when unbound. This is what we observe for Orc2 and Orc5. Overexpressed Orc2 has a shorter overall half-life, presumably because excess Orc2 is degraded. Orc5 has a longer overall half-life when overexpressed because excess Orc5 aggregates. What was surprising was that a change in protein half-life of two-fold appeared to determine whether an overexpressed protein aggregates or is eliminated by degradation. What determines the half-life of a protein and what minimal half-life is required to be eliminated by degradation when in excess remains to be determined.

Dosage compensation by protein aggregation

Our results indicate that protein aggregation is nearly as effective as protein degradation at lowering levels of excess proteins. Using stringent cutoffs to designate a protein being dosage compensated, we found that 12% of disome-encoded proteins were depleted from the soluble fraction by 1.32-fold due to aggregation; depletion by protein degradation occurred for 18.4% of proteins.

We hypothesize that protein aggregation could serve a cytoprotective function that shields aneuploid and euploid cells from toxic activities of excess protein complex subunits. Testing this hypothesis requires the development of methods that prevent protein aggregation in aneuploid cells, a task that has proven difficult. Protein aggregation as a cytoprotective mechanism has been proposed during cellular and organismal aging (David et al., 2010; Walther et al., 2015), suggesting that it could also serve this function in aneuploid cells. However, it is worth noting that the aggregation that occurs during aging does not function to normalize stoichiometric imbalances but rather to protect cells from highly abundant proteins that exceed their solubility during aging. Protein aggregation has also been found to protect cells during stress by sequestering misfolded proteins and targeting them to specific subcellular compartments (Escusa-Toret et al., 2013; Miller et al., 2015). Aggregation of disease inducing proteins such as amyloid β, associated with Alzheimers Disease, has also been proposed to protect neurons from toxic oligomers (Caughey and Lansbury, 2003). Our observation that widespread aggregation serves the purpose of protecting cells from gene dosage excess raises the intriguing possibility that aggregation may be one mechanism that allows cancer cells to tolerate aneuploidy. Analysis of the degree of protein aggregation in cancer cell lines that are highly aneuploid and thus experience large scale stoichiometric imbalances yet have high proliferative potential, could shed light on this question.

MATERIALS AND METHODS

Yeast strains, plasmids, and growth conditions

All yeast strains are derivatives of W303 and are described in Table S1. Primers are listed in Table S2 and plasmids are listed in Table S3. Yeast strains were generated and manipulated as

described previously (Guthrie and Fink, 1991). Cells were grown at 30°C in YEP supplemented with 2% raffinose (YEP-R), 2% raffinose + 2% galactose (YEP- RG), or 2% glucose (YEP-D), or in synthetic complete medium supplemented 2% glucose (SC-D). For SILAC experiments, cells were grown overnight in SC-D medium containing 100 µg/mL of heavy (13C6 15N2 (K8), Cambridge Isotope Labs) or light (K4) (Sigma-Aldrich) L-lysine. Cells were then diluted to OD_{600}=0.05 to 0.1 and grown to OD_{600}=0.4 to 1.0 before harvesting lysates. Strains harboring temperature sensitive mutations in *NDC10* or *RPN6* were grown over night at room temperature, grown to exponential phase at room temperature before shifting to the semi-permissive temperature of 30°C. *ndc10-1* strains were grown for 4 hours and *rpn6-1* strains were grown for 6 hours at 30°C before harvesting for aggregate purification. For radicicol treatments, cells harboring a deletion in the gene encoding the drug transporter Pdr5 were grown to exponential phase in SC medium at 30°C and radicicol (Cayman Chemical Company) or DMSO was added to the culture at a final concentration of 70 µM. Cells were grown in the presence of radicicol or DMSO for 30 minutes at 30°C before harvesting for aggregate purification.

Disomes used in this study are derivatives of those published in Torres et al. (2007). Gene deletions, fusion proteins, and promoter swaps were generated using PCR-based methods (Longtine et al., 1998) in a wild-type W303 yeast strain. Disomes carrying gene manipulations were constructed by crosses. Karyotypes of key disomic strains were verified by whole genome sequencing.

GCD11-HA and *GCD11*Δ were generated by PCR based methods (Longtine et al., 1998). *GCD11-HA* with its native promoter and terminator was cloned into p24 by Gibson cloning (Gibson et al., 2009) and integrated at *ura3* by NcoI digestion. The p*SUI2/SUI3* CEN plasmid

was constructed by Gibson cloning (Gibson et al., 2009) *SUI2* and *SUI3* with their native promoters and terminators into p158.

Cell culture and SILAC labeling of RPE-1 cells

RPE-1 wild type (control) cells were grown in heavy (K8) SILAC DMEM medium (DMEM with high glucose—minus glutamine, lysine and arginine, supplemented with 10% dialyzed FBS (Gibco), 50μg/mL K8 lysine (Cambridge Isotope Labs), 85μg/mL arginine (Sigma-Aldrich), and 2mM glutamine/1mM sodium pyruvate (Thermo Fisher). RPE-1 cells trisomic for chromosome 12 or 21 were grown in light (K0) SILAC DMEM (same formulation as above, with 50μg/mL K0 lysine; Sigma-Aldrich). Cells were cultured for 8 generations prior to testing K8 labeling efficiency by mass spectrometry, then expanded for harvest at passage 10. Prior to harvest, cells were treated with 25μM chloroquine for 16 hours, followed by addition of 1μM MG-132 (Calbiochem) for 6 hours to induce protein aggregation. Cells were harvested by trypsinization, neutralized in SILAC DMEM, and washed with PBS; an equal number of heavy and light cells were mixed (wild type + Ts12; wild type + Ts21), treated with sodium azide, and snap frozen for aggregate purification.

Aggregate purification and analysis

Aggregate purification was carried out based on methods published by Koplin et al., (2010), with some modifications to the cell lysis procedure. Cells were grown to exponential phase in SC medium under the conditions indicated above. Sodium azide was added to a final concentration of 50 mM in the cultures before pelleting cells. 10 OD_{600} units of cell pellets were washed with 50 mM sodium azide and snap frozen in liquid nitrogen. Frozen pellets were then

resuspended in 1 mL of buffer containing 1M sorbitol, 100mM sodium citrate, 60mM EDTA, 10mM DTT, pH 7.0 (SCED), 1mM PMSF, and 2x protease inhibitor tablets (Roche). Cells were pelleted then resuspended in 1 mL of SCED buffer containing 1 mg/mL 20T Zymolyase (MP Biomedicals) and rotated gently for 8 minutes at room temperature. Digested cells were pelleted by spinning at 250xg for 5 minutes at 4°C. Pellets were washed gently in ice-cold SCED buffer containing 1mM PMSF and 2x protease inhibitors. Cells were lysed by resuspending in 1.5 mL of ice-cold lysis buffer containing 20mM NaPi pH 6.8, 10mM DTT, 1mM EDTA, 0.1% Tween 20 (Sigma-Aldrich), 1 mM PMSF, and 5x protease inhibitors and then sonicated for 8 pulses at output level 4, 50% duty cycle. Lysates were cleared by spinning for 2 minutes at 850xg at 4°C yielding the *total lysate fraction*. Protein concentrations of the cleared lysate were determined by Bradford (Bio-Rad). For quantitative Western blotting and Coomassie staining, samples were diluted to the same protein concentration. A small aliquot of the total lysate fraction was taken for analysis and aggregates were pelleted from the remainder of the total lysate by spinning the extract for 20 minutes at 16,000xg at 4°C. The supernatant was removed and reserved as the *soluble fraction*. Pellets were then washed twice by resuspending in buffer containing 20mM NaPi pH 6.8, 2% NP40, 1mM PMSF, and 2x protease inhibitors then sonicating for 6 pulses at output level 4, 50% duty cycle. A final wash was carried out in buffer containing 20mM NaPi pH 6.8, 1mM PMSF, and 2x protease inhibitors then sonicating for 4 pulses at output level 2, 60% duty cycle yielding the *aggregate fraction*. All centrifugations during the wash steps were 20 minutes at 16,000xg at 4°C.

