Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast

Adam M. Deutschbauer,[*,†] **Daniel F. Jaramillo,**[†] **Michael Proctor,**[†] **Jochen Kumm,**[†]
Maureen E. Hillenmeyer,[†] **Ronald W. Davis,**[*,†] **Corey Nislow**[†] **and Guri Giaever**[†,1]

*Department of Genetics, Stanford University School of Medicine, Stanford, California 94305 and
†Stanford Genome Technology Center, Palo Alto, California 94304-1103

ABSTRACT

Haploinsufficiency is defined as a dominant phenotype in diploid organisms that are heterozygous for a loss-of-function allele. Despite its relevance to human disease, neither the extent of haploinsufficiency nor its precise molecular mechanisms are well understood. We used the complete set of *Saccharomyces cerevisiae* heterozygous deletion strains to survey the genome for haploinsufficiency via fitness profiling in rich (YPD) and minimal media to identify all genes that confer a haploinsufficient growth defect. This assay revealed that ~3% of all ~5900 genes tested are haploinsufficient for growth in YPD. This class of genes is functionally enriched for metabolic processes carried out by molecular complexes such as the ribosome. Much of the haploinsufficiency in YPD is alleviated by slowing the growth rate of each strain in minimal media, suggesting that certain gene products are rate limiting for growth only in YPD. Overall, our results suggest that the primary mechanism of haploinsufficiency in yeast is due to insufficient protein production. We discuss the relevance of our findings in yeast to human haploinsufficiency disorders.

O RGANISMS heterozygous for a loss-of-function allele often have no discernible phenotype. This observation has been attributed to the metabolic theory of dominance described over 30 years ago. Briefly, this model states that the phenotypic consequences of heterozygous loss-of-function alleles are masked by the presence of one wild-type allele due to the redundancy of cellular physiology (KACSER and BURNS 1981). There are, however, exceptions to this rule where deletion of a single gene copy leads to an abnormal phenotype. Such haploinsufficiency is observed in all eukaryotes from yeast to humans. In yeast, the phenomenon has been characterized in gene-by-gene analyses of, for example, cytoskeletal components such as actin (*ACT1*) (DRUBIN *et al.* 1993), tubulin (*TUB1*) (SCHATZ *et al.* 1986), and components of the spindle pole body (*NDC1*) (CHIAL *et al.* 1999). An explanation for the observation of reduced fitness in these heterozygous strains is provided by biochemical studies showing that a balance of protein levels is required to maintain cytoskeletal integrity.

The relevance of haploinsufficiency in human disease has become increasingly apparent. Many of the haploinsufficient mutations in humans are observed in transcription factors including *TWIST* (JOHNSON *et al.* 1998) and *GATA3* (MUROYA *et al.* 2001). It is not surprising that such haploinsufficiencies are detrimental, as they should result in multiple transcription defects of diverse downstream targets (for review see SEIDMAN and SEID-

MAN 2002). Haploinsufficiency has also been implicated in cancer; heterozygous mutations in both *ATM* (SPRING *et al.* 2002) and *BLM* (GOSS *et al.* 2002) result in increased cancer susceptibility.

We sought to place our observations of yeast haploinsufficiency into the context of the prevailing theories of haploinsufficiency. One theory posits that deviations from the normal stoichiometry of members in a protein complex cause haploinsufficiency. This theory, the balance hypothesis (PAPP *et al.* 2003), predicts that the haploinsufficient phenotype will be the same as the overexpression phenotype, as both scenarios should result in an imbalance of subunits of a protein complex. The opposing theory states that haploinsufficiency may simply be due to the reduced levels of protein produced in the heterozygous state. In this scenario, overexpression of haploinsufficient genes is not expected to be deleterious and should result in the wild-type phenotype being restored. Our genome-wide haploinsufficiency profiling allows us to discriminate between these two hypotheses.

The yeast deletion collection is a valuable resource for examining loss-of-function phenotypes of yeast genes. The majority of studies to date have used either the haploid or the homozygous deletion collections (see, for example, BIRRELL *et al.* 2001; OOI *et al.* 2001; TONG *et al.* 2001; DEUTSCHBAUER *et al.* 2002; GIAEVER *et al.* 2002). While powerful, these studies do not include the essential genes. The use of the heterozygous deletion collection circumvents this problem and adds to the growing list of studies that address the essential genes (HAZBUN *et al.* 2003; MNAIMNEH *et al.* 2004). Because the majority

[1]*Corresponding author:* Stanford Genome Technology Center, 855 California Ave., Palo Alto, CA 94304-1103.
E-mail: ggiaever@stanford.edu

of heterozygotes show no obvious growth defects [only a single gene in the genome, *MLC1*, is lethal as a heterozygote (STEVENS and DAVIS 1998)] a high-resolution method for phenotyping must be employed. To identify the complete set of genes in yeast that are haploinsufficient under standard laboratory growth conditions we used the yeast gene-deletion collections combined with automated robotic sample handling in parallel fitness assays (WINZELER *et al.* 1999; GIAEVER *et al.* 2002). Combining this genome-wide approach with confirmation by individual strain analysis, we find that 184 genes (~3% of the genome) exhibit significant haploinsufficiency under optimal growth conditions. This gene set is highly enriched for genes that are functionally annotated in basic cellular processes (*e.g.*, protein biosynthesis and mRNA processing), indicating that many core cellular events are sensitive to gene dosage. Our data strongly suggest that the majority of haploinsufficiency in yeast results from insufficient protein production in heterozygous deletion strains.

MATERIALS AND METHODS

Strains and plasmids: Yeast deletion strains were constructed by an international consortium in the S288C background (BRACHMANN *et al.* 1998). Plasmid pTN001 (a gift from Taavi Neklesa) was generated by replacing the *URA3* gene in pRS426 with the *hph* gene encoding resistance to hygromycin B (GOLDSTEIN and McCUSKER 1999).

Media and growth conditions: Yeast extract/peptone/dextrose (YPD) and minimal media were prepared as described (GUTHRIE and FINK 1991). YPD was supplemented with hygromycin B (300 µg/ml) for all overexpression experiments with plasmid pTN001.

Deletion pool construction, growth, and chip experiments: Deletion pool construction and pool growth were as described (GIAEVER *et al.* 2002) with the following modifications of growth conditions. Frozen aliquots of the two independently constructed heterozygous and homozygous pools were diluted in YPD or minimal media to an OD_{600} of 0.0625, and 0.7 ml was pipetted into a well of a 48-well microplate. Cells were grown in a Tecan (Durham, NC) GENios microplate reader and every 5 generations cells were automatically pipetted into 0.7 ml of fresh media using a Packard Multiprobe II four-probe liquid handling system (Perkin-Elmer Life Sciences, Norwalk, CT) controlled by custom LABVIEW software (National Instruments, Austin, TX). Over the course of 20 generations of growth, cells were saved every 5 generations and frozen at $-20°$ for subsequent preparation of genomic DNA. Genomic DNA preparation, PCR, and chip hybridization to the Affymetrix (Santa Clara, CA) TAG3 array were as described (GIAEVER *et al.* 2002).

YPD fitness analysis: *Preprocessing and normalization:* For the primary calculation of YPD fitness values, we analyzed the data collected from time points using four independently constructed deletion pools (heterozygous A and B, homozygous C and D) separately. Each deletion strain is typically represented by four hybridization signals, corresponding to "tags," on the array. For each of the four pools, we determined a "present" tag set based on six hybridizations from time zero (aliquots hybridized straight from $-80°$). For all deletion pools ~8000 tags failed to meet the present criteria defined by a mean hybridization intensity greater than fourfold over the

mean array background. These low-intensity tags were removed from the analysis and each array was normalized to have a standard mean intensity across all tags in the corresponding pool.

