Main Manuscript for

Genetic architecture of a mutation's expressivity

Rachel Schell1, Joseph J. Hale1, Martin N. Mullis1, Takeshi Matsui1,2, Ryan Foree1, and Ian M. Ehrenreich1,*

1 Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

2 Present address: Joint Initiative for Metrology in Biology, Stanford University, Stanford, CA 94305, USA

*Correspondence to: Ian M. Ehrenreich

Email: ian.ehrenreich@usc.edu

Author Contributions: R.S. and I.M.E. conceptualized this project. R.S., J.J.H., M.N.M., T.M., and R.F. performed the experiments. R.S. and I.M.E. analyzed the data. R.S. and I.M.E. wrote the paper. R.S., J.J.H., M.N.M., T.M., R.F., and I.M.E. edited the manuscript.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences. Genetics.

Keywords: genetic background, background effects, spontaneous mutation, expressivity, complex traits, genetic architecture

This PDF file includes:

Main Text
Figures 1 to 5

Abstract

Genetic background often influences the phenotypic consequences of mutations, resulting in variable expressivity. How standing genetic variants collectively cause this phenomenon is not fully understood. Here, we comprehensively identify loci in a budding yeast cross that impact the growth of individuals carrying a spontaneous missense mutation in the nuclear-encoded mitochondrial ribosomal gene *MRP20*. Initial results suggested that a single large effect locus influences the mutation's expressivity, with one allele causing inviability in mutants. However, further dissection revealed that in fact many additional loci shape the mutation's expressivity, collectively leading to a wide spectrum of mutational responses. These results exemplify how complex combinations of loci can produce a diversity of qualitative and quantitative responses to the same mutation.

Significance Statement

Mutations can show different phenotypic effects depending on the individuals in which they occur (or background effects). These background effects are important to understand because they impact the contributions of mutations to traits of biomedical, evolutionary, and agricultural significance. While it is known that a main cause of background effects is genetic interactions between mutations and pre-existing genetic variants, the architectures of these interactions are not well characterized. Here, we use show that a large number of genetic variants influence the effect of a spontaneous mutation in a cross of two yeast strains. These loci individually tend to have small impacts, but collectively they cause the mutation to show a range of effects that extends from lethal to near zero.

Main Text

Introduction

Mutations frequently exhibit different effects in genetically distinct individuals (or 'background effects') (1-3). For example, not all people with the same disease-causing mutations manifest the associated disorder or exhibit identical symptoms. A commonly observed form of background effect among individuals carrying the same mutation is different degrees of response to that mutation (or 'variable expressivity') (4). Variable expressivity can arise due to a myriad of reasons, including genetic interactions (or epistasis) between a mutation and segregating loci (1), dominance (1), stochastic noise (5), the microbiome (6), and the environment (1).

The genetic architecture underlying expressivity has proven difficult to study, in part because natural populations harbor substantial genetic diversity, which can facilitate complex forms of epistasis between segregating loci and mutations (7-20). Mapping the loci involved in these genetic interactions is technically challenging. However, the introduction of mutations into diverse strains and crosses in the laboratory can enable the identification and genetic dissection of such background effects (7, 12, 17-19, 21).

In this paper, we use a series of controlled crosses in the budding yeast *Saccharomyces cerevisiae* to comprehensively characterize the genetic basis of a mutation's expressivity. We focus on a missense mutation in *MRP20*, an essential nuclear-encoded subunit of the mitochondrial ribosome (22). This mutation occurred by chance in a cross between the reference strain BY4716 ('BY') and a clinical isolate 322134S ('3S'), and was found to show variable expressivity among BYx3S cross progeny. This presented an opportunity to determine how loci segregating in the BYx3S cross individually and collectively influence this mutation's expressivity.

Results

A spontaneous mutation increases phenotypic variance in the BYx3S cross

In the BY/3S diploid progenitor of haploid BYx3S segregants, a spontaneous mutation occurred in a core domain of Mrp20 that is conserved from bacteria to humans (Fig. 1A, fig. S1, table S1) (22, 23). This mutation resulted in an alanine to glutamine substitution at amino acid 105 (*mrp20-105E*) and showed variable expressivity among segregants carrying it. Specifically, segregants with this mutation showed increased phenotypic variance relative to wild type segregants when ethanol was provided as the carbon source, the condition used hereafter (Fig. 1B; Levene's test, $p = 5.9 \times 10^{-22}$). Mutant segregants exhibited levels of growth ranging from inviable to wild type, and fit a bimodal distribution that centered on 10% and 57% growth relative to the haploid BY parent strain (bimodal fit log likelihood 30; fig. S2).

A large effect locus shows epistasis with *mrp20-105E*

Loci contributing to this variable expressivity should be detectable through their genetic interactions with *MRP20*. To find such loci, we performed linkage scans for two-way epistasis with *MRP20*. We identified a single locus on Chromosome XIV (ANOVA, interaction term $p = 4.3 \times 10^{-16}$; Fig. 2A). Individuals with XIV^{BY} showed reduced growth among both *MRP20* and *mrp20-105E* segregants, but to a greater degree among the latter (Fig. 2B). The Chromosome XIV locus explained 79% of the phenotypic variance among *mrp20-105E* segregants (ANOVA, $p = 3.2 \times 10^{-31}$) and accounted for all observed cases of inviability (Fig. 2B).

To further resolve the Chromosome XIV locus, we crossed an *mrp20-105E* XIV^{BY} F_2 segregant and an *mrp20-105E* XIV^{3S} F_2 segregant (supplementary text 2; Fig. 2C, table S1). 361 F_3 progeny were

genotyped by low-coverage whole genome sequencing and phenotyped for growth. Linkage mapping with these data reidentified the Chromosome XIV locus at a p-value of 2.50 x 10^{-43} (ANOVA; Fig. 2D, fig. S3) and resolved it to a single SNP in the coding region of *MKT1* (Fig. 2E). This SNP, which encodes a glycine in BY and a serine in 3S at amino acid 30, was then validated by nucleotide replacement in *mrp20-105E* segregants (Fig. 2F). Notably, this specific SNP was previously shown to play a role in mitochondrial genome stability (24), suggesting epistasis between *MRP20* and *MKT1* involves mitochondrial dysfunction, impairing growth on non-fermentative carbon sources such as ethanol.

Epistasis between *MRP20* and *MKT1* differs in cross parents and segregants

We attempted to validate the epistasis between *MRP20* and *MKT1* by introducing all four possible combinations of the causal nucleotides at these two genes into haploid versions of both BY and 3S (Fig. 3A). The *mrp20-105E* mutation affected growth in both parent strains (ANOVA, p = 4.3 x 10^{-24} and p = 4.0 x 10^{-4}). However, the magnitude of the effect differed between the two: *mrp20-105E* caused inviability in BY but had a more modest effect in 3S. In addition, *MKT1* influenced response to *mrp20-105E* in the 3S background (ANOVA, p = 0.01) but not in the BY background (ANOVA, p = 0 .99). Furthermore, the phenotypic consequence of epistasis between *MRP20* and *MKT1* differed between parent and segregant strains. Specifically, the phenotypes of BY *mrp20-105E* $MKT1^{3S}$, 3S *mrp20-105E* $MKT1^{BY}$, and 3S *mrp20-105E* $MKT1^{3S}$ all differed from the expectations established by BYx3S *mrp20-105E* segregants. These departures from expectation imply that additional unidentified loci also influence response to *mrp20-105E*.