For Western blot analysis and Coomassie staining analysis, total lysate fractions were boiled in SDS loading buffer. Aggregate fractions were resuspended in buffer containing 200mM NaPi pH 6.8, 5% SDS, 8M Urea, 10mM DTT, and 0.01% bromophenol blue (HU buffer) then

vortexed at 1500 rpm for 5 minutes at 60°C. For coomassie staining, SDS-PAGE gels were stained with Imperial protein stain (Thermo Scientific) per manufactures instructions. For Western blotting, proteins were separated by SDS-PAGE then transferred to nitrocellulose membranes and detected by the following antibodies. Hsp104-eGFP was detected using a mouse anti-GFP antibody (JL-8 epitope, Clontech) at a dilution of 1:1000. Pgk1 was detected using a mouse anti-Pgk1 antibody (Invitrogen) at a 1:10,000 dilution. Gcd11-HA was detected using a mouse anti-HA antibody (HA.11 epitope, BioLegend). Quantification was performed using Fiji software (Schindelin et al., 2012).

Ribosome purification

Cells were grown to exponential phase, pelleted by centrifugation, washed with water, then snap frozen in liquid nitrogen. Cells were incubated in lysis buffer (20 mM HEPES pH 7.4, 100mM potassium acetate, 2mM magnesium acetate, 3mM DTT, protease inhibitor tablets, and 1 mg/mL Zymolyase) for 5 minutes at 4°C then lysed by French Press (Sim-Aminco). The lysate was cleared by centrifuging for 20 minutes at 19,000 RPM (Beckman Coulter JA25.50 rotor) at 4°C. The cleared lysates were applied to a 30% sucrose gradient containing 20 mM HEPES pH 7.4, 500mM potassium acetate, 2mM magnesium acetate, and 3mM DTT. Ribosomes were pelleted by centrifugation for 4 hours at 4°C at 50,000 RPM (Beckman Coulter Type 70 Ti rotor). The pellet was dissolved in lysis buffer and analyzed by SDS-PAGE.

Quantitative Proteomics of Yeast Aggregates

5 OD_{600} units of heavy-labeled cells were mixed with 5 OD_{600} units of light-labeled cells and aggregates were harvested as described above. Purified SILAC-labeled aggregates were

boiled in SDS lysis buffer (50mM Tris pH 7.5, 2mM EDTA, 2% SDS) and separated on 4-12% Tris-Glycine by SDS-PAGE gels (Invitrogen). Proteins were visualized by Coomassie staining (45% methanol, 5% glacial acetic acid, 3g/L Coomassie brilliant blue G-250) and subjected to gel band cutting (Shevchenko et al., 2006), destaining, and in-gel Lys-C (Wako Chemicals) digest at a 50:1 protein to protease ratio to liberate peptides. Following digest, the buffer containing peptides and gel slices were acidified using 50% acetonitrile/5% formic acid, and recovered peptides were vacuum centrifuged to near dryness. Each gel fraction was desalted via StageTip, dried via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing.

Quantitative proteomics of RPE-1 cell aggregates

RPE-1 cells were cultured and treated as described above. Aggregate purification was carried out as described for yeast cells with the following modifications. Cells were collected, treated with sodium azide and snap frozen. Frozen cell pellets were resuspended in cold lysis buffer (described above) and sonicated. Samples were analyzed using the same mass spectrometric method and data processing workflow as yeast aggregates (see below).

Liquid chromatography and mass spectrometry

MS data were collected using a Q Exactive mass spectrometer (Thermo Fisher) coupled with a Famos autosampler (LC Packings) and an Accela600 liquid chromatography (LC) pump (Thermo Fisher). Peptides were separated on an ~18cm column with 100μm inner diameter packed with of Accucore150 resin (2.6 μm, 150 Å, ThermoFisher Scientific, San Jose, CA); ~1μg peptides were loaded onto the column for each analysis. Peptides were separated with a

2hour gradient of 5-26% acetonitrile in 0.125% formic acid with a flow rate of ~300nL/min. For the MS1 scan, resolution was set to 70,000 with an automatic gain control (AGC) target 1×10^6 and a maximum injection time of 250 ms. We selected the top 20 precursors for HCD MS2 analysis with the following parameters: resolution, 17,500; AGC 1×10^5; maximum injection time, 90 ms; isolation window, 2 Th; normalized collision energy (NCE), 25; and centroid spectrum data type. In addition, unassigned and singly charged species were excluded from MS2 analysis, and dynamic exclusion was set to automatic.

For data analysis, mass spectra were processed using a Sequest-based in-house software pipeline. MS spectra were converted to mzXML using a modified version of ReAdW.exe. MS2 spectra were searched with a database including all predictive ORFs for entries from the yeast SGD (http://www.yeastgenome.org/download-data; downloaded March 12, 2014). For human aggregate samples, Database searching included all entries from the human UniProt database (10 August 2011). Both yeast and human databases were concatenated with a reverse database composed of all protein sequences in reverse order. Searches were performed using a 50ppm precursor ion tolerance. Product ion tolerance was set to 0.03 Th. Oxidation of methionine residues (+15.995 Da) and heavy lysine (K8) incorporation (+8.0142) were set as a variable modification. Peptide spectral matches (PSMs) were altered to a 1% False discovery rate (FDR) (Elias and Gygi, 2007; 2010). PSM filtering was performed using a linear discriminant analysis, as described previously (Huttlin et al., 2010), while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy. Peptide-spectral matches were identified, quantified, and collapsed to a 1% FDR and then further collapsed to a final protein-level FDR of 1%. Furthermore, protein assembly was guided by principles of parsimony to produce the smallest set of proteins necessary to account for all

observed peptides. 4 gel band regions were cut per sample (individual lanes) of the SDS-PAGE gel; these 4 regions were processed separately and each is an independent MS sample (to decrease complexity of peptides in each run). Protein assembly was utilized to group the 4 runs of a given sample together for batched analysis and calculation of $\log_2$(Heavy/Light) ratios. The resulting data were filtered for SILAC quantified proteins based on the summed signal to noise for heavy and light peptides; contaminant peptides identified in the search were removed at this step. An intensity cutoff of 10 was applied for summed heavy plus light channels, and each individual channel had an intensity cutoff of 5 (to avoid identifying proteins with an intensity of 0 in either the light or heavy channel; i.e. proteins found only in one condition or the other). A SILAC-labeled total lysate reference sample (to allow for normalization of SILAC mixing) was obtained as described above. Proteins were pelleted by TCA precipitation, dried via vacuum centrifugation, and digested with Lys-C. Resulting peptides were desalted using StageTips, and MS data collection and analysis were performed as described above for aggregate samples.

SILAC mass spectrometry of total lysate and soluble fractions

Total lysate samples were prepared, and soluble fractions (supernatant post aggregate isolation) were collected as described above. ~800µg of protein per sample was pelleted via TCA precipitation and dried via vacuum centrifugation. Samples were reduced and alkylated followed by digestion with Lys-C. Resulting peptides were desalted using StageTips and samples were dried via vacuum centrifugation.

Each sample was fractionated using off-line basic pH reversed-phase chromatography (BPRP HPLC) (Wang et al., 2011). We used an Agilent 1200 pump equipped with a degasser and a photodiode array (PDA) detector (set at 220 and 280 nm wavelength) from ThermoFisher

Scientific. Peptides were subjected to a 50-min linear gradient from 5% to 35% acetonitrile in 10 mM ammonium bicarbonate pH 8 at a flow rate of 0.6 mL/min over an Agilent 300Extend C18 column (3.5 μm particles, 4.6 mm ID and 220 mm in length). The peptide mixture was fractionated into a total of 96 fractions, which were consolidated into 24, from which 12 non-adjacent samples were analyzed (Paulo et al., 2016). Samples were subsequently acidified with 1% formic acid and vacuum centrifuged to near dryness. Each consolidated fraction was desalted via StageTip, dried again via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing. Samples were analyzed using a Q Exactive mass spectrometer using the same instrument method and data processing as described above for yeast aggregate samples.

