Uncovering Genes Underlying Quantitative Trait Variation Using S. cerevisiae as a Model System

Uncovering genes underlying quantitative trait variation using S. cerevisiae as a model system

A dissertation presented

by

Bo Hua

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Systems Biology

Harvard University
Cambridge, Massachusetts

March 2017

Uncovering genes underlying quantitative trait variation using S. cerevisiae as a model system

Abstract

Individuals from the same species often differ quantitatively in many phenotypes. Understanding the genetic causes of such individuality may improve our understanding of the corresponding phenotypes, and our ability to make models to predict phenotypes using individual's genotype. In this thesis, I would like to present two works related to this question. In Chapter 2, I used a few highly sensitive experimental systems to ask, what are "all" the genes that can affect a given quantitative trait. By analyzing these results, I found that a large number of yeast genes can affect four studied traits, namely two galactose response traits, unfolded protein response and growth rate in rich medium. These genes consist of a few large effect-size genes that are typically directly involved in the process related to the quantitative trait, and a large number of small effect-size genes that are enriched in a number of core cellular processes. This implies that genetic variation in one process has the potential to influence behaviors in seemingly unconnected processes, and a considerable proportion of trait variation in natural populations may be caused by the cumulative effects of many small effect-size genetic variants. The core processes that are discovered in Chapter 2 can be affected by many types of genetic changes. In Chapter 3, in collaboration with Angelika Amon lab, I followed a specific type of mutation, aneuploidy, and experimentally measured the effects of aneuploidy-associated stresses on three different yeast traits, namely galactose response, DTT response and heat shock response. The

results showed that, aneuploidy can generally increase cell-to-cell variability in an isogenic population. In Chapter 4, I will present a few potentially interesting directions. Overall the results presented in this thesis improved our understanding about potential source of quantitative trait variation.

Table of Contents

Acknowledge

This thesis would not be possible without the help from my Ph.D. advisor, friends, and colleagues. I would like to express my sincere thanks here.

First and foremost, I would like to thank my Ph.D. advisor professor Michael Springer. Mike was literally the person who led me into the field of experimental biology. Beyond experimental techniques, I learned from him how to conduct research, from generating ideas to finishing a paper. I am also deeply impressed by his constant passion and interest in science and education. In addition, I am grateful for his patience to allow me to try different things in the lab, one of which fortunately turned out to be the main chapter of this thesis. It is also the many discussions with Mike that help me shape my own views about science and technology.

Mike also cares about creating a healthy lab environment, and thank for that, I benefited tremendously from the people in the Springer lab. I would like to thank Jue Wang for being a great friend to chat about science and being critical and sharp on many of our discussions. I would like to thank Sarah Boswell and Christine DeGennaro, both of whom are great experimental experts in the lab and are patient enough to allow me repeatedly bothering them about details of experiment protocols. I would like to thank Yonatan Savir, who always reminds me thinking biology in the eyes of a physicist. I would like to thank two former post-docs in the lab, Jacky Jui-yu Chou and Ilan Wapinski – both are not just colleagues but also friends, showing me the life outside of the lab. I want to thank Yarden Katz, who introduced me many great AI ideas and tools. I would like to thank Chiara Ricci-Tam, Kayla Lee, and Renan Escalante-Chong,

who made the Springer lab a nicer place to work in.

As a graduate student who has never stepped a step in the US before Ph.D., I would have a lot more trouble if I did not have great friends around me. I would like to thank my classmates in the Systems Biology program, for their constant and inspiring love for both life and science. I would like to thank the Systems Biology Department, for its extremely friendly environments.

Last but not the least, I would like to thank my parents and parents-in-law for their constant support. I would like to thank my wife, Mingcong Ma, who I met and married with during my Ph.D. study. Without her great sacrifice and countless love, this thesis would not be possible.

Chapter 1. Introduction

Although individuals within the same species share highly similar genetic material, it is clear that they also vary substantially in many phenotypes. Such phenotypes include morphological traits, cellular responses to stress, propensity to various diseases and etc. Because such variation is often quantitative, rather than qualitative, the corresponding phenotypes are called quantitative traits. Understanding the pattern, causes and consequences of this variation is an important task to biologists.

In this thesis, I would like to present two works related to this topic - one as my main Ph.D. project and another one as a collaborative project. Before going into details of each work, I would like to present a short overview of this field to motivate the rest of the thesis, followed by an outline for the remaining chapters.

1.1 Quantitative trait variation as a hundred-year-old topic

While much of what we know today about quantitative trait variation, such as the identity of causal genetic variants, relies on modern technologies, many insights in this field can be traced back to century-old observations. One of the first lessons that researchers learned about quantitative traits is that traits do not vary randomly in a population. Such non-random patterns reflect on the phenotypic similarity among family members. For example, in 1886, Francis Galton described such a pattern in height as,

> *"... we define the law of regression very briefly. It is that the height-deviate of the offspring is, on the average, two-thirds of the height-deviate of its mid-*

parentage." [1]

Based on the phenotype-level characterization, the genetic cause of trait correlation among family members was later correctly proposed by Ronald Fisher, who attributed this pattern to the results of "a large number of Mendelian factors". In 1918, he wrote that,

> *" The simplest hypothesis ... is that such features as stature are determined by a large number of Mendelian factors, and that the large variance among children of the same parents is due to the segregation of those factors in respect to which the parents are heterozygous. " [2]*

The idea that there are genetic factors underlying quantitative trait variation in a population is further supported by a series of artificial selection experiments on various traits. For a given trait, selectively breeding individuals that exhibit extreme trait values led researchers and breeders to observe a continuous response to selection. The explanation is that the genetic variants that are related to the trait of interest, either de novo mutations or existing variants before selection, are favored in the selection process and cause a continuous response to selection. For example, for abdominal bristle number of *Drosophila melanogaster*, exerting artificial selection pressure on this trait produced the observation that,

> *"response continued for at least 75 generations and average total response was in excess of 36 additive genetic standard deviations of the base population ..." [3]*

All these theories and experiments lead to an analytic framework to decompose quantitative trait variation in terms of the contributions from genetic factors and environmental factors. In population genetics, heritability is often used to analyze the relative importance of these two factors and is broadly applied in many traits. Technically, heritability is defined as the proportion

of trait variation determined by genetic factors. A number of methods were used to quantify this trait-specific population-level metric [4]. For example, it is estimated that about 80% of human height variation is determined genetically, hence human height heritability is about 0.8. In this thesis, the main relevance of this number is that it sets a goal for future endeavors to identify causal genetic variants underlying quantitative traits. If the total contribution of all identified variants is close to the estimated heritability, one can conclude that we are close to identifying all the important variants for this trait in the studied population.

It is worthwhile to mention that many human diseases are heritable, and can be analyzed in the same framework as continuous traits. With a few exceptions, most human disease traits are binary, i.e. individuals are either affected or not affected. In 1965, Falconer proposed a liability model [5], in which individuals are assigned a 'liability' score based on the genotypes and environments. If liability score is above a certain threshold, the individual will be affected by the disease.

A lot of recent progress in understanding quantitative trait variants focused on locating the causal genetic variants for a given trait of interest. While much of the progress is empowered by the newly developed genotyping techniques, the idea to use genetic markers to locate causal variants is not new. As an experimental example, David Botstein et al. proposed a method in 1980, to use restriction fragment length as genetic marker [6] to detect potential linkage to human disease. From the theory side, Risch and Merikangas showed in 1996 that association studies should in principle be more powerful than family linkage studies to detect small effect polymorphisms. In the conclusion part of that work, they insightfully mentioned that,

> *"the primary limitation of genome-wide association tests is not a statistical one but a technological one. A large number of genes (up to 100,000) and polymorphisms ... must first be identified, and an extremely large number of such polymorphisms will need to be tested." [7]*

Such "an extremely large number" of polymorphisms indeed became possible to genotype in the next few years. After finishing sequencing the human genome, the pattern of common human genetic variants was mapped systematically. In the HapMap project[8], which aimed to develop a haplotype map of the human genome, “a block-like structure of linkage disequilibrium” was found. The “substantial correlation of SNPs” discovered in this project allowed using a relatively small fraction of common SNPs as genetic markers to locate causal genetic variants in human. There are enormous efforts to use common genetic variants to study human traits – by the end of 2008, 531 SNP-trait associations were discovered in 151 studies [9]; by the time of writing this thesis (Feb 2017), over thirty thousand of such association were identified in 2732 studies [10].

Despite such vast advances in techniques, our understanding of the source of quantitative trait variation is still far from complete, which is the main motivation of this thesis. As mentioned, heritability for a given trait sets a target of the total contribution of genetic factors that can be discovered to explain trait variation. However, when summing up the total contribution of identified SNPs, the explained fraction of trait variation is often much lower than the corresponding estimated heritability. Such problem is often referred as the “missing heritability” problem [11].

1.2 Outline of this thesis

This thesis is about two projects completed during my Ph.D. studies to study the genetic causes

of quantitative trait variation using yeast as a model system. The main project in my Ph.D. study is to understand what are the genes that possibly affect quantitative traits. The second project is a collaboration project, which studied how aneuploidy state can affect quantitative traits.

For a given trait, supposing that there is a gene list that contains all the genes that when mutated can affect the trait of interest, the causal genetic variants in a natural population that affect the quantitative trait must act on all or a subset of such genes. Although it is still unclear which subset of genes harbor causal genetic variants, studying such a gene list might shed lights on the "missing heritability" problem to clarify the target size of potential mutations. In Chapter two, we set up a highly sensitive assay to identify the most important genes that are capable of affecting two yeast quantitative traits related to galactose signaling. To aid the analysis, we also analyzed two additional traits that were previously studied in other labs. Together, the experiments and analysis suggest that a large number of genes can affect quantitative traits, and many of such genes might be related to core cellular processes.

Chapter 3 is a collaboration with the Angelika Amon lab. Previously, they found that aneuploidy-associated stress can increase the cell-to-cell variability of the cell cycle progress in an isogenic population, yet it was unclear if such stress can affect any other traits. To address this question, I studied the effects of adding additional chromosome on yeast galactose response, heat shock response, and unfolded protein response. The results will be presented in Chapter 3.

Reference

1. Galton F. Regression Towards Mediocrity in Hereditary Stature. The Journal of the Anthropological Institute of Great Britain and Ireland. 1886;15: 246. doi:10.2307/2841583

2. Fisher RA. XV.---The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. Cambridge Univ Press; 1919;52: 399–433.

3. Yoo BH. Long-term selection for a quantitative character in large replicate populations of Drosophila melanogaster: 1. Response to selection. Genet Res. Cambridge, UK: Cambridge University Press; 1980;35: 1–17.

4. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. Nat Rev Genet. Nature Publishing Group; 2013;14: 139–149. doi:10.1038/nrg3377

5. Falconer DS. The inheritance of liability to certain diseases. Ann Hum Genet. Wiley Online Library; 1965;29: 51–76.

6. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980;32: 314–331.

7. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273: 1516–1517. doi:10.1126/science.273.5281.1516

8. International HapMap Consortium. The International HapMap Project. Nature. 2003;426: 789–796. doi:10.1038/nature02168

9. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. National Academy of Sciences; 2009;106: 9362–9367.

10. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research. 2014;42: D1001–6. doi:10.1093/nar/gkt1229

11. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461: 747–753. doi:10.1038/nature08494

Chapter 2. Small effect-size mutations cumulatively affect yeast quantitative traits

Bo Hua[1,2] and Michael Springer[1,*]

[1] Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115

[2] Systems Biology Graduate Program, Harvard University, Cambridge, Massachusetts, United States of America

This chapter is adapted from a manuscript written by the authors listed above.

2.1 Introduction

What are all the genes that are involved in a trait? Classically, the pathways that contribute to a trait, like those involved in signaling or development, were defined by genetic screens that identified genes with loss- or gain-of-function phenotypes [1]. As screens became more quantitative, many alleles of both small and large effects size were identified [2,3]. But, the methods to validate and then determine the molecular function have remained laborious. Hence, research has typically focused on genes on characterizing genes with large effect size. This has led to a potential bias that these large effect size genes dominate the behavior and variability in a

pathway. An alternative view is that cumulatively, the mainly overlooked small effect size genes significantly shape pathway function and population-level trait variation, and hence the genetic architecture of a pathway is distributed not centralized (Figure 2.1). Until recently, it wasn't possible to easily and comprehensively identify genes implicated in quantitative traits, making it difficult to distinguish between these two hypotheses concerning the architecture of most pathways.

The genetic architecture of quantitative traits has taken on increased importance as it has become clear that many human traits, such as body mass index and traits that underlie heritable human disease, are also quantitative. Numerous human traits and disease have been studied using genome-wide association studies (GWAS) to uncover the loci containing causative variants that are responsible for the genetic component of these traits [4]. If the genetic architecture of the underlying pathway were centralized, one would expect GWAS would yield a small number of large effect-size genes typically of related function; if the genetic architecture of a quantitative trait were distributed, one would expect GWAS would yield a large number of small effect-size genes of often seemingly unrelated function. In some diseases, e.g. age-related macular degeneration (AMD), GWAS indeed identified several common alleles of large effect size that explain about half of the disease risk to siblings of affected individuals [5]. This would support the view of centralized signaling pathways. But, in many cases, GWAS has yielded many small effect-size variants with low odd ratios [4], and additionally many identified loci have not included genes with an obvious connection to disease [6,7]. These results are consistent with the hypothesis that the gene architecture of some pathways underlying human traits is distributed.

Direct experiments to separate between these two hypotheses can help frame our expectation for the results from these association studies.

Model organisms should be a powerful set of tools for defining the architecture of quantitative traits. Several studies in yeast [2,8] show that linkage analysis has the potential to identify most of the causative loci needed to explain trait variation between two natural yeast isolates. But these studies, like human GWAS, are limited by recombination block size and sample size, and hence are not ideal for identifying causative genes or the exact number and identity of small effect-size loci. As an alternative approach, deletion libraries have been used to assess the role of every yeast gene. These studies have been transformative for defining the function of unknown genes [9] and for showing that many processes in yeast are genetically interconnected [10,11]. While informative, the assays that are typically performed, e.g. colony size assay, are not quantitative enough to accurately determine the effect size of every mutant. Hence whether this interconnectedness has a significant role in pathway function is still unclear.

In this work, we quantified the effect sizes of all non-essential yeast genes on several traits. Instead of identifying existing genetic variation in natural populations, we used a yeast deletion library to measure with high precision the magnitude of effect of all non-essential genes on a quantitative trait, which we refer to as gene effect size. By its design, this approach identifies all the genes whose loss-of-function has the potential to influence a trait, and the effect size distribution of these genes. We found that all four traits we analyzed have an exponential distribution of effect sizes. The consequence of these results is that cumulatively, small effect-

size can significantly contribute to pathway function. Gene Ontology (GO) analysis and additional experiments showed that many of these small effect-size mutations are involved in core cellular processes and affect quantitative traits in a trait-specific, not generic, manner. In natural populations, phenotypic variation is influenced by the actual existing variants; this natural variation is more complex than our deletion library. We showed through simulation that our analysis based on deletion mutants, given modest assumptions, yields an effect size distribution that is close to the distribution that would be observed for other sources of genetic variation.

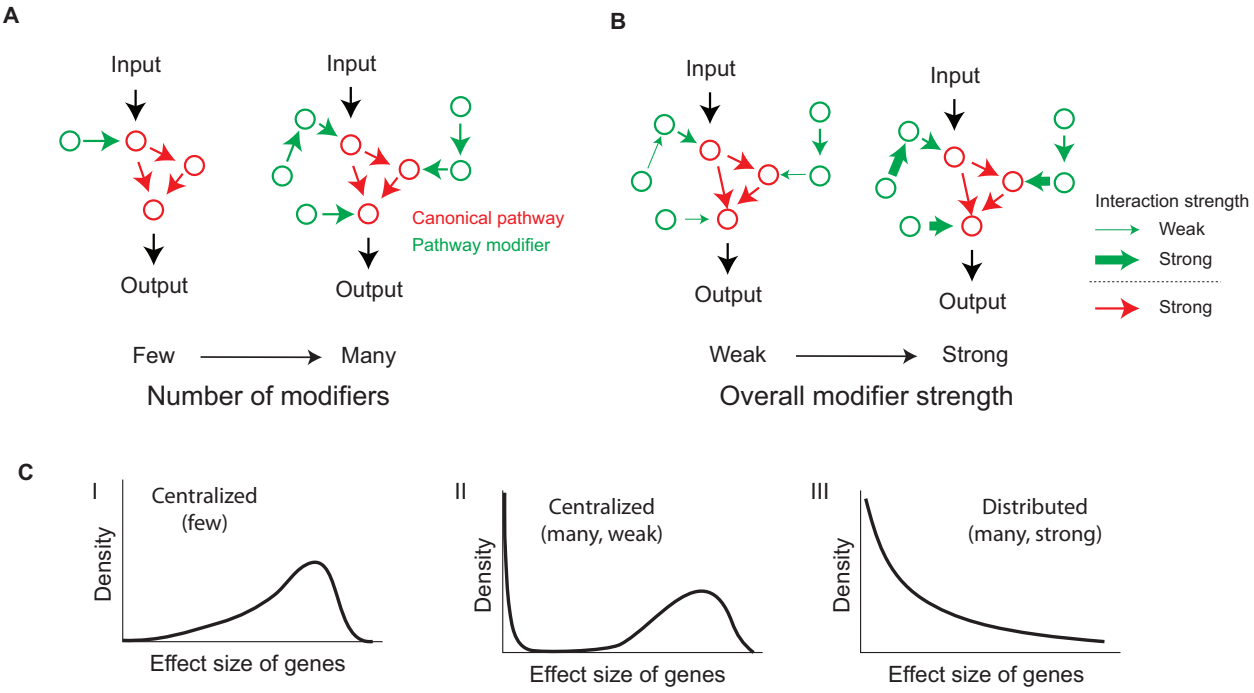


Figure 2.1 Genes outside of the canonical signaling pathways have the potential to substantially influence pathway function

(**A**) A canonical pathway (red circles) can be modified by anywhere from a small number to large number of currently unidentified genes (green circles). (**B**) Regardless of the number of modifiers, the modifiers could range from having a weak to strong effect on the pathway (represented by arrow thickness). (**C**) If the number of modifiers is small (I) or if the effect size of the modifiers is small (II) the genetic architecture of the pathway will be centralized, i.e. a small number of genes will control the function of and variation in the pathway. If the number of modifiers is large and the effect size of the modifiers is sufficiently large (III) the genetic architecture will be distributed; i.e. a large number of genes will control the function of and variation in the pathway.

2.2 A large fraction of genes can influence multiple biological processes

A large number of screens have been performed with the yeast deletion library [12]. These screens could potentially serve as a rich source of data for determining the effect size of each gene on many traits. Reanalyzing this data, we found that, due to measurement noise, most of these studies do not have the power to determine the full gene-level effect size distribution (See Appendix A). This is not surprising as the goal of most studies was to identify genes of large effect size rather than attempting to identify all genes of any effect size. Therefore, to determine the number of genes that can affect a pathway, we created a reporter library with which we could quantitatively measure the response of cells to galactose (GAL). We systematically constructed a library of strains deleted for all non-essential yeast genes each containing a YFP reporter driven by the GAL1 promoter (GAL1pr-YFP). We then assayed the bimodal YFP response [13,14] in

single cells growing in mixtures of glucose and galactose by flow cytometry (Figure 2.2 A-B). Additionally, to supplement the analysis, we identified and re-analyzed two deletion studies [15,16], one on growth rate in rich medium and one on the unfolded protein response (UPR), which had a signal-to-noise ratio that was sufficiently large to determine the effect size distribution.

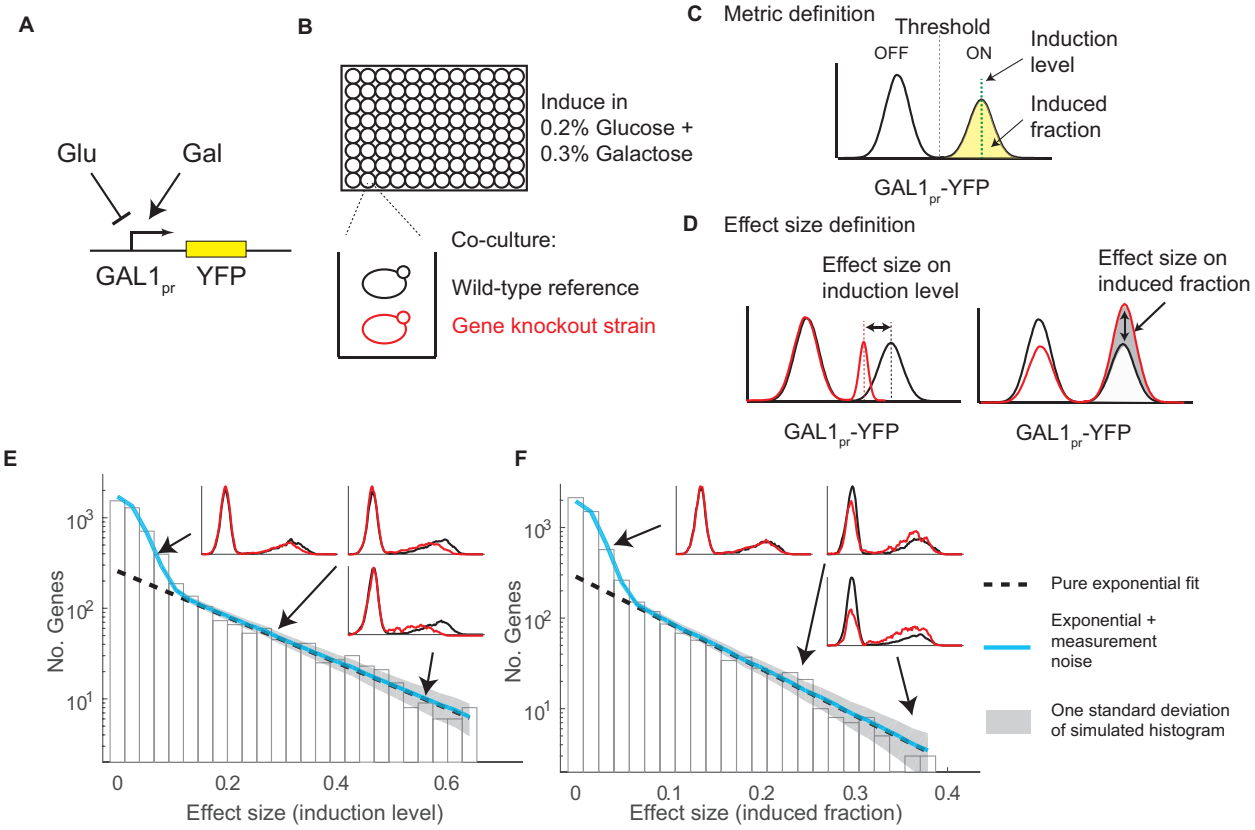


Figure 2.2 Quantitative genetic screen determines that a large number of genes quantitatively affects the yeast galactose response

(**A**) Galactose (Gal) activates while glucose (Glu) inhibits transcription from a GAL1 promoter

YFP fusion. **B**) A mCherry expressing mutant strain (red) was co-culture with a wild-type reference strain (black); both strains contained the reporter construct from **A**. Each well contained a distinct deletion mutant. (**C**) We defined two metrics to characterize the bimodal response of the GAL pathway. We defined the induced fraction (yellow area versus total area under the curve) as the percent of cells whose YFP expression level was above a threshold (black dotted line). We also defined the induction level as the mean YFP expression of all induced cells (green dotted line). (**D**) Mutant effect sizes for the induction level (**D**, left) and for induced fraction (**D**, right) are defined as the relative change in each metric between mutant (red) and the co-cultured wild-type reference strain (black). (**E**-**F**) Effect size distribution for two GAL traits. Effect sizes of all mutants were binned and plotted as a histogram (black bars). Mutant that passed a 0.5% false discovery rate cut-off were well fit with an exponential distribution using maximum likelihood estimation (dashed black line, R^2 =0.96 for each, see Method). The full distribution is parsimonious with a convolution of experimental noise and an exponential distribution (blue line is the average distribution of 100,000 simulations, R^2=0.92-0.98; gray shading is one standard deviation around the mean).

Principal component analysis of the results from our GAL response screen highlighted three distinct traits (Figure S2.1). These traits, corresponding to: 1) the fraction of cells that are induced above background; 2) the induction level of the induced ('on') peak; and 3) the background level of the uninduced ('off') peak (Figure 2.2C, Supplemental Information). The signal-to-noise ratio of the first two metrics was sufficient to calculate an effect size distribution for a large number of genes. We will refer to these two separable GAL traits as the "induced

fraction" and the "induction level" (Figure 2.2D).

Each of the four traits - the induced fraction, induction level, growth rate, and UPR - considered in isolation, was influenced by a large number of deletion strains (Figure 2.2E and F); the distribution of mutant effects was continuous. Based on a comparison of the measured effect sizes and the measurement noise estimated from biological replicates, 19% (796 of 4201), 16% (735 of 4562), 16% (689 of 4162), and 20% (849 of 4162) of non-essential genes screened, at a 0.5% false discovery rate, affect the growth rate in rich media, unfolded protein response, induced fraction in GAL, and induction level in GAL respectively. Together the two GAL traits are composed of 1104 unique genes. Interestingly, if we used a single composite trait, i.e. mean expression, to quantify the GAL response, fewer genes (593 of 4162) were identified, highlighting the utility of sub-classifying higher-level phenotypes that might be composed of separable traits each controlled by distinct genetic factors (Appendix A). To obtain a more accurate estimate of how many genes can quantitatively affect each of the traits, at the sacrifice of knowing the identity of the genes, we determined the area of the normalized effect size distribution that is outside the normalized measurement noise distribution (Figure S2.2). From this, we estimate that the fraction of genes affecting the growth rate in rich media is 62%, unfolded protein response is 23%, induced fraction in GAL is 28%, and induction level in GAL is 34% (Appendix A). Together these results highlight that a large fraction of the protein-coding genes has the potential to quantitatively affect a trait.

As a final method to determine the number of genes that influence our four traits we determined whether the effect size distributions could be explained by a simply analytical function. To

minimize the effect of measurement noise on measured effect sizes, we first focused our analysis on the genes whose effect size was significantly different from measurement noise. Interestingly, we found that the effect size distribution for all four traits was well fit by an exponential distribution (R^2=0.91-0.96, Figure 2.2E an F, dotted line). When extrapolating the exponential fit into the measurement noise, it predicts that 27-33% of genes affect each of our four traits, similar to the orthogonal estimates above. Adding measurement noise to the exponential distribution (Figure 2.2E and F, blue line) well fit the full measurement distribution (R^2=0.92-0.98). Therefore, a parsimonious explanation of our data is that the effect size distribution of a quarter to half of genes is exponential. Half to three quarters of all genes have little to no effect.

2.3 Small effect-size genes can influence pathway function

The shape of the determined effect size distributions implies that each of the four traits is affected by genes with a continuous distribution of effect sizes ranging from a small number of large effect-size genes to a large number of small effect-size genes. It has been questioned whether even such a large number of small effect-size mutants could substantially contribute to the functionality of a pathway [17]. The answer to this question depends on the exact shape of the measured effect size distribution (e.g. Figure 2.1C II versus III). We therefore determined the number of genes that are cumulatively important for pathway function. To do so, we devised a method to quantify the impact of each gene, which is similar to the one used to quantify allelic contribution to narrow-sense heritability in a GWAS [18]. In the calculation, we first assumed a population of cells with independent and randomly assorting alleles. We assumed only two

possible alleles for each gene, i.e. deletion or wild-type (a more complex model will be considered below). We then calculated each gene's contribution to the trait variation in the population as $2\beta^2 f(1-f)$ [18], where β is the effect size and f is the allele frequency, assuming that each allele has a frequency of 50% and no epistasis (Figure 2.3A). For our four traits, using the measured effect size for each gene, we find that 257-352 genes with the largest effect sizes, representing 5.6-8.5% of screened genes, are needed to explain 80% of total computed variation (Figure 2.3B and C, and Figure S2.4). If human traits behave similarly to our yeast deletions, we would estimate that the number of genes required to explain most of the heritability of a quantitative trait is in the range of 1200-1900 genes. Interestingly, our estimate is concordant with estimations from GWAS. For example, the current estimate for human height, the best characterized human trait, is that 423 1Mb loci are involved. Yet this explains only 20% of the heritability. This result suggests that both in yeast and humans, some pathways and traits resemble the distributed architecture from Figure 2.1C III; i.e. a large number of genes of slowly diminishing effect size contribute to pathway function and trait variation.

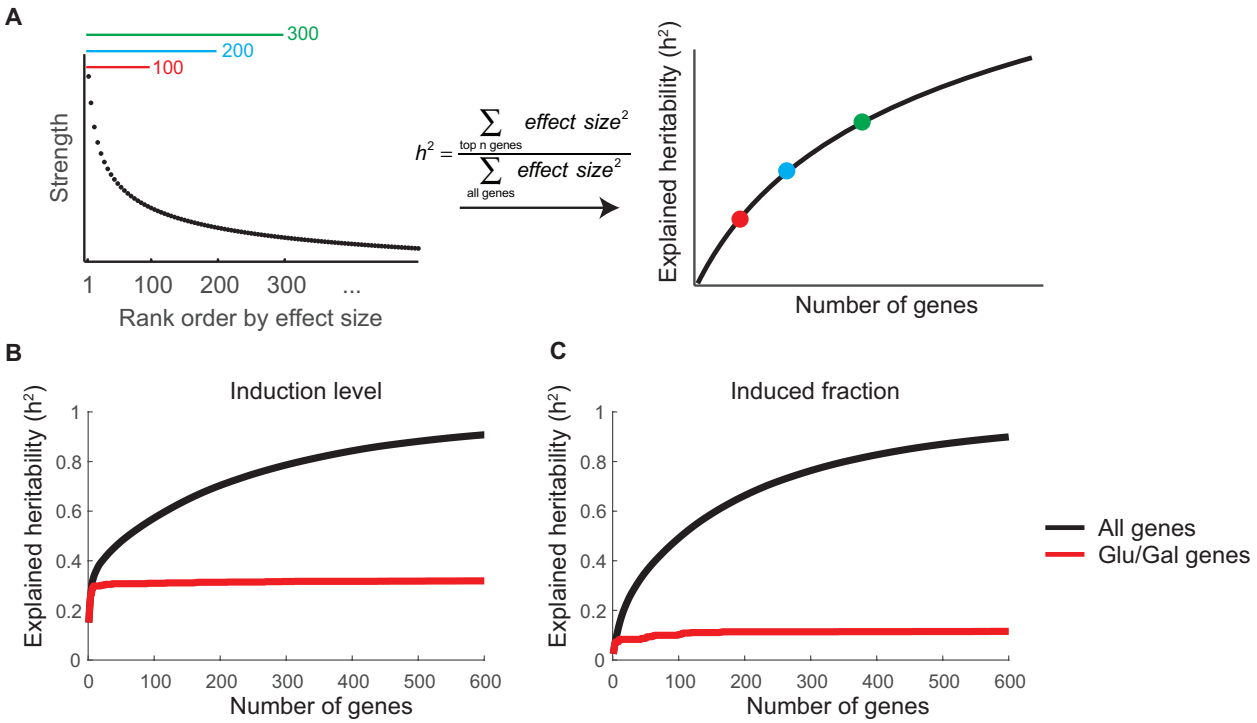


Figure 2.3 Pathway modifiers can significantly contribute to heritability

(**A**) Methods to estimate the heritability explained by a set of deletion mutants. Genes were sorted based on their effect size when deleted. The heritability was calculated as the sum of the squares of the effect sizes for the top n genes compared to all genes. The heritability (right) for the top 100 (red), 200 (blue), and 300 (green) mutant strains (left) is shown. (**B-C**) The contribution to explained heritability, as calculated in **A**, from GAL genes (red) or all genes (black) for induction level (**B**) and induced fraction (**C**).

Given their individual small effect size, our analysis also suggests that a significant portion of the genes that account for pathway function would not typically be considered to be contributing to each trait. Classical genetic screens identified only a fraction of the genes that have the potential

to significantly affect each of the two GAL traits. A compiled list of the 50 genes previously identified as affecting the GAL pathway (Supplemental Information) explained only 32.0% and 11.7% of variation in the induction level and induced fraction traits respectively. Similarly, in the unfolded protein response, genes whose products localize throughout the secretory pathway (ER and Golgi) explain only 27.1% of variation, further suggesting substantial roles of additional genes/processes. Hence, much of the variance occurs in genes we term non-trait-specific, i.e. genes that are not typically considered to be physiologically related to the trait. This is consistent with previous GWAS that identified putative causative loci that in some cases contained genes that were obviously trait-specific but in other cases were involved in general cellular processes. For example, human height is affected by variants in genes that underlie skeletal growth defects (trait-specific), as well as general pathways such as the Hedgehog pathway (non-trait-specific) [19]. Surprisingly, our analysis suggests that the non-trait-specific processes can have a larger aggregate effect than trait-specific pathways.

A potential caveat to these estimates is that the GAL phenotypes of ten mutants are either fully induced or uninduced, causing the effects of these genes to be underestimated. These ten genes have previously described influences on the GAL pathway. *GAL1*, *GAL3*, *GAL4*, *GAL80*, *REG1*, and *SNF3* are involved in either glucose or galactose signaling. *HSC82* and *STI1* interact with the *HSP90* co-chaperone that has been shown to influence the GAL pathway. *SNF2* is a SWI/SNF chromatin remodeling complex that was previously suggested to be involved in nucleosome occupancy on GAL promoter [20]. *GCN4,* is a general transcription factor that responds to amino acid starvation. We believe in most cases this caveat does not affect our results. Because, the

loss-of-function effect size of these alleles is effectively infinite, they will behave as Mendelian not quantitative alleles. Instead, for any quantitative traits, the predominant alleles of these Mendelian loss-of-function genes must be hypomorphic alleles. Indeed, when we assume hypomorphic allele effect sizes for these genes by randomly sampling from the tail of the fitted exponential distribution, we only observed a modest increase in total trait variation (< 3%).

