nature genetics

Transcription control reprogramming in genetic backup circuits

Ran Kafri, Arren Bar-Even & Yitzhak Pilpel

A key question in molecular genetics is why severe mutations often do not result in a detectably abnormal phenotype. This robustness was partially ascribed to redundant paralogs[1,2] that may provide backup for one another in case of mutation. Mining mutant viability and mRNA expression data in *Saccharomyces cerevisiae*, we found that backup was provided predominantly by paralogs that are expressed dissimilarly in most growth conditions. We considered that this apparent inconsistency might be resolved by a transcriptional reprogramming mechanism that allows the intact paralog to rescue the organism upon mutation of its counterpart. We found that in wild-type cells, partial coregulation across growth conditions predicted the ability of paralogs to alter their transcription patterns and to provide backup for one another. Notably, the sets of regulatory motifs that controlled the paralogs with the most efficient backup activity deliberately overlapped only partially; paralogs with highly similar or dissimilar sets of motifs had suboptimal backup activity. Such an arrangement of partially shared regulatory motifs reconciles the differential expression of paralogs with their ability to back each other up.

Functionally redundant gene duplicates are inherently evolutionarily unstable; consequently, in many duplications, one of the duplicates is silenced[3,4]. Retention of duplicates over long evolutionary time scales was therefore suggested to require either degenerative subfunctionalizing mutations or introduction of new functions[3,5–8]. We aim here to understand both the relevance of transcription regulation to duplicate retention in evolution and its role in controlling expression of genes that provide backup in case of mutation.

Mining single gene–knockout phenotype data and annotations of molecular functions of all yeast genes, we found high correlation between the essentiality of genes and the similarity of molecular function between themselves and their paralogs (**Supplementary Figs. 1** and **2** online). We also found that only 4% of the dispensable paralogs did not colocalize[9] in the same organelles (**Supplementary Fig. 3** online). These observations corroborate the notion that 'dispensability' may be explained by backup between paralogs.

A priori, it might seem that backup requires the paralogs' mRNAs to be coregulated. To examine this possibility, we calculated, for each pair of paralogs, 40 correlation coefficients of mRNA expression corresponding to 40 different experiments. We define the means and standard deviations of such correlations, for each pair, as their mean expression similarity and partial coregulation (PCoR) values, respectively. We refer to the standard deviation of the correlations as PCoR because its value is high for pairs that have interchangeably high and low correlations across different conditions. **Figure 1** shows the proportion of dispensable genes in sets of gene pairs versus their mean expression similarity and PCoR. We inspected close and remote paralogous pairs separately and found markedly different trends. Among remote paralogs, we found that the essentiality of coexpressed pairs was very high, implying that there is little backup activity among them. In remote pairs, backup was most efficient among transcriptionally noncorrelated pairs, as their essentiality was substantially lower than that of single genes. **Supplementary Figures 4** and **5** online show the increase in protein-protein interaction among paralogs and the decrease in similarity of Gene Ontology–annotated molecular function between them, respectively, as a function of coexpression. These results provide a potential explanation for the observed decrease in backup capacity with increased coexpression. In contrast to remote pairs, close pairs showed an almost opposite, more intuitive trend, in which dispensability increased somewhat with expression similarity (in agreement with refs. 1,10).

Backup among naturally dissimilarly expressed genes A and B may suggest that, upon mutation in gene A, expression of gene B is reprogrammed to acquire a profile that is similar to the wild-type expression profile of gene A. Such reprogramming has been experimentally verified for the Acs1 and Acs2 isoenzymes. Wild-type Acs1 is subject to glucose repression[11] (**Fig. 1**), but upon deletion of Acs2, the repression of Acs1 is relieved, and Acs1 acquires an Acs2-like responsiveness to glucose[12]. Despite dissimilar expression, the two genes share a promoter motif (CSRE) and also have unique motifs[12]. As befits a genuine backup circuit, Acs1 and Acs2 are synthetically lethal[11]. Additional examples of reprogramming in response to mutations in prokaryotes, yeast and mammals are given in **Supplementary Note** online.