Real-time RT-qPCR

Cells were grown to exponential phase in SC medium containing 2% glucose as for aggregate purification. 5 OD_{600} units of culture were pelleted by centrifugation at 4°C (3000 rpm, 2 min). The pellet was resuspended in 1 ml cold SC medium, transferred to a 2 ml microfuge tube and pelleted by centrifugation at 4°C (3000 rpm, 2 min). The pellet was flash-frozen in liquid nitrogen and stored at -80°C. To extract total RNA, ~200 μl of glass beads, 400 μl TES buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS) and 400 μl acid phenol:chloroform (pH 4.5) were added to the cell pellet and the tubes were vortexed for 30 minutes at 65°C. The phases were separated by centrifugation, and the top phase was transferred to a new tube containing 1 ml of 120 mM sodium acetate in ethanol to precipitate RNA at 4°C. Precipitates were collected by centrifugation and resuspended in 100 μl DEPC-treated water. Total RNA was further purified using the RNeasy Mini kit (Qiagen), including DNase treatment, according to the

manufacturer's instructions. cDNA was synthesized from 750 ng total RNA using the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen) with random hexamer primers according to the manufacturer's instructions. Real-time qPCR reactions were run using the SYBR Premix Ex Taq Perfect Real Time kit (TaKaRa Bio) and a Roche LightCycler 480 (Roche) according to manufacturer's instructions. qPCR primers are listed in Table S2.

Turnover measurements for aggregated proteins

Protein turnover for aggregated proteins in disome II cells was determined by examining protein levels relative to WT before treatment with MG-132 and choloroquine, and 90 seconds and 300 seconds thereafter. Data were generated by Dephoure et al. (2014).

Orc protein half-life measurements

Cells were grown to exponential phase at 30°C in YEP medium containing 2% raffinose. Expression of HA tagged *ORC2* and *ORC5* was induced by diluting cultures into YEP medium containing 2% raffinose and 2% galactose for 20 minutes. The 0 minute time point was taken and protein synthesis was halted by the addition of 2% glucose and 0.5 mg/mL cycloheximide (Sigma-Aldrich). Time points were taken at the indicated times after the addition of glucose and cycloheximide. For each time point, approximately 0.5 OD_{600} units of culture were pelleted by centrifugation at 4°C (3000 rpm, 2 min). Cells were incubated at 4°C in 5% trichloroacetic acid for at least 10 minutes. Cell pellets were washed once with acetone and dried. Cells were lysed with glass beads in 100 µL of lysis buffer (50 mM Tris-Cl at pH 7.5, 1 mM EDTA, 2.75 mM dithiothreitol [DTT], complete protease inhibitor cocktail [Roche]) with a bead beater. Samples were boiled in 1X SDS loading buffer. Following SDS-PAGE and transfer of proteins to a

nitrocellulose membrane, proteins were detected with the following antibodies. Pgk1 was detected using a mouse anti-Pgk1 antibody (Invitrogen) at a 1:10,000 dilution. Orc2-HA and Orc5-HA were detected using a mouse anti-HA antibody (HA.11 epitope, BioLegend). Quantification was performed using Fiji software (Schindelin et al., 2012). Protein levels were calculated by subtracting a background measurement for each band then dividing the intensity of the Orc band by the intensity of the Pgk1 band and normalizing all time points to the 0 minute time point. Half-life calculations were made by fitting curves to an exponential decay function in Prism (Graphpad).

SILAC ms data normalization

To account for imperfect mixing of heavy and light-labeled cells, log2 ratios of proteins identified in aggregates were normalized by subtracting the average log2 ratio of all euploid encoded proteins in the total lysate obtained before pelleting aggregates for each experiment.

To control for non-biological batch effects between disomic cell-lines, each experiment was mean-centered to 0 by subtracting the mean of all SILAC ratios in that experiment from all data points. To return the normalized values to a baseline that more closely resembles the increase in protein aggregation in disomic strains observed in the raw data, a constant (log2 0.27) was added to all normalized data points. This constant is the mean log2 ratio of all euploid encoded proteins in the dataset prior to normalization.

Cutoff determination for aggregation

To decide what level of enrichment in disome aggregates constituted calling a protein aggregated, we used a 5% false discovery rate (FDR) obtained from the analysis of aggregates

obtained from a WT(heavy)-WT(light) SILAC ms analysis. In this analysis, only 5% of identified proteins were enriched by greater than log2 1.27 in WT(heavy) aggregates.

Signal to noise quantification of relative aggregate burden

To account for the fact that some proteins are more abundant in aggregates than others, aggregate burden for each disome relative to WT was calculated by comparing the signal to noise ratio (S:N) of all heavy-labeled proteins to the S:N of all light-labeled proteins in protein aggregates. The sum of all light-labeled S:N was divided by the sum of all heavy-labeled S:N. This ratio was then normalized by dividing the ratio of light S:N to the heavy S:N for the total lysate.

Gene Ontology (GO) analysis

GO component analysis was performed using the Saccharomyces Genome Database (SGD) GO term finder (accessed 10/12/2018).

Aggregate property calculations

Disorder predictions for protein sequences were calculated using IUPRED on the 'long' setting. For each sequence, per-residue disorder scores were averaged across the full length of the protein. Grand average hydropathy (GRAVY) scores, isoelectric points, aliphatic indices, and aromaticity scores were calculated using the YeastMine tool on Saccharomyces Genome Database (SGD) (accessed 10/31/2018).

Proteins not quantified due to low signal in WT channel

In our mass spectrometry analysis pipeline, peptides for which either individual analysis channel had an intensity of less than 5 were discarded to avoid identifying proteins with an intensity of 0 in either the light or heavy channel; i.e. proteins found only in one condition or the other because a SILAC ratio cannot be calculated. We examined proteins that failed to pass the signal to noise threshold in one channel, but had a signal to noise ratio (S:N) of at least twice the threshold in the other channel. Using this method, we found 320 proteins that were quantified only in the disome channel compared to 72 quantified in only the WT channel (Supplemental Data S1). Importantly, 92 of the proteins quantified only in the disome channel were encoded by the duplicated chromosome as opposed to just 2 duplicated gene-encoded proteins quantified in only the WT channel. It is possible that proteins that are identified exclusively in the disome channel never aggregate in WT cells. It is also possible that with increased coverage these proteins would pass the signal threshold in both channels. To distinguish between these possibilities, we took advantage of a difference in coverage between two of our replicate experiments for disome II. In the low coverage experiment (366 proteins quantified), 9 proteins were not quantified due to low signal in the WT channel. In the high coverage experiment (847 proteins quantified), we identified 6 of these proteins, and all 6 passed the detection threshold in both channels allowing us to calculate a SILAC ratio. 4 of these 6 proteins were considered aggregating using 2.4-fold [log2 of 1.27] as a cutoff. We conclude that proteins that cannot be quantified by SILAC ms due to low signal in WT cells are also likely enriched in aneuploid aggregates.

Heteromeric interface size determination

We searched for subunits of heteromeric protein complexes in the Protein Data Bank (on February 2, 2017) with >90% sequence identity to *S. cerevisiae* genes. We selected a single heteromeric structure for each gene – for genes with multiple structures available, we selected the structure with the greatest number of unique subunits in the complex, followed by the greatest number of atoms present in the polypeptide chain for ties. The total amount of heteromeric interface formed by each subunit was calculated using AREAIMOL (Winn et al., 2011).

Statistical analysis

The statistical tests used are indicated in the Figure Legend, Materials and Methods, and/or in the Results section. Values of n, definition of center, error bars (e.g. SD and confidence intervals), and significance levels are reported in the Figures and/or in the Figure Legends. All box plots represent median and interquartile range (IQR) with whiskers indicating the 10-90 percentile. All indicated statistical tests were performed using MATLAB or Prism.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository in seven parts with the following dataset identifiers: PXD011874, PXD011875, PXD011876, PXD011877, PXD011878, PXD011915.

Supplemental Table S1. Yeast strains used in this study, Related to Materials and Methods.