Fixation of *mrp20-105E* and *MKT1* genotypes further expands phenotypic variance

To enable the identification of other loci underlying response to *mrp20-105E*, we generated two new BYx3S crosses (Fig. 3B, table S1). In both crosses, the BY and 3S parents were engineered to carry *mrp20-105E*. Further, one cross was engineered so that both parents carried $MKT1^{BY}$ and the other cross was engineered so that both parents carried $MKT1^{3S}$. By altering the parent strains in this manner, we increased the chance of detecting additional loci contributing to the variable expressivity of *mrp20-105E*. From these engineered crosses, 749 total segregants were obtained through tetrad dissection, genotyped by low-coverage genome sequencing, and phenotyped for growth on ethanol.

The new crosses exhibited continuous ranges of phenotypes, in contrast with the bimodal phenotypic distribution observed in the original segregants (Fig. 3C). In both the $MKT1^{BY}$ and $MKT1^{3S}$ crosses, *mrp20-105E* segregants ranged from inviable to nearly wild type. The distributions of phenotypes in the two crosses differed in a manner consistent with their *MKT1* alleles, with the mean of the $MKT1^{BY}$ segregants lower than the $MKT1^{3S}$ segregants (t-test, p = 4.8 x 10^{-34}). These data show that regardless of the *MKT1* allele present, additional loci can cause *mrp20-105E* to show phenotypic effects ranging from lethal to benign.

Many additional loci affect the expressivity of *mrp20-105E*

We used the new crosses to map other loci contributing to response to *mrp20-105E*. Excluding *MKT1*, which explained 18% of the phenotypic variance in the new crosses, linkage mapping identified 16 new loci (Fig. 3D, fig. S4, and table S2). We found no evidence for genetic interactions among the loci (pairs and trios examined with fixed effects linear models, Bonferroni threshold). Of these loci, the BY allele was inferior at 10 and superior at six. These loci individually explained between 0.79% and 14% of the phenotypic variance. 13 of these loci resided on a subset of chromosomes but were distantly linked: four on Chromosomes XII, three on XIII, two on XIV, and four on XV. The three remaining loci were detected on Chromosomes IV, VII, and XI. Recombination breakpoints delimited these loci to small genomic intervals spanning one (12 loci), two (3 loci) or three (1 locus) genes (table S3). These candidate genes functioned in many

compartments of the cell and implicated a diversity of cellular pathways and processes (table S4). Thus, the molecular basis of *mrp20-105E*'s expressivity is complex and cannot be ascribed to a single coherent mechanism.

The Chromosome XIV locus contains multiple causal variants

Among the newly detected loci, the largest effect (14% phenotypic variance explained) was on Chromosome XIV. The position of maximal significance at this site was located two genes away from the end of *MKT1*, with a 99% confidence interval that did not encompass the causal variant in *MKT1* (table S3). Thus, the originally identified large effect Chromosome XIV locus in fact represents multiple distinct closely linked nucleotides that both genetically interact with *MRP20* and occur in different genes (Fig. 3E). The new locus on Chromosome XIV was delimited to two genes, one of which was *SAL1*, encoding a mitochondrial ADP/ATP transporter that physically interacts with Mrp20. A SNP in *SAL1* that segregates in this cross was previously linked to increased mitochondrial genome instability in BY (24), suggesting it is likely also causal in our study. For this reason, we refer to this additional Chromosome XIV locus as *'SAL1'*. We find no evidence for epistasis between *MKT1* and *SAL1* (ANOVA, p = 0.77).

Although the *MKT1-SAL1* locus had a large effect, it explained a minority of the phenotypic variance among *mrp20-105E* segregants in a model including all detected loci (32% for *MKT1-SAL1* vs. 36% for all other loci collectively). Thus, by enabling *MKT1* and *SAL1* to segregate independently through genetic engineering and examining a large number of *mrp20-105E* segregants with different *MKT1-SAL1* genotypes, we observed a greater diversity of mutational responses than was originally seen and detected many additional loci.

Aneuploidy also contributes to the variable expressivity of *mrp20-105E*

Despite the fact that the identified loci explain most of *mrp20-105E*'s expressivity, some individuals exhibited unexpectedly poor growth (Fig. 3F). This finding led to the identification of a Chromosome II duplication that reduced growth (ANOVA, 1.2×10^{-48}). The aneuploidy was common among *mrp20-105E* segregants, with a higher prevalence when $MKT1^{3S}$ was also present (Fisher's exact test, $p = 1.5 \times 10^{-43}$; table S5). The Chromosome II aneuploidy was not seen among wild type segregants. These data suggest that *mrp20-105E* increases the rate of aneuploidization and that genetic variation in *MKT1* influences the degree to which *mrp20-105E* segregants duplicate Chromosome II. The aneuploidy's contribution to phenotypic variation was relatively minor, explaining 5% of phenotypic variance among *mrp20-105E* segregants in a model also including all identified loci.

Multiple mechanisms underlie poor growth in the presence of *mrp20-105E*

Evidence suggests mitochondrial genome instability contributes to the variable expressivity of *mrp20-105E*. First, mitochondrial genome instability is known to cause poor growth on non-fermentative carbon sources, such as ethanol (25, 26). Second, the exact variants that segregate in our cross at *MKT1* and *SAL1* were previously linked to mitochondrial genome instability (24). Third, both Mrp20 and Sal1 function in the mitochondria (22, 27). Fourth, two other candidate genes in the newly detected loci encode proteins that function in the mitochondria (table S4).

To determine the role of mitochondrial genome instability in the variable expressivity of *mrp20-105E*, we measured petite formation, a proxy for spontaneous mitochondrial genome loss (Fig. 4) (28). In addition to *MRP20* and *mrp20-105E* BY and 3S parent strains, 16 *MRP20* segregants and 42 *mrp20-105E* segregants were examined. Despite causing reduced growth in both parents, *mrp20-105E* only led to elevated mitochondrial genome instability in BY (t-test p = 0.013 in BY and p = 0.39 in 3S; Fig. 4A). This result indicates that mitochondrial genome instability explains part, but not all, of response to *mrp20-105E*. Supporting this point, *mrp20-105E* segregants exhibited

increased mitochondrial genome instability relative to *MRP20* segregants (Wilcoxon rank-sum test $p = 0.023$), especially at lower levels of growth, but a subset of inviable segregants did not show elevated petite formation (Fig. 4B and C). These results suggest that *mrp20-105E* can lead to poor growth through mitochondrial genome instability, but this is not the only mechanism determining response to *mrp20-105E*.

Genetic underpinnings of *mrp20-105E*'s expressivity in segregants and parents

We determined the extent to which our identified loci explained phenotypic variability among mutants. Modeling growth as a function of all identified loci and the aneuploidy accounted for the majority (78%) of the broad-sense heritability among *mrp20-105E* segregants (ANOVA, $p = 5.2 \times 10^{-188}$). Further, phenotypic predictions for segregants based on their genotypes were strongly correlated with their observed phenotypes ($r = 0.85$, $p = 4.4 \times 10^{-209}$; Fig. 5A).

The model was also effective for other genotypes that were not present in the new crosses, but had been generated throughout the course of this work. For instance, the model accurately predicted the phenotypes of the original *mrp20-105E* segregant population ($r = 0.90$, $p = 1.6 \times 10^{-39}$), as well as the phenotypes of cross parents engineered to carry *mrp20-105E* (Fig. 5A). Moreover, the model explained both qualitative and quantitative variation within and between the two Chromosome XIV classes that were originally seen among *mrp20-105E* segregants. These results confirm that the variable expressivity of *mrp20-105E* is driven by many loci that collectively produce a spectrum of mutational responses.