In search for a mechanism that may regulate switching between dissimilar and similar expression in response to mutation, we examined the dependence of gene essentiality on PCoR. We asked whether backup occurs among paralogs that show high PCoR in

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. Correspondence should be addressed to Y.P. (pilpel@weizmann.ac.il).

Published online 20 February 2005; doi:10.1038/ng1523

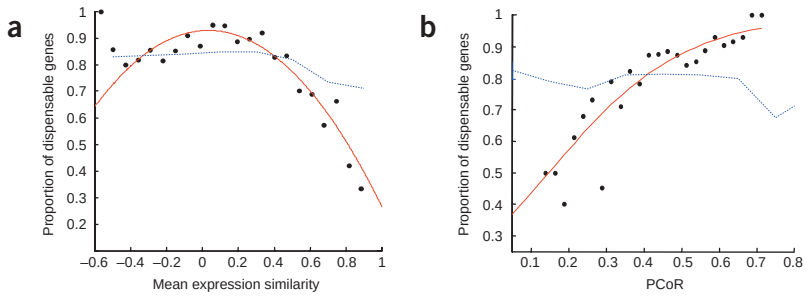

Figure 1 Dependence of backup on expression similarity between paralogs. Proportion of dispensable genes as a function of the mean expression similarity (**a**) and PCoR (**b**) in sets of paralogs. Results are shown only for genes having remote paralogous partners ($K_s > 1$). Results for close pairs ($K_s < 1$) showed a marginally significant opposite trend (**Supplementary Figs. 11** and **12** online). The red line represents a quadratic regression fit (**a**), scoring an adjusted R^2 value of 0.83, and a logistic regression fit (**b**) with a P value of 10^{-25}. To exclude the possibility that the trend in **a** reflects a tendency for genes that belong to major expression clusters to be essential, we repeated the analysis using random pairing of genes and observed a nonsignificant trend (blue). Examination of remnants of whole-genome duplication[26] showed similar trends to that observed with all remote pairs, but with marginal significance.

the wild type. We reasoned that because PCoR represents the ability to switch between similar and dissimilar expression profiles in a condition-dependent manner, it may be predictive of switching between similar and dissimilar expression in response to mutation. We found that PCoR was a very strong predictor of backup (**Fig. 1b**).

We next investigated the promoter architecture of backup-providing paralogs. The possibility that the partial overlap in the sets of regulatory motifs controlling Acs1 and Acs2 accounts for their wild-type differential expression and for reprogramming upon mutation prompted us to inspect the similarity of motif content of all paralogs. To quantify the extent to which promoters of paralogs are arranged to obtain partial coexpression, we defined O, a normalized measure of the overlap between the sets of promoter motifs that regulate two genes:

$$O = \frac{|m_1 \cap m_2|}{\max(|m_1|, |m_2|)},$$

where m_1 and m_2 are the sets of motifs that regulate genes 1 and 2, respectively, and $|x|$ is the size of a set x. By plotting gene dispensability versus motif-content overlap O, we found that maximal backup coincided with intermediate levels of motif sharing (**Fig. 2**). Pairs with high or low promoter similarity had suboptimal backup activity. These observations confirm that optimal backup is obtained when two paralogs share some, but not all, motifs. We propose that the unique motifs of each paralog provide differential expression in the wild type and that the shared motifs allow paralogs to respond to the same conditions. This situation allows for reprogramming in response to mutations. We plotted the number of shared transcription factor binding sites against the rate of substitutions per synonymous position, K_s (a rough duplication-age surrogate), and found nearly identical average numbers of shared motifs across the entire range of K_s values ($R = 0.025$, $P > 0.29$; **Supplementary Figs. 6** and **7** online). This indicates that sharing of transcription factor binding motifs results either from restricted divergence or from convergence and is not an evolutionary artifact that is likely to dissipate on an evolutionary time scale.