Strain number	Aneuploidy	Genotype
A2587	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [phi+]*
A35797		*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], ade1::HIS3, lys2::KAN*
A13413	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, ndc10-1*
A28265	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker)*
A23504	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], ade1::HIS3, lys2::KAN, xxx::URA3 (marker inserted at random sequence on chromosome VI)*
A23487	I	*MATalpha, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], ade1::KAN, ade1::HIS, lys2::URA3*
A23506	II	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], lys2::KAN, lys2::HIS3, xxx::URA3 (marker inserted at random sequence on chromosome VI)*
A23489	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker), lys2::URA3*
A23490	VIII	*MATalpha, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, xxx::HIS3 (markers inserted on chromosome VIII; 119778-119573 deleted between ERG11 and STP2 on chromosome VIII), lys2::URA3*
A23491	IX	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, xxx::HIS3 (markers inserted on chromosome IX; 342433-342832 deleted between FAA3 and URM1 on chromosome IX), lys2::URA3*
A23492	X	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::HIS, xxx::KanMX (markers inserted on chromosome X; 322250-322350 deleted between NUP82 and BNA3 on chromosome X), lys2::URA3*
A23493	XI	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, xxx::HIS3 (markers inserted on chromosome XI; 430900-431000 deleted between SFT1 and RPL14A on chromosome XI), lys2::URA3*
A23494	XII	*MATalpha, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], ade16::KAN, ade16::HIS3, lys2::URA3*

A23495	XIII	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KanMX, xxx::HIS (markers inserted on chromosome XIII; 309200-309300 deleted between SPO20 and YMR018W on chromosome XIII), lys2::URA3*
A23496	XIV	*MATalpha, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, xxx::HIS3 (markers inserted on chromosome XIV; 622880-622980 deleted between MRP7 and HRB1 on chromosome XIV), lys2::URA3*
A23497	XV	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], leu9::KAN, leu9::HIS3, lys2::URA3*
A23498	XVI	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], met12::KAN, met12::HIS3, lys2::URA3*
A13413	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, ndc10-1*
A15548	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], ade1::HIS3, lys2::KAN, pdr5::TRP1*
A38239	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], rpn6-1::URA*
A40189	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [phi+], ade1::HIS3, lys2::KAN, gcd11-HA::NATMX6, p158*
A40190	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker), GCD11/GCD11-HA::NATMX6, p158*
A40195	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker), GCD11/GCD11-HA::NATMX6, pSUI2/SUI3*
A40193	-	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [phi+], ade1::HIS3, lys2::KAN, gcd11-HA::NATMX6*
A40192	-	*MATalpha, ade2-1, leu2-3, trp1-1, his3-11,15, can1-100, GAL, [phi+], ura3::gcd11-HA::NATMX6*
A40194	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker), GCD11/GCD11-HA::NATMX6*
A40191	V	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], xxx::KAN, can1::HIS3 (markers inserted on chromosome V; 187520-187620 deleted between FAA2 and BIM1 on chromosome V for KAN marker), gcd11Δ::TRP/GCD11-HA::NATMX6*

A40196	II	*MATa, ade2-1, leu2-3, ura3, trp1-1, his3-11,15, can1-100, GAL, [psi+], lys2::KAN, lys2::HIS3, rpn6-1::URA*
A40197	-	*MATa/alpha, ade2-1/ade2-1, leu2-3/leu2-3, ura3/ura3, trp1-1/trp1-1, his3-11,15/his3-11,15, can1-100/can-100, GAL/GAL, [phi+], ORC2/orc2::pGAL1-3HA-ORC2:KanMX6*
A40198	-	*MATa/alpha, ade2-1/ade2-1, leu2-3/leu2-3, ura3/ura3, trp1-1/trp1-1, his3-11,15/his3-11,15, can1-100/can-100, GAL/GAL, [phi+], ORC5/orc5::pGAL1-3HA-ORC5:KanMX6*

Supplemental Table S2. Primers used in this study, Related to Materials and Methods.

Primer name	Target	Used for	Sequence
10467	GCD11/ p2426	GCD11 HA tag	GGCAACCATTAAAAAGGGTACTACATTGGAA CCCATCGCTcggatccccgggttaattaa
10468	GCD11 p2426	GCD11 HA tag	ATGAAATTTTTGTCTTTGCAGTGGTTTTATTG GTTCCTTAgaattcgagctcgtttaaac
10967	GCD11/ p2424	GCD11 deletion	GAGCGTAATACACCGTTAACATCGCGCATTA GAGGTAGACcggatccccgggttaattaa
10968	GCD11/ p2424	GCD11 deletion	ATTATGAAATTTTTGTCTTTGCAGTGGTTTTAT TGGTTCCgaattcgagctcgtttaaac
10809	GCD11/ p24	GCD11-HA cloning	agcttgcatgcctgcaggtcgactctagagCCTGCTGTACC TCATGGCTTTGGAATAGGG
9387	ADH1term/ p24	GCD11-HA cloning	acgacgttgtaaaacgacggccagtgaattGGTGTGGTCA ATAAGAGCGACCTCATACTA
10663	pMOBY-SUI2	SUI2/3 cloning	agcttgcatgcctgcaggtcgactctagagCAGACGTATCA GTACATCACGAGACTACTA_
10893	SUI2, SUI3	SUI2/3 cloning	AAAGTAACTCGGGACTGACTAAAGTTGTAAAT TACAGCATCACTACCTTTTGAGGTAAAA
10894	SUI2, SUI3	SUI2/3 cloning	TTTTACCTCAAAAGGTAGTGATGCTGTAATTT ACAACTTTAGTCAGTCCCGAGTTACTTT
10664	pMOBY-SUI3	SUI2/3 cloning	acgacgttgtaaaacgacggccagtgaattACTTACCACA TCACGATAGGTCTCACGATC
11026	GCD11	qPCR f	AGGTACCATTGCTGACGGTG
11027	GCD11	qPCR r	TCGATTTCAGCACCTGGCTT
11030	SUI2	qPCR f	TTAAATTAGTCGCCGCCCCA
11031	SUI2	qPCR r	AGCCTTTGGTGGCATGGTAA
11034	SUI3	qPCR f	CGTTTCTGCCGATGCTGAAG
11035	SUI3	qPCR r	AGCACTTGGAGTGCCTTCTT
8464	ACT1	qPCR forward	GTACCACCATGTTCCCAGGTATT
8465	ACT1	qPCR reverse	CAAGATAGAACCACCAATCCAGA
11992	ORC2/ p244	ORC2 tagging	CTTTAAAACAAGTTTTTGTAGTACTGCGAATT GCCATAACgaattcgagctcgtttaaac
11993	ORC2/ p244	ORC2 tagging	GGATATCATTATGCTCTACAAAGTCTTCCCCA TTTAGCATgcactgagcagcgtaatctg
12000	ORC5/ p244	ORC5 tagging	ACTTTTCTTTTATTTTCTTCTTAATACTTTTGG AAATAAAgaattcgagctcgtttaaac
12001	ORC5/ p244	ORC5 tagging	GATATTCCCTAAAAGCAACTTCCGGAGTGGT CACATTCATgcactgagcagcgtaatctg

Supplemental Table S3. Plasmids used in this study, Related to Materials and Methods.

Plasmid name	Backbone	Yeast genes cloned in
p1522	pFA6a-mCherry-natMX6	-
pMoBY-SUI2	MoBY	SUI2
pMoBY-SUI3	MoBY	SUI3
p158	YCplac33	-
pSUI2/SUI3	p158	SUI2 and SUI3
p2426	pFA6a-3HA-natMX6	-
p24	YIplac211	-
pGCD11-HA	p24	GCD11-HA
p2424	pFA6a-natMX6	-
p244	pFA6a-kanMX6-PGAL1-3HA	-

Acknowledgments

We thank Arzu Sandikci for providing purified ribosomes, Steve Bell for assistance with ORC experiments, David Pincus, Luke Berchowitz, and members of the Amon lab for suggestions and critical reading of this manuscript. This work was supported by NIH grants CA206157 and GM118066 to A.A., who is an investigator of the Howard Hughes Medical Institute and the Paul F. Glenn Center for Biology of Aging Research at MIT. J.A.M. is supported by a Medical Research Council Career Development Award (MR/M02122X/1). S. S. was supported by the American Italian Cancer Foundation (AICF), by a Fellowship in Cancer Research from Marie Curie Actions and the Italian Association for Cancer Research (AIRC), and by a KI Quinquennial Cancer Research Fellowship.