Finally, we examined how diverse combinations of loci collectively produced similar phenotypic responses to *mrp20-105E*. We examined the relationship between growth and the total number of detrimental alleles carried by *mrp20-105E* segregants, keeping track of each individual's genotype at *MKT1* and *SAL1*, the largest effect loci (Fig. 5B). The number of detrimental alleles carried by a segregant showed a strong negative relationship with growth, which was not observed in wild type segregants (fig. S5). Further, regardless of genotype at *MKT1* and *SAL1*, the effect of *mrp20-105E* ranged from lethal to benign in a manner dependent on the number of detrimental alleles present at other loci. These findings demonstrate that many segregating loci beyond the large effect *MKT1-SAL1* locus influence the expressivity of *mrp20-105E* and enable different genotypes in the cross to exhibit a broad range of responses to the mutation.

Discussion

We have provided a detailed genetic characterization of the expressivity of a spontaneous mutation. Response to this mutation in a budding yeast cross is influenced by at least 18 genetic factors in total, with the largest effect due to two closely linked variants. However, at least 15 additional loci segregate and jointly exert larger effects than the largest two. Different combinations of alleles across these loci produce a continuous spectrum of mutational responses. Due to tight linkage between *MKT1* and *SAL1* in the original cross parents, the full extent of this continuum was not originally observed, leading to an initial understanding of the expressivity of the *mrp20-105E* mutation that was simplistic.

These findings also show how quantitative variation in mutational response can produce seemingly discrete outcomes. In part, whether responses appear qualitative depends on the configuration of mutationally responsive alleles in examined mutants. Approaches such as crossing of genetically engineered strains can be used to disrupt these configurations that mask the full extent of variation. However, another part of this expressivity is the tolerance of a system to quantitative variation in key processes, for example mitochondrial genome stability in the case of *mrp20-105E*. Our data suggest that these processes can only tolerate quantitative variation to a point, but also indicate

that lethality to the same mutation may arise in different genetic backgrounds due to impairment of distinct cellular processes.

Our results inform efforts to understand expressivity in other systems, including humans. For example, there is interest in determining why people who carry highly penetrant alleles known to cause disease do not develop pathological conditions (3, 29, 30). Such resilience, as observed here, may involve numerous loci. This speaks to the complicated and unexpected epistasis that can arise between mutations and segregating loci in genetically diverse populations (7-20). It also illustrates the importance of characterizing epistasis (31-41), including background effects, as these forms of genetic interactions are immediately relevant to evolution and disease, and may not emerge from studies that do not directly interrogate natural variation in genetically diverse populations.

Materials and Methods

Generation of segregants

The haploid BYx3S segregants in which *mrp20-105E* was identified were the *hos3*Δ F_2 segregants generated and described in Mullis et al (17). In brief, a BY *MAT**a** can1*Δ::*STE2pr-SpHIS5 his3*Δ strain was mated to a 3S *MATα ho*::*HphMX his3*Δ*::NatMX* strain to generate a wild type BY/3S diploid. PCR-mediated, targeted gene disruption was then used to produce a BY/3S *HOS3/hos3*Δ*::KanMX* strain. Both the wild type and hemizygous deletion strains were sporulated, and random BYx3S *MAT**a*** spores were obtained from each using the magic marker system with plating on His- plates containing canavanine (42). Following discovery of the *mrp20-105E* mutation, we performed tetrad dissected of this diploid to obtain *mrp20-105E* segregants in both *HOS3* and *hos3*Δ genetic backgrounds.

To produce haploid *mrp20-105E* F_3 segregants, we deleted *URA3* from a BYx3S F_2 *MAT**a** can1*Δ*::STE2pr-SpHIS5 his3*Δ *hos3*Δ*::KanMX mrp20-105E XIV*^{3S} segregant. We then mating type switched the strain by transforming it with a *URA3* plasmid containing an inducible *HO* endonuclease, inducing *HO*, and plating single cells. The mating type-switched BYx3S F_2 *MATα can1*Δ*::STE2pr-SpHIS5 his3*Δ *ura3*Δ *hos3*Δ*::KanMX mrp20-105E XIV*^{3S} segregant was then mated to a BYx3S F_2 *MAT**a** can1*Δ*::STE2pr-SpHIS5 his3*Δ *hos3*Δ*::KanMX mrp20-105E XIV*^{BY} segregant. The resulting diploid was sporulated and random segregants were obtained by plating on His- media.

To obtain additional haploid *mrp20-105E MKT1*^{BY} and *MKT1*^{3S} F_2 segregants, we engineered *mrp20-105E*, as well as the 3S and BY causal variants at *MKT1* position 467,219 into BY and 3S, respectively. BY *mrp20-105E* was independently mated to 3S *mrp20-105E MKT1*^{BY} twice. Two resultant diploids were sporulated and tetrads were dissected to obtain BYx3S *mrp20-105E MKT1*^{BY} haploid segregants. The same process was followed with BY mrp20-105E *MKT1*^{3S} and 3S *mrp20-105E* strains to obtain BYx3S *mrp20-105E MKT1*^{3S} haploid segregants.

Genotyping

F_2 segregants shown in Fig 1 and Fig 2 A-B were previously genotyped in Mullis et al. using the same techniques described below (17). In this paper, F_3 segregants and all remaining F_2 segregants shown in Figs 1 and 3-5 were genotyped by low coverage whole genome sequencing. Freezer stocks of strains were inoculated into liquid overnight cultures and grown to stationary phase at 30°C. DNA was extracted using Qiagen 96-well DNeasy kits (Qiagen P/N 69581). Sequencing libraries were prepared using the Illumina Nextera Kit and custom barcoded adapter sequences. Segregants from each respective cross (361 F_3s and 872 F_2s) were pooled in equimolar fractions into three separate multiplexes, run on a gel, size selected, and purified with the Qiagen Gel Extraction Kit. F_2 and F_3 segregants were sequenced by Novogene on Illumina HiSeq 4000 lanes using 150 bp x 150 bp paired-end reads.

Sequencing reads were mapped against the S288C genome (version S288C_reference_sequence_R64-2-1_20150113.fsa from the Saccharomyces Genome Database https://www.yeastgenome.org) using BWA version 0.7.7-r44 (43). Samtools v1.9 was then used to create a pileup file for each segregant (44). For both BWA and Samtools, default settings were employed. Base calls and coverages were gathered for 44,429 SNPs that segregate in the cross (14). Low coverage individuals (<0.7x average per site coverage) were removed from analyses. Diploid and contaminated individuals were identified by abnormal patterns of heterozygosity or sequencing coverage, and were also excluded. For each segregant, a raw genotype vector was determined by the percent of calls at each site for the 3S allele. We then used a Hidden Markov Model (HMM) implemented in the 'HMM' package v 1.0 in R to correct each raw genotype vector using the following probability matrices (45):
transitionProbabilitiy = matrix(c(.9999,.0001,.0001,.9999),2)
and
emissionProbability = matrix(c(.0.25,0.75,0.75,0.25),2).

Aneuploidies were identified based on elevated sequencing coverage at particular chromosomes within each individual sample. This identified a chromosome II duplication event in a subset of BYx3S *mrp20-105E MKT1^{BY}* and BYx3S *mrp20-105E MKT1^{3S}* segregants. The BY *mrp20-105E MKT1^{3S}* x 3S *mrp20-105E* cross had the highest prevalence (50%), and thus individuals from this cross were further examined. We employed the normalmixEM() function from the mixtools library in R (46) to determine that coverage on Chr II was bimodal and centered on 0.98 and 1.8 (log likelihood of 237). Posterior probabilities were used to call aneuploid individuals which that had an average per site coverage of 1.5x or greater. This threshold was also applied to other crosses to identify aneuploid individuals.