To corroborate the hypothesis that PCoR underlies reprogramming and, ultimately, backup, we examined three predictions. First, one member of a pair with high PCoR should be upregulated transcriptionally in response to the deletion of its paralog. To investigate this prediction, we used the Rosetta Compendium[13] containing genome-wide expression response to single-gene deletions. Of the 259 knockouts in the Compendium, 76 have paralogs in our data set. Of these, 18 share high similarity in molecular function, and another 5 are synthetically lethal. We reasoned that if such potential backup-providing pairs undergo reprogramming then the transcriptional level of the intact paralog should increase as a function of the pair's PCoR. In fact, we found a significant correlation between PCoR and the logarithm of transcriptional response to deletion among these backup-providing candidates ($R = 0.67$, $P = 0.002$; **Fig. 3**). As a negative control, functionally similar nonparalogs and random pairs showed no correlation between PCoR and transcriptional response to deletion ($R = -0.02$ and $R = 0.01$, respectively). Therefore, we conclude that PCoR measured across wild-type conditions predicts backup capacity or the ability of a gene to respond, by upregulation, to deletion of its counterpart.

Our second prediction addresses 478 paralogs in which only one of the two genes is essential. We tested our ability to predict which of the two genes in such asymmetric pairs is essential by inspecting their regulatory motifs. Our reprogramming scenario predicts that the more motifs control a gene, the better its reprogramming and backup-providing capacity will be. Therefore, for paralogous pairs, we expect a negative correlation between number of motifs controlling a gene and its dispensability. As expected, the more essential of the two genes tended to have more motifs (**Fig. 4**). As a negative control, we repeated the analysis with random pairing of the paralogs to determine whether this observation merely reflected the potential bias that essential genes are regulated by a larger number of motifs. This analysis with random pairing resulted in no signal (**Fig. 4**).

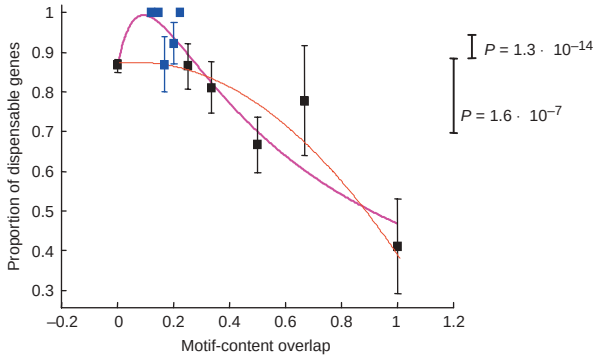

Figure 2 Gene dispensability as a function of the regulatory motif–content overlap O between genes and their closest paralogs. By fitting these data to a linear function (not shown), a quadratic function (red) and the rational function (purple; $y = (ax + b)/(x^2 + cx + d)$), we obtained adjusted r^2 values of 0.56, 0.72 and 0.82, respectively. A binomial test showed that the proportion of dispensable genes with O values between 0 and 0.25 (blue bars) was significantly higher than that of genes with O values of either 0 ($P = 1.6 \times 10^{-7}$) or >0.25 ($P < 1.3 \times 10^{-14}$).

