ANALYSIS

nature genetics

Multiple knockout analysis of genetic robustness in the yeast metabolic network

David Deutscher[1], Isaac Meilijson[2], Martin Kupiec[3] & Eytan Ruppin[1,4]

Genetic robustness characterizes the constancy of the phenotype in face of heritable perturbations. Previous investigations have used comprehensive single and double gene knockouts to study gene essentiality and pairwise gene interactions in the yeast *Saccharomyces cerevisiae*. Here we conduct an *in silico* multiple knockout investigation of a flux balance analysis model of the yeast's metabolic network. Cataloging gene sets that provide mutual functional backup, we identify sets of up to eight interacting genes and characterize the '*k* robustness' (the depth of backup interactions) of each gene. We find that 74% (360) of the metabolic genes participate in processes that are essential to growth in a standard laboratory environment, compared with only 13% previously found to be essential using single knockouts. The genes' *k* robustness is shown to be a solid indicator of their biological buffering capacity and is correlated with both the genes' environmental specificity and their evolutionary retention.

In laboratory conditions, about 19% of the genes in the yeast *S. cerevisiae* are essential[1]; that is, their null mutation is lethal to the organism (see also the *Saccharomyces* Genome Database (SGD) (http://www.yeastgenome.org)). All other genes are apparently dispensable, demonstrating genetic robustness[2,3]. Several authors[3–8] have provided three explanations accounting for this observed dispensability: (i) a gene's function might be buffered by duplication or overlap at either the sequence or the molecular function levels (also termed degeneracy, genetic buffering[5] or, often, redundancy[3,9]); (ii) a gene's function might be buffered by an alternative biochemical pathway (functional complementation[5]); or (iii) a gene might be involved in processes that are required only under untested environmental conditions[4]. The first two mechanisms involve functional backup interactions between genes, the main subject of this study.

Gene essentiality and pairwise genetic interactions have been previously investigated using large-scale single and double knockout studies in yeast[1,10–14]. Here we go beyond gene essentiality and chart the architecture of robustness against gene knockouts of the yeast metabolic network, employing large-scale deep multiple knockouts in an *in silico* model. Such knockouts have been used experimentally to study small-scale networks[7], but large-scale multiple knockouts[11] are still scarce owing to the high combinatorial number of experiments involved. Two recent papers performed all double knockouts of yeast and the bacterium *H. pylori* metabolic genes using *in silico* models[12,13].

Multiple knockouts, essential sets and *k* robustness

Extending the common notion of essentiality to the realm of genetic robustness via multiple knockouts, we define a gene as 'contributing' to the organism's viability and growth if it is a member of an 'essential gene set'. This denotes a set of genes whose combined knockout results in a mutant strain with very slow or no growth (relative to the wild-type growth rate) but where the growth rate of a mutant missing only a subgroup of these genes remains high. Hence, the functioning of any one gene in an essential set buffers against the concomitant knockout of all other genes in the set, providing a basic functional backup and indicating the existence of pairwise backup interactions (also termed synthetic[11], aggravating[12] or synergistic[14] interactions). We denote the system as '*k* robust' to a specific gene knockout according to the size *k* of the smallest essential gene set that includes the knocked-out gene (its interaction depth). Thus, the system is 1-robust to knockout of an essential gene, 2-robust to knockout of any nonessential gene that is involved in a synthetic lethal pair[11], and so on. This definition of *k* robustness subsumes the set of essential genes, creating a higher-level dichotomy of contributing versus noncontributing genes. It extends the classical notion of an essential contribution of a gene to its potential contribution in face of possibly larger genetic or environmental perturbations. We further denote as 'coessential' genes that are in a common essential set. Our definition of essential gene sets is similar to that of minimal cut sets introduced in ref. 15, but the calculation in that work relies on the use of elementary modes[16], currently feasible only for small-scale networks.