2.4 Gene deletions in core cellular processes affect quantitative traits

What are the functions of these 'pathway modifiers' we identified? Are they genes that have general effects on all traits or are they specific to one or a subset of traits? We found that non-trait-specific processes often affect more than one trait. All pairs of traits share significantly more genes that affect their behaviors than expected ($p < 10^{-65}$, one-tailed hypergeometric test). While only 2 genes would be expected by chance, 113 genes were shared by all traits (Figure 2.4A). These genes also overlap significantly with "hub" genes identified from genetic interaction network (between 140 and 257 out of 380 hubs genes are significant for each of the four traits, $p < 10^{-47}$, hypergeometric test) [12]. We used Gene Ontology to ask if shared non-trait-specific genes were enriched for specific biological processes. Indeed, many processes were enriched (Table S2.1), including translation (GO:0006412), regulation of metabolism (GO:0031323), and transcription (GO:0006351). Although this has not previously been extensively characterized, it is not surprising that these traits might be altered by perturbations in some core cellular processes.

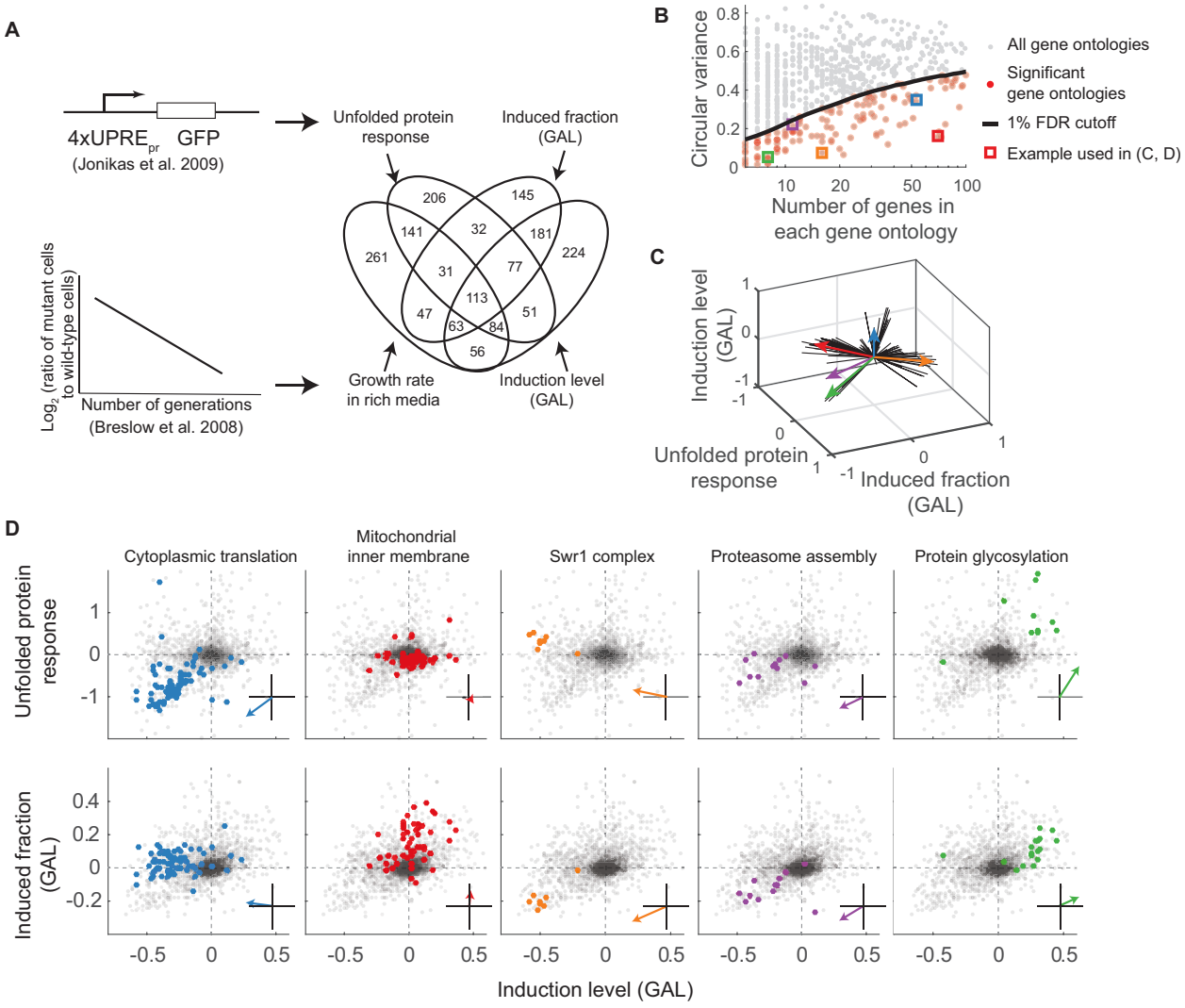


Figure 2.4 Core cellular processes affect quantitative traits

(**A**) Venn diagram showing the overlap between genes that significantly affect each of our four quantitative traits. Effect size for the unfolded protein response and growth rate in rich media was determined by reanalyzing data from Jonikas et al. and Breslow et al. [15,16]. Only genes that were assayed for all four traits are included in the Venn diagram. (**B**) Identification of gene

ontologies (GOs) that are significantly clustered in the 4-D trait space. For each GO the mean circular variance in the 4-D trait space was determined (Methods) and plotted against the corresponding number of genes in that GO (orange dots are significant, gray dots are not). To determine the 1% false discovery rate (FDR < 1%, black line), gene names were permuted (10000 bootstraps) before calculating the circular variance. GOs displayed in **C** and **D** are shown as squares. (**C**) The average effect size vector for each significant GO in **B** projected into the 3-D induced fraction-induction level-UPR response space. (**D**) Examples of GO with distinct spatial clustering. The effect of gene deletion on the unfolded protein response vs. GAL induction level (top) and on the GAL induced fraction vs. GAL induction level (bottom) was plotted for all genes from five different significant GOs from **B** (GO genes in color, all other genes in gray). (Inset) Average mutant vector of GO.

The identification of these core cellular processes as having potential to explain a significant amount of trait variation could be fundamental or trivial. It could reflect an architecture where many biological traits integrate many external and internal factors as inputs (e.g. the GAL pathway responding not just to galactose but glucose, redox status, ribosome capacity, ER capacity, etc.). Alternatively, as the expression of a large fraction of yeast genes is affected by growth rate control [21-23], a trivial explanation could be that the effect on the UPR and GAL traits is solely an indirect effect of a growth rate defect (Figure S2.4). Our data do not support growth rate as the sole factor explaining our results. Between 40% and 60% of gene deletions affect our GAL and UPR traits without affecting growth and vice versa (Table S2.2). Furthermore, for genes that affect both growth rate and any of the other traits, there is no

correlation in effect size between the two effects (R^2<0.02, Figure S2.7B-D). These observations argue against the idea that defects in growth are the main reason that non-trait-specific genes affect the behavior of traits. The involvement of many non-trait-specific genes instead suggests that many signaling pathways integrate a much larger set of cellular inputs than the single input for which the pathways are named.

2.5 Perturbation of core cellular processes can have trait-specific effects

Consistent with the idea that traits integrate a number of inputs in a trait-specific manner, we find that biological processes often affect more than one trait, but importantly not all traits. Using a spatial clustering algorithm in the four-trait space (Figure 2.4B-C), we found an enrichment in core cellular components (Table S2.3), such as ribosomal genes (GO:0002181), mitochondrial genes (GO:0005743), mannosyltransferases (GO:0000030), genes that affect histone exchange (GO:0000812) or proteasome assembly (GO:0043248), and genes involved in peptidyl-diphthamide synthesis (GO:0017183). Each of these sets of genes had a separable direction in this 4-dimensional space suggesting each process is responding distinctly to the mutations (Figure 2.4D). For example, the 89 genes involved in cytoplasmic translation (GO:0002181) were enriched in 3 out of 4 quantitative traits, namely the unfolded protein response, growth rate in rich media, and GAL induction level, but not GAL induced fraction (40, 21, 37 and 3 genes respectively out of the top 300 genes). Conversely, mitochondrial inner membrane genes (GO:0005743) are enriched in the GAL induced fraction but not growth rate, unfolded protein response, nor GAL induction level (21 versus 3, 1, and 4 respectively out of the top 300 genes).

Furthermore, the same core processes can have distinct effects on different traits. For example, at first glance, one might expect mutations in ribosomal genes to affect the level of induction of a pathway (e.g., by altering the expression level of all genes) but not the fraction of cells induced. Indeed, this is the case for the GAL response. But, when we examined the effect of the same mutants on a phosphate responsive (PHO) promoter, PHO84pr, we obtained a different result (Figure 2.5). The *PHO84* promoter responds to phosphate limitation in a bimodal manner and can therefore be characterized in the same way as we characterize the GAL response. The effects of ribosomal mutants on the induction level of PHO84pr-YFP are significantly less than for GAL1pr-YFP (Figure 2.5B versus C; Figure 2.5E and Figure S2.5, $p=3x10^{-10}$, two-tailed t-test). Instead, ribosomal mutants affect the PHO induced fraction and the level of expression of the uninduced cells (Figure 2.5C and examples in Figure 2.5D). In support that these results are a direct consequence of perturbation of ribosomal function, cycloheximide, a small molecule inhibitor of the ribosome, phenocopies the results of ribosomal gene deletions on both the GAL and PHO pathways (Figure 2.5B-D). While this result at first may seem counter-intuitive, these results could be explained if ribosomal proteins differentially impacted the expression level of positive versus negative regulators of a trait. In total, this suggests that variation in genes involved in core cellular processes could have both generic and pathway specific effects.

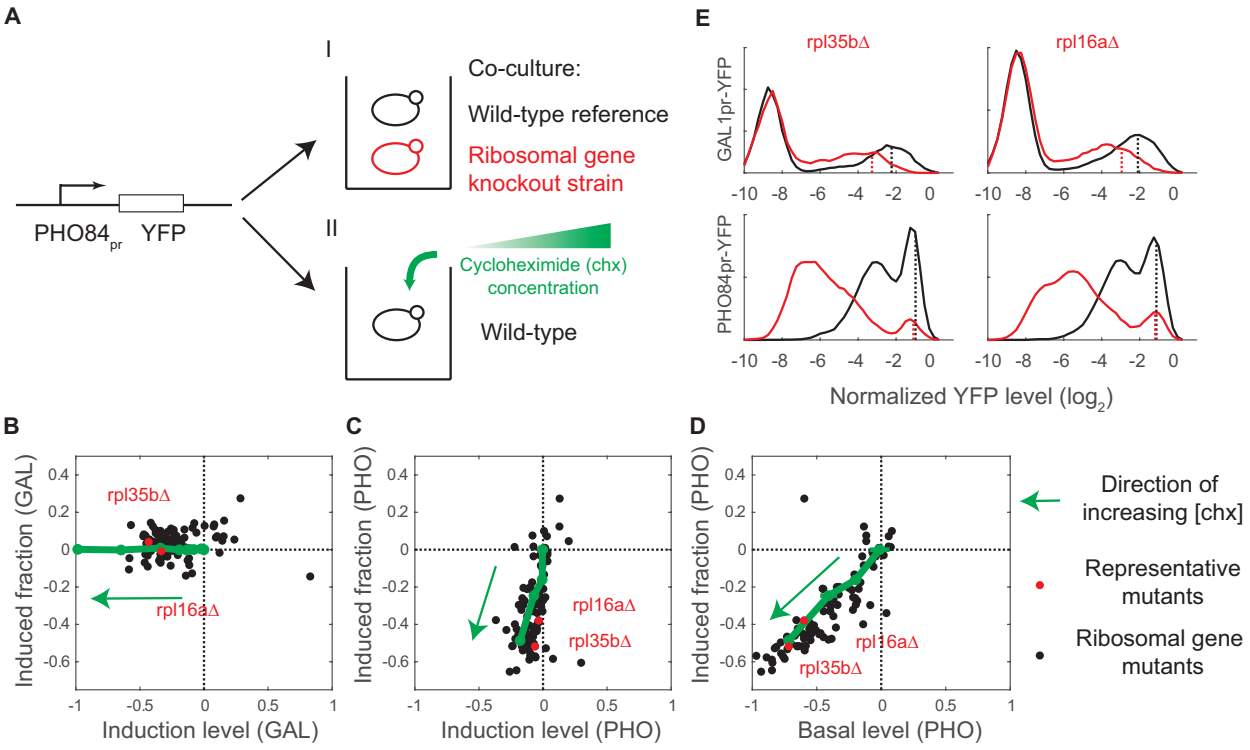


Figure 2.5 Effects of protein synthesis perturbation on the phosphate response (PHO) are distinct from the effects on the galactose response (GAL)

(**A**) Schematic of experiment to quantify the effects of perturbing protein synthesis on the PHO response. A $PHO84_{pr}$-YFP reporter was used to quantify PHO pathway activation in single cells. Protein synthesis was perturbed by either (I) knocking out genes involved in protein synthesis or (II) treating our wild-type strain with a titration of cycloheximide. (**B**-**D**) The effects of perturbing protein synthesis are different between the GAL and PHO response. Perturbation phenotypes were quantified by: 1) induced fraction, 2) induction level and 3) for the PHO response, basal expression level. A set of 95 strains each deleted for a gene involved protein synthesis (black dots) was assayed (GAL in **B**; PHO in **C** and **D**). Cycloheximide (chx), a protein

synthesis inhibitor, was added at 11 different concentrations to a wild-type strain (green dots; green arrow denotes direction of increasing chx concentration). Cycloheximide has a dose-dependent affect on both the GAL and PHO response that phenocopies the effect of protein synthesis mutants. (**E**) The expression distribution for two representative mutants, *rpl16a*Δ and *rpl35b*Δ in (red dots in **B-D**). The GAL1pr-YFP (top) and PHO84pr-YFP (bottom) distributions of *rpl16a*Δ and *rpl35b*Δ mutants are shown (red), together with the co-cultured wild-type strain (black). The induction level metric is denoted (dashed line). The induction level is not change in PHO (bottom) while it is in GAL (top).

2.6 Extension to other sources of genetic variation through simulation

We next wished to determine to what extent our results generalize to genetic variation beyond the complete loss-of-function variants we experimentally measured. Genetic variation in natural population is more complex genetically than the deletion library we analyzed. To generalize our results to account for a broader range of genetic variation, we developed a model where we accounted for 1) other types of alleles, i.e. hypermorphs and neomorphs as originally proposed by Muller [24], 2) variable number of alleles per gene, and 3) variable allele frequencies in the population (Figure 2.6). While the actual molecular cause of the variation can come from many sources, e.g. single nucleotide polymorphisms (SNPs), copy number variation, and indels, for the purpose of understanding the genetic architecture, it is only important to understand the effect of the genetic change on the trait, and hence for simplicity we will refer to all genetic variants as SNPs. Additionally, we assumed all SNPs contribute linearly to the trait with no epistasis. This

assumption is based on the fact that a linear model using all SNPs genotyped in human height GWAS can explain a large fraction of height heritability [25,26].

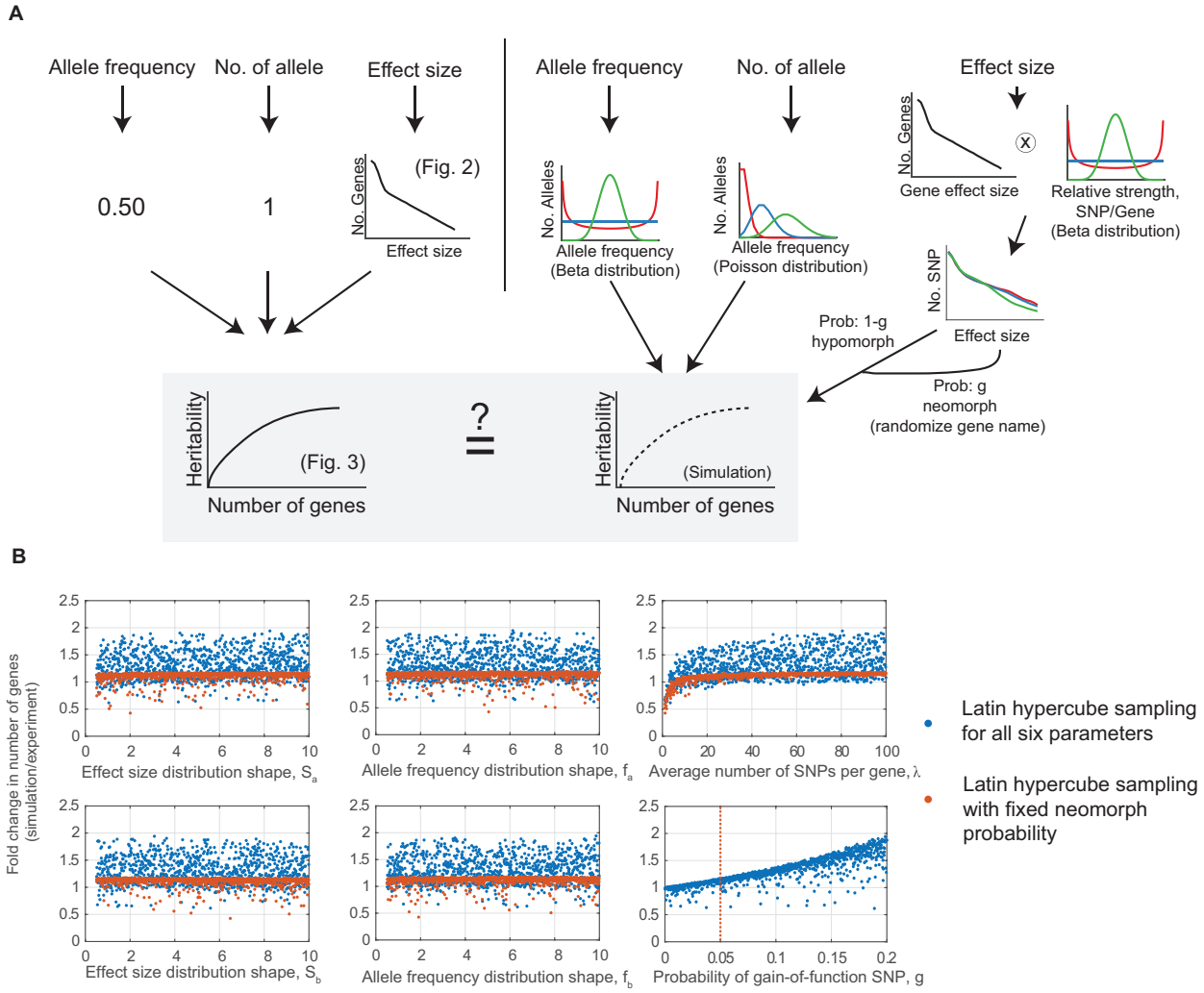


Figure 2.6 Effect size distribution estimated from gene deletions is informative for more complex genetic scenarios

(**A**) In figure 2.3, heritability versus gene number was estimated assuming an allele frequency of 0.5, exactly 1 SNP per gene, and the effect size distribution measured in figure 2.2. To simulate

more complex biological scenarios, we sampled allele frequency from a Beta distribution, Beta(f_a, f_b); number of SNPs from a Poisson distribution, Poisson(λ); simulated hypomorphs by convolving the measured effect size distribution for amorphs with a Beta distribution, Beta(S_a, S_b); and neomorphs by randomly sampling from the hypomorph effect size distribution. The frequency of hypomorphs versus neomorphs was a constant, g, for each simulation. The extremes and middle of the range of each distribution are shown (red, blue, and green) (**B**) Comparison of the number of genes required to explain 80% of the heritability in the experimental and simulated data. Simulate data was generated by Latin hypercube sampling of the six parameters (1000 iterations; blue dots). The fraction of neomorphs (g) had the largest affect on the model. To examine the effect of the rest of parameters, g was set to 5% (vertical red dashed line), and Latin hypercube sampling was used was used to scan the remaining five parameter space 1000 times (red dots).

To instantiate the model (Figure 2.6) a series of functional forms and constants were assumed for each of the potential variables. The number of SNPs that affect a given gene was chosen from a Poisson distribution to reflect variable number of alleles observed in human genome [27]. The effect size of hypomorphic SNP was modeled by multiplying a beta distributed random variable by the actual measured effect size of each affected gene. In this way, the maximum effect size was the complete loss-of-function and the minimum effect size was zero. A beta distribution was chosen to allow modeling of a wide range of different shaped distributions (Figure 2.6). To simulate neomorphic (gain-of-function) SNPs, we randomly selected a fraction SNPs, and reassigned their effect sizes with the effect size of randomly chosen SNPs. Lastly, the allele

frequency for each SNP was chosen from a beta distribution.

In each simulation, we calculated the explained trait variation for each SNP, and then summed up all the SNPs for a single gene to obtain the explained variation for each gene. We then varied the parameters in each of the distributions of the variables introduced above. Specifically, we used Latin hypercube sampling to scan the parameter space of the distributions (blue dots, Figure 2.6), and then compared the number of genes that explain 80% of trait variation obtained from this model and our experimental results (Figure 2.3). The results from our simulation show variation in the fraction of neomorphs dominates variation in the model. However, as long as this fraction is below 5%, the results of the simulation do not vary from the experimental results by more than 17%. Neomorphic alleles are typically assumed to be rare. In order to determine the potential impact of the other parameters in our model, we fixed gain-of-function rate to be 5%. Resampling the other five parameters, we found that the average number of SNPs per gene is the second largest source of variation in our model. However, as long as on average 5 SNPs exist per gene in the population, the effect is negligible (orange samples in Figure 2.6B). Large-scale sequencing efforts have now identified ~20 million genetic variants in humans [28]. Even if 99% of these variants were neutral, there would still be enough SNPs per gene on average to support our conclusions. From this, we determined that our estimate of number of genes that influence a trait from our knockout data is largely insensitive to the parameters of our model, and quantitative analysis of complete loss-of-function alleles should be informative even for the analysis of less severe and of rare alleles.

2.7 Discussion

In this work we sought to determine all the genes that can influence a pathway underlying a quantitative trait. Depending on the number of genes and the magnitude of the effect, pathways could in principle have a centralized or distributed architecture (Figure 2.1). To address this question we determined the effect size distribution of deletion mutants for four quantitative traits in yeast. We did this by measuring the response to galactose at the single-cell level for each deletion strain from the yeast library and by reanalyzing two additional quantitative screens that measure competitive growth rates [15] and the unfolded protein response (UPR) [16] in the deletion library. We found that in all four cases, the distribution of effect sizes is such where a quarter to half of the genes follow an exponential distribution, with the rest of the genes having a negligible effect size (Figure 2.2). Based on a simple model to calculate heritability, we found this result implies that a large number of genes (5-9% of all genes) would be needed to cumulatively explain at least 80% of trait variation (Figure 2.3). Our results imply that there is a significantly larger subset of genes that affect each trait than previously appreciated, but that individually their effect is difficult to detect by less quantitative experimental methods. The results provide evidence for pathways having a distributed, rather than centralized, genetic architecture.

A distributed pathway architecture suggests that many genes that are not typically considered part of a pathway, such as the GAL pathway, could still play an important role in pathway function. We found that these “pathway modifiers” were enriched in several core cellular

processes (Figure 2.4). Given the pleiotropic nature of these processes, it is not surprising that these genes can influence multiple traits. Unexpectedly, however, we found that a mutant in a core cellular process can have trait-specific consequences; e.g. a ribosomal mutant affects the induction level in the GAL response, but the induced fraction in the PHO response (Figure 2.5). This implies that, instead of making cells 'sick', biological processes underlying quantitative traits are likely affected by a large number of inputs that have the potential to act in a trait-specific manner.

Previous work had found that yeast traits were affected by fewer genes than we report here. Work by Bloom et al. used linkage analysis to identify quantitative loci underlying 46 yeast traits, and found a median of 12 loci affected each trait [8]. While it is possible that our four traits happen to be more complex than the traits that were analyzed by Bloom et al., we believe the differences result from the applied methods. If either the two yeast strains used in the linkage analysis of Bloom et al. are more related than two random isolates in a natural population or if the traits analyzed were under strong selection, this would lead to an underestimation of the number of genes. Because we are using a deletion library, we avoid the confounding effect of selection and the biases due to the limited number of alleles between two natural isolates. We therefore believe that the discrepancy between the results of these two works is at least in part due to the applied methods, in particular selection on growth rates.

2.7.1 Mendelian vs. quantitative trait

A distributed genetic architecture, as observed in this study, has implication for patterns of

genetic inheritance. Different individuals can have different numbers of alleles and the effect size of the strongest alleles can be different. Therefore, the expectation should be that the same trait, when examined in a pairwise manner between many individuals, should exhibit a range of segregation patterns from Mendelian to quantitative depending on the number and strength of the alleles. Indeed, this exactly what was recently observed for multiple traits in crosses between yeast strains [29]. Furthermore, one should expect a smaller number of genes that contribute to a quantitative trait will have rare alleles that make the trait behave as a Mendelian trait. Indeed, this has also been found that many quantitative loci associated with normal human height variation contain genes underlying syndromes characterized by abnormal skeletal growth [19].

2.7.2 Application to human genetics

Our results suggest that the number of genes that can influence a trait, when extrapolated to humans, is ~ 1500. However, our results were focused on a single-celled microbe that has a more compact genome with a smaller number of protein-coding genes than metazoan genomes. To what extent might our observations generalize to human genetic variation, given the differences in genome architecture and complexity? Do human traits, especially ones involved in important human disease, also have such a distributed underlying genetic architecture? One way to assess whether the genetic architecture of yeast and human traits is different is to compare the number of genes and their corresponding effect size distribution.

While a small number of human diseases or traits can be explained by a small number of causative genes, e.g. three genes explain 50% of the genetic risk in macular degeneration [5],

many traits are poorly explained by a small number of genes. For example, a GWAS on human height found that 423 loci explained less than 20% of total heritability [30]. Similarly, 163 loci only explain 14% of heritability in Crohn's disease [31], and 100 loci, excluding major histocompatibility complex, explain less than 6% of heritability in rheumatoid arthritis [32]. Since the explained fraction of heritability is far less than 100% in all these studies, it is difficult to accurately estimate the number of loci required to explain a majority of heritability in a human trait, but a reasonable estimate would be in the thousands. This suggests that the fraction of genes involved in a quantitative trait is similar in yeast and humans.

While the effect size distribution of human traits is poorly defined it is consistent with our results. Park et al. devised a method to determine the effect size distribution by taking into account all identified alleles and the power to have detected these alleles. From this they concluded that the effect size distribution alleles affecting human traits are monotonically increasing [33]. The range of possible distribution discussed in that work is consistent with an exponential distribution. While there is no good human data exists on the distribution of small effect size alleles, gene essentiality can be used as a rough comparison of the relative distribution of strong effect size allele between yeast and humans. Further supporting the similarity in effect size distributions, the number of essential genes in yeast and humans is similar. In total we believe this supports the idea that while the human genome is more complex than yeast, differences in genetic architecture are likely subtle and quantitative not large and qualitative.

2.7.3 Implication of a distributed genetic architecture on human disease

High-throughput genetic interaction maps have suggested that cellular processes are deeply interconnected [10,11]. But, it was not determined whether these connections were strong enough to be physiologically relevant. Our results demonstrate that cumulatively many genes of small effect size can make significant contributions to quantitative traits. Importantly, the effect sizes of these variants are not infinitesimal, and therefore we believe that increased power in GWAS would likely capture a significant portion of the missing heritability. This conclusion is consistent with work from Yang et al., which has shown that human genetic variants tagged in GWAS on body mass index is capturing the vast majority of heritability even if it is underpowered to identify the causative loci [26]. Of course, increased power alone will not help identify which SNPs within a locus is causative.

Given that so many genes can affect a trait, a second expectation is that causative small effect size loci should be shared between many but not all traits. Indeed, correlation among genetic variants has been observed in a recent study using 24 human traits [34]. Interpreting these results has been challenging as these genetic correlations could arise from either a direct causative link between the two diseases or shared genetic factors. Our results suggest that these correlations can result from shared genetic factors that are enriched in core cellular processes. This means that there could be power in searching for processes that are significantly enriched between diseases that wouldn't typically be thought of as related. Finally, the spectrum of defects seen in some complex diseases could arise from the specific combination of small effect alleles in each individual.

In summary, our work provides a system-level perspective into the architecture of a quantitative trait. In contrast to most other works that focused on existing genetic variants, our work quantitatively determined the contribution of loss-of-function alleles. With further development of gene editing technologies and disease models, it will be interesting to test these conclusions in more complex systems.

Methods

Re-analysis of quantitative screening that used the yeast deletion collection. Genome-wide screens that used the yeast deletion collection (reviewed in Giaever 2014) were re-analyzed. After downloading available effect size measurements for individual mutants, the measurement error of each assay in Table S2.4 was determined as the standard deviation of the differences of replicate measurements for identical strains (see Supplemental Information for details). The effect sizes were compared to measurement noise distribution, $\sim \boldsymbol{N}$ (0, measurement noise), to assign p-values for mutants. False discovery rates (FDR) were used to correct for the multiple hypothesis test problem. Significant mutants were defined as ones with FDR less than 0.5%.

Plasmid and strain construction. We constructed a plasmid containing the GAL1 promoter driving YFP with a Zeocin resistance marker all flanked by regions that are homologous to the HO locus (A65V). This plasmid was digested with Not1 and transformed into the parental SGA strain (B56Y, MATx ura3Δ leu2Δ his3Δ met15Δ can1Δ::ste2pr-spHIS5 lyp1Δ::Ste3pr-LEU2 LYS2+ cyh2) [1] to construct a base strain (D62Y), which was used to create both query and

reference strains used in the GAL screen. Query strains (library SLL14) were constructed using the SGA techniques[1] on the deletion collection and base strain D62Y. A reference strain (F59Y) was constructed by a second transformation with a TDH3pr-mCherry construct. The PHO84 promoter driving YFP reporter (E40B) was constructed using similar method by using PHO84pr PCR-ed from FY4 (using primer CGTACGCTGCAGGTCGACGGATCCCGTTTTTTTACCGTTTAGTAGACAG and TAATTCTTCACCTTTAGACATTTTGTTATTAATTAATTGGATTGTATTCGTGGAGTTTTG) instead of the GAL1 promoter. The resulting PHO library (SLL15) and reference strain (I32Y) were used in the PHO screen.

Galactose induction assay. Mutant strains from the deletion library that contains GAL1pr-YFP reporter and the corresponding reference strain were pinned onto YEPD agar plate before being inoculated into synthetic complete 2% raffinose medium to allow growth till saturation. Mutants and the reference strain were pinned together into 150 µl of fresh raffinose medium and grown for another seven hours, before being inoculated into 150 ul of synthetic complete 0.2% glucose and 0.3% galactose. After induction for eight hours, 10 ul of cultures were analyzed by flow cytometry LSRII with HTS. Each plate ran for ~20 minutes on the instrument. To ensure that all mutants underwent roughly the same induction time, no more than four plates were inoculated at a time. The induction level and induced fraction trait were based on measurements from two biological replicates in two separate days. Data were analyzed using a Matlab script (for representative raw data, see Figure S2.6).

Phosphate starvation induction assay. Mutant strains from the library that contain the

PHO84pr-YFP reporter and the corresponding reference strain were pinned on YEPD plate before being inoculated into synthetic glucose medium (SD). Mutants and the reference strain were then co-cultured in SD for 12 hours before washing in water twice and transferred into induction medium - synthetic glucose medium supplemented with 200 μM of K_2HPO_4. Medium recipe is from Wykoff et al. [2]. Cultures were analyzed by a Stratedigm S1000EX cytometer cytometry. The three PHO traits were based on measurements from two biological replicates in two separate days. Data was analyzed using Matlab scripts.

Fitting the effect size distribution. As the measured effects of most strains are close to measurement error, we first analyzed the effect size distribution of strains with significant measured effect sizes (FDR<0.5%). Mutant effect sizes were binned and fitted to exponential distributions. The only fitting parameter is the scale of the exponential distribution, which was estimated by maximizing the following log-likelihood function.

$$\log L = \sum_{significant\ genes} log\ (P(ES_i|\theta))$$

, where ES_i is the effect size of the i^{th} significant gene, and θ is the the scale of the exponential distribution. The probability distribution is an exponential distribution defined over a range of effect size, i.e.:

$$P(x|\theta) = \frac{\frac{1}{\theta}\int_{x_{min}}^{\infty} exp\left(-\frac{x}{\theta}\right)dx}{exp\ (-\frac{x_{min}}{\theta})}$$

Parameters that maximize the likelihood of measurements were used for Figure 2.2. The fitted

exponential distribution was extrapolated into the small effect size region to estimate the number of genes that are likely to follow the distribution. As a parsimonious model to explain our effect size measurements for four traits, we assumed the rest of mutants to have effect sizes as zero. Using this model, we predicted the expected effect size measurement distribution by convolving the true effect size distribution with the measurement noise of each assay (solid blue line in Figure 2.2). This distribution was then randomly sampled in 10,000 simulations and the standard deviation of the simulation was used as the confident zone of our estimation.