Calculation of regression coefficients from time-course data in YPD: Regression slopes were determined using a linear model corresponding to a multiple-regression model on time (measured in generations and treated as a quantitative predictor) and replicate series (treated as a categorical predictor) simultaneously. This model represents an analysis of covariance (ANCOVA) for (1) time effects, (2) replicate series effects, and (3) series-time interactions. This analysis provides estimates of statistical significance using the *F*-statistic. We interpret the results of the ANCOVA as a linear regression where the *F*-statistic provides *P*-values. This analysis was performed using an additive linear dummy regression model with interactions in R (Fox 2002, p. 133) for each tag. We added 1 to the tag regression slope to obtain a relative tag fitness where values <1 indicate a fitness defect.

Calculation of strain fitness: We averaged individual tag fitness values for each deletion strain in a given pool. We referred to these fitness values as heterozygous (het)_ypd_A, het_ypd_B, homozygous (hom)_ypd_C, and hom_ypd_D. The following criteria were used to identify heterozygous strains with reduced fitness: both het_ypd_A and het_ypd_B fitness values had to be <0.98 and at least one tag for that gene had to be statistically significant ($P < 0.05$) in both pools as determined by the ANCOVA. To calculate the HET_AV fitness measure for each gene, we averaged the fitness values from all present molecular tags across both the A and B heterozygous pools. We used the same criteria for selecting strains with reduced fitness for homozygous diploids.

Minimal media fitness analysis: We repeated the ANCOVA analysis for two time series performed in minimal media (MM). Preprocessing, normalization, and the multiple-regression model were analogous to the YPD analysis. Regression slopes and *F*-statistic *P*-values were obtained for these samples. Due to fewer replicate experiments compared to the YPD data and unexpected auxotrophic mutations in the deletion collection, our analysis of minimal media strain fitness differed slightly from our YPD analysis. We calculated het_mm_B and hom_mm_D values for each strain on the basis of two replicate experiments in minimal media supplemented with histidine, leucine, and uracil. We identified heterozygous and homozygous slow-growth strains as those with het_mm_B or hom_mm_D, respectively, values <0.95 and at least one statistically significant ($P < 0.05$) tag. Our more stringent criteria for slow growth in minimal media (fitness <0.95) compared to YPD (fitness <0.98) reflect the increased confidence in many YPD replicate experiments and the results of individual strain examination. During the course of the analysis, we identified ~100 heterozygous strains with severe fitness defects in minimal media while the corresponding homozygous deletions grew normally. We confirmed that 22 of these strains (see supplementary Table 4 at http://chemogenomics.stanford.edu/supplements/01yfh/) were erroneously homozygous for *met15Δ* due to an observed increase in loss of heterozygosity at this locus (McMURRAY and GOTTSCHLING 2003). To address the extent of this phenomenon, we fitness profiled the heterozygous and homozygous deletion pools in minimal media supplemented with histidine, leucine, uracil, lysine, and methionine. On the basis of the restoration of wild-type growth in the added presence of lysine and methionine, we estimate that ~100 strains are homozygous for the *met15Δ* marker (and have forwarded this information to the Yeast Deletion Collection curator).

Construction of overexpression plasmids: The coding regions for haploinsufficient genes with 500 bp upstream of the start codon and 200 bp downstream of the stop codon were

PCR amplified from total genomic DNA and cloned into vector pTA2.1 (Invitrogen, San Diego). Inserts were restriction enzyme digested from pTA2.1 and subcloned into vector pTN001 using *Sbf*I and *Not*I sites introduced into the amplification primers.

Individual growth curves: Strains were diluted in YPD or minimal media to an OD_{600} of 0.0625 and 0.7 ml was pipetted into a well of a 48-well microplate. Cells were grown in a Tecan GENios microplate reader and monitored for growth. Calculations of doubling times were based on exponential fits to the growth curves using custom software. For all overexpression experiments with plasmid pTN001 and for haploinsufficient strains with slight growth defects, we performed 20-generation growth curves using the automation described for the deletion pool experiments.

Gene expression experiments: Total RNA collection from exponentially growing cells and probe synthesis was performed as described (PRIMIG *et al.* 2000). Labeled cDNA probes were hybridized to S98 oligonucleotide arrays and processed according to the Affymetrix expression manual. Affymetrix CEL files from six YPD experiments and two minimal media experiments were transferred to the dCHIP software package for normalization and analysis (LI and WONG 2001). Using expression indices calculated from a model-based analysis, we calculated fold expression changes between YPD and minimal media for all genes. We identified differentially expressed genes as those with significant ($P < 0.05$) fold changes.

RESULTS

Approximately 3% of the yeast genome displays haploinsufficiency in rich media: The yeast deletion strains are tagged with unique molecular bar codes or tags, enabling pooling of the deletion strains and parallel phenotypic analysis. We determined fitness values for both the heterozygous and homozygous deletion collections in rich media (YPD) by monitoring the abundance of the molecular bar codes over time in YPD as previously described (WINZELER *et al.* 1999; GIAEVER *et al.* 2002). To control for variations introduced by pool construction, we profiled two independently constructed pools of both the heterozygous (referred to as pools A and B) and homozygous deletion collections (referred to as pools C and D). To minimize variation due to inconsistency in sample collection (*e.g.*, capturing cultures at precise generation times), all samples were collected robotically and each experiment was replicated 6 times per pool across 5 time points (comprising a total of 24 time series with individual data points collected at 0, 5, 10, 15, and 20 generations). Genomic DNA was then prepared from samples and molecular tags were PCR amplified and hybridized to a high-density array of oligonucleotides with sequence complementary to the bar codes. The relative growth rate of all strains in the pool was calculated using a linear-regression model and analysis of covariance (see MATERIALS AND METHODS). The use of robotics and rigorous statistical analysis of the data allowed us to reproducibly detect small fitness defects that were then verified by individual strain analysis.

Fitness profiling of the heterozygous deletion collection grown in YPD identified 272 genes with a growth defect (relative fitness <0.98, $P < 0.05$, see MATERIALS AND METHODS for criteria) in both the A and B heterozygous pools (Figure 1A). Seventy-five of the 272 haploinsufficient genes are neither essential nor required for normal growth rate as homozygotes, suggesting that these strains were defective in other ways. Indeed, many of these 75 strains (1) failed to sporulate, (2) mated as diploids, and (3) produced inviable meiotic progeny and were therefore eliminated from our analysis (for a complete list of strains and their defects, see supplementary Table 1 at http://chemogenomics.stanford.edu/supplements/01yfh/). Strains deleted for dubious open reading frames, which often physically overlap with characterized genes, were also removed. After this data reduction we were confident that the slow-growth phenotype of the remaining 184 heterozygous strains was associated with the intended gene deletion. This gene set of 184 is a robust reflection of haploinsufficiency in the yeast genome and indicates that ∼3% of the yeast genome is haploinsufficient under optimal conditions (see supplementary Table 2 at http://chemogenomics.stanford.edu/supplements/01yfh/ for a complete gene list). These haploinsufficient strains are enriched for essential genes (53.3%, $N = 98$) compared to the genome (18.7%; $P < 3.3e$-16 by hypergeometric distribution). The remaining 86 haploinsufficient genes are nonessential and also have growth defects as homozygous diploids (see below).

