UC Davis
UC Davis Previously Published Works

Title

Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss

Permalink

https://escholarship.org/uc/item/153538kt

Journal

Molecular Systems Biology, 3(1)

ISSN

1744-4292

Authors

Ihmels, Jan
Collins, Sean R
Schuldiner, Maya
et al.

Publication Date

2007

DOI

10.1038/msb4100127

Peer reviewed

Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss

Jan Ihmels[1,2,3,*], Sean R Collins[1,2,3], Maya Schuldiner[1,2,3], Nevan J Krogan[1,2] and Jonathan S Weissman[1,2,3]

[1] Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, USA, [2] The California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA, USA and [3] Howard Hughes Medical Institute, University of California, San Francisco, CA, USA
* Corresponding author. Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, University of California at San Francisco, San Francisco, 1700 4th street, CA 94143-2542, USA. Tel.: + 1 415 502 7642; Fax: + 1 415 514 2073; E-mail: jan.ihmels@gmail.com

Many genes can be deleted with little phenotypic consequences. By what mechanism and to what extent the presence of duplicate genes in the genome contributes to this robustness against deletions has been the subject of considerable interest. Here, we exploit the availability of high-density genetic interaction maps to provide direct support for the role of backup compensation, where functionally overlapping duplicates cover for the loss of their paralog. However, we find that the overall contribution of duplicates to robustness against null mutations is low ($\sim 25\%$). The ability to directly identify buffering paralogs allowed us to further study their properties, and how they differ from non-buffering duplicates. Using environmental sensitivity profiles as well as quantitative genetic interaction spectra as high-resolution phenotypes, we establish that even duplicate pairs with compensation capacity exhibit rich and typically non-overlapping deletion phenotypes, and are thus unable to comprehensively cover against loss of their paralog. Our findings reconcile the fact that duplicates can compensate for each other's loss under a limited number of conditions with the evolutionary instability of genes whose loss is not associated with a phenotypic penalty.
Molecular Systems Biology 27 March 2007; doi:10.1038/msb4100127
Subject Categories: functional genomics
Keywords: duplication; evolution; genetic interactions; redundancy

Introduction

Much of our understanding of biological processes has been derived from the characterization of the functional consequence to an organism of altering one or more of its genes. Recent progress in high-throughput approaches now make it possible to systematize such classical genetic efforts and carry out phenotypic analysis of compromised alleles on a genomic scale. Specifically, deletion libraries for model organisms like the budding yeast *Saccharomyces cerevisiae* and RNAi-based screens in metazoans have greatly facilitated efforts to define gene function (Winzeler *et al*, 1999; Giaever *et al*, 2002; Steinmetz *et al*, 2002; Kamath *et al*, 2003). The ability to measure the phenotypic consequences of gene deletions on a genomic scale has also provided a broad range of systems-level insights, including the link between network connectivity and essentiality (centrality–lethality) in protein interaction networks (Jeong *et al*, 2001) as well as the relationship between essentiality and cell-to-cell variability (noise) (Fraser *et al*, 2004; Newman *et al*, 2006). Finally, the quantitative cost of gene loss has far-reaching implications for evolutionary theory, including the connection between gene importance, rates of evolution (Hirsh and Fraser, 2001) and patterns of conservation among related phyla (Krylov *et al*, 2003).

Efforts to systematically evaluate the phenotypic effects of gene loss, however, have been hampered by the fact that the disruption of most genes has surprisingly modest effects on cell growth and viability. In *S. cerevisiae*, for example, less than 20% of genes are essential, and the large majority of the remaining deletions have little or no detectable effect on growth in rich media (Giaever *et al*, 2002). Similar observations have been made for other eukaryotes (Kamath *et al*, 2003) and prokaryotes (Kobayashi *et al*, 2003). The high proportion of genes with no apparent deletion effect has wide-ranging practical and theoretical implications and has been the subject of considerable interest (Wagner, 2000, 2005; Giaever *et al*, 2002; Gu *et al*, 2003; Papp *et al*, 2004; Kafri *et al*, 2005).

One factor that has been implicated as contributing to the high degree of dispensability is the abundance of closely related paralogs present in most genomes (Winzeler *et al*, 1999; Wagner, 2000; Giaever *et al*, 2002). Indeed, recent work in *S. cerevisiae* has shown that the existence of a paralog elsewhere in the genome significantly increases the chance that deletion of a given gene has little effect on growth (Gu *et al*, 2003). The prevailing explanation for this excess dispensability among duplicates is that it is due to backup compensation in which duplicate genes with overlapping functionality cover for the loss of their paralogous partner gene.

While excess dispensability of duplicates compared to singletons is well documented, the magnitude and underlying mechanism of such effects remain unclear. Specifically, backup compensation is only one possible way to explain the observed difference in mutant fitness between duplicates

and single-copy genes. A recent study, for example, suggests that the difference arises because genes with severe deletion phenotypes are less likely to have undergone duplication or have their duplicates retained (He and Zhang, 2006a). Another possibility is that specialization following the duplication event may have allowed paralogs to distribute functions among them such that each duplicate is required in a more limited set of conditions than the ancestor gene, as appears to have occurred with ubiquitin ligases (Pickart, 2001) or nuclear import receptors (Nakielny and Dreyfuss, 1999). Analyses to date have been mostly correlative, and direct mechanistic evidence supporting or refuting the role of backup compensation in mutational robustness is still largely missing. Furthermore, backup between duplicates is not easily justified in evolutionary terms, in that a genuine ability to comprehensively cover for the loss of another gene is evolutionarily unstable (Brookfield, 1992). Finally, even if one accepts the prevailing model of backup compensation, current estimates for the contribution of duplicates to robustness against deletions cover a wide range ($\sim$20–60%) (Gu *et al*, 2003), and perhaps less (Papp *et al*, 2004).

Recently, two approaches (synthetic genetic arrays (SGA) and diploid-based synthetic lethality analysis on microarrays (dSLAM)) have been developed to identify synthetic sickness/ lethal (SSL) relationships in *S. cerevisiae* by systematic generation of double mutant strains (Tong *et al*, 2001, 2004; Pan *et al*, 2006). These large-scale techniques provide a unique opportunity to address these issues directly. Genetic interactions quantify the extent to which the phenotype of mutating one gene is modulated by the absence or presence of another. In particular, SSL interactions occur between genes whose simultaneous deletion phenotype is much stronger than expected from the two single deletions. A clear prediction of the backup model is therefore that buffering duplicates should exhibit a strong SSL interaction with their paralog. Using two recent data sets of high-density epistatic mini-array profiles (E-MAPs) (Schuldiner *et al*, 2005; Collins *et al*, 2007), we provide direct experimental evidence for duplicate compensation, but find that the contribution of this mechanism to robustness against deletion is close to the lower-bound estimate given previously (Gu *et al*, 2003).