Extrapolate the number of genes that affect quantitative traits to human traits. The number of significant genes were corrected by a factor determined by the gene number ratio between known human genes and screened yeast genes. The number of human genes is estimated as 22,500 [3]. The number of screened yeast genes was determined as the number of genes that passed quality control.

Simulation of potential biases from the study of amorphs. In our model, we defined the explained heritability as the total explained heritability by all SNPs that affect each gene. As described in the main text, we simulated the number of SNPs that affect each gene as a Poisson distribution. The allele frequency and relative effect size are modeled using beta distribution. Gain-of-function SNPs were modeled by re-assigning effect sizes of a fraction of all SNPs by randomly sampling from the effect size distribution of all SNPs. In our Latin hypercube sampling, parameters in the two beta distributions ranged from 0.5 to 9, the fraction of gain-of-function SNPs ranged from 0 to 50%, and the average number of SNP per gene ranged from 1 to 100. The heritability of each SNP is modeled as $2*S^2*f*(1-f)$, where S is effect size and f is

allele frequency. Measured knockout effect size on induced level is used in the model as complete loss-of-function effects. The code used for this simulation is available at Dryad.

Gene Ontology analysis. Genes that are significant for all four traits (FDR<0.5%) were used as a hit list; all the genes that passed quality control were used as a background list. Gene Ontology analyses were done using GO TermFinder [4].

Spatial clustering algorithm. Each gene was represented a 4-dimensional effect size vector using the effect size measured for each of the four yeast traits. Since different traits have different units, we normalized each dimension of the effect size vectors by its scale, which is defined as the root mean square of the effect sizes of all the genes that significantly affect that trait. For any gene set, we determined the similarity of their effects on four traits by 1) filtering out all genes that are not significant to any of our traits; 2) calculate the circular mean of the normalized effect size vectors (e) as: $R = \sum \vec{e} / |\vec{e}|$; 3) calculate the circular deviation as $Var = 1 - R$. To determine the significance of this, we repeated the calculation 10,000 times after randomizing the gene names. Gene Ontologies that have at least five genes significant for any of the four traits were analyzed using the method above. Significantly clustered processes were defined as FDR < 0.01.

Cycloheximide effect on GAL and PHO. Cycloheximide was purchased from Sigma (C7698). Cycloheximide was added directly to the induction media and this was the only change in the protocol from strains that were not exposed to cycloheximide. Cells were grown in a two-fold dilution series of cycloheximide with the highest concentration of cycloheximide being 20

μg/ml. Cycloheximide effects in Figure 2.5 were based on at least three biological replicates.

Reference

1. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature. 1980;287: 795–801.

2. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. 2010;464: 1039–1042. doi:10.1038/nature08923

3. Friedman A, Perrimon N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature. Nature Publishing Group; 2006;444: 230–234.

4. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. 2009;106: 9362–9367. doi:10.1073/pnas.0903103106

5. Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nature Genetics. 2006;38: 1055–1059. doi:10.1038/ng1873

6. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93: 779–797. doi:10.1016/j.ajhg.2013.10.012

7. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. Nature Publishing Group; 2011;12: 628–640. doi:doi:10.1038/nrg3046

8. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. 2013;494: 234–237. doi:10.1038/nature11867

9. Botstein D, Fink GR. Yeast: An Experimental Organism for 21st Century Biology. Genetics. 2011;189: 695–704. doi:10.1534/genetics.111.130765

10. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. AAAS; 2010;327: 425–431.

11. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353. doi:10.1126/science.aaf1420

12. Giaever G, Nislow C. The yeast deletion collection: a decade of functional genomics. Genetics. 2014;197: 451–465.

13. Acar M, Becskei A, van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228–232. doi:10.1038/nature03524

14. Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proceedings of the National Academy of Sciences. National Academy of Sciences; 2015.

15. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nature Methods. 2008;5: 711–718. doi:10.1038/nmeth.1234

16. Jonikas MC, Collins SR, Denic V, Oh E, Quan EM, Schmid V, et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science. 2009;323: 1693–1697. doi:10.1126/science.1167983

17. Crow TJ. 'The missing genes: what happened to the heritability of psychiatric disorders?'. Mol Psychiatry. 2011;16: 362–364. doi:10.1038/mp.2010.92

18. Lynch M, Walsh B. Genetics and Analysis of Quantitative Traits. Sinauer; 1998.

19. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. Nature Publishing Group; 2010;467: 832–838.

20. Bryant GO, Prabhu V, Floer M, Wang X, Spagna D, Schreiber D, et al. Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol. 2008;6: 2928–2939. doi:10.1371/journal.pbio.0060317

21. Regenberg B, Grotkjær T, Winther O, Fausbøll A, Åkesson M, Bro C, et al. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biology. BioMed Central; 2006;7: 1. doi:10.1186/gb-2006-7-11-r107

22. Slavov N, Botstein D. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. 2011.

23. Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, et al. Promoters maintain their relative activity levels under different growth conditions. Molecular Systems Biology. 2013;9: 701. doi:10.1038/msb.2013.59

24. Muller HJ. Further studies on the nature and causes of gene mutations. 1932. pp. 213–255.

25. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics. 2010;42: 565–569. doi:10.1038/ng.608

26. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47: 1114–1120.

27. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409: 928–933. doi:10.1038/35057149

28. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research. Oxford Univ Press; 2001;29: 308–311.

29. Hou J, Sigwalt A, Fournier T, Pflieger D, Peter J, de Montigny J, et al. The Hidden Complexity of Mendelian Traits across Natural Yeast Populations. CellReports. 2016;16: 1106–1114. doi:10.1016/j.celrep.2016.06.048

30. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics. 2014;46: 1173–1186. doi:10.1038/ng.3097

31. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491: 119–124. doi:10.1038/nature11582

32. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506: 376–381. doi:10.1038/nature12873

33. Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. Nature Publishing Group; 2010;42: 570–575.

34. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of

genetic correlations across human diseases and traits. Nature Genetics. 2015;47: 1236–1241. doi:10.1038/ng.3406

Chapter 3. Aneuploidy causes non-genetic individuality

Rebecca R. Beach[1], Chiara Ricci-Tam[2], Christopher M. Brennan[1], Christine A. Moomau[1], Pei-hsin Hsu[1], Bo Hua[2], Rebecca E. Silberman[1], Michael Springer[2], and Angelika Amon[1,3,*]

[1]David H. Koch Institute for Integrative Cancer Research, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

[2]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

In this work, I collected data for Figure 3.6 and Figure S3.5, S3.6 with a lab mate Chiara Ricci-Tam, conducted all related data analysis, and I am an author of the final manuscript. This chapter is adapted from a manuscript written by the authors listed above.

3.1 Introduction

Aneuploidy, a genomic state defined by chromosome gains and losses, is the leading cause of miscarriages and mental retardation in humans and a hallmark of cancer. A remarkable characteristic of diseases caused by unbalanced karyotypes is high variability in presentation. In Down Syndrome (Trisomy 21), for example, 87.5 percent of fetuses die in utero, while some individuals reach ages of 60 and more (reviewed in [1]. High levels of variability are also observed in the degree of facial dysmorphology, cognitive abilities, hypotonia, and

dermatoglyphic features at birth. Disease predisposition is also variable. This phenotypic variability has been attributed to allelic variance among individuals (reviewed in [1]. Here we show that phenotypic variability is inherent to the aneuploid state, providing an additional potential explanation for the heterogeneity observed in diseases caused by large-scale karyotypic abnormalities.

Gains and losses of genes encoded on autosomes generally lead to corresponding changes in gene expression (reviewed in [2]); therefore, it is not surprising that the consequences of mis-segregating entire chromosomes are severe. Aneuploid cells exhibit proliferation defects, form protein aggregates at an elevated level, and are genomically unstable, which are collectively known as the aneuploidy-associated stresses (reviewed in [3]). In addition to the traits observed in a broad range of aneuploidies, aneuploid cells exhibit gene-specific phenotypes where changes in dosage of specific genes cause a specific phenotype [4].

Studies of aneuploidy to date have largely been confined to cell lines that have gained chromosomes [5–7] because cell lines harboring chromosome losses are difficult to maintain [8,9]. Thus, it is not known how chromosome loss impacts cell physiology nor whether monosomic cells exhibit some or all of the aneuploidy-associated stresses that are so widespread among cells with chromosome gains. Not only do we lack an understanding of how monosomy affects cell physiology, but we also do not know how quickly aneuploidy-associated phenotypes develop following chromosome mis-segregation.

Here we employ a system that allows us to examine the immediate consequences of defined

chromosome gains and losses in budding yeast. Our analyses reveal that the adverse effects of aneuploidy on cell proliferation are immediate for both chromosome gains and losses. Our analyses further revealed high cell-to-cell variability in cell cycle progression among cells harboring the same aneuploidies. We identify gene copy number imbalances caused by the aneuploid state as one source of this variability and show that the high variance in S phase and early mitosis duration is due to stochastic DNA damage caused by these gene imbalances. Responses to environmental perturbations are also highly variable among cells of the same aneuploid karyotype, indicating that aneuploidy impacts the robustness of multiple—perhaps all—biological processes. Finally, we find that inbred trisomic mouse embryos exhibit phenotypic variability. Our results show that while subtle changes in gene dosage of individual genes have little effect on the robustness of biological networks, they can promote alternate behaviors when they occur at a large scale.

3.2 A system to examine the immediate consequences of chromosome mis-segregation

In budding yeast, chromosomes that cannot attach to the mitotic spindle due to centromere failure are retained in the mother cell [10,11]. To induce chromosome mis-segregation, we replaced the endogenous centromere of each chromosome with a conditional centromere: a construct containing the galactose inducible/glucose repressible *GAL1* promoter adjacent to the centromere [10]. Using this strategy, we generated a set of strains with chromosomes whose segregation can be controlled by carbon source (Figure S3.1A). To follow chromosome mis-

segregation, we introduced an *E. coli* lac operator array near the centromere of the chromosome harboring the conditional centromere and expressed a GFP-LacI fusion to visualize the array (GFP dot; [12]; Figure S3.1A). Using this system, we found that upon addition of galactose to the medium, 67 - 93 percent of mother cells retained both sister chromatids harboring the conditional centromere (Figure S3.1B).

In a haploid strain, chromosome mis-segregation generates nullisomes (N-1)—which are inviable—and disomes (N+1); in a diploid strain, monosomes (2N-1) and trisomes (2N+1) are generated by chromosome mis-segregation. To ensure that induction of mis-segregation of one chromosome does not lead to non-disjunction of other chromosomes, we followed the segregation of a large (chromosome IV) and a small chromosome (chromosome V) one (3 hours) and 2-3 cell divisions (5 hours) after induction of mis-segregation of an unrelated chromosome. Inducing mis-segregation of chromosomes I, II, VIII, IX, XI or XIII did not cause mis-segregation of chromosome IV or V (Figure S3.1C), indicating that inactivation of one centromere does not cause substantial destabilization of other chromosomes.

We generated a large number of strains harboring single or multiple conditional centromeres (Supplemental Experimental Procedures). This allowed us to examine the consequences of mis-segregating up to 8 chromosomes and enabled us—for the first time—to systematically analyze the consequences of chromosome losses on yeast cell physiology. We conducted all of our cytological analyses in strains harboring GFP dots on the mis-segregating chromosome to ensure that we only analyzed cells that had indeed gained or lost a chromosome.

Quantitative proteomic analyses of budding yeast cells harboring an additional chromosome showed that approximately 80 percent of proteins encoded on the disomic chromosome are expressed according to their gene dosage [13,14]. To determine how quickly protein levels change upon gain of an additional chromosome, we measured protein content of haploid cells that had gained a copy of chromosome IV (disome IV) immediately after chromosome mis-segregation (1 hour) and 2 cell divisions later (after 5 hours). As a control we measured protein content of cells constitutively harboring an extra copy of chromosome IV (henceforth chronic disomes). Proteome quantification by TMT showed that levels of proteins encoded on the disomic chromosome were increased in this control strain. However, the increase was only 1.4-fold instead of two-fold, likely the result of the 'ratio compression' effect, a well-known artifact occurring in isobaric labeling analyses (Figure 3.1A; reviewed in [15]. We corrected for this ratio compression in our analysis.

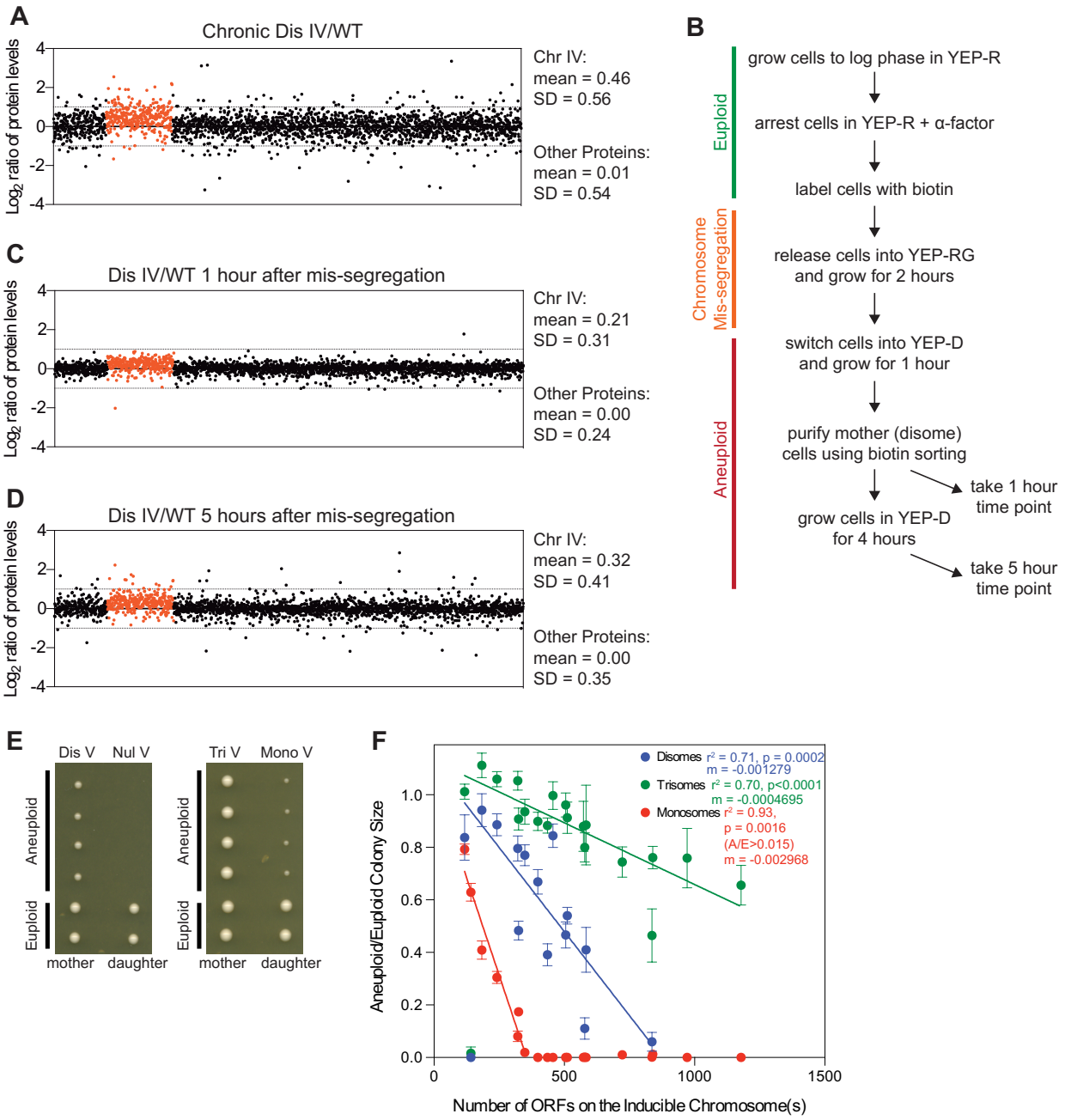


Figure 3.1 The proteome rapidly adjusts to changes in gene dosage following chromosome mis-segregation

(A) Plots show $\log_2$ ratios of the relative protein abundance of a strain harboring chronic disomy IV (A24367) grown in YEP-D at 30°C compared to wild-type cells (A38452) from the 5-hour time-point in (D). (B) Experimental setup used for the experiment shown in (C,D). Mother cells were isolated by sorting for biotin-labeled cells (1 hour; C) and then grown in YEP-D at 30°C for 4 hours (5 hours; D). (C,D) Log_2 ratios of the relative protein abundance of disome IV cells (A38455) compared to wild-type cells (A38452) 1 hour (C) and 5 hours (D) following chromosome mis-segregation. Proteins encoded on chromosome IV are in red. (E-F) Cells were grown in YEP-R and then switched to YEP-RG and grown at 30°C for 160 minutes to induce chromosome mis-segregation. Mother-daughter pairs of cells that had just completed a cell division were separated and placed side-by-side on YEP-D plates. Colony area was measured after growth at 30°C for 40-48 hours. Euploid colonies were grown on the same plate and were either from a wild-type strain or from cells in the experimental strain that did not missegregate a chromosome. Disome (blue), trisome (green), and monosome (red) colony size is shown as a ratio of aneuploid to euploid colony area as a function of the number of open reading frames (ORFs) on the aneuploid chromosome(s). Error bars indicate standard deviation. Linear regressions are shown for disomes ($r^2 = 0.71$, excluding disome VI due to lethality of excess *TUB2*), trisomes ($r^2 = 0.70$, excluding trisome VI), and monosomes ($r^2 = 0.93$, for colonies with aneuploid to euploid ratios > 0.015). Abbreviations: WT, wild-type; Dis, disome; Nul, nullisome; Tri, trisome; Mono, monosome; Chr, chromosome; SD, standard deviation; A/E, aneuploid to euploid colony size; m, slope. See also Figure S3.1.

In cells induced to mis-segregate chromosome IV, 67% of cells were disomic for chromosome

IV, 20% euploid, and 13% cells lacked chromosome IV one hour following chromosome mis-segregation. Five hours later, the population was comprised of 61% disomes, 34% euploid cells, and 5% nullisomes. Based on these distributions we calculated the $\log_2$ ratio of the expected protein abundance ratio to be 0.12 at 1 hour and 0.28 at 5 hours in the extreme where proteins are only diluted by proliferation and 0.27 at 1 hour and 0.28 at 5 hours in the other extreme where protein turnover is instantaneous (see Supplemental Experimental Procedures). Our experimental analysis revealed that one hour following chromosome mis-segregation, proteins encoded on chromosome IV were elevated compared to proteins encoded by the euploid chromosomes, with a mean $\log_2$ protein abundance ratio of 0.21 for proteins on chromosome IV and of 0.32 five hours following chromosome mis-segregation (Figure 3.1B-D). These results argue that protein composition had not quite adjusted to the genome composition of the disome IV cells within one hour of chromosome mis-segregation, but protein composition is effectively equivalent to a chronic disome by five hours. We conclude that chromosome mis-segregation leads to a rapid remodeling of the proteome to reflect the cell's genetic makeup.

3.3 Chromosome gains and losses cause defects in cell division

Given that chromosome mis-segregation results in the rapid adjustment of the cell's proteome to the new aneuploid karyotype (Figure 3.1B-D), we next investigated how quickly aneuploidy-associates stresses such as cell proliferation defects arise following chromosome mis-segregation. To measure cell proliferation following chromosome mis-segregation, we first examined colony formation 40-48 hours following chromosome mis-segregation, when colony

size was still increasing exponentially (Figure 3.1E). This analysis led to two conclusions. First, the growth defects observed in monosomes were more severe than the growth defects observed in disomes and trisomes (Figures 3.1F, B.1B; Supplemental Experimental Procedures). Second, colony size was inversely correlated with the degree of aneuploidy (Figure 3.1F). The observation that the negative fitness slope of trisomes is less steep than that of disomes further indicates that it is the relative ratio of aneuploid to euploid gene dosage that is in large part responsible for the observed fitness defects. Doubling the gene dosage relative to base ploidy of the cell, as occurs in the disomes, is proportionally more detrimental than increasing the gene dosage by 50 percent relative to base ploidy, as occurs in the trisomes. We conclude that (1) both chromosome gain and loss impair cellular fitness; (2) changes in relative expression levels of genes are responsible for the fitness defects observed in cells with chromosome gains; and (3) the fitness defect of monosomes is likely a composite of aneuploidy-associated stresses and haploinsufficiency.

To more carefully define the cell cycle defects caused by chromosome loss, we analyzed cell cycle progression following chromosome loss by time-lapse microscopy. We used mCherry-Cdc3 and Spc42-dsRed or Spc42-GFP fusion proteins along with GFP dots on the mis-segregating chromosome(s) to follow cell cycle progression. Cdc3 is a component of the septin ring that forms at the site of bud formation from the time of entry into the cell cycle until exit from mitosis [16]. Thus, the time during which a cell lacks a Cdc3 ring at the plasma membrane estimates the time a cell spends in G1. Spc42 is a component of the spindle pole body (SPB; [17]. The distance between SPBs and the separation of sister chromatids as analyzed by GFP dot

separation indicates anaphase onset.

Doubling-time measurements revealed that cell proliferation was delayed in all monosomic strains analyzed (Figure 3.2A). This analysis agreed with our colony size measurements but was not perfectly correlated (i.e. monosomes XII and XIV) because the doubling time measurements do not take into account cells that stop dividing, which is a frequent occurrence in some monosomic strains. We further observed that the cell cycle delay and number of cells that did not divide at all increased with degree of monosomy (Figure 3.2A). Note, however, that strains with the highest degree of monosomy appeared to have less of a cell cycle delay than strains harboring lower levels of monosomy. This is because cells that stop dividing, which are a frequent occurrence in strains with multiple monosomies (Figure 3.B.2A), cannot be included in the doubling time measurements.

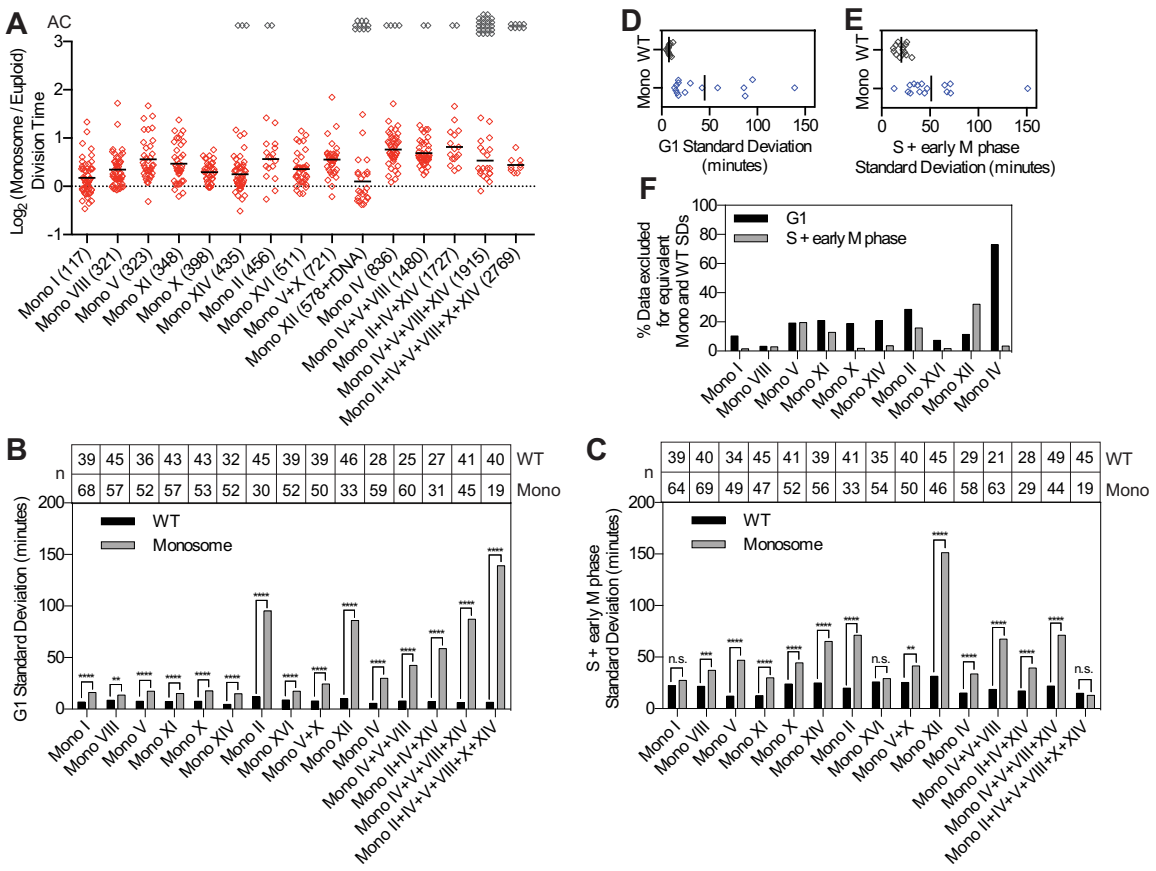


Figure 3.2 Chromosome loss leads to cell cycle delays and cell-to-cell variability

(A-F) Cells were grown to mid-log phase in SC-R and then switched to SC-RG for 160 minutes to induce chromosome mis-segregation. Cells were then plated on SC-D solid medium and imaged every 5 minutes for 8-10 hours at 25°C. Division time (A) was calculated for monosomes and normalized to euploid cells imaged during the same time-lapse. Euploid cells were either from the wild-type control strain or from cells in the experimental strain that did not mis-segregate a chromosome. Log_2 transformed aneuploid to euploid ratios are plotted. Lines are at the mean. Numbers in parentheses on the x-axis labels indicate number of open reading frames

on the aneuploid chromosome(s). Standard deviations for euploid (WT; black bars in B,C and grey points in D,E) and monosome (grey bars in B,C and blue points in D,E) populations were measured and an F-test was used to test for equality of variance between the monosome and the euploid population from the same experiment (** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$). (F) The number of most slowly dividing cells (as percent of the total population) that need to be excluded from the monosome population to obtain equal variance with the euploid populations is shown for G1 (black bars) and S+early M phase (grey bars). Note that we did not analyze strains harboring multiple monosomies. Such cells essentially only undergo one or two cell divisions before arresting—the duration of which underreports on variability because long lived proteins have not yet adjusted to the monosomic state and become limiting. Abbreviations: AC, arrested cells; WT, wild-type; Mono, monosome. See also Figure S3.2.

Next we examined the effects of monosomies on specific cell cycle stages. This analysis revealed that 14 of 15 monosomes analyzed exhibited a G1 delay that increased with degree of monosomy ($r^2 = 0.89$; Figure S3.2B,E). Delays in other cell cycle stages were also common (Figure S3.2C,D). We conclude that loss of most chromosomes impacts multiple cell cycle stages.

Our live-cell analyses also revealed that monosomic cells of the same karyotype exhibited dramatic cell-to-cell variability in cell cycle duration (Figure 3.2A-C; B.2B-D). To quantify this heterogeneity, we determined the standard deviation of G1 duration and of the time it took cells to go through S phase and early mitosis (bud emergence to anaphase onset, henceforth S+early M phase). This analysis revealed high variability in these two cell cycle periods (Figure 3.2B-E;

S2B,C). Furthermore, we found variability in G1 duration to increase with degree of monosomy (Figure 3.2B). S+early M phase duration variability appeared less correlated with degree of monosomy, but this appearance is misleading as cells harboring multiple monosomies only undergo one or two cell divisions before arresting in G1 (Figure S3.2A). Variabilities in the first and second divisions following chromosome mis-segregation are likely smaller than they would be in later divisions, were complex monosomies to continue to divide, because levels of long-lived proteins have not yet adjusted to the monosomic state and have not yet become limiting.

To determine whether variability was driven by a small number of extremely slowly dividing cells or it reflected increased variability amongst all members of the population, we asked how many of the most slowly dividing cells had to be eliminated from monosomic cell populations for variance of the monosome population to resemble that of euploid control cells (Figure 3.2F). This analysis revealed that variability in S+early M duration in half of the monosomes was driven by few cells with extreme cell cycle delays (Figure 3.2F). In the other half of monosomes, a higher percentage of the population (approximately 20 percent) conferred heterogeneity. A similar fraction of cells was responsible for the variability in G1 duration in most monosomes (Figure 3.2F). We conclude that monosomy interferes with multiple aspects of cell proliferation and that the degree of the defect varies greatly among cells harboring the same monosomy.

To determine whether cell-to-cell variability in cell cycle duration was specific to chromosome loss, we examined cell cycle kinetics immediately following chromosome gain(s) in haploid cells. Consistent with previous results, we found cell division to be slowed in all disomes, with most disomes exhibiting delays in both G1 and S+early M phase (Figures 3.3A, S3.3A-C).

Interestingly, like cells carrying monosomies, disomic cells exhibited significant cell-to-cell variability in G1 length and S+early M phase duration that increased with degree of disomy (Figure 3.3B, C; S3.3D). In the majority of strains harboring single chromosome gains, variability in S+early M and G1 length was driven by 20 percent of the population or less (Figure 3.3D). In strains harboring multiple disomes, variability reflected increased variability among the entire population (Figure 3.3D).

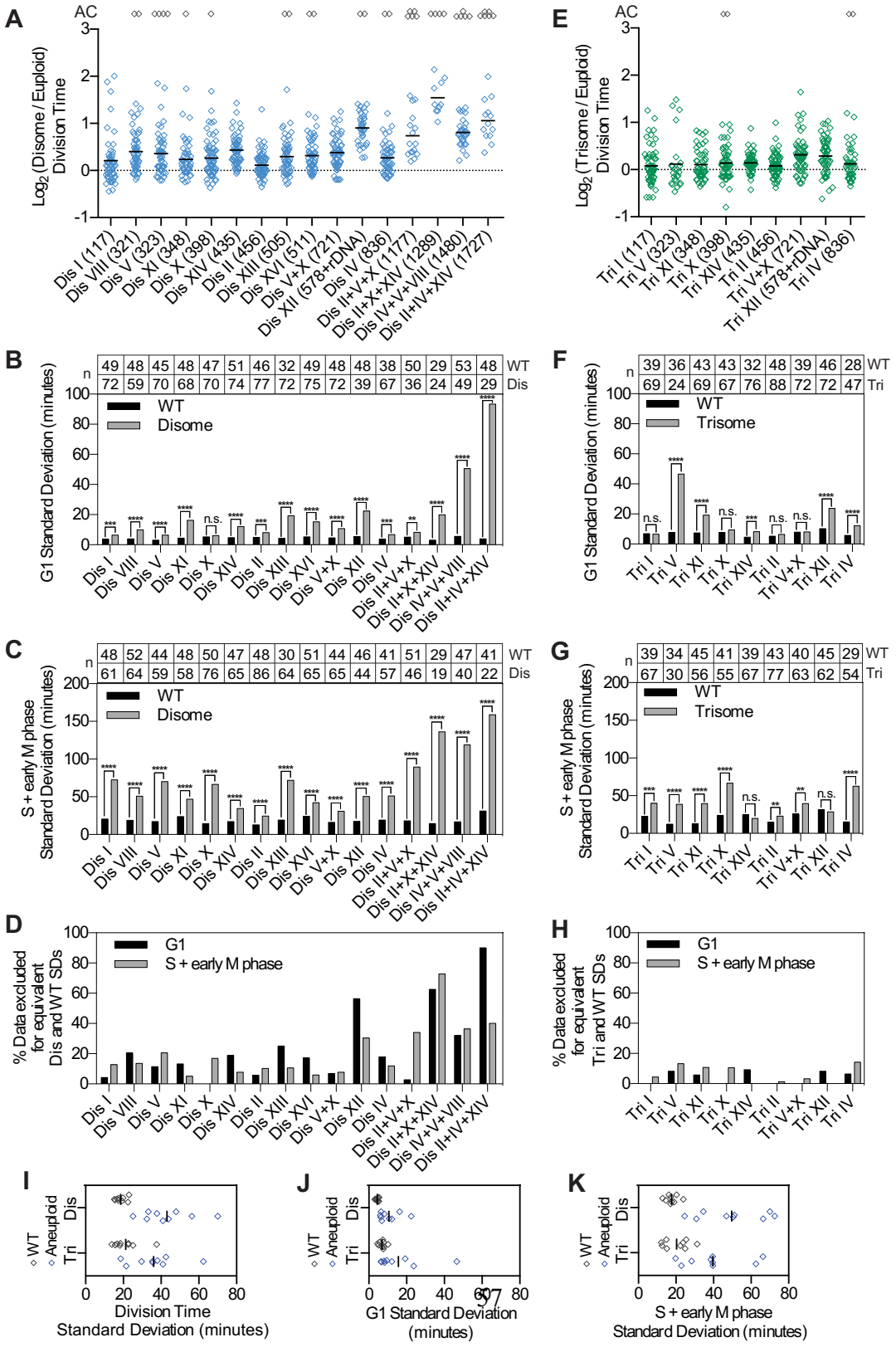

A
AC
Log2 (Disome / Euploid) Division Time
Dis I (117)
Dis VIII (321)
Dis V (323)
Dis XI (348)
Dis X (398)
Dis XIV (435)
Dis II (456)
Dis XIII (505)
Dis XVI (511)
Dis V+X (721)
Dis XII (578+rDNA)
Dis IV (836)
Dis II+V+X (1177)
Dis II+X+XIV (1289)
Dis IV+V+VIII (1480)
Dis II+IV+XIV (1727)
E
AC
Log2 (Trisome / Euploid) Division Time
Tri I (117)
Tri V (323)
Tri XI (348)
Tri X (398)
Tri XIV (435)
Tri II (456)
Tri V+X (721)
Tri XII (578+rDNA)
Tri IV (836)
B
n
WT
Dis
WT
Disome
G1 Standard Deviation (minutes)
F
n
WT
Tri
WT
Trisome
G1 Standard Deviation (minutes)
C
n
WT
Dis
WT
Disome
S + early M phase Standard Deviation (minutes)
G
n
WT
Tri
WT
Trisome
S + early M phase Standard Deviation (minutes)
D
G1
S + early M phase
% Data excluded for equivalent Dis and WT SDs
H
G1
S + early M phase
% Data excluded for equivalent Tri and WT SDs
I
WT
Aneuploid
Dis
Tri
Division Time Standard Deviation (minutes)
J
WT
Aneuploid
Dis
Tri
G1 Standard Deviation (minutes)
K
WT
Aneuploid
Dis
Tri
S + early M phase Standard Deviation (minutes)

Figure 3.3 Chromosome gain leads to cell-to-cell variability that is partially attenuated by increased ploidy

Cells were grown and imaged as described in legend to Figure 3.2. Disome division times (A) and standard deviations (B,C,I-K) and trisome division times (E) and standard deviations (F,G,I-K) were calculated as described in Figure 3.2 (** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p < 0.0001$). The number of most slowly dividing cells (as percent of the total population) that need to be excluded from the disome (D) or trisome (H) population to obtain equal variance with the euploid population is shown for G1 (black bars) and S+early M phase (grey bars). Plots in I-K show only common aneuploidies between disomes and trisomes (chromosomes I, II, IV, V, X, XI, XII, XIV, and V+X). See also Figure S3.3.