We study genetic robustness using a previously reconstructed[17–19] flux balance analysis[20] (FBA) model of the metabolic network of the yeast, incorporating 708 genes, 1,175 reactions and 584 metabolites. Our investigation is focused on those 484 model genes with known ORFs whose product enzyme is not on a dead-end pathway in the model[4] (**Supplementary Table 1** online). (The analysis excludes fictitious genes, which catalyze reactions that are known or assumed to be available to the yeast according to biochemical literature but that are not annotated to any known ORF.) The FBA analysis takes into consideration the structure, stoichiometry and basic thermodynamics of

[1]School of Computer Science, [2]School of Mathematical Sciences, [3]Department of Molecular Microbiology and Biotechnology and [4]School of Medicine, Tel Aviv University, PO Box 39040, Tel Aviv 69978, Israel. Correspondence should be addressed to E.R. (ruppin@post.tau.ac.il).

Published online 29 August 2006; doi:10.1038/ng1856

Table 1 The overlap between several gene or protein pair characteristics (C) and the 'coessentiality' property (B)

Characteristic C	B only	C only	B&C	Neither	*P* value	Large-scale experimental *P* value	*P* value for non-isoenzymes
Sequence homology	2,837	386	328	113,335	5×10^{-314}	4×10^{-22}	0.002
Similar biological process	2,751	3,249	431	110,455	8×10^{-147}	$<2 \times 10^{-322}$	7×10^{-22}
Same biological process	2,877	1,511	305	112,193	2×10^{-145}	5×10^{-296}	6×10^{-7}
Same subcellular localization	2,188	24,210	994	89,494	4×10^{-38}	2×10^{-70}	3×10^{-7}
Common regulatory motifs	1,439	8,072	379	51,535	3×10^{-17}	-	2×10^{-6}
Same MIPS mutant phenotype	3,155	129	27	113,575	2×10^{-14}	9×10^{-316}	5×10^{-11}
Physical interaction (DIP)	3,148	59	17	113,662	2×10^{-11}	-	0.11
Physical interaction: same MIPS complex (mostly TAP, HMS-PCI)	3,100	1,131	65	112,590	2×10^{-7}	4×10^{-6}	0.01
Correlated expression, (Rosetta) CC > 0.7	3,048	131	6	112,255	0.16	0.79	0.64
Correlated expression, (Rosetta) CC < –0.7	3,054	23	0	112,363	1	0.37	1

Table entries indicate the number of pairs that have the property or combination of properties indicated, and *P* values are from Fisher's exact test. The next-to-last column indicates, for comparison, the results obtained by ref. 11 measuring the overlap between experimental genetic interactions and the corresponding characteristics. The last column lists corresponding *P* values obtained by considering the 2,866 (90%) non-isoenzyme coessential gene pairs alone. CC: correlation coefficient. The data sets are detailed in Methods.

the metabolic network, applying mass-balance constraints to predict phenotypes and other properties with general prediction accuracy of 70–90% (ref. 20) and single-deletion mutant viability with 89% accuracy (**Supplementary Note**). Our analysis is performed in two stages: in the first, we exhaustively search through the space of all possible combinations of concomitant, multiple knockouts of genes, up to the concomitant knockout of four genes. We record the essential sets found and list the contributing genes with their *k* robustness levels. Because further exhaustive search is currently computationally infeasible, the second stage uses a stochastic sampling method to identify genes with *k* robustness levels >4 (see Methods). The FBA model estimates the organism's potential to grow under various conditions, though in reality the organism may not use all this potential owing to additional non-modeled constraints (for example, non-optimal gene expression resulting from regulatory constraints). Therefore, the *k* robustness we record is actually an approximation of the true, experimental value, reflecting the backup potential provided by the network structure and stoichiometry. Finally, essential sets are marked as based on functional duplication if all genes in the set catalyze the same essential reaction, as alternative pathways if there are no isoenzymes in the set or as a mixed mechanism otherwise. Each gene is tagged with one or both types of functional backup.

Coessential genes and their corroboration

Our study focused on a standard synthetic rich medium[17] (see Methods). Using an exhaustive multiple knockout search, we found 48 essential genes, 14 essential pairs, 17 triplets and 39 essential quadruples, overall involving 159 contributing genes. The gene knockout sampling method identified an additional 173 contributing genes with *k* robustness levels >4, the vast majority of which are 10-robust or less. Inspection of the list of reactions in the metabolic network identified an additional 28 contributing genes that catalyze essential reactions but that are backed up by at least four duplicated isoenzymes. We repeated the same procedure using a glucose minimal medium for comparison. The essential sets found are detailed in **Supplementary Table 1**.