Phenotyping
Segregants were inoculated into rich media containing glucose ('YPD'), which was comprised of 1% yeast extract (BD P/N 212750), 2% peptone (BD P/N: 211677), and 2% dextrose (BD P/N 15530). Cultures were grown to stationary phase (two days at 30°C). Strains were then pinned onto YP + 2% agar (BD P/N 214050) rich media containing ethanol ('YPE'). The YPE recipe was 1% yeast extract (BD P/N 212750), 2% peptone (BD P/N: 211677), and 2% ethanol (Koptec P/N A06141602W). Plates were then grown at 30°C for two days. Growth assays were conducted in a minimum of three replicates across three plates. On each plate, a BY control was included. Plates were imaged with the BioRAD Gel Doc XR+ Molecular Imager at a standard size of 11.4 x 8.52 cm^2 (width x length) and imaged with epi-illumination using an exposure time of 0.5 seconds. Images were saved as 600 dpi tiffs. ImageJ (http://rsbweb.nih.gov/ij/) was used to quantify pixel intensity of each colony through the Plate Analysis JRU v1 plugin (https://research.stowers.org/imagejplugins/zipped_plugins.html), as described in Matsui et al. (47). Growth values were normalized against the same plate BY control, then averaged across replicates to produce a single growth value for each segregant.

Linkage mapping
Initial linkage mapping was conducted with F_2 segregants. Initial discovery of the spontaneous *mrp20-105E* mutation resulted from linkage mapping with 385 F_2 segregants (164 wild type and 221 *hos3*Δ) from Mullis et al. (17). We employed the linear model *growth ~ hos3*Δ *+ locus + hos3*Δ*x locus + error*, from which the *hos3*Δ*x locus* interaction term was used to identify loci that differentially explained growth in *hos3*Δ segregants. Examination of the *hos3*Δ*x locus* interaction term led to discovery of the spontaneous *mrp20-105E* mutation on *the MRP20^{BY}* allele present in *hos3*Δ segregants. Following discovery of *mrp20-105E*, we used the fixed effects linear model *growth ~ MRP20 + locus + MRP20 x locus + error* using only *hos3*Δ individuals from Mullis et al. (17). From this scan, we examined the *MRP20* x *locus* interaction term. 361 *mrp20-105E* F_3 segregants were used to better resolve the Chromosome XIV locus. We employed the model *growth ~ locus + error* and examined the *locus* term. We examined the minimum observed test on chromosome XIV to delimit that locus.

To find loci affecting growth in the *mrp20-105E* background, we generated new populations of *mrp20-105E MKT1*^{BY} (353) and *mrp20-105E MKT1*^{3S} (396) haploid segregants. The combined 749 *mrp20-105E* segregants were used for linkage mapping that followed a forward regression approach. We first obtained residuals from the linear model *growth ~ MKT1 + error*, and then implemented a genome-wide scan using the model *residuals ~ locus + error*. We examined the locus term and significance was determined by using 1,000 permutations with the threshold set at the 95th quantile of observed -$\log_{10}$(p-values) (48). A maximum of one locus per chromosome per scan was identified as significant. Following the identification of additional loci, we accounted for the newly detected loci in a new model, *residuals ~ locus 1 + locus 2 + … locus n + error* and obtained the residuals. These new residuals were used in another genome-wide scan using the model *residuals ~ locus + error*. Permuted thresholds were calculated for each scan. This process was repeated for a total of 5 iterations at which point no loci were detected above our significance threshold. Chromosome II was excluded from linkage mapping due to the presence of a chromosomal duplication in a subset of individuals. The Chromosome II duplication was tested for significance using the model *growth ~ MKT1 + ChromosomeII + error,* from which the Chromosome II term was examined.

All linkage mapping was performed in R. Linear models were implemented using the lm() function. To call peaks for each scan we required that the local minimum position within each peak be a minimum of 150,000 kb away from any other peak. We also required peaks to be more than 20kB from the edge of a chromosome. We report 99% confidence intervals as 2-lod intervals surrounding the peak position at each locus.

Classification of inviable segregants

Initial discovery of the *MRP20 x MKT1* genetic interaction suggested that expressivity of *mrp20-105E* was largely determined by variation at *MKT1*. Furthermore, *mrp20-105E MKT1*^{BY} segregants exhibited very poor growth, while *mrp20-105E MKT1*^{3S} segregants showed more tolerant, variable growth. We termed this initial *mrp20-105E MKT1*^{BY} segregant population as 'inviable'. Figures 4 and 5 include a gray dashed line to denote the highest growth value observed among the original inviable segregants.

Delimiting loci with recombination breakpoints

For each locus examined, we split the appropriate segregants into two groups: individuals carrying the BY allele and individuals carrying the 3S allele. Segregants' haplotypes across the adjacent genomic window were then examined. The causal region was determined by identifying the SNPs fixed for BY among all BY individuals and fixed for 3S among all 3S individuals. Raw Illumina sequencing reads were examined to confirm the delimit of *IV* to *MRP20* among original F_2 segregants, the delimit of *XIV* to the *MKT1* coding SNP at 467,219 among F_3 segregants, and the delimit of the secondary *XIV* locus to *SAL1* and *PMS1* among the new F_2 segregants.

Reciprocal hemizygosity experiments

Four *hos3*Δ F_2 *MAT**a*** segregants were used in all reciprocal hemizygosity (RH) experiments (49): two were *hos3*Δ *IV*^{BY} *XIV*^{BY} and two were *hos3*Δ *IV*^{3S} *XIV*^{BY}. The four segregants were first mating type switched to enable mating of these segregants to produce homozygous *IV*^{BY}/*IV*^{BY}, homozygous *IV*^{3S}/*IV*^{3S}, or heterozygous *IV*^{BY}/*IV*^{3S} diploids. Each pairwise mating was performed and confirmed by plating on mating type tester plates. These diploid strains were then phenotyped on agar plates containing ethanol, which verified that *IV*^{BY} has an effect in diploids and acts in a recessive manner. Using the haploid *MAT**a*** and *MATα* versions of these four segregants, we individually engineered premature stop codons into *DIT1*, MRP20 and *PDR15* using CRISPR-mediated targeted gene disruption and lithium acetate transformations (50). Plasmid-based CRISPR-Cas9 was employed to target the beginning of each coding region and 20bp repair templates which contained a premature stop codon followed by 1bp deletions were incorporated. Each sgRNA and repair template was designed so that only the first 15 (of 537), 26 (of 264), and

33 (of 1,530) amino acids would be translated for *DIT1, MRP20* and *PDR15*, respectively. Engineered strains were confirmed by PCR and Sanger sequencing. After confirmation, wild type and knockout strains for each gene were then mated in particular combinations to produce reciprocal hemizygotes that were otherwise isogenic. A minimum of two distinct hemizygotes were generated for each allele of each gene.