References

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nature Publishing Group *18*, 285–298.

Bell, S.P., Kobayashi, R., and Stillman, B. (1993). Yeast origin recognition complex functions in transcription silencing and DNA replication. Science *262*, 1844–1849.

Caughey, B., and Lansbury, P.T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu. Rev. Neurosci. *26*, 267–298.

David, D.C., Ollikainen, N., Trinidad, J.C., Cary, M.P., Burlingame, A.L., and Kenyon, C. (2010). Widespread protein aggregation as an inherent part of aging in C. elegans. PLoS Biol. *8*, e1000450.

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Elife *3*, e03023.

Donnelly, N., and Storchová, Z. (2014). Dynamic karyotype, dynamic proteome: buffering the effects of aneuploidy. Biochim. Biophys. Acta *1843*, 473–481.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods *4*, 207–214.

Elias, J.E., and Gygi, S.P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol. Biol. *604*, 55–71.

Escusa-Toret, S., Vonk, W.I.M., and Frydman, J. (2013). Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. Nat Cell Biol *15*, 1231–1243.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods *6*, 343–345.

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., and Beltrao, P. (2017). Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. Cell Syst *5*, 386–398.e4.

Guthrie, C., and Fink, G.R. (1991). Methods in enzymology: molecular biology of Saccharomyces cerevisiae (San Diego: Academic Press).

Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. Cell *143*, 1174–1189.

Isono, E., Saito, N., Kamata, N., Saeki, Y., and Toh-E, A. (2005). Functional analysis of Rpn6p, a lid component of the 26 S proteasome, using temperature-sensitive rpn6 mutants of the yeast Saccharomyces cerevisiae. J. Biol. Chem. *280*, 6537–6547.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487–498.

Koplin, A., Preissler, S., Ilina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. J. Cell Biol. *189*, 57–68.

Levy, E.D., De, S., and Teichmann, S.A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc. Natl. Acad. Sci. U.S.a. *109*, 20461–20466.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Liu, B., Larsson, L., Caballero, A., Hao, X., Oling, D., Grantham, J., and Nyström, T. (2010). The polarisome is required for segregation and retrograde transport of protein aggregates. Cell *140*, 257–267.

Longtine, M.S., McKenzie, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Marsh, J.A., Hernández, H., Hall, Z., Ahnert, S.E., Perica, T., Robinson, C.V., and Teichmann, S.A. (2013). Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. Cell *153*, 461–470.

McShane, E., Sin, C., Zauber, H., Wells, J.N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchová, Z., Marsh, J.A., et al. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. Cell *167*, 803–815.e821.

Miller, S.B.M., Mogk, A., and Bukau, B. (2015). Spatially Organized Aggregation of Misfolded Proteins as Cellular Stress Defense Strategy. Journal of Molecular Biology 1–11.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell Proteomics *1*, 376–386.

Ori, A., Iskar, M., Buczak, K., Kastritis, P., Parca, L., Andrés-Pons, A., Singer, S., Bork, P., and Beck, M. (2016). Spatiotemporal variation of mammalian protein complex stoichiometries. Genome Biol. *17*, 47.

Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. Genes & Development *26*, 2696–2708.

Paulo, J.A., O'Connell, J.D., Everley, R.A., O'Brien, J., Gygi, M.A., and Gygi, S.P. (2016). Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 S. cerevisiae proteins across 10 carbon sources. J Proteomics *148*, 85–93.

Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., and Li, R. (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature *468*, 321–325.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. *37*, 825–831.

Rachidi, M., Lopes, C., Costantine, M., and Delabar, J.-M. (2005). C21orf5, a new member of Dopey family involved in morphogenesis, could participate in neurological alterations and mental retardation in Down syndrome. DNA Res. *12*, 203–210.

Santaguida, S., and Amon, A. (2015). Short- and long-term effects of chromosome mis-segregation and aneuploidy. Nature Publishing Group *16*, 473–485.

Santaguida, S., Vasile, E., White, E., and Amon, A. (2015). Aneuploidy-induced cellular stresses limit autophagic degradation. Genes & Development *29*, 2010–2021.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676–682.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc *1*, 2856–2860.

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchová, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Molecular Systems Biology *8*, 608.

Taggart, J.C., and Li, G.-W. (2018). Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst *7*, 580–589.e584.

Tang, Y.-C., Williams, B.R., Siegel, J.J., and Amon, A. (2011). Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell *144*, 499–512.

Torres, E.M., Dephoure, N., Panneerselvam, A., Tucker, C.M., Whittaker, C.A., Gygi, S.P., Dunham, M.J., and Amon, A. (2010). Identification of aneuploidy-tolerating mutations. Cell *143*, 71–83.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916–924.

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. *44*, D447–D456.

Walther, D.M., Kasturi, P., Zheng, M., Pinkert, S., Vecchi, G., Ciryam, P., Morimoto, R.I., Dobson, C.M., Vendruscolo, M., Mann, M., et al. (2015). Widespread Proteome Remodeling and Aggregation in Aging C. elegans. Cell *161*, 919–932.

Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T., et al. (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics *11*, 2019–2026.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Chapter 3: Conclusions and Future Directions

Understanding the effects of aneuploidy on cellular physiology is of critical importance for human health because it causes Down syndrome and is prevalent in cancer. The development of stable models of aneuploidy with many different karyotypes in multiple organisms has led to the discovery of phenotypes that are common among aneuploid cells, the vast majority of which confer fitness penalties (reviewed in (Santaguida and Amon, 2015)). Advances in quantitative proteomics have allowed for accurate measurements of protein levels in aneuploid cells relative to euploid controls, revealing that an imbalanced genome results in a proportionally imbalanced proteome (Dephoure et al., 2014; Pavelka et al., 2010; Torres et al., 2010). Protein quality control mechanisms must cope with the resultant increased demand to fold and degrade proteins. In both yeast and mammals, exhaustion of chaperones and the ubiquitin proteasome system cause decreases in proliferation (Donnelly et al., 2014; Oromendia et al., 2012; Torres et al., 2010). Understanding how an imbalanced proteome causes phenotypes observed in aneuploid cells is crucial for resolving why aneuploidy causes fitness penalties in most contexts besides cancer, where it is generally associated with more aggressive disease. Aneuploidy further represents an ideal model for studying how cells cope with imbalances in their proteome as it allows for interrogation into the fate of hundreds to thousands of imbalanced proteins simultaneously. This is particularly true in disomic yeast where 13 out of 16 disomies were created making it possible to examine the fate of nearly every yeast protein when it is duplicated.

Stoichiometric imbalance causes protein aggregation

In Chapter 2 of this thesis I have described our study in which we systematically identified proteins that are enriched in aggregates isolated from aneuploid cells. A small number of proteins aggregate in multiple aneuploid lines. Such proteins are likely aggregation prone, an

idea that is supported by the fact that they also aggregate in euploid cells with defective folding or degradative capacity. However, the proteins most enriched in aneuploid cell aggregates are encoded by the duplicated chromosome. Using stringent cut offs, we estimated that ~9% of yeast proteins become significantly enriched in aggregates when duplicated. Interestingly, nearly half of these aggregate-enriched proteins are known subunits of heteromeric protein complexes. With a more complete list of protein complexes, I anticipate that an even greater number of these proteins are in fact subunits of complexes. This supports the hypothesis that stoichiometric imbalance causes protein aggregation observed in aneuploid cells. This hypothesis predicts that restoring the stoichiometry of the complex should inhibit aggregation of the excess subunit. We demonstrated that this was indeed the case for Gcd11, as doubling the expression of Gcd11's two binding partners, Sui2 and Sui3, prevented Gcd11 from aggregating when encoded in excess in disome V cells. Thus, stoichiometric imbalance of protein complexes causes protein aggregation. This explains why protein aggregation occurs as a result of both chromosome gain and chromosome loss (Beach, 2016; Oromendia et al., 2012). In light of this, I speculate that this is the major cause of proteotoxic stress and associated proliferation defects observed in aneuploid cells.