Validating these model predictions is not straightforward, as almost no experimental multiple knockouts of the yeast's metabolic genes are available. Considering the very few relevant known synthetic lethal interactions, FBA predictions of coessential genes (pairs that are in the same set; see **Supplementary Table 1**) achieve good recall, given our use of sampling (59%; **Supplementary Note**). To further test the model's accuracy, one can measure the percentage of experimentally essential genes in each *k* robustness level. Ideally, one would expect that all 1-robust genes, but no other genes, be experimentally identified as essential if the model were completely accurate. The true picture (**Fig. 1**) depicts a rapid decrease in the fraction of essential genes with rising *k* robustness levels, showing that *k* robustness is indeed a clear indicator of the biological buffering capacity. To further corroborate the validity of the pairwise interactions predicted by the model between members of the same essential set, we followed the procedures laid out in ref. 11 to search for other possible biological pairwise relations that correspond with these interactions. The list of predicted interacting gene pairs is indeed significantly enriched with many experimentally measured pairwise characteristics (**Table 1**). Although this enrichment is expected for isoenzyme pairs, it remains valid even when considering only non-isoenzyme coessential genes. In addition, we find that the expression patterns of coessential pairs are

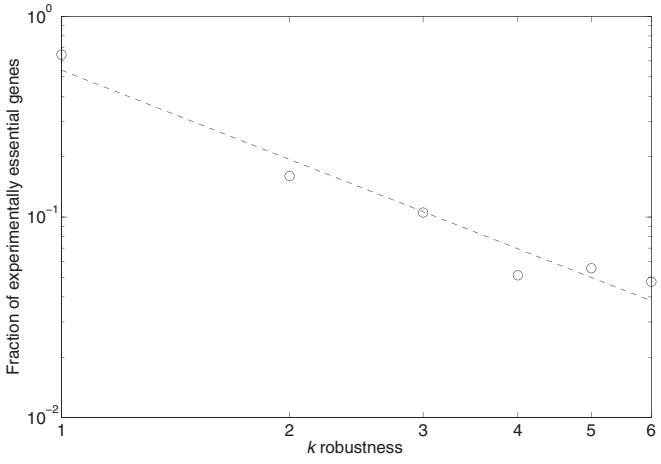

Figure 1 Fraction of essential genes in each *k* robustness level. Essentiality is determined according to large-scale experiments (see the SGD). The straight line is the linear regression fit. Data is presented for $k \leq 6$, as the number of genes in higher levels is very small.