More broadly, the ability to identify the subset of duplicates with buffering capacity allowed us to explore their properties and how they differ from non-buffering paralogs. In particular, we investigate to what extent the capability to mitigate the deletion defect of paralogs in rich media reflects genuine redundancy, where one gene comprehensively covers for the loss of another. To this end, in a distinct use of *EMAP* data, we employ patterns of genetic interactions as high-resolution phenotypes to compare the functional role of buffering duplicates. Strikingly, this more detailed phenotypic readout reveals that even SSL duplicates show rich spectra of genetic interactions with other genes, implying that their ability to provide functional backup is not upheld in the presence of additional deletions. Similarly, deletions of buffered duplicates resulted in measurable growth defects when a larger variety of environments are taken into account. Lastly, patterns of interactions are divergent between duplicates and more similar to non-paralogous genes in general.

Taken together, our results indicate that although a fraction of duplicates can provide buffering compensation in optimal growth environments, in most cases they are functionally divergent and unable to provide genuine backup against the loss of their paralogous copy over a range of compromising conditions, represented here through additional gene deletions as well as environmental perturbations. We discuss the physiological relevance of deletion backgrounds, and demonstrate that synthetic interactions provide an evolutionarily significant deletion phenotype for the majority of genes whose deletion in rich media has little phenotypic effect.

The role of duplicate genes in robustness against deletions

To explore the contribution of duplicates to robustness against deletion, we used two data sets of high-density genetic interaction maps, one centered around genes of the endoplasmic reticulum (ER; Schuldiner *et al*, 2005) and a more recent one of genes involved in chromosome biology (Collins *et al*, 2007). Both sets consist of quantitative measures of alleviating (positive) and aggravating (negative, SSL) interactions between gene pairs. Among the 1136 genes covered, 300 genes were classified by sequence similarity as unambiguously having no paralogs (singletons) and 90 were found to have exactly one duplicate copy (duplicates). To ensure only pairwise interactions, gene families of more than two paralogs were excluded from the analysis.

Duplicates whose lack of a strong deletion phenotype is due to buffering compensation are expected to be synthetically sick or lethal with their corresponding paralog. In line with previous results for a smaller data set (Tong *et al*, 2004), we find that duplicate pairs of genes are significantly more likely to interact negatively than unrelated pairs of genes (Figure 1A). To test whether SSL paralogs can account for the excess fitness of duplicates, we classified genes into fitness categories according to their deletion growth defect (Materials and methods). The subset of genes covered by our combined data set exhibits an over-representation of duplicate genes in the weak/no deletion phenotype (WNP) class similar to that reported previously (Gu *et al*, 2003) (Figure 1B). Specifically, the proportion of genes in the WNP class was 67% for duplicates compared to 47% for singletons. Strikingly, this difference (17 genes) corresponds to the number of WNP duplicates (17 genes) that have an SSL interaction with their corresponding paralog (Figure 1C). Notably, this result is not sensitive to the exact definition of the WNP class (Figure 1D). Similar results are observed when the two data sets of genetic interactions are analyzed separately (Supplementary Figure 1). Although the discrepancy between duplicate and singleton dispensability is substantially different for the two data sets (26% compared to 14%), this difference is closely matched by the number of WNP SSL duplicates in both cases. Furthermore, these results are robust to random subsampling of the data (Supplementary Figure 2).

Most buffering duplicates have a number of SSL interactions in addition to those between the paralogs themselves, each of which could provide backup compensation. Further analysis, however, distinguishes the interactions between duplicate
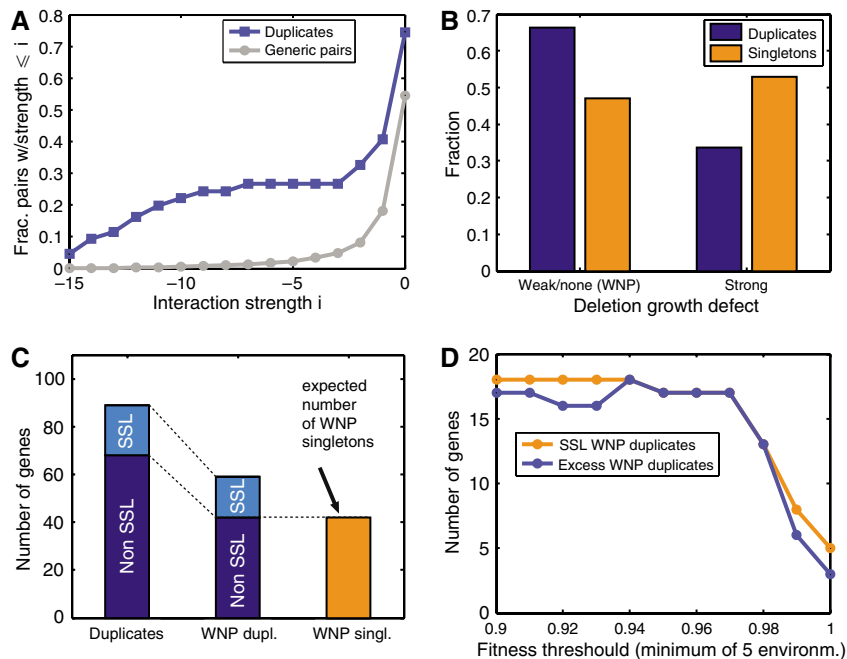
Figure 1 (**A**) Enrichment of duplicates with SSL interactions. Shown is the fraction of duplicates (blue) and random gene pairs with interaction strength less than the threshold value *s*, as a function of *s*. The threshold used in this work to define SSL interactions is $s_{thr}=-3$. (**B**) The subset of duplicates and singletons for which interaction data are available exhibits an excess of duplicate fitness similar to that reported earlier for a genome-wide set by Gu *et al* (2003). Shown is the number of genes assigned to the two fitness classes (Materials and methods), for duplicates and singletons. (**C**) The excess number of duplicate genes in the WNP class compared to singletons corresponds to the number of SSL duplicates in the data set. Shown is the total number of duplicates covered by the genetic interaction data set (left column) and the number assigned to the WNP class (middle column). The number of SSL duplicates is indicated in light blue. The right column shows how many WNP singletons are expected for the same number of genes, based on the proportion of singletons assigned to that class. (**D**) The observed correspondence between excess fitness and the number of backup duplicates remains stable over a range of fitness thresholds defining the WNP class (Materials and methods). Shown is the number of SSL duplicates assigned to the WNP class (orange) and the difference between the observed number of WNP duplicates and the expected number of WNP singletons (blue), as a function of the fitness threshold.

genes. Specifically, the number of interactions between paralogous pairs is highly significant and cannot be attributed to chance or a higher number of interaction partners for duplicates. Assuming a model where each ER duplicate interacts with the duplicate average of 8 genes, the probability of obtaining the observed number of interactions (or more) between duplicates by chance is $P<10^{-13}$ (Materials and methods). Furthermore, the interactions between most SSL duplicates are significantly stronger in magnitude than the remaining interactions with other genes (Supplementary Table 1).