Most phenotypes associated with aneuploidy are attenuated when base ploidy is increased—that is, they are more severe in disomes than in trisomes [6]. To some extent, increased ploidy also buffered against cell cycle duration heterogeneity (Figure 3.3; S3.3). Variability in total cell cycle length and S+early M phase duration was partially attenuated by increased ploidy, but variability in G1 length was not (Figure 3.3I-K). Our data demonstrate that increased cell-to-cell variability with regards to cell cycle progression is a universal feature of the aneuploid state in yeast. Our data further indicate that this heterogeneity is caused in part by the amount of overproduced protein with respect to base ploidy and implicate overexpressed proteins or proteins that are not correctly assembled into their biological complexes as contributing to non-genetic variability seen in aneuploid cells.

What are the origins of the high cell-to-cell variability in cell cycle duration in aneuploid cells? Although aneuploid cells are genomically unstable [18–22], mutation-induced population heterogeneity is an unlikely cause. The system we employ examines the immediate effects of mis-segregating a specific chromosome, precluding the accumulation of suppressor mutations. Nevertheless, to exclude the possibility that heterogeneity is due to genetic or perhaps epigenetic changes, we examined whether cell cycle timing was consistent from one division to the next. We measured division time of individual cells during two consecutive cell divisions following chromosome mis-segregation for cells disomic for chromosomes I, II, V, VIII, or XIV. This analysis revealed that a substantial delay during one cell cycle was not predictive of a slow subsequent cell cycle, as single-cell division times in two consecutive divisions were not correlated ($r^2 \leq 0.019$; Figure S3.4A-E). Little correlation between consecutive division times was also observed in euploid cells [23]. We conclude that neither genetic nor epigenetic changes cause population heterogeneity in aneuploid cells.

3.4 Stochastic DNA damage contributes to population heterogeneity in aneuploid strains

If genetic and epigenetic alterations are not responsible for cell-to-cell variability, stochastic events must be the source of heterogeneity. To test this idea we first asked whether stochastic DNA damage could explain aspects of the observed cell-to-cell variability, as a previous study had shown that yeast strains harboring chronic disomies experience increased DNA damage [18]. We deleted *RAD9*, a gene encoding a component of the DNA damage checkpoint pathway that is

responsible for delaying cell cycle progression in response to DNA damage (Figure 3.4). Deletion of *RAD9* decreased cell-to-cell variability in S+early M phase duration in wild-type cells (Figure 3.4G). This is not surprising because even wild-type cells experience DNA damage, especially during live-cell microscopy. The effects of deleting *RAD9* were even more dramatic in disomes and monosomes. Cell-to-cell variability in S+early M duration was dramatically decreased in all strains except monosome XI (Figure 3.4F). Deletion of *RAD9* only caused a minor decrease in G1 length variability in some aneuploid strains (Figure 3.4D,E). This minor effect on G1 duration variability is not surprising given that DNA damage is predominantly repaired in S phase and G2 in yeast [24]. We conclude that stochastic DNA damage is largely responsible for variability in S+early M phase length in all but one aneuploid strain analyzed.

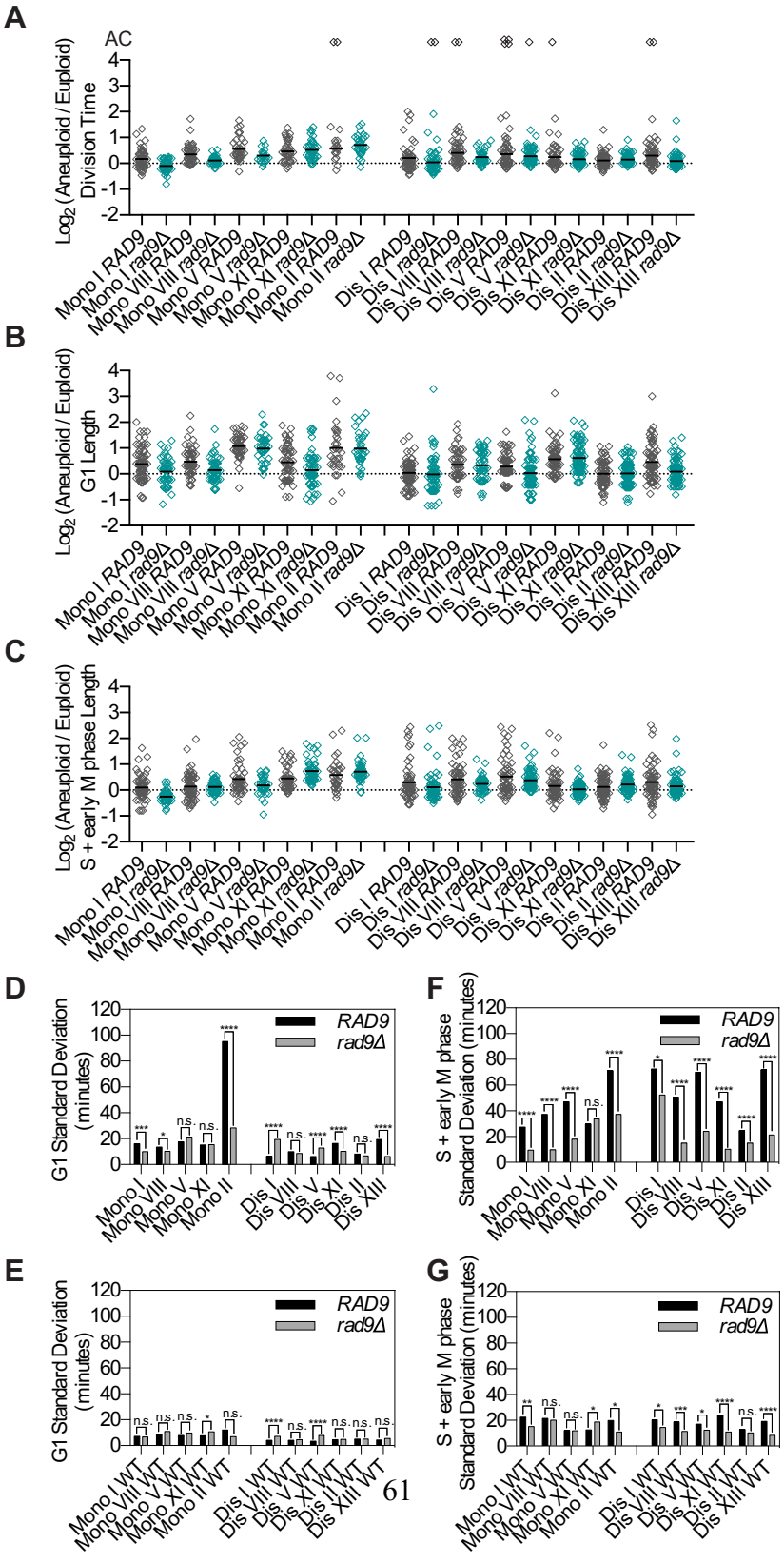

A
AC
Log2 (Aneuploid / Euploid) Division Time
B
Log2 (Aneuploid / Euploid) G1 Length
C
Log2 (Aneuploid / Euploid) S + early M phase Length
Mono I RAD9
Mono I rad9Δ
Mono VIII RAD9
Mono VIII rad9Δ
Mono V RAD9
Mono V rad9Δ
Mono XI RAD9
Mono XI rad9Δ
Mono II RAD9
Mono II rad9Δ
Dis I RAD9
Dis I rad9Δ
Dis VIII RAD9
Dis VIII rad9Δ
Dis V RAD9
Dis V rad9Δ
Dis XI RAD9
Dis XI rad9Δ
Dis II RAD9
Dis II rad9Δ
Dis XIII RAD9
Dis XIII rad9Δ
D
G1 Standard Deviation (minutes)
RAD9
rad9Δ
F
S + early M phase Standard Deviation (minutes)
E
Mono I WT
Mono VIII WT
Mono V WT
Mono XI WT
Mono II WT
Dis I WT
Dis VIII WT
Dis V WT
Dis XI WT
Dis II WT
Dis XIII WT
G

Figure 3.4 Cell-to-cell variability is in part due to stochastic DNA damage

Cells were grown and imaged as described in legend to Figure 3.2. Division time (A), G1 duration (B), and S+early M duration (C) were calculated as described in Figure 3.2. Standard deviations for euploid control populations (WT; E,G) and aneuploid (D,F) populations were measured and an F-test was used to test for equality of variance between *RAD9* (black bars, D-G) and *rad9*Δ (grey bars, D-G) populations (* = $p \leq 0.05$, (** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p < 0.0001$). The data for strains harboring a wild-type *RAD9* gene are the same data as shown in Figures 3.2 and 3.3 and are duplicated here for comparison to *rad9*Δ.

Deletion of *RAD9* not only decreased variability in S+early M phase length, it also slightly decreased average S+early M duration in some aneuploid strains (Figure 3.4C). To determine whether accelerating cell proliferation *per se* led to a decrease in cell-to-cell variability, we examined the consequences of deleting *UBP6* on cell cycle length heterogeneity. Deletion of *UBP6* attenuates levels of proteins whose abundance changes the most in disomic yeast strains and increases proliferative potential of most chronic disomes at 37°C and of three disomes (V, VIII, and XI) at 25°C [13,14]. Deletion of *UBP6* also decreased average G1 length immediately following mis-segregation of chromosomes V, VIII and XI but it did not affect cell cycle duration variability (Figure B.4F-L). Our results indicate that increased variability in cell cycle duration in disomic yeast strains is not solely a consequence of increased cell cycle length. Based on the observation that deletion of *UBP6* attenuates levels of the proteins most mis-regulated in the disomes yet does not affect variability, we further conclude that cell cycle length variability

is not caused by the most extremely deregulated genes in disomes.

The heterogeneity in cell cycle duration immediately following chromosome mis-segregation was dramatic, ranging in some instances from near wild-type division times to permanent cell cycle arrest in others. Is this heterogeneity maintained as cells continue to divide with an unbalanced karyotype? To address this question we measured cell cycle duration in chronic disomes.

Disomy of chromosome I, VIII, XIV or XVI causes significant variability in G1 and S+early M duration immediately following chromosome mis-segregation ("Acute Disomes"; Figure 3.5). Variability in G1 duration was attenuated in three of the four corresponding chronic disomic strains (Figure 3.5). Variability in S +early M phase was only attenuated in chronic disome I. We conclude that cell-to-cell variability in G1 duration is initially high following chromosome mis-segregation but is attenuated upon continuous selection for an aneuploid karyotype. These findings further suggest that G1 heterogeneity is selected against. Identifying the genetic and perhaps epigenetic alterations that reduce G1 variability in the chronic disomes will be of great interest.

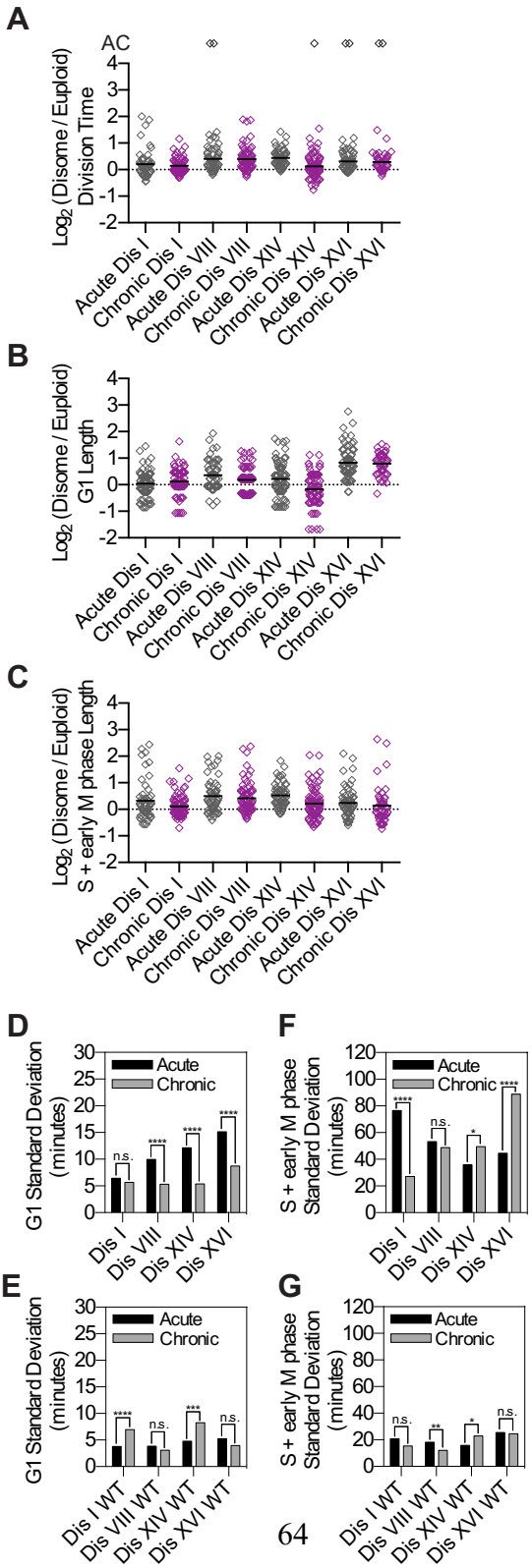

A
AC
Log2 (Disome / Euploid) Division Time
B
Log2 (Disome / Euploid) G1 Length
C
Log2 (Disome / Euploid) S + early M phase Length
Acute Dis I
Chronic Dis I
Acute Dis VIII
Chronic Dis VIII
Acute Dis XIV
Chronic Dis XIV
Acute Dis XVI
Chronic Dis XVI
D
G1 Standard Deviation (minutes)
Acute
Chronic
n.s.
****
Dis I
Dis VIII
Dis XIV
Dis XVI
F
S + early M phase Standard Deviation (minutes)
E
Dis I WT
Dis VIII WT
Dis XIV WT
Dis XVI WT
***
**
*
G

Figure 3.5 G1 length variability is attenuated in yeast strains harboring chronic disomies

Chronic disomes were grown to mid-log phase in SC-D and then plated on SC-D solid medium and imaged as described in legend to Figure 3.2. Division time (A), G1 duration (B), and S+early M phase duration (C) were measured as described in legend to Figure 3.2. Standard deviations for control euploid (WT; E,G) and disome (D,F) populations were measured and an F-test was used to test for equality of variance between the inducible (black bars, D-G) and chronic (grey bars, D-G) populations (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p < 0.0001$). The data shown for acute disomes are the same as in Figure 3.3. However, only data from the second cell division and beyond were included in this analysis to eliminate variation due to the carbon source switch used to induce chromosome mis-segregation, which was not done in the analysis of the chronic disomes.

3.5 Aneuploidy causes increased variability in the response to environmental perturbations

Next we wished to determine whether aneuploidy also causes increased cell-to-cell variability in other biological processes. We analyzed the response of disomes I, II, VIII, IX, XI, XIII, and XVI to three perturbations: galactose, DTT, and heat shock. Strains were analyzed at single-cell resolution (by flow cytometry) with fluorescent reporter fusions that are responsive to each perturbation (see Supplemental Experimental Procedures; Figure 3.6), as gene expression is a classic model in the study of population heterogeneity [25]. Additionally, disomes V, X, XII, and XIV were analyzed in a subset of the three environments (Figure B.5). Strains with *GAL1pr-YFP*

were grown in medium containing 0.25% galactose with no other sugars—a maximally inducing condition (Figure 3.6A-B). Strains with heat shock element (HSE) motifs (P_{4xHSE}-*YFP*), which are controlled by the Hsf1 transcription factor, were grown at 39°C (a temperature where most euploid cells robustly responded to heat-shock with minimal cell death; Figures 3.6C-D and S3.5A-B). Strains with a promoter containing four unfolded protein response element (UPRE) motifs, which are regulated by Hac1 (P_{4xUPRE}-*GFP*) were grown in medium containing 0.625mM DTT, a concentration where cells robustly respond (Figures 3.6E-F and S5C-D). While the mean significantly varied in 14 of the 26 strain condition pairings (p-values < 0.05 by Wilcoxon rank sum test; Figures 3.6A,C,E and B.5A,C) there was no generic pattern of whether the mean increased or decreased (Table S3.1). The standard deviation on the other hand increased in all 18 experimental-condition pairs where variation in the standard deviations was observed (p-values < 0.05 by Wilcoxon rank sum test; Figures 3.6B,D,F and S3.5B,D and Table S3.1). The fold increases in standard deviation varied from 1.03 to 1.5-fold. In total, our results indicate that aneuploidy leads to an increase in cell-to-cell variability in a number of different biological responses.

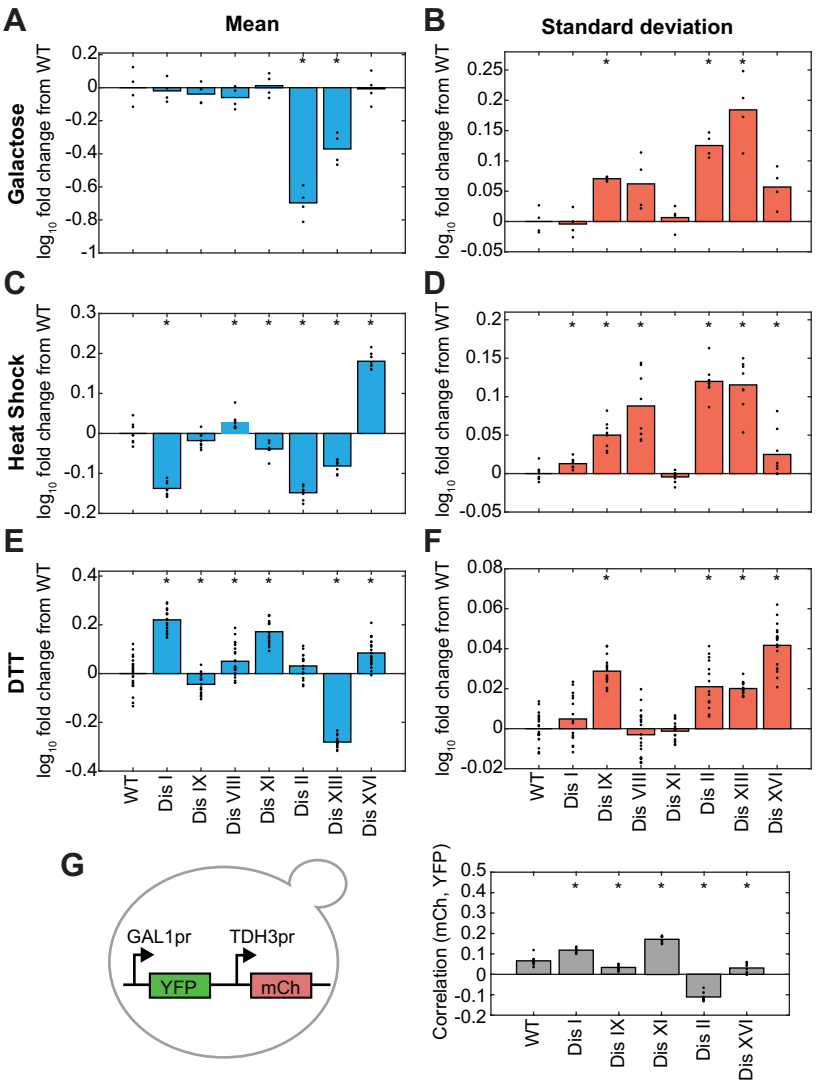


Figure 3.6 Disomes show increased cell-to-cell variability in response to environmental perturbations

Mean (A,C,E) and standard deviation (B,D,F) of *GAL1pr-YFP* (A-B), P_{4xHSE}-*YFP* (C-D), and P_{4xUPRE}-*GFP* (E-F) expression at steady-state in conditions that robustly induce each reporter construct. Error bars represent the standard error of the mean (n=4).

(G) Six strains were constructed that contained both a *GAL1pr-YFP* and a *TDH3pr-mCherry* construct. The correlation between the two fluorescent signals was calculated (n=4). All measurements (A-G) were normalized to SSC to account for differences in cell size. Asterisks (*) in A-G indicate statistical significance between the disome and euploid populations by Wilcoxon rank sum test. Abbreviations: WT, wild-type; Dis, disome; Glu, glucose; Gal, galactose. See also Figures B.5, B.6 and Table S3.1.

In addition to variations in the maximal response, variations can also exist in the kinetics of response to a signal and the concentration at which cells respond. Therefore, we titrated glucose concentrations to measure the concentration of glucose at which cells repress the *GAL1pr-YFP* reporter construct in medium that contains a constant amount of 0.25% galactose (Figure S3.6). We found that all disomes were more sensitive to glucose inhibition (p-values < 0.05 by Wilcoxon rank sum test; Figure S3.6B). In addition, we measured the time to induce *GAL1pr-YFP* following shift from raffinose to galactose (time = 0 in Figure S3.6C-D). All disomic populations, except for disome I, harbored a substantial fraction of cells that were delayed in induction of *GAL1pr-YFP* compared to euploid populations (Figure S3.6C-D). The time for all cells to completely induce the GAL response also varied among the disomes, in some cases increasing and in others decreasing. Together, these results support that the vast majority of disomes are more variable in their response to galactose.

The increase in standard deviation of the response could result from pathway specific increases in variation (intrinsic) or global increases in variation (extrinsic). Because we had measured a number of the strains harboring extra chromosomes under multiple conditions we were able ask

whether an increase in variation in response to one stimulus was correlated with an increase in variation in a second stimulus. The results suggest both extrinsic and intrinsic components. The extrinsic component is supported by several strains having higher standard deviations in all conditions (Figure 3.6B,D,F; significance of 0.06 by Wilcoxon rank sum test). The intrinsic component is supported by strains such as disome VIII that had increased variation in two conditions but not in a third (Figure 3.6B,D,F).

As an alternative way to assess the degree of intrinsic versus extrinsic noise, we created a wild-type and five disomes containing both a *GAL1pr-YFP* and a *TDH3pr-mCherry* construct. The correlation in expression of the two promoters in a single cell could then be assessed across the six strains (Figure 3.6G). After binning for cell size (see Supplemental Experimental Procedures), two of the strains had an increased correlation while three of the strains had a decreased correlation (p-values < 0.05 by Wilcoxon rank sum test). Our data show that aneuploid cells exhibit increased variability in their response to multiple different environmental stimuli caused by both intrinsic and extrinsic sources of variation.

3.6 Aneuploidy causes increased phenotypic variability in mammals.

Is increased population heterogeneity a unique feature of aneuploid budding yeast strains or a more general property of the aneuploid state? To address this question we examined the effects of chromosome gain on phenotypic variability in inbred mouse strains. We used Robertsonian translocations that were backcrossed at least 10 times into the C57BL/J6 background to generate isogenic euploid, trisomy 13, and trisomy 19 embryos [7,26]. In crosses involving Robertsonian

translocations, litter size is usually small and trisomic embryos represent less than 16% of the offspring [27], severely limiting the number of animals we were able to analyze. It was nevertheless evident that phenotypic variability was high among trisomy 19 embryos compared to isogenic euploid littermate controls. We observed variability in facial morphology, degree of nuchal edema and hemorrhaging at embryonic day 15.5 (Figure 3.7A). To quantify this variability we measured the maximal width of nuchal translucency, a classic prenatal test to assess degree of nuchal edema, which is an indicator of chromosomal abnormalities in human pregnancies. This analysis revealed that variability in nuchal edema thickness was significantly higher in trisomy 19 embryos than euploid control animals ($p<0.05$ by permutation analysis and $p<0.005$ by F-test on the variance; Figure 3.7B). Variability in morphology was also observed among trisomy 13 embryos even though we were only able to obtain one litter harboring trisomy 13 embryos (Figure 3.7C). We conclude that aneuploidy causes phenotypic variability in inbred trisomy 19 embryos and likely also in trisomy 13 embryos.

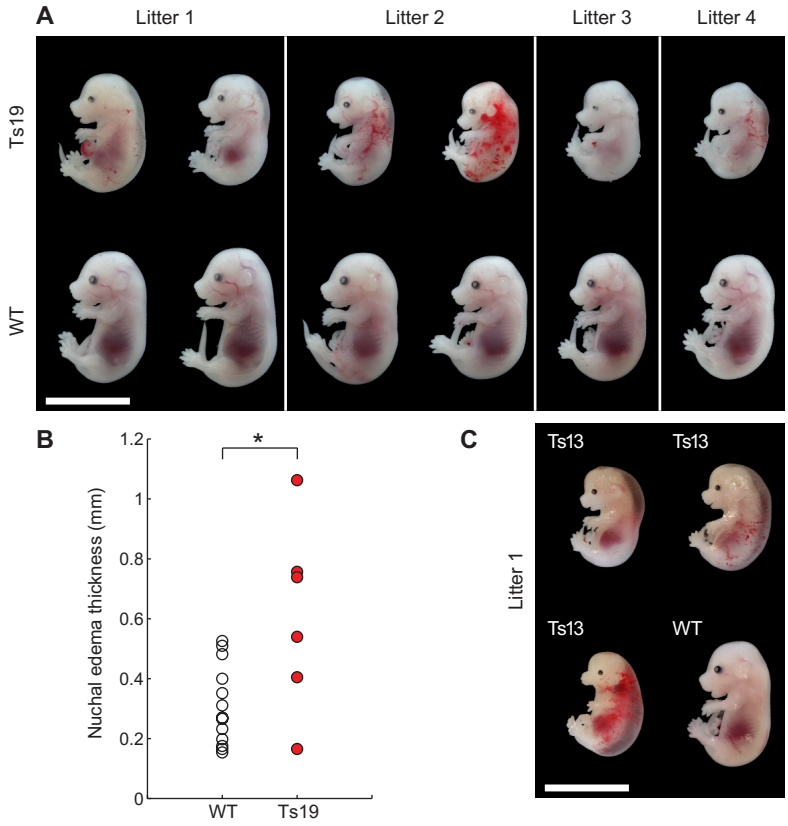


Figure 3.7 Non-genetic individuality in trisomic mice

(A) Images of trisomy 19 embryos and their representative euploid littermates at gestational stage E15.5. Bar, 10 mm. (B) Thickness of nuchal edema in wild-type and trisomy 19 E15.5 embryos shown in (A). (C) Images of trisomy 13 embryos and a WT littermate at E15.5. Bar, 10 mm.

3.7 Discussion

Phenotypic differences in genetically identical populations have been observed in a variety of experimental settings [23,25,28]. Variability can arise from stochastic synthesis and degradation

of molecules [29,30] as well as binomial partitioning at cell division [31]. Fluctuations are exacerbated when molecules are at low copy number or are segregated as a unit [32,33]. Mechanisms have been described that counteract these fluctuations to make biological processes robust (reviewed in [34], but when buffering mechanisms are perturbed, regulatory networks can exhibit alternate behaviors [35]. We show here that chromosome-scale changes in gene dosage severely impact the robustness of every biological network that we analyzed. We propose that this universal characteristic of the aneuploid state could help explain the variability in presentation and treatment responses of diseases caused by karyotypic abnormalities.

Aneuploidy decreases robustness of biological networks

Recent work has shown that karyotypically heterogeneous populations of aneuploid *S. cerevisiae* and *C. albicans* cells exhibit a wide range of phenotypes, particularly while under stress, and that this range of phenotypes is beneficial in adaption of the population to different environments [36,37]. To our knowledge, our work is the first to report that *karyotypically homogeneous* populations of aneuploid cells exhibit increased phenotypic variation as compared to euploid cells. All chromosome gains and losses that we analyzed cause a dramatic increase in variance in G1 and/or S+early M phase duration, with increases in both cell cycle stages for many aneuploidies. Two lines of evidence indicate that this variability is not due to genetic heterogeneity. First, the inducible system to generate aneuploid cells has the advantage that strains are propagated in the euploid state and are acutely induced to mis-segregate specific chromosomes, which avoids the accumulation of genetic alterations selected for by an aneuploid karyotype. Second, we find that cell cycle delays observed in an individual cell are not consistent

from one division to the next, indicating that the delays are caused by stochastic and not heritable events.

Heterogeneity in cell cycle duration was also observed in strains with chronic aneuploidies. Strains harboring chronic aneuploidies are variable for a number of different cellular responses in addition to cell cycle timing. We found that population variation in gene expression increased in multiple aneuploid strains in response to an alternate carbon source and in response to two different cellular stresses: heat shock and DTT. This increase in variability occurred across all tested conditions that led to a response in our reporters. Together, these results lead to the unanticipated conclusion that aneuploidy causes an increase in non-genetic variability in many different cellular responses. In fact, unaffected pathways may be in the minority.

Origins of non-genetic individuality

How does aneuploidy cause increased cellular variability? Previous studies suggested that deterministic processes are responsible for cell-to-cell variability in cell cycle progression in wild-type cells [38]. Our analyses indicate that stochastic DNA damage, which leads to activation of the DNA damage checkpoint pathway, is a significant source of heterogeneity in S+early M duration. The subtle S phase defects and genomic instability that aneuploid cells experience [18,21,22] are likely the cause for activation of the DNA damage checkpoint pathway in a fraction of aneuploid cells.

In contrast to variability in S+early M duration, inactivation of the DNA damage checkpoint pathway had little effect on G1 length heterogeneity in most aneuploid strains. What additional

factors contribute to the variability in this cell cycle stage remains to be determined. It is however worth noting that most stress response pathways delay cells in G1 [39] and cellular stresses have been shown to increase phenotypic variability under some conditions [40]. We propose that various stresses associated with the aneuploid state, such as protein folding, metabolic and oxidative stress, could contribute to the G1 length variability we observe in aneuploid strains. We further note that many different aneuploidies interfere with macromolecule biosynthesis [41], which controls G1 length (reviewed in [42]. Cell-to-cell variability in biosynthetic rates could thus also contribute to G1 length variability in aneuploid strains.

We also do not yet know how aneuploidy causes heterogeneity in gene expression. While change in gene dosage might affect a response—for instance disome XIII might affect the response to galactose by changing the dosage of the Gal80 repressor encoded on chromosome XIII—it is unclear why dosage changes would lead to an increase in population variability. Instead, the increase in variability may be caused by an indirect source such as cellular stress, which is a universal feature of the aneuploid state. Previous studies showed that environmental stresses cause a decrease in growth rate. This decreased growth rate can lead to lower levels of expression of some transcription factors, increasing their variability and thereby the variability of the responses they control [43]. In microorganisms, multiple lines of evidence support the idea that variability in response to cellular stress is advantageous [28]. Given that aneuploidy causes a variety of cellular stresses, this alone is likely to be a key contributor to the increased cell-to-cell variability that we observe. We also note that multiple lines of evidence indicate that phenotypic heterogeneity allows populations to bet-hedge under stressful situations, which decreases the

mean fitness of the population but increases the geometric mean of the population, thereby allowing cells to survive in extreme situations [44–49]. The increased variability in aneuploid cells could also be related to this phenomenon.

Ultimately, general features of the aneuploid state are likely to be the source of population heterogeneity. Many aneuploid yeast strains experience decreased protein synthesis rates [41] which could lead to decreased production of rare proteins. Unequal partitioning of such rare proteins could cause phenotypic variability. Changes in relative ratios of gene products also contribute to cell-to-cell variability. Higher base ploidy partially buffers against aneuploidy-induced cell-to-cell variability in S+early M phase. We do not yet know which genes decrease robustness upon doubling or halving of gene dosage. Very few genes exist in yeast that cause substantive cell cycle defects when their gene dosage is altered by the addition of a single extra copy [50] and haploinsufficiency is equally rare [51]. Subtle changes in the dosage of network nodes, such as transcription factors or regulatory enzymes (i.e. protein kinases) may nevertheless have wide-spread effects on the activity of the biological pathway that they control [35,52], making cells more susceptible to intracellular and extracellular noise.