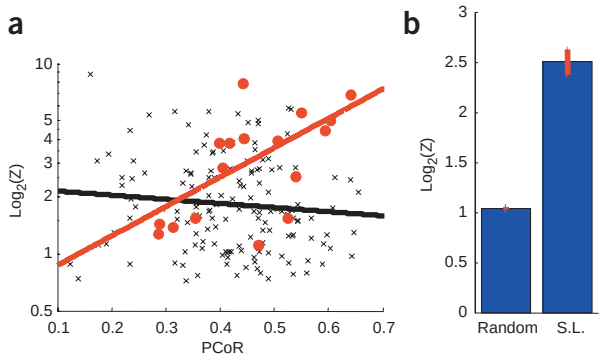

Figure 3 Transcriptional response of backup-providing genes to the deletion of the counterparts. (**a**) Transcriptional responses of genes to the deletion of their functionally similar paralogs (red, R = 0.67, $P = 0.002$) and functionally similar nonparalogs (black, R = −0.02) as a function of their PCoR (data obtained from the Rosetta Compendium[13]). Response is depicted as average $\log_2$ relative change of the expression level of a gene in the mutant strain lacking its paralog divided by the expression level of the gene in the wild type. Decreased reads in response to deletion may result from artifacts owing to potential cross-hybridization in the wild type. This effect was excluded by analyzing only genes that are upregulated after the deletion. Only functionally similar paralogs were analyzed, defined either as genes encoding enzymes with the same EC classification or, for nonenzymes, as genes with high Gene Ontology–based semantic similarity[27] (where high similarity indicates similarity exceeding that observed at the 90th percentile of similarities of all gene pairs in the genome). (**b**) Average upregulation of functionally similar pairs that are also synthetically lethal (from the BIND database) compared with randomly selected gene pairs.

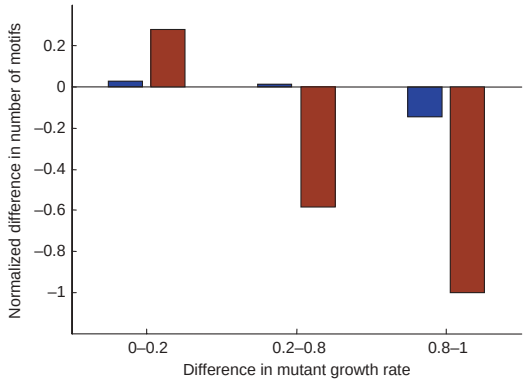

Figure 4 Difference in the number of motifs regulating paralogous pair members as a function of the difference in the growth rates of mutants lacking them. For each pair of paralogs, the number of motifs contained by the gene with the higher growth rate was subtracted from the number of motifs in the promoter of the gene with the lower growth rate. This difference was normalized to the size of the larger of the two motif sets. All paralogs were then grouped into three categories on the basis of the absolute value of the difference between their growth rates, and for each category, the mean normalized difference in the number of motifs was calculated. The analysis was done separately for all paralogs (blue) and for paralogs with similar molecular functions (red; defined as in the legend to **Fig. 3**).

Third, our proposed model predicts synthetic lethal interactions. We embedded the paralogous pairs in a plane spanned by their mean expression similarity and PCoR (**Fig. 5**). We gathered evidence for synthetic lethality for certain pairs of paralogs and observed that the prevalence of backup depended on both mean expression similarity and PCoR score. Backup was maximal among pairs with high PCoR and low coexpression. Physical interactions between paralogs showed an opposite trend (**Supplementary Fig. 4** online), in agreement with previous observations[14,15]. The model also includes verified cases of reprogramming.

A crucial question is what controls reprogramming of a gene upon mutation of its paralog. We propose a kinetic model, or reprogramming switch, consisting of two genes, G1 and G2, that encode enzymes E1 and E2, which interconvert metabolite M1 into metabolite M2. In the wild type, only E1 is active. Assuming that the two genes contain binding sites for a shared transcription factor T that is induced by M1, T reprograms (*i.e.*, activates) G2 and hence maintains the level of M2 upon knockout of G1 (**Fig. 6**). Upon silencing of G1, M1 accumulates and the concentration of T increases, resulting in more efficient activation of G2 (**Fig. 6**). Consequently, the level of E2 increases and the level of M2 returns to its original value after a transient decrease (**Fig. 6**). This model provides appropriate control of backup as it couples response of G2 to an environmental condition (*i.e.*, the accumulation of M1) with response to an internal perturbation (*i.e.*, silencing of G1). The model describes enzymatic reactions; enzymes are over-represented in our data set (34%; **Supplementary Fig. 8** online). Backup among paralogous transcription factors may use alternative architectures (**Supplementary Note** online).