Our data thus provide direct evidence that it is indeed duplicate compensation that accounts for the observed difference in deletion growth defect between duplicates and singletons, at least for the genes covered by our data set. However, out of 59 WNP duplicates, only 17 (29%) are SSL with their paralog. Assuming that the observations made for our data set hold on a genomic scale, this suggests that the contribution of duplication to overall robustness against deletions is close to the lower-bound estimation (23%) given previously (Gu *et al*, 2003).

Can duplicates provide genuine backup?

Apart from the mechanism itself, the characteristic features of buffering duplicates have received considerable attention

(Gu *et al*, 2003; Kafri *et al*, 2005; Wagner, 2005). For example, it might be expected that buffering duplicates have diverged less in key properties than those that are unable to provide backup (Wagner, 2005). Other reports have suggested that backup pairs are typically not coexpressed but rely on feedback causing the upregulation of the duplicate (Kafri *et al*, 2005). Also in these studies, conclusions drawn from correlative arguments, based on the enrichment with dispensable genes, have been indirect and subject to debate (Wong and Roth, 2005; He and Zhang, 2006b). In contrast, our data allowed us to unambiguously distinguish the subset of duplicates whose dispensability can be attributed to the existence of a backup paralog. The ability to identify backup duplicates directly put us in a position to study their features, and how they differ from other duplicates without buffering properties. In particular, we asked to what extent the observed buffering in rich media reflects functional similarity and a genuine ability to cover for the loss of a paralog in a broader range of conditions.

Immediately following a duplication event, duplicate gene pairs are expected to be fully redundant, before undergoing divergence in sequence, function and regulation. The capability of some duplicates to compensate for each other's deletion in rich media suggests that for SSL duplicates, at least some functional overlap has been retained long after the duplication event. Indeed, in line with earlier results (Gu *et al*, 2003), we find that buffering genes are on average more
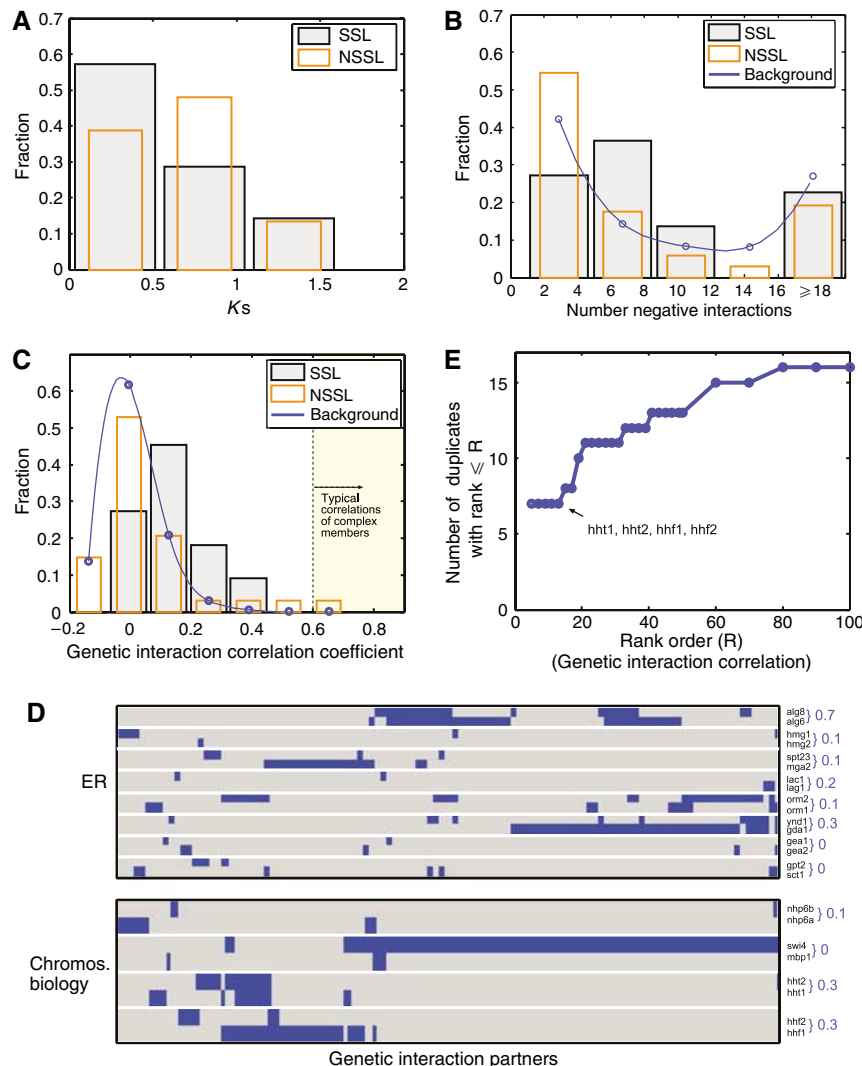
Figure 2 (**A**) Backup (WNP) SSL duplicates have more similar sequences than non-SSL duplicates. Shown is the distribution of non-synonymous substitution rates k_a for both sets of genes. A corresponding plot for the same measure normalized by the rate of silent substitutions is shown in Supplementary Figure 3. (**B**) Backup SSL duplicates have no less negative interactions than non-SSL duplicates and generic pairs of genes. No significant difference in the distributions of the number of interactions was found between the three groups (Kolmogorov–Smirnov test, $P > 0.4$ and $P > 0.1$). The number of interactions was normalized between the two data sets (Materials and methods). (**C**) Genetic correlation coefficients were evaluated as described in Materials and methods and by Schuldiner *et al* (2005). Shown are histograms of the distributions associated with SSL backup genes, non-SSL duplicates and random pairs of genes. The distributions between SSL and non-SSL duplicates are significantly different (Kolmogorov–Smirnov test $P < 0.002$). (**D**) Genetic interactions of SSL duplicate pairs. Blue boxes in each lane indicate genes with an SSL interaction. Numbers next to the gene pairs represent the Pearson correlation coefficient of their genetic interaction profiles. The two matrices correspond to the two interaction data sets used. See Supplementary Table 2 for a list of specific and common interaction partners for each duplicate pair. An example of duplicates that are highly correlated in their genetic patterns but perform different functions is provided by alg6 and alg8, which are performing different functions within the same pathway. It is interesting to note that the interaction strength between these is significantly lower than for the remaining SSL duplicates (below the threshold of -3 used in this study), and that at least one of the genes (alg6) has a detectable deletion growth defect. (**E**) Duplicates are less correlated in their patterns of genetic interactions with their paralog than with other genes in the data set. For each duplicate, correlation coefficients between its epistatic profile and of each of the remaining genes in the data set were calculated and the resulting coefficients were rank-ordered. The rank R represents the rank of the correlation with the corresponding paralog in this sequence, for example, $R = 3$ if the correlation with the duplicate copy was the third highest. Shown is the number of buffering duplicates for which the rank is at most R, as a function of R.

similar in sequence than non-SSL duplicates (Figure 2A). The extent of functional overlap between them, however, remains unclear.

