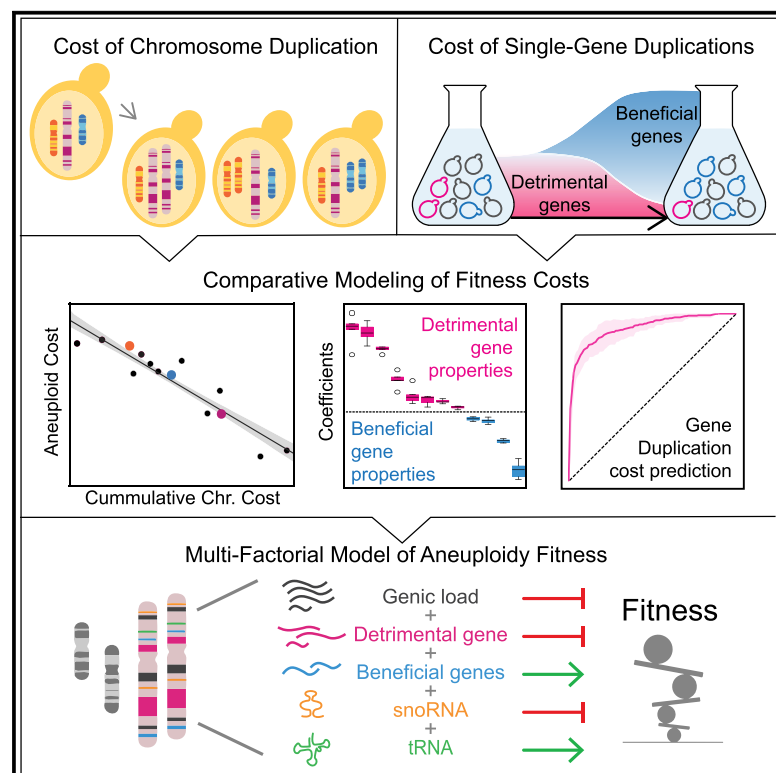


Comparative modeling reveals the molecular determinants of aneuploidy fitness cost in a wild yeast model

Graphical abstract



Authors

Julie Rojas, James Hose, H. Auguste Dutcher, Michael Place, John F. Wolters, Chris Todd Hittinger, Audrey P. Gasch

Correspondence

agasch@wisc.edu

In brief

Why extra chromosomes are harmful to cells has been difficult to understand. Rojas et al. use molecular analysis, statistical modeling, and model validation in eukaryotic budding yeast. A multi-factorial model explains a large portion of the variance in chromosome costs. This model includes the cumulative cost of single-gene duplications, deleterious effects of snoRNAs, and beneficial effect of tRNAs on each chromosome. This work presents a framework for understanding the cost of aneuploidy, with implications for evolution, developmental disorders, and cancer.

Highlights

- Measured costs of chromosome duplication are explained by a multi-factorial model
- The cumulative cost of single-gene duplications drives aneuploidy cost
- Duplication of tRNAs improves, and snoRNAs worsen, the cost of aneuploidy
- Gene length was identified as the best predictor of deleterious gene duplications



Article

Comparative modeling reveals the molecular determinants of aneuploidy fitness cost in a wild yeast model

Julie Rojas,¹ James Hose,¹ H. Auguste Dutcher,¹ Michael Place,^{1,2} John F. Wolters,³ Chris Todd Hittinger,^{1,2,3,4} and Audrey P. Gasch^{1,2,3,4,5,*}

¹Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53706, USA

²Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

³Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706, USA

⁴J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA

⁵Lead contact

*Correspondence: agasch@wisc.edu

<https://doi.org/10.1016/j.xgen.2024.100656>

SUMMARY

Although implicated as deleterious in many organisms, aneuploidy can underlie rapid phenotypic evolution. However, aneuploidy will be maintained only if the benefit outweighs the cost, which remains incompletely understood. To quantify this cost and the molecular determinants behind it, we generated a panel of chromosome duplications in *Saccharomyces cerevisiae* and applied comparative modeling and molecular validation to understand aneuploidy toxicity. We show that 74%–94% of the variance in aneuploid strains' growth rates is explained by the cumulative cost of genes on each chromosome, measured for single-gene duplications using a genomic library, along with the deleterious contribution of small nucleolar RNAs (snoRNAs) and beneficial effects of tRNAs. Machine learning to identify properties of detrimental gene duplicates provided no support for the balance hypothesis of aneuploidy toxicity and instead identified gene length as the best predictor of toxicity. Our results present a generalized framework for the cost of aneuploidy with implications for disease biology and evolution.

INTRODUCTION

Aneuploidy, when cells carry an abnormal number of one or more chromosomes, can produce different outcomes depending on the environmental and cellular context. On the one hand, aneuploidy is broadly considered deleterious. Whole autosome duplication in humans is generally lethal, with the primary exception of trisomy of chromosome 21 that causes Down syndrome (DS).¹ The deleterious effects of chromosome duplication can also be seen at the cellular level in most organisms.^{2,3} On the other hand, aneuploidy is often beneficial during evolution. Chromosome amplifications are frequently selected in drug-resistant human pathogens and represent a major source of drug evasion.^{4,5} Furthermore, aneuploidy is observed in ~20% of non-laboratory *S. cerevisiae* isolates^{6–9} and is associated with adaptive traits in natural and industrial environments.^{10–15} Aneuploidy is also found in >88% of cancers: tumors with high levels of aneuploidy display poorer patient prognosis, respond less well to treatment, and have a higher rate of relapse.^{16,17} Recent studies show that specific chromosome amplifications underlie these benefits.^{17–21} Thus, aneuploidy can be a fast route to adaptation in a changing environment. Whether cells can use aneuploidy to evolve to a new environment depends on the balance between aneuploidy cost and potential benefit—if the benefit under the

conditions at hand outweighs the cost, aneuploidy will be maintained.

However, a major limitation in predicting the impact of aneuploidy is that we lack a mechanistic understanding of why aneuploidy is deleterious under optimal conditions, especially in the case of chromosome amplifications. Previous studies showed that large mammalian chromosomes transformed into yeast and lacking coding potential do not incur the same fitness cost as duplicating native chromosomes, strongly implicating protein-coding sequences as a major contributor.^{22–24} Two mutually exclusive models have been proposed to explain the inherent cost of duplicating chromosomes (herein referred to as aneuploidy). On one end of the spectrum is what we refer to as the genic load model, in which aneuploidy cost is driven by the burden of making extra gene products, independent of their functions or properties.^{25,26} Multiple studies, from yeast to mammals, suggest that larger chromosomes with more genes incur a larger cost.^{2,6} In yeast, chromosome length and gene number negatively correlate with the growth defect of aneuploid strains^{27–29} and with the number of aneuploid strains found outside of the lab, which presumably reflects the strength of negative selection.^{6,9,30} The magnitude of that correlation varies for different studies, which analyze incomplete sets of strains often isolated from multiple sources. In another study, the impact



of large segmental duplications on yeast growth was partly correlated with the number of genes in those segments, although discrepancies were identified.³¹

On the other end of the spectrum is the driver gene model that predicts that aneuploidy toxicity is due to a handful of dosage-sensitive genes encoded on each chromosome. This model prevails in the study of DS, where research often focuses on one or a few specific genes on human chromosome 21.^{32,33} In yeast, one of the most striking examples of a driver gene is thought to be beta-tubulin, *TUB2*, encoded on chromosome VI (chrVI, chr6): chr6 duplication is viable only in the presence of other chromosome duplications that encode Tub2-interacting proteins; these chromosome amplifications occur spontaneously when *TUB2* is duplicated on a plasmid.^{34,35} In cancer cells, the frequency of segmental gains and losses found in The Cancer Genome Atlas database can be partially modeled by the number of tumor suppressors and oncogenes amplified in those regions.^{36–38} These models are based on <8% of genes scored at the time as tumor suppressors and oncogenes, suggesting that only a subset of human gene amplifications contributes a major impact. Other recent studies provide evidence for a mixed model of aneuploidy cost. For example, Keller et al. analyzed a suite of segmental chromosome amplifications in yeast and showed that fitness cost partially correlated with the length of the amplification; however, several outliers implicated that other effects must be at play.³¹ A major limitation in distinguishing any of these models is a lack of systematic study measuring the cost of each chromosome's duplication in a controlled environment and then modeling the mechanistic basis for that cost.

Both models above are compatible with a prominent view of deleterious effects known as the balance hypothesis. This hypothesis posits that duplication of genes encoding proteins with many protein interactions or that participate in multi-subunit protein complexes can produce stoichiometric imbalance, disrupting protein interaction networks and causing downstream stress on protein folding, degradation, and management known as proteostasis stress.^{39–41} Proteostasis stress can be exacerbated by an increased burden produced by many gene amplifications, overloading cellular machineries.^{2,23,42,43} Indeed, yeast aneuploids are sensitive to conditions that interfere with proteostatic functions, including protein translation, folding, and degradation.^{2,23,42–44} However, these models are heavily influenced by results from a laboratory strain of yeast, W303, which is highly sensitized to chromosome duplication. The genetic basis for this sensitivity is a hypomorphic variant of an RNA-binding protein, *Ssd1*, that is required for yeast to tolerate extra chromosomes.⁴⁵ Most non-laboratory strains studied to date are significantly more tolerant of chromosome amplification, although a detectable fitness cost remains, raising questions about the cost and effect of aneuploidy in more representative non-laboratory strains.^{6,7,9}

In this study, we used comparative modeling and molecular validation to distinguish and quantify models of aneuploidy cost under optimal growth conditions in a natural oak-soil isolate of *S. cerevisiae* YPS1009, with and without *SSD1*. In doing so, we leveraged a pooled library of cloned genes to measure the cost of duplicating each gene individually. Our results indicate that a multi-factorial model incorporating the cumulative cost

of individual gene duplications on each chromosome, plus the impact of several non-coding RNA (ncRNA) classes, explains a large proportion of aneuploid growth defects. Surprisingly, we found no evidence for the balance hypothesis in aneuploidy cost and propose that yeast cells have evolved to manage mere duplication of most genes. We used machine learning approaches to identify other features associated with deleterious single-gene duplication. Surprisingly, the most impactful feature predicting the fitness effect of a gene's duplication is its length, since deleterious genes are on average significantly longer than non-deleterious genes. Together, our results raise important considerations regarding the effects of gene and chromosome amplification.

RESULTS

Fitness costs of chromosome duplication vary by chromosome

We began by generating a panel of haploid strains in the oak-soil YPS1009 background in which each chromosome is duplicated. YPS1009 was selected because its response to aneuploidy is representative of that of other strain backgrounds with similar aneuploid content^{9,45,46} and because of existing tool sets in our lab. We used the method of Hill and Bloom⁴⁷ to generate aneuploid cells by integrating a galactose-inducible promoter facing each centromere. Cells were shifted to galactose medium for one generation to induce transcription, which blocks kinetochore assembly and thus causes chromosome retention in the mother cell during mitosis (see STAR Methods). We generated aneuploid strains in which each of 15 of the 16 yeast chromosomes is duplicated. The exception was chr6, proposed previously to be lethal due to amplification of tubulin *TUB2*.^{34,35,48,49} Most of the chromosome duplications were stable over many generations (see STAR Methods). We generated a comparable strain background that was sensitized to aneuploidy through the deletion of *SSD1*.⁴⁵ We were unable to isolate an *ssd1Δ* strain with chr16 duplicated, suggesting that this specific chromosome duplication is not viable in this strain background without *Ssd1*.

The strain panel affords an opportunity to sensitively measure the fitness cost of aneuploidy under standardized conditions. We measured the growth rate of wild-type and *ssd1Δ* strains in the panel, in biological quadruplicate. Not surprisingly, different chromosome duplications inflict different levels of fitness defects (Figure 1A; Table S1). We observed a range of growth rates, from 96% of the euploid growth rate for duplication of chr3 (the second smallest chromosome) to 65% for duplication of chr15, which falls among the larger chromosomes but is not the largest in size or gene content. These results already highlight an imperfect relationship between chromosome size and its fitness cost. *Ssd1* was previously shown to be important for some chromosome duplications in multiple wild strains,⁴⁵ but the breadth of its impact on other chromosomal aneuploidies was not known. We discovered that 8 of the 15 aneuploids (53%) incurred significantly greater growth defects in the *ssd1Δ* background (Figure 1A). Most of the other chromosomes were also more deleterious in the *ssd1Δ* strain but missed the threshold for statistical significance. Thus, *Ssd1* is important for tolerating most

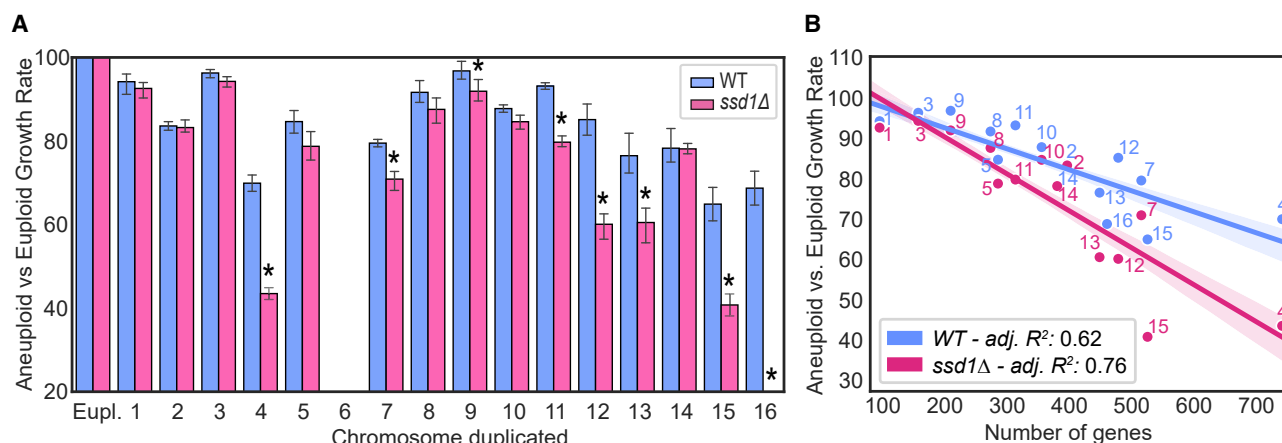


Figure 1. Chromosome duplications inflict variable fitness costs in wild-type and *ssd1Δ* cells

(A) Average and standard deviation ($n = 4$) of aneuploid growth rates relative to isogenic euploid. All *SSD1*⁺ (“WT,” blue) aneuploids grew slower than the euploid ($p < 0.05$, replicate-paired t test); *ssd1Δ* aneuploids that grew significantly slower than their wild-type aneuploid equivalent are indicated with an asterisk ($p < 0.05$, t test).