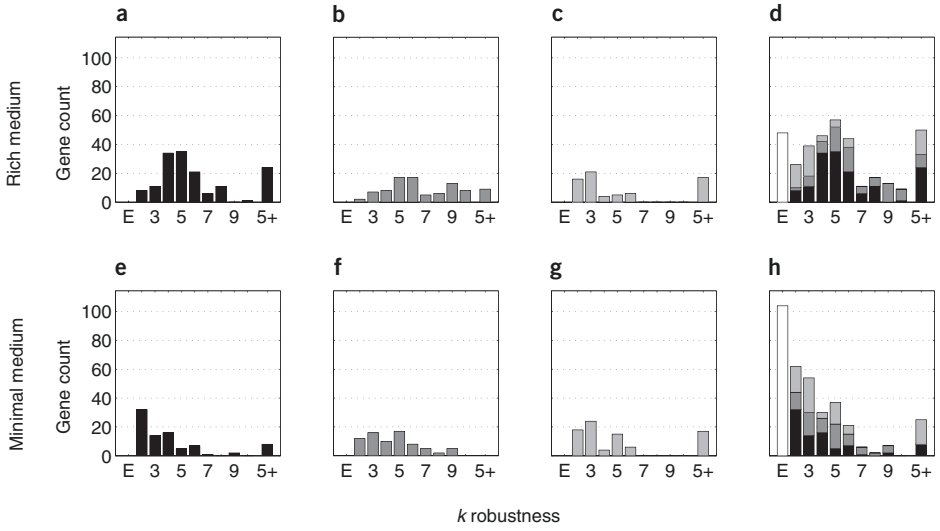

Figure 2 *k* robustness gene histograms and the distribution of backup mechanisms for contributing genes at different levels of *k* robustness, on rich (**a**–**d**) and minimal media (**e**–**h**). Backed up genes are black if backed up by alternative pathways (**a**,**e**), light gray if backed up by duplication (**c**,**g**) and dark gray if backed by both mechanisms (**b**,**f**). **d** and **h** present the total counts, with the leftmost column in each panel depicting essential genes with no backups. Genes with robustness levels 5 and up are found using stochastic search, and their robustness level might be overestimated (Methods). The rightmost column in each pane counts contributing genes whose robustness level remains undetermined or is >10, including 17 genes encoding various hexose transporters comprising a single duplicated-function essential set.

more coherent and similar than that of random pairs (**Supplementary Note**). Finally, *k* robustness values (either for all genes or only for non-isoenzyme pairs) are correlated with several other properties of genes, including evolutionary conservation, environmental specificity and expression levels (**Supplementary Table 2** online), showing that *k* robustness indeed has a biological meaning, related to other genetic properties. Nonetheless, this does not suggest any causal relation between genetic robustness, or *k* robustness, and other genetic qualities, a subject that is still a contentious issue[2,21–24].

The nature of the genetic interactions depicted in essential sets is demonstrated in the following example concerning the pentose phosphate pathway: ribose 5-phosphate is a critical precursor in the synthesis of nucleic acids, which are needed in high amounts in growing cells, and is produced by the pentose phosphate pathway using either the oxidative or nonoxidative branches[25]. Therefore, it is not surprising that the disruption of both branches is predicted to be lethal, giving rise to several essential sets, such as the combination of glucose 6-phosphate dehydrogenase (*ZWF1*) and the two transketolases (*TKL1*, *TKL2*), a combination that previously has been found experimentally[26] (see additional examples in the **Supplementary Note** and **Supplementary Fig. 1** online).

The architecture of metabolic robustness

We analyzed these results on a large scale (**Fig. 2** shows a histogram of the genes' robustness levels). The contributing genes total 74% (360) of the tested genes, compared with only 10% of these genes that are identified *in silico* as essential using traditional single knockouts[17] (and 13% previously found *in vivo*[1]; see also the SGD). This indicates that a large majority of the genes are involved in processes already required in the standard laboratory rich environment, even though the individual genes are not essential. Using a glucose minimal medium, a slightly smaller set of 72% of the genes is uncovered at markedly lower *k* robustness levels. These differences arise mainly from the more extensive activity of membrane transporters and catabolic pathways in the rich medium, increasing the number of contributing genes and the overall k robustness (as synthesis and transport buffer each other). These media-dependent changes are described in more detail in the **Supplementary Note**.

Backups arise more often from alternative pathways than from functional gene duplication (**Fig. 2**), the former being solely responsible for 45% of the backed up genes and partially responsible for 33% more. Furthermore, considering all coessential gene pairs, only 10% involve genes coding for duplicated isoenzymes. Alternative pathways are particularly dominant in genes with high *k* robustness levels, suggesting that their role in genetic robustness might be underestimated when the investigation is limited to shallow knockout depths. Another notable quality of backup interactions is transitivity, or the formation of dense neighborhoods[11]: we find that the probability of a backup interaction between two genes is significantly higher ($P < 10^{-323}$) if both genes are backed up by a common third gene (a fivefold increase, from 2.7% in general to 14% among genes with a common neighbor). In agreement with ref. 7, we find that the number of essential sets per gene is usually small, averaging 8 sets or 22 pairwise interactions per gene, although a few genes are involved in many interactions (**Supplementary Fig. 2** online).