To assess the extent to which SSL duplicates can provide genuine backup under compromising conditions, we used genetic interaction profiles as a more stringent test for redundancy that assesses the effect of gene loss in the background of additional gene deletions. Duplicates that are

comprehensively covered by backup paralogs should be characterized by a very small number of negative interactions, as in the presence of full backup their loss should have little phenotypic effect. In the extreme case of perfect redundancy, the only expected interaction is between the duplicates themselves. In striking contrast, we find that SSL duplicates have a substantial number of synthetic interactions that often exceeds that of random genes and non-SSL duplicates

(Figure 2B). Thus, the backup capacity of SSL duplicates is limited and not indicative of a comprehensive ability to cover for the loss of the paralogous partner.

We next used patterns of genetic interactions to test the degree of functional similarity of buffering duplicates. Correlated interaction profiles have been shown to be a strong indicator of shared functionality (Tong *et al*, 2004; Schuldiner *et al*, 2005; Pan *et al*, 2006). However, in spite of their rich media buffering properties, we find that the interaction patterns of most SSL duplicates are divergent (Figure 2C and D) and profile correlations between them are usually exceeded by correlations with other, non-paralogous genes (Figure 2E). For example, the highest correlation coefficient between SSL duplicates in our data set was $c=0.3$ for histones (see below). In contrast, genes coding for constituents of the same complex (whose deletion therefore has highly similar effects) typically reach much higher correlation values (Figure 2C) (Schuldiner *et al*, 2005; Collins *et al*, 2007). This suggests that even duplicates that are substantially different in function or regulation can nevertheless provide some backup at least in standard laboratory conditions. Conversely, other duplicates with less divergent properties lack buffering capability.

Role of duplicates in dosage amplification

Although the epistatic profiles of most duplicates are divergent, a notable exception is provided by the four histone genes hht1/hht2 and hhf1/hhf2 (Figure 2D), whose patterns of interactions are significantly correlated ($P<10^{-21}$ and $P<10^{-12}$, respectively). While the majority of duplicates with distinct sets of genetic interactions are likely fixed in the population because of divergent function and/or regulation, such functionally similar paralogs may be retained because their gene product is required at high abundance (Figure 3A). In this scenario, both genes should be subject to similar regulation in addition to their correlated interaction profiles. They should also be abundant. To test this, we used a database of >1000 expression profiles measured across a variety of cellular conditions (Ihmels *et al*, 2002), as well as data from a genome-wide study of protein abundance (Ghaemmaghami *et al*, 2003). Supporting their role in dosage amplification, we find that, in contrast to other pairs of buffering duplicates, histones are indeed strongly correlated in their expression patterns and expressed at high copy number (Figure 3B). Even in this case, however, both duplicates have a number of specific genetic interactions, suggesting a degree of functional diversification even between these proteins. This is consistent with a related recent analysis of another pair of coexpressed and abundant histones (hta1–htb1 and hta2–htb2) (Libuda and Winston, 2006).

As these observations are based on the limited and specific subset of duplicates that are represented in our genetic interaction data sets, we asked if a similar correlation between protein abundance and coexpression holds on a genomic scale. To this end, we plotted protein abundance values for the full set of duplicates and singletons in the genome (Figure 3C). As had been observed previously for mRNA copy number (Seoighe and Wolfe, 1999), we find that duplicates are significantly enriched in the high-abundance regime. We then divided duplicates into low- and high-abundance classes, and

considered the distribution of expression correlations for both separately. Remarkably, we find that most abundant duplicates have highly similar expression patterns (Figure 3D).

Apart from absolute copy number, duplicates that are simultaneously required for reasons of amplification might also be expected to be expressed at similar levels. In support of this, we find that abundance levels are significantly more similar between duplicates (and in particular SSL duplicates) than random gene pairs (Figure 3E). Consistent with our hypothesis, this effect is especially pronounced for duplicates with correlated expression patterns (Figure 3F).

Rich deletion phenotypes based on genetic interaction spectra

Our finding that even buffering duplicates have rich deletion phenotypes in the presence of additional mutations prompted us to use genetic interaction spectra more broadly to assign deletion phenotypes to generic genes whose deletion in standard conditions has little measurable effect. Apart from duplicate compensation, the effectiveness of single deletions is limited by other compensation mechanisms like distributed robustness (backup pathways) (Wagner, 2000, 2005), or condition-specific gene requirement (Papp *et al*, 2004). Systematic exploration of growth defects in double deletions is capable of overcoming these limitations. First, genes buffered through alternative pathways are synthetically sick or lethal with members of these pathways. This is one of the classical ideas underlying the study of genetic interactions. Second, many cellular conditions are characterized by specific stresses or availability of substrates that can be mimicked using genetic perturbations, thus eliciting a phenotype from genes that are not required under rich media conditions. For example, gene deletions of transporter genes in rich media have similar effects as the absence of the corresponding nutrients in a specific environment. Biosynthetic genes that may be dispensable in rich media are essential in these environments, and correspondingly exhibit SSL interactions with transporters. Similarly, cellular stress conditions can be mimicked by deletion backgrounds. For example, genes involved in the unfolded protein response (UPR) display little fitness defect in rich media. However, additional deletion of chaperones or genes involved in glycosylation is lethal, reflecting the requirement of the UPR in conditions that compromise ER folding activity (Figure 4A). Finally, many drugs inhibit specific genes and therefore have an effect similar to that of a deletion.