Implications for cell-to-cell variability in wild-type populations

Fluctuations in mRNA synthesis, protein synthesis and degradation, and unequal segregation of proteins have all been implicated in producing cells with alternate fates in populations of genetically identical cells [29,31,53,54]. Our analyses of aneuploid cells are consistent with variability in these processes being sources of biological noise. Our data furthermore lead to the

remarkable conclusion that subtle changes in gene expression (50% decreases or increases) can cause significant cell-to-cell variability when they occur on a chromosome-wide scale. In contrast, changing the dosage of a single gene by 50% usually has little effect on the robustness of biological networks. In other words, simultaneously changing multiple nodes of a biological network in subtle ways can make the behavior of a pathway unpredictable. We propose that simultaneous subtle fluctuations in the levels and activities of network components could contribute to heterogeneity observed in wild-type cells.

Implications for disease presentation and treatment

Phenotypic variability is not confined to aneuploid yeast strains. Trisomic embryos obtained from crosses of inbred mouse strains exhibit variability in embryonic morphology. Based on this observation, we propose that non-genetic heterogeneity could contribute to the dramatic phenotypic variability seen in constitutional aneuploidies in humans and, of course, cancer. Comparing phenotypic variability between inbred and outbred aneuploid yeast strains may in fact provide an opportunity to assess the degree of variability conferred by the aneuploid state itself and allelic variation in the population.

Non-genetic heterogeneity could also provide an explanation for why aneuploidy is such a frequent occurrence in experimental evolution. The phenotypic variability associated with the aneuploid state could well provide opportunity for adaptation. In this regard, it is interesting to note that increased variance has the potential to enhance population fitness in euploid populations [23]; therefore, it is possible that aneuploidy occurs in evolution as a mechanism to

further increase population variance.

We note that aneuploidy-associated non-genetic variability has profound implications for aneuploidy as a therapeutic target in cancer and fungal infections. Non-genetic heterogeneity could make therapies targeting specific aneuploidies difficult. Treatments that target general features of the aneuploid state may thus be more likely to be effective. In summary, our results indicate that decreased robustness of many, if not all, biological processes is a general feature of the aneuploid state. It will be interesting to determine how increased variability in single cell behavior leads to increased variability of multicellular processes.

AUTHOR CONTRIBUTIONS

R.B., C.R., B.H., C.B., C.M., P.H., R.S., M.S. and A.A. designed the experiments. R.B, C.R., B.H., C.B., C.M., P.H. and R.S. performed the experiments and analyzed results. R.B., C.R., B.H., C.B., C.M., P.H., R.S., M.S. and A.A. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Kirk Anders, Doug Koshland and David Pincus for reagents. This work was supported by the National Institutes of Health (CA206157-22 to A.A.), the Curt Marble Cancer Research Fund (to A.A.), in part by the Koch Institute Support Grant P30-CA14051, National Science Foundation Graduate Research Fellowships (C.R., C.M.) and a National Science Foundation grant 1349248 (M.S. and B.H.). A.A. is an investigator of the Howard Hughes Medical Institute

and of the Glenn Foundation for Biomedical Research.

Method

Yeast strains, plasmids and growth conditions

Yeast strains were generated and manipulated as described previously (Guthrie and Fink, 1991). All yeast strains used in this study are listed in Table S2.2. 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). For live-cell imaging, cells were grown in synthetic complete medium at 25°C supplemented with 2% raffinose (SC-R), 2% raffinose + 2% galactose (SC-RG), or 2% glucose (SC-D). Growth conditions for individual experiments are described in the Figure legends.

Mis-segregation frequency of non-targeted chromosomes

After mis-segregation in SC-RG, cultures were released into SC-D and fixed with 3.7% formaldehyde after 3 and 5 hours of growth at room temperature. The number of GFP dots per cell was counted for 100 unbudded cells at each time point. Unbudded cells containing 2 GFP dots were classified as having mis-segregated the GFP marked chromosome.

Galactose and stress response reporter analyses

All *GAL1pr-YFP, UPRE-GFP*, and *HSE-YFP* reporter experiments were performed in synthetic dropout medium containing 1.7g/L yeast nitrogen base (BD Difco), 1g/L monosodium glutamate, 200µg/mL Geneticin (G418), and histidine dropout complete supplement mix

(Sunrise Science), plus carbon sources. In the galactose response experiments, outgrowth cultures were grown in synthetic dropout medium supplemented with 2% raffinose, and induction assays were performed in synthetic dropout base supplemented with 0.25% galactose and a two-fold gradient of glucose ranging from 1.0%-0.002% (including a 0% condition). For the stress response experiments, all cultures were grown in synthetic dropout medium supplemented with 2% glucose. All 30ºC incubations for the *GAL1pr-YFP, UPRE-GFP*, and *HSE-YFP* reporter experiments were performed in a humidified incubator (Infor Multitron) with rotary shaking at 230 rpm (tubes and flasks) or 999 rpm (deep 96-well plates). Higher-temperature incubations were conducted in an incubating microplate shaker (VWR) with rotary shaking at 1000rpm.

Construction of conditional centromere strains

Strains with conditional centromeres were constructed using a PCR-based method [9,56]. Briefly, the conditional centromere construct was amplified from plasmid p1888 using primers designed to target the conditional centromere construct to a particular. Strain A2587 was transformed with the conditional centromere constructs as previously described [56]. We replaced the endogenous centromere with the conditional centromere construct for every single chromosome (I-XVI). With the exception of cells harboring a conditional centromere on chromosome XV, all strains containing the conditional centromere construct exhibited doubling times equal to that of wild-type cells when grown in medium containing glucose (data not shown). We also generated strains harboring multiple conditional centromeres. This allowed us to examine the consequences of mis-segregating multiple (up to 8) chromosomes. Importantly,

this system also enabled us—for the first time—to systematically analyze the consequences of chromosome losses on yeast cell physiology.

CEN-LacO plasmid construction

To follow chromosome mis-segregation, we introduced a lac operator (*E. coli*, *lacO*) array near the centromere of the chromosome harboring the conditional centromere (Figure S3.1A). When a GFP-LacI fusion protein is expressed in these cells, it binds to the *lacO* sequences and forms a GFP dot at *lacO* array sites visible by fluorescence microscopy [12].

Plasmids targeting the LacO array to various chromosomes (CEN-LacO plasmids;) were constructed by cloning a homology region to the specific target site with XhoI restriction sites into the SalI cut plasmid p1499 (pCM40). Plasmids were integrated at the target site by restriction enzyme digest using the enzymes. Transformants were screened for gain of nourseothricin resistance (100 μg/ml). All plasmids containing the LacO array were propagated in Max Efficiency Stbl2 competent cells (Life Technologies, Grand Island, NY) due to the high propensity of the LacO array to recombine. All enzymes used for cloning were obtained from New England BioLabs (Ipswich, MA).

HYGRO-LacO plasmid construction

A plasmid targeting the LacO array to the *hph* hygromycin B resistance gene (HYGRO-LacO) was constructed in the same way as the Cen-LacO plasmids, using homology to the *hph* hygromycin B resistance gene. Integration of the HYGRO-LacO plasmid used a two-step process. First, the *hph* hygromycin B resistance gene was integrated into the desired site in the

genome using a PCR-based method [56]. Second, the HYGRO-LacO plasmid was integrated at the *hph* site by EagI digest. Transformants were screened for loss of hygromycin resistance and gain of nourseothricin resistance.

GFP dot strain construction

Strains containing the LacO array integration were crossed to a strain containing the GFP-LacI fusion protein. To integrate the GFP-LacI construct into the genome, plasmid p1801was linearized with NheI and transformed into A2587 cells as previously described (Guthrie and Fink, 1991).

Other constructs

Spc42-dsRed and *Spc42-GFP* were constructed by PCR-based methods [56]. *mCherry-Cdc3* was constructed using the *YIp211-CDC3-mCherry* integrating plasmid [57]. The *GAL1pr-YFP* reporter was integrated at the *HO* locus [58]. The P_{4xHSE}-*YFP* and P_{4xUPRE}-*GFP* reporters were constructed from a single-integrating plasmid [59]. To make the dual YFP-mCherry reporter for the assays described in Figure 3.6G, a constitutively-expressed *TDH3pr-mCherry* construct was integrated in tandem with the *GAL1pr-YFP* reporter at the *HO* locus.

Fluorescence microscopy

For GFP dot imaging, cells were either imaged live or fixed in 3.7% formaldehyde for 15 minutes at room temperature and then resuspended in 0.1 M potassium phosphate (KPi pH 6.4)/1.2 M sorbitol. To visualize biotin labeling, cells were fixed in 3.7% formaldehyde for 15 minutes at room temperature and then resuspended in 0.1 M potassium phosphate (KPi pH

6.4)/1.2 M sorbitol with 1 μg/ml streptavidin conjugated to Alexa 568 fluorophore (Molecular Probes). Cells were then washed once in 0.1 M potassium phosphate (KPi pH 6.4)/1.2 M sorbitol. Cells were imaged using a Zeiss Axioplan 2 microscope (Carl Zeiss) and a Hamamatsu OCRA-ER digital camera (Hamamatsu).

Biotin labeling and sorting

Cells were arrested with α-factor in YEP-R medium at 30°C and labeled with biotin [60]. Upon release from the G1 arrest, cells were switched to medium containing galactose (YEP-RG) to inactivate *CEN4*. Cells were then transferred to YEP-D medium at 30°C to stop further chromosome mis-segregation and grown for 1 or 5 hours. To purify populations of disomic cells for proteomic quantification following chromosome mis-segregation, we used a previously described method to purify mother cells [60]. Cells were grown to mid log phase in YEP-R and then arrested in G1 with the yeast mating pheromone α-factor (5 μg/ml). Following G1 arrest, cells were washed 3 times in phosphate buffered saline (PBS) pH 8.0 and labeled with EZ-Link Sulfo-NHS-LC-Biotin (1 mg biotin in PBS pH 8.0 per 1 OD_{600} unit of cells; Thermo Fisher Scientific) at 4°C for 30 minutes. Biotin forms permanent amide bonds with primary amines on cell surface proteins, thus labeling the cell wall of the G1 arrested cells. As the cell wall is built *de novo* during budding, release from the G1 block in the absence of biotin results in the mother cell being labeled with biotin while the bud is not labeled. Following biotin labeling, cells were washed 4 times with PBS pH 8.0/100 mM glycine to remove excess biotin.

After biotin labeling, cells were transferred into YEP-RG medium to induce chromosome mis-

segregation. Cells were grown for 120 minutes in YEP-RG at 30°C (one cell division). After nearly all cells had finished dividing, cells were transferred to YEP-D and grown for 1 hour. To purify biotin labeled cells, cells were first washed with PBS pH 7.4/0.5% bovine serum albumin (BSA) and then incubated with anti-biotin magnetic beads (Miltenyi Biotec) in PBS pH 7.4/0.5% BSA for 15 minutes at 4°C. Cells were washed once more with PBS pH 7.4/0.5% BSA and then resuspended in PBS pH 7.4/0.5% BSA and run over LS columns on a QuadroMACS separator (Miltenyi Biotec) to retain biotin-labeled cells. Biotin labeled cells were eluted from the columns with PBS pH 7.4/0.5% BSA following removal from the QuadroMACS separator. Cells were then washed once with PBS pH 7.4 and resuspended in YEP-D and grown at 30°C. Samples for analysis were taken at times indicated in Figure 3.1B.

Protein quantification by mass spectrometry

Preparation of protein extracts was adapted from Dephoure et al. (2014). Briefly, cells were lysed using a FastPrep-24 (MP Biomedicals) in a buffer containing 8 M urea, 75 mM NaCl, and 50 mM Tris-Cl pH 8.2 using 3-6 cycles of 45 s at 4°C separated by 5 minute incubations at 4°C. Samples were broken until ~70% of cells had lysed as confirmed by light microscopy. Lysates were cleared by centrifugation at 14,000x*g* for 15 minutes at 4°C. Protein concentrations were determined using the Bradford protein assay (Bio-Rad). Further sample preparation, TMT labeling, and sample fractionation was performed as described in Dephoure et al. (2014). Samples were run with 3 technical replicates on a Thermo QExactive mass spectrometer (Thermo Fisher Scientific). Database searching was conducted using Mascot (Matrix Science).

Protein ratios were normalized to account for minor variations in sample mixing by centering the $\log_2$ protein abundance ratio distribution over zero, as described in Dephoure et al. (2014). When calculating the normalization factor, proteins on the disomic chromosomes were excluded. We expect most proteins (except those on the disomic chromosome) to be present at a one-to-one ratio.

Calculation of expectation value for protein abundance ratio after chromosome mis-segregation.

While mis-segregation occurs in many of the cells in the population, mis-segregation itself and the isolation of disomic cells are not fully efficient. One hour following chromosome mis-segregation, the culture contained 67% cells disomic for chromosome IV, 20% euploid cells, and 13% cells lacking chromosome IV (nullisomes). Five hours following chromosome mis-segregation, the population was comprised of 61% disomes, 34% euploid cells, and 5% nullisomes. The slight decrease in the proportion of disomic cells at the 5 hour time point compared to the 1 hour time point is likely due to the reduced proliferation rate of disome IV cells compared to euploid cells.

The expected ratio of proteins encoded on chromosome IV from disome IV and euploid control strains depends on two factors: 1) the percent of disomes, euploid cells, and nullisomes, and 2) the rate of protein turnover from degradation versus dilution. Given the measured percent of each karyotype at one and five hours, an expected value for the protein abundance ratio can be calculated for different models of protein turnover. In the extreme where proteins are only diluted by proliferation (using proliferation rate measured in Figure 3.3A and correcting for the

ratio compression; see below) one would expect a $\log_2$ protein abundance of 0.12 at 1 hour and 0.28 at 5 hours. In the other extreme where protein turnover is instantaneous, the expectation is a $\log_2$ protein abundance of 0.27 at 1 hour and 0.28 at 5 hours. One hour following chromosome mis-segregation, proteins encoded on chromosome IV were elevated compared to proteins encoded by the euploid chromosomes, with a mean $\log_2$ protein abundance ratio of 0.21 for proteins on chromosome IV, compared to a mean of 0 for proteins on all other chromosomes (Figure 3.1C). Levels of chromosome IV-encoded proteins increased further to a mean $\log_2$ protein abundance ratio of 0.32 five hours following chromosome mis-segregation (Figure 3.1D).

Calculation of expected ratios:

Given that previous work had shown that proteins encoded on disomic chromosomes exhibit a protein ratio of 2:1 versus proteins encoded by euploid chromosomes [13], we assumed that the observed $\log_2$ protein abundance ratio of 1.4 for proteins encoded on disomic chromosome IV was due to ratio compression. We can therefore calculate a ratio compression factor for our samples which when can then be use to calculate the actual ratio using the following formula $\frac{2+x}{1+x}=1.4$ which yields a ratio compression factor (x in the equation) of 1.66.

To calculate the expected $\log_2$ protein abundance ratio we first used the following relationship:

$$\text{ChromosomeContentInDisome} = \%\text{Disome} * \left(1 * 2^{-\left(\frac{time}{growthrateofdisome}\right)} + 2 * \left(1 - 2^{-\left(\frac{time}{growthrateofdisome}\right)}\right)\right)$$

$$+\%\text{Euploid} * (1) + \%\text{Nullisome} * \left(1 * 2^{-\left(\frac{time}{growthrateofnullisome}\right)}\right)$$

When protein turnover is high this relationship simplifies to:

$$\text{ChromosomeContentInDisome} = \%\text{Disome} * \left(2 * \left(1 - 2^{-\left(\frac{time}{growthrateofdisome}\right)}\right)\right) + \%\text{Euploid} * (1)$$

The expectation value for the $\log_2$ protein abundance ratio is calculated with the following formula

$$\log_2\left(\frac{\text{ChromosomeContentInDisome} + 1.66}{1 + 1.66}\right)$$

Aneuploid fitness colony measurements

Cells were grown in YEP-R and then transferred to YEP-RG for 160 minutes to induce chromosome mis-segregation. Cells were then plated on YEP-D agar plates. Mother-daughter pairs that had just completed a cell division were separated by micromanipulation and allowed to grow into colonies at 30°C. Colony size was measured 40 – 48 hours later when colony size was still increasing exponentially. Colony area was measured using ImageJ.

To calculate aneuploid to euploid colony area ratio, euploid colony area was measured and averaged per plate, with a separate average for euploid mothers and euploid daughters. Euploid cells on the same plate as the aneuploid cells were used for analysis to control for plate-to-plate variability. A value for each individual aneuploid colony was calculated as a ratio of the aneuploid colony area to the mean euploid colony area for euploid cells. For disomes and trisomes, the mean colony area for euploid mothers was used. For monosomes, the mean colony area for euploid daughters was used. Each circle in Figure 3.1F represents the mean of all

aneuploid/euploid ratios for a particular karyotype. Error bars indicate the standard deviation of the aneuploid/euploid ratios, not accounting for error propagation from euploid cells.

Chromosome mis-segregation was confirmed by determining whether colonies could grow on plates lacking uracil. The inducible chromosome is marked with *URA3* at the conditional centromere, so colony pairs in which one colony (arising from the daughter cell) is auxotrophic for growth on uracil and the other colony (arising from the mother cell) is prototrophic for growth on uracil indicates a chromosome mis-segregation event. For aneuploid strains with severe fitness defects, colonies were grown at 30°C until colonies were large enough to test for uracil prototrophy via replica-plating. Colonies were also tested for growth on plates containing glycerol as the sole carbon source. Colonies that did not grow on this medium were excluded from analysis as they are respiratory deficient, which causes slow growth. Rarely, the inducible chromosome was segregated into the daughter cell instead of the mother; these mis-segregation events were excluded from analysis.

The colony mis-segregation assay in haploid cells was used to determine the chromosome mis-segregation rates presented in Figure S3.1B. The growth defects observed in monosomes were more severe than the growth defects observed in disomes and trisomes (Figures 3.1F, S3.1B). While the sizes of many monosomic colonies were less than 1% of the sizes of colonies formed by euploid cells after ~48 hours of growth, nearly all monosomic strains eventually formed colonies after many days of growth. The only monosome that was inviable was chromosome XIII monosomy, due to monosomy of *TUB1* which is lethal [61]. Disomy and trisomy were better tolerated than monosomy, with the exception of disomy VI and trisomy VI. These

chromosomal gains are inviable due to an increase in *TUB2* copy number [61].

Differentiating aneuploid from euploid cells in live-cell cell cycle analysis

For cell cycle analysis, aneuploid cells were differentiated from euploid cells by counting the number of GFP dots present in a cell. In all euploid strains, 1 GFP dot per cell indicates euploidy. For disomes and trisomes of single chromosomes, the presence of 2 GFP dots in the cell indicates disomy or trisomy while 1 GFP dot indicates euploidy. For disomes and trisomes of 2 chromosomes, 4 GFP dots indicates disomy or trisomy. For disomes and trisomes of 3 chromosomes, cells with 4, 5 or 6 GFP dots were scored as disomes and trisomes, as it was difficult to accurately count GFP dots at numbers greater than 4 dots per cell. For all monosomes (single or multiple chromosomes), 0 GFP dots per cell indicates monosomy. In all analyses euploid cells were either from the wild-type control strain or from cells in the experimental strain that did not mis-segregate a chromosome.

Cell cycle analysis image acquisition

For inducible aneuploids, cells were grown in SC-R to mid-log phase then transferred to SC-RG and grown for 160 minutes to induce chromosome mis-segregation. Chronic disomic strains were grown in SC-D medium to mid-log phase. Then, for both inducible aneuploids and chronic disomes, cells were transferred to SC-D and layered on an agar pad (2% agarose, SC-D) affixed to a glass slide and covered with a cover slip. Cells were imaged at 5-minute intervals for 8-10 hours using a Zeiss Axio Observer-Z1 with a 63x objective (Carl Zeiss), equipped with a Zeiss Definite Focus module (Carl Zeiss) and a Hamamatsu ORCA-AG digital camera (Hamamatsu).

9 Z-stacks (0.6 microns apart) were acquired and maximally projected. MetaMorph software (Molecular Devices) was used for image acquisition and processing.

Cell cycle analysis data processing

For cell cycle analyses of aneuploid cells, approximately 20 aneuploid cells and 10 euploid cells were scored per experiment. G1 duration, S+early M phase duration, anaphase onset to cytokinesis duration, and division time (cytokinesis to cytokinesis) were measured for each cell following chromosome mis-segregation. Analyses of cells started either at bud emergence or at anaphase onset in the first division following chromosome mis-segregation. We were not able to image cells during the first G1 phase following chromosome mis-segregation because of the time needed to prepare the cells for imaging. Thus, division time analyses start in the second division following chromosome mis-segregation.

At the end of the time-lapse, cells that were in G1 and had spent at least 100 minutes in G1 were scored as arrested and included in G1 duration calculations. Cells that were in G1 at the end of the time-lapse and had spent less than 100 minutes in G1 were not considered arrested and excluded from analysis for calculating G1 duration. At the end of the time-lapse, cells that had passed bud emergence but had not reached anaphase onset and had spent at least 200 minutes in that stage were scored as arrested and included in S+early M phase duration calculations. Cells that had passed bud emergence at the end of the time-lapse and had not reached anaphase onset but had spent less than 200 minutes in this stage were not considered arrested and excluded from analysis for calculating S+early M phase duration.

To calculate standard deviation of cell populations and compare variance using the F-test, raw data (in minutes) were used (Figures 3.2B-E, 3.3B,C,F,G,I-K, 3.4D-G, 3.5D-G, and S3.4I-L). These analyses were done on the combined data from all divisions measured following chromosome mis-segregation for all cells of a given karyotype for a particular cell cycle stage (G1, S+early M phase, or division time duration). Aneuploid cells were compared to euploid cells from the same experiment to account for minor day-to-day variability.

In Figures 3.2A, 3.3A,E, 3.4A-C, 3.5A-C, S3.2B-D, S3.3A-C,E-G, and S3.4F-H, cell cycle measurements for aneuploid cells were normalized to euploid cells from the same experiment to allow us to compare cell cycle measurements among experiments. Each aneuploid cell measurement for G1 duration, S+early M phase duration, anaphase duration, or total division time was taken as a ratio of the mean of all euploid cells for the same measurement from the same division. For example, the G1 duration measurement for an inducible aneuploid cell in its third division following chromosome mis-segregation would be normalized to the mean of the G1 duration for all euploid cells in the third division following chromosome mis-segregation. Normalization to the euploid strain controls for day-to-day variability in experiments as well as changes in cell cycle durations upon shift from raffinose+galactose (SC-RG) as the carbon source to glucose (SC-D) as the carbon source. In the same manner, measurements of cell cycle duration in chronic disomes were normalized to euploid controls by cell cycle stage and division. Log_2 transformed aneuploid to euploid ratios are plotted in the figures.

In Figures 3.2F and 3.3D,H, the longest G1 or S+early M phases were removed from the aneuploid data set until the standard deviation of the aneuploid cell cycle stage was equal to that

of the corresponding euploid cell cycle stage in the same experiment. Standard deviations were considered equal when no longer significantly different by F-test. The percent of the total number of measured cell cycle stages (G1 or S+early M phase) that were removed to equalize standard deviations of euploid and aneuploid data is reported.

Galactose response assays

Strains were inoculated from colonies streaked on a G418/-his plate into liquid dropout medium supplemented with 2% glucose and incubated at 30ºC for 24 hours until cultures reached saturation. Cultures were diluted 1:100-1:400 into medium containing 2% raffinose and grown for 16-18 hours. For the steady-state glucose gradient assay, raffinose cultures at a final $OD_{600} \leq$ 0.1 were pelleted, washed twice, and then resuspended 1:1 in synthetic dropout medium without carbon and diluted 1:100 into the glucose gradient to measure expression of the *GAL1* reporter.

For the kinetic response assays, raffinose cultures with a final OD_{600} 0.1-0.3 were selected; a 0 hour time point of the raffinose culture was harvested by addition of a concentrated TE/sodium azide 'stop solution' to a final (1X) concentration of 0.1% sodium azide. The remaining culture was pelleted, washed twice, and then resuspended at a 1:4 dilution into G418/-his liquid medium supplemented with 2% galactose. Aliquots were harvested every 30 min for up to 8 hours after the medium switch, and at 9 and 10 hours post medium switch.

Dual-reporter noise assay

Strains bearing a tandem *GAL1pr-YFP-TDH3pr-mCherry* reporter were inoculated from colonies streaked on a G418/-his plate into liquid dropout medium supplemented with 2% glucose and

incubated at 30ºC for 24 hours until cultures reached saturation. Cultures were diluted 1:200 into medium supplemented with 2% galactose and grown to saturation. Cultures were then diluted 1:10 into fresh 2% galactose medium and grown for 8 hours before they were analyzed by flow cytometry. Each strain was measured in eight replicates.

Stress response assays

Strains were inoculated from colonies streaked on a G418/-his plate into liquid G418/-his medium and incubated at 30ºC for 24 hours until cultures reached saturation. Cultures were diluted 1:100-1:400 and grown for 8-12 hours until they reached an OD_{600} of ~0.1-0.2.

For the heat-shock experiment, the outgrowth cultures were diluted 1:50 into pre-warmed medium in a 96 deep well plate, and grown for 4 hours with rotary shaking at the specified temperature. After incubation, cultures were pelleted by centrifugation, washed twice, and resuspended in Tris-EDTA pH 7.5 (TE) before analysis on the flow cytometer.

For the DTT stress response experiments, the outgrowth cultures were diluted 1:4 into medium containing a final concentration of either 0.625mM DTT or a 2-fold serial dilution ranging from 10-0.0098mM (including a 0mM condition). All cultures were grown for 4 hours at 30ºC before being harvested by addition of concentrated TE/sodium azide 'stop solution' to a final (1X) concentration of 0.1% sodium azide.

Flow cytometry and data analysis

All samples were run on a Stratedigm S1000EX cytometer with FSC/SSC thresholds set to 200. PMT gain settings of FITC 0.5%, FSC 34.3%, SSC 40.0% were used for the dual-reporter noise

assay. PMT gain settings of FITC 2.0%, FSC 33.5%, SSC 55.3% were used for all other assays. Flow cytometry data were analyzed using custom MATLAB code (https://github.com/springerlab/Flow-Cytometry-Toolkit). As a general workflow, raw FCS data were processed by gating out debris and bacterial contamination events based on FSC/SSC measurements, normalizing the YFP values to cell size by dividing by SSC values, rescaling the values by an arbitrary factor of $10^{3.5}$ for visualization purposes, and computing population statistics, expression histograms, and other metrics of induction from the $\log_{10}$-transformed data.

For the steady-state galactose and stress response experiments, induced fractions were calculated relative to the probability distribution of cells grown only in glucose-containing medium in the absence of stressors. Separate "off", "inducing", and "on" fractions were similarly defined relative to the probability distributions from maximally induced and uninduced conditions (see diagram), with threshold concentrations being calculated from a cubic spline (MATLAB function csaps with smoothing parameter of 0.9) that was fit to the data by interpolating between consecutive points of normalized log-transformed YFP values plotted against concentration (glucose or DTT).

Statistical significance testing

For analyses conducted in Figure 3.6A-F (all galactose induction and stress response assays), statistical significance of variability between wild-type and aneuploid was assessed by two-sided Wilcoxon rank sum tests on the values of each given metric (mean expression level and standard deviation of the expression level). Results of the significance testing are summarized in Table

S3.1.

For the dual-reporter variability assay (Figure 3.6G), data were analyzed using customized *MATLAB* code. For each sample, all events were equally binned into 30 bins by SSC, as a proxy for cell size. Events from each bin were treated as if they were from cells with similar cell sizes. The Pearson correlation coefficients between mCherry and YFP were calculated for events in each SSC bin, and averaged among bins representing intermediate cell sizes, defined as 40~60 percentile.

To determine whether the difference in variance in nuchal edema thickness (Figure 3.7B) was statistically significant, we performed two statistical tests: 1) we calculated an F-statistic on the variance ($p<0.005$) and 2) we permuted all the nuchal edema thickness values, randomly split them into two groups - 'wt' and 'trisomy 19' - and determined whether the variance was at least as large as observed in the real data (39500 out of 1,000,000, or $p = 0.0395$).

Mouse embryo imaging and analysis

Trisomic embryos were generated by crossing C57BL/6J mice with congenic mice heterozygous for two Robertsonian chromosomes [7,62]. The following Robertsonian fusion chromosome were used for these crosses: B6.Cg-Rb(11.13)4Bnr/JAmonJ; B6.Cg-Rb(13.16)1Mpl/JAmonJ; B6.Cg-Rb(16.17)7Bnr/JAmonJ; B6.CgRb(5.19)1Wh/JAmonJ; B6Ei.Cg-Rb(9.19)163H/J.

To isolate embryos of a certain gestational stage, the day on which a copulatory plug was observed was defined as embryonic day E0.5. Embryos were collected at the indicated gestational stages. Amniotic membranes and portions of the umbilical cords were carefully

removed, then rinsed twice with PBS before placing on a 10-cm petri dish for imaging. A Canon EF 100mm f/2.8L Macro IS USM lens was used to attain the close-up magnification of 1x at the closest focusing. Aperture-stop was chosen to cover sufficient depth of field while keeping the macrophotography diffraction limited. Thickness of the nuchal edema, defined as the translucent region in the neck and head region, was measured in the region of the embryo, where the width was maximal. The nuchal edema thickness of an embryo, in pixel number, was calculated using the Ruler Tool of Adobe Photoshop 6.0, and then converted to metrics given the pixel pitch of the camera sensor.

Reference

1. Roper RJ, Reeves RH. Understanding the Basis for Down Syndrome Phenotypes. PLoS Genet. 2006;2: e50. doi:10.1371/journal.pgen.0020050

2. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. Hum Mol Genet. 2009;18: R1–R8. doi:10.1093/hmg/ddp011

3. Siegel JJ, Amon A. New insights into the troubles of aneuploidy. Annu Rev Cell Dev Biol. 2012;28: 189–214. doi:10.1146/annurev-cellbio-101011-155807

4. Dodgson SE, Kim S, Costanzo M, Baryshnikova A, Morse DL, Kaiser CA, et al. Chromosome-Specific and Global Effects of Aneuploidy in Saccharomyces cerevisiae. Genetics. 2016;202: 1395–1409. doi:10.1534/genetics.115.185660

5. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012;8: 608. doi:10.1038/msb.2012.40

6. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science. 2007;317: 916–924. doi:10.1126/science.1142210

7. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science. 2008;322: 703–709. doi:10.1126/science.1160058

8. Alvaro D, Sunjevaric I, Reid RJD, Lisby M, Stillman DJ, Rothstein R. Systematic hybrid LOH: a new method to reduce false positives and negatives during screening of yeast gene deletion libraries. Yeast Chichester Engl. 2006;23: 1097–1106. doi:10.1002/yea.1423

9. Anders KR, Kudrna JR, Keller KE, Kinghorn B, Miller EM, Pauw D, et al. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. BMC Genet. 2009;10: 36. doi:10.1186/1471-2156-10-36

10. Hill A, Bloom K. Genetic manipulation of centromere function. Mol Cell Biol. 1987;7: 2397–2405.

11. Murray AW, Szostak JW. Pedigree analysis of plasmid segregation in yeast. Cell. 1983;34: 961–970.

12. Straight AF, Belmont AS, Robinett CC, Murray AW. GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. Curr Biol. 1996;6: 1599–1608. doi:10.1016/S0960-9822(02)70783-5

13. Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, et al. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife. 2014;3: e03023. doi:10.7554/eLife.03023

14. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of aneuploidy-tolerating mutations. Cell. 2010;143: 71–83. doi:10.1016/j.cell.2010.08.038

15. Christoforou AL, Lilley KS. Isobaric tagging approaches in quantitative proteomics: the ups and downs. Anal Bioanal Chem. 2012;404: 1029–1037. doi:10.1007/s00216-012-6012-9

16. Kim HB, Haarer BK, Pringle JR. Cellular morphogenesis in the Saccharomyces cerevisiae cell cycle: localization of the CDC3 gene product and the timing of events at the budding site. J Cell Biol. 1991;112: 535–544.

17. Donaldson AD, Kilmartin JV. Spc42p: a phosphorylated component of the S. cerevisiae spindle pole body (SPD) with an essential function during SPB duplication. J Cell Biol. 1996;132: 887–901.

18. Blank HM, Sheltzer JM, Meehl CM, Amon A. Mitotic entry in the presence of DNA

damage is a widespread property of aneuploidy in yeast. Mol Biol Cell. 2015;26: 1440–1451. doi:10.1091/mbc.E14-10-1442

19. Nicholson JM, Macedo JC, Mattingly AJ, Wangsa D, Camps J, Lima V, et al. Chromosome mis-segregation and cytokinesis failure in trisomic human cells. eLife. 2015;4. doi:10.7554/eLife.05068

20. Passerini V, Ozeri-Galai E, de Pagter MS, Donnelly N, Schmalbrock S, Kloosterman WP, et al. The presence of extra chromosomes leads to genomic instability. Nat Commun. 2016;7: 10754. doi:10.1038/ncomms10754

21. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy drives genomic instability in yeast. Science. 2011;333: 1026–1030. doi:10.1126/science.1206412

22. Zhu J, Pavelka N, Bradford WD, Rancati G, Li R. Karyotypic determinants of chromosome instability in aneuploid budding yeast. PLoS Genet. 2012;8: e1002719. doi:10.1371/journal.pgen.1002719

23. Cerulus B, New AM, Pougach K, Verstrepen KJ. Noise and Epigenetic Inheritance of Single-Cell Division Times Influence Population Fitness. Curr Biol CB. 2016;26: 1138–1147. doi:10.1016/j.cub.2016.03.010

24. Weinert T, Hartwell L. The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. Science. 1988;241: 317–322. doi:10.1126/science.3291120

25. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science. 2005;309: 2010–2013. doi:10.1126/science.1105891

26. Gropp A, Winking H, Zech L, Müller H. Robertsonian chromosomal variation and identification of metacentric chromosomes in feral mice. Chromosoma. 1972;39: 265–288.