To compare the heterozygous deletion strains with reduced fitness to the homozygous deletion strains with reduced fitness, we profiled the homozygous deletion collection using identical assay conditions. In this case, we identified 891 slow-growing homozygous deletion strains (∼20% of the genome; see supplementary Table 3 at http://chemogenomics.stanford.edu/supplements/01yfh/ for a complete gene list). To facilitate downstream analysis of the fitness values as with the heterozygous data, we averaged the data from across both pools (see MATERIALS AND METHODS) to obtain a single value for each gene, HOM_AV. Likewise, we averaged the fitness values across the A and B heterozygous pools to generate a HET_AV value for each gene. A genome-wide comparison between the heterozygous and homozygous pools reveals that the slow growth of heterozygous strains is generally less severe than that of homozygotes (compare Figure 1B and 1D). An examination of the distribution of the HOM_AV values reveals that 77 of the 891 have "severe" slow growth (relative fitness <0.80) and thus are "nearly" essential (Figure 1D). The combination of the 891 genes with a homozygous growth defect with the 1102 essential genes demonstrates that greater than one-third (1993 genes) of the yeast genome is either essential or required for optimal growth in rich media.

Verification of haploinsufficiency results: In many cases the haploinsufficiency observed for particular strains is not severe. To confirm these array results we analyzed
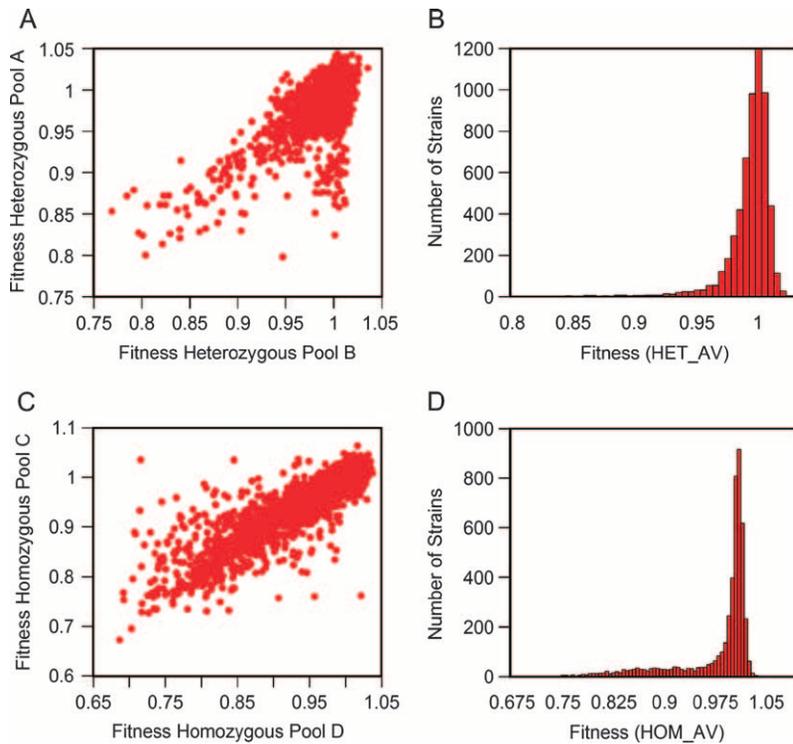
FIGURE 1.—Fitness profiling of yeast deletion strains in rich media. (A) Scatter plot comparing the fitness values for 5668 genes detected in two independent pools (A and B) of the heterozygous deletion collection. (B) Distribution of fitness values averaged across the A and B pools for 5668 heterozygous strains. (C) Scatter plot comparing the fitness values for 4624 genes detected in two independent pools (C and D) of the homozygous diploid deletion collection. The minor discrepancies between the two homozygous deletion pools likely result from unequal starting amounts of each strain between the pools. (D) Distribution of fitness values averaged across the C and D pools for 4624 homozygous diploid strains.

these strains individually. Thirty strains heterozygous for haploinsufficient genes covering a range of fitness values were grown over a 24-hr period with optical density measurements collected automatically every 15 min, providing highly accurate and reproducible growth rate measurements. Each curve was performed in triplicate and representative samples are shown in Figure 2A. In some cases, particularly for strains with HET_AV values >0.97, a 24-hr growth assessment was insufficient to distinguish the growth defect compared to wild type. In these cases we verified the slow growth of these strains (such as *mcm2Δ/MCM2,* Figure 2B), using the automated growth assay over 20 population doublings (equivalent to the number of generations that takes place in a full-genome screen). By 20 generations, the growth defect of *mcm2Δ/MCM2* is reproducibly detected (similar results were observed for *rap1Δ/RAP1* and *glc7Δ/GLC7*). This finding validates the use of molecular bar coding followed by highly quantitative individual growth assays to reproducibly detect changes in growth rate as small as 2%.

Quantitative fitness profiling yields a relative value representing the fitness of each strain (GIAEVER *et al.* 2002). The relationship between these values, generated from microarray data, and actual generation times, however, is not clear. To directly address this relationship, we plotted the generation times for the 30 strains confirmed as haploinsufficient above as a function of their fitness value. As expected, strains with longer generation times are associated with lower fitness values, demonstrating that relative fitness values can be used as a quantitative measure of actual generation times (Figure 2C). The majority of the 184 haploinsufficient strains have

a HET_AV value between 0.90 and 0.98 (152 genes), suggesting that most haploinsufficient genes have a doubling time between 85 min (wild type) and 110 min.

Haploinsufficient genes encode subunits of protein complexes involved in metabolic processes: We compared the broad functional distribution of the ∼1100 essential genes with the ∼900 nonessential homozygous genes that, when deleted in the homozygote, result in slow growth [Saccharomyces Genome Database (SGD) gene ontology (GO) term mapper (DWIGHT *et al.* 2002)]. This comparison showed both gene sets to be quite similar to one another but distinct from the genome as a whole. Both sets are enriched for genes involved in metabolism ($P < 0.001$ for both gene classes by the hypergeometric distribution) and contain few uncharacterized genes (Figure 3A). These differences are more pronounced when the functions of the 184 haploinsufficient genes are examined; over half of the annotated functions are involved in metabolism and even fewer genes remain unclassified (Figure 3A). In addition, genes involved in "broad" cellular functions such as signal transduction and amino acid metabolism are underrepresented in the haploinsufficient class of genes.

A more precise functional map [SGD GO term finder (DWIGHT *et al.* 2002)] of the genes sensitive to haploinsufficiency revealed processes related to protein metabolism, including rRNA processing, ribosome biosynthesis, translation control, and protein folding (Figure 3B). Other haploinsufficient genes are involved in complex processes such as transcription from the RNA polymerase II promoter (*RPB5*, *RPB8*, *SRB7*, *RPO26*, *RPC10*, *RPB4*, *RPB7*, and *RPB3*), DNA replication (*MCM2* and

CDC47), ER-to-Golgi transport (*SEC23* and *SEC34*), nuclear import/export (*NUP60, NUP145, NUP120,* and *NIC96*), and cytoskeletal function (*ACT1, TUB1, SPC97,* and *SPC98*).

Many haploinsufficient genes encode components of multisubunit protein complexes. We find that a higher percentage of haploinsufficient genes are in GO-annotated protein complexes (~77%) compared to the genome in general (~20%, $P < 3.3e$-16; the list of GO-annotated protein complexes can be found at http://www.yeastgenome.org/). A striking example is the haploinsufficiency we observe for structural components of the ribosome. Over 57% (32 of 56 genes) of the components of the small cytosolic subunit and 49% (40 of 81 genes) of the large cytosolic subunit components are haploinsufficient under optimal growth conditions. Other complexes with multiple haploinsufficient subunits include the CCT folding chaperone, the exosome, the core subunit of RNA polymerase II, *SPC97* and *SPC98* of the spindle pole body, and the α- (*SUI2*) and β- (*SUI3*) subunits of the translation initiation factor eIF2 (see Table 1 for complexes haploinsufficient in YPD). In contrast, single genes encoding enzymatic activities (such as glycolytic enzymes) are rarely haploinsufficient in YPD.