Genetic interaction profiles are thus expected to reveal a phenotype in many instances for which single gene deletions in rich media would not. Indeed, the interaction spectra of genes covered by our data sets reveal that 80–90% of genes have at least one significant genetic interaction, whereas single gene deletions in five growth environments (Steinmetz *et al*, 2002) yield a detectable growth defect only for 30–40% (Figure 4B). Thus for most genes, there is a substantial cost of gene loss, even though this is often not reflected in single gene deletion tests carried out in rich media. A similar observation was made in bacteria, where the effect of many mutations was found to depend on the environmental and genetic context in which they were tested (Remold and Lenski, 2004).
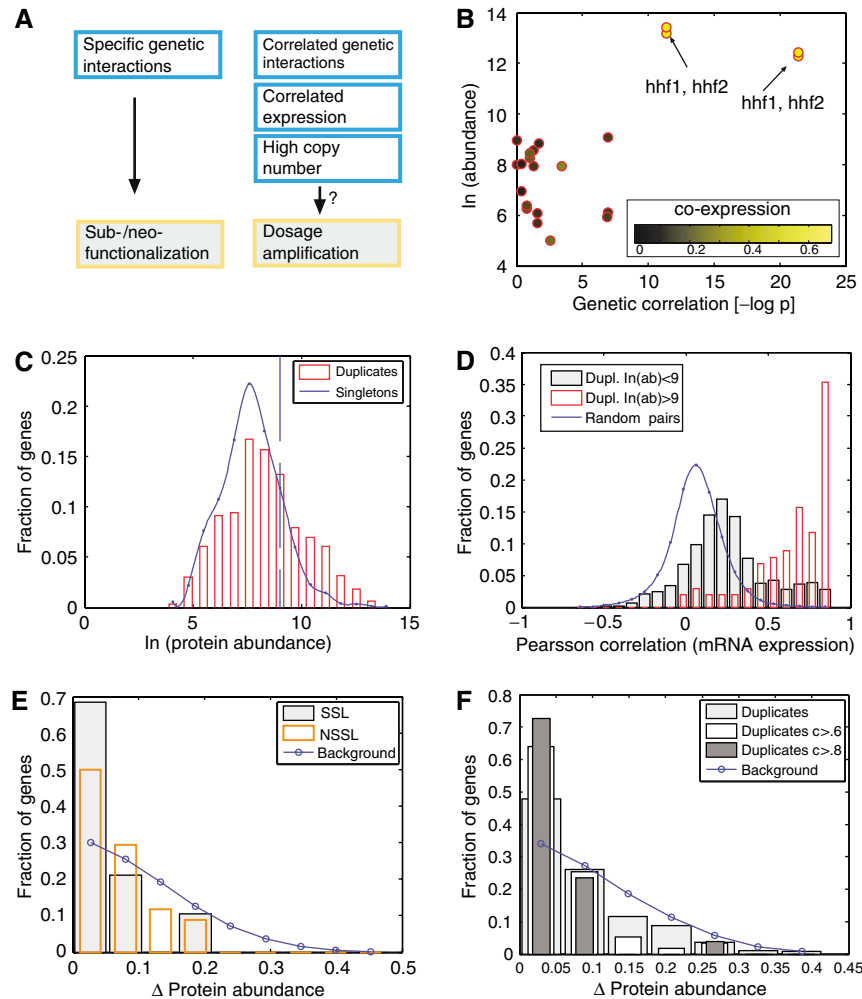
Figure 3 (**A**) Two distinct reasons for duplicate retention: functional divergence and dosage amplification for high copy numbers. (**B**) Relationship between similarity in genetic interaction patterns, protein abundance and mRNA coexpression. Shown is a scatter plot of genetic profile correlation (expressed as *P*-value to correct for different-size data sets) and protein copy numbers for backup SSL duplicates. Points are color-coded according to their expression correlation coefficients, as indicated. (**C**) Distribution of genome-wide protein copy numbers of the full set of duplicates and singletons (>2000 genes). (**D**) Expression profiles of abundant duplicates are significantly correlated (Kolmogorov–Smirnov test $P=10^{-126}$). Duplicates were partitioned by their protein copy numbers, using a cutoff ln(abundance$-$cut)$=9$ (dashed line in (C)). Shown is the distribution of Pearson correlation coefficients between expression profiles of random sets of genes (blue line), abundant duplicates (red bars) and duplicates where at least one paralog is less abundant than the cutoff (gray bars). As in (C), the full set of duplicates and singletons was used. The effect remained qualitatively similar and highly significant ($P=10^{-95}$) after ribosomal proteins were removed from the analysis (Supplementary Figure 4). Distributions for abundant and non-abundant random pairs of genes are shown in Supplementary Figure 5. (**E**) Abundance values of duplicates and in particular backup SSL are significantly more similar than those of unrelated pairs of genes ($P=10^{-4}$). The similarity in protein abundances of a pair of genes is represented by the quantity Δabundance$=(ab(a)-ab(b))/(ab(a)+ab(b))$, where *a* and *b* represent the protein copy numbers of each paralog. Shown are the distributions for random pairs of genes (blue line), backup SSL (gray bars) and NSSL genes (orange bars). (**F**) On a genomic scale, the tendency of duplicates toward similar copy numbers is significant ($P=10^{-13}$) and greatest for duplicates with correlated expression profiles. Shown are the distributions of abundance similarity for generic duplicates (light gray), duplicates whose expression correlation is at least 0.6 (white bars), duplicates whose expression correlation is at least 0.8 (dark gray bars) and random pairs of genes (blue lines). The full set of duplicates and singletons was used. Removal or ribosomal genes from the analysis resulted in similar distributions (Supplementary Figure 6), albeit with lower *P*-value ($<10^{-7}$).

Are genetic interactions indicative of gene importance? To address this question, we asked if there is a relationship between the probability of a gene being retained between related species and its number of genetic interactions (genetic interactivity). As has been noted before, genes that are essential for viability have a higher probability of retention than those with a viable deletion phenotype. However, among non-essential genes, we find that genetic interactivity exhibits a strong correlation with gene retention across related phyla

(Figure 4C and Supplementary Figure 7), and predicts the likelihood of gene loss better than lethality/viability, quantitative growth deficiency or environmental specificity (Supplementary Figure 8). These results suggest that double deletions reveal a cost of gene loss that is physiologically relevant and effectively recapitulates evolutionary constraints.