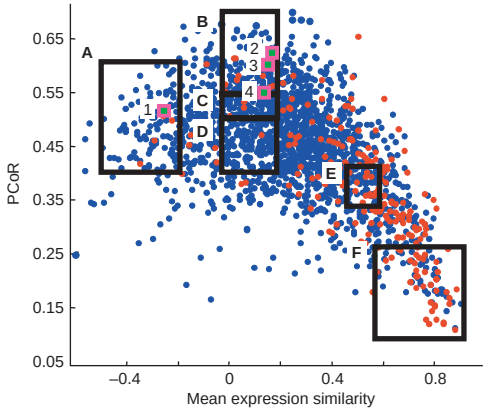

Figure 5 Confirmation and characterization of genetic backup circuits. Paralogous gene pairs are plotted as a function of their mean expression similarity and PCoR. Pairs are colored red if both members are essential or blue if both are dispensable. Black rectangles (A–F) enclose sets of genes whose functional redundancy was confirmed or disputed using the Proteome and BIND databases. In this analysis, pairs were considered to back each other up only if they have similar molecular activities and are synthetically lethal. The number of such backup-providing pairs was divided by the total number of functionally characterized pairs in each of the marked rectangles individually (A: 10/29 = 0.34; B: 14/30 = 0.47; C: 12/35 = 0.34; D: 10/45 = 0.22; E: 8/30 = 0.27; F: 1/42 = 0.02). The highest probability for verified backup coincides with cases where mean expression similarity is ~0 and PCoR >0.4. The placement of the rectangles reflects our desire to examine how incidence of backup depends on the *x*,*y* coordinates of the pairs. Four examples of paralogs that show transcriptional reprogramming in response to gene deletion are also shown (green squares): 1: Acs1-Acs2 (ref. 11); 2: Hxt2-Hxt10 (ref. 28); 3: Idp1-Idp2 (ref. 29); 4: Fks1-Gsc2 (ref. 30).

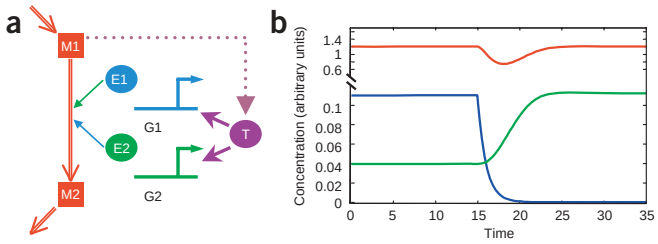

Figure 6 Schematic and dynamics of the reprogramming switch. (**a**) The reprogramming switch. (**b**) Simulated dynamics of the switch before and after knockout, at time point 15. The blue and green curves represent the concentrations of E1 and E2, respectively; the red curve represents the concentration of M2. The dynamics were calculated from the differential equations describing the system using the ode23 solver of Matlab's simulink.

The different behavior of close and remote paralogs probably stems from the profoundly different evolutionary regimens acting on them[5]. Focusing first on remote pairs, we propose that preservation of high coexpression in a subset of these pairs was predominantly due to evolutionary pressures that are inconsistent with, and compromise, backup. One such effect is evolving protein-protein interactions between paralogs, which requires coexpression but precludes backup (**Supplementary Fig. 4** online). Second, subfunctionalization of proteins may alleviate the pressure to diverge in expression[3], but that, too, precludes backup between coexpressed pairs (**Supplementary Fig. 5** online). Third, quantitative subfunctionalization[16] that may result in regulatory motif degeneration[10,17] accounts for both coexpression and lack of backup (*e.g.*, due to low dosage of each of the coacting paralogs[18,19]).