We used the Gene Ontology (GO)-Slim biological process annotations from the SGD (http://www.yeastgenome.org; November 2005) to test if any biological process category is significantly enriched or depleted with backed-up genes, as portrayed in **Figure 3** (see Methods). Indeed, two main metabolic functions, amino acid metabolism and generation of precursor metabolites and energy are highly backed up ($P = 3 \times 10^{-11}$ and $P = 0.01$, respectively). In contrast, genes functioning in lipid metabolism contain significantly more essential genes than expected by chance ($P = 1 \times 10^{-10}$), comprising a particularly non-robust functional category. The backup interactions between the functional categories in the metabolic network are shown in **Figure 4**. The functional categories of precursors and energy generation, carbohydrate metabolism, amino acid metabolism and transport processes have notable interfunctional backups (which may be quite intricate when examined in detail; see examples in **Supplementary Note**). These and other categories also have significantly elevated levels of intrafunctional backup interactions, although overall, interfunctional backup interactions are abundant. This is evident also with the higher resolution possible by using the full GO annotation: defining two GO terms as similar if they are annotated with significantly overlapping gene sets[11,27], we find that only 18% of backup interactions are between genes annotated with similar GO terms, comparable to the 27% found experimentally in ref. 11 and differing from previous observations in small-scale systems[7] (see also **Supplementary Fig. 3** and **Supplementary Methods** online). It should be noted that the qualitative similarities between the findings in ref. 11 and our

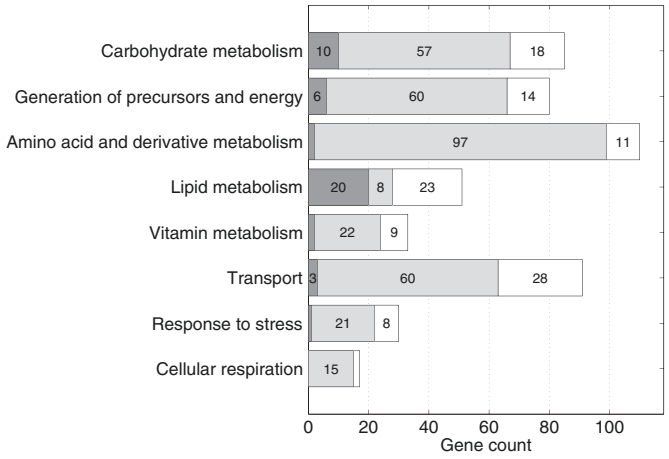

Figure 3 Metabolic network robustness across different functional GO-Slim categories on rich medium, showing for each category the proportions of essential genes (dark), backed up genes (light), and genes not found to contribute in our analysis (white). Superimposed numbers indicate gene counts (for clarity, only counts of 3 or more are indicated). Only categories annotated with at least ten genes are included. The respective measurements in glucose minimal medium are very similar, except that many more of the genes involved in amino acid and derivative metabolism are essential (45/110).

findings also include the existence of dense neighborhoods and arise even though the studies involve different subsets of the yeast genome and use *in silico* versus *in vivo* knockouts of different depths.

Robustness, dispensability and evolution

Two previous FBA-based studies[4,28] of the mechanisms for dispensability reported that the majority of dispensable metabolic genes are specific to certain environmental conditions, concluding that environmental specificity is the dominant explanation behind dispensability. They further showed that gene duplication is the second common explanation. We find that 91% of the condition-specific genes identified in ref. 4 are contributing genes (as defined above) already in the standard rich environment. There is a significant correlation between the *k* robustness of genes and their environmental specificity, measured as the number of environments where the gene is dispensable ($R = 0.39$, $P = 9 \times 10^{-11}$, $N = 252$ (B. Papp, personal communication; **Fig. 5a** and **Supplementary Note**)). That is, genes with many backups tend to catalyze reactions that are essential in only a few specific environments. This may suggest that the availability of backups allows for the functional divergence and specification of genes with high *k* robustness to specific environments during evolution. It has been reported[28] that redundancy (duplication) is an important cause of metabolic network robustness to single-gene deletions during growth on glucose (minimal medium). This conclusion can also be seen in our results (**Fig. 2**). However, when extending the analysis to multiple gene knockouts, we find that at higher depths, and especially in the more complex rich medium, the role of alternative pathways towards genetic robustness is more prominent than that of duplication.