The fate of excess protein: to aggregate or to degrade

It was previously observed that ~20% of proteins that are duplicated in disomic yeast fail to accumulate to their expected doubled levels because some fraction of the excess protein is degraded (Dephoure et al., 2014). Intriguingly, like aggregated proteins, these degraded proteins are enriched for subunits of protein complexes. This prompted us to examine whether the same proteins are aggregated and degraded. This was not the case, as aggregation and degradation

were mutually exclusive for most proteins. We believe that this principle is generalizable, and the propensity to aggregate or be degraded is an intrinsic property of proteins when they are in excess. In support of this, proteins within the same complex can have different fates when in excess, e.g. two subunits of the origin recognition complex, Orc2 which is degraded, and Orc5 which aggregates. Using bioinformatic approaches, we were unable to determine what properties (for example abundance, size, amino acid composition, disorder, etc.) predisposes certain proteins to aggregation and others to degradation. By measuring the half-life of excess Orc2 and Orc5, we observed that the aggregated subunit is more stable. Therefore, the characteristic that ultimately determines whether a protein is aggregated or degraded may simply be whether the excess subunit is a good substrate for the ubiquitin proteasome system (UPS) or not. We propose that aggregation is the default fate for excess proteins, but many proteins are readily recognized by the UPS and degraded before they have a chance to aggregate. This is supported by our observation that proteins that are normally degraded when in excess aggregate when the proteasome is defective.

What determines why some proteins are better proteasome substrates when in excess than others is still unclear. Both yeast and mammals contain ubiquitin conjugation machinery that recognize unassembled subunits (Sung et al., 2016; Xu et al., 2016; Yanagitani et al., 2017). For example, a ubiquitin ligase called Tom1 in yeast mediates the degradation of unassembled ribosomal proteins by ubiquitinating residues that are not accessible in fully assembled ribosomes (Sung et al., 2016). It will be interesting to determine what the molecular basis for substrate selection is beyond just abnormally exposed residues since not all unassembled protein complex subunits are recognized by these ubiquitin ligases. An additional factor may derive from features of co-translational complex assembly. It was observed that some protein subunits act as

scaffolds for complex assembly by localizing to the site of translation of their binding partners (Shiber et al., 2018). Perhaps scaffold subunits have evolved to be more stable, allowing sufficient time for them to exist in an unbound state in order to translocate to the site of translation of their binding partner. In this scenario, I would predict that when scaffolds are produced in excess, they would be likely to aggregate. Subunits that do not act as scaffolds could be rapidly degraded when in excess since they should never exist in the absence of a scaffold. Doing such an analysis would require a more comprehensive dataset on the directionality of complex formation to identify which subunits function as scaffolds.

Subunits that are neither aggregated nor degraded

While much of my work focused on proteins that are aggregated or degraded, it is worth noting that 27% of complex subunits that are in excess are neither aggregated nor degraded using the thresholds we established (Fig. 1). How these subunits escape degradation yet remain soluble is an important question that may provide insight into the properties that govern aggregation and degradation. We were able to uncover one property by using computationally determined protein-protein interface size calculations (Marsh et al., 2013). Protein complex subunits that are aggregated or degraded have significantly larger interfaces with their binding partners than proteins that are neither aggregated nor degraded. Thus, the exposed interface of excess subunits is likely to be recognized by the UPS or mediate aggregation by forming inappropriate interactions with other proteins. Excess proteins with smaller interfaces are not readily degraded and not aggregation prone. Lacking a binding partner, these proteins may constantly rely on chaperones to help maintain their solubility. This could explain the observation that aneuploid cells have decreased Hsp90 activity (Donnelly et al., 2014; Oromendia et al., 2012).

It is also possible for these unbound subunits to engage in ectopic interactions. Theoretically, excess protein can partake in novel interactions that rarely occur in euploid cells, modulating the function of the new binding partner or taking on an entirely new function. Such interactions have been proposed as a possible consequence of the imbalance in protein levels in aneuploid cells but never directly observed (Sheltzer and Amon, 2011). This work demonstrates that many proteins are in fact soluble and stable when in excess and so have the potential to participate in ectopic interactions. These proteins could serve as a starting point for determining whether these interactions occur and how they affect the physiology of aneuploid cells.

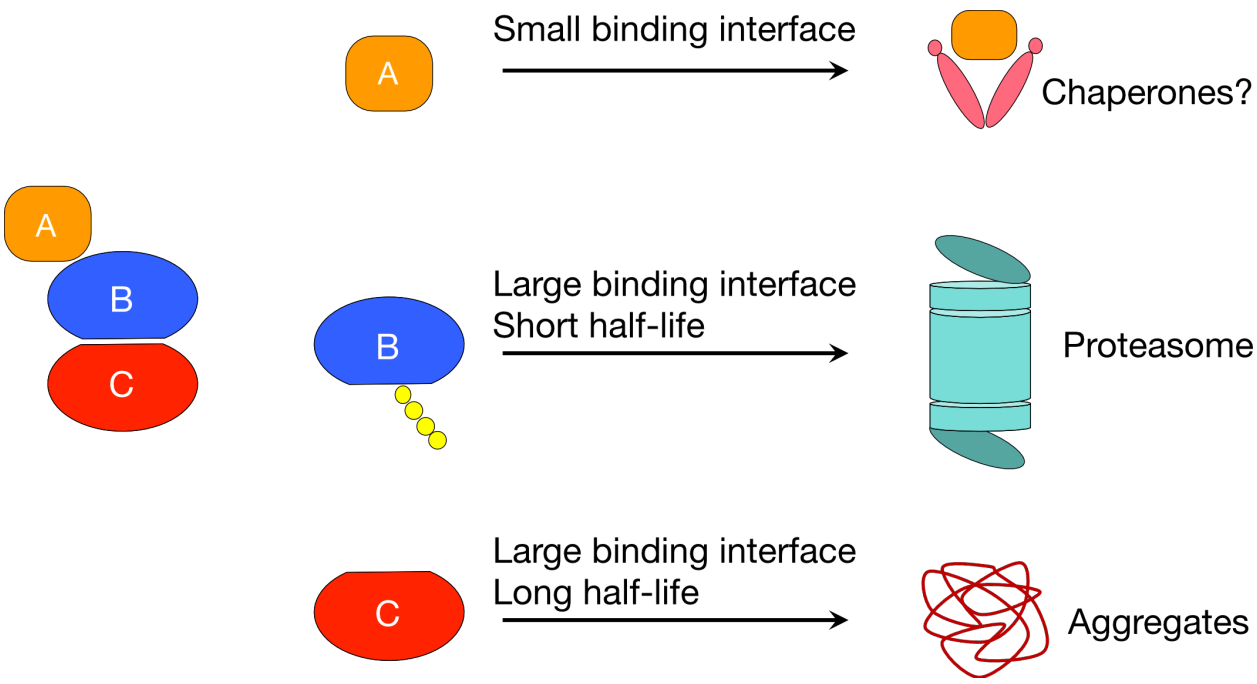

Figure 1. The fate of excess subunits of protein complexes

For a given heteromeric protein complex, A-B-C, each subunit can have one of three distinct fates when it is present in excess relative to its binding partners. In this example, subunit A represents a member of the complex with a smaller protein binding interface. These types of subunits are often soluble and not degraded. It is possible that chaperones may help to maintain

their solubility, but they have the capacity to interact with other proteins in the cell. Subunit B and C are core members of the complex, both containing large binding interfaces that become exposed when in excess. These types of subunits are most often degraded or aggregated, but rarely both. The molecular features that distinguish degraded subunits from aggregated subunits is still unknown. In this example, subunit B is a good substrate for the ubiquitin proteasome system resulting in its degradation while subunit C is not leading to its aggregation.