(B) Mean relative growth rate of each aneuploid strain (numbered by duplicated chromosome) relative to the isogenic euploid plotted against the number of genes per amplified chromosome. Ordinary least-squares regression with 95% confidence interval (shaded) and adjusted R^2 indicated in the box. See also Table S1 and source data for Figures 1, 2, and 3 in S6.

chromosome duplications, with greater impacts for chromosomes that cause a greater defect in wild-type cells.

Genic load partly explains the fitness costs of chromosome duplication

With the fitness costs of each chromosome duplication in hand, we developed mathematical models to understand the determinants of aneuploidy toxicity. For optimal modeling, we first sequenced the YPS1009 genome (see STAR Methods). This produced a high-quality genome of 7,362 annotated genes and non-genic features across 16 assembled chromosomes.

We began by calculating the linear relationship between the relative fitness cost measured for each chromosome duplication (taken as the aneuploid versus euploid growth rates) and the number of genes per chromosome (model 1), which in yeast is highly correlated with the chromosome length ($R^2 = 0.99$). Excluding chr6, which could not be generated, the fit for the remaining chromosomes explains 62% (adjusted [adj] $R^2 = 0.62$) of the variance in relative fitness costs. Thus, the number of genes per chromosome alone explains a significant proportion of the variance of aneuploid fitness cost (Figure 1B), confirming previous implications in various organisms.^{6,9,27,28} The fit was even higher for *ssd1Δ* strains, explaining 76% of the variance in fitness costs of cultivatable chromosome duplications (Figure 1B). The increased slope reflects the stronger fitness costs in *ssd1Δ* aneuploids, suggesting that, in the absence of *Ssd1*, cells are more sensitive to the genic load.

The cumulative effect of single-gene duplications accurately models whole chromosome gain

An open question in the aneuploidy field is the degree to which specific genes on each duplicated chromosome contribute to the fitness cost of aneuploidy. We therefore set out to determine the fitness impact of duplicating each gene individually in the

YPS1009 euploid strain using a single-copy gene duplication library. Each centromeric plasmid contains one yeast gene with its native upstream and downstream regulatory sequence along with a unique DNA barcode for tracking.⁵⁰ To measure the fitness cost of duplicating each gene individually, we transformed euploid YPS1009 with the pooled library and grew it competitively for 10 generations, taking the $\log_2$ (fold change) in barcode abundance after competitive outgrowth as the relative fitness cost (see STAR Methods). Genes whose barcode abundance significantly decreased during growth are considered detrimental to fitness, while those whose frequency increased are considered beneficial. Of the 4,369 YPS1009 yeast genes for which fitness could be measured, 25.5% were beneficial and 28% were detrimental (false discovery rate [FDR] < 0.05 , Figure 2A; Figure S1A), as validated for several representative genes (Figure S1B and Source Table S6).

Because genes with noisy measurements are statistically insignificant but can have artificially skewed mean measurements, we replaced insignificant scores (FDR > 0.05) with the mean cost of all measured genes ($\log_2$ value of -0.33 , see STAR Methods). Hence, genes without a significant effect are considered to have a mild negative impact. We then computed the fitness cost of each chromosomal duplication (chr cost) as the sum of $\log_2$ fitness effects of all genes encoded on that chromosome.

The cumulative-effects model based on single-gene costs (model 2) significantly improved the fit compared to model 1, which considers only the number of genes per chromosome (adj R^2 for wild type = 0.69, *ssd1Δ* = 0.84, Figure 2B). The improvement in the fit was highly significant as assessed in two ways. First, a nested model that includes both the number of genes per chromosome and the cumulative gene cost (normalized to chromosome gene number) improves the fit, since both factors are significant ($p < 3.9 \times 10^{-2}$, likelihood-ratio test, see STAR Methods). Second, the observed fit for model 2

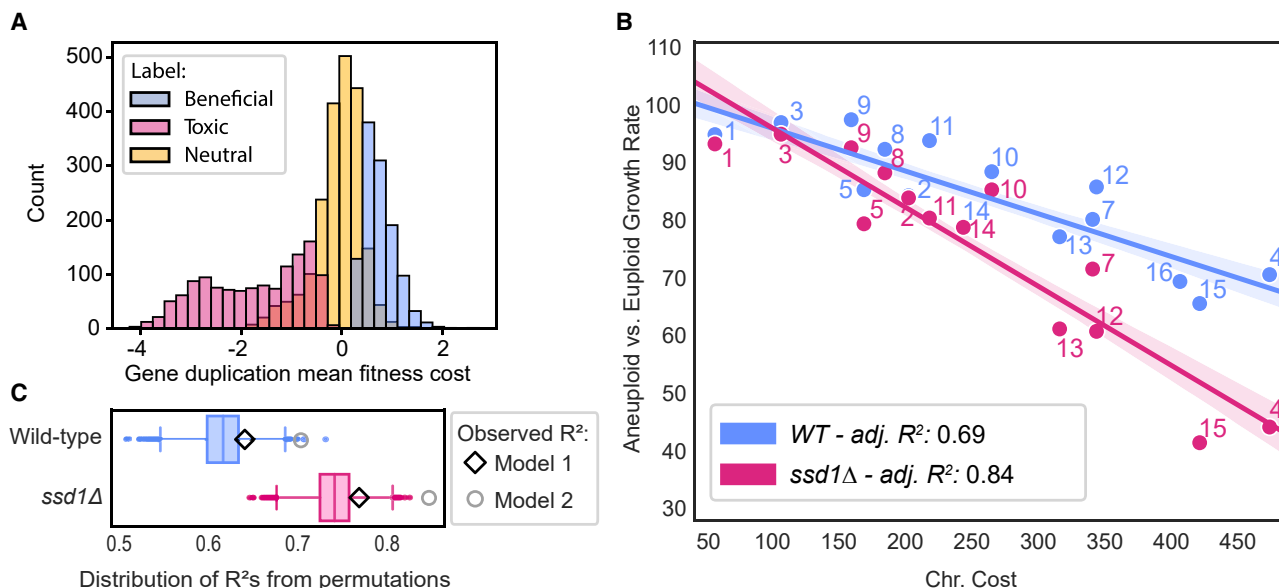


Figure 2. Considering gene-specific fitness costs improves the modeling

(A) Distribution of $\log_2$ fitness scores for single-gene duplications for gene groups in the key. (B) Linear fit of the mean relative growth rate as in Figure 1 plotted against the sum of the $\log_2$ fitness costs for genes encoded on each chromosome ("Chr. Cost"). (C) Distribution of R^2 values from 10,000 random permutations of gene fitness scores affiliated with each chromosome (whiskers – 1.5 times the interquartile range). The observed adjusted- R^2 values for model 1 and model 2 are shown for each strain panel. See also Tables S1, S2, and S6.

was better than nearly all of the 10,000 random permutations of gene fitness costs (preserving the number of genes per chromosome in each trial). Of 10,000 permutations, only 4 met the observed model 2 fit for wild-type aneuploids ($p = 0.0004$) and none for the *ssd1 Δ* strains ($p < 0.0001$, Figure 2C). Together, these results show that the identity of duplicated genes makes an important contribution to the cost of aneuploidy and is predictive of the fitness effect of whole-chromosome duplication. Interestingly, the fit for model 1 that simply counts the number of genes per chromosome was better than 88% and 82% of random trials for the wild-type and *ssd1 Δ* strains, respectively, which were close to statistical significance ($p = 0.17$ for wild type, $p = 0.11$ for *ssd1 Δ*). This raises the intriguing possibility that fungal evolution has optimized gene content on each chromosome to minimize the cost of chromosome duplication, which is relatively frequent in yeast. Regardless, these results show that the combination of single-gene fitness effects is predictive of the fitness effect of whole-chromosome duplication.

We devised an independent experimental approach to test the models using strains carrying two chromosome duplications. These dual-chromosome duplications were not stable in *ssd1 Δ* cells, and thus we focused on *SSD1 $^+$* strains. Those with multiple chromosome duplications grew slower than corresponding single-chromosome duplication strains, as expected (Figure S2A). We assessed the variance in growth rates of dual-chromosome duplications explained by the models trained on single-chromosome duplications. Indeed, model 2 was significantly better (adj $R^2 = 0.54$) than model 1 (adj $R^2 = 0.34$, Figures S2B and S2C). Thus, the model does not overfit the training data and instead shows that the cost of chromosome duplication is significantly influenced by the genes encoded on each chromosome.

Beneficial gene duplications alleviate the cost of chromosome duplication

Studies in cancer cells suggested that beneficial oncogenes on amplified chromosomes counteract tumor suppressors on the same segments.^{36,37} We wondered if genes whose duplication is beneficial to YPS1009 are important for aneuploidy fitness. To test this, we excluded beneficial genes from the chr cost, which decreased the model performance (adj R^2 of 0.67 for wild type and 0.80 for *ssd1 Δ* compared to 0.69 and 0.83, respectively, for model 2). The contribution of beneficial genes is statistically significant in a nested model in which their cumulative effect was added as a separate feature ($p = 4.7 \times 10^{-2}$, likelihood test). Hence, genes scored as beneficial when duplicated in isolation contribute to aneuploidy fitness, likely because they collectively counter some of the aneuploidy fitness cost.

Non-coding features contribute to aneuploidy fitness effects

While it is clear that gene fitness costs explain much of the cost of chromosome duplication, non-coding features could also contribute. We therefore compiled a set of non-genic features per chromosome based on the YPS1009 genome sequence and used lasso regression to identify additional features that improve predictions. The input set included the numbers of small nucleolar RNAs (snoRNAs), tRNAs, other ncRNAs, autonomous replicating sequences (ARSs), retrotransposons, and long terminal repeats (LTRs), all normalized by the total number of features encoded on each chromosome (see STAR Methods). Aside from LTR and retrotransposon numbers, most of the features were not confounded by co-variation (Figure S3).

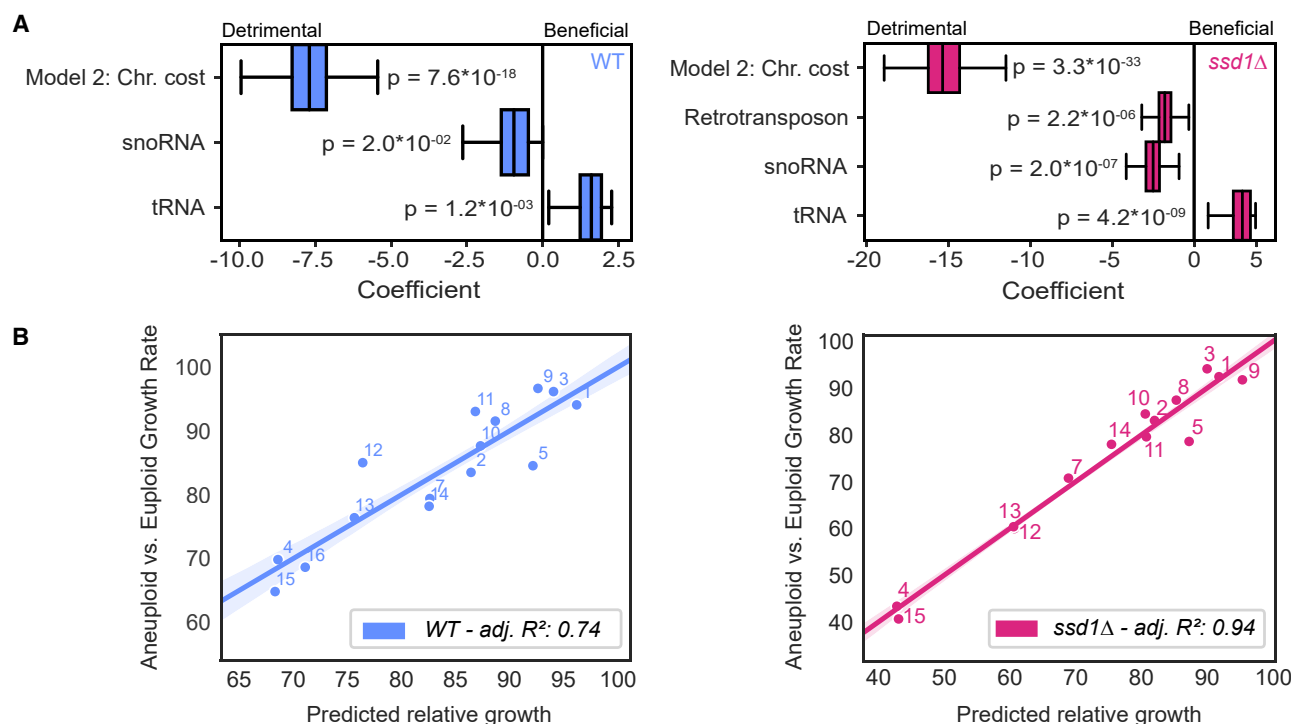


Figure 3. A multi-factorial model best explains the costs of chromosome duplication

(A) Distribution of coefficients obtained from 1,000 lasso regression bootstrap iterations (whiskers – 1.5 times the interquartile range). Only features exhibiting non-zero weights in more than 90% of bootstrap resamples are depicted. The likelihood-ratio test's p values for each selected feature for the wild-type (blue) and *ssd1Δ* (pink) regression models are displayed.

(B) Linear fit of the mean relative growth rates as in Figure 1 against model 3 predictions (using significant features for each strain as shown in A). The adjusted R^2 value is indicated in the lower right corner. See also Tables S1, S2, S5, and S6.

We used a bootstrap-lasso (Bolasso-S)⁵¹ approach to select features that contribute significant explanatory power to the modeling of measured aneuploidy growth defects, selecting from non-genic features as well as model 2 chr costs. Features selected by lasso regression in 90% of 1,000 bootstrap trials (lasso alpha factor = 0.7, see STAR Methods) were retained and incorporated into multi-factorial model 3. For both wild-type and *ssd1Δ* models, lasso chose the chr costs from model 2 as the most impactful factor but also the normalized number of snoRNAs per chromosome as deleterious to fitness and the normalized number of tRNAs per chromosome as beneficial (Figure 3A). The method also chose the normalized number of retrotransposons as a deleterious predictor only for the *ssd1Δ* strain. All selected features were significant (chi-squared test, Figure 3A). Remarkably, the multi-factorial model 3 explains 74% of the growth rate variance for wild-type and 94% for *ssd1Δ* aneuploids (Figure 3B). When the trained models were assessed on dual-chromosome duplication strains, model 3 improved the predictions compared to model 2 (adj R^2 = 0.7 compared to 0.54 for model 2, Figure S2D).