An alternative way of overcoming environmental specificity is to carry out deletion assays in a larger number of cellular environments and stresses. Indeed, in a recent study,
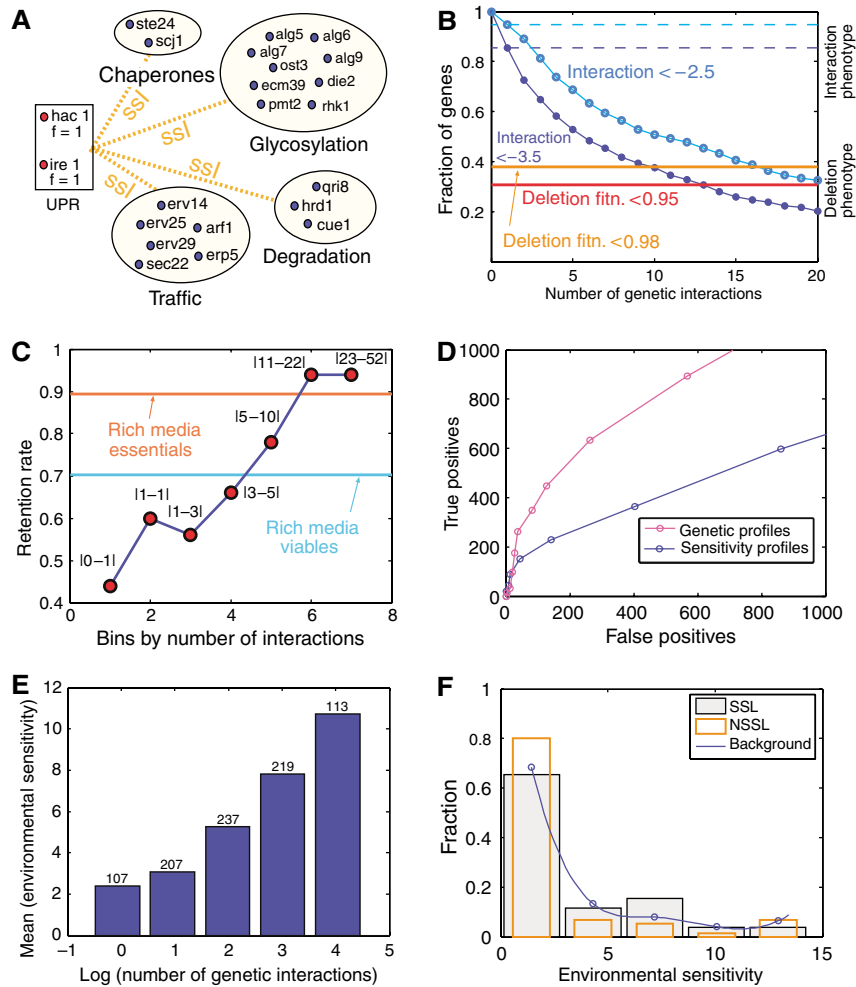
Figure 4 (**A**) Genetic interactions can elicit phenotypes from genes required only in specific conditions. The two genes hac1 and ire1 are inducers of the unfolded protein response, whose deletion has little or no effect on cellular growth rate (deletion fitness $f=1$ in both cases). However, simultaneous deletion of genes affecting protein folding results in a strong growth defect (synthetic interaction). (**B**) Genetic interactions reveal a phenotype for many genes that are missed in single gene deletion assays under the same conditions (rich media). Shown in blue is the fraction of genes with at least × SSL interactions, as a function of *x*. Two different significance cutoffs for negative interactions were used (light and dark blue). The red and orange lines represent the fraction of genes with a deletion growth defect, for two choices of the fitness threshold (red and orange). (**C**) Genetic interactivity (number of negative interactions) correlates with probability of gene retention between *S. cerevisiae* and *C. albicans*. Comparisons with other yeast species produce similar results (Supplementary Figure 7). Genes covered by the ER interaction data set were arranged by the number of negative interactions and partitioned into bins of 50 genes each. The two lines correspond to genes of the same data set that are annotated as either essential or viable in the SGD database. For each bin, the fraction of genes shared between the species is shown. The range of the number of interactions is indicated above next to the data points. Results obtained using the chromosome biology data are similar (data not shown). The correlation between phylogenetic retention and genetic interactivity is stronger than that between retention and quantitative growth defects in rich media or across a range of environments (Supplementary Figure 8; correlations between binned quantities are $r^2=0.93$, $r^2=0.56$ and $r^2=0.89$, respectively). (**D**) Comparison between the ability of sensitivity and genetic interaction profiles to cluster functionally similar genes. Gene associations based on profile similarity were evaluated against GO functional annotations (Materials and methods). The number of correct predictions was plotted against the number of false positives for a range of thresholds. Genes were limited to those assigned to the chromosome biology data set for both methods. (**E**) The number of genetic interactions is related to sensitivity of deletion mutants in response to 51 different drugs and environments. Genes were assigned to bins according to the logarithm of their number of genetic interactions. Each gene is associated with a score representing its combined sensitivity to the different environments (Materials and methods). Shown is the mean sensitivity score of genes assigned to each bin. The number of genes is indicated above the corresponding bars. The Pearson correlation coefficient between the unbinned quantities is $c=0.36$ ($P<10^{-26}$). A similar result is obtained when only naturally occurring environments are considered (Supplementary Figure 9), with a correlation coefficient of $c=0.32$ ($P<10^{-21}$). (**F**) Deletions of SSL duplicates have a comparable effect on growth rate across a range of environments as deletions of non-SSL duplicates and random genes. A similar result is obtained when only naturally occurring environments are considered (Supplementary Figure 6).

sensitivity profiles of deletion strains to a range of agents and environments were shown to provide numerous functional predictions of genes with unknown functions (Brown *et al*, 2006). It is interesting to compare this large data set with our genetic interaction spectra. Although both genetic and

sensitivity profiles successfully cluster functionally related genes, the most comprehensive data set of sensitivity profiles currently available (51 conditions centered around DNA-damaging agents) is not sufficient to give a comparable degree of functional enrichment to that seen with genetic interactions

(Figure 4D). Future studies assaying a larger number of diverse conditions could overcome this limitation and provide similar functional enrichment as that obtained from genetic inter-action profiles on a genome-wide scale. Consistent with our argument, the propensity of a gene to show sensitivity to these drugs and environments increases with the number of its genetic interaction partners (Figure 4E, $P < 10^{-26}$). This correlation is likely because many drugs inhibit specific genes and thus have an effect similar to that of a mutation. Likewise, as mentioned above, certain physiological environments are effectively mimicked by gene deletions. This is further supported by the observation that the correlation between the number of genetic interactions and sensitivity profiles remains stable when drug treatments are eliminated from the data set, such that only naturally occurring environments remain (Supplementary Figure 6).

Importantly, environmental sensitivity profiles provide independent evidence for the inability of buffering paralogs to comprehensively cover for the loss of their partner gene: the deletion of SSL duplicates across a range of environments has on average no weaker (and in fact a slightly stronger) effect on cellular growth rate than that of non-SSL duplicates or random genes (Figure 4F). Likewise, when sensitivity profiles are used as a functional signature that complements that of genetic interaction spectra, we similarly find that profiles between most SSL duplicates have substantially diverged (Supplementary Figure 10) and, with one exception, display greater similarity to those of other, non-paralogous genes in the data set.

Discussion

The high proportion of dispensable genes in yeast as well as other eukaryotes (Kamath *et al*, 2003) and prokaryotes (Kobayashi *et al*, 2003) represents both a theoretical and practical challenge. In practical terms, the fact that thousands of genes fail to exhibit a detectable growth defect under multiple conditions substantially limits efforts of systematic phenotyping. Elucidation of gene function, in particular, relies critically on the ability to elicit a rich range of phenotypes. Conceptually, the high degree of dispensability has widely been taken as evidence for mutational robustness (i.e. the ability of the system to function after genetic changes), similar to robustness observed in biochemical networks (Kitano, 2004). Direct mechanistic evidence for the underlying causes, however, has largely been missing. In addition, true dispensability and redundancy are difficult to justify because of their evolutionary instability.

One factor that has been implicated in the high degree of dispensability is the presence in the yeast genome of numerous paralogs, originating both from the large-scale duplication event more than 100 million years ago as well as from smaller scale duplications (Wolfe and Shields, 1997; Kellis *et al*, 2004). Although such duplicates are often lost rapidly, in a fraction of cases they are retained in functional form. The reason and consequence of such retention has been the subject of considerable interest. For example, duplications could have provided an opportunity to greatly increase mutational robustness of the organism. However, although our data

provide direct support for the role of duplicate buffering, we find that its total contribution to dispensability is small. Together with the observation that the majority of duplicates are not synthetic with their paralog, this argues that the evolutionary pressure to maintain similar functions between duplicates is low at best. Instead, our findings suggest that the predominant reason for the retention of duplicates is for functional innovation and refinement (Ohno, 1970).