Why do remote pairs back each other up? Although it is difficult to imagine that backup by duplicates is evolutionarily selectable[3], we propose that backup-providing duplicates may be retained during evolution if their retention is coupled to other selectable traits, such as acquisition of new regulatory capabilities[10]. Such novelties do not preclude backup, provided that shared functionalities are preserved. Our finding that backup is optimal among pairs that maintain high partial coregulation provides considerable support to this notion. Notwithstanding this, however, backup has a profound impact on an organism's robustness, whether selected for its own sake or not. But apparent dispensability may be partially due to limited coverage of growth conditions tested in the laboratory, and a recent computational study[19] estimated that this factor accounts for 37–68% of dispensable genes, compared with the 15–28% that are estimated to be compensated by a duplicate. Gu *et al.* estimated a similar lower bound of 25% (ref. 2).

Many of the close paralogs that represent recent duplications[5] are assumed to be under free selection, meaning that they have not yet undergone either sub- or neofunctionalization; hence, they are redundantly similarly expressed. This probably explains their somewhat more intuitive behavior (backup increases with coexpression), which does not depend on evolving reprogramming.

METHODS

Set of analyzed genes and definition of paralogs. To ensure that we analyzed genuine genes, we discarded from the list of *S. cerevisiae* open reading frames (ORFs) all entries corresponding to spurious ORFs[20] and all transposon-derived genes as annotated by Saccharomyces Genome Database. This resulted in a list of 5,862 ORFs. We defined paralogs as pairs of ORFs that, by BLASTP with standard parameters, had E valued $<10^{-20}$, provided that the ratio of the length of the long protein to the length of the short protein was not larger than 1.33. For each pair of paralogs, we calculated the number of synonymous and nonsynonymous substitutions (K_s and K_a, respectively)[21]. We defined remote paralogs as pairs with $K_s > 1$ and close paralogs as pairs with $K_s \leq 1$. To avoid potential misclassification of borderline cases, we also adopted an alternative definition in which we regarded remote pairs as those with $K_s > 1.2$ and close pairs as those with $K_s < 0.8$ and found that the same trends characterized the two sets (**Supplementary Fig. 9** online). **Supplementary Figure 9** contains additional cutoff justifications including a systematic assessment of the robustness of the results to changes of threshold value and to use of alternative measures of sequence similarity (*e.g.*, K_a). To remove from the set of close pairs ($K_s < 1$) any paralogs that represent old duplications, we removed close paralogs in which at least one of the genes had a low (<32) effective number of codons[22].

Gene essentiality data. We defined dispensable genes as all genes with a viable gene-deletion phenotype that were not included in the lists of spurious ORFs or transposon-derived ORFs. Additionally, we obtained data on growth rates of mutants lacking each of the ORFs in the genome in five different growth media[23].

mRNA expression data. We obtained whole-genome mRNA expression data of 40 natural and perturbed time series and the Rosetta Compendium data, which measures genome-wide transcription response to gene deletions[13], from ExpressDB. We normalized all expression profiles of genes in each time series with respect to mean and variance. Detailed descriptions of all analyzed conditions is presented on our project website (see URL below).

We obtained expression data from either Affymetrix chips (seven experiments) or PCR product-based microarrays (33 experiments). Because the latter technology is more prone to cross-hybridization errors, we used only data derived from Affymetrix chips when analyzing close paralogs.

A nonredundant set of promoter regulatory motifs in *S. cerevisiae*. We compiled a nonredundant set of 112 yeast regulatory motifs, along with their gene assignments, from three different sources: ChIP-chip (originally augmented with phylogenetic conservation of motifs across multiple yeast species)[24], expression data[25] and phylogenetic conservation[20]. We included motifs derived from the last two computational methods only if they corresponded to experimentally known motifs and had a significance score higher than the 90th percentile in their respective methods.