Imbalanced duplication of snoRNAs is detrimental

The lasso predictions above improve the modeling, but is the model correct? We set out to experimentally verify several of the model predictions. We first tested the predicted deleterious impact of duplicating snoRNAs. snoRNAs guide catalytic modifications of other RNAs, such as ribosomal RNAs (rRNAs) and

tRNAs. snoRNAs can be split into C/D box snoRNAs, which direct 2'-hydroxyl methylation of their RNA targets, and H/ACA box snoRNAs involved in pseudouridylation.⁵² The two groups were combined into one for modeling given their relatively small numbers in the genome (45 C/D and 29 H/ACA). To test predictions, we cloned seven C/D snoRNAs present in an array on chr13 or seven H/ACA snoRNAs from a single region on chr15 onto centromeric plasmids (see STAR Methods). Duplication of either snoRNA cassette significantly reduced growth of the euploid strain, validating that duplication of these cassettes is indeed deleterious (Figure 4A). Furthermore, the growth rates of haploid YPS1009 carrying duplications of chr4 (among the most deleterious chromosomes, which also encodes fewer snoRNAs than others) or chr15 were also reduced upon duplication of these snoRNAs (despite missing the significance threshold in one case, Figure 4A).

Reciprocally, if snoRNAs contribute to aneuploidy toxicity, then restoring them to euploid copy number should partially alleviate the aneuploidy fitness costs. With that aim, we deleted from one of the chr13 copies a segment of six of its nine C/D snoRNAs (see STAR Methods). Although there was no significant effect in the wild type, deleting the extra C/D snoRNA copies from the *ssd1Δ* chr13 aneuploid strain significantly improved its growth rate (Figure 4B). The increased sensitivity of *ssd1Δ* aneuploids may provide more power to detect improvements than in the wild type, where snoRNA imbalance was also predicted to be

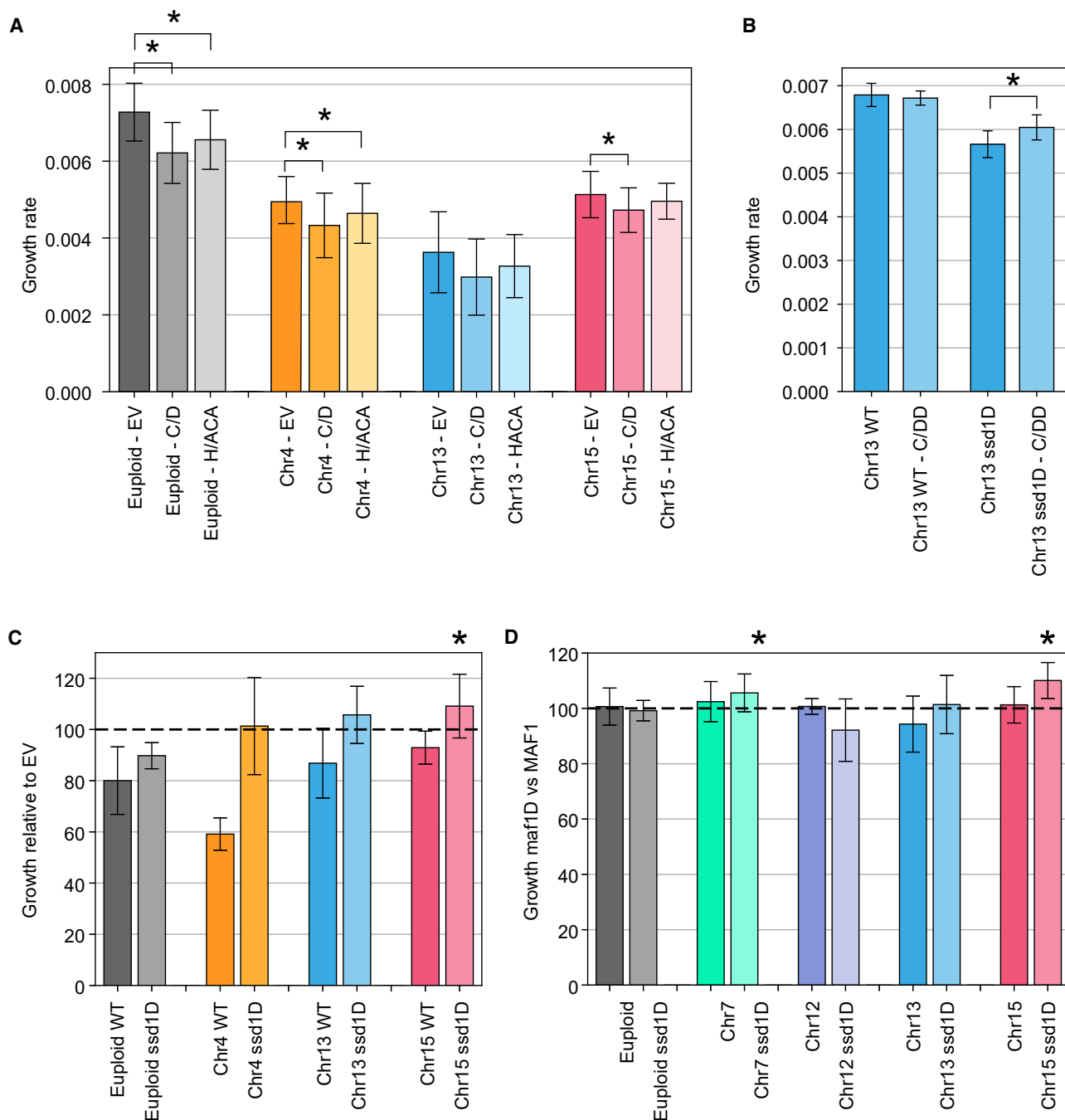


Figure 4. Duplication of select snoRNAs and tRNAs contributes to aneuploidy fitness

(A) Average and standard deviation of growth rates of strains containing the empty vector (EV) or plasmids encoding either seven C/D box snoRNAs or seven H/ACA snoRNAs as described in the text (* $p < 0.05$, replicate-paired t test versus empty vector, $n > 6$).

(B) Average and standard deviation of growth rates of chr13 aneuploids with or without restoration of the seven C/D box snoRNAs' copy number to euploid levels (* $p < 0.05$, replicate-paired t tests, $n > 7$).

(C) Average and standard deviation of relative growth rates of strains harboring chr12 tRNA cassette versus strain with the empty vector (* $p < 0.01$, replicate paired t tests, between each aneuploid and the corresponding euploid, $n > 3$).

(D) Average and standard deviation of relative growth rates of each strain in the *maf1Δ* versus *MAF1*⁺ background (* $p < 0.05$, replicate-paired t tests between *MAF1* and *maf1Δ*, $n > 4$). See source data in [Table S6](#).

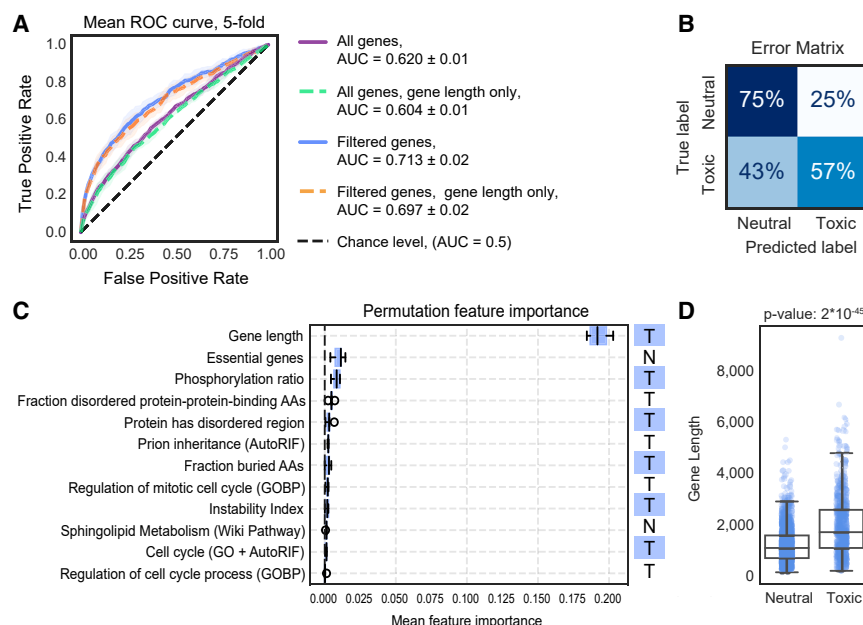


Figure 5. Gene length is the main predictor of deleterious gene duplications

(A) Mean receiver operating characteristic (ROC) curve for 5-fold cross validation of the logistic regression model using the top 12 features (see STAR Methods), applied to 1,177 deleterious and 3,028 neutral gene duplications (all genes) or the restricted set of 613 substantially deleterious genes and 1,472 clearly neutral genes (filtered genes). Dashed, colored lines show the fit when only gene length is considered in the model. The mean area under the curve (AUC) is shown in the key.

(B) Error matrix shows the percentage recovery of true labels by the predicted labels of the combined 5-fold cross-validation test sets.

(C) Boxplot of the mean feature importance ($n = 10$) for the 5-fold cross-validation measured with respect to ROC-AUC gain (whiskers – 1.5 times the interquartile range, see STAR Methods). Features associated with or higher in the deleterious gene duplication group are labeled with a “T”, while enrichment in the neutral group is indicated with an “N”.

(D) Distribution of gene lengths for the 613 deleterious (“toxic”) and 1,472 neutral gene duplications (p value, Wilcoxon rank-sum test). See also Tables S2 and S3 as well as corresponding source data in Table S6.

deleterious. Nonetheless, together, these results confirm that snoRNA duplication is deleterious to both euploid and aneuploid cells and contributes to the cost of chromosome duplication in at least the *ssd1Δ* background.

Increasing tRNA copy number benefits *ssd1Δ* aneuploid cells

Model 3 above predicts that chromosomes with more tRNAs are less deleterious than otherwise predicted. We tested this in several ways. First, we introduced an available plasmid carrying 21 tRNAs encoded on chr12⁵³ into the YPS1009 euploid and a subset of aneuploid strains. The tRNA plasmid decreased proliferation in the euploid and chr4 aneuploid wild-type cells, indicating that an imbalanced set of these tRNAs is deleterious (Figure 4C). However, their duplication had a less detrimental effect in the other aneuploids, especially strains lacking *SSD1*. In fact, duplication of these tRNAs was beneficial to varying degrees in *ssd1Δ* aneuploids with chr13 and chr15 duplications.

As an alternative approach, we assessed the effect of upregulating all tRNAs by deleting the RNA polymerase III repressor, *Maf1*. *MAF1* deletion leads to an accumulation of tRNAs,⁵⁴ which we confirmed (Figure S4). We found that *MAF1* deletion improved growth rates for chr7 and chr15 aneuploids in the *ssd1Δ* background ($p < 0.05$, Figure 4D). Although the effects were somewhat mixed, these results suggest that several aneuploidy-sensitized *ssd1Δ* strains benefited from extra tRNAs but that the effect could be specific to certain chromosomes or tRNAs (see the discussion).

Machine learning implicates properties common to duplication-sensitive genes

Although the cumulative cost of gene duplications explains a significant proportion of the cost of aneuploidy, some gene dupli-

cates are more deleterious than others. To further explore this, we sought properties that can predict deleterious genes. We focused on 1,177 genes scored as deleterious when duplicated in euploid YPS1009 ($FDR < 0.05$), compared to 3,028 genes whose duplication was neutral or beneficial ($FDR > 0.05$, herein referred to as “neutral”). Consistent with other studies using gene duplication libraries,^{55–59} we found only a handful of functional terms enriched in the deleterious group, including several categories linked to cell-cycle regulation. We next compiled a list of 120 gene and protein properties and selected those that differentiated the deleterious gene duplications from the neutral set (Wilcoxon rank-sum test, Figure S5A; Tables S2 and S3). The group of deleterious genes displayed a slightly higher proportion of intrinsically disordered regions, marginally more phosphorylated sites, a higher proportion of serine residues, lower translation rates as indicated by ribosome profiling,⁶⁰ and longer length (Figure S6); however, several of these features are correlated with one another (Figure S5A), confounding interpretation. Notably, the group of genes that are deleterious when duplicated was not enriched for those encoding proteins involved in complexes or with a high number of protein-protein interactions (see the discussion).

We used a machine-learning approach to identify the most impactful gene properties and determine if their combination can accurately differentiate deleterious gene duplications from those that are neutral or beneficial (see STAR Methods). A logistic regression classifier was trained on significant biophysical and functional enrichment (Figure S5A). Five-fold cross validation revealed that the model performed relatively poorly, with a mean area under the curve (AUC) of 0.62 (Figure 5A). Restricting the classification to the 613 most deleterious genes (bottom 15% quantile) and the 1,472 genes most confidently called

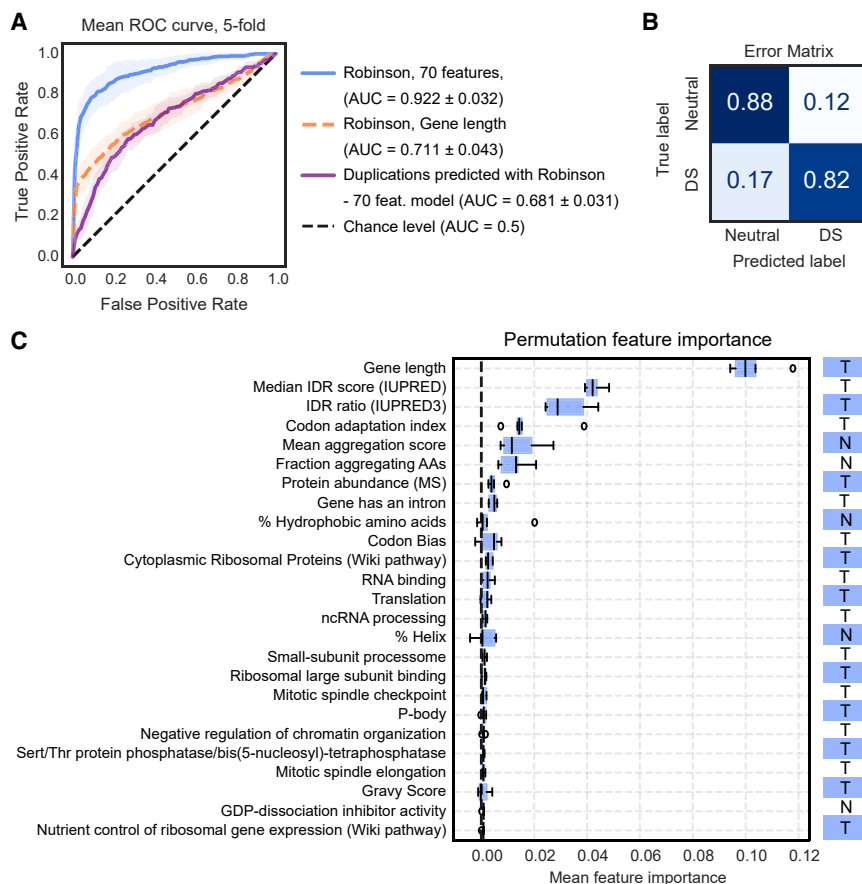


Figure 6. Model predictions applied to 2- μ m overexpression dataset

(A) As shown in [Figure 5](#) but using the top 70 identified features applied to 400 commonly deleterious genes versus 1,657 commonly neutral genes based on data from Robinson et al.⁶¹ (blue curve). Robinson data fitted only with gene length (dashed line). Gene-duplication data from this study (“duplications,” purple curve) predicted using the 70 feature-model trained on the Robinson data.