The fact that our study is confined to the genes of the two currently available large-scale genetic interaction data sets raises concerns about the generality of our results. However, it should be noted that genes in the first data set (ER) were chosen by their cellular localization and included both soluble and membrane-bound proteins of diverse and often unknown function. In contrast, the genes of the second data set were selected based on their known functions centered around chromosome biology (including DNA damage repair, transcriptional regulation, chromosome segregation, telomere regulation as well as the cell cycle), and comprised largely soluble genes of diverse localizations. Importantly, the close correspondence between the excess robustness of duplicates and the number of those that are SSL with each other (Figure 1C and D) is observed for each of the two data sets separately as well as in combination (Supplementary Figure 1). This is particularly noteworthy as the fitness distributions themselves differ substantially between the two data sets. The fact that these two very different data sets separately support our conclusions argues in favor of their generality. This is furthermore supported by the observation that these results are robust against subsampling of the data (Supplementary Figure 2). However, there may be other subsets of genes for which the observed relationships will not hold. The present analysis can serve as a framework to explore this question as more genetic interaction data become available.

Is the subset of paralogs that do contribute to dispensability functionally redundant? We find that even duplicates with strong SSL interactions are far from a state of redundancy and rarely, if ever, exhibit a capacity to broadly cover for the loss of paralogous partner genes. Using epistatic interaction spectra as an indicator of function, we ascribe this lack of generic backup capacity to two distinct reasons, namely functional divergence and dosage amplification. In the first case, most duplicates are largely uncorrelated in both interaction spectra and expression profiles. These duplicates either overlap only in specific functions or are subject to different regulation, or both. Although deletion of these genes has little effect in rich media, they exhibit a considerable number of synthetic interactions. Previous analyses had suggested that the degree of divergence between duplicates is such that most are unlikely to serve as backup copies (Wagner, 2005). In contrast to such expectations, our experimental data show that high similarity in sequence, regulation or interaction patterns *per se* appears not to be a prerequisite for backup capability.

In the case of dosage amplification, a small subset of duplicates with a high degree of functional and regulatory similarity is likely involved in processes where their gene product is required at high copy number. In spite of their functional similarity, loss of one of the duplicates thus generally has a deleterious effect. This is illustrated by the example of the histone genes hht1/hht2 and hhf1/hhf2, which

are identical in coding sequence, expressed at high and similar copy number and capable of buffering in rich media conditions. In spite of their great similarity, however, cells are far from robust against the loss of one of the copies in general conditions, as evidenced by their large number of SSL interactions. A role of duplicates in dosage amplification complements previous results in yeast metabolism, where most isozymes were found to be differentially coregulated with separate pathways (Ihmels *et al*, 2004). Although no genome-wide genetic interaction data are available, the observation that the relationship between coexpression, abundance and abundance similarity holds on a genomic scale supports the role of dosage amplification for a significant subset of duplicates in the genome (Seoighe and Wolfe, 1999; Kondrashov and Kondrashov, 2006).

The ability to elicit numerous genetic interactions from ostensibly dispensable genes, including those buffered by a paralog, raises the question of the physiological relevance of deletion backgrounds. Rich genetic interaction spectra of buffering duplicates could point to pleiotropic genes, where each subfunction is buffered by a separate backup gene. Alternatively, as detailed above, deletion backgrounds can provide cellular stresses that mimic physiological environments and thus reveal phenotypes for genes required only in specific conditions. The use of genetic interactions as a reporter of such condition specificity is supported by the correlation between the number of synthetic interactions, phylogenetic retention and environmental sensitivity, as well as the observation that similarity in genetic interaction spectra is strongly indicative of similarity in sensitivity profiles (Supplementary Figure 11, $P < 10^{-21}$). A related connection was found in bacteria, where the effect of mutations that exhibit epistasis were shown to be more likely to simultaneously depend on the environment (Remold and Lenski, 2004).

Previously known examples of duplicates that provide compensation in some environments and not others include the three yeast A kinase isoforms tpk1, tpk2 and tpk3, which are separately dispensable in rich media, but have different functions under conditions of pseudohyphal growth, where tpk2 is essential (Robertson and Fink, 1998). From an evolutionary point of view, condition-specific gene requirement and backup capacity that is limited to some environments offer a way to reconcile the concept of robustness against deletions with the constraint that genes whose loss is not associated with a phenotypic penalty cannot be maintained in the population. In addition to the rich genetic interaction spectra of SSL duplicates, this view is further supported by the observation that growth rates of buffering paralogs are no less and in fact slightly more affected by environmental perturbations than non-buffering duplicates or random sets of genes.

In silico models of cellular metabolism based on flux balance analysis (FBA) have similarly predicted a large fraction of seemingly dispensable genes that are in fact required, but only in specific environments (Papp *et al*, 2004). Our findings provide experimental support for these predictions in the context of deletion backgrounds, and demonstrate that they are not confined to metabolic genes. On the issue of duplicates, however, our results differ from those of FBA-based models,

where full redundancy is an explicit assumption. Similarly, previous analyses view duplicate buffering, condition-specific gene requirement and alternative pathways as separate mechanisms, whereas our data suggest that they are interrelated in the sense that the function of backup genes and pathways is often dependent on the cellular environment as well.

Materials and methods

Genetic interaction data

Two separate data sets of quantitative genetic interaction profiles were used, one of 424 genes involved in endoplasmic reticulum function (Schuldiner *et al*, 2005) and a more recent data set of 743 genes centered around DNA damage and transcription (Collins *et al*, 2007). To control for the different sizes of the two sets, the number of interactions in the second set was scaled by a factor of 424/743 in Figure 4B. Details of how the data were generated can be found in literature (Schuldiner *et al*, 2005; Collins *et al*, 2006). Briefly, colony sizes of double mutants were measured under identical conditions, and size measurements were normalized to correct for systematic errors. Interaction scores were calculated for each pair of genes using a modified *t*-statistic, based on the means and variances of the normalized double and single mutant sizes. The interaction score reflects the fitness of the double mutant, relative to the fitness that would be expected given the fitness of each single mutant. This expected fitness (assuming no genetic interactions) is determined empirically from the data by considering the full set of double mutants involving each single deletion. Negative and positive scores indicate aggravating (SSL) and alleviating interactions, respectively. Both data sets contain genetic interaction scores between the possible pairwise combinations of genes in each set. The genes contained in each data set are listed in Supplementary Table 3.

Genetic interaction correlations

Each gene was assigned an interaction profile, that is, a vector containing the genetic interaction score with all other genes in the data set. Genetic interaction correlations were calculated between these profiles using the Pearson correlation coefficient. Calculations of the corresponding *P*-values were made using the Matlab function *corrcoef*, which uses the correlation to generate a *t*-statistic.

Definition of SSL interactions

Following Schuldiner *et al* (2005), negative interactions were considered significant beyond a threshold value of −3, unless otherwise stated.