Duplicated ribosomal genes have less severe haploinsufficiency than their nonduplicated partners: Over 50% of the structural components of the cytoplasmic ribosome are haploinsufficient. Because ~75% of the ribosomal genes are duplicated (59 are duplicated, 19 are not duplicated) (Planta and Mager 1998) with high conservation, we can ask why certain ribosomal components are duplicated and others are not. For example, there may be evolutionary benefits for maintaining duplications to buffer the cell from the deleterious effects of haploinsufficiency. If this is the case, the evolutionary cost of maintaining the duplications should outweigh the costs of producing additional transcripts from a single-gene copy and therefore, the haploinsufficiency of the 19 nonduplicated ribosomal components should be more pronounced compared to that of the 59 duplicated components. This prediction is supported by the data; the mean haploinsufficiency of the 19 nonduplicated ribosomal genes (HET_AV = 0.92) is more severe than that of the 59 duplicated ribosomal genes (HET_AV = 0.95; *t*-test, $P = 0.036$, Figure 4A). This analysis demonstrates that (1) ribosomal genes are exceedingly sensitive to gene dosage regardless of whether the gene is duplicated and (2) having a duplicated partner provides a small buffer against the effects of haploinsufficiency. In addition, these data support the observation that ribosomal genes are significantly overrepresented among the complete set of duplicated genes in yeast (Papp *et al.* 2003).

This finding raises the question of whether strains deleted for nonduplicated ribosomal genes exhibit decreased fitness over those deleted for duplicated ribosomal genes simply because the majority of them are
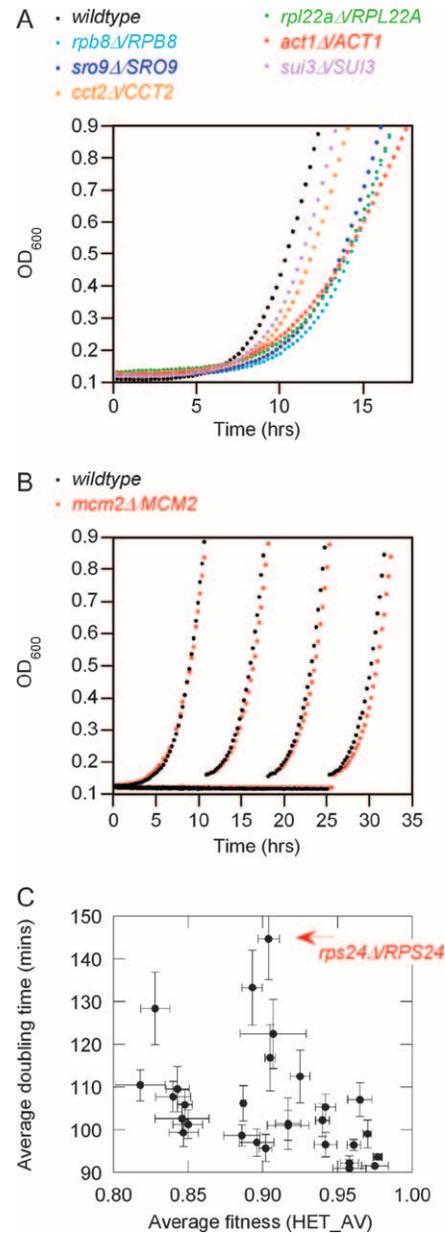

FIGURE 2.—Verification of haploinsufficiency results. (A) Representative deletion strains detected as haploinsufficient by hybridization of amplified bar codes to the TAG3 microarray were grown individually for ~20 hr in a microplate reader. (B) The *mcm2Δ/MCM2* deletion strain, which exhibits a slight haploinsufficient phenotype, was monitored for growth over 20 population doublings in a Tecan GENios microplate reader. Every 5 generations, cells were robotically transferred to a well containing fresh media. By 20 generations, the growth defect of the *mcm2Δ/MCM2* heterozygote becomes apparent. (C) Thirty putative haploinsufficient strains were grown individually as described in A and B to determine doubling times. For each heterozygous deletion strain, the growth curve was performed in triplicate. The calculated doubling times for these strains were then plotted against their corresponding HET_AV values as determined by microarray analysis. A positive correlation between doubling time and calculated fitness (HET_AV) is observed (Pearson correlation coefficient = −0.44, $P = 0.01$). Outliers in this correlation, such as the *rps24aΔ/RPS24A* strain, reflect a limitation of our high-throughput approach in accurately measuring doubling times for very slow-growing strains.
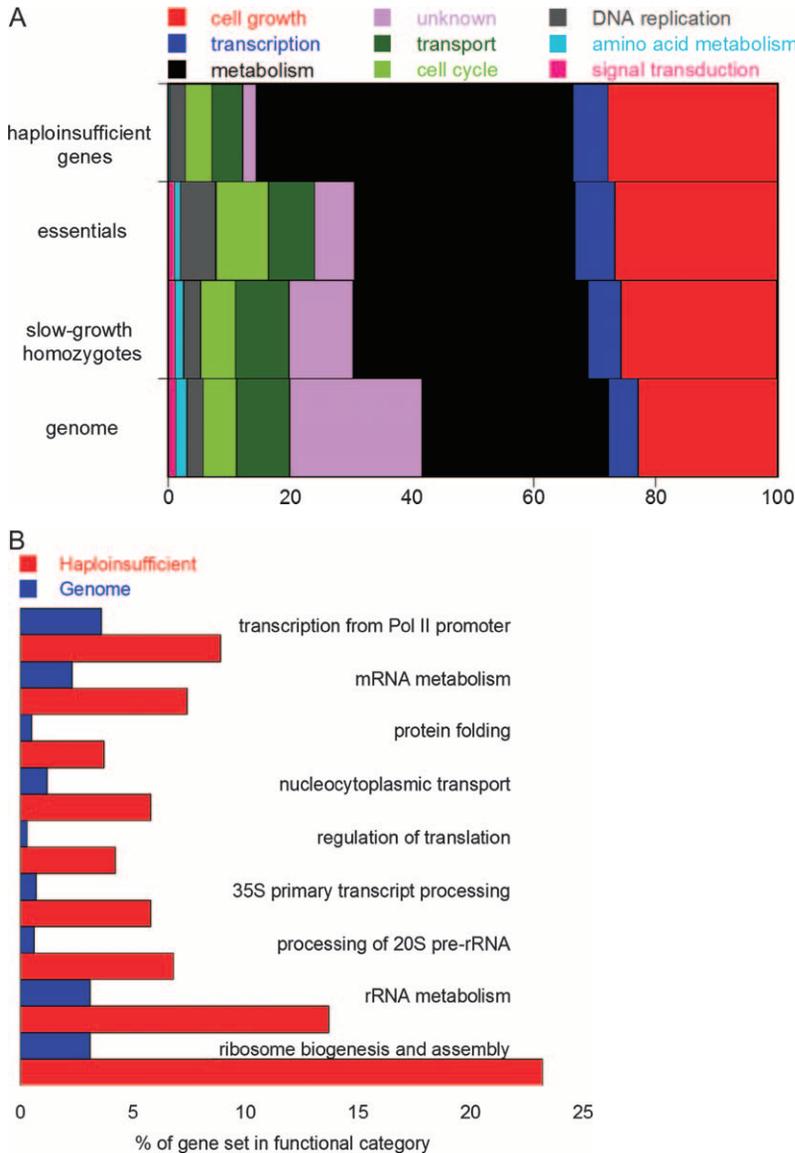
FIGURE 3.—Haploinsufficient genes are enriched for metabolic functions. (A) Broad functional classification of the yeast genome, 1102 essential genes, 891 nonessential slow-growth homozygous genes, and 184 haploinsufficient genes derived from the gene ontology (GO) term mapping function available on the SGD website (www.yeastgenome.org). The classifications represent general GO parental terms (GO slim) for the process ontology. (B) Detailed GO processes enriched in the haploinsufficient gene class as determined through the use of the GO term finder function available on the SGD website (http://www.yeastgenome.org/ help/goTermFinder.html, $P < 0.0001$ for all processes shown, calculated using the binomial approximation of the hypergeometric distribution). Some GO processes with high similarity to the processes shown were omitted.