(B) Error matrix for the Robinson et al. model as described in [Figure 5](#).

(C) Boxplot of the mean feature importance ($n = 10$) for the 5-fold cross-validation, measured with respect to ROC-AUC gain (whiskers – 1.5 times the interquartile range, see [STAR Methods](#)) for Robinson’s model with the top 25 features, as shown in [Figure 5](#). See the source data in [Table S6](#). A complete report of the permutation feature importance for all 70 features of the model is available for this figure in [Table S6](#).

rious genes (Figures 6A and 6B). Thus, the poor performance in predicting duplication-sensitive genes is not due to our methods. In fact, the model trained on the higher-copy 2- μ m library performed relatively poorly when applied to the gene-duplication datasets (Figure 6A), with an AUC of 0.68 that was once again no better than considering gene length alone. The only common predictor between the models trained on duplicated

neutral/beneficial (upper 65% quantile) improved performance (AUC = 0.713), correctly predicting 57% of deleterious gene duplications (Figures 5A and 5B). Surprisingly, by far the most impactful feature in explaining deleterious genes was gene length; deleterious genes were significantly longer than neutral genes (Figures 5C and 5D). A model considering only gene length had nearly equal predictive power as the more complex model (Figure 5A). In an attempt to identify other gene properties that could in combination supplant gene length in the model, we trained a classifier without considering gene length; but the classifier performed worse (mean AUC = 0.66) than when fitted on gene length alone, and the most impactful features selected (ratio of buried residues and the presence of disordered regions) both correlate with gene length (Figure S5A). Thus, gene length distinguishes the deleterious gene set better than any other combination of considered features.

These results were especially surprising because past work from our lab using a higher-copy library identified shared features among genes that are deleterious when overexpressed, including genes encoding proteins with many protein interactions, higher expression, intrinsic disorder, and other features.⁶¹ We therefore applied our modeling approach to discriminate 400 genes whose higher-copy expression on a 2- μ m plasmid is deleterious to many strain backgrounds from genes that are neutral or beneficial in most strains (1,657 genes).⁶¹ This classifier was highly accurate (AUC = 0.92), correctly predicting 82% of deleterious

genes versus the 2- μ m overexpression experiment⁶¹ was gene length, along with different measures of intrinsically disordered regions, suggesting it as a common factor (Figure 6C). However, the latter features made only a marginal contribution to explaining deleterious gene duplicates while being very prominent features for gene overexpression.

We conclude that most biological features that account for deleterious effects when genes are overexpressed to higher levels may not be relevant for mere gene duplications. In both models, but especially in the case of gene duplication, gene length is the single best predictor of whether a gene duplication will be deleterious to strain fitness (see the [discussion](#)).

DISCUSSION

Through systematic experimental and mathematical analysis, our results present a clarified view of the cost of chromosome duplication and the molecular properties behind it. Under standard growth conditions, the cost of aneuploidy cannot be fully explained by generic gene load nor by a handful of duplication-sensitive genes. Instead, our results quantitatively confirm previous suppositions^{31,62} that both the generalized burden of aneuploidy load coupled with combinatorial effects of the specific suite of genes and non-genic features on each chromosome explain 74%–94% of the aneuploidy costs measured here. Some duplicated genes are more deleterious than others, while

beneficial genes help to counteract the burden of deleterious genes on the same chromosomes. Thus, the cost of chromosome duplication is an emergent property of the affected genes and the collective burden of amplifying coding and non-coding genetic elements. Although not investigated here, it is likely that genetic interactions among genes duplicated together also contribute, albeit to a lesser extent than the simple cumulative effects modeled here, perhaps explaining a portion of the 6%–26% variance not explained by our models.

Although the cost of chromosome duplication is explained by these combined effects, it is important to highlight that duplication of single genes on a chromosome can have a disproportional impact on specific phenotypes. This may explain arsenic resistance contributed by amplification of *S. cerevisiae* chr16, which encodes arsenic resistance genes,⁹ or fluconazole evasion by amplification of *C. albicans* chr5, which encodes drug pumps and their regulators.⁴ A similar implication was made for trisomy 21, by correlating specific DS phenotypes to genes amplified in subsets of people with partial-chromosomal trisomies.^{63,64} These single-gene effects almost certainly contribute to chromosome-specific impacts observed for different karyotypes.^{3,6} In terms of evolution, if the benefit provided by the resulting phenotypes outweighs the underlying cost of chromosome amplification, aneuploidy will be maintained. Notably, this cost-benefit analysis is heavily dependent on the environmental context, and that balance can shift with changing environments or genetic background. Indeed, genes with detrimental or beneficial effects can vary substantially across strain backgrounds and in a strain-by-environment manner.^{61,65} Hence, while the principles outlined here are likely generalizable across strains and potentially other species, the impacts of specific genes and chromosomal features, as well as their relationship with environment or genetic background, are likely to differ across systems.

The contribution of snoRNAs and tRNAs points to aneuploidy impacts on translation

Our work implicates the contribution of ncRNAs to the cost of chromosome duplication. The modeling predicted, and experimental analysis confirmed, that imbalanced expression of tested snoRNAs incurs a fitness cost in euploids and select aneuploids, whereas restoration of their balance can alleviate toxicity in the *ssd1Δ* chr13 aneuploid. The altered abundance of specific snoRNAs can produce cellular phenotypes. For instance, overexpression of snoRNA *SNR51* in budding yeast increases binding to its target RNAs.⁶⁶ RNA-mediated modifications can be heterogeneous in the population of substrate molecules, such that modifications could contribute to cellular heterogeneity, including ribosome functions.⁶⁷ Thus, aneuploidy-induced imbalance could change the landscape of rRNA and tRNA modifications, leading to broader effects on translation.⁶⁸ In cancer, snoRNA dysregulation has been associated with both tumor-suppressing and tumor-promoting effects (reviewed in Zhang et al.⁶⁹). In one example, snoRNA overexpression was shown to upregulate ribosome biogenesis, which interferes with the p53 protective role,⁷⁰ hence linking snoRNA misexpression to the disruption of translation in humans. The adverse effect of

snoRNA overexpression might therefore be generalizable to other organisms.

In contrast, chromosomes with more tRNAs were less toxic than the model otherwise predicted, in both wild-type and *ssd1Δ* strains, pointing to a role for tRNAs in alleviating the cost of aneuploidy. We confirmed this prediction experimentally in several sensitized *ssd1Δ* aneuploids, albeit with mixed results in the wild-type strain. We considered all tRNAs together in the modeling, as our relatively small dataset does not have the statistical power to test individual tRNA contributions, but different tRNA duplications may differentially benefit different chromosome amplifications. What could be the reason? The abundance of specific tRNAs correlates with the frequency of their cognate codons in the transcriptome, since higher abundance of those tRNAs facilitates translational efficiency through their codons. In fact, tRNA pools can shift composition to accommodate a changing transcriptome.⁷¹ In recent years, tRNA overexpression has emerged as an important feature of cancer,^{72–74} since upregulation of specific tRNAs increases translation of transcripts enriched for their cognate codons, thereby promoting metastasis.^{75,76} Thus, the benefits of specific chromosome arm gains could be partially linked to specific tRNA duplications.

The implication of snoRNAs and tRNAs adds to a growing body of evidence that aneuploids may have a liability related to translation. First, *Ssd1* is required to manage the stress of chromosome duplication across strain backgrounds and amplified chromosomes.⁴⁵ *Ssd1* has been implicated in translational repression and mRNA localization,^{45,77–79} among other processes. Intriguingly, *SSD1* deletion sensitizes euploid strains to mutation of the elongator complex as well as *Deg1* tRNA pseudouridine synthase, both of which modify tRNAs to promote translational fidelity.^{80,81} These links connect *Ssd1* to aneuploidy and translation, but also to snoRNAs and tRNAs that are implicated in our modeling. Recent work from our lab shows that overexpression of genes involved in translation or translation quality control can partly complement *ssd1Δ* aneuploid growth defects during the exponential phase or *SSD1*⁺ aneuploid defects entering quiescence during the stationary phase.^{82,83} Both *SSD1*⁺ and, especially, *ssd1Δ* aneuploids are inherently more sensitive to translation elongation inhibitors,^{24,45} suggesting that translational stress is likely at play in wild-type aneuploids. We proposed that *SSD1*⁺ strains can largely buffer the cost of most chromosome duplications unless otherwise compromised by translational stress.⁴⁵ Evidence in aneuploid yeast and trisomy 21 cells indicates that protein dosage control is mostly post-translational at the level of protein turnover, suggesting that overabundant mRNAs are indeed translated.^{22,84,85} It is possible that it is the translation of duplicated genes, rather than the overabundance of encoded proteins, that contributes significantly to the fitness defects in aneuploid cells.

Gene length is the strongest predictor of deleterious gene duplications

The cost of chromosome duplication is well modeled by the cumulative cost of duplicating individual genes on each chromosome; thus, considering the features of deleterious gene duplications can further our understanding of aneuploidy. We expected that genes encoding multi-subunit complexes and

with multiple protein-protein interactions would be among the most deleterious, thus validating long-standing models of protein imbalance as a major cause of aneuploidy toxicity. However, deleterious gene duplications were not enriched for either feature. This recapitulates several other studies that also saw no enrichment for components of protein complexes among duplication-sensitive genes.^{55,59,86} The absence of these signatures indicates that the balance hypothesis,^{41,87} often invoked to explain aneuploidy toxicity, may well be true for high-level protein imbalance but not for mere duplication of genes and their native regulatory sequences. The reason is likely to be dosage control, which has been observed repeatedly for multi-subunit proteins amplified in yeast and human cells.^{22,55,85,88–91} While some dosage control can happen at the transcriptional level,⁷ much occurs post-translationally. For example, proteins encoded by human chromosome 21 show increased turnover rates.⁸⁵ Genes encoded by other aneuploid chromosomes in human cell lines also show increased degradation rates according to their role in the complex.⁹² Hence, cells likely have evolved mechanisms to manage stoichiometric balance of important proteins, at least when their genes are merely duplicated.

We were surprised that modeling predicted a single major feature—gene length—as the strongest predictor of deleterious gene duplicates, with longer genes associated with dosage sensitivity. This is unlikely due to DNA/plasmid burden, since aneuploidy-sensitized yeast are susceptible to chromosome duplications, but not large artificial chromosomes without coding potential.^{24,93,94} Furthermore, although deleterious genes tend to be longer, many long genes are still scored as neutral, which is not expected if DNA burden is the driving cause. Re-analysis of previously published overexpression screens, in both wild yeast isolates and the laboratory strain, indicates that deleterious overexpressed genes identified in each study are longer ($p = 1 \times 10^{-7}$ to 4×10^{-43} , Wilcoxon rank test).^{56,57,59,61} The patterns we observe may be conserved in higher organisms. Indeed, Ni et al. found that a compiled list of reliable dosage-sensitive genes is significantly longer than one of genes reported as dosage insensitive.⁹⁵ Thus, altered copy number of longer genes is more likely to cause fitness problems across organisms.

There are several possible reasons longer genes tend to be more deleterious when duplicated. First, gene length is correlated with multiple other biophysical features: larger proteins are more likely to contain an intrinsically disordered region, have more phosphorylated sites, and have a higher fraction of buried residues. One possibility is that gene length is simply a proxy for a multitude of other gene properties that are each mildly deleterious. However, we did not find strong support for this hypothesis: when gene length was omitted from the model, several features correlated with gene length were selected, but the model did not perform as well as using gene length alone. It remains possible, however, that longer protein primary sequences are more likely to capture some deleterious features.

Another possibility is that longer genes and transcripts create more chances for error during protein synthesis. Longer genes typically display slower translation initiation and elongation rates, a relationship conserved across organisms.^{96–100} This relation-

ship could reflect higher-order RNA structure or other features of long mRNAs^{101,102}; indeed, of the subset measured, deleterious gene duplicates do have more structure ($p = 0.0008$).¹⁰³ Longer transcripts also increase the probability of translation errors, including tRNA/amino acid misincorporation, ribosome frameshifting, premature termination, and co-translational protein folding errors, all of which are influenced by sequence but are also proportional to transcript length.^{102,104–106} On the one hand, long coding sequences, independent of other problematic sequences, are preferentially targeted by surveillance mechanisms. This may be driven by the reduced translation rate of long mRNAs, which could emerge from an increased probability of translational errors as coding length increases.¹⁰⁷ On the other hand, nonsense-mediated decay (NMD), a pathway that responds to translational errors including frameshifting and premature termination, is less efficient when NMD-triggering sequences are introduced into longer open reading frames.¹⁰⁸ Thus, long open reading frames could lead to more translational errors, both probabilistically due to length and from increased chance of escape of surveillance systems. Translational errors in turn can lead to proteostasis stress and an energy burden to manage that stress.^{96,102} Indeed, managing proteostasis stress through quality control pathways such as the ubiquitin proteasome system is important in sensitized aneuploid strains^{109,110}; however, the direct source of the proteostasis stress remains unclear—our results suggest that translational errors could contribute.

In all, our study presents a quantitative assessment of aneuploidy cost, in a single strain, single growth phase, and controlled environment. Given that many principles in yeast are conserved in higher organisms, the principles reported here are likely conserved; however, the details, including precise fitness costs of specific genes and non-genic features, as well as the generalized sensitivity of strains to translational and proteotoxic stress, could vary significantly across strains, organisms, and conditions.^{61,65} It will be interesting to see if the results observed here pertain to cancer cells, which often benefit from amplified chromosomes.^{17–21} Pioneering work by Davoli et al. showed that the sum of oncogenes and tumor suppressors can partially predict the gain and loss of that chromosome in cancer cells.³⁶ Including our findings in predictive models could improve accuracy in modeling specific copy-number variants in cancer.