Kinetic analysis of the reprogramming switch. We modeled the concentration of induced transcription factor (T*) and the fractions of time in which genes G1 and G2 were transcribed, denoted as G1* and G2*, respectively, with three saturation equations:

$$T^* = T^{Tot}\frac{(M1/K_M)^{n_M}}{(M1/K_M)^{n_M}+1};$$

$$G1^* = \frac{(T^*/K_1)^n}{(T^*/K_1)^n+1};$$

and

$$G2^* = \frac{(T^*/K_2)^n}{(T^*/K_2)^n+1},$$

where T^{Tot} is the concentration of total transcription factor; K_M is the affinity between the transcription factor T and the inducing metabolite M1; K_1 and K_2 are the transcription factor's affinities to G1 and G2, respectively; and the powers n_M and n represent binding cooperativity Hill coefficients of M1 to T and of T* to the two promoters, respectively.

The concentration of the enzymes and the metabolites are described with time-dependent differential equations:

$$\frac{dE1}{dt} = \beta \cdot G1^* - \alpha E1;$$

$$\frac{dE2}{dt} = \beta \cdot G2^* - \alpha E2;$$

$$\frac{dM1}{dt} = \beta_M - \phi(E1+E2)M1;$$

and

$$\frac{dM2}{dt} = \phi(E1 + E2)M1 - \alpha_M M_2,$$

where β and β_M are the maximal production rate of E1 and E2 and of M1, respectively; α and α_M represent the degradation and dilution of E1 and E2 and of M2, respectively; and ϕ represents the conversion rate of M1 to M2. Values of the coefficients, the present simulation, are $T^{Tot} = 1$, $K_M = 20$, $K_1 = 0.1$, $K_2 = 0.3$, $n_M = 4$, $n = 1$ (*i.e.*, no cooperativity is assumed) and $\alpha = \beta = \alpha_M = \beta_M = \phi = 1$. There are three reasonable assumptions in the model. First, we assume that the binding of M1 to T and of T* to G1 and G2 occurs on a short time scale compared with the other reactions; therefore, these reactions are in a quasi-steady state. Second, we assume that M1 > T (the total number of free M1 molecules is roughly the same as the total number of M1 molecules). Finally, we assume that E1 and E2 work linearly with respect to M1 as a substrate.

Statistical analyses. We computed proportions of dispensable genes as a function of mean expression similarity, PCoR and motif-content overlap *O* (**Figs. 1** and **2**) by binning genes into groups according to each of these three variables and then calculating the frequency of viable mutants in each bin. We counted each gene in each bin only once to avoid repetitions caused by one gene having multiple paralogs. To establish that our results are independent of the particular choice of binning strategy, we verified that the observed trends were valid under any relevant bin-size choice (**Supplementary Fig. 10** online).

We tested the significance of trends observed for the proportions of dispensable genes against mean expression similarity (**Fig. 1a**), PCoR (**Fig. 1b**) and motif-content overlap (**Fig. 2**) using logistic regressions analyses (in **Fig. 1a**, only the declining portion of the curve, with positive expression similarity values, was used). Further statistical analyses of the results are shown in **Supplementary Figures 9** and **10** online.

URLs. We downloaded gene sequences from the Saccharomyces Genome Database (http://www.yeastgenome.org/) and retrieved gene knockout phenotype data from http://sequence-www.stanford.edu/group/yeast_deletion_project/Essential_ORFs.txt. Functional annotations for all genes came from the Gene Ontology annotation scheme at http://www.yeastgenome.org/. We downloaded synthetic lethal interactions and physical interactions between proteins from the BIND database (http://bind.ca/). We collected gene expression data from ExpressDB (http://arep.med.harvard.edu/ExpressDB/) and used the Proteome database (http://proteome.incyte.com/) to collect synthetic lethal pairs manually. Our project website is http://longitude.weizmann.ac.il/BackUpCircuits/. We obtained EC classifications from http://mips.gsf.de/genre/proj/yeast/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all members of the laboratory of Y.P. for discussions; I. Pechersky for computational assistance; and Y. Garten, N. Barkai, J. Berman, B. Shilo, A.M. Dudley, I. Yanai, O. Man, S. Shen-Orr, D. Graur, D. Lancet, M. Levy and D. Artzi for critical review of the manuscript. Y.P. is an incumbent of the Aser Rothstein Career Development Chair in Genetic Diseases and is a Fellow of the Hurwitz Foundation for Complexity Sciences. We thank the Leo and Julia Forchheimer Center for Molecular Genetics and the Ben May Foundation for grant support. This paper is dedicated to the memory of I. Kafri.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 November 2004; accepted 25 January 2005
Published online at http://www.nature.com/naturegenetics/