Protein aggregation mediates protein complex stoichiometry

Degradation of excess protein in aneuploid cells functionally performs dosage compensation by lowering protein levels to near their levels seen in a euploid cell. Having established that degradation and aggregation are unique fates for excess proteins, we investigated whether aggregation could also perform dosage compensation by rendering excess protein insoluble. To do this, we compared the relative amount of excess protein in total lysates (soluble fraction plus aggregates) to the relative amount of excess protein in the soluble fraction alone. Remarkably, the amount of excess protein in the soluble fraction was substantially lower than in the total lysates, indicating that aggregation can perform dosage compensation. For the duplicated proteins that we examined, 12% had nearly half of their excess protein depleted by aggregation. By comparison, 18% of proteins are depleted by exclusively degradation to the same degree. Thus, degradation is a more common mechanism of dosage compensation, but aggregation plays a much larger role than anticipated. This surprising result raises several questions about the role of aggregation in protein quality control, both in aneuploid and euploid cells that I will discuss below.

Is aggregation protective in aneuploid cells?

The most important question raised by my work is whether aggregation is protective for aneuploid cells, and for euploid cells that experience transient imbalances. Proteins that are in excess in the cytoplasm with large, exposed protein-protein interfaces present problems for the cell as discussed in the introduction to this thesis, particularly their ability to interfere with the folding of nascent chains. Since increasing the cells degradative capacity by deleting *UBP6* benefits many aneuploid cells (Torres et al., 2010), we reason that increasing the capacity to perform dosage compensation by aggregation should also be beneficial. It is difficult to imagine a mutation or chemical treatment that would enhance the cells ability to specifically aggregate only excess proteins and not affect the rest of the proteome. Perhaps deletion of *HSP104* would prevent cells from resolubilizing aggregated excess protein that is doomed to aggregate again without a binding partner. Since Hsp104 consumes ATP, this may also have an indirect effect of ameliorating the energy stress experienced by aneuploid cells (Torres et al., 2007; Williams et al., 2008). Such a result would be interesting but also complicates any interpretation that aggregation is beneficial to aneuploid cells. Further, it is unclear whether any benefits would outweigh penalties caused by loss of function of *HSP104*.

Aggregation appears to play a beneficial role in other contexts, challenging the notion that aggregates are toxic for cells. In amyloid diseases, it is becoming evident that the fibrillar form of the protein that was once considered the cause of disease due to its appearance in autopsies, is in fact protective against the soluble, toxic, oligomeric form (Caughey and Lansbury, 2003). In *C. elegans*, a mutation that increases longevity also increases the formation of chaperone-containing aggregates, suggesting that aggregation may protect cells against the toxic effects of proteins that aberrantly increase in levels during aging (Walther et al., 2015).

Finally, the fact that cells employ small heat shock proteins (sHSPs) to facilitate aggregation at specific sites within the cell points to aggregation as a regulated branch of the protein quality control network (reviewed in (Miller et al., 2015)). Another potential way to address whether protein aggregation is protective for aneuploid cells is to delete these sHSPs. The prediction being that deletion of sHSPs should prevent the formation of chaperone-containing aggregates and also decrease their proliferation.

Current evidence does not suggest that the aggregates observed in aneuploid cells are toxic. While all disomes contain protein aggregates as measured by Hsp104 focus formation, the percent of cells with aggregates within a strain does not scale with its proliferation delay (Oromendia et al., 2012; Torres et al., 2007). For example, disome IV has the worst proliferation defect but among the fewest percentage of cells containing Hsp104 foci. This is also true when estimating aggregate burden using mass spec (Chapter 2 of this thesis), e.g. disome XV has very little highly insoluble protein by this method but among the slowest doubling times. In conclusion, I propose that aggregations function as a mechanism of dosage compensation that is protective the way that degradation is. However, further experiments that modulate the capacity of cells to perform dosage compensation by aggregation will be required to directly demonstrate its benefit to cell physiology.

How do cancer cells cope with proteotoxic stress?

Given that cancer cells are highly aneuploid, a major question is how they either evade or cope with the proteotoxic stress inherent to the aneuploid state in order to maintain their proliferative advantage. Since stoichiometric imbalance of protein complexes is a major source of proteotoxic stress, perhaps tumor evolution favors karyotypes that minimize imbalances. Such

a mechanism could avoid chromosome gains or losses that result in the excess production of particularly deadly subunits such as β-tubulin. It could also result in co-gain or co-loss of particular chromosomes or pieces of chromosomes such that stoichiometry of complexes is maintained. Evidence for this type of selection exists (Beroukhim et al., 2010; Davoli et al., 2013; Knouse et al., 2017; Taylor et al., 2018; Zack et al., 2013), however the complexity of cancer and the large number of genes affected by chromosomal changes makes it difficult to assess the role that stoichiometric imbalance has on selective pressure. Perhaps the most compelling piece of evidence is that, in The Cancer Genome Atlas, chromosomes with more genes are lost at lower frequencies (Knouse et al., 2017). This supports the idea that large amounts of imbalance are not well tolerated even by cancer cells. Beyond this, it is also true that specific chromosomes (e.g. chromosome 4) are rarely gained while others are rarely lost (such as 7 and 20), but these preferences are more likely driven by the presence of tumor suppressors and oncogenes (Knouse et al., 2017).

While it is possible that cancer karyotypes avoid imbalances, some amount of imbalance is inevitable if they become aneuploid and cells must cope with the ensuing challenge to their protein quality control network. One of the most frequently observed genomic alterations in cancer is the gain of the q arm of chromosome 8 (Beroukhim et al., 2010; Davoli et al., 2013). This region of the genome contains the oncogene Myc, but also notably the transcription factor HSF1. HSF1 transcribes HSP70 and HSP90 chaperones as well as other co-chaperones to aid in protein folding (Solís et al., 2016). Indeed, it has been shown that deletion of HSF1 prevents cancer development because cancer cells are much more dependent on HSF1 activity for their proliferation and survival than untransformed cells (Dai et al., 2007; Jin et al., 2011). Further, increasing expression levels of HSF1 causes resistance to HSP90 inhibitors and improves

proliferation of human colorectal cancer cell lines (HCT116) that had been made aneuploid by microcell mediated chromosome transfer (Donnelly and Storchová, 2015). Perhaps one reason why chromosome 8q is gained so frequently in cancer is to increase expression of HSF1 targets that help to manage stoichiometric imbalances. Interestingly, the gain of chromosome 8q results in cells containing an extra copy of Myc and HSF1, thus increasing their expression simultaneously. It is possible that increased HSF1 activity can help cope with the increase in protein production that occurs due to increased Myc activity (Hsieh et al., 2015).

Finally, it is tempting to speculate that dosage compensation by aggregation could help cancer cells cope with proteome imbalances. A first step in addressing this would be to assess whether highly aneuploid, but rapidly proliferating cancer cells contain high levels of protein aggregates. Taken together with more evidence that dosage compensation by aggregation is in fact cytoprotective, this would argue that aggregation may allow these highly aneuploid cells to avoid complications of proteome imbalance. If transformation enhances cells ability to utilize dosage compensation by aggregation, a prediction of this hypothesis is that aneuploid cancer cells should have more aggregated protein than karyotype-matched untransformed cells. One possible way to create such a pair of cell lines would be to transform trisomic mouse embryonic fibroblasts (MEFs), and then compare their aggregation levels with the untransformed line.

Does aggregation commonly control protein stoichiometry in euploid cells?

Using aneuploidy as a model has uncovered protein degradation and aggregation as two mechanisms to cope with imbalances in the proteome. The fact that ubiquitin ligases exist that target unassembled subunits for degradation argues that even euploid cells encounter imbalances and employ degradation as a strategy to control subunit stoichiometry (Sung et al., 2016; Xu et

al., 2016; Yanagitani et al., 2017). Further, since individual proteins tend to have one fate or the other when in excess, it stands to reason that proteins that are not degraded when in excess could be controlled instead by aggregation in euploid cells.