Limitations of the study

One limitation of our study is that we focused on fitness costs during exponential growth and under standard, optimal conditions—certainly, the fitness cost of aneuploidy varies under alternate conditions, where amplification of specific genes can be beneficial.^{4,5,9,31} Other evidence from our lab shows that the fitness cost is dramatically different in stationary phase, when starved haploids enter quiescence.⁸³ Thus, an important area for future investigation is to quantify fitness costs under the full range of natural growth phases and conditions, to enable dynamic modeling of evolutionary patterns in real-world situations. Another limitation is that our work is specific to one *S. cerevisiae* strain. While the general principles observed here may pertain to other strains and species, the cost of individual

gene duplications—and thus the cost of specific chromosome duplications—is likely different for other genetic backgrounds. Finally, while machine learning implicated trends among deleterious genes, most notably the association with gene length, the mechanistic basis behind that association will require molecular dissection.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and strains and resources should be directed to and will be fulfilled by the lead contact, Audrey P. Gasch (agash@wisc.edu).

Materials availability

Plasmids and yeast strains generated in this study are available upon demand. This study did not generate new unique reagents.

Data and code availability

- The YPS1009-derivative strain genome assembly used to assign genes to chromosomes and detect non-coding genes has been deposited at NCBI, BioProject, and is publicly available. Accession numbers are listed in the [key resources table](#). The gene duplication screen raw sequencing and barcode counts have been deposited in GEO and are publicly available. Accession numbers are listed in the [key resources table](#). This paper analyzes existing, publicly available data. The accession numbers or references for the datasets are listed in the [key resources table](#).
- The code used to generate all findings and figures, together with the datasets required to run all codes, is available at <https://zenodo.org/records/12701832> (<https://doi.org/10.5281/zenodo.12701832>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

Thanks to Dr. Patrick Cai for sharing his tRNA plasmids, Michael Newton for statistical advice, and the Gasch Lab for useful discussions. This study was funded by the NIH (R01GM147271). H.A.D.'s research is funded by an NHGRI training grant to the Genomic Sciences Training Program (T32HG002760) and an NIH training grant (T32GM007133). Research in the Hittinger lab is funded by the National Science Foundation (DEB-2110403), by the USDA National Institute of Food and Agriculture (Hatch Project 7005101), in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409), and by an H.I. Romnes Faculty Fellowship (Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation).

AUTHOR CONTRIBUTIONS

Conceptualization, J.R. and A.P.G.; manuscript writing, J.R. and A.P.G.; investigation, J.R., J.H., and H.A.D.; methodology, J.R., J.H., and H.A.D.; formal analysis, J.R., J.H., and H.A.D.; YPS1009 genome assembly, M.P. and J.F.W.; YP1009 genome annotation, M.P.; mentorship of J.F.W., C.T.H.; resource sharing, C.T.H.; supervision, A.P.G.; funding acquisition, A.P.G.; project administration, A.P.G.

DECLARATION OF INTERESTS

The authors have no competing interests to declare.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Strains and plasmid
- [METHOD DETAILS](#)
 - Growth conditions
 - YPS1009 genome sequencing
 - Gene duplication fitness cost measurements
 - Modeling aneuploidy fitness costs
 - Deleterious gene duplications classifier
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2024.100656>.

Received: April 23, 2024

Revised: July 10, 2024

Accepted: August 20, 2024

Published: September 23, 2024

REFERENCES

- Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* 2, 280–291. <https://doi.org/10.1038/35066065>.
- Torres, E.M., Williams, B.R., and Amon, A. (2008). Aneuploidy: Cells Losing Their Balance. *Genetics* 179, 737–746. <https://doi.org/10.1534/genetics.108.090878>.
- Zhu, J., Tsai, H.-J., Gordon, M.R., and Li, R. (2018). Cellular Stress Associated with Aneuploidy. *Dev. Cell* 44, 420–431. <https://doi.org/10.1016/j.devcel.2018.02.002>.
- Selmecki, A., Forche, A., and Berman, J. (2006). Aneuploidy and Isochromosome Formation in Drug-Resistant *Candida albicans*. *Science* 313, 367–370. <https://doi.org/10.1126/science.1128242>.
- Vande Zande, P., Zhou, X., and Selmecki, A. (2023). The Dynamic Fungal Genome: Polyploidy, Aneuploidy and Copy Number Variation in Response to Stress. *Annu. Rev. Microbiol.* 77, 341–361. <https://doi.org/10.1146/annurev-micro-041320-112443>.
- Gilchrist, C., and Stelkens, R. (2019). Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast* 36, 525–539. <https://doi.org/10.1002/yea.3427>.
- Hose, J., Yong, C.M., Sardi, M., Wang, Z., Newton, M.A., and Gasch, A.P. (2015). Dosage compensation can buffer copy-number variation in wild yeast. *Elife* 4, e05462. <https://doi.org/10.7554/eLife.05462>.
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freil, K., Llored, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. <https://doi.org/10.1038/s41586-018-0030-5>.
- Scopel, E.F.C., Hose, J., Bensasson, D., and Gasch, A.P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics* 217, iyab015. <https://doi.org/10.1093/genetics/iyab015>.
- Chen, G., Bradford, W.D., Seidel, C.W., and Li, R. (2012). Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* 482, 246–250. <https://doi.org/10.1038/nature10795>.
- Lauer, S., Avecilla, G., Spealman, P., Sethia, G., Brandt, N., Levy, S.F., and Gresham, D. (2018). Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS Biol.* 16, e3000069. <https://doi.org/10.1371/journal.pbio.3000069>.
- Linder, R.A., Greco, J.P., Seidl, F., Matsui, T., and Ehrenreich, I.M. (2017). The Stress-Inducible Peroxidase TSA2 Underlies a Conditionally Beneficial Chromosomal Duplication in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 7, 3177–3184. <https://doi.org/10.1534/g3.117.300069>.

13. Millet, C., Ausiannikava, D., Le Bihan, T., Granneman, S., and Makovets, S. (2015). Cell populations can use aneuploidy to survive telomerase insufficiency. *Nat. Commun.* 6, 8664. <https://doi.org/10.1038/ncomms9664>.
14. Selmecki, A.M., Maruvka, Y.E., Richmond, P.A., Guillet, M., Shores, N., Sorenson, A.L., De, S., Kishony, R., Michor, F., Dowell, R., and Pellman, D. (2015). Polyploidy can drive rapid adaptation in yeast. *Nature* 519, 349–352. <https://doi.org/10.1038/nature14187>.
15. Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. USA* 109, 21010–21015. <https://doi.org/10.1073/pnas.1211150109>.
16. Lukow, D.A., and Sheltzer, J.M. (2022). Chromosomal instability and aneuploidy as causes of cancer drug resistance. *Trends Cancer* 8, 43–53. <https://doi.org/10.1016/j.trecan.2021.09.002>.
17. Ben-David, U., and Amon, A. (2020). Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* 21, 44–62. <https://doi.org/10.1038/s41576-019-0171-x>.
18. Girish, V., Lakhani, A.A., Thompson, S.L., Scaduto, C.M., Brown, L.M., Hagenson, R.A., Sausville, E.L., Mendelson, B.E., Kandikuppa, P.K., Lukow, D.A., et al. (2023). Oncogene-like addiction to aneuploidy in human cancers. *Science* 381, eadg4521. <https://doi.org/10.1126/science.adg4521>.
19. Huth, T., Dreher, E.C., Lemke, S., Fritzsche, S., Sugiyanto, R.N., Castven, D., Ibberson, D., Sticht, C., Eiteneuer, E., Jauch, A., et al. (2023). Chromosome 8p engineering reveals increased metastatic potential targetable by patient-specific synthetic lethality in liver cancer. *Sci. Adv.* 9, eadh1442. <https://doi.org/10.1126/sciadv.adh1442>.
20. Lukow, D.A., Sausville, E.L., Suri, P., Chunduri, N.K., Wieland, A., Leu, J., Smith, J.C., Girish, V., Kumar, A.A., Kendall, J., et al. (2021). Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev. Cell* 56, 2427–2439.e4. <https://doi.org/10.1016/j.devcel.2021.07.009>.
21. Su, X.A., Ma, D., Parsons, J.V., Replogle, J.M., Amatruda, J.F., Whitaker, C.A., Stegmaier, K., and Amon, A. (2021). RAD21 is a driver of chromosome 8 gain in Ewing sarcoma to mitigate replication stress. *Genes Dev.* 35, 556–572. <https://doi.org/10.1101/gad.345454.120>.
22. Dephore, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *Elife* 3, e03023. <https://doi.org/10.7554/eLife.03023>.
23. Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* 26, 2696–2708. <https://doi.org/10.1101/gad.207407.112>.
24. Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. *Science* 317, 916–924. <https://doi.org/10.1126/science.1142210>.
25. Bonney, M.E., Moriya, H., and Amon, A. (2015). Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. *Genes Dev.* 29, 898–903. <https://doi.org/10.1101/gad.261743.115>.
26. Krivega, M., and Storchova, Z. (2023). Consequences of trisomy syndromes – 21 and beyond. *Trends Genet.* 39, 172–174. <https://doi.org/10.1016/j.tig.2022.11.004>.
27. Larrimore, K.E., Barattin-Voyanova, N.S., Reid, D.W., and Ng, D.T.W. (2020). Aneuploidy-induced proteotoxic stress can be effectively tolerated without dosage compensation, genetic mutations, or stress responses. *BMC Biol.* 18, 117. <https://doi.org/10.1186/s12915-020-00852-x>.
28. Sheltzer, J.M., and Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet.* 27, 446–453. <https://doi.org/10.1016/j.tig.2011.07.003>.
29. Pompei, S., and Cosentino Lagomarsino, M. (2023). A fitness trade-off explains the early fate of yeast aneuploids with chromosome gains. *Proc. Natl. Acad. Sci. USA* 120, e2211687120. <https://doi.org/10.1073/pnas.2211687120>.
30. Zhu, Y.O., Sherlock, G., and Petrov, D.A. (2016). Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3 (Bethesda)* 6, 2421–2434. <https://doi.org/10.1534/g3.116.029397>.
31. Keller, A., Gao, L.L., Witten, D., and Dunham, M.J. (2023). Condition-dependent fitness effects of large synthetic chromosome amplifications. Preprint at bioRxiv. <https://doi.org/10.1101/2023.06.08.544269>.
32. Antonarakis, S.E., Skotko, B.G., Rafii, M.S., Strydom, A., Pape, S.E., Bianchi, D.W., Sherman, S.L., and Reeves, R.H. (2020). Down syndrome. *Nat. Rev. Dis. Primers* 6, 9. <https://doi.org/10.1038/s41572-019-0143-7>.
33. Lana-Elola, E., Watson-Scales, S.D., Fisher, E.M.C., and Tybulewicz, V.L.J. (2011). Down syndrome: searching for the genetic culprits. *Dis. Model. Mech.* 4, 586–595. <https://doi.org/10.1242/dmm.008078>.
34. Anders, K.R., Kudrna, J.R., Keller, K.E., Kinghorn, B., Miller, E.M., Pauw, D., Peck, A.T., Shelloe, C.E., and Strong, I.J.T. (2009). A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. *BMC Genet.* 10, 36. <https://doi.org/10.1186/1471-2156-10-36>.
35. Katz, W., Weinstein, B., and Solomon, F. (1990). Regulation of tubulin levels and microtubule assembly in *Saccharomyces cerevisiae*: consequences of altered tubulin gene copy number. *Mol. Cell Biol.* 10, 5286–5294.
36. Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative Haploinsufficiency and Triplo-sensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. *Cell* 155, 948–962. <https://doi.org/10.1016/j.cell.2013.10.011>.
37. Sack, L.M., Davoli, T., Li, M.Z., Li, Y., Xu, Q., Naxerova, K., Wooten, E.C., Bernardi, R.J., Martin, T.D., Chen, T., et al. (2018). Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* 173, 499–514.e23. <https://doi.org/10.1016/j.cell.2018.02.037>.
38. Solimini, N.L., Xu, Q., Mermel, C.H., Liang, A.C., Schlabach, M.R., Luo, J., Burrows, A.E., Anselmo, A.N., Bredemeyer, A.L., Li, M.Z., et al. (2012). Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential. *Science* 337, 104–109. <https://doi.org/10.1126/science.1219580>.
39. Veitia, R.A. (2002). Exploring the etiology of haploinsufficiency. *Bio-essays* 24, 175–184. <https://doi.org/10.1002/bies.10023>.
40. Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197. <https://doi.org/10.1038/nature01771>.
41. Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. USA* 109, 14746–14753. <https://doi.org/10.1073/pnas.1207726109>.
42. Santaguida, S., and Amon, A. (2015). Aneuploidy triggers a TFEB-mediated lysosomal stress response. *Autophagy* 11, 2383–2384. <https://doi.org/10.1080/15548627.2015.1110670>.
43. Tsai, H.-J., Nelli, A.R., Choudhury, M.I., Kuchavay, A., Bradford, W.D., Cook, M.E., Kim, J., Mair, D.B., Sun, S.X., Schatz, M.C., and Li, R. (2019). Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. *Nature* 570, 117–121. <https://doi.org/10.1038/s41586-019-1187-2>.
44. Donnelly, N., and Storchová, Z. (2015). Causes and consequences of protein folding stress in aneuploid cells. *Cell Cycle* 14, 495–501. <https://doi.org/10.1080/15384101.2015.1006043>.
45. Hose, J., Escalante, L.E., Clowers, K.J., Dutcher, H.A., Robinson, D., Bouriaikov, V., Coon, J.J., Shishkova, E., and Gasch, A.P. (2020). The