To examine the extent to which the *k* robustness of genes may actually confer them with a functional backup from an evolutionary perspective, we compared the genes' *k* robustness with the propensity for gene loss[21,22] (PGL data courtesy of Y. Wolf, personal communication; **Fig. 5b**), which is an (inverse) measure of the evolutionary conservation of genes. The resulting significant correlation ($R = 0.23$, $P = 1 \times 10^{-4}$, $N = 278$) shows that genes with high *k* robustness are less conserved and hence testifies that they are indeed functionally buffered, permitting their divergence. This conclusion is further strengthened by the finding that the PGL scores of coessential genes are significantly more similar, or coherent, than those of random gene pairs. This is unsurprising for homologous or isoenzyme pairs but is true even when disregarding them: the average absolute difference in PGL scores of non-isoenzyme coessential genes is 27% lower than the average for all gene pairs ($P = 3 \times 10^{-66}$, Wilcoxon's rank-sum test; similar results hold considering nonhomologous coessential genes). It seems that common evolutionary forces were imposed on backup gene pairs to channel them in similar evolutionary paths. Although still controversial, previous studies have found that environmental specificity and gene expression are both correlated with evolutionary conservation of genes[4,21–24]. As both are also correlated with *k* robustness (**Supplementary Table 2**), we verified that *k* robustness and PGL are correlated even when statistically controlling for these variables (Pearson's partial $R = 0.22$, $P = 5 \times 10^{-4}$, $N = 233$; **Supplementary Note**).

It is important to note that genetic robustness did not necessarily evolve because it was favored by natural selection[2]. This explanation, termed 'adaptive' robustness, claims that for well-adapted traits, mutations derive a non-optimal phenotype and hence decrease fitness. However, alternative 'intrinsic' theories—often raised in the context of dominancy but relevant to genetic robustness in general—view robustness as a correlated side effect of the evolution of other properties, such as higher metabolic efficiency[4], or even as an inherent property of complex, evolving systems[3]. An intermediate, 'congruent' possibility points to the tight coupling between genetic robustness and environmental robustness (buffering of non-heritable perturbations), as many mechanisms allow both (for instance, buffering between transport and synthesis). As environmental perturbations occur at a higher frequency, this view posits the evolution of genetic robustness as a side effect of the evolution of environmental robustness[29]. Our findings do not contradict any of these possibilities.

Our investigation leaves the contribution of 26% of the genes undetected, and assuming that genes retained by evolution do fulfill some

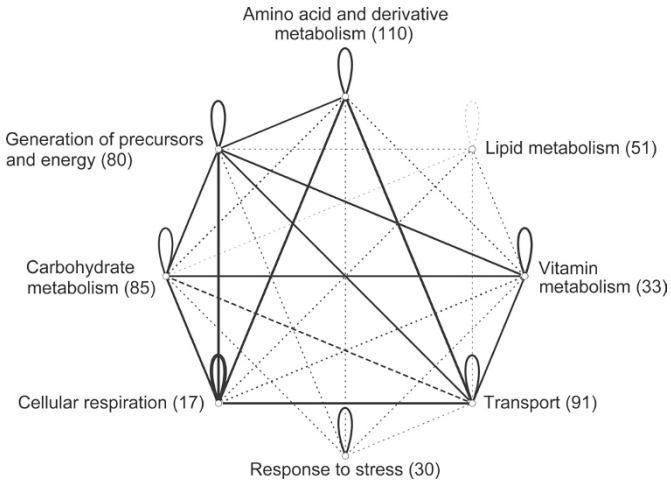

Figure 4 Functional backup capacity on rich medium. Vertices of the graph represent GO-Slim biological process categories (annotated with at least ten genes). Dotted edges connect categories if there are any two genes in a common essential set that are annotated one to each category. Dashed edges indicate a higher-than-average frequency of such gene pairs, whereas solid edges indicate a statistically significant high frequency (see Methods). Edge width correlates with the logarithm of the frequency. Numbers in brackets indicate how many genes are annotated to each functional category.