essential (16 of 19 nonduplicated ribosomal genes are essential while only 5 of 118 duplicated ribosomal genes are essential). We tested whether this was the case by comparing the haploinsufficiency of the 98 essential genes *vs.* that of the 86 nonessential genes. Interestingly, a comparison of the observed haploinsufficiencies of

TABLE 1

Complexes overrepresented among 184 haploinsufficient genes

Complex name	GO complex ID	No. of haploinsufficient genes (total genes in complex)	Enrichment P-value [a]
Chaperonin-containing T-complex	5832	7 (8)	1.1*e*-10
DNA-directed RNA polymerase II, core complex	5665	7 (12)	4.9*e*-08
Exosome (RNAse complex)	178	6 (13)	5.5*e*-06
Ribosome	5840	79 (273)	2.2*e*-16
Small nucleolar ribonucleoprotein complex	5732	11 (31)	6.4*e*-09

[a] Enrichment was calculated by the hypergeometric distribution, with Bonferroni correction for number of complexes tested (167).
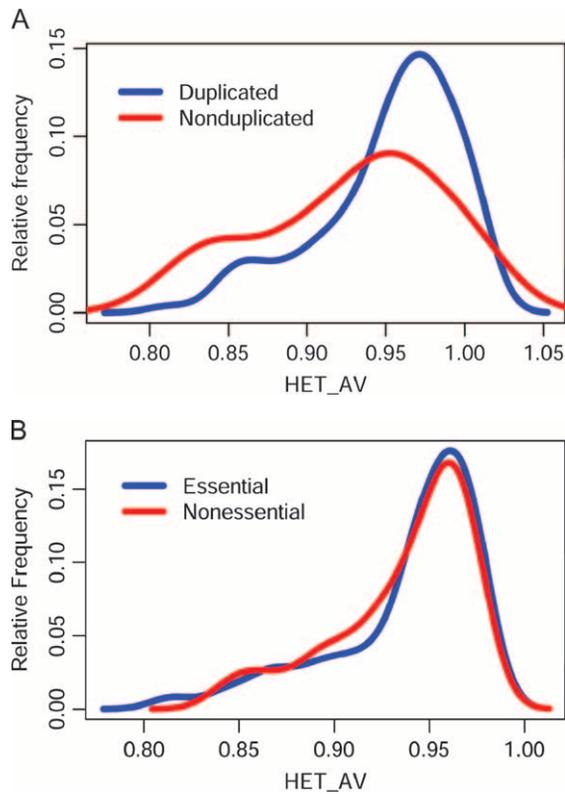
FIGURE 4.—Degree of haploinsufficiency for duplicated ribosomal and essential genes. (A) Frequency plot of median HET_AV values for 19 nonduplicated ribosomal genes and 59 duplicated ribosomal genes. (B) Frequency plot of median HET_AV values for 98 essential haploinsufficient genes and 86 nonessential haploinsufficient genes.
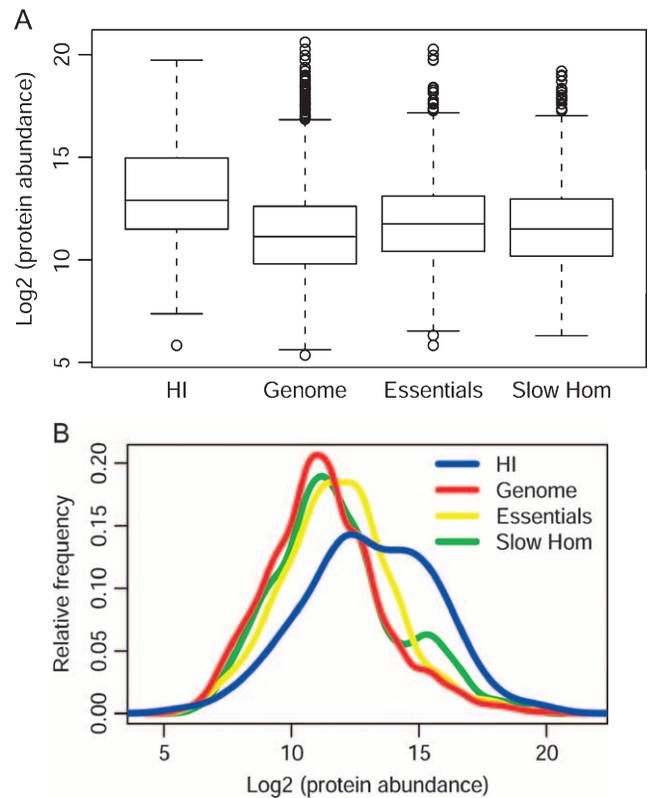

FIGURE 5.—Haploinsufficient genes are highly expressed. (A) Mean protein abundance of different gene classes defined by fitness in YPD: haploinsufficient, essential, and slow-growth homozygous. Only nonzero protein abundance values derived from a whole-genome study (GHAEMMAGHAMI *et al.* 2003) were analyzed. Haploinsufficient genes are more highly expressed when compared to the genome (*t*-test, $P < 2.2e$-16), essential genes (*t*-test, $P < 1.4e$-8), and slow-growth homozygous genes (*t*-test, $P < 2.9e$-8). (B) Frequency plot of observed protein abundances in the same gene classes as in A.

the essential genes is virtually identical to those of the nonessential genes (mean HET_AV for both classes = 0.94, *t*-test, $P = 0.78$, Figure 4B).

Haploinsufficient genes are highly expressed: Heterozygous deletion strains may exhibit a deleterious phenotype because their gene products are required at high levels in actively dividing cells. To address this hypothesis we examined two genome-wide data sets: mRNA expression levels (see MATERIALS AND METHODS) and protein expression levels (GHAEMMAGHAMI *et al.* 2003). In general, the protein levels of haploinsufficient genes are higher than those of (1) the essential genes, (2) the nonessential genes conferring slow growth as homozygotes, and (3) the entire genome (Figure 5A). A similar trend is observed for mRNA abundance where a fivefold higher mean mRNA expression level is observed for haploinsufficient genes compared to the genome as a whole (*t*-test, $P = 1e$-127).

Our finding that most ribosomal components are haploinsufficient, combined with the fact that the ribosomal genes are transcribed at a high level (WARNER 1999), prompted us to ask if the ribosomal genes are distinct in their requirement for high levels of protein expression. This appears to be the case as we observe

a bimodal distribution in a frequency plot of median protein abundance for haploinsufficient genes, where the peak of lower height is due to the ribosomal genes (Figure 5B; this bimodality due to ribosomal genes is also apparent in the slow-growth homozygous strains). These ribosomal genes are not the sole determinant for the observation that haploinsufficient genes are expressed at high protein levels, however, as nonribosomal haploinsufficient genes are also highly expressed compared to the genome in general (Mann-Whitney *U*-test, $P = 0.0013$).