Construction of nucleotide replacement strains

Single nucleotide replacement strains were generated for *MRP20* and *MKT1* using a CRISPR/Cas9-mediated 10pproach. For a given replacement, the appropriate strain was first transformed with a modified version of pML104 that constitutively expresses Cas9 using LiAc transformation (50, 51). We then inserted the KanMX gene using co-transformation of a double-stranded DNA containing KanMX with 30bp upstream and 30bp downstream homology tails and gRNAs targeting the region containing the site of interest (52). DNA oligos and PCR were used to construct custom sgDNA templates which included crRNA and tracrRNA in a single molecule. Next, we employed T7 RNA Polymerase to express sgDNA templates *in vitro*. Dnase treatment and phenol extraction were used to obtain purified sgRNAs. Transformants were selected on media containing G418, and KanMX integration was confirmed by PCR. Next, KanMX was replaced with the nucleotide of interest. To do this, integrants were co-transformed with four gRNAs targeting KanMX, a 60 bp single-stranded DNA repair oligo, and a marker plasmid expressing either HygMX or NatMX using electroporation (53). Marker plasmids were constructed by Gibson assembly with HygMX or NatMX and pRS316 (54, 55). Repair constructs were 60bp ssDNA oligos ordered from Integrated DNA technologies that included upstream homology, the desired nucleotide at the site of interest, and downstream homology. Transformants were selected on media containing either hygromycin or nourseothricin, depending on what marker plasmid was used. Replacement strains were then confirmed by sanger sequencing.

Following this strategy, *the mrp20-105E* nucleotide was engineered into two *hos3*Δ IV^{3S} XIV^{BY} segregants, and two *hos3*Δ IV^{BY} XIV^{BY} segregants were restored to *MRP20*. Similarly, at *MKT1* the causal, nearest upstream and downstream SNPs were engineered into two *hos3*Δ IV^{BY} XIV^{3S} segregants. Similarly, we generate BY *mrp20-105E*, BY $MKT1^{3S}$, 3S *mrp20-105E*, and 3S $MKT1^{BY}$ strains in this manner. Each single nucleotide parental replacement strain was then backcrossed to its own progenitor. Each subsequent diploid was sporulated and tetrad dissected, and we confirmed haploid genotypes by sequencing. The same approach was used to generate 3S mrp20-105E $MKT1^{BY}$ haploids by crossing 3S *mrp20-105E* and 3S $MKT1^{BY}$ strains. However, this strategy could not be followed to generate BY *mrp20-105E* $MKT1^{3S}$ haploids, because, crossing BY *mrp20-105E* and BY $MKT1^{3s}$ strains failed to produce any tetrads with 4 viable spores. Instead, we took BY $MKT1^{3S}$ strains and converted *MRP20* to *mrp20-105E*.

Mitochondrial genome instability experiments

We performed *petite* frequency assays as described in Dimitrov et al. (24) In brief, freezer stocks were streaked onto solid YPD media and grown for two days at 30°C. Single colonies were then resuspended in PBS, plated across dilutions onto YPDG plates (1% yeast extract, 2% peptone, 0.1% glucose, and 3% glycerol) and grown for five days at 30°C. Plates were then imaged with the BioRAD Gel Doc XR+ Molecular Imager at a standard size of 12.4 x 8.9 cm^2 (width x length) and imaged with epi-illumination using an exposure time of 0.5 seconds. Images were saved as 600 dpi tiffs. ImageJ (http://rsbweb.nih.gov/ij/) was used to examine growth and quantitate colony size as described in Dimitrov et al. (24). Colonies were then classified as *petite* and *grande* using a threshold defined as the maximum colony diameter of observed *petites* among BY and 3S wild type strains. *Petite* frequency is the ratio of small colonies to total colonies.

Modeling growth and examining the model in additional segregant populations

We modeled growth for *mrp20-105E* segregants from the Byx3S crosses fixed for *mrp20-105E* and engineered at *MKT1*. We incorporated MKT1, the 16 detected loci and the Chromosome II duplication in the linear model *growth ~ MKT1 + locus1 + locus2 + … locus16 + Chromsome II +*

error. This model was used to generate predicted growth values. We then compared our observed growth values to these predictions. Next, we sought to determine whether loci influencing the expressivity of *mrp20-105E* also affected growth in other strains. To accomplish this, we input the genotype information for each strain into our model to obtain predictions for its growth. We then compared the predicted values to the observed growth values and obtained Pearson correlations when possible.

Relationship between detrimental alleles, growth, and inviability

At each detected locus influencing response to *mrp20-105E*, we determined the allele associated with worse growth ('detrimental allele'). Next, we counted the number of detrimental alleles carried by each *mrp20-105E* segregant and examined how phenotypic response to *mrp20-105E* related to it. The *MKT1* and *SAL1* loci were not included when counting detrimental alleles, so that this relationship could be examined across different *MKT1-SAL1* genotype classes.

Funding

This work was funded by grants R01GM110255 and R35GM130381 from the National Institutes of Health to I.M.E., as well as Research Enhancement Fellowships from the University of Southern California Graduate School to R.S. and M.N.M.

Acknowledgments

We thank A. Dudley and G. Cromie for input at an early stage of this project, A. Mahableshwarkar and J. Sloan for assistance with experiments, and A. Coradini, I. Goldstein, and C. Hull for feedback during the implementation and writing phases of this work. We also thank the anonymous reviewers of this manuscript for their valuable feedback and Steven Finkel and his lab for allowing use of laboratory equipment.

Data Availability Statement

All raw and processed data used in this work is publicly available. Processed data tables and all code used for analyses are included in supplementary Data S1-9 and will be uploaded to the Ehrenreich Lab GitHub page. Raw sequencing data will be available through Bioproject accession PRJNA739014 in the NCBI Short Read Archive.

References

1. J. H. Nadeau, Modifier genes in mice and humans. *Nat Rev Genet* **2**, 165-174 (2001).
2. C. H. Chandler, S. Chari, I. Dworkin, Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet* **29**, 358-366 (2013).
3. J. D. Riordan, J. H. Nadeau, From peas to disease: modifier genes, network resilience, and the genetics of health. *Am J Hum Genet* **101**, 177-191 (2017).
4. A. J. F. Griffiths, S. R. Wessler, S. B. Carroll, J. Doebley, *Introduction to genetic analysis* (Macmillian Publishers, New York, NY, ed. 12th, 2015).
5. A. Raj, S. A. Rifkin, E. Andersen, A. van Oudenaarden, Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-918 (2010).
6. M. R. Wagner *et al.*, Microbe-dependent heterosis in maize. *Proc Natl Acad Sci U S A* **118** (2021).