- genetic basis of aneuploidy tolerance in wild yeast. *Elife* 9, e52063. <https://doi.org/10.7554/eLife.52063>.
46. Gasch, A.P., Hose, J., Newton, M.A., Sardi, M., Yong, M., and Wang, Z. (2016). Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution. *Elife* 5, e14409. <https://doi.org/10.7554/eLife.14409>.
47. Hill, A., and Bloom, K. (1987). Genetic Manipulation of Centromere Function. *Mol. Cell Biol.* 7, 2397–2405.
48. Liu, H., Krizek, J., and Bretscher, A. (1992). Construction of a Gal1-Regulated Yeast Cdna Expression Library and Its Application to the Identification of Genes Whose Overexpression Causes Lethality in Yeast. *Genetics* 132, 665–673.
49. Weinstein, B., and Solomon, F. (1990). Phenotypic consequences of tubulin overproduction in *Saccharomyces cerevisiae*: differences between alpha-tubulin and beta-tubulin. *Mol. Cell Biol.* 10, 5295–5304.
50. Ho, C.H., Magtanong, L., Barker, S.L., Gresham, D., Nishimura, S., Natarajan, P., Koh, J.L.Y., Porter, J., Gray, C.A., Andersen, R.J., et al. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat. Biotechnol.* 27, 369–377. <https://doi.org/10.1038/nbt.1534>.
51. Bach, F. (2008). Bolasso: model consistent Lasso estimation through the bootstrap. Preprint at arXiv. <https://doi.org/10.48550/arXiv.0804.1302>.
52. Bacherle, J.-P., Cavallé, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* 84, 775–790. [https://doi.org/10.1016/S0300-9084\(02\)01402-5](https://doi.org/10.1016/S0300-9084(02)01402-5).
53. Schindler, D., Walker, R.S.K., Jiang, S., Brooks, A.N., Wang, Y., Müller, C.A., Cockram, C., Luo, Y., García, A., Schraivogel, D., et al. (2022). Design, Construction, and Functional Characterization of a tRNA Neochromosome in Yeast. Preprint at bioRxiv. <https://doi.org/10.1101/2022.10.03.510608>.
54. Pluta, K., Lefebvre, O., Martin, N.C., Smagowicz, W.J., Stanford, D.R., Ellis, S.R., Hopper, A.K., Sentenac, A., and Boguta, M. (2001). Maf1p, a Negative Effector of RNA Polymerase III in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 21, 5031–5040. <https://doi.org/10.1128/MCB.21.15.5031-5040.2001>.
55. Ascencio, D., Diss, G., Gagnon-Arsenault, I., Dubé, A.K., DeLuna, A., and Landry, C.R. (2021). Expression attenuation as a mechanism of robustness against gene duplication. *Proc. Natl. Acad. Sci. USA* 118, e2014345118. <https://doi.org/10.1073/pnas.2014345118>.
56. Douglas, A.C., Smith, A.M., Sharifpoor, S., Yan, Z., Durbic, T., Heisler, L.E., Lee, A.Y., Ryan, O., Göttert, H., Surendra, A., et al. (2012). Functional Analysis With a Barcoder Yeast Gene Overexpression System. *G3 (Bethesda)* 2, 1279–1289. <https://doi.org/10.1534/g3.112.003400>.
57. Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., et al. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* 19, 2816–2826. <https://doi.org/10.1101/gad.1362105>.
58. Morrill, S.A., and Amon, A. (2019). Why haploinsufficiency persists. *Proc. Natl. Acad. Sci. USA* 116, 11866–11871. <https://doi.org/10.1073/pnas.1900437116>.
59. Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping Pathways and Phenotypes by Systematic Gene Overexpression. *Mol. Cell* 21, 319–330. <https://doi.org/10.1016/j.molcel.2005.12.011>.
60. Diamant, A., Feldman, A., Schochet, E., Kupiec, M., Arava, Y., and Tuller, T. (2018). The extent of ribosome queuing in budding yeast. *PLoS Comput. Biol.* 14, e1005951. <https://doi.org/10.1371/journal.pcbi.1005951>.
61. Robinson, D., Place, M., Hose, J., Jochem, A., and Gasch, A.P. (2021). Natural variation in the consequences of gene overexpression and its implications for evolutionary trajectories. *Elife* 10, e70564. <https://doi.org/10.7554/eLife.70564>.
62. Shen, Y., Gao, F., Wang, Y., Wang, Y., Zheng, J., Gong, J., Zhang, J., Luo, Z., Schindler, D., Deng, Y., et al. (2023). Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast. *Cell Genom.* 3, 100364. <https://doi.org/10.1016/j.xgen.2023.100364>.
63. Lana-Elola, E., Watson-Scales, S., Slender, A., Gibbins, D., Martineau, A., Douglas, C., Mohun, T., Fisher, E.M., and Tybulewicz, V.L. (2016). Genetic dissection of Down syndrome-associated congenital heart defects using a new mouse mapping panel. *Elife* 5, e11614. <https://doi.org/10.7554/eLife.11614>.
64. Lyle, R., Béna, F., Gagos, S., Gehrig, C., Lopez, G., Schinzel, A., Lespinasse, J., Bottani, A., Dahoun, S., Taine, L., et al. (2009). Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. *Eur. J. Hum. Genet.* 17, 454–466. <https://doi.org/10.1038/ejhg.2008.214>.
65. Robinson, D., Vanacloig-Pedros, E., Cai, R., Place, M., Hose, J., and Gasch, A.P. (2023). Gene-by-environment interactions influence the fitness cost of gene copy-number variation in yeast. *G3 (Bethesda)* 13, jkad159. <https://doi.org/10.1093/g3journal/jkad159>.
66. Buchhaupt, M., Sharma, S., Kellner, S., Oswald, S., Paetzold, M., Peifer, C., Watzinger, P., Schrader, J., Helm, M., and Entian, K.-D. (2014). Partial Methylation at Am100 in 18S rRNA of Baker's Yeast Reveals Ribosome Heterogeneity on the Level of Eukaryotic rRNA Modification. *PLoS One* 9, e89640. <https://doi.org/10.1371/journal.pone.0089640>.
67. Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. <https://doi.org/10.1080/15476286.2016.1259781>.
68. Gay, D.M., Lund, A.H., and Jansson, M.D. (2022). Translational control through ribosome heterogeneity and functional specialization. *Trends Biochem. Sci.* 47, 66–81. <https://doi.org/10.1016/j.tibs.2021.07.001>.
69. Zhang, X., Wang, C., Xia, S., Xiao, F., Peng, J., Gao, Y., Yu, F., Wang, C., and Chen, X. (2023). The emerging role of snoRNAs in human disease. *Genes Dis.* 10, 2064–2081. <https://doi.org/10.1016/j.gendis.2022.11.018>.
70. Su, Z., Wilson, B., Kumar, P., and Dutta, A. (2020). Noncanonical Roles of tRNAs: tRNA Fragments and Beyond. *Annu. Rev. Genet.* 54, 47–69. <https://doi.org/10.1146/annurev-genet-022620-101840>.
71. Percudani, R., Pavesi, A., and Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae* Edited by J. Karn. *J. Mol. Biol.* 268, 322–330. <https://doi.org/10.1006/jmbi.1997.0942>.
72. Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M.R., and Pan, T. (2009). tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.* 37, 7268–7280. <https://doi.org/10.1093/nar/gkp787>.
73. Pinzaru, A.M., and Tavazoie, S.F. (2023). Transfer RNAs as dynamic and critical regulators of cancer progression. *Nat. Rev. Cancer* 23, 746–761. <https://doi.org/10.1038/s41568-023-00611-4>.
74. Santos, M., Fidalgo, A., Varanda, A.S., Oliveira, C., and Santos, M.A.S. (2019). tRNA Deregulation and Its Consequences in Cancer. *Trends Mol. Med.* 25, 853–865. <https://doi.org/10.1016/j.molmed.2019.05.011>.
75. Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation. *Cell* 158, 1281–1292. <https://doi.org/10.1016/j.cell.2014.08.011>.
76. Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H., and Tavazoie, S.F. (2016). Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* 165, 1416–1427. <https://doi.org/10.1016/j.cell.2016.05.046>.

77. Hu, G., Luo, S., Rao, H., Cheng, H., and Gan, X. (2018). A Simple PCR-based Strategy for the Introduction of Point Mutations in the Yeast *Saccharomyces cerevisiae* via CRISPR/Cas9. *Biochem Mol Biol J* 04. *Biochem. Mol. Biol. J.* 4, 9. <https://doi.org/10.21767/2471-8084.100058>.
78. Jansen, J.M., Wanless, A.G., Seidel, C.W., and Weiss, E.L. (2009). Cbk1 Regulation of the RNA-Binding Protein Ssd1 Integrates Cell Fate with Translational Control. *Curr. Biol.* 19, 2114–2120. <https://doi.org/10.1016/j.cub.2009.10.071>.
79. Wanless, A.G., Lin, Y., and Weiss, E.L. (2014). Cell morphogenesis proteins are translationally controlled through UTRs by the Ndr/LATS target Ssd1. *PLoS One* 9, e85212. <https://doi.org/10.1371/journal.pone.0085212>.
80. Khonsari, B., Klassen, R., and Schaffrath, R. (2021). Role of SSD1 in Phenotypic Variation of *Saccharomyces cerevisiae* Strains Lacking DEG1-Dependent Pseudouridylation. *Int. J. Mol. Sci.* 22, 8753. <https://doi.org/10.3390/ijms22168753>.
81. Xu, F., Byström, A.S., and Johansson, M.J.O. (2019). SSD1 suppresses phenotypes induced by the lack of Elongator-dependent tRNA modifications. *PLoS Genet.* 15, e1008117. <https://doi.org/10.1371/journal.pgen.1008117>.
82. Dutcher, H.A., Hose, J.A., Howe, H., Rojas, J., and Gasch, A. (2024). The response to single-gene duplication implicates translation as a key vulnerability in aneuploid yeast. Preprint at bioRxiv. <https://doi.org/10.1101/2024.04.15.589582>.
83. Escalante, L.E., and Gasch, A.P. (2021). The role of stress-activated RNA-protein granules in surviving adversity. *RNA* 27, 753–762. <https://doi.org/10.1261/rna.078738.121>.
84. Muenzner, J., Trébulle, P., Agostini, F., Zauber, H., Messner, C.B., Steger, M., Kilian, C., Lau, K., Barthel, N., Lehmann, A., et al. (2024). Natural proteome diversity links aneuploidy tolerance to protein turnover. *Nature* 630, 149–157. <https://doi.org/10.1038/s41586-024-07442-9>.
85. Liu, Y., Borel, C., Li, L., Müller, T., Williams, E.G., Germain, P.-L., Buljan, M., Sajic, T., Boersema, P.J., Shao, W., et al. (2017). Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nat. Commun.* 8, 1212. <https://doi.org/10.1038/s41467-017-01422-6>.
86. Semple, J.I., Vavouri, T., and Lehner, B. (2008). A simple principle concerning the robustness of protein complex activity to changes in gene expression. *BMC Syst. Biol.* 2, 1. <https://doi.org/10.1186/1752-0509-2-1>.
87. Veitia, R.A., Bottani, S., and Birchler, J.A. (2008). Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 24, 390–397. <https://doi.org/10.1016/j.tig.2008.05.005>.
88. Chen, Y., Chen, S., Li, K., Zhang, Y., Huang, X., Li, T., Wu, S., Wang, Y., Carey, L.B., and Qian, W. (2019). Overdosage of Balanced Protein Complexes Reduces Proliferation Rate in Aneuploid Cells. *Cell Syst.* 9, 129–142.e5. <https://doi.org/10.1016/j.cels.2019.06.007>.
89. Geiger, T., Cox, J., and Mann, M. (2010). Proteomic Changes Resulting from Gene Copy Number Variations in Cancer Cells. *PLoS Genet.* 6, e1001090. <https://doi.org/10.1371/journal.pgen.1001090>.
90. Jüschke, C., Dohnal, I., Pichler, P., Harzer, H., Swart, R., Ammerer, G., Mechtler, K., and Knoblich, J.A. (2013). Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome Biol.* 14, r133. <https://doi.org/10.1186/gb-2013-14-11-r133>.
91. Stingeles, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* 8, 608. <https://doi.org/10.1038/msb.2012.40>.
92. McShane, E., Sin, C., Zauber, H., Wells, J.N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J.A., et al. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* 167, 803–815.e21. <https://doi.org/10.1016/j.cell.2016.09.015>.
93. Thorburn, R.R., Gonzalez, C., Brar, G.A., Christen, S., Carlile, T.M., Ingolia, N.T., Sauer, U., Weissman, J.S., and Amon, A. (2013). Aneuploid yeast strains exhibit defects in cell growth and passage through START. *MBoC* 24, 1274–1289. <https://doi.org/10.1091/mbc.e12-07-0520>.
94. Sunshine, A.B., Ong, G.T., Nickerson, D.P., Carr, D., Murakami, C.J., Wasko, B.M., Shemorry, A., Merz, A.J., Kaeberlein, M., and Dunham, M.J. (2016). Aneuploidy shortens replicative lifespan in *Saccharomyces cerevisiae*. *Aging Cell* 15, 317–324. <https://doi.org/10.1111/ace1.12443>.
95. Ni, Z., Zhou, X.-Y., Aslam, S., and Niu, D.-K. (2019). Characterization of Human Dosage-Sensitive Transcription Factor Genes. *Front. Genet.* 10, 1208.
96. Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 100, 3889–3894. <https://doi.org/10.1073/pnas.0635171100>.
97. MacKay, V.L., Li, X., Flory, M.R., Turcott, E., Law, G.L., Serikawa, K.A., Xu, X.L., Lee, H., Goodlett, D.R., Aebersold, R., et al. (2004). Gene Expression Analyzed by High-resolution State Array Analysis and Quantitative Proteomics: Response of Yeast to Mating Pheromone. *Mol. Cell. Proteomics* 3, 478–489. <https://doi.org/10.1074/mcp.M300129-MCP200>.
98. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 324, 218–223. <https://doi.org/10.1126/science.1168978>.
99. Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. *PLoS Biol.* 7, e1000238. <https://doi.org/10.1371/journal.pbio.1000238>.
100. Laccina, J.R., LaMonte, G., Nicchitta, C.V., and Chi, J.-T. (2011). Poly-some profiling of the malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 179, 42–46. <https://doi.org/10.1016/j.molbiopara.2011.05.003>.
101. Fernandes, L.D., Moura, A.P.S.d., and Ciandrini, L. (2017). Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Sci. Rep.* 7, 17409. <https://doi.org/10.1038/s41598-017-17618-1>.
102. Guo, J., Lian, X., Zhong, J., Wang, T., and Zhang, G. (2015). Length-dependent translation initiation benefits the functional proteome of human cells. *Mol. Biosyst.* 11, 370–378. <https://doi.org/10.1039/C4MB00462K>.
103. Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107. <https://doi.org/10.1038/nature09322>.
104. Kurland, C.G. (1992). Translational accuracy and the fitness of bacteria. *Annu. Rev. Genet.* 26, 29–50. <https://doi.org/10.1146/annurev.ge.26.120192.000333>.
105. Zhang, G., Fedyunin, I., Miekley, O., Valleriani, A., Moura, A., and Ignatova, Z. (2010). Global and local depletion of ternary complex limits translational elongation. *Nucleic Acids Res.* 38, 4778–4787. <https://doi.org/10.1093/nar/gkq196>.
106. Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 16, 274–280. <https://doi.org/10.1038/nsmb.1554>.
107. Aliouat, A., Hatini, I., Bertin, P., François, P., Stierlé, V., Namy, O., Salhi, S., and Jean-Jean, O. (2020). Divergent effects of translation termination factor eRF3A and nonsense-mediated mRNA decay factor UPF1 on the expression of uORF carrying mRNAs and ribosome protein genes. *RNA Biol.* 17, 227–239. <https://doi.org/10.1080/15476286.2019.1674595>.
108. Decourty, L., Doyen, A., Malabat, C., Frachon, E., Rispal, D., Séraphin, B., Feuerbach, F., Jacquier, A., and Saveanu, C. (2014). Long open