Complementation of haploinsufficiency by overexpression: We overexpressed haploinsufficient genes to complement the haploinsufficient growth defect in individual strains to ensure that a single-gene deletion is responsible for the observed phenotype. Further, to test the validity of the balance hypothesis, we tested these strains for growth defects. Each of 16 essential genes identified as haploinsufficient that interact in com-
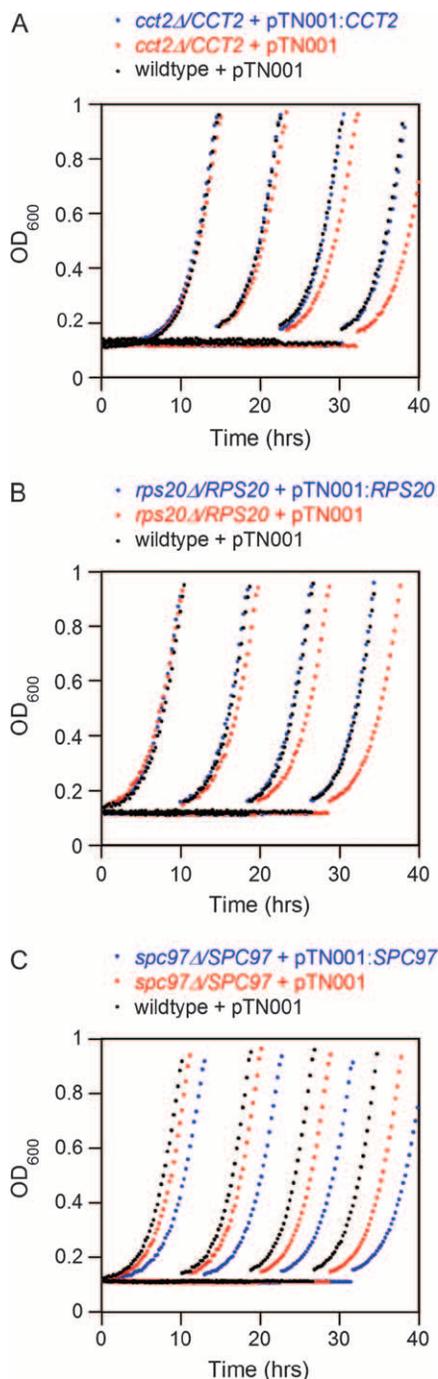
FIGURE 6.—Overexpression of haploinsufficient genes. (A) High-copy overexpression of the *CCT2* gene complements the haploinsufficiency of a *cct2Δ/CCT2* heterozygous deletion strain. Plotted are 20-generation growth curves performed identically to those described in Figure 2B for three strains: wild type with pTN001, a *cct2Δ/CCT2* heterozygous strain containing pTN001, and a *cct2Δ/CCT2* heterozygous strain carrying a pTN001:*CCT2* overexpression plasmid. Overexpression of *CCT2* in a wild-type background does not cause a growth defect (data not shown). (B) The same as in A for *RPS20A*. Overexpression of *RPS20A* in a wild-type background does not cause a growth defect (data not shown). (C) The same as in A for *SPC97*. Overexpression of *SPC97* in a wild-type background results in a severe growth defect (data not shown).

plexes with other proteins was overexpressed from its native promoter on a 2-μm vector with a dominant drug marker. This strategy allowed: (1) physiologically relevant transcription from the endogenous promoter, (2) high-copy overexpression above physiological levels, and (3) growth in YPD for accurate comparisons to the haploinsufficiency data. These strains were then subject to precise 20-generation growth curves as shown in Figure 2B. For 13 genes, the high-copy overexpression under the native promoter complemented the growth defect of the heterozygous strain and did not result in a growth defect after 20 generations of growth (Figure 6, A and B). In addition, for these 13 genes overexpression in a wild-type background did not lead to a phenotype. Genes in this class include components of the large ribosome subunit (*RPL17A* and *RPL18A*), small ribosome subunit (*RPS15* and *RPS20*), CCT folding chaperone (*CCT2*, *CCT4*, and *CCT6*), RNA polymerase II core complex (*RPB3* and *RPB7*), the translation initiation factor *SUI2*, the transcriptional regulator *RVB2*, and genes involved in the biogenesis of the large ribosomal subunit (*RLP24* and *NOG1*). In contrast to the above cases, for 3 genes (*SPC97*, *TUB1*, and *ACT1*), overexpression was toxic (Figure 6C), consistent with published observations (Schatz *et al.* 1986; Drubin *et al.* 1993).

Fitness profiling in minimal media alleviates the haploinsufficiency associated with YPD: The growth rate of yeast in culture depends on favorable environmental conditions and on available nutrients. In the case of rich media (YPD), the generation time reflects the optimal growth rate of yeast. Such optimal, fast cell cycle transits could emphasize the requirement for particular genes, such as ribosomal genes, which could become rate limiting. To test whether slowing the growth rate of the culture might eliminate some of the YPD haploinsufficiency, we profiled both the heterozygous deletion collection and homozygous deletion collection in minimal media where the generation time of wild-type yeast is approximately twice that in YPD. The positive correlation observed in a plot of fitness of the 184 haploinsufficient strains grown in YPD as a function of their fitness in minimal media indicates that the majority of the strains (136 genes, see materials and methods for criteria) are either partially or completely relieved of their haploinsufficiency when grown in minimal media. Moreover, the strains most likely to remain haploinsufficient in minimal media generally have a more severe growth defect in YPD (Figure 7A, Pearson correlation, $P = 2.2e$-16).

Growth in minimal media could alleviate the observed YPD haploinsufficiency by a reduced requirement for gene product. At the transcript level, a significant expression change ($P < 0.05$) was observed for 30 of the 184 haploinsufficient genes (see materials and methods and Figure 7B). The majority of these 30 genes (27) are repressed in minimal media relative to YPD. These results suggest that many genes that confer
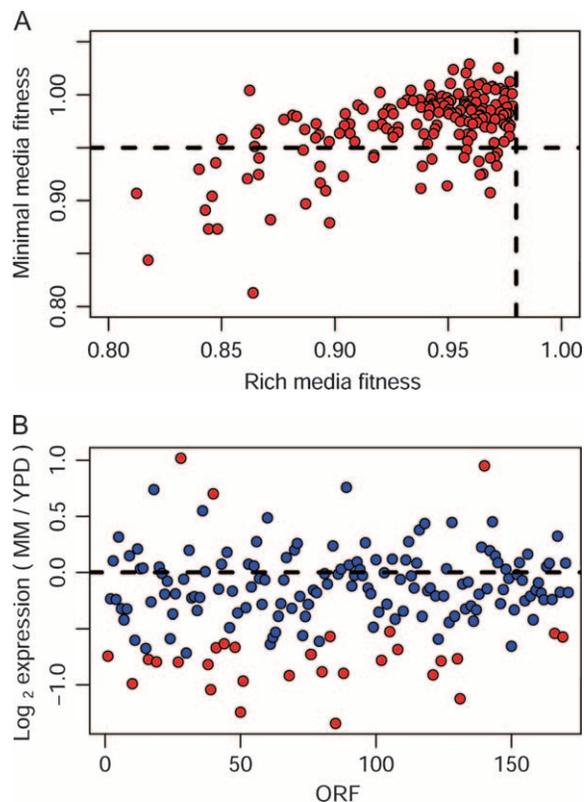
FIGURE 7.—Fitness profiling of heterozygous deletions in minimal media. (A) Plot of minimal media (MM) fitness values for 172 YPD haploinsufficient genes *vs.* their rich media (YPD) fitness values. Minimal media fitness values were not obtained for 12 YPD haploinsufficient genes. Criteria for YPD haploinsufficiency (HET_AV <0.98) and minimal media haploinsufficiency (fitness <0.95) are shown as dashed lines. (B) Log_2 expression ratio (minimal media/YPD) as a function of ORF name (ordered alphabetically) for 172 YPD haploinsufficient genes. Expression ratios in red are statistically significant ($P <$ 0.05).

haploinsufficiency in YPD produce sufficient transcript from a single-gene copy for optimal growth under minimal media conditions.