7. R. D. Dowell *et al.*, Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
8. S. Chari, I. Dworkin, The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen. *PLoS Genet* **9**, e1003661 (2013).
9. C. H. Chandler, S. Chari, D. Tack, I. Dworkin, Causes and consequences of genetic background effects illuminated by integrative genomic analysis. *Genetics* **196**, 1321-1336 (2014).
10. A. B. Paaby *et al.*, Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *Elife* **4** (2015).
11. M. B. Taylor, I. M. Ehrenreich, Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet* **10**, e1004324 (2014).
12. M. B. Taylor, I. M. Ehrenreich, Transcriptional derepression uncovers cryptic higher-order genetic interactions. *PLoS Genet* **11**, e1005606 (2015).
13. V. Vu *et al.*, Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* **162**, 391-402 (2015).
14. M. B. Taylor, J. Phan, J. T. Lee, M. McCadden, I. M. Ehrenreich, Diverse genetic architectures lead to the same cryptic phenotype in a yeast cross. *Nat Commun* **7**, 11669 (2016).
15. J. T. Lee, M. B. Taylor, A. Shen, I. M. Ehrenreich, Multi-locus genotypes underlying temperature sensitivity in a mutationally induced trait. *PLoS Genet* **12**, e1005929 (2016).
16. C. H. Chandler *et al.*, How well do you know your mutation? Complex effects of genetic background on expressivity, complementation, and ordering of allelic effects. *PLoS Genet* **13**, e1007075 (2017).
17. M. N. Mullis, T. Matsui, R. Schell, R. Foree, I. M. Ehrenreich, The complex underpinnings of genetic background effects. *Nat Commun* **9**, 3548 (2018).
18. J. Hou, G. Tan, G. R. Fink, B. J. Andrews, C. Boone, Complex modifier landscape underlying genetic background effects. *Proc Natl Acad Sci U S A* **116**, 5045-5054 (2019).
19. J. T. Lee, A. L. V. Coradini, A. Shen, I. M. Ehrenreich, Layers of cryptic genetic variation underlie a yeast complex trait. *Genetics* **211**, 1469-1482 (2019).
20. L. Parts *et al.*, Natural variants suppress mutations in hundreds of essential genes. *Mol Syst Biol* **17**, e10138 (2021).
21. M. Galardini *et al.*, The impact of the genetic background on gene deletion phenotypes in Saccharomyces cerevisiae. *Mol Syst Biol* **15**, e8831 (2019).
22. K. Fearon, T. L. Mason, Structure and function of MRP20 and MRP49, the nuclear genes for two proteins of the 54 S subunit of the yeast mitochondrial ribosome. *J Biol Chem* **267**, 5162-5170 (1992).
23. E. C. Koc *et al.*, The large subunit of the mammalian mitochondrial ribosome. Analysis of the complement of ribosomal proteins present. *J Biol Chem* **276**, 43958-43969 (2001).
24. L. N. Dimitrov, R. B. Brem, L. Kruglyak, D. E. Gottschling, Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of Saccharomyces cerevisiae S288C strains. *Genetics* **183**, 365-383 (2009).
25. K. A. Lipinski, A. Kaniak-Golik, P. Golik, Maintenance and expression of the S. cerevisiae mitochondrial genome--from genetics to evolution and systems biology. *Biochim Biophys Acta* **1797**, 1086-1098 (2010).

26. G. S. Shadel, Yeast as a model for human mtDNA replication. *Am J Hum Genet* **65**, 1230-1237 (1999).
27. B. Kucejova, L. Li, X. Wang, S. Giannattasio, X. J. Chen, Pleiotropic effects of the yeast Sal1 and Aac2 carriers on mitochondrial function via an activity distinct from adenine nucleotide transport. *Mol Genet Genomics* **280**, 25-39 (2008).
28. B. Ephrussi, P. P. Slonimski, Subcellular units involved in the synthesis of respiratory enzymes in yeast. *Nature* **176**, 1207-1208 (1955).
29. R. Chen *et al.*, Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol* **34**, 531-538 (2016).
30. V. M. Narasimhan *et al.*, Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474-477 (2016).
31. O. Carlborg, C. S. Haley, Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* **5**, 618-625 (2004).
32. H. Shao *et al.*, Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc Natl Acad Sci U S A* **105**, 19910-19914 (2008).
33. T. F. Mackay, Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet* **15**, 22-33 (2014).
34. T. F. Mackay, J. H. Moore, Why epistasis is important for tackling complex human disease genetics. *Genome Med* **6**, 124 (2014).
35. M. L. Siegal, J. Y. Leu, On the nature and evolutionary impact of phenotypic robustness mechanisms. *Annu Rev Ecol Evol Syst* **45**, 496-517 (2014).
36. M. B. Taylor, I. M. Ehrenreich, Higher-order genetic interactions and their contribution to complex traits. *Trends Genet* **31**, 34-40 (2015).
37. S. K. Forsberg, J. S. Bloom, M. J. Sadhu, L. Kruglyak, O. Carlborg, Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet* **49**, 497-503 (2017).
38. I. M. Ehrenreich, Epistasis: searching for interacting genetic variants using crosses. *Genetics* **206**, 531-535 (2017).
39. R. F. Campbell, P. T. McGrath, A. B. Paaby, Analysis of epistasis in natural traits using model organisms. *Trends Genet* **34**, 883-898 (2018).
40. M. Costanzo *et al.*, A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
41. E. Kuzmin *et al.*, Systematic analysis of complex genetic interactions. *Science* **360**, aao1729 (2018).
42. A. H. Tong, C. Boone, Synthetic genetic array analysis in Saccharomyces cerevisiae. *Methods Mol Biol* **313**, 171-192 (2006).
43. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
44. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
45. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE*, 257-286 (1989).
46. T. Benaglia, D. Cahauveau, D. R. Hunter, D. S. Young, mixtools: an R package for analyzing mixture models. *Journal of Statistical Software* **32**, 1-29 (2009).
47. T. Matsui, I. M. Ehrenreich, Gene-environment interactions in stress response contribute additively to a genotype-environment interaction. *PLoS Genet* **12**, e1006158 (2016).
48. G. A. Churchill, R. W. Doerge, Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971 (1994).

49. L. M. Steinmetz *et al.*, Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326-330 (2002).
50. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87-96 (2002).
51. M. F. Laughery *et al.*, New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. *Yeast* **32**, 711-720 (2015).
52. K. Kannan *et al.*, One step engineering of the small-subunit ribosomal RNA using CRISPR/Cas9. *Sci Rep* **6**, 30714 (2016).
53. J. R. Thompson, E. Register, J. Curotto, M. Kurtz, R. Kelly, An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* **14**, 565-571 (1998).
54. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
55. R. S. Sikorski, P. Hieter, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19-27 (1989).

Figures

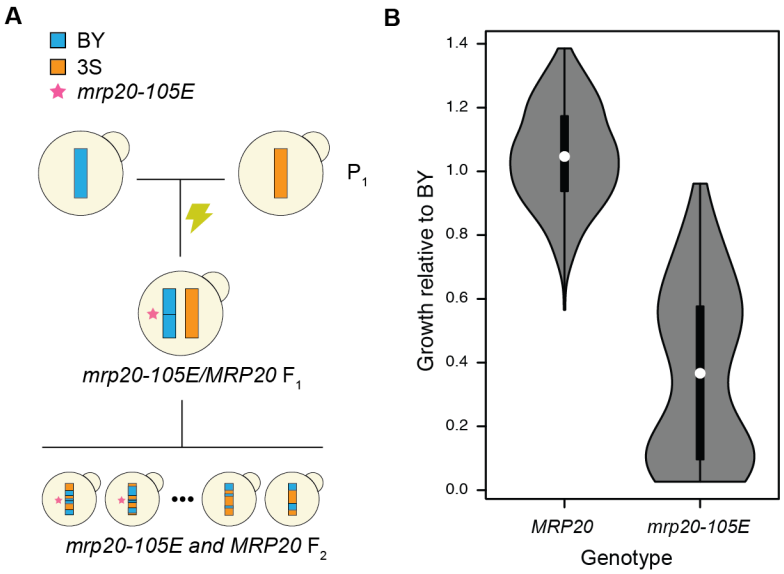

Figure 1. The *mrp20-105E* mutation occurred spontaneously, increasing phenotypic variance in the BYx3S cross. (A) A spontaneous mutation in a BY/3S diploid gave rise to a BYx3S segregant population in which *mrp20-105E* segregated. (B) The *mrp20-105E* segregants exhibited increased phenotypic variance and a bimodal distribution of phenotypes. Throughout the paper, blue and orange are used to denote BY and 3S genetic material, respectively. All growth data presented in the paper are measurements of colonies on agar plates containing rich medium with ethanol as the carbon source.