27. Pfau SJ, Amon A. A System to Study Aneuploidy In Vivo. Cold Spring Harb Symp Quant Biol. 2015;80: 93–101. doi:10.1101/sqb.2015.80.027193

28. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;135: 216–226. doi:10.1016/j.cell.2008.09.050

29. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297: 1183–1186. doi:10.1126/science.1070919

30. Paulsson J. Summing up the noise in gene networks. Nature. 2004;427: 415–418.

doi:10.1038/nature02257

31. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat Genet. 2011;43: 95–100. doi:10.1038/ng.729

32. Jajoo R, Jung Y, Huh D, Viana MP, Rafelski SM, Springer M, et al. Accurate concentration control of mitochondria and nucleoids. Science. 2016;351: 169–172. doi:10.1126/science.aaa8714

33. Uphoff S, Lord ND, Okumus B, Potvin-Trottier L, Sherratt DJ, Paulsson J. Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. Science. 2016;351: 1094–1097. doi:10.1126/science.aac9786

34. Arias AM, Hayward P. Filtering transcriptional noise during development: concepts and mechanisms. Nat Rev Genet. 2006;7: 34–44. doi:10.1038/nrg1750

35. Mileyko Y, Joh RI, Weitz JS. Small-scale copy number variation and large-scale changes in gene expression. Proc Natl Acad Sci. 2008;105: 16659–16664. doi:10.1073/pnas.0806239105

36. Chen G, Mulla WA, Kucharavy A, Tsai H-J, Rubinstein B, Conkright J, et al. Targeting the Adaptability of Heterogeneous Aneuploids. Cell. 2015;160: 771–784. doi:10.1016/j.cell.2015.01.026

37. Hickman MA, Paulson C, Dudley A, Berman J. Parasexual Ploidy Reduction Drives Population Heterogeneity Through Random and Transient Aneuploidy in Candida albicans. Genetics. 2015;200: 781–794. doi:10.1534/genetics.115.178020

38. Sandler O, Mizrahi SP, Weiss N, Agam O, Simon I, Balaban NQ. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. Nature. 2015;519: 468–471. doi:10.1038/nature14318

39. Herrero E, Angeles de la Torre M, Torres J, Belli G. Regulation of G1 phase of yeast cells by stress. In: Boonstra J, editor. G1 Phase Progression. New York: Kluwer Academic/Plenum; 2003. pp. 165–186.

40. Axelrod K, Sanchez A, Gore J. Phenotypic states become increasingly sensitive to perturbations near a bifurcation in a synthetic gene network. eLife. 2015;4. doi:10.7554/eLife.07935

41. Thorburn RR, Gonzalez C, Brar GA, Christen S, Carlile TM, Ingolia NT, et al. Aneuploid yeast strains exhibit defects in cell growth and passage through START. Mol Biol Cell. 2013;24: 1274–1289. doi:10.1091/mbc.E12-07-0520

42. Turner JJ, Ewald JC, Skotheim JM. Cell Size Control in Yeast. Curr Biol. 2012;22: R350–R359. doi:10.1016/j.cub.2012.02.041

43. Kashiwagi A, Urabe I, Kaneko K, Yomo T. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. PloS One. 2006;1: e49. doi:10.1371/journal.pone.0000049

44. Acar M, Mettetal JT, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. Nat Genet. 2008;40: 471–475. doi:10.1038/ng.110

45. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. Nature. 2009;462: 90–93. doi:10.1038/nature08504

46. Casanueva MO, Burga A, Lehner B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. Science. 2012;335: 82–85. doi:10.1126/science.1213491

47. Fridman O, Goldberg A, Ronin I, Shoresh N, Balaban NQ. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. Nature. 2014;513: 418–421. doi:10.1038/nature13469

48. Levy SF, Ziv N, Siegal ML. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol. 2012;10: e1001325. doi:10.1371/journal.pbio.1001325

49. New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, et al. Different levels of catabolite repression optimize growth in stable and variable environments. PLoS Biol. 2014;12: e1001764. doi:10.1371/journal.pbio.1001764

50. Bonney ME, Moriya H, Amon A. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. Genes Dev. 2015;29: 898–903. doi:10.1101/gad.261743.115

51. Deutschbauer AM. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics. 2005;169: 1915–1925. doi:10.1534/genetics.104.036871

52. Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, et al. Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. Cell. 2008;135: 879–893. doi:10.1016/j.cell.2008.09.039

53. Blake WJ, KÆrn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422: 633–637. doi:10.1038/nature01546

54. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. Nature. 2006;444: 643–646. doi:10.1038/nature05316

55. Guthrie C, Fink GR, editors. Guide to yeast genetics and molecular biology. San Diego, Calif.: Acad. Press; 1991.

56. Longtine MS, McKenzie A, Demarini DJ, Shah NG, Wach A, Brachat A, et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast Chichester Engl. 1998;14: 953–961. doi:10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U

57. Fang X, Luo J, Nishihama R, Wloka C, Dravis C, Travaglia M, et al. Biphasic targeting and cleavage furrow ingression directed by the tail of a myosin II. J Cell Biol. 2010;191: 1333–1350. doi:10.1083/jcb.201005134

58. Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proc Natl Acad Sci U S A. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

59. Damon JR, Pincus D, Ploegh HL. tRNA thiolation links translation to stress responses in Saccharomyces cerevisiae. Mol Biol Cell. 2015;26: 270–282. doi:10.1091/mbc.E14-06-1145

60. Smeal T, Claus J, Kennedy B, Cole F, Guarente L. Loss of transcriptional silencing causes sterility in old mother cells of S. cerevisiae. Cell. 1996;84: 633–642.

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

62. Pfau SJ, Silberman RE, Knouse KA, Amon A. Aneuploidy impairs hematopoietic stem cell fitness and is selected against in regenerating tissues in vivo. Genes Dev. 2016;30: 1395–1408. doi:10.1101/gad.278820.116

Chapter 4. Conclusion

The genetic architecture of a quantitative trait often refers to the number and identity of causal genetic variants in a natural population, as well as the interaction among them. While many methods have been used to identify such variants in nature, most methods like GWAS and linkage analysis require testing a large number of genetic variants. This makes the identification of small effect-size variants difficult. In this thesis, by experimentally testing the effects of individual genetic changes, either single gene knockout (in Chapter 2) or adding a whole chromosome (in Chapter 3), we were able to unbiasedly measure the impact of many mutations on multiple quantitative traits. We found that the number of genes that are capable of affecting quantitative traits is large (a quarter to half of assayed non-essential genes), and many of such genes are functionally related to core cellular processes. These observations lead to a few interesting research directions that I would like to present in this chapter.

The work presented in Chapter 2 suggests a necessity to understand quantitative traits in the context of an interconnected genetic interaction network, i.e. genes outside of canonical signaling pathway can also influence quantitative traits. A limitation of this work is that all analyses were conducted using traits in yeast, which is a relatively simple organism compared to higher eukaryote systems. Given the current advances in gene editing technologies, like CRISPR, it would be interesting to perform similar quantitative screens in a more complex system, in order to examine the generality of the conclusion. Such efforts would be one step closer to elucidate the genetic architecture of important traits in humans.

While this work tried to uncover all the important non-essential genes that have the potential to affect quantitative traits, the actual genetic variants in a natural population might only be a subset of what we uncovered in this work. This is especially important as 1) it is possible that many genes that are involved in core cellular processes are more conserved than other genes, given their functional importance, and 2) it is unclear how genetic changes on essential genes affect quantitative traits. The exact identity of genetic variants in a natural population depends on selection and genetic drifts, and hence may be difficult (if not impossible) to understand. It would be interesting to conduct lab evolution experiments on quantitative traits, with controlled selection conditions and population history. By examining the genetic differences between evolved strains and parental strain, especially those genetic variants of small effects, it would be interesting to understand which genes actually carry mutations that affect quantitative traits.

Another observation from this work is that the effect size distributions of all genes that have measurable effects on quantitative traits are close to an exponential distribution. It is unclear if this is a coincidence or not. It would be interesting to examine what are the necessary and sufficient conditions to get an exponentially-distributed effect size distribution. More specifically, it would be interesting to set up a model that consist of interactions similar to known biological networks, like metabolic network, and examine the effect size of each node in the network by simulating the impact on the network after removing that node. By changing the network structure/topology and the function form of these interactions, one could examine the conditions under which the effect size distribution is similar to those observed experimentally.

How do the genes identified in Chapter 2 affect each of the studied quantitative traits? The

answer to this question might seem trivial as these core cellular processes are expected to affect many biological processes. However, we often observed that perturbing a particular cellular process only affects specific traits. It would be interesting to study how this is achieved from a mechanistic perspective.

Finally, it would also be interesting to ask if signaling pathway can make better decisions by taking inputs from those identified core processes. Cells need to make decisions in complicated environments, where multiple conditions can change simultaneously. It is hence possible that by integrating additional information from those core processes, signaling pathways can make smarter decisions and hence gain fitness advantages in complex environments.

Appendix A: Supplementary material for Chapter 2

Supplemental figures and legends

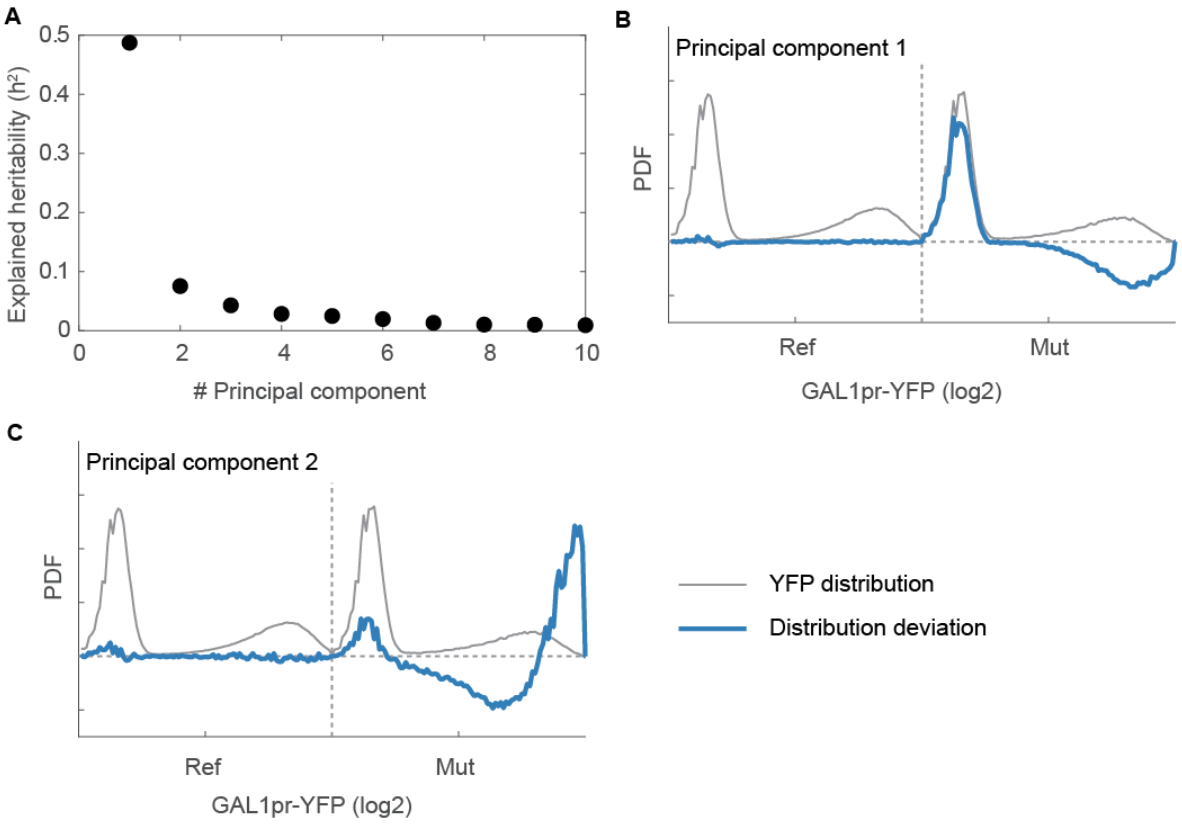


Figure S2.1. Determining modes of response with principal component analysis

After data segmentation, histograms of GAL1pr-YFP for the mutant and reference strain for each sample were normalized, concatenated, and then analyzed using principal component analysis. (**A**) The fraction of variation explained by the first ten principal components. (**B-C**) Effects on GAL1pr-YFP distribution by the top two principal components. The average GAL1pr-YFP distribution of all reference and mutant strains are concatenated (gray). The principal component

(blue) from the PCA analysis is the deviation from this average profile due to mutant effects. The horizontal line y=0 means no effects; i.e. the behavior of the wild-type strain. Note that the first two principal components correspond to biological properties, i.e. the induced fraction and induction level.

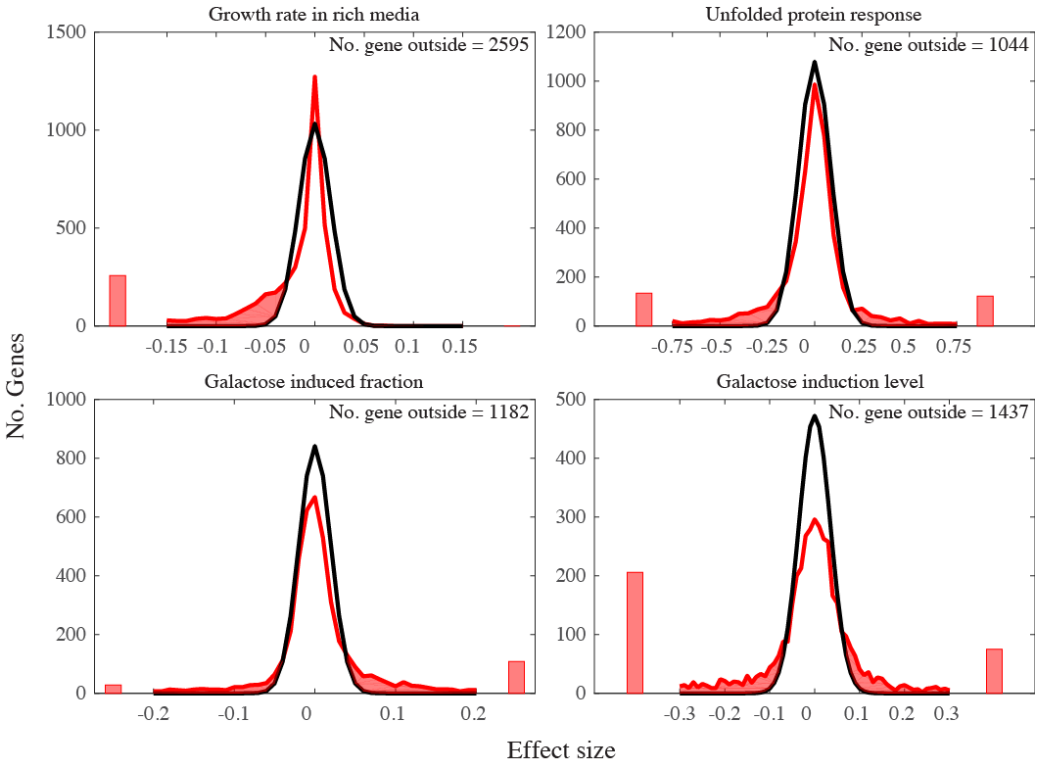


Figure S2.2. Effect size distribution versus measurement noise for four traits

As many mutants have effect sizes that are close to or smaller than average measurement noise, the total number of genes that affects each quantitative trait was estimated by comparing the measured effect size distribution (red) and measurement noise effect size distribution (black).

The measurement noise effect size distribution is the distribution of measurement noises between all replicated samples. The total number of genes that affect each trait was estimated from the number of genes in the shaded region.

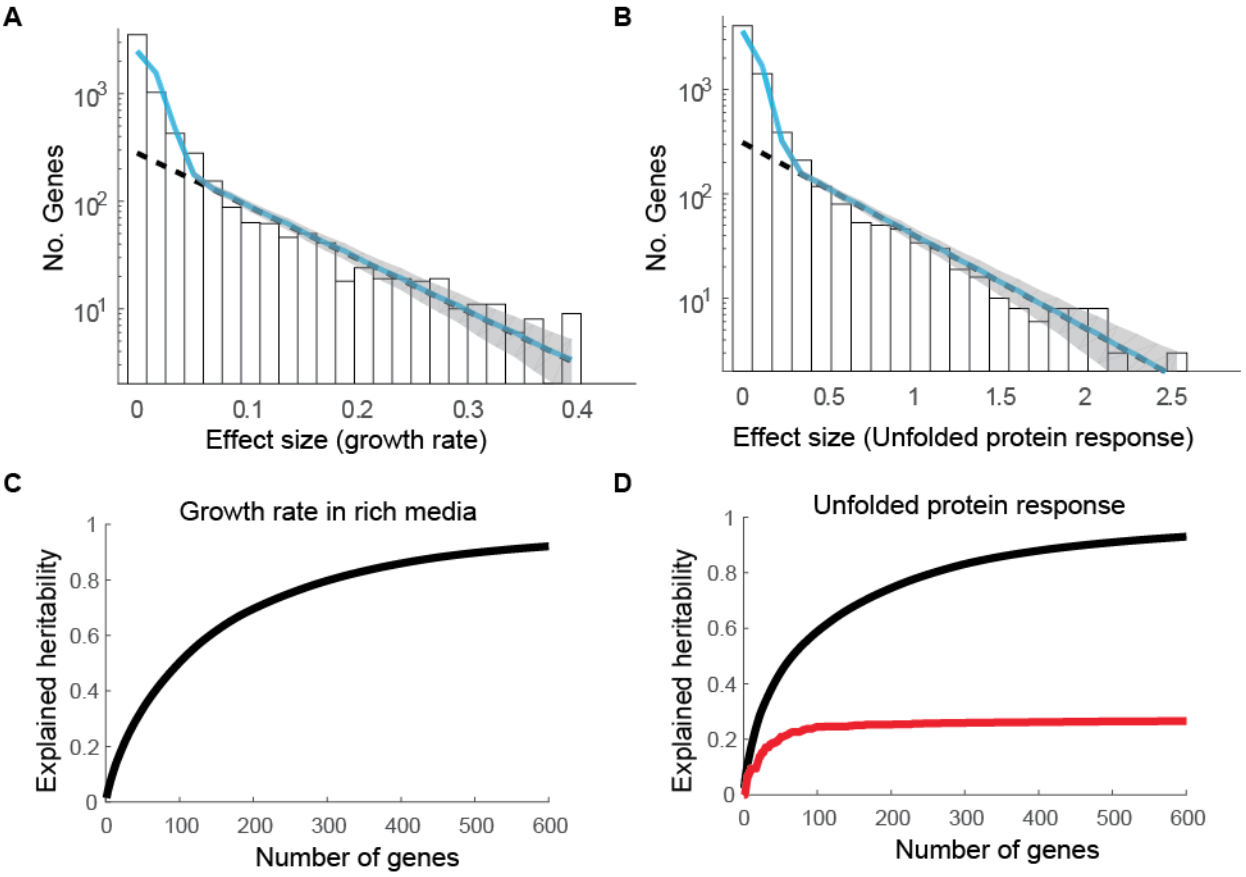


Figure S2.3. Reanalysis of two screens confirms that a large number of genes quantitatively affect yeast galactose response

Data from two deletion studies [5,6], one on growth rate in rich medium and one on the unfolded protein response (UPR), were reanalyzed. Both the effect size distribution (**A-B**) and explained heritability (**C-D**) were calculated as in Figure 2.2 and 2.3. Fit of the significant genes to an

exponential (dashed line) has an R^2 of 0.91 for growth rate (**A**) and 0.94 for UPR (**B**). The fit of the full data to an exponential plus noise had an R^2 of 9.2 (**A**) and 0.95 (**B**). (**C-D**) The contribution to explained heritability, as calculated in 2.3A, from UPR genes (red) or all genes (black) for growth rate in rich media (**C**) and UPR (**D**).

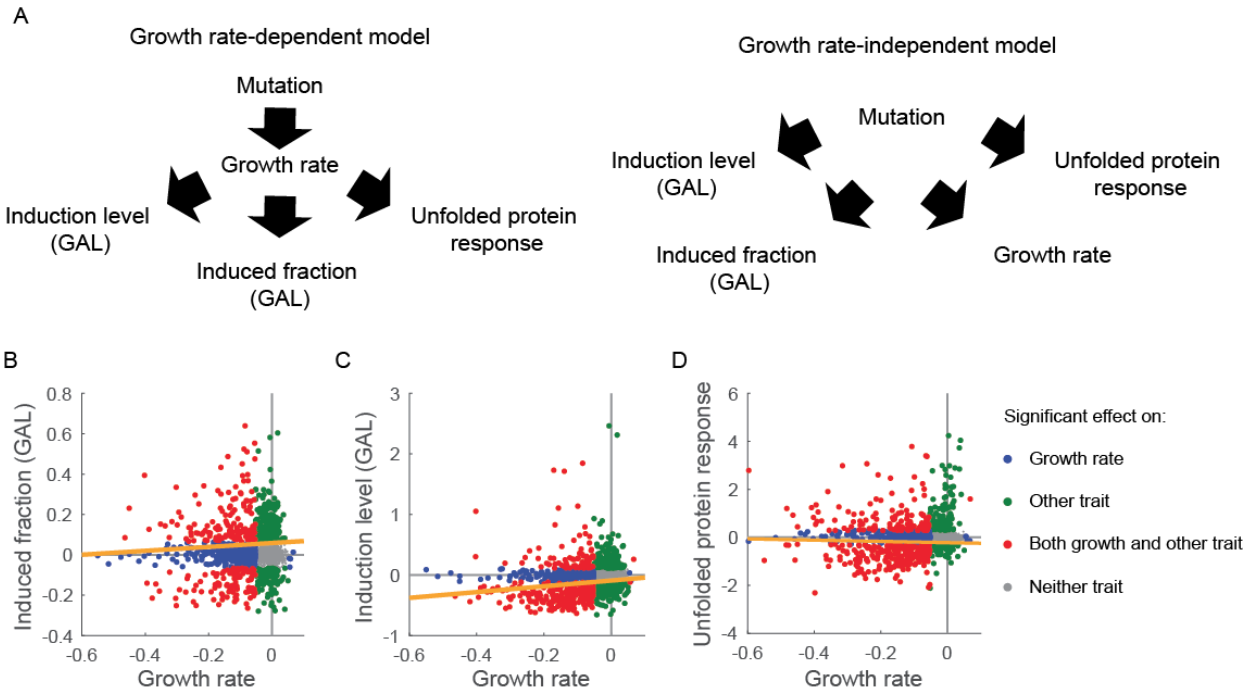


Figure S2.4. Affecting growth rate is not the sole mechanism for significant mutants to affect yeast GAL response and unfolded protein response

(**A**) Two alternative models of how quantitative traits can be affected by gene deletion. In the growth rate-dependent model (left), mutants affect growth rate that in turn affects other traits. In the growth rate-independent model (right), mutants directly affect quantitative traits including growth rate. These two models can be distinguished by determining whether mutant effects on

growth rate and other traits are correlated. (**B-D**) Mutant phenotypes for the unfolded protein response (**B**), GAL induced fraction (**C**) and GAL induction level (**D**) are plotted against the growth rate data reported by Breslow et al. Mutants were segmented into four quadrants based on whether the mutant had a significant effect (based on 0.5% FDR cut-off) on growth rate and non-growth rate trait: growth rate (blue), other non-growth rate trait (green), both (red), neither (gray). A linear fit of the points that are significant for both traits (red) is plotted (orange line).

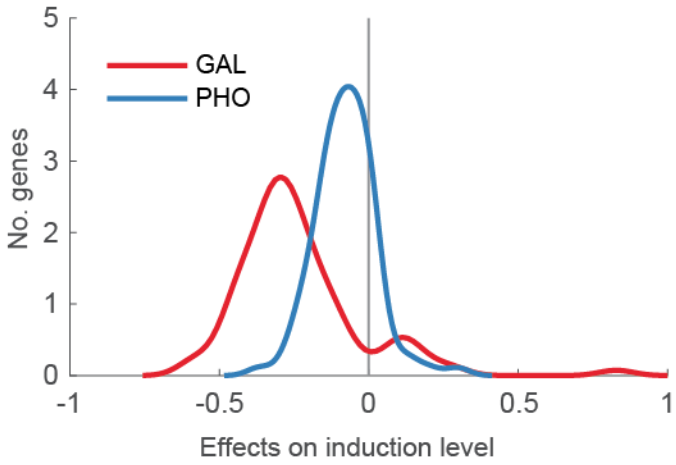


Figure S2.5. The difference of effects on GAL and PHO response by deleting genes involved in protein synthesis

For each of the 95 mutants we tested that are involved in protein synthesis, the mutant effects on the induction level were quantified for the PHO (blue) and GAL (red) responses. The effect size distribution was smoothened with kernel smoothing with a bandwidth of 0.05. The two distributions are extremely unlike to have results from noise in a single distribution (p-value $3*10^{-10}$, two-tailed t-test). The magnitude of the average difference in effect size between the two

distributions is 3 fold.

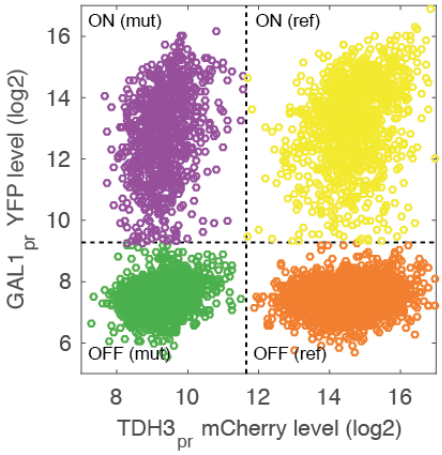


Figure S2.6. Data segmentation example

An example is shown here using data from the first replicate sample for mutant *yal068c*Δ. Cell debris is filtered from the raw data using a FSC/SSC gate, and the mCherry vs. YFP values of remaining events are plotted. Data is segmented on the mCherry channel to separate reference and mutant strain, and on the YFP channel to separate the induced cells and uninduced cells. The horizontal and vertical dashed lines show the threshold used for segmentation.

Supplemental tables and legends

Table S2.1. Enriched Gene Ontology for genes that significantly affect all four yeast traits

GO TermFinder [4] was used to analyze GO enrichment. The p-values are corrected for multiple hypotheses. Significant GOs are defined by the ones with corrected p-value less than 0.01.

GO ID	GO Term	Corrected P-Value
GO:0010467	gene expression	9.1E-21
GO:0016070	RNA metabolic process	2.1E-15
GO:0034641	cellular nitrogen compound metabolic process	2.8E-14
GO:0006807	nitrogen compound metabolic process	2.4E-13
GO:0090304	nucleic acid metabolic process	5.8E-13
GO:0044260	cellular macromolecule metabolic process	1.6E-12
GO:0044271	cellular nitrogen compound biosynthetic process	2.2E-12
GO:0043170	macromolecule metabolic process	9.1E-12
GO:0034645	cellular macromolecule biosynthetic process	1.2E-10
GO:0009059	macromolecule biosynthetic process	1.8E-10
GO:0006139	nucleobase-containing compound metabolic process	4.2E-10
GO:0043933	macromolecular complex subunit organization	2.3E-09
GO:0046483	heterocycle metabolic process	3.1E-09
GO:0006725	cellular aromatic compound metabolic process	4.2E-09

Table S2.1 (Continued)

GO ID	GO Term	Corrected P-Value
GO:1901360	organic cyclic compound metabolic process	1.4E-08
GO:0044238	primary metabolic process	2.7E-08
GO:0006396	RNA processing	1.2E-07
GO:0044237	cellular metabolic process	1.2E-07
GO:0044249	cellular biosynthetic process	3.1E-07
GO:0002181	cytoplasmic translation	3.5E-07
GO:0006355	regulation of transcription, DNA-templated	3.8E-07
GO:1903506	regulation of nucleic acid-templated transcription	3.8E-07
GO:2001141	regulation of RNA biosynthetic process	3.8E-07
GO:0071824	protein-DNA complex subunit organization	4.1E-07
GO:0010468	regulation of gene expression	5.8E-07
GO:0044267	cellular protein metabolic process	6.7E-07
GO:0051252	regulation of RNA metabolic process	7.2E-07
GO:1901576	organic substance biosynthetic process	8.1E-07
GO:0006351	transcription, DNA-templated	8.8E-07
GO:0032774	RNA biosynthetic process	8.8E-07
GO:0097659	nucleic acid-templated transcription	8.8E-07
GO:2000112	regulation of cellular macromolecule biosynthetic process	1.2E-06
GO:0071704	organic substance metabolic process	1.3E-06

Table S2.1 (Continued)

GO ID	GO Term	Corrected P-Value
GO:0010556	regulation of macromolecule biosynthetic process	1.7E-06
GO:0019219	regulation of nucleobase-containing compound metabolic process	2.0E-06
GO:0009058	biosynthetic process	2.0E-06
GO:0019538	protein metabolic process	2.1E-06
GO:0031326	regulation of cellular biosynthetic process	4.3E-06
GO:0009889	regulation of biosynthetic process	5.1E-06
GO:0022613	ribonucleoprotein complex biogenesis	6.2E-06
GO:0034654	nucleobase-containing compound biosynthetic process	7.4E-06
GO:0051171	regulation of nitrogen compound metabolic process	8.4E-06
GO:0006325	chromatin organization	9.1E-06
GO:0008152	metabolic process	1.3E-05
GO:0006412	translation	1.3E-05
GO:0060255	regulation of macromolecule metabolic process	1.8E-05
GO:0043043	peptide biosynthetic process	1.8E-05
GO:0018130	heterocycle biosynthetic process	2.8E-05
GO:0034728	nucleosome organization	3.3E-05
GO:0043604	amide biosynthetic process	3.9E-05
GO:0034660	ncRNA metabolic process	3.9E-05
GO:0019438	aromatic compound biosynthetic process	4.0E-05

Table S2.1 (Continued)

GO ID	GO Term	Corrected P-Value
GO:0019222	regulation of metabolic process	6.8E-05
GO:0034622	cellular macromolecular complex assembly	6.9E-05
GO:0080090	regulation of primary metabolic process	8.8E-05
GO:1901362	organic cyclic compound biosynthetic process	9.1E-05
GO:0034470	ncRNA processing	9.3E-05
GO:0016568	chromatin modification	1.0E-04
GO:0042254	ribosome biogenesis	1.1E-04
GO:0051276	chromosome organization	1.1E-04
GO:0006364	rRNA processing	1.2E-04
GO:0031323	regulation of cellular metabolic process	1.3E-04
GO:0043486	histone exchange	1.3E-04
GO:0006518	peptide metabolic process	1.4E-04
GO:0071840	cellular component organization or biogenesis	2.4E-04
GO:0030490	maturation of SSU-rRNA	3.5E-04
GO:0016072	rRNA metabolic process	4.7E-04
GO:0071822	protein complex subunit organization	4.8E-04
GO:0043603	cellular amide metabolic process	7.3E-04
GO:0065003	macromolecular complex assembly	1.1E-03
GO:0044085	cellular component biogenesis	1.1E-03

Table S2.1 (Continued)

GO ID	GO Term	Corrected P-Value
GO:0043044	ATP-dependent chromatin remodeling	1.3E-03
GO:0010629	negative regulation of gene expression	1.7E-03
GO:0045892	negative regulation of transcription, DNA-templated	2.0E-03
GO:0051253	negative regulation of RNA metabolic process	2.0E-03
GO:1902679	negative regulation of RNA biosynthetic process	2.0E-03
GO:1903507	negative regulation of nucleic acid-templated transcription	2.0E-03
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2.1E-03
GO:0006338	chromatin remodeling	3.7E-03
GO:0042274	ribosomal small subunit biogenesis	3.8E-03
GO:0010558	negative regulation of macromolecule biosynthetic process	4.0E-03
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	4.0E-03
GO:0016569	covalent chromatin modification	5.0E-03
GO:0016570	histone modification	5.0E-03
GO:0006357	regulation of transcription from RNA polymerase II promoter	5.7E-03
GO:0051172	negative regulation of nitrogen compound metabolic process	8.8E-03
GO:0031327	negative regulation of cellular biosynthetic process	9.4E-03
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	9.6E-03

Table S2.2. The number of genes that affect growth rate and each of the three non-growth traits

Genes that significantly affect the unfolded protein response, induced fraction (GAL), and induction level (GAL) were compared to the genes that significantly affect growth rate. The total number of genes that were measured in both growth rate and the other trait is listed. Of this total number, the number that significantly affected growth rate, significantly affected the non-growth rate trait, and significantly affected both traits is listed.

	No. genes also assayed in growth rate screen	No. genes significantly affect growth rates	No. genes significantly affect the non-growth rate trait	No. genes significantly affect both
Unfolded protein response	4152	779	594	369
Induced fraction (GAL)	3869	634	595	254
Induction level (GAL)	3869	634	744	316

Table S2.3. Significantly spatially clustered Gene Ontology

For each Gene Ontology that is spatially clustered, the direction in the four-trait space is shown. Significant GOs were defined by FDR < 0.01.