Thirty-three genes appear haploinsufficient only in minimal media. However, individual examination of these strains revealed that the majority were either false negatives in the YPD experiments (*FIT3, MED8, ARP9,* and *UGO1*) or false positives in the minimal media experiments (*INO2, YRB1, SPT20, CHS2,* and *PHO85*). Thus minimal media-specific haploinsufficiency is extremely rare. Indeed only a single gene, *FUR4* (a uracil transporter), showed minimal media-specific haploinsufficiency.

DISCUSSION

We used yeast as a model organism to perform a genome-wide survey of haploinsufficiency, a basic and increasingly important biological phenomenon. We assayed the relative growth of all strains as heterozygous and homozygous deletions using parallel fitness profiling (GIAEVER *et al.* 2002) and identified 184 genes (∼3% of the genome) as haploinsufficient and 891 genes (or ∼20% of the genome) that exhibit slow growth as homozygotes. The 184 genes likely represent a robust estimate of the total number of haploinsufficient genes in the yeast genome during growth in YPD because (1) independently constructed pools of both the heterozygous and homozygous collections were used, (2) robotic handling of all experiments ensured robust reproducibility and reduced statistical error, (3) individual growth assays verified the results obtained from array analysis even for the most subtle growth differences, and (4) strains with secondary mutations were eliminated from our analyses (these defective strains are currently being reconstructed; R. W. DAVIS, J. BOEKE and M. SNYDER, personal communication).

The haploinsufficient genes are overwhelmingly involved in core metabolic processes, carried out by molecular complexes such as the ribosome. Few genes encoding enzymes are haploinsufficient. This is somewhat unexpected because haploinsufficient genes tend to be highly expressed, and many of the glycolytic enzymes are expressed at very high levels in YPD and, likewise, many biosynthetic enzymes are highly expressed in minimal media. However, while many haploinsufficient genes are highly expressed, the converse is not the case. This observation is addressed by the metabolic theory of dominance (MTD), which argues that most mutations are recessive because metabolic pathways are adequately buffered from quantitative changes in any one enzyme in the pathway (KACSER and BURNS 1981). Although this theory was developed largely on single enzymatic pathways, our application to the entire genome supports the MTD.

An important question arising from our observations is whether or not haploinsufficiency is condition dependent. Although the significant overlap between the YPD and minimal media data sets might argue that haploinsufficiency is not condition dependent, there is ample evidence of condition-specific haploinsufficiency induced by stress. This is clearly the case for drug-induced haploinsufficiency (GIAEVER *et al.* 2004; LUM *et al.* 2004) as well as for other stress conditions such as salt and high pH (G. GIAEVER, unpublished results).

The underlying cause(s) of haploinsufficiency is under question (VEITIA 2002). Our goal was to distinguish between two hypotheses, the balance hypothesis (PAPP *et al.* 2003) and the insufficient amounts hypothesis. Our analysis of mRNA and protein expression levels revealed that haploinsufficient genes are, on average, more highly expressed when compared to all analyzed gene classes, suggesting that these genes are needed at abnormally high levels and are therefore more sensitive to a reduction in gene dosage. The balance hypothesis argues that the relative amounts of complex subunits are under tight regulation and overexpression of haploinsufficient genes is more likely to lead to a fitness defect

(Papp *et al.* 2003). We find that for 13 genes, including components of the large cytosolic ribosome subunit, the small cytosolic ribosome subunit, CCT folding chaperone, and RNA polymerase II core complex, high-copy overexpression is not deleterious. Moreover, growth in minimal media, in which the generation time is approximately double that in YPD, alleviates most YPD haploinsufficiency. We have also observed that this is the case for other conditions that increase the doubling time such as high pH, high osmolarity, and a variety of drugs (Giaever *et al.* 2004; Lum *et al.* 2004; our unpublished results). As these experiments are also performed in YPD, it suggests that the alleviation of haploinsufficiency in minimal media is not due to differences in the nutrient environment, but rather the population doubling time. Under the balance hypothesis, we would expect genes haploinsufficient in YPD to maintain haploinsufficiency in minimal media because these complexes would still be poisoned by subunit imbalance regardless of the growth condition. In addition, the alleviation of haploinsufficiency in minimal media is associated with a repression of gene expression in minimal media relative to YPD, suggesting that absolute transcript level is critical. Taken together, these data suggest that the majority of haploinsufficiency in yeast is due to insufficient amounts of protein. The exceptions to this are the cytoskeletal genes *ACT1*, *TUB1*, and *SPC97*. For these genes, our results and other published reports suggest that the balance hypothesis best explains their haploinsufficiency. Indeed, both *ACT1* and *SPC97* maintain haploinsufficiency in minimal media. In addition, overexpression of *ACT1*, *TUB1*, and *SPC97* in an otherwise wild-type background leads to severe fitness deficiency. Interestingly, these three genes are involved in cytoskeletal function, demonstrating that this cellular component is sensitive to gene dosage in a manner distinct from other haploinsufficient complexes. The toxic consequences of unassembled cytoskeletal proteins are reflected by the evolution of cellular mechanisms to prevent imbalance [*e.g.*, the binding of free β-tubulin by the chaperone Rbl2p (Abruzzi *et al.* 2002)].

The continued value of yeast as a model organism to address basic biological questions is, in part, a reflection of its ability to provide insights into the functions of gene products in higher eukaryotes. A total of 107 of the 184 (59%) haploinsufficient genes (in YPD) have homologs in humans (Wall *et al.* 2003). All complexes that are haploinsufficient in yeast (Table 1) are present in humans. The importance of ribosomal haploinsufficiency in multicellular eukaryotes is illustrated by the *Minute* mutations of Drosophila (Cramton and Laski 1994; Enerly *et al.* 2003) that lead to a variety of developmental abnormalities. In zebrafish, a recent study revealed that haploinsufficiency in 11 ribosomal genes led to an increase in tumor formation (Amsterdam *et al.* 2004). In addition, haploinsufficiency of structural proteins (Andersen *et al.* 2004), signaling factors (How-

ard *et al.* 2004), and cell cycle tumor suppressor regulators (Lam *et al.* 2004) has been directly implicated in a number of inherited mammalian diseases. Comparing our genome-wide study of haploinsufficiency in yeast to more complex metazoans will provide a valuable reference genome for understanding newly discovered diseases that arise as a consequence of haploinsufficiency.

LITERATURE CITED

Abruzzi, K. C., A. Smith, W. Chen and F. Solomon, 2002   Protection from free beta-tubulin by the beta-tubulin binding protein Rbl2p. Mol. Cell. Biol. **22:** 138–147.

Amsterdam, A., K. C. Sadler, K. Lai, S. Farrington, R. T. Bronson *et al.*, 2004   Many ribosomal protein genes are cancer genes in zebrafish. PLoS Biol. **2:** E139.