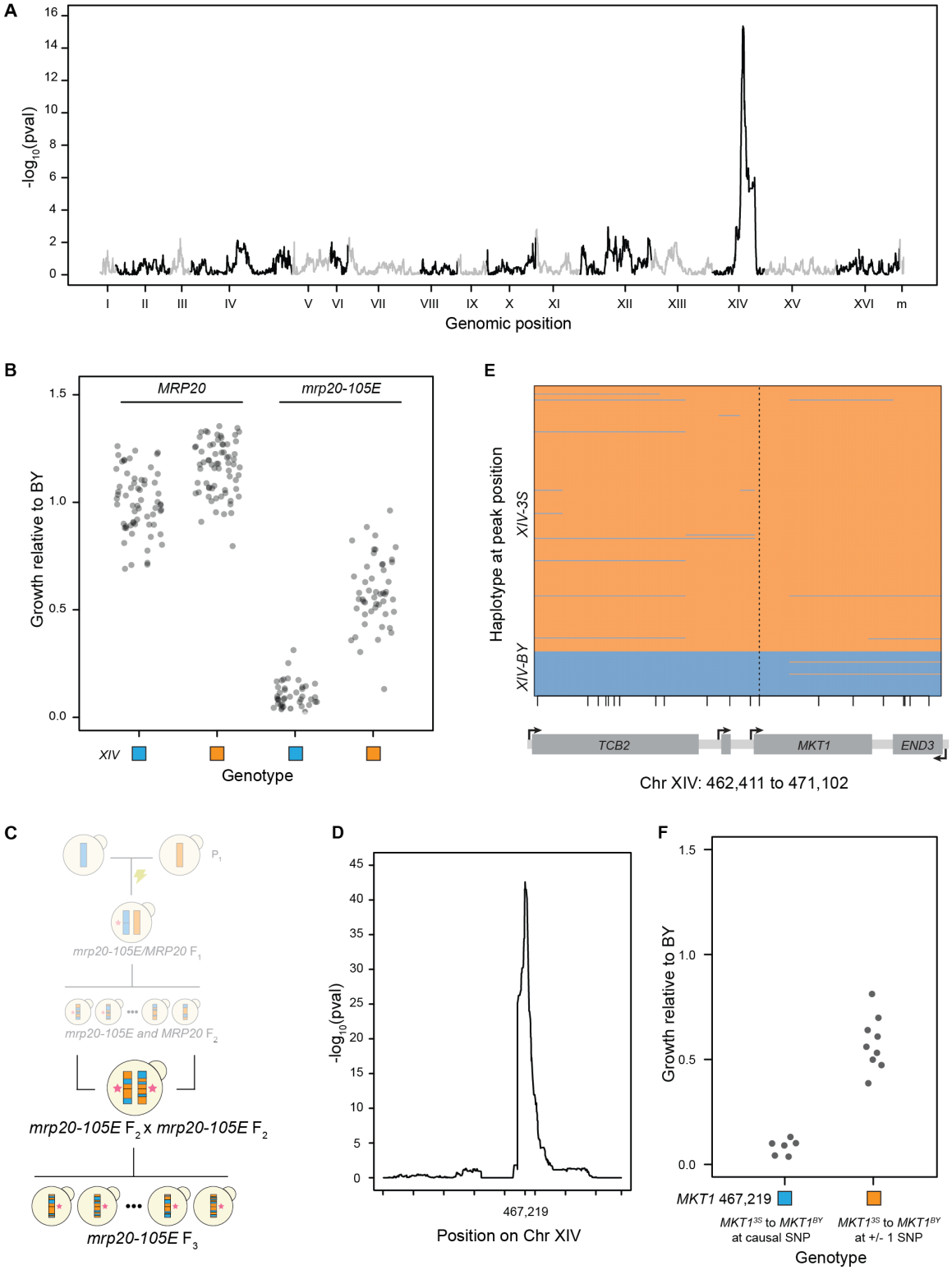

Figure 2. Epistasis between *MRP20* and *MKT1* appears to mostly explain response to *mrp20-105E*. (A) Linkage mapping in the BYx3S segregants shown in Fig 1 identified a locus on Chromosome XIV that exhibits a two-way genetic interaction with *MRP20*. (B) The

Chromosome XIV locus had effects in both *MRP20* and *mrp20-105E* segregants but had a greater effect among *mrp20-105E* segregants. (C) To identify the causal gene, we crossed two *mrp20-105E* F_2 segregants that differed at the Chromosome XIV locus and gathered a panel of F_3 segregants. (D) Linkage mapping in the F_3 segregants identified the Chromosome XIV locus at high resolution, with a peak at position 467,219. Tick marks denote every 100,000 bases along the chromosome. (E) Recombination breakpoints in the F_3 segregants delimited the Chromosome XIV locus to a single SNP in *MKT1* at position 467,219. Vertical dashed line highlights the delimited causal polymorphism, while small vertical lines along the x-axis indicate different SNPs in the window that is shown. (F) Engineering the BY allele into *mrp20-105E* XIV^{3S} segregants changed growth (left), while substitutions at the nearest upstream and downstream variants did not (right).