1. Conant, G.C. & Wagner, A. Duplicate genes and robustness to transient gene knock-downs in Caenorhabditis elegans. *Proc. R. Soc. Lond. B Biol. Sci.* **271**, 89–96 (2004).
2. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
3. Nowak, M.A., Boerlijst, M.C., Cooke, J. & Smith, J.M. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
4. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
5. Lynch, M. & Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
6. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
7. Wagner, A. The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* **154**, 1389–1401 (2000).
8. Gu, Z., Nicolae, D., Lu, H.H. & Li, W.H. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* **18**, 609–613 (2002).
9. Huh, W.K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
10. Papp, B., Pal, C. & Hurst, L.D. Evolution of cis-regulatory elements in duplicated genes of yeast. *Trends Genet.* **19**, 417–422 (2003).
11. van den Berg, M.A. *et al.* The two acetyl-coenzyme A synthetases of Saccharomyces cerevisiae differ with respect to kinetic properties and transcriptional regulation. *J. Biol. Chem.* **271**, 28953–28959 (1996).
12. Kratzer, S. & Schuller, H.J. Transcriptional control of the yeast acetyl-CoA synthetase gene, ACS1, by the positive regulators CAT8 and ADR1 and the pleiotropic repressor UME6. *Mol. Microbiol.* **26**, 631–641 (1997).
13. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
14. Jansen, R., Greenbaum, D. & Gerstein, M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* **12**, 37–46 (2002).
15. Ge, H., Liu, Z., Church, G.M. & Vidal, M. Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. *Nat. Genet.* **29**, 482–486 (2001).
16. Lynch, M. & Katju, V. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* **20**, 544–549 (2004).
17. Teichmann, S.A. & Babu, M.M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).
18. Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, RESEARCH0008 (2002).
19. Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme 'dispensability' in yeast. *Nature* **429**, 661–664 (2004).
20. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E.S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
21. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
22. Cavalcanti, A.R., Ferreira, R., Gu, Z. & Li, W.H. Patterns of gene duplication in Saccharomyces cerevisiae and Caenorhabditis elegans. *J. Mol. Evol.* **56**, 28–37 (2003).
23. Steinmetz, L.M. *et al.* Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400–404 (2002).
24. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
25. Pilpel, Y., Sudarsanam, P. & Church, G.M. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* **29**, 153–159 (2001).
26. Kellis, M., Birren, B.W. & Lander, E.S. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* **428**, 617–624 (2004).
27. Lord, P.W., Stevens, R.D., Brass, A. & Goble, C.A. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* **19**, 1275–1283 (2003).
28. Ozcan, S. Two different signals regulate repression and induction of gene expression by glucose. *J. Biol. Chem.* **277**, 46993–46997 (2002).
29. McCammon, M.T. & McAlister-Henn, L. Multiple cellular consequences of isocitrate dehydrogenase isozyme dysfunction. *Arch. Biochem. Biophys.* **419**, 222–233 (2003).
30. Garcia-Rodriguez, L.J. *et al.* Characterization of the chitin biosynthesis process as a compensatory mechanism in the fks1 mutant of Saccharomyces cerevisiae. *FEBS Lett.* **478**, 84–88 (2000).