Andersen, P. S., O. Havndrup, H. Bundgaard, L. A. Larsen, J. Vuust *et al.*, 2004   Genetic and phenotypic characterization of mutations in myosin-binding protein C (MYBPC3) in 81 families with familial hypertrophic cardiomyopathy: total or partial haploinsufficiency. Eur. J. Hum. Genet. **12:** 673–677.

Birrell, G. W., G. Giaever, A. M. Chu, R. W. Davis and J. M. Brown, 2001   A genome-wide screen in Saccharomyces cerevisiae for genes affecting UV radiation sensitivity. Proc. Natl. Acad. Sci. USA **98:** 12608–12613.

Brachmann, C. B., A. Davies, G. J. Cost, E. Caputo, J. Li *et al.*, 1998   Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast **14:** 115–132.

Chial, H. J., T. H. Giddings, Jr., E. A. Siewert, M. A. Hoyt and M. Winey, 1999   Altered dosage of the Saccharomyces cerevisiae spindle pole body duplication gene, NDC1, leads to aneuploidy and polyploidy. Proc. Natl. Acad. Sci. USA **96:** 10200–10205.

Cramton, S. E., and F. A. Laski, 1994   *string of pearls* encodes Drosophila ribosomal protein S2, has *Minute*-like characteristics, and is required during oogenesis. Genetics **137:** 1039–1048.

Deutschbauer, A. M., R. M. Williams, A. M. Chu and R. W. Davis, 2002   Parallel phenotypic analysis of sporulation and postgermination growth in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA **99:** 15530–15535.

Drubin, D. G., H. D. Jones and K. F. Wertman, 1993   Actin structure and function: roles in mitochondrial organization and morphogenesis in budding yeast and identification of the phalloidin-binding site. Mol. Biol. Cell **4:** 1277–1294.

Dwight, S. S., M. A. Harris, K. Dolinski, C. A. Ball, G. Binkley *et al.*, 2002   Saccharomyces Genome Database (SGD) provides secondary gene annotation using the gene ontology (GO). Nucleic Acids Res. **30:** 69–72.

Enerly, E., J. Larsson and A. Lambertsson, 2003   Silencing the Drosophila ribosomal protein L14 gene using targeted RNA interference causes distinct somatic anomalies. Gene **320:** 41–48.

Fox, J., 2002   *An R and S-PLUS Companion to Applied Regression.* Sage Publications, Thousand Oaks, CA.

Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle *et al.*, 2003   Global analysis of protein expression in yeast. Nature **425:** 737–741.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002   Functional profiling of the Saccharomyces cerevisiae genome. Nature **418:** 387–391.

Giaever, G., P. Flaherty, J. Kumm, M. Proctor, C. Nislow *et al.*, 2004   Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc. Natl. Acad. Sci. USA **101:** 793–798.

Goldstein, A. L., and J. H. McCusker, 1999   Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast **15:** 1541–1553.

Goss, K. H., M. A. Risinger, J. J. Kordich, M. M. Sanz, J. E.

STRAUGHEN et al., 2002   Enhanced tumor formation in mice heterozygous for Blm mutation. Science **297:** 2051–2053.

GUTHRIE, C., and G. R. FINK (Editors), 1991   *Guide to Yeast Genetics and Molecular and Cell Biology.* Academic Press, San Diego.

HAZBUN, T. R., L. MALMSTROM, S. ANDERSON, B. J. GRACZYK, B. FOX et al., 2003   Assigning function to yeast proteins by integration of technologies. Mol. Cell **12:** 1353–1365.

HOWARD, J. K., B. J. CAVE, L. J. OKSANEN, I. TZAMELI, C. BJORBAEK et al., 2004   Enhanced leptin sensitivity and attenuation of diet-induced obesity in mice with haploinsufficiency of Socs3. Nat. Med. **10:** 734–738.

JOHNSON, D., S. W. HORSLEY, D. M. MOLONEY, M. OLDRIDGE, S. R. TWIGG et al., 1998   A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. Am. J. Hum. Genet. **63:** 1282–1293.

KACSER, H., and J. A. BURNS, 1981   The molecular basis of dominance. Genetics **97:** 639–666.

LAM, M. H., Q. LIU, S. J. ELLEDGE and J. M. ROSEN, 2004   Chk1 is haploinsufficient for multiple functions critical to tumor suppression. Cancer Cell **6:** 45–59.

LI, C., and W. H. WONG, 2001   Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. USA **98:** 31–36.

LUM, P. Y., C. D. ARMOUR, S. B. STEPANIANTS, G. CAVET, M. K. WOLF et al., 2004   Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell **116:** 121–137.

MCMURRAY, M. A., and D. E. GOTTSCHLING, 2003   An age-induced switch to a hyper-recombinational state. Science **301:** 1908–1911.

MNAIMNEH, S., A. P. DAVIERWALA, J. HAYNES, J. MOFFAT, W. T. PENG et al., 2004   Exploration of essential gene functions via titratable promoter alleles. Cell **118:** 31–44.

MUROYA, K., T. HASEGAWA, Y. ITO, T. NAGAI, H. ISOTANI et al., 2001   GATA3 abnormalities and the phenotypic spectrum of HDR syndrome. J. Med. Genet. **38:** 374–380.

OOI, S. L., D. D. SHOEMAKER and J. D. BOEKE, 2001   A DNA microar-ray-based genetic screen for nonhomologous end-joining mutants in Saccharomyces cerevisiae. Science **294:** 2552–2556.

PAPP, B., C. PAL and L. D. HURST, 2003   Dosage sensitivity and the evolution of gene families in yeast. Nature **424:** 194–197.

PLANTA, R. J., and W. H. MAGER, 1998   The list of cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. Yeast **14:** 471–477.

PRIMIG, M., R. M. WILLIAMS, E. A. WINZELER, G. G. TEVZADZE, A. R. CONWAY et al., 2000   The core meiotic transcriptome in budding yeasts. Nat. Genet. **26:** 415–423.

SCHATZ, P. J., F. SOLOMON and D. BOTSTEIN, 1986   Genetically essential and nonessential alpha-tubulin genes specify functionally interchangeable proteins. Mol. Cell. Biol. **6:** 3722–3733.

SEIDMAN, J. G., and C. SEIDMAN, 2002   Transcription factor haploinsufficiency: when half a loaf is not enough. J. Clin. Invest. **109:** 451–455.

SPRING, K., F. AHANGARI, S. P. SCOTT, P. WARING, D. M. PURDIE et al., 2002   Mice heterozygous for mutation in Atm, the gene involved in ataxia-telangiectasia, have heightened susceptibility to cancer. Nat. Genet. **32:** 185–190.

STEVENS, R. C., and T. N. DAVIS, 1998   Mlc1p is a light chain for the unconventional myosin Myo2p in Saccharomyces cerevisiae. J. Cell Biol. **142:** 711–722.

TONG, A. H., M. EVANGELISTA, A. B. PARSONS, H. XU, G. D. BADER et al., 2001   Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science **294:** 2364–2368.

VEITIA, R. A., 2002   Exploring the etiology of haploinsufficiency. BioEssays **24:** 175–184.

WALL, D. P., H. B. FRASER and A. E. HIRSH, 2003   Detecting putative orthologs. Bioinformatics **19:** 1710–1711.

WARNER, J. R., 1999   The economics of ribosome biosynthesis in yeast. Trends Biochem. Sci. **24:** 437–440.

WINZELER, E. A., D. D. SHOEMAKER, A. ASTROMOFF, H. LIANG, K. ANDERSON et al., 1999   Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science **285:** 901–906.

Communicating editor: M. JOHNSTON