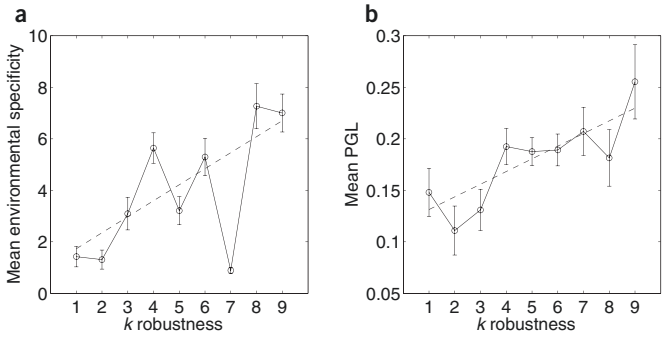

Figure 5 Environmental specificity (ES) and propensity for gene loss (PGL) as a function of robustness level. Means ± s.e.m. are shown for the ES (**a**) and PGL (**b**) measures at each *k* robustness level. The dashed lines are the least squares linear regression through the original data points. Owing to their small number and the uncertainty in their robustness level estimation, we do not consider genes with *k* robustness >9, although the significant correlations found remain valid across *k* robustness thresholds from 5–12. The correlation between *k* robustness and PGL goes beyond the previously reported correlation between essentiality and evolutionary conservation[21–24], as it remains significant even when considering only nonessential genes ($R = 0.26$, $P = 5 \times 10^{-5}$, $N = 235$).

function, they should be accounted for. First, we verified that at least 12% of the genes are non-contributing in rich medium (see Methods). Second, some genes might be heavily backed up and escape detection because of the depth limit of our investigation, the 'optimistic' bias of the model and its inaccuracies (see Methods). Last, some tested genes might be backed up by model genes without a known ORF, which were excluded from the current study. Even so, the contributions of almost three-quarters of yeast metabolic genes are detected, uncovering the underlying architecture of robustness of yeast metabolism.

METHODS

The model. We use the constraint-based model of ref. 18, focusing on the 484 genes with known ORFs that are not on a dead-end pathway[4] (that is, at least one of their catalyzed reactions' products is a substrate for another reaction that is itself not on a dead-end and vice versa). Growth on both rich and minimal media was simulated under aerobic conditions. The minimal medium included glucose, oxygen, ammonia, phosphate, sulfate and potassium. The rich medium included, in addition, 20 amino acids, purines and pyrimidines[17].

The FBA finds an upper bound on the obtainable growth rate of the organism and hence has an optimistic bias, falsely predicting viability more often than falsely predicting lethality (80% of the errors are false positives[17]). Hence, rather than falsely attributing contribution, we are more likely to miss some contributing genes and detect the contributing genes at *k* robustness levels higher than their real level.

Search for backed up genes and essential sets. We performed an exhaustive search, which included all gene sets of up to four genes. Each set marked essential had a lethal knockout phenotype (growth rate <20% of the wild-type rate) with all subset knockout mutants viable (growth rate >80%; see **Supplementary Methods** and **Supplementary Fig. 4**). If all genes in an essential set were isoenzymes catalyzing the same essential reaction, their backup was attributed to duplication. Alternatively, if no isoenzymes were found in an essential set, backup was attributed to alternative pathways. When both isoenzymes and other genes were found in a common essential set, the genes encoding the isoenzymes were tagged with both types of backup mechanisms, whereas the other genes were obviously tagged as buffered by alternative pathways solely.

The exhaustive search took a week using a cluster of ten computers. Testing all combinations of five knockouts would have increased the computational resources needed by two orders of magnitude and would have required about two years. Thus, we searched for genes that are more than 4-robust using stochastic sampling methods, requiring an additional 2 weeks on the computer cluster.

To stochastically test whether gene X is contributing, we repeatedly tested random knocked-out mutants, each missing a large number of knocked-out genes but leaving gene X intact. Finding such a knockout configuration that is itself viable but then becomes lethal when gene X is knocked out (meaning that all its backups are already silenced) provides proof of X's contribution. As the probability of finding such an event can be estimated analytically assuming that gene X is *k* robust (**Supplementary Methods**), one can bound the probability that the said gene is contributing at a given *k* robustness level by repeating this stochastic test a sufficient number of times. We calibrated parameters of the stochastic testing for a misdetection rate of 10^{-2} for 8-robust genes (on rich medium, or 6-robust genes on minimal medium, according to computational feasibility) and validated them on the sets of contributing genes found with the exhaustive search. This implies an extremely high detection rate for genes with robustness levels <8, a 45% misdetection probability of 9-robust genes, and an 85% misdetection probability of 10-robust genes. We were usually able to extract essential sets from the large knockout groups during the stochastic search (**Supplementary Methods**), additionally giving upper bounds on *k* robustness levels.