- p>reading frame transcripts escape nonsense-mediated mRNA decay in yeast.
- Cell Rep.*
- 6, 593–598.
- <https://doi.org/10.1016/j.celrep.2014.01.025>
- .
109. Dodgson, S.E., Santaguida, S., Kim, S., Sheltzer, J., and Amon, A. (2016). The pleiotropic deubiquitinase Ubp3 confers aneuploidy tolerance. *Genes Dev.* 30, 2259–2271. <https://doi.org/10.1101/gad.287474.116>.
 110. Torres, E.M., Dephoure, N., Panneerselvam, A., Tucker, C.M., Whittaker, C.A., Gygi, S.P., Dunham, M.J., and Amon, A. (2010). Identification of Aneuploidy-Tolerating Mutations. *Cell* 143, 71–83. <https://doi.org/10.1016/j.cell.2010.08.038>.
 111. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30, 187–200. <https://doi.org/10.1002/pro.3978>.
 112. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. <https://doi.org/10.1093/nar/gkac1000>.
 113. Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res.* 37, 825–831. <https://doi.org/10.1093/nar/gkn1005>.
 114. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. <https://doi.org/10.1093/nar/gkac1052>.
 115. Badaczewska-Dawid, A.E., Kuriata, A., Pintado-Grima, C., Garcia-Pardo, J., Burdukiewicz, M., Iglesias, V., Kmiecik, S., and Ventura, S. (2024). A3D Model Organism Database (A3D-MODB): a database for proteome aggregation predictions in model organisms. *Nucleic Acids Res.* 52, D360–D367. <https://doi.org/10.1093/nar/gkad942>.
 116. Zhao, B., Katuwawala, A., Oldfield, C.J., Dunker, A.K., Faraggi, E., Gsponer, J., Kloczkowski, A., Malhis, N., Mirdita, M., Obradovic, Z., et al. (2021). DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Res.* 49, D298–D308. <https://doi.org/10.1093/nar/gkaa931>.
 117. Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. *Genetics* 169, 1915–1925. <https://doi.org/10.1534/genetics.104.036871>.
 118. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
 119. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654. <https://doi.org/10.1093/molbev/msab199>.
 120. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21, 245. <https://doi.org/10.1186/s13059-020-02134-9>.
 121. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
 122. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ -mer weighting and repeat separation. *Genome Res.* 27, 722–736. <https://doi.org/10.1101/gr.215087.116>.
 123. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics* 37, 1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>.
 124. Riehl, K., Riccio, C., Miska, E.A., and Hemberg, M. (2022). TransposonUltimate: software for transposon classification, annotation and detection. *Nucleic Acids Res.* 50, e64. <https://doi.org/10.1093/nar/gkac136>.
 125. Fang, Z., Liu, X., and Peltz, G. (2023). GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39, btac757. <https://doi.org/10.1093/bioinformatics/btac757>.
 126. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
 127. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
 128. Erdős, G., Pajkos, M., and Dosztányi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* 49, W297–W303. <https://doi.org/10.1093/nar/gkab408>.
 129. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
 130. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
 131. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705. <https://doi.org/10.1093/nar/gkr1029>.
 132. Magtanong, L., Ho, C.H., Barker, S.L., Jiao, W., Baryshnikova, A., Bahr, S., Smith, A.M., Heisler, L.E., Choy, J.S., Kuzmin, E., et al. (2011). Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nat. Biotechnol.* 29, 505–511. <https://doi.org/10.1038/nbt.1855>.
 133. Piotrowski, J.S., Simpkins, S.W., Li, S.C., Deshpande, R., McIlwain, S.J., Ong, I.M., Myers, C.L., Boone, C., and Andersen, R.J. (2015). Chemical Genomic Profiling via Barcode Sequencing to Predict Compound Mode of Action. In *Chemical Biology: Methods and Protocols* Methods in Molecular Biology, J.E. Hempel, C.H. Williams, and C.C. Hong, eds. (Springer), pp. 299–318. https://doi.org/10.1007/978-1-4939-2269-7_23.
 134. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>.
 135. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 136. Bach, F.R. (2008). Bolasso: model consistent Lasso estimation through the bootstrap. In *Proceedings of the 25th international conference on Machine learning - ICML '08* (ACM Press), pp. 33–40. <https://doi.org/10.1145/1390156.1390161>.
 137. Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* 137, 146–158. <https://doi.org/10.1016/j.cell.2009.02.044>.

138. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. <https://doi.org/10.1038/75556>.
139. Gene Ontology Consortium; Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. *Genetics* 224, iyad031. <https://doi.org/10.1093/genetics/iyad031>.
140. Kuleshov, M.V., Diaz, J.E.L., Flamholz, Z.N., Keenan, A.B., Lachmann, A., Wojciechowicz, M.L., Cagan, R.L., and Ma'ayan, A. (2019). modEnrichr: a suite of gene set enrichment analysis tools for model organisms. *Nucleic Acids Res.* 47, W183–W190. <https://doi.org/10.1093/nar/gkz347>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
DH5 alpha + pJH1	Hose et al. ⁴⁵	AGB90
DH5 alpha + pJR1	This paper	AGB350
DH5 alpha + pJR2	This paper	AGB360
DH5 alpha + pJR3	This paper	AGB361
Chemicals, peptides, and recombinant proteins		
Yeast extract	Fisher/BD	DF0127179 and 212750
Peptone	Fisher/BD	DF0118-17-0 and 211677
D Raffinose	Fisher	10279780
Galactose	Fisher	BP656-500
Nourseothricin-dihydrogen sulfate (clonNAT)	Werner BioAgents	5002000
G-418 disulfate	Fisher	bp673-1
Yeast Synthetic Drop-Out Media w/o Histidine (SC-His)	Sigma	Y2001-20G
Zymolyase	Zymo Research	E1004 or E1005
Critical commercial assays		
DNeasy Blood and Tissue Kit	Qiagen	69581
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®	NEB	E7103
MiSeq	Illumina	
Zymoprep Yeast Plasmid Miniprep II	Zymo Research	D2004-A
AxyPrep Mag PCR Clean-Up Kit	Fisher	nc9959336
Deposited data		
YPS1009-derivative strain genome assembly in NCBI, BioProject	This paper	PRJNA984736
gene duplication screen raw sequencing and barcode counts, on GEO	This paper	GSE263221
Biogrid	Oughtred et al. ¹¹¹	http://thebiogrid.org/
STRING database	Szklarczyk et al. ¹¹²	https://string-db.org/
YPS1009 Euploid Proteomics data	Hose et al. ⁴⁵	https://cdn.elifesciences.org/articles/52063/elifesciences-suppl3-v2.xlsx
Yeast ribosome profiling dataset	Diament et al. ⁶⁰	https://doi.org/10.1371/journal.pcbi.1005951.s002
CYC2008_complex	Pu et al. ¹¹³	https://wodaklab.org/cyc2008/resources/CYC2008_complex.tab
Uniprot	The UniProt Consortium ¹¹⁴	https://www.uniprot.org/
A3D Database	Badaczewska-Dawid et al. ¹¹⁵	https://biocomp.chem.uw.edu.pl/A3D2/yeast
DescribeProt database	Zhao et al. ¹¹⁶	http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/main.php
Haploinsufficient genes	Deutschbauer et al. ¹¹⁷	https://yeastmine.yeastgenome.org/yeastmine/report.do?id=996557687
Experimental models: Organisms/strains		
S.cerevisiae, wild-type yeast strain YPS1009	Hose et al., 2020 ⁴⁵	N/A
All other YPS1009 strains used in this paper are listed in Table S3	See Table S3	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Yeast ORF library (MoBY 1.0)	Ho et al. ⁵⁰	https://horizondiscovery.com/en/non-mammalian-research-tools/products/molecular-barcoded-yeast-moby-orf-library
Software and algorithms		
MUMmer version 4.0.0 rc1	Kurtz et al. ¹¹⁸	https://github.com/mummer4/mummer , RRID:SCR_018171
BUSCO version 5.4.4	Manni et al. ¹¹⁹	https://busco.ezlab.org/ , RRID:SCR_015008
Merqury and Meryl	Rhie et al. ¹²⁰	https://github.com/marbl/merqury , RRID:SCR_022964
Pilon version 1.23	Walker et al. ¹²¹	https://github.com/broadinstitute/pilon/releases , RRID:SCR_014731
Guppy version 6.2.1 (ONT)	Oxford Nanopore Technology	https://nanoporetech.com/
Canu version 1.9	Koren et al. ¹²²	https://github.com/marbl/canu
Liftoff	Shumate et al. ¹²³	https://github.com/agshumate/Liftoff
ReasonaTE	Riehl et al. ¹²⁴	https://github.com/DerKevinRiehl/TransposonUltimate
Python 3	Python™	https://www.python.org/ , RRID:SCR_008394
GSAEpy version 1.0.6	Fang et al. ¹²⁵	https://gseapy.readthedocs.io/
EdgeR version 3.36.0	Robinson et al. ¹²⁶	https://bioconductor.org/ , RRID:SCR_012802
Statsmodels, version 0.13.5	N/A	https://www.statsmodels.org/v0.13.5/ , RRID:SCR_016074
Sciklearn version 1.3.0	Pedregosa et al. ¹²⁷	https://scikit-learn.org/stable/index.html , RRID:SCR_019053
IUPRED3	Erdos et al. ¹²⁸	https://iupred3.elte.hu/
AlphaFold	Jumper et al. ¹²⁹	https://alphafold.ebi.ac.uk/ , RRID:SCR_023662
SnEff version 5.0	Cingolani et al. ¹³⁰	https://pcingola.github.io/SnpEff/
Other		
Saccharomyces Genome Database (SGD)	SGD community ¹³¹	https://yeastgenome.org/
Code repository for aneuploidy model and gene machine learning model	This paper	https://doi.org/10.5281/zenodo.12701832

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Strains and plasmid

Strains and plasmids used are listed in the Table S4. YPS1009 aneuploids were generated using the methods of Hill and Bloom⁴⁷ except Chr12 aneuploidy described in Hose et al.⁴⁵ Briefly, a DNA cassette including the *GAL1-10* promoter (*GAL1* oriented toward the centromere), *HphMX6* gene for hygromycin resistance, and terminator *P_{TDH3}-GFP-T_{CYC1}* (except for Chr3, 9, and 16 where GFP was omitted) was integrated at 60 bp from each centromere of interest and selected on hygromycin medium. Each resulting euploid strain was grown for 16 h in YP (1% yeast extract and 2% peptone) medium with 2% raffinose and switched to YP with 2% galactose for one doubling based on optical density, and then plated for single colonies. For transformants carrying the GFP cassette, colonies were initially screened for 1X (euploid) versus 2X (aneuploid) GFP fluorescence on a flow cytometer, and colonies with 2X fluorescence were selected. Aneuploid colonies were selected via qPCR of genes on and off the amplified chromosome to confirm duplication of the amplified chromosome; selected colonies used in this study were confirmed by low-coverage whole genome sequencing, confirming that genes spanning the entire chromosome were present on average 2X higher copy than genes on all other chromosomes. *ssd1Δ* aneuploids were obtained by crossing aneuploids selected above to the euploid *ssd1Δ* and selecting resulting *ssd1Δ* aneuploid clones. YPS1009 with a duplication of Chr6 could not be generated in YPS1009, and duplication of Chr16 in *ssd1Δ* produced very sick colonies that could not be cultivated. Genomic DNA was isolated with the DNeasy Blood and Tissue Kit modified for yeast (Qiagen) and sequenced using the NEBNext Ultra II DNA Library Prep Kit on the Illumina MiSeq. Nine of the aneuploids (Chr1, 4, 5, 7, 10, 13, 14, 15 and 16) were backcrossed to remove the centromere-proximal cassette. Euploids and aneuploids with the cassette had no difference in growth rate compared to an isogenic strain without the cassette, confirming that the cassette does not influence fitness.

The pJR1 plasmid expressing 7 C/D box snoRNAs encoded on Chr13 (*snr72*, *snr73*, *snr74*, *snr75*, *snr76*, *snr77*, *snr78*) was obtained by amplifying 2017 bp containing the polycistronic C/D snoRNAs region from Chr13 (coordinates 280,245–282,261 from the YPS1009 genome assembly) and ligating it into pJH1 plasmid. The pJR2 plasmid containing 7 H/ACA snoRNAs was obtained by ligating a fragment containing *SNR36*, *SNR8*, *SNR31*, *SNR5*, *SNR81*, *SNR9* (synthesized by Twist Bioscience) and *SNR35* (amplified from YPS1009) into pJH1. A fragment from the Yce1313 plasmid (shared by the Cai Lab) containing all Chr12 tRNAs was cloned

into pJH1 to obtain the pJR3 plasmid. The amplified snoRNA and tRNA were picked for ease of cloning, all plasmids were verified by Sanger sequencing. MAF1 was deleted by homologous recombination of the *HphMX6* cassette and verified by diagnostic PCR; aneuploid strains were generated by crossing the euploid *maf1*Δ to aneuploids.

METHOD DETAILS

Growth conditions

Strain passaging was minimized to ensure maintenance of the aneuploidies. Freshly streaked colonies were used to inoculate liquid YPD and cultured for ~1 generation to allow cells to exit lag phase before changes in optical density (OD₆₀₀) were scored every ~15–20 min for ~140 min. Care was taken to ensure that strains did not pass OD₆₀₀ of 0.8, ensuring they were far from the diauxic shift in this media. An exponential curve was then fit to calculate growth rates; most R₂ were above 99% (except for some slow-growing strains that had a R₂ ~0.98) indicating exponential growth. The growth curves, growth rates and OD are available in Table S1. The maintenance of aneuploidy was periodically checked through diagnostic qPCR of one or two genes on the amplified chromosome normalized to a single-copy gene elsewhere in the genome (*ERV25* or *ACT1*), taking ~2X higher copy of the amplified genes to confirm aneuploidy. Detectable loss of the extra chromosome at the culture level was rarely observed, but cultures for which >20% of final colonies reverted to euploidy were excluded from analysis. Significant differences in observed versus expected growth rate were assessed with replicate-paired t tests. Unless otherwise noted, all studies used 4 biological replicates.

For strains transformed with plasmids (pJH1, pJR1, pJR2, pJR3), cells were cultured for 2 h in YPD +100ug/ml nourseothricin media then shifted to YPD without antibiotics and grown for another hour before OD₆₀₀ measurements were collected for growth rates. The biological replicates represent the growth of at least two different transformants, transformed on different days.