Table S2.3 (Continued)

GO ID	GO Term	Effect direction			
		Growth rates	Unfolded protein response	GAL induced fraction	GAL Induction level
GO:0000027	ribosomal large subunit assembly	-0.39	-0.63	-0.13	-0.66
GO:0000030	mannosyltransferase activity	-0.03	0.78	0.42	0.45
GO:0000070	mitotic sister chromatid segregation	-0.63	-0.23	0.66	0.33
GO:0000153	cytoplasmic ubiquitin ligase complex	0.01	0.99	0.04	0.09
GO:0000280	nuclear division	-0.85	-0.24	0.42	0.21
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	-0.67	-0.5	0.22	-0.5
GO:0000470	maturation of LSU-rRNA	-0.36	-0.64	0.11	-0.67
GO:0000479	endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	-0.59	-0.5	0.12	-0.62
GO:0000812	Swr1 complex	-0.36	0.18	-0.55	-0.73
GO:0000819	sister chromatid segregation	-0.7	-0.26	0.61	0.27
GO:0000839	Hrd1p ubiquitin ligase ERAD-L complex	0.01	0.99	0.04	0.09
GO:0001300	chronological cell aging	-0.1	0.02	0.99	0.08
GO:0002181	cytoplasmic translation	-0.49	-0.57	0.14	-0.65
GO:0003723	RNA binding	-0.69	-0.55	0.19	-0.42
GO:0003735	structural constituent of ribosome	-0.55	-0.57	0.18	-0.58
GO:0005198	structural molecule activity	-0.62	-0.55	0.18	-0.52

Table S2.3 (Continued)

GO:0005730	nucleolus	-0.72	-0.62	-0.01	-0.3
GO:0005740	mitochondrial envelope	-0.64	-0.05	0.76	0.08
GO:0005743	mitochondrial inner membrane	-0.49	-0.09	0.86	0.05
GO:0005746	mitochondrial respiratory chain	-0.18	-0.1	0.95	0.24
GO:0005768	endosome	-0.93	0.33	0.17	-0.03
GO:0005770	late endosome	-0.9	0.39	0.12	-0.14
GO:0005777	peroxisome	-0.97	-0.09	0.23	-0.06
GO:0005778	peroxisomal membrane	-0.97	-0.16	0.14	-0.09
GO:0005794	Golgi apparatus	-0.69	0.63	0.26	0.23
GO:0005829	cytosol	-0.71	-0.37	0.04	-0.6
GO:0005840	ribosome	-0.64	-0.53	0.14	-0.54
GO:0005856	cytoskeleton	-0.72	0.28	0.6	0.21
GO:0006099	tricarboxylic acid cycle	-0.12	-0.11	0.98	-0.07
GO:0006101	citrate metabolic process	-0.12	-0.11	0.98	-0.07
GO:0006364	rRNA processing	-0.66	-0.58	0.22	-0.43
GO:0006407	rRNA export from nucleus	-0.54	-0.63	0.15	-0.54
GO:0006412	translation	-0.57	-0.56	0.2	-0.57
GO:0006486	protein glycosylation	-0.02	0.83	0.34	0.44
GO:0006487	protein N-linked glycosylation	-0.01	0.8	0.35	0.48
GO:0006490	oligosaccharide-lipid intermediate biosynthetic process	0.04	0.94	0.22	0.25
GO:0006605	protein targeting	-0.97	0.11	0.15	-0.17
GO:0006623	protein targeting to vacuole	-0.88	0.42	0.23	0.02
GO:0006625	protein targeting to peroxisome	-0.96	-0.18	0.17	-0.14
GO:0006629	lipid metabolic process	-0.58	0.6	0.48	0.28
GO:0006886	intracellular protein transport	-0.95	0.26	0.17	-0.09
GO:0007031	peroxisome organization	-0.96	-0.05	0.14	-0.24

Table S2.3 (Continued)

GO:0007034	vacuolar transport	-0.9	0.37	0.22	-0.02
GO:0007059	chromosome segregation	-0.75	-0.27	0.53	0.28
GO:0007062	sister chromatid cohesion	-0.63	-0.18	0.7	0.28
GO:0007064	mitotic sister chromatid cohesion	-0.6	-0.18	0.7	0.33
GO:0007126	meiotic nuclear division	-0.91	-0.27	0.24	0.18
GO:0007127	meiosis I	-0.92	-0.28	0.27	0.11
GO:0008104	protein localization	-0.94	0.25	0.22	-0.06
GO:0009055	electron carrier activity	-0.1	-0.13	0.94	0.29
GO:0012505	endomembrane system	-0.71	0.65	0.26	0.08
GO:0015031	protein transport	-0.93	0.33	0.18	-0.04
GO:0015934	large ribosomal subunit	-0.49	-0.65	0.13	-0.56
GO:0015935	small ribosomal subunit	-0.66	-0.48	0.21	-0.55
GO:0016072	rRNA metabolic process	-0.67	-0.59	0.26	-0.38
GO:0016192	vesicle-mediated transport	-0.82	0.48	0.26	0.17
GO:0016197	endosomal transport	-0.94	0.31	0.17	-0.02
GO:0016558	protein import into peroxisome matrix	-0.97	-0.17	0.14	-0.12
GO:0016758	transferase activity, transferring hexosyl groups	-0.19	0.73	0.45	0.47
GO:0017183	peptidyl-diphthamide biosynthetic process from peptidyl-histidine	-0.27	0.04	-0.7	-0.66
GO:0019843	rRNA binding	-0.63	-0.48	0.22	-0.57
GO:0022618	ribonucleoprotein complex assembly	-0.64	-0.59	-0.08	-0.49
GO:0022625	cytosolic large ribosomal subunit	-0.44	-0.64	0.06	-0.63
GO:0022626	cytosolic ribosome	-0.56	-0.57	0.1	-0.6
GO:0022627	cytosolic small ribosomal subunit	-0.64	-0.47	0.2	-0.58
GO:0030490	maturation of SSU-rRNA	-0.7	-0.53	0.2	-0.44

Table S2.3 (Continued)

GO:0030529	intracellular ribonucleoprotein complex	-0.71	-0.52	0.1	-0.46
GO:0030684	preribosome	-0.67	-0.55	0.1	-0.49
GO:0030687	preribosome, large subunit precursor	-0.52	-0.66	0.01	-0.54
GO:0032266	phosphatidylinositol-3-phosphate binding	-0.94	0.11	0.31	0.03
GO:0033365	protein localization to organelle	-0.96	0.19	0.2	-0.09
GO:0034613	cellular protein localization	-0.95	0.22	0.23	-0.07
GO:0034728	nucleosome organization	-0.39	0.29	-0.45	-0.75
GO:0042175	nuclear outer membrane-endoplasmic reticulum membrane network	-0.37	0.89	0.19	0.2
GO:0042254	ribosome biogenesis	-0.63	-0.59	0.14	-0.48
GO:0042255	ribosome assembly	-0.52	-0.6	0.06	-0.6
GO:0042273	ribosomal large subunit biogenesis	-0.42	-0.66	-0.03	-0.63
GO:0042274	ribosomal small subunit biogenesis	-0.72	-0.51	0.23	-0.4
GO:0043043	peptide biosynthetic process	-0.57	-0.55	0.2	-0.57
GO:0043044	ATP-dependent chromatin remodeling	-0.34	0.23	-0.61	-0.68
GO:0043486	histone exchange	-0.42	0.19	-0.59	-0.67
GO:0043603	cellular amide metabolic process	-0.61	-0.52	0.22	-0.56
GO:0044255	cellular lipid metabolic process	-0.57	0.61	0.49	0.26
GO:0044743	intracellular protein transmembrane import	-0.93	-0.2	0.16	-0.26
GO:0045184	establishment of protein localization	-0.92	0.33	0.21	-0.04
GO:0048193	Golgi vesicle transport	-0.7	0.64	0.18	0.26
GO:0070469	respiratory chain	-0.16	-0.08	0.96	0.21

Table S2.3 (Continued)

GO:0072665	protein localization to vacuole	-0.89	0.4	0.23	-0.01
GO:0097502	mannosylation	0.01	0.89	0.32	0.33
GO:0098588	bounding membrane of organelle	-0.76	0.56	0.31	0.12
GO:0000028	ribosomal small subunit assembly	-0.62	-0.55	0.23	-0.51
GO:0000329	fungal-type vacuole membrane	-0.94	0.12	0.3	-0.09
GO:0005788	endoplasmic reticulum lumen	-0.01	0.96	0.16	0.24
GO:0005789	endoplasmic reticulum membrane	-0.34	0.89	0.19	0.22
GO:0006892	post-Golgi vesicle-mediated transport	-0.85	0.46	0.23	0.08
GO:0007067	mitotic nuclear division	-0.67	-0.22	0.63	0.31
GO:0016050	vesicle organization	-0.94	0.28	0.17	0.08
GO:0031090	organelle membrane	-0.78	0.37	0.48	0.12
GO:0031491	nucleosome binding	-0.28	0.22	-0.51	-0.78
GO:0032527	protein exit from endoplasmic reticulum	-0.1	0.95	0.16	0.23
GO:0045324	late endosome to vacuole transport	-0.92	0.25	0.29	0
GO:1990415	Pex17p-Pex14p docking complex	-0.95	-0.25	0.16	0.14
GO:1990429	peroxisomal importomer complex	-0.95	-0.25	0.16	0.14
GO:0000001	mitochondrion inheritance	-0.88	0	0.48	-0.03
GO:0000460	maturation of 5.8S rRNA	-0.67	-0.54	0.2	-0.46
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	-0.67	-0.54	0.2	-0.46
GO:0000469	cleavage involved in rRNA processing	-0.67	-0.54	0.2	-0.46

Table S2.3 (Continued)

GO:0006644	phospholipid metabolic process	-0.56	0.56	0.55	0.26
GO:0006650	glycerophospholipid metabolic process	-0.5	0.66	0.49	0.26
GO:0007005	mitochondrion organization	-0.58	-0.12	0.8	0.06
GO:0007131	reciprocal meiotic recombination	-0.92	-0.12	0.37	0.06
GO:0008610	lipid biosynthetic process	-0.61	0.43	0.6	0.29
GO:0031966	mitochondrial membrane	-0.67	-0.03	0.75	0.01
GO:0000725	recombinational repair	-0.89	-0.33	0.32	-0.07
GO:0005783	endoplasmic reticulum	-0.43	0.85	0.29	0.08
GO:0006334	nucleosome assembly	-0.46	0.43	-0.27	-0.73
GO:0006396	RNA processing	-0.72	-0.5	0.05	-0.47
GO:0006810	transport	-0.94	0.17	0.29	0.01
GO:0043248	proteasome assembly	-0.33	-0.35	-0.62	-0.61
GO:0046488	phosphatidylinositol metabolic process	-0.44	0.79	0.23	0.36
GO:0090305	nucleic acid phosphodiester bond hydrolysis	-0.8	-0.42	0.14	-0.41
GO:0090502	RNA phosphodiester bond hydrolysis, endonucleolytic	-0.58	-0.42	0.14	-0.68
GO:0000291	nuclear-transcribed mRNA catabolic process, exonucleolytic	-0.45	-0.8	0.02	-0.4
GO:0032008	positive regulation of TOR signaling	-0.74	0.11	0.39	-0.54
GO:0034427	nuclear-transcribed mRNA catabolic process, exonucleolytic, 3'-5'	-0.45	-0.8	0.02	-0.4

Table S2.3 (Continued)

GO:0043928	exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay	-0.45	-0.8	0.02	-0.4
GO:0070478	nuclear-transcribed mRNA catabolic process, 3'-5' exonucleolytic nonsense-mediated decay	-0.45	-0.8	0.02	-0.4
GO:0005935	cellular bud neck	-0.77	0.28	0.41	0.39
GO:0006643	membrane lipid metabolic process	-0.48	0.69	0.38	0.39
GO:0006896	Golgi to vacuole transport	-0.89	0.27	0.36	0.07
GO:0055085	transmembrane transport	-0.9	-0.02	0.44	0
GO:0030686	90S preribosome	-0.85	-0.37	0.19	-0.31
GO:0034975	protein folding in endoplasmic reticulum	-0.13	0.75	0.13	0.63
GO:0006281	DNA repair	-0.91	-0.32	0.14	-0.24
GO:0006302	double-strand break repair	-0.93	-0.3	0.18	-0.09
GO:0016757	transferase activity, transferring glycosyl groups	-0.28	0.66	0.44	0.54
GO:0044271	cellular nitrogen compound biosynthetic process	-0.77	-0.38	0.11	-0.5
GO:0000724	double-strand break repair via homologous recombination	-0.88	-0.35	0.31	-0.09
GO:0001104	RNA polymerase II transcription cofactor activity	-0.76	-0.27	-0.15	-0.58
GO:0010008	endosome membrane	-0.93	0.3	0.19	-0.11
GO:0046907	intracellular transport	-0.95	0.13	0.25	-0.11

Table S2.3 (Continued)

GO:0043628	ncRNA 3'-end processing	-0.69	-0.48	0.23	-0.49
GO:0005933	cellular bud	-0.79	0.23	0.37	0.42
GO:0006310	DNA recombination	-0.9	-0.28	0.28	-0.16
GO:0006448	regulation of translational elongation	-0.75	-0.4	-0.07	-0.52
GO:0006913	nucleocytoplasmic transport	-0.67	-0.61	0.13	-0.4
GO:0006807	nitrogen compound metabolic process	-0.83	-0.31	0.22	-0.4
GO:0015629	actin cytoskeleton	-0.68	0.44	0.56	0.14
GO:0000139	Golgi membrane	-0.64	0.63	0.25	0.36
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	-0.52	-0.69	0.17	-0.48
GO:0000278	mitotic cell cycle	-0.87	-0.01	0.49	0.07
GO:0035091	phosphatidylinositol binding	-0.93	0.2	0.26	0.16
GO:0005934	cellular bud tip	-0.82	0.26	0.15	0.49
GO:0042147	retrograde transport, endosome to Golgi	-0.91	0.34	0.12	0.2
GO:0071555	cell wall organization	-0.62	0.51	0.46	0.38
GO:0005938	cell cortex	-0.68	0.44	0.47	0.35

Table S2.4. Quantitative screens that are analyzed for gene effect size distribution

We manually scanned over 200 published deletion library screens to identify datasets that could be reanalyzed to potentially determine an effect size distribution. Of these 200 papers, we found only 6 that contained datasets in a form that was suitable for our reanalysis.

First author, publication year	Trait	# Genes	Source of published data	Reference
Breslow, 2008	Growth rate	4204	Supplementary table S5	Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5,** 711–718 (2008).
Schluter, 2008	Endosomal protein sorting	4814	Supplementary table 1	Schluter, C. et al. Global analysis of yeast endosomal transport identifies the vps55/68 sorting complex. Mol. Biol. Cell 19, 1282–1294 (2008).
Vizeacoumar, 2010	Spindle morphogenesis	4286	Supplementary table S6	Vizeacoumar, F. J. et al. Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J. Cell Biol. 188, 69–81 (2010).
Cooper, 2010	Amino acid level	4382	Supplementary table 4	Cooper, S. J. et al. High-throughput profiling of amino acids in strains of the Saccharomyces cerevisiae deletion collection. Genome Res. 20, 1288–1296 (2010).
Jonikas, 2009	Unfolded protein response	4563	Supplementary table 1	Jonikas, M. C. et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science 323, 1693–1697 (2009).
Hillenmeyer, 2008	Chemical genomic profile	5337	http://chemogenomics.stanford.edu/supplements/global/download.html	Hillenmeyer, M. E. et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320, 362–365 (2008).

Supplemental Text

Re-analysis of previous quantitative screening using yeast deletion collections

Since the release of the yeast deletion collection, a large number of studies have been performed [7] potentially providing a rich source to understand the quantitative effects of gene deletions on traits. Unfortunately, the raw data was not published and readily available for all but a small handful of these studies (**Table S2.4**).

Data from each screen in the **Table S2.4** was analyzed using the following method: 1) download raw data; 2) determine the measurement error; 3) calculate p value for each gene by comparing effect size measurement to measurement error (two-tailed t-test, assuming measurement error is Gaussian distributed); 4) correct the p values for multiple hypothesis tests by calculating false discovery rate; 5) identify the number of significant genes as ones with FDR<0.5%.

The measurement error for individual assays were determined as below. We assume that the true effects of deleting the i^{th} gene is x^i. The two independent measurements, $x^{i,j} = x^i + \epsilon^{i,j}$ for $j = 1,2$, where Ɛ is the measurement noise term. Assuming that measurement noise follows a Gaussian distribution, i.e. $\epsilon^{i,j} \sim N(0, \sigma)$. The difference of the two measurements on the identical strain will reveal information about the standard deviation of measurement noise. Specifically, since

$$x^{i,1} - x^{i,2} \sim N(0, \sqrt{2}\sigma)$$

we can derive the following estimate of the standard deviation of measurement noise:

$$\hat{\sigma} = \frac{\sum\left(x^{i,1} - x^{i,2}\right)^2}{2 * (N_{gene} - 1)}$$

This method was applied to the raw data from the six assays in **Table S2.4**. Breslow et al. had different number of replicates [5], and hence the measurement error for the individual mutants varied depending on the number of replicates. Specifically, among 4204 assayed genes, 2809 genes have one measurement, 874 genes have two replicates and 521 genes have at least three replicates. To avoid this complication, we only used the data from the first measurements and used the remaining data to estimate the measurement error. We first estimated the measurement error by applying the equation above to the replicate measurements of 874 strains with two measurements and determined measurement error as 0.015. Then we calculate the measurement error the 521 strains for which three measurements had been made. This yielded a measurement error of 0.017. As these two estimations are close, we use the average (0.016) as the measurement error for the assay. We observed that the measurement noise tends to be larger for strains with large effect size, which means that most strains with moderate effect sizes probably have smaller than estimated measurement error. Hence, we do not believe that this method will overestimate the number of genes affecting the growth rate trait.

Similarly, mutants in Jonikas et al. had different numbers of replicates [6]. Measurement noise decreased as the number of replicate increased. As a conservative estimate of effect size measurements, we treated all measurements as if they had only two replicate data. To estimate the measurement error, we used the data from 541 strains with exactly two replicate data. In the original paper, the standard deviation of measurements for each strain was reported. Since there

were only two measurements for these strains, the standard deviation equals the half of the difference between two measurements. Assuming that measurement noise of each replicate data followed $N\ (0, \sigma)$, the expectation of the half of the difference of two independent measurements is $\sigma/\sqrt{\pi}$. When plotting the histogram of this data, we found that a number of measurement have exceptionally large measurement error, which artificially increased our estimation. After removing strains with measurement error larger than 0.5, the resulting measurement standard deviation has an average as 0.0678. Hence, we estimated the measurement noise as $\frac{\sigma}{\sqrt{2}} =$

$0.0678 * \sqrt{\frac{\pi}{2}} = 0.085.$

Mutants in Schluter et al. were assayed in replicates for both haploid and diploid strains[8]. We applied the equation above to this data and determined that the measurement error was 0.027 for the MATa, haploids, 0.021 for the MATalpha haploids, and 0.042 for diploids. We used the average of these three to estimate measurement error (0.030). Vizeacoumar et al. provided p values for each mutants in the assay [9]. We convert the p value back to a *z*-score using Matlab function *norminv()*. While Copper et al. published raw data, they did so for only one replicate and hence we did not proceed with further analysis on this data set [10].

Furthermore, we analyzed the raw data from Hillenmeyer by comparing the measured effect sizes in independent experiments using the same condition (drug name, dosage and the duration to apply the drug) in separate batches [11]. We found a large variation of the reproducibility between these replicates, determined as the pair-wise Pearson correlation coefficient (ranging

from -0.2 to 0.99 with a median of 0.36 depending on the condition used). Hence we did not analyze the data further more.

To evaluate the number of gene deletions that significantly affected each of the quantitative trait, we first considered a null model where all gene deletions had no effects on the assayed traits. We expected the measured effect sizes to follow a normal distribution determined by measurement noise, i.e. $\sim N(0, \text{measurement noise})$. However, we found this was not the case for all the traits that we analyzed. To better illustrate this, we re-scaled the effect sizes by measurement error for each trait, and plotted the histograms of the re-scaled effect sizes for the gene deletions that have effect sizes at least 3-fold of the estimated measurement noise in Figure S2.2. We found the distributions were continuous. Note that only about (1-99.7%)*5000=15 genes were expected from the noise distribution. This suggests that the measured effects of most of the plotted genes were not from the measurement noise. Note that the data from Vizeacoumar was not shown here as the majority of genes have effects that are within three-fold of measurement noise.

To identify assays that are sensitive enough the measure the effect sizes of as many genes as possible. We estimated the number of genes that significantly affect each of the analyzed traits by comparing the measured effect sizes to measurement noise. Using a cutoff of FDR < 0.5%, we determined that two screens by Jonikas et al. and Breslow et al. are suitable for effect size distribution analysis as they have smallest measurement errors.

Flow cytometry data processing

Raw data was exported from an LSRII or Stratedigm in fcs3.0 file format. All data was loaded

using customized *MATLAB* code. In briefly, data from each sample was first filtered on FSC/SSC channel to remove cell debris, and on SSC channel to normalize for cell size. The FSC/SSC gates were drawn manually on pooled samples. The SSC gate was determined to include events between the 25^{th} to 75^{th} percentiles of the pooled sample. Pooled samples were also used to find thresholds on YFP and mCherry channels to segment induced vs. uninduced cells, and reference vs. mutant cells (Figure S2.6 as an example). Mutants were filtered to ensure that there are at least 700 events for both reference and mutant cells in at least one biological replicates. Mutants in twelve plates in replicate one of the GAL screen have higher induced fraction than the reference strain in the same sample. Data from the second replicate were used for these mutants in the future analysis. For the PHO screen, we calculated the standard deviation of the effect size differences between two replicates for each of the three traits. The effect size measurements for fourteen mutants are greater than five-fold of these standard deviations. These strains were filtered from future analysis.

Principal component analysis on reporter expression distribution of the entire deletion collection

Yeast responds to a mixture of glucose and galactose in a bimodal way. We measured expression level of GAL1pr-YFP in single cells for each of the mutant strain in the deletion collection. We generally observed that the reproducibility was higher when normalizing the distribution by comparing the mutant distribution to the reference distribution in the same well (see the section **Data Normalization** for details); as opposed to analyzing the mutant data directly. This is presumably due to slight variation between wells, plates, and days.

To find appropriate metric by which to analyze the mutant strains, we performed PCA analysis. We did this by pooling reference and mutant YFP distribution from two replicates. After data segmentation, the GAL1pr-YFP distributions of both reference strain and mutant strain were binned into 92 equally size log2 bins ranging from the maximum to minimum value. Data was normalized to probability distribution, separately for reference strain and mutant strain in each sample. PCA results were shown in Figure S2.1. The first three principle components explain ~ 60% variation. By manually examining the shape of each principle component, we could provide a plausible biological explanation for the major components. The first vector affects the induced fraction without affecting the expression level. The second vector has two effects, shifting the expression level of induced cells as well as changing the fraction of induced cells. The third vector change the expression level of both uninduced cells (basal level) and induced cells. In further analysis, we found that the expression level of uninduced cells could not be accurately determined for the majority of strains in our assay for GAL1pr-YFP reporter, and hence only the induced fraction and the induction level are used in the main text. This third metric was used for analysis of the PHO response.

Data normalization

The induced fraction and induction level traits were calculated for each mutant strain using the following method. First, the induced fraction and induction level were calculated for reference strains and query strain in each sample. The induced fraction was calculated as the ratio of the number of induced events over the number of all events. The induction level was calculated as the average level of YFP of the induced cells. For both traits, the mutant value was regressed

against the reference value using the *Matlab* function *robustfit()*. The residual of each measurement from the fit was averaged between two replicates to determine the final values of the induced fraction and the induction level.

Estimate the number of genes that affect yeast quantitative traits

The noise distribution determined from measurement noise estimation was overlaid with the actual effect size measurements. Both curves were normalized to the total number of genes. The area of the region where the actual effect size distribution was outside the measurement noise distribution was determined for estimating the number of genes that affected each of the four yeast traits (Figure S2.2).

Compare the number of detected mutants by using induced fraction and induction level vs. average expression level

Our screening data on the yeast galactose response provided a test for estimating the total number of significant mutants using different metrics. This is interesting as many biological traits could usually be defined in different ways, yet it was unclear to our knowledge how much potentially subtle differences in metric could influence genes identified. Here when we are referring to different metrics it is probably easiest to think of them as different sub measurements. For example, if one measured standing height as opposed to sitting height, would one uncover different sets of genes. In our case, the effect of gene deletion on galactose response can be represented as the two GAL traits as used in the main text, or alternatively we could simply use the average YFP level as used in Jonikas et al [6]. To estimate such effects, we re-

analyzed our data by quantifying not just the two GAL traits, but also the average YFP level. After applying the same method to detect mutants that significantly affect yeast GAL response, we found that the two-traits method detected more mutants (1104) than the average YFP method (593). In addition, the one-trait method could not reveal the distinct modes by which different mutants worked; i.e. 50% reduction in average can come because 50% of cells don't induce or 100% of cells are 50% less induced. Hence our data suggested that, biological meaningful decomposition of a complex trait will increase detection sensitivity, and provides new biology insights to understand traits.

Genes that saturated our assay

Our GAL assay was designed to detect genes of small effect size, and as a result, ten genes of larger effect size saturated our assay. These genes were manually verified by inspecting the YFP distribution of the raw data. These genes are: *GAL4 (YPL248C), GCN4 (YEL009C), GAL80 (YML051W), GAL1 (YBR020W), SNF3 (YDL194W), STI1 (YOR027W), REG1 (YDR028C), GAL3 (YDR009W), SNF2 (YOR290C), HSC82 (YMR186W)*. This is important when calculating the explained heritability for top N genes (see main text). One of our main arguments is that the number of genes that affect a quantitative trait is around 8% of the genome. If the true effects of these ten genes is much larger than what we estimated, the number of genes that affect a quantitative trait could be smaller.

When using the nominal values of the measurements as effect sizes of these genes, we determined that the total contribution of these genes are 25.2% and 7.5% for induction level and

induced fraction respectively. As another way to estimate the effect sizes of these genes, we randomly sampled the effect size distribution. The average contribution of these genes is 27.8% and 10.5% respectively, suggesting that this alternate method does not strongly affect conclusion.

Overlapping among genes that are significant for each of the four studied traits

We examined the overlap between significant genes that affect growth rate and ones that affect each of the three other non-growth traits. To do so, genes with missing data in one of the data sets were removed. The result is in **Table S2.2**. The p-value was calculated between each pair of growth rate and non-growth rate trait, using a hypergeometric test (one-tailed).

Compare the effects on GAL and PHO response by deleting genes involved in protein synthesis

For 95 genes involved in protein synthesis, we compared their effects on GAL and PHO traits in the main text and Figure S2.5 using t-test (two-tailed). The average difference between the effects on GAL and PHO is 0.15. The standard deviations of effects on GAL and PHO are 0.21 and 0.10. As an alternative method to test for significance, we pooled the measured effects on GAL and PHO response and randomly split the pooled data into two groups for 1,000,000 times and calculated the difference between two groups. The observed difference (0.15) is not observed in the randomized sample. Hence we determined that $p < 10^{-6}$ using this method.

Canonical genes involved in galactose signaling and unfolded protein response

Glu/Gal gene list: GPB2, IRA1, TOS1, GLK1, GPA2, GAL83, SAK1, GLC7, YCK1, BCY1,

RGT1, ELM1, TPK3, HXK1, GPR1, RGT2, SNF3, REG1, MTH1, MSN5, SIP1, SNF1, MIG1, SNF4, SIP2, PDE1, HXK2, CYR1, TPK1, GRR1, SDC25, CDC25, SIP5, RAS2, YCK2, IRA2, STD1, RAS1, RGS2, PDE2, GPB1, TPK2, GAL1, GAL3, GAL80, GAL4, SNF2, GCN4, HSC82, STI1

Gene localized in ER, Golgi, and early Golgi are (298 genes): YEL031W, YJR117W, YFL025C, YJL062W, YML012W, YAL023C, YJR118C, YML055W, YML013W, YOR002W, YGL084C, YCR044C, YER122C, YNL219C, YNR030W, YDL095W, YML115C, YGL020C, YGL054C, YIL039W, YEL036C, YPL227C, YOL013C, YMR022W, YMR161W, YKL212W, YDL192W, YLR110C, YGL167C, YMR264W, YAL058W, YER083C, YDR027C, YLR372W, YCR094W, YLR268W, YNL238W, YMR307W, YJL029C, YBR171W, YDL100C, YGL226C-A, YBR106W, YJR073C, YNL322C, YGR229C, YGR284C, YJR010C-A, YML128C, YFR041C, YNL323W, YEL042W, YMR123W, YBR015C, YJR075W, YBR162W-A, YCR067C, YJL004C, YCR017C, YAL026C, YOR216C, YIL090W, YAL007C, YNL041C, YJL123C, YIL040W, YBR164C, YCL045C, YNL051W, YIR004W, YPL050C, YPL051W, YGL126W, YCR034W, YMR292W, YDR233C, YNL297C, YGL005C, YDR245W, YBR036C, YDR221W, YPL192C, YLL014W, YDR508C, YEL001C, YER005W, YDR137W, YDL099W, YGL231C, YHR108W, YMR238W, YAL053W, YIL027C, YER072W, YML038C, YER120W, YEL027W, YIL030C, YDR492W, YJR131W, YMR010W, YHR181W, YPR063C, YIL124W, YLR350W, YJR088C, YBL011W, YML048W, YNL044W, YDR358W, YOR311C, YDR411C, YMR272C, YNL049C, YMR015C, YDL052C, YJR134C, YKL096W, YNL280C, YLR194C, YER113C, YDR077W, YDR055W, YNR021W, YNL327W, YLR130C, YNR039C, YJL099W,

YKL146W, YPR003C, YHL017W, YOR245C, YER166W, YBR132C, YOR016C, YPR090W, YNL300W, YLR250W, YGR038W, YPL259C, YPR071W, YKL065C, YKL046C, YPL274W, YEL048C, YOR317W, YDR100W, YNL146W, YMR253C, YJR031C, YER011W, YJL078C, YIL016W, YML037C, YGR247W, YFL004W, YBR023C, YIL044C, YMR052W, YDL204W, YBR067C, YDR153C, YIL043C, YNL095C, YDR476C, YOR307C, YOR321W, YCR011C, YMR237W, YMR071C, YER004W, YPR028W, YGL255W, YPL170W, YKL063C, YJL044C, YLR023C, YMR215W, YMR251W-A, YGR261C, YPR091C, YDR056C, YLL028W, YLR330W, YBL010C, YNR019W, YGL124C, YDR294C, YNL046W, YDR519W, YKR088C, YLR042C, YKL094W, YCR048W, YCR043C, YDR084C, YKR067W, YJL196C, YLL061W, YML101C, YDL232W, YOL030W, YMR054W, YDR410C, YBR273C, YLR120C, YHR110W, YOR044W, YDL137W, YJL171C, YOR285W, YMR029C, YLR064W, YPL137C, YOR092W, YBR159W, YGL083W, YNL156C, YDL128W, YBR296C, YOR175C, YJL198W, YOL101C, YHL019C, YJL117W, YGR263C, YML059C, YOR214C, YNR013C, YOR087W, YJL192C, YGR177C, YBL102W, YPL195W, YLL052C, YLR390W-A, YDR264C, YOR299W, YMR152W, YLL055W, YDR424C, YBR287W, YEL040W, YNL125C, YHL003C, YBR283C, YDL121C, YHR045W, YNR075W, YOR377W, YHR039C, YGL010W, YCL025C, YNR044W, YLR050C, YOL137W, YOL107W, YDL018C, YDR307W, YDR297W, YNL190W, YDR503C, YBR177C, YGR266W, YER019C-A, YLR034C, YOR322C, YGR260W, YDR349C, YJR015W, YPL246C, YMR058W, YBR290W, YLL023C, YDR205W, YHR123W, YJL024C, YJL212C, YLR292C, YPL207W, YKR027W, YIL076W, YBR288C, YJL183W, YKL008C, YJL207C, YML067C, YGR089W, YOR291W, YNL111C, YEL043W,

YPL234C, YLR056W, YKL096W-A, YGR157W, YHR060W, YLR039C, YHR079C

Reference

1. Tong AHY, Boone C. Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol Biol. 2006;313: 171–192.

2. Wykoff DD, Rizvi AH, Raser JM, Margolin B, O'Shea EK. Positive feedback regulates switching of phosphate transporters in S. cerevisiae. MOLCEL. 2007;27: 1005–1013. doi:10.1016/j.molcel.2007.07.022

3. Pertea M, Salzberg SL. Between a chicken and a grape: estimating the number of human genes. Genome Biology. 2010;11: 206. doi:10.1186/gb-2010-11-5-206

4. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics. 2004;20: 3710–3715. doi:10.1093/bioinformatics/bth456

5. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nature Methods. 2008;5: 711–718. doi:10.1038/nmeth.1234

6. Jonikas MC, Collins SR, Denic V, Oh E, Quan EM, Schmid V, et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science. 2009;323: 1693–1697. doi:10.1126/science.1167983

7. Giaever G, Nislow C. The yeast deletion collection: a decade of functional genomics. Genetics. 2014;197: 451–465.

8. Schluter C, Lam KKY, Brumm J, Wu BW, Saunders M, Stevens TH, et al. Global analysis of yeast endosomal transport identifies the vps55/68 sorting complex. Mol Biol Cell. 2008;19: 1282–1294.