In yeast, the majority of stoichiometry control takes place at the level protein synthesis. Ribosome profiling experiments have demonstrated that synthesis rates of obligate members of protein complexes are proportional to their stoichiometry in the complex (Li et al., 2014; Taggart and Li, 2018). For example, the α-ketoglutarate dehydrogenase complex is composed of one molecule of Kgd1 and two molecules of Kgd2. Kgd2 is synthesized at double the rate of Kgd1 as would be predicted if synthesis is proportional to stoichiometry (Li et al., 2014). While exceptions to this rule are rare, subunits that are produced in excess tend to have shorter half-lives than subunits that are synthesized proportionally, suggesting that dosage compensation by degradation plays a role in tuning stoichiometry in these cases (Taggart and Li, 2018). It would be interesting to examine whether any of these subunits produced in excess that do not have shorter half-lives are dosage compensated by aggregation.

A larger set of exceptions to proportional synthesis exists during meiosis. A study in which mRNA, translation, and protein were measured in meiotic yeast found that a majority of protein complexes exhibit imprecise synthesis of complex subunits, while protein levels reflect the stoichiometry of the subunits (Eisenberg et al., 2018). This indicates that under some conditions, protein degradation controls protein stoichiometry in euploid cells. Additionally, a study utilizing pulse-chase labeling and mass spectrometry in mammalian cells found that ~10% of proteins have much shorter half-lives immediately after being synthesized compared to when they are mature (McShane et al., 2016). These proteins were enriched for subunits of protein complexes that the authors demonstrated were produced in excess relative to their

stoichiometries. Thus, in higher eukaryotes, degradation acts to control dosage. It would be interesting to compare proteins that are aggregated or degraded in disomes to subunits that have been demonstrated to be synthesized in excess under normal, steady-state conditions. For proteins that are synthesized in excess yet do not appear to be degraded, considering dosage compensation by aggregation could provide a more complete picture of stoichiometry control. It is also possible that aggregation prone subunits are the ones that are proportionally synthesized, thus avoiding the use of aggregation as a widespread mechanism for tuning protein stoichiometries under steady-state conditions.

The above studies were conducted on unperturbed populations of cells. Cells may encounter conditions that cause transient imbalances even for proportionally synthesized subunits. Such imbalances could occur as the result of changes in gene expression or protein mistargeting. Many of these conditions result in activation of stress responses that increase protein degradation via the UPS or autophagy (Harper and Bennett, 2016). It would be interesting to assess if perturbations that result in transient proteomic imbalance, such as increased gene expression induced by the oncogene, Myc, or disruptions in mitochondrial import also lead to increased protein aggregation as this could provide evidence that dosage compensation by aggregation occurs in euploid cells as well.

In conclusion, at least in yeast, protein stoichiometries of obligate complex subunits are controlled primarily by proportional synthesis. In cases where they are not, or where proportional synthesis is disrupted, they can be controlled by degradation. I propose that many proteins that are not controlled by degradation are controlled by aggregation, however future experiments will be needed to test this once methods to disrupt or enhance aggregation are established.

Concluding Remarks

The imbalances in the proteome induced by aneuploidy represent a distinct challenge to the protein quality control machinery. In this thesis I have demonstrated that stoichiometric imbalances of protein complexes cause proteotoxic stress in aneuploid cells. I have also utilized aneuploidy as a model to determine that excess proteins are dosage compensated by either degradation or aggregation. I hope that this work not only informs our comprehension of aneuploid cell physiology, but also provides a more complete understanding of how aneuploid and euploid cells cope with stoichiometric imbalances, namely that protein aggregation can function as protein quality control mechanism in this regard.

References

Beach, R.R. (2016). Insights into the consequences of chromosome gains and losses in *S. cerevisiae*. Massachusetts Institute of Technology.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Caughey, B., and Lansbury, P.T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu. Rev. Neurosci. *26*, 267–298.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948–962.

Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Elife *3*, e03023.

Donnelly, N., and Storchová, Z. (2015). Causes and consequences of protein folding stress in aneuploid cells. Cell Cycle *14*, 495–501.

Donnelly, N., Passerini, V., Dürrbaum, M., Stingele, S., and Storchová, Z. (2014). HSF1 deficiency and impaired HSP90-dependent protein folding are hallmarks of aneuploid human cells. The EMBO Journal *33*, 2374–2387.

Eisenberg, A.R., Higdon, A., Keskin, A., Hodapp, S., Jovanovic, M., and Brar, G.A. (2018). Precise Post-translational Tuning Occurs for Most Protein Complex Components during Meiosis. Cell Rep *25*, 3603–3617.e2.

Harper, J.W., and Bennett, E.J. (2016). Proteome complexity and the forces that drive proteome imbalance. Nature *537*, 328–338.

Hsieh, A.L., Walton, Z.E., Altman, B.J., Stine, Z.E., and Dang, C.V. (2015). MYC and metabolism on the path to cancer. Semin. Cell Dev. Biol. *43*, 11–21.

Jin, X., Moskophidis, D., and Mivechi, N.F. (2011). Heat shock transcription factor 1 is a key determinant of HCC development by regulating hepatic steatosis and metabolic syndrome. Cell Metab. *14*, 91–103.

Knouse, K.A., Davoli, T., Elledge, S.J., and Amon, A. (2017). Aneuploidy in Cancer: Seq-ing Answers to Old Questions. Annu. Rev. Cancer Biol. *1*, 335–354.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Marsh, J.A., Hernández, H., Hall, Z., Ahnert, S.E., Perica, T., Robinson, C.V., and Teichmann, S.A. (2013). Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. Cell *153*, 461–470.

McShane, E., Sin, C., Zauber, H., Wells, J.N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchová, Z., Marsh, J.A., et al. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. Cell *167*, 803–815.e821.

Miller, S.B.M., Ho, C.-T., Winkler, J., Khokhrina, M., Neuner, A., Mohamed, M.Y.H., Guilbride, D.L., Richter, K., Lisby, M., Schiebel, E., et al. (2015). Compartment-specific aggregases direct distinct nuclear and cytoplasmic aggregate deposition. The EMBO Journal *34*, 778–797.

Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. Genes & Development *26*, 2696–2708.

Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., and Li, R. (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature *468*, 321–325.

Santaguida, S., and Amon, A. (2015). Short- and long-term effects of chromosome mis-segregation and aneuploidy. Nature Publishing Group *16*, 473–485.

Sheltzer, J.M., and Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends Genet. *27*, 446–453.

Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., and Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. Nature *561*, 268–272.

Solís, E.J., Pandey, J.P., Zheng, X., Jin, D.X., Gupta, P.B., Airoldi, E.M., Pincus, D., and Denic, V. (2016). Defining the Essential Function of Yeast Hsf1 Reveals a Compact Transcriptional Program for Maintaining Eukaryotic Proteostasis. Molecular Cell *63*, 60–71.

Sung, M.-K., Porras-Yakushi, T.R., Reitsma, J.M., Huber, F.M., Sweredoski, M.J., Hoelz, A., Hess, S., and Deshaies, R.J. (2016). A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. Elife *5*, 3429.

Taggart, J.C., and Li, G.-W. (2018). Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Syst *7*, 580–589.e584.

Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell *33*, 676–689.e3.

Torres, E.M., Dephoure, N., Panneerselvam, A., Tucker, C.M., Whittaker, C.A., Gygi, S.P., Dunham, M.J., and Amon, A. (2010). Identification of aneuploidy-tolerating mutations. Cell *143*, 71–83.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916–924.

Walther, D.M., Kasturi, P., Zheng, M., Pinkert, S., Vecchi, G., Ciryam, P., Morimoto, R.I., Dobson, C.M., Vendruscolo, M., Mann, M., et al. (2015). Widespread Proteome Remodeling and Aggregation in Aging C. elegans. Cell *161*, 919–932.

Williams, B.R., Prabhu, V.R., Hunter, K.E., Glazier, C.M., Whittaker, C.A., Housman, D.E., and Amon, A. (2008). Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science *322*, 703–709.

Xu, Y., Anderson, D.E., and Ye, Y. (2016). The HECT domain ubiquitin ligase HUWE1 targets unassembled soluble proteins for degradation. Nature Publishing Group *2*, 16040.

Yanagitani, K., Juszkiewicz, S., and Hegde, R.S. (2017). UBE2O is a quality control factor for orphans of multiprotein complexes. Science *357*, 472–475.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.-Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.