Identification of duplicates and singletons and calculation of substitution rates

Gene pairs were defined as paralogs if their BLASTP *E*-values was $< 10^{-20}$ and whose protein lengths differed by no more than one-third. Following Gu *et al* (2003), singletons were defined as genes with no hits in a FASTA search at $E = 0.1$. Rates of synonymous and non-synonymous substitution were calculated using an estimation algorithm (Li, 1993) implemented in the Matlab package MBEToolbox (Cai *et al*, 2005). Gene families of more than two paralogs were excluded from the analysis.

mRNA expression profiles

We used a compilation of several published microarray data sets taken from Ihmels *et al* (2002). These data comprise genome-wide transcription profiles under a large variety of cellular conditions, including gene deletions, environmental stresses, different growth

media, cell–cycle progression, etc. Log 2 ratios across 1011 available conditions were used to calculate Pearson correlation coefficients between pairs of genes.

GO term annotations

GO term annotation files were downloaded from http://www. geneontology.org. The assignment of genes to the different categories was extended to include parent terms, that is, genes assigned to a given category were assigned to the parent categories as well. Genes were considered positive if they co-appeared in at least one functional category of the *biological process* ontology with no more than 300 genes. Negative pairs consisted of genes whose most specific co-annotation occurs in terms containing at least 1000 genes.

Definition of WNP and other fitness classes

Following Gu *et al* (2003), the genes whose minimum deletion growth rate in five environments exceeded a threshold value of 0.95 were assigned to the *weak/no phenotype* (WNP) class. Genes with lower fitness were assigned to the *strong phenotype* class. Deletion growth rates in five environments (YPD, YPDGE, YPG, YPE, YPL) were downloaded from the Yeast Deletion Project database at the URL http://www-deletion.stanford.edu/YDPM/YDPM_index.html.

Definition of gene sensitivity

Genome-wide sensitivity profiles were taken from Brown *et al* (2006). Relative abundance of each strain in a pool of deletion mutants was measured using oligonucleotide arrays. The data generated in each experimental array were normalized to that of a control array, resulting in logarithmic ratios S_i^j for each gene i in condition j. A combined sensitivity coefficient was calculated for each gene as

$$s_i = \sum_j (S_i^j)^2 \cdot \theta(-S_i^j)$$

where θ is the unit step function, such that the sum is only over negative entries (corresponding to growth defects). This coefficient provides a measure of how strongly the deletion of each gene affects cellular growth over the 51 conditions studied.

Removal of drug additions from the data set resulted in the following 15 environmental conditions: YPD, RafA, GlyE, Alk-5g, Alk-15g, Gal-5g, Gal-15g, Lys, Min-5g, Min-15g, NaCl-5g, NaCl-15g, SC, Sorb-5g, Sorb-15g.

Significance of SSL interactions between duplicates

The chance of finding at least n_{ssl} duplicates out of n_d with a negative interaction between them can be expressed as

$$p = \sum_{x=n_{ssl}}^{n_d} \binom{n_d}{x} \left(\frac{Q}{n}\right)^x \left(1 - \frac{Q}{n}\right)^{n_d-x} = 1 - \sum_{x=0}^{n_{ssl}-1} \binom{n_d}{x} \left(\frac{Q}{n}\right)^x \left(1 - \frac{Q}{n}\right)^{n_d-x}$$

where Q is the average number of negative interactions of duplicates assigned to the WNP class and n is the number of genes in the E-MAP. The right-hand side can be evaluated in a numerically stable form using the Matlab function *binocdf*.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

References

Brookfield J (1992) Can genes be truly redundant? *Curr Biol* **2:** 553–554

Brown JA, Sherlock G, Myers CL, Burrows NM, Deng C, Wu HI, McCann KE, Troyanskaya OG, Brown JM (2006) Global analysis of gene function in yeast by quantitative phenotypic profiling. *Mol Syst Biol* **2:** 1

Cai JJ, Smith DK, Xia X, Yuen KY (2005) MBEToolbox: a MATLAB toolbox for sequence data analysis in molecular biology and evolution. *BMC Bioinform* **6:** 64

Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol* **7:** R63

Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles J, Boone C, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ (2007) Functional dissection of yeast chromosome biology complexes using a genetic interaction map. *Nature* (in press)

Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol* **2:** e137

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737–741

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418:** 387–391

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH (2003) Role of duplicate genes in genetic robustness against null mutations. *Nature* **421:** 63–66

He X, Zhang J (2006a) Higher duplicability of less important genes in yeast genomes. *Mol Biol Evol* **23:** 144–151

He X, Zhang J (2006b) Transcriptional reprogramming and backup between duplicate genes: is it a genomewide phenomenon? *Genetics* **172:** 1363–1367

Hirsh AE, Fraser HB (2001) Protein dispensability and rate of evolution. *Nature* **411:** 1046–1049

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet* **31:** 370–377

Ihmels J, Levy R, Barkai N (2004) Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat Biotechnol* **22:** 86–92

Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* **411:** 41–42

Kafri R, Bar-Even A, Pilpel Y (2005) Transcription control reprogramming in genetic backup circuits. *Nat Genet* **37:** 295–299

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P,

Ahringer J (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421:** 231–237

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae. Nature* **428:** 617–624

Kitano H (2004) Biological robustness. *Nat Rev Genet* **5:** 826–837

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100:** 4678–4683

Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. *J Theor Biol* **239:** 141–151

Krylov DM, Wolf YI, Rogozin IB, Koonin EV (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res* **13:** 2229–2235

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* **36:** 96–99

Libuda DE, Winston F (2006) Amplification of histone genes by circular chromosome formation in *Saccharomyces cerevisiae. Nature* **443:** 1003–1007

Nakielny S, Dreyfuss G (1999) Transport of proteins and RNAs in and out of the nucleus. *Cell* **99:** 677–690

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441:** 840–846

Ohno S (1970) *Evolution by Gene Duplication*. Springer, New York

Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD (2006) A DNA integrity network in the yeast *Saccharomyces cerevisiae. Cell* **124:** 1069–1081

Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429:** 661–664

Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70:** 503–533

Remold SK, Lenski RE (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli. Nat Genet* **36:** 423–426

Robertson LS, Fink GR (1998) The three yeast A kinases have specific signaling functions in pseudohyphal growth. *Proc Natl Acad Sci USA* **95:** 13783–13787

Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123:** 507–519

Seoighe C, Wolfe KH (1999) Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* **2:** 548–554

Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002) Systematic screen for human disease genes in yeast. *Nat Genet* **31:** 400–404

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364–2368

Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C (2004) Global mapping of the yeast genetic interaction network. *Science* **303:** 808–813

Wagner A (2000) Robustness against mutations in genetic networks of yeast. *Nat Genet* **24:** 355–361

Wagner A (2005) Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27:** 176–188

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285:** 901–906

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387:** 708–713

Wong SL, Roth FP (2005) Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in *Saccharomyces cerevisiae. Genetics* **171:** 829–833