9. Vizeacoumar FJ, van Dyk N, Vizeacoumar FS, Cheung V, Li J, Sydorskyy Y, et al. Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. The Journal of Cell Biology. Rockefeller University Press; 2010;188: 69–81. doi:10.1083/jcb.200909013

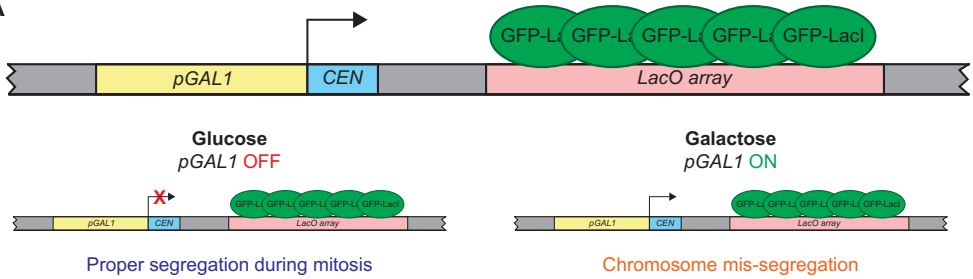
10. Cooper SJ, Finney GL, Brown SL, Nelson SK, Hesselberth J, MacCoss MJ, et al. High-throughput profiling of amino acids in strains of the Saccharomyces cerevisiae deletion

collection. Genome Research. 2010;20: 1288–1296. doi:10.1101/gr.105825.110

11. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science. 2008;320: 362–365. doi:10.1126/science.1150021

Appendix B: Supplementary material for Chapter 3

A



B **Colony size measurements and mis-segregation rates for inducible aneuploidies**

Inducible Chromosome(s)	ORFs	Percent Mis-segregation*	Disome / Euploid Colony Area			Trisome / Euploid Colony Area			Monosome / Euploid Colony Area		
			Mean	SD**	n	Mean	SD**	n	Mean	SD**	n
1	117	86	0.837	0.086	14	1.012	0.029	16	0.793	0.020	16
2	456	76	0.844	0.045	22	0.997	0.052	25	0.000	0.000	25
3	183	73	0.941	0.063	15	1.112	0.047	20	0.409	0.035	20
4	836	67	0.060	0.036	15	0.465	0.102	26	0.001	0.002	26
5	323	73	0.483	0.036	18	0.908	0.042	27	0.174	0.016	27
6	141	70	0.000	0.000	13	0.016	0.025	19	0.629	0.033	19
7	583	70	0.410	0.085	10	0.885	0.152	30	0.000	0.000	30
8	321	69	0.795	0.048	13	1.054	0.036	25	0.081	0.020	25
9	241	83	0.886	0.041	22	1.059	0.028	31	0.305	0.023	31
10	398	86	0.669	0.047	19	0.899	0.035	22	0.000	0.000	22
11	348	77	0.770	0.040	23	0.936	0.048	29	0.019	0.004	29
12	578	93	0.111	0.041	21	0.800	0.056	27	0.001	0.005	27
13	505	80	0.466	0.050	22	0.962	0.046	28	0.000	0.000	28
14	435	70	0.392	0.042	18	0.883	0.029	24	0.001	0.001	24
16	511	82	0.540	0.032	24	0.913	0.059	26	0.000	0.000	26
1+2	573	NA	NA	NA	NA	0.880	0.095	12	0.000	0.000	12
5+10	721	NA	NA	NA	NA	0.744	0.058	8	0.010	0.003	8
2+5+10	1177	NA	NA	NA	NA	0.656	0.076	18	0.000	0.000	18
1+5+10	838	NA	NA	NA	NA	0.762	0.043	9	0.012	0.011	9
1+2+10	971	NA	NA	NA	NA	0.759	0.114	14	0.000	0.000	14

* Percent mis-segregation calculated in haploid strains (disomes)
** SD = standard deviation
NA = not analyzed

C **Mis-segregation rates of chromosomes IV and V in inducible aneuploidies**

Inducible Chromosome	ORFs	Percent ChrIV Mis-segregation (3 hours)	Percent ChrIV Mis-segregation (5 hours)	Percent ChrV Mis-segregation (3 hours)	Percent ChrV Mis-segregation (5 hours)
1	117	0	0	1	0
2	456	0	0	0	0
8	321	0	0	0	0
9	241	0	0	0	1
11	348	0	0	0	0
13	505	0	0	0	0

Chr = chromosome

Figure S3.1 A conditional centromere system to induce chromosome mis-segregation.

(A) Schematic of the inducible aneuploidy system. The *GAL1* promoter was inserted adjacent to the *CEN* sequence on a particular chromosome. In medium containing glucose, the *GAL1* promoter is repressed and chromosome segregation occurs faithfully. In medium containing galactose, the *GAL1* promoter is activated, and transcription through the *CEN* sequence prevents kinetochore assembly, leading to chromosome mis-segregation. (B) Percent mis-segregation rates for each chromosome in a haploid cell and colony area measurements for disomes, trisomes and monosomes (normalized to euploid colony sizes). Data obtained for these measurements were obtained from the experiment shown in Figure 3.1F. Strains used in this assay: haploid wild-type (A2587), diploid wild-type (A16629); disomes I (A38370), II (A38372), III (A38374), IV (A38376), V (A38378), VI (A38380), VII (A38382), VIII (A38384), IX (A38386), X (A38388), XI (A38390), XII (A38392), XIII (A38394), XIV (A38396), and XVI (A38398); trisomes and monosomes I (A38401), II (A38402), III (A38403), IV (A38404), V (A38405), VI (A38406), VII (A38407), VIII (A38408), IX (A38409), X (A38410), XI (A38411), XII (A38412), XIII (A38413), XIV (A38414), and XVI (A38415). Trisomes and monosomes I+II (A38753), V+X (A38755), II+V+X (A38756), I+V+X (A38758), and I+II+X (A38759) were used in this assay with a different wild-type strain (A38751). (C) Mis-segregation rates (in percent) for chromosome IV and chromosome V following chromosome mis-segregation of chromosomes I, II, VIII, IX, XI, and XIII. Cells were grown to mid-log phase in SC-R and then transferred into SC-RG for 160 minutes to induce mis-segregation of the inducible chromosome. Cells were then transferred to SC-D, and mis-segregation frequencies of chromosomes IV or V

(not inducible chromosomes in these strains) were determined 3 and 5 hours after mis-segregation.

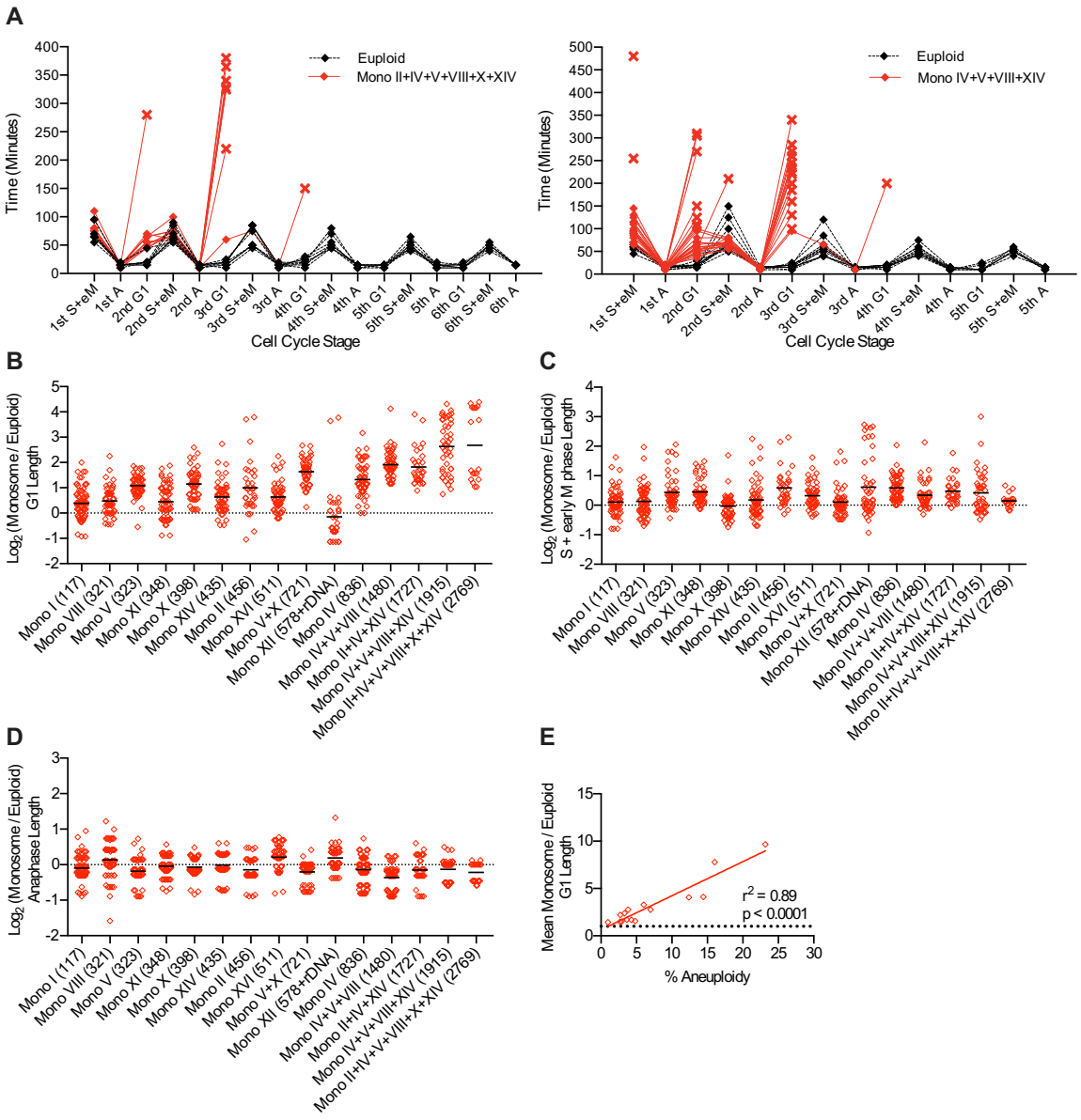


Figure S3.2 Chromosome loss increases duration and variability of multiple cell cycle

stages.

Cells were grown to mid-log phase in SC-R, then transferred to SC-RG for 160 minutes to induce chromosome mis-segregation and then plated on SC-D solid medium and imaged every 5 minutes for 8-10 hours at 25°C to monitor mCherry-Cdc3 to measure the time from bud emergence to cytokinesis, Spc42-dsRed or Spc42-GFP to monitor anaphase onset, and GFP dots to monitor chromosome mis-segregation. (A) Durations of cell cycle stages for single cells following chromosome missegregation are shown for euploid cells (black lines) and cells monosomic for chromosomes II+IV+V+VIII+X+XIV (red lines, left panel) and IV+V+VIII+XIV (red lines, right panel). Each line represents a single cell. An "X" indicates that the cell arrests for the remainder of the movie. (B-D) G1 duration (B), S+early M phase duration (C), and anaphase duration (D) were calculated for monosomes and normalized to euploid cells imaged during the same time-lapse. Log_2 transformed aneuploid to euploid ratios are plotted. Lines shown are at the mean. Numbers in parentheses on the x-axis labels indicate number of open reading frames on the aneuploid chromosome(s). rDNA copy number is estimated at 150 copies in the W303 strain background. Euploid cells were either from the wild-type control strain or from cells in the experimental strain that did not mis-segregate a chromosome. (E) G1 duration correlates with the degree of monosomy. Mean G1 lengths are plotted as a function of the degree of aneuploidy (calculated by number of open reading frames encoded by the mis-segregating chromosome(s); linear regression, $r^2 = 0.89$, $p<0.0001$).

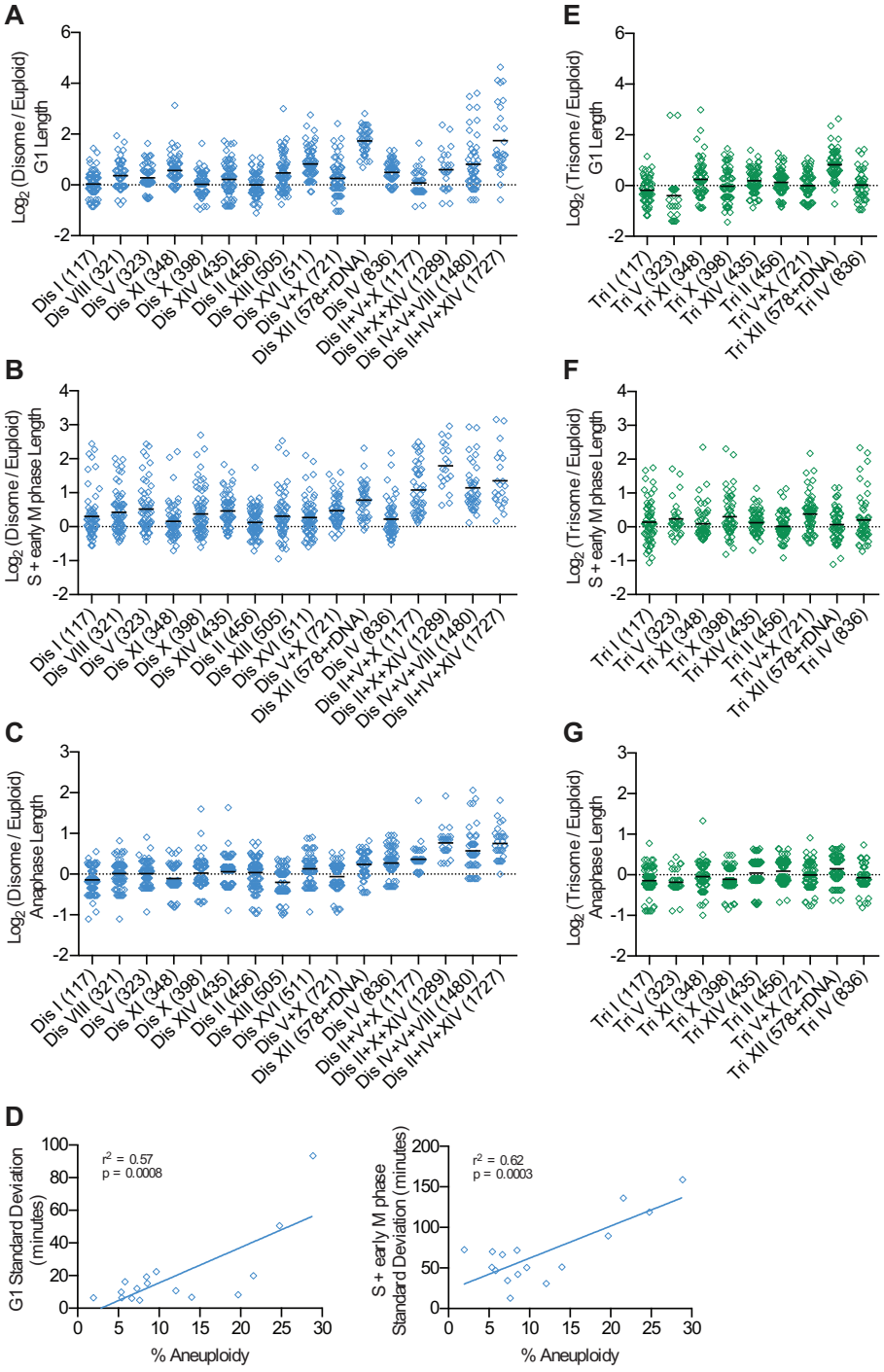

A
B
C
D
E
F
G
Log2 (Disome / Euploid) G1 Length
Log2 (Disome / Euploid) S + early M phase Length
Log2 (Disome / Euploid) Anaphase Length
Log2 (Trisome / Euploid) G1 Length
Log2 (Trisome / Euploid) S + early M phase Length
Log2 (Trisome / Euploid) Anaphase Length
Dis I (117)
Dis VIII (321)
Dis V (323)
Dis XI (348)
Dis X (398)
Dis XIV (435)
Dis II (456)
Dis XIII (505)
Dis XVI (511)
Dis V+X (721)
Dis XII (578+rDNA)
Dis IV (836)
Dis II+V+X (1177)
Dis II+X+XIV (1289)
Dis IV+V+VIII (1480)
Dis II+IV+XIV (1727)
Tri I (117)
Tri V (323)
Tri XI (348)
Tri X (398)
Tri XIV (435)
Tri II (456)
Tri V+X (721)
Tri XII (578+rDNA)
Tri IV (836)
G1 Standard Deviation (minutes)
r2 = 0.57
p = 0.0008
% Aneuploidy
S + early M phase Standard Deviation (minutes)
r2 = 0.62
p = 0.0003
% Aneuploidy

Figure S3.3 Chromosome gain increases duration and variability of multiple cell cycle stages.

Cells were grown to mid-log phase in SC-R and then transferred into SC-RG for 160 minutes to induce chromosome mis-segregation. Cells were then plated on SC-D solid medium mounted on a slide. Cells were imaged every 5 minutes for 8-10 hours at 25°C to monitor mCherry-Cdc3 to measure the time from bud emergence to cytokinesis, Spc42-dsRed or Spc42-GFP to monitor anaphase onset, and GFP dots to monitor chromosome mis-segregation. G1 duration (A, E), S+early M phase duration (B, F), and anaphase duration (C, G) were calculated for disomes and trisomes and normalized to euploid cells imaged during the same time-lapse. Log_2 transformed aneuploid to euploid ratios are plotted. Lines shown are at the mean. Numbers in parentheses on the x-axis labels indicate number of open reading frames on the aneuploid chromosome(s). rDNA copy number is estimated at 150 copies in the W303 strain background. Euploid cells were either from the wild-type control strain or from cells in the experimental strain that did not mis-segregate chromosomes. The graph in (D) shows that variabilities in G1 (left panel) and in S+early M phase duration (right panel) correlate with the degree of disomy. Mean G1 lengths are plotted as a function of the degree of aneuploidy (calculated by number of open reading frames encoded by the mis-segregating chromosome(s); linear regression, $r^2 = 0.57$, $p = 0.0008$; $r^2 = 0.62$, $p = 0.0003$). Abbreviations: Dis, disome; Tri, trisome.

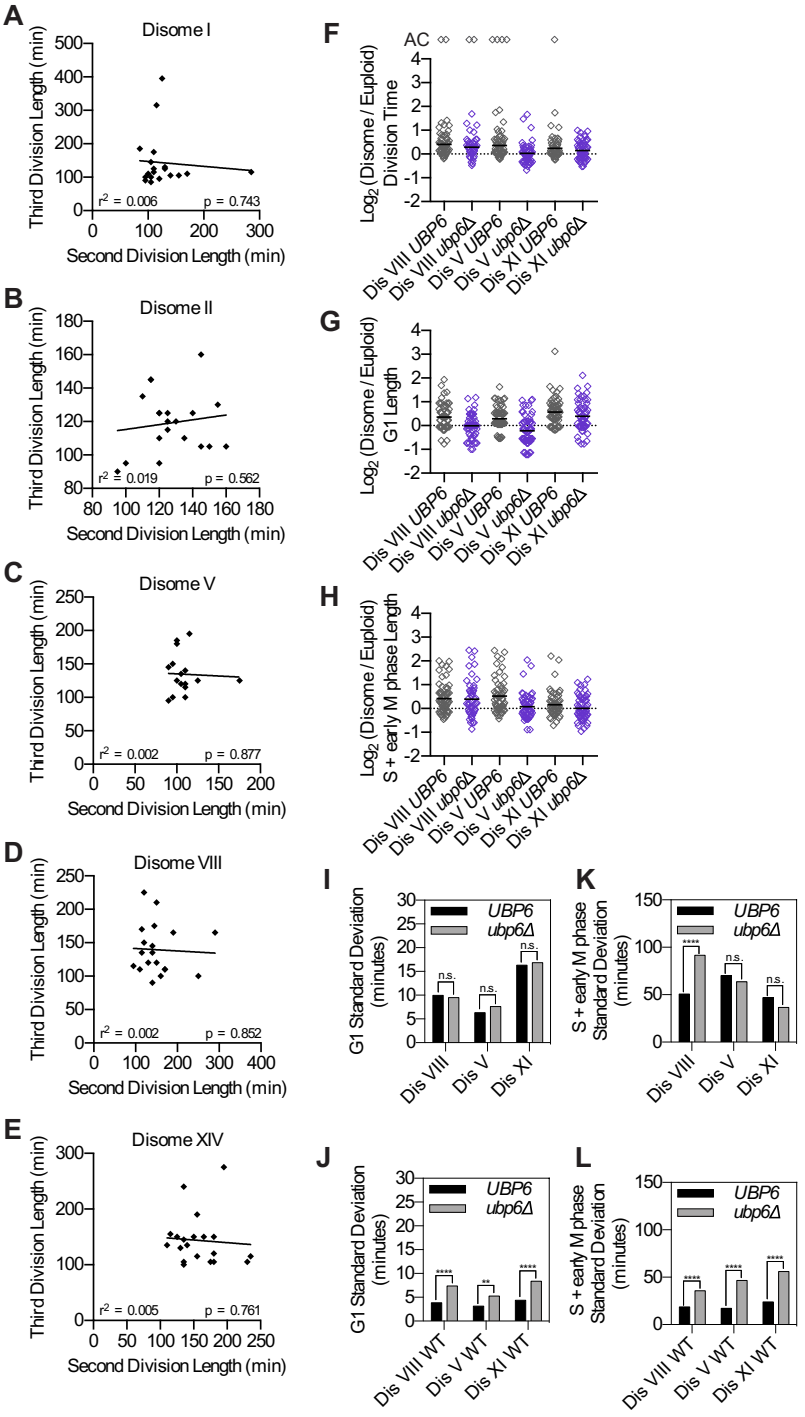


Figure S3.4 Cell division length variability is stochastic and not affected by deletion of

***UBP6*.**

(A-E) Cell cycle duration of the second and third cell divisions following chromosome missegregation from live-cell imaging analyses shown in Figure 3.3A. r^2 values shown are from linear regression analyses. Strains used for this analysis are: disomes I (A38660; A), II (A38662; B), V (A38666; C), VIII (A38668; D), and XIV (A38678; E). (F-L) Cells were grown to mid-log phase in SC-R and then transferred into SC-RG for 160 minutes to induce chromosome mis-segregation. Cells were then plated on SC-D solid medium mounted on a slide. Cells were imaged every 5 minutes for 8-10 hours at 25°C to monitor mCherry-Cdc3 to measure the time from bud emergence to cytokinesis, Spc42-dsRed or Spc42-GFP to monitor anaphase onset, and GFP dots to monitor chromosome mis-segregation. Division times (F), G1 durations (G), and S+early M phase durations (H) were calculated for disomes and normalized to euploid cells imaged during the same time-lapse. Log_2 transformed aneuploid to euploid ratios are plotted. Lines in F-H are at the mean. Standard deviations for euploid (WT; J,L) and disomic (I,K) populations were measured (in minutes), and an F-test was used to test for equality of variance between the *UBP6* (black bars, I-L) and *ubp6*Δ (grey bars, I-L) populations. Asterisks (*) in I-L indicate statistical significance between *UBP6* and *ubp6*Δ population variances as calculated by an F-test (** = $p \leq 0.01$, **** = $p \leq 0.0001$). The data for strains harboring a wild-type *UBP6* gene shown here are the same data as shown in Figure 3.3 and are duplicated here for comparison to *ubp6*Δ.

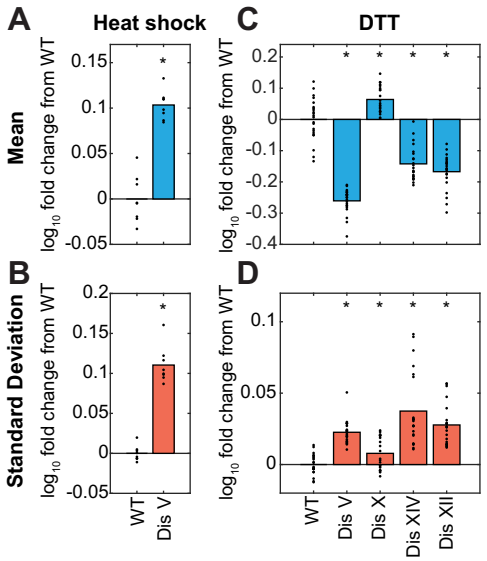


Figure S3.5 Disomes show increased cell-to-cell variability in response to DTT and heat shock.

(A-B) Mean (A) and standard deviation (B) of P_{4xHSE}-*YFP* expression in steady-state heat shock conditions (39ºC). Points indicate individual biological replicates (n=8), and the asterisks (*) represent statistical significance between the disome and euploid populations as calculated by Wilcoxon rank sum test ($p<0.05$). Strains used: wild-type (A38735) and disome V (A38737). (C-D) Mean (C) and standard deviation (D) of P_{4xUPRE}-*GFP* expression in steady-state oxidative stress conditions (growth in the presence of 0.625mM DTT for 4hrs). Points indicate individual biological replicates (n=21-22), and the asterisks (*) represent statistical significance between the disome and euploid populations as calculated by Wilcoxon rank sum test ($p<0.05$).

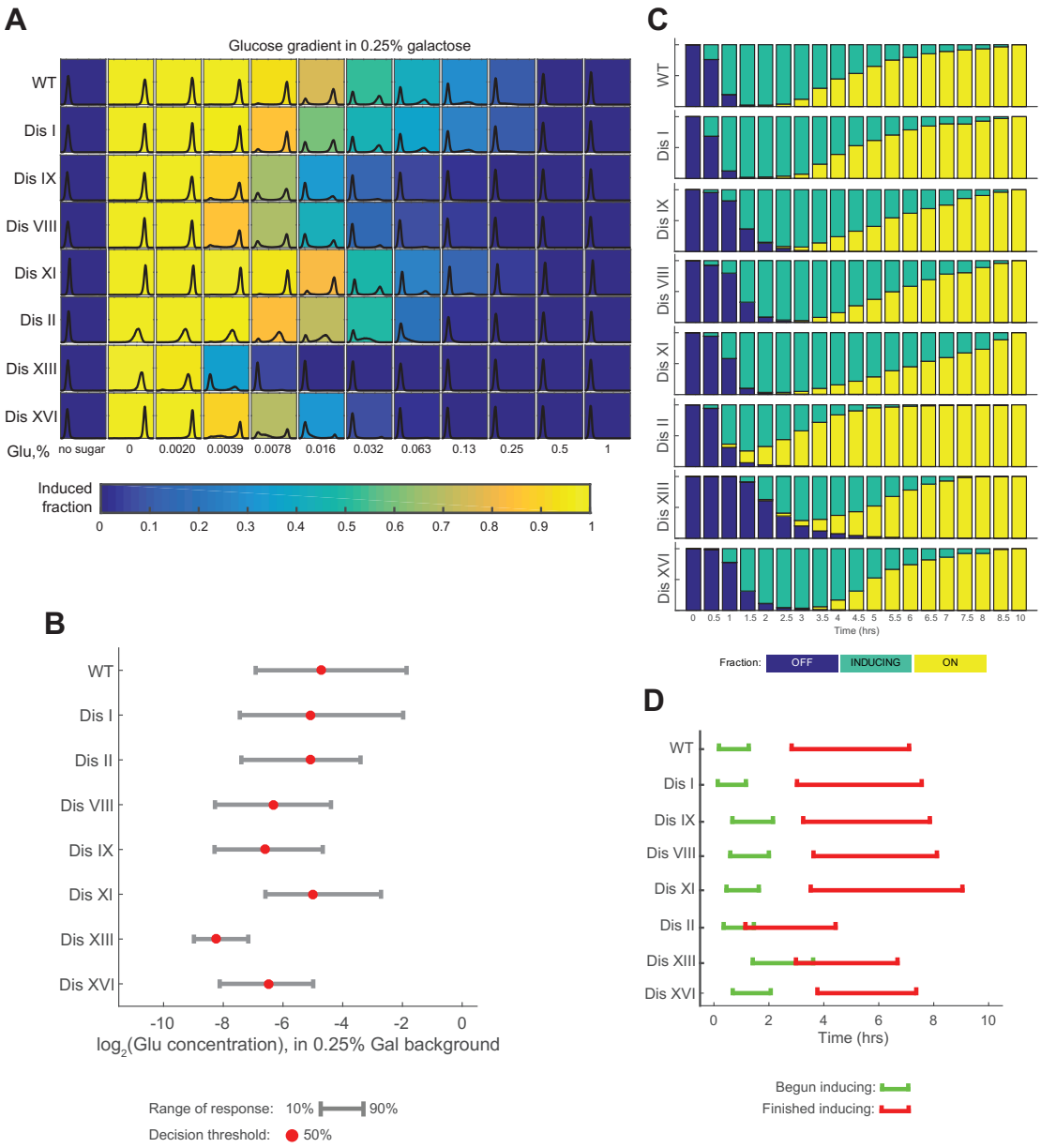

A
Glucose gradient in 0.25% galactose
WT
Dis I
Dis IX
Dis VIII
Dis XI
Dis II
Dis XIII
Dis XVI
Glu,%
no sugar
0
0.0020
0.0039
0.0078
0.016
0.032
0.063
0.13
0.25
0.5
1
Induced fraction
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
C
Time (hrs)
Fraction:
OFF
INDUCING
ON
B
log2(Glu concentration), in 0.25% Gal background
Range of response: 10% 90%
Decision threshold: 50%
D
Time (hrs)
Begun inducing:
Finished inducing:

Figure S3.6 GAL1 induction is variable in aneuploid cells.

(A-B) *GAL1pr-YFP* expression in disomes grown at multiple glucose concentrations in a background of 0.25% galactose. The induced fraction is computed by estimating the YFP fluorescence probability distribution for each well (black histogram trace) and calculating the fraction of area outside the probability distribution of fully repressed cells (far right; Escalante-Chong et al., 2015). Range of response in (B) is calculated from cubic spline curve fitting to the induced fraction values across all glucose concentrations, and is the range from which 10% to 90% of the population is induced (n=4). Data from one representative dataset are shown, and data for all replicates are available on Dryad (identifier number will be provided). (C-D) Kinetics of induction of the galactose pathway when switched from raffinose (non-repressed but uninduced) to galactose medium. Here, the relative percentage of three different population fractions are computed from the YFP probability distribution at each time point: the "on" fraction - the fraction of the area that overlaps with probability distribution at full induction (t=10hr time point); the "off" fraction - the fraction of the area that overlaps with the probability distribution at the uninduced (t=0hr time point); and the "inducing" fraction - the remaining fraction of the area besides the "on" and "off" fractions that corresponds to an intermediate state of induction. The ranges for beginning induction in (D) are calculated based on cubic spline curve fitting to "off" fraction measurements and defined as the time range over which the fraction in the "off" state goes from 90% to 10%; similar fitting is done with "on" fraction measurements to estimate the range for finishing induction (n=3). Strains used in these assays: wild-type (A38340); disomes I (A38341), II (A38342), VIII (A38343), IX (A38344), XI

(A38345), XIII (A38346), and XVI (A39284). Abbreviations: WT, wild-type; Dis, disome; Glu, glucose; Gal, galactose.

Strain	Galactose				Heat shock				DTT			
	Mean	p	SD	p	Mean	p	SD	p	Mean	p	SD	p
WT	3.224	1.0000	0.1708	1.0000	2.423	1.0000	0.1156	1.0000	1.236	1.0000	0.1568	1.0000
Dis I	3.205	0.8857	0.1666	0.8857	2.286	1.55E-04	0.1285	0.0104	1.457	2.12E-07	0.1617	0.2805
Dis IX	3.186	0.8857	0.2414	0.0286	2.405	0.1605	0.1656	1.55E-04	1.192	0.0237	0.1856	2.12E-07
Dis VIII	3.165	0.4857	0.2329	0.0571	2.452	0.0499	0.2035	1.55E-04	1.287	0.0378	0.1538	0.2962
Dis XI	3.236	0.8857	0.1772	0.8857	2.384	1.86E-03	0.1112	0.5737	1.408	3.97E-07	0.1555	0.4688
Dis II	2.527	0.0286	0.2961	0.0286	2.275	1.55E-04	0.2355	1.55E-04	1.268	0.1785	0.1778	2.08E-06
Dis XIII	2.854	0.0286	0.3550	0.0286	2.342	1.55E-04	0.2309	1.55E-04	0.955	2.12E-07	0.1769	2.12E-07
Dis XVI	3.217	1.0000	0.2277	0.0571	2.604	1.55E-04	0.1404	0.0281	1.321	1.13E-04	0.1985	1.44E-08
Dis V					2.527	1.55E-04	0.2260	1.55E-04	0.976	1.44E-08	0.1794	1.89E-08
Dis X									1.300	1.80E-03	0.1646	0.0425
Dis XIV									1.069	6.29E-08	0.1845	2.48E-08
Dis XII									1.094	4.78E-07	0.1942	2.48E-08

Table S3.1: Related to Figure 6. The effects of aneuploidy on the mean and standard deviation of three reporter constructs.

Strains containing an extra copy of chromosome I, II, V, VIII, IX, X, XI, XII, XIII, XIV, or XVI were monitored by single cell fluorescent reporters in galactose, heat shock, and/or DTT (Figure 6, Figure S5). Reporter levels were normalized by SSC as a proxy for cell size, scaled by an arbitrary factor of $10^{3.5}$ for visualization purposes, and $\log_{10}$-transformed. Shading indicates statistically significant increase (red) or decrease (green) from WT ($p<0.05$) by Wilcoxon rank sum test.

Appendix C: Other publications from my graduate study

Wang, J., Atolia, E., Hua, B., Savir, Y., Escalante-Chong, R., & Springer, M. (2015). Natural variation in preparation for nutrient depletion reveals a cost–benefit tradeoff. PLoS Biol, 13(1), e1002041.