YPS1009 genome sequencing

A highly contiguous assembly of YPS1009 strain AGY731 was prepared through a hybrid approach of Oxford Nanopore (ONT, Oxford, UK) and Illumina (San Diego, California) sequencing. High molecular weight DNA was prepared for ONT sequencing by harvesting cells from an overnight YPD culture, spheroplasting, and gently lysing cells followed by phenol:chloroform extraction and ethanol precipitation of DNA. The preparation was enriched for high molecular weight DNA >1.5 kb by bead cleanup using a custom buffer (10mM Tris-HCl, 1mM EDTA pH 8.0, 1.6M NaCl, 11% PEG8000). DNA was prepared for sequencing using sequencing kit LSK-110 (ONT) and sequenced on a single flongle flow cell (ONT). ONT sequencing produced 175 Mb resulting in ~14x coverage of the yeast reference genome. Initial base calling was done using guppy v.6.2.1 (ONT) retaining reads with Q > 7. The initial assembly was done using ONT reads with Canu v.1.9.¹²² This assembly was polished using Illumina data pooled from 32,723,650 reads of all YPS1009 aneuploid strains (211X YPS1009 genome coverage) using pilon v.1.23 iteratively three times.¹²¹

The assembly resulted in 23 contigs with sizes ranging from 1,061 to 1,482,091 bp of which 11,353,357 bp had homology to the S288c genome. Each of the 23 contigs was aligned to the S288c chromosome to which it had shown maximal homology using MUMmer,¹¹⁸ with -c parameter set for each chromosome based on aligning the S288c chromosome sequence to the S288c reference genome, to minimize short off-target alignments. Four chromosomes (Chr7,12,13,16) were spanned by two contigs and one (Chr15) was spanned by 4 contigs. To evaluate alignment gaps on those chromosomes, we considered Illumina DNA read coverage from the aneuploid YPS1009 strain in which that chromosome was duplicated. We did not find support for the S288c sequence being present in YPS1009 at any of these gaps, strongly suggesting that the S288c sequence in those gaps is truly missing from YPS1009. Contigs for these chromosomes were joined by {N}₁₀ representing those gaps. The final assembly resulted in 16 assembled chromosomes.

We assessed the quality of the assembly in several ways. First, the median percent identity for MUMmer-aligned segments was 99.25%, showing high similarity to the S288c genome as expected. Second, we considered the coverage of known universal single-copy orthologs from the OrthoDB database BUSCO.¹¹⁹ BUSCO analysis identified 99.2% (2119 out of 2137) of the universal single-copy genes from the saccharomycetes_odb10 ortholog database, of which 2074 were in single copy and 45 were duplicated, indicating high coverage of expected genes. Base-level accuracy and completeness were measured with Merqury.¹²⁰ An optimal k-mer size (16) was generated using best_k.sh (provided by Merqury suite) and a k-mer database created with Meryl.¹²⁰ This k-mer database was used to evaluate the assembly, which returned a completeness score of 99.502%.

Finally, we annotated the gene content using LiftOff.¹²³ Multiple genes and other genomic elements with high level of homology were annotated to the same region, we filtered out the annotation with the lowest homology. LiftOff identified 6,552 genes, 277 tRNAs, 77 snoRNAs, 21 ncRNA, and 354 ARS in the YPS1009 genome. For transposable elements (TE), we combined LiftOff identification with ReasonaTE,¹²⁴ and collapsed TEs that were mapped to the same region. There are 23 retrotransposons containing functional GAG-POL open reading frames. Among the 6552 genes annotated by LiftOff, 57 are missing a start codon, 331 are missing a stop codon, and 70 have an in-frame stop codon.

89 genes from S288C were missing in YPS1009: 48 of them were mapped to other ORFs and filtered out, which likely correspond to genes present in multiple copies in S288C. The remaining 41 missed genes were used as BLAST queries to the YPS1009 assembly: 4 small genes aligned to multiple loci (>8) in the YPS1009 assembly while 38 genes were not identified by LiftOff or BLAST of the YPS1009 contigs; the position of 19 of these genes in the S288C genome reside in 3 suspected gaps between YPS1009 contigs that were supported by the absence of Illumina reads mapping to those YPS1009 regions as described above. 12 genes mapped

to a gap on Chr12 that was corroborated by an absence of Illumina reads. Thus, the draft assembly of the YPS1009 genome is close to complete, barring small-scale errors whose correction is beyond the scope of this study, and is available on BioProject (accession number: PRJNA984736).

Gene duplication fitness cost measurements

The euploid YPS1009 strain (AGY1611) was transformed with a pool of the molecular barcoded yeast ORF library (MoBY 1.0) containing 5,037 barcoded CEN plasmids.⁵⁰ At least 25,000 transformants were scraped from agar plates per transformation, and frozen glycerol stocks were made. Three independent transformations of the pooled library were performed, for a total of ~52X coverage of the library. Competitive growth was done in liquid synthetic media lacking histidine (SC-His) and with 100 mg/L nourseothricin and 200 mg/L G418 to maintain the plasmids. Competition experiments were performed as previously described.^{50,61,132,133} Briefly, 1 mL frozen glycerol stocks of library-transformed cells were thawed into 100 mL of liquid medium at a starting OD₆₀₀ of 0.05, then grown in shake flasks at 30°C with shaking. The remaining cells from the frozen stocks were pelleted by centrifugation and represented the starting pool (generation 0) for each strain. After five generations, each pooled culture was diluted to an OD₆₀₀ of 0.05 in fresh media, to maintain cells in log phase. Cells were harvested and stored at –80°C after 10 generations. 7 biological replicates from 3 independent library transformations were collected and analyzed. Plasmids were recovered from each pool using ZymoPrep Yeast Plasmid Miniprep II (Zymo Research D2004-A) with the following modifications: samples were incubated with 15 units zymolyase at 37°C for 1 h, with inversion every 15 min; incubation in cell lysis buffer was extended to 10 min; after neutralization, samples were put on ice for 30 min, then centrifuged at 4°C. Plasmid barcodes were amplified using primers containing Illumina multiplex adaptors as described in.¹³³ The number of PCR cycles was reduced to 20. Barcode amplicons were pooled and purified using AxyPrep Mag beads (1.8X volume beads per sample volume) according to the manufacturer's instructions. Pooled amplicons were sequenced on one lane of an Illumina HiSeq 4000 to generate single-end 50 bp reads. The screen raw sequencing and barcode counts were deposited on GEO (GSE263221). The data analysis was performed as follows: barcodes with no valid values and the bottom 5% of barcodes based on read abundance at generation 0 were removed from the total counts. Generation 0 includes other strains not exploited in here but in Dutcher et al.⁸² A pseudo-count of 1 was added to each gene in every sample in the dataset. Barcode counts were normalized using the TMM method¹³⁴ and analyzed in EdgeR¹²⁶ version 3.36.0 using a gene-wise negative binomial generalized linear model with quasi-likelihood tests. Results were similar when normalized by total reads per sample. Significant differences between experiment endpoint and generation 0 were defined as those with FDR < 0.05 using the Benjamini-Hochberg procedure for multiple test correction.¹³⁵ Fitness scores of 4,462 genes were calculated as the log₂ of the ratio of normalized reads after 10 generations divided by reads at generation 0 (Source data S6 for Figure 2; Table S2). Significant fitness scores are highly correlated with those from comparable YPS1009 Moby 1.0 library grown in YPD medium ($R^2 = 0.8$), but not with YPS1009 transformed with the Moby 2.0 library grown under similar conditions as used here,⁶¹ confirming that media differences between this study and Robinson et al. do not explain modeling differences.

Modeling aneuploidy fitness costs

Model 1 fits the measured growth rates (4 per strain) for each aneuploid relative to euploid cells as a function of the sum number of verified and uncharacterized genes per chromosome, according to the YPS1009 genome annotation. A total of 4,369 measured genes are mapped to the YPS1009 genome and included for further analyses. We did not consider dubious genes. Linear regression was performed using the ordinary least square (OLS) method (Statsmodels, version 0.13.5). All codes and models were written in Python 3 and are available (<https://doi.org/10.5281/zenodo.12701832>).

Model 2 fit measured growth rates described above as a function of the measured fitness costs for genes duplicated on each chromosome as follows. For measured genes that were statistically significant (FDR < 0.05), the fitness cost was taken as the fitness scores described above. Genes with missing values (848 genes) or that were not statistically different from neutral (FDR > 0.05) were scored with the mean log₂ fitness score across all measured genes = –0.33. For 624 genes that are in the collection but were not detected in our experiment, we assumed their fitness cost was too toxic to make it to the starting pool in this strain background and thus imputed values with the 2.5% lower quantile value of all genes = –3.2. Each chromosome cost was estimated based on the sum of these log₂ values for genes on that chromosome. The linear fit was calculated as described for Model 1. The improvement of Model 2 compared to Model 1 was estimated in two ways. First, we used a nested model and Chi-square test, considering the contribution of Model 1 (gene number) plus the contribution of Model 2 costs normalized to each chromosome's gene number, then fitted in an OLS model. We then perform a likelihood-ratio test (Chi-Square test, degree of freedom = 1) to show that both features are significant (number of genes/Chromosome p value: 1.2×10^{-13} , normalized Chr. cost p value: 0.045). Second, we performed 10,000 random permutations of gene fitness cost labels across chromosomes, while preserving the number of genes per chromosome in each trial and summed the permuted Chr. costs. We then fitted the aneuploid relative growth rate against every permuted Chr. cost iteration and compared the R^2 values to Model 2 R^2 . Out of 10,000 permutations, only 4 met the observed Model2 fit for wild-type aneuploids ($p = 0.0004$) and none for the *ssd1Δ* strains. The importance of beneficial genes was estimated by summing detrimental/neutral genes and beneficial genes separately and fitting a multifactorial linear regression. A Chi-square test showed that both features are significantly contributing to the fit.

Model 3 was assessed by first compiling a list of non-genic features from the YPS1009 Liftoff feature detection (Table S5) and normalized to the total number of features per chromosome to prevent high correlations in between features. Features were

selected using a bootstrap-Lasso approach¹³⁶: 10000 random subsets of 60 relative growth measurements were fitted using Lasso ($\alpha = 0.7$), and features that had a non-zero coefficient for 90% or more iteration were incorporated into a multi-linear regression model (OLS) to get model performance.

Deleterious gene duplications classifier

Gene biophysical features considered in the modeling are described in Table S3 and are available together with the gene duplication fitness costs in Table S2. These features regroup datasets from several publications,^{45,60,113,115,117,137} public databases (SGD, UniProt, Biogrid, STRING, A3D, Gene Ontology and DescribeProt),^{111,112,114–116,131,138,139} and prediction software (IUPRED3, AlphaFold, SnpEff).^{128–130} Functional enrichments using GSAEpy python library¹²⁵ (version 1.0.6) and the ontologies from Yeast modEnrich¹⁴⁰ were performed in 2 ways. First, we performed a hypergeometric test to compare genes whose duplication was deleterious (FDR < 0.05) versus the background genes set (all barcoded genes with a measured logFC). Second, we used a GSEA rank test: genes were ranked on their $\log_2$ fitness scores * $\log_{10}$ (FDR) values. Enrichments with an adjusted p value < 0.05 were included as categorical features for the modeling and are available in Table S3. For numerical features, a Wilcoxon rank test was performed with Benjamini-Hochberg correction.¹³⁵ To train the gene classifier to predict deleterious genes, we reduced the number of features to only those that were significant (adjusted p value < 0.05) and removed features that were highly correlated (Spearman correlation > 0.70, see Figure S5A), keeping the feature most strongly distinguishing detrimental genes (Figure S5A). All models were trained and tested using a stratified 5-fold cross-validation approach: for 5 iterations, the dataset was randomly split into training and test sets while maintaining the proportion of deleterious and neutral genes. We then computed the mean and standard deviation receiver-operator curves and area under the curve (AUC-ROC) for analysis of the test set. Confusion matrices also were computed from the aggregated test set predictions. We used a seed of 17 for the k-fold splitting and all models. The following model and parameters from Sklearn¹²⁷ (version 1.3.0) were used: Logistic regression with l2 penalty (maximum iteration = 500, solver = newton-cholesky, and balanced class weight), Random Forest classifier (n estimators = 100, minimum sample per leaf = 24, max depth = 8, minimum impurity decrease = 0.01), XGBoost Classifier (number of estimators = 100, minimum child weight = 250, subsample = 0.8, maximum depth = 4, balanced weight (0.7)), Gradient Boosting Classifier (number of estimators = 100, subsample = 0.8, minimum impurity decrease = 4, maximum depth = 6). Parameters were manually selected to reduce overfitting; Overfitting was assessed by comparing the ROC-AUC for the training and testing sets.

Models were first trained on the whole gene fitness screen from which genes with more than 6 missing biophysical features were removed (1,177 detrimental genes and 3,028 neutral/beneficial genes remaining). Due to poor predictions on the whole dataset, we focused on training binary classifiers to distinguish between medium-highly detrimental genes ($\log_2$ fitness score < -1.54 (quantile = 0.15) and FDR < 0.05 = 613 genes (29%)) and neutral genes ($\log_2$ fitness score > 0.27 (quantile 0.65), 1,472 genes). The logistic regression classifier performed better than tree classifiers or Neural networks. Features were sorted by their mean coefficients (Figure S5B) and we observed that the 12 top features were sufficient to maintain maximal model performance with an AUC-ROC of 0.713. Features importance was assessed using a permutation feature importance strategy (Sklearn version 1.3.0, permutation_importance)¹²⁷: each feature is randomly shuffled and the resulting degradation of the model's score is used to compare features. Values were shuffled 10 times for each 5-fold validation dataset splitting. Feature coefficients were analyzed to assess if a feature was associated with detrimental genes or with the neutral group.

A similar classifier (Logistic regression with l2 penalty, maximum iteration = 500, solver = newton-cholesky, balanced class weight) was trained on Robinson et al. data to discriminate the commonly deleterious gene overexpression (400 genes, detrimental at FDR < 0.05 in at least 10 yeast isolates) from commonly neutral or beneficial gene overexpression (1,657 genes, not detrimental (FDR > 0.05) in at least 12 yeast isolates) measured in our lab under slightly different growth conditions.⁶¹ In that case, no features were filtered out based on correlation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses are described in the result section and figure legends. In Figure 1A significant difference of growth rates between euploid versus aneuploid and WT versus *ssd1Δ* for each aneuploid was tested using a replicate-paired, 2-sided, t test. The significance threshold was p value < 0.05. The original OD measurements used to compute the growth rates, growth rates and R_2 are available in Table S1. All linear regressions were fitted from the datasets available in the supplemental source table (Source data S6 - Figures 1, 2, and 3). For aneuploids with more than 4 measured growth rates, we randomly selected 4 relative growth values (seed = 0, see code) so that all aneuploids have the same weight. All linear regression reports the adjusted r-square in the figures. The effect of gene duplications measured using MoBy 1.0 plasmid library was tested using gene-wise negative binomial generalized linear model with quasi-likelihood tests (as described in the methods above). Beneficial genes are defined as genes with an FDR < 0.05 and a positive log(fold-change) between generation 0 and 10 while detrimental genes have a negative logFC and FDR < 0.05. Comparisons between linear regression models were tested using a Chi-square test as specified in the result section.