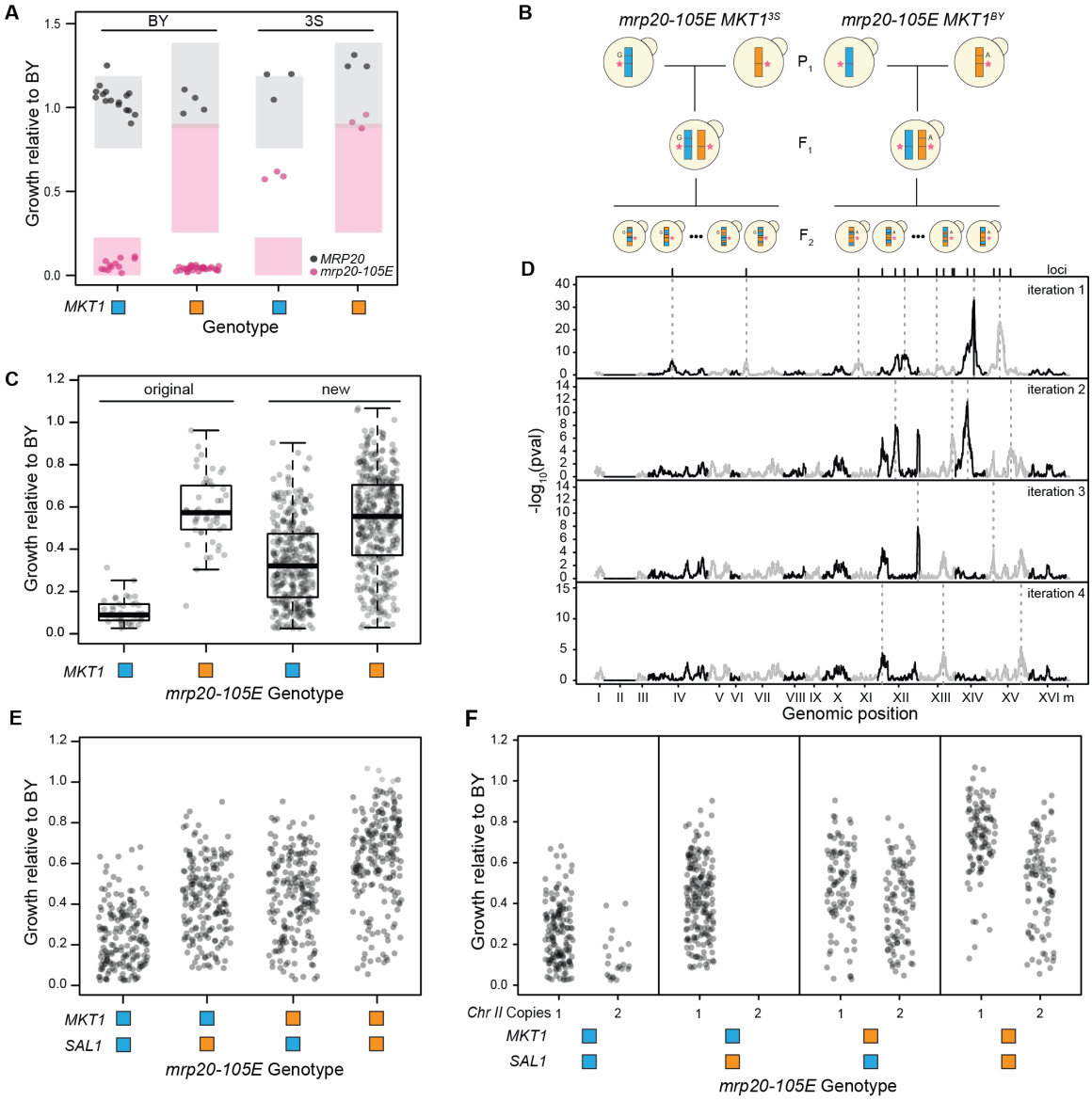

Figure 3. Additional loci govern response to the mutation. (A) We engineered all combinations of *MRP20* and *MKT1* into the BY and 3S cross parents. Expected phenotypes are shown as shaded boxes denoting 95% confidence interval based on the originally obtained segregant phenotypes. (B) We generated BY x 3S crosses in which all segregants carried *mrp20-105E*. Two crosses were performed: one in which all segregants carried $MKT1^{BY}$ and one in which all segregants carried $MKT1^{3S}$. Tetrads were dissected and spores were phenotyped for growth on ethanol. (C) Each of the new crosses showed increased growth that extended from inviable to wild type, differing from the more qualitative bimodal phenotypes seen among the original *mrp20-105E MKT1* segregant populations. (D) Linkage mapping identified a total of 16 loci that influenced growth. After four iterations of a forward regression, no additional loci were identified. (E) Inviable segregants were present among all *mrp20-105E MKT1 SAL1* genotype classes. (F) Aneuploid individuals with duplicated Chromosome II showed reduced growth. Aneuploid individuals were not evenly detected across the different *MKT1-SAL1* genotype classes.

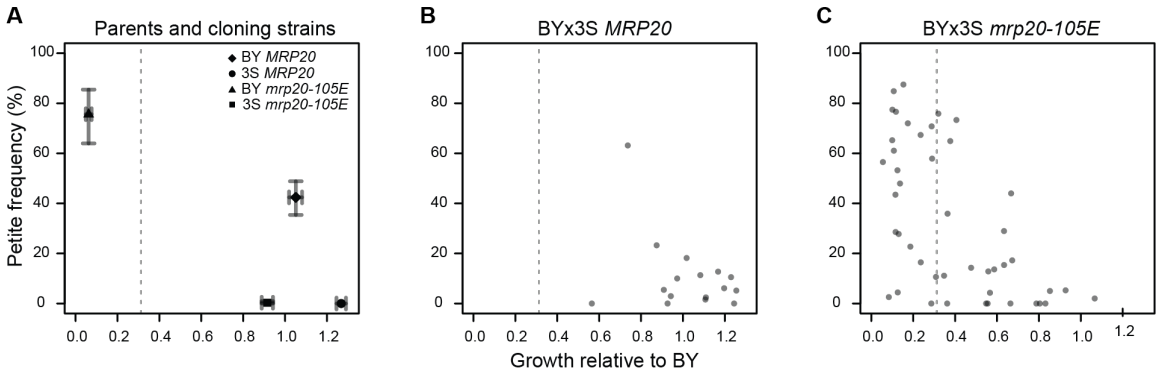

Figure 4. Mitochondrial genome instability partially underlies the expressivity of *mrp20-105E*. We measured petite formation frequency, which estimates the proportion of cells within a clonal population capable of respiratory growth. Higher petite frequency is a proxy for greater mitochondrial genome instability. (A) We examined *MRP20* and *mrp20-105E* versions of the BY and 3S parent strains. For each, average values and 95% bootstrapped confidence intervals are shown. BY showed elevated mitochondrial genome instability in the presence of *mrp20-105E*, while 3S showed no change. (B) We examined 16 BYx3S *MRP20* segregants. These segregants were randomly selected and spanned the range of growth values for *MRP20* segregants. (C) 45 BYx3S *mrp20-105E* segregants. Poorer growing segregants tended to exhibit higher mitochondrial genome instability, though some exhibited wild type levels of mitochondrial genome instability. The gray dashed line indicates the threshold used to call inviability.

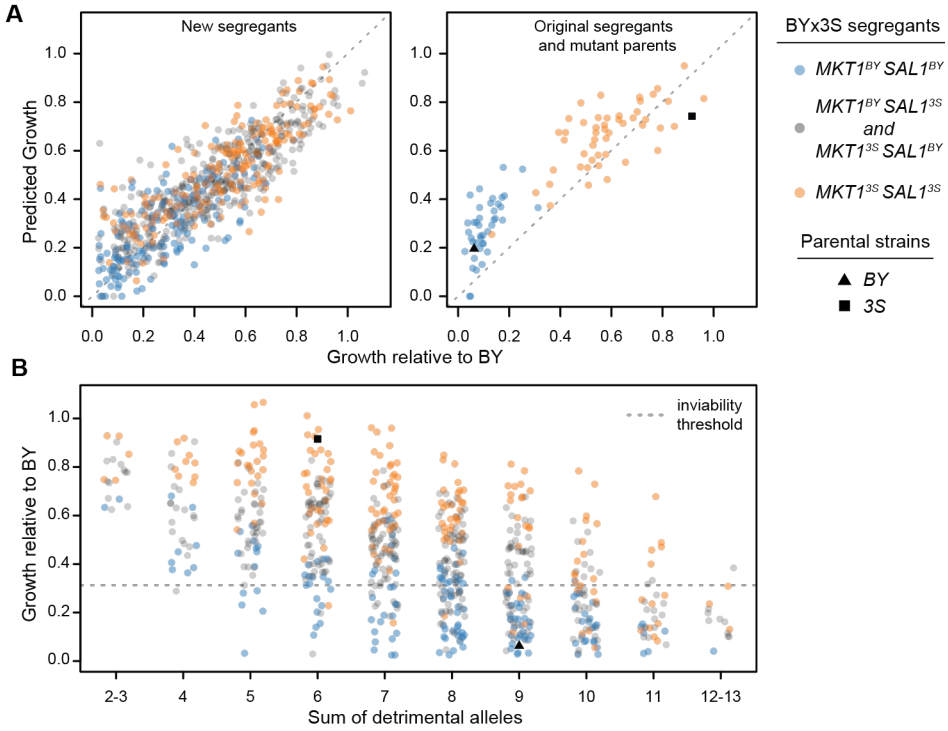

Figure 5. Detected loci quantitatively and qualitatively explain mutant phenotypes. (A) We fit a linear model accounting for the effects of all detected loci and the aneuploidy on the growth of *mrp20-105E* segregants. This model not only explained the growth of the new BY x 3S *mrp20-105E* crosses generated in this paper, but also accurately predicted the phenotypes of the mutant parents and previously generated segregants. (B) We examined growth relative to the sum of detrimental alleles carried by a segregant. This relationship shows how collections of loci produce a quantitative spectrum of phenotypes, including instances of qualitative phenotypic responses. This relationship explains the full range of responses, from inviable to wild type growth, across *MKT1-SAL1* genotypes. The gray dashed line indicates the threshold used to call inviability.