We have used the model to compute the maximal and minimal possible flux through any reaction, given that the growth rate is at least 80% of the wild-type growth. This computation enables us to identify reactions that always have a zero flux under these conditions, even after multiple knockouts, and hence identify genes that are noncontributing in the rich environment tested.

Experimental gene pair characteristics. We evaluated the statistical significance of the overlaps in **Table 1** using Fisher's exact test. The experimental data sets are as follows. (i) Sequence homology using BLAST *E* values below 10^{-4}. (ii) Same or similar biological process GO annotation (see the SGD): genes were considered to have the same process annotation if they shared at least one direct biological process. Two GO annotations were considered similar if the sets of genes annotated to each one (including genes annotated to descendent terms in the ontology) were significantly and strongly overlapping. Significance was evaluated using Fisher's exact test, corrected for multiple testing by limiting the false discovery rate (FDR)[30] to 10%. Strength of association was determined by LOD[11] values >3. Genes were considered similarly annotated if at least one of their annotations (one annotation of each gene) was similar. (iii) Same subcellular localization (that is, sharing at least one direct cellular component GO annotation; see the SGD). (iv) Common regulatory motifs. Motifs are from ref. 31. Considering only genes that have at least one regulatory motif attached in the data, we listed all gene pairs that have at least one common motif. (v) Same MIPS mutant phenotype. This was determined according to the list of phenotypes in the MIPS database (http://mips.gsf.de/genre/proj/yeast/, August 2005), excluding nonspecific phenotype categories (categories with names including the word 'other', categories with more than 200 genes, and those at the least specific level of the hierarchy). (vi) Physical interaction (DIP). The protein-protein interactions, based on the DIP database, were taken from ref. 32 (data courtesy of R. Sharan). Only interactions with a positive probability were considered. (vii) Physical interaction (MIPS complex). This means participation in at least one protein complex listed in MIPS (usually from large-scale tandem affinity purification (TAP) or high-throughput mass spectrometric protein complex identification (HMS-PCI) experiments). (viii) Correlated expression (Rosetta). This was computed among the expression vectors of each gene in the 300 conditions of the Rosetta compendium[33] (ignoring missing values). Correlation coefficients >0.7 or <–0.7 were considered.

Functional qualities of essential sets. For each GO-Slim biological process category at the SGD, we tested the number of backed up genes out of all genes annotated to that category, compared with a random distribution of the contributing genes across categories. *P* values are from Fisher's exact test, corrected for multiple testing by controlling the FDR[30] at 10%. Two GO terms were defined as similar if they had a significant overlap of annotated genes[11] (after genes were annotated with all ancestor terms in the GO hierarchy) using the same statistical test. We counted the percentage of gene pairs annotated with such similar terms, out of all coessential gene pairs, to find that most coessential pairs are not annotated with similar terms. Similar results were obtained when considering the semantic similarity[27] of GO terms (**Supplementary Methods**). The existence of the dense neighborhoods' quality was tested by considering all coessential gene pairs using the procedure of ref. 11 (when examining a specific pair of interacting genes, care

was taken to exclude all other interactions arising from their common essential set). For pairwise backups between two GO-Slim categories, we counted the number of gene pairs, one from each category, and the proportion of such pairs that are coessential, compared with a random distribution of such coessential gene pairs (Fisher's exact test, corrected for multiple testing).

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

ACKNOWLEDGMENTS

We thank the Tauber fund for supporting D.D. Discussions with and comments of A. Hirsh, A. Kaufman, O. Meshi, Y. Pilpel, T. Pupko, R. Sharan, T. Shlomi and I. Venger are much appreciated. Figure 4 was drawn using Pajek from http://vlado.fmf.uni-lj.si/pub/networks/pajek/. M.K.'s work was supported by grants from the Israeli Science Foundation (ISF) and the Israeli Ministry of Health. E.R.'s research is supported by the Yishayahu Horowitz Center for Complexity Science, the Israeli Science Foundation (ISF), and the German-Israeli Foundation for scientific research and development (GIF).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
2. De Visser, J.A. *et al.* Perspective: evolution and detection of genetic robustness. *Evolution Int. J. Org. Evolution* **57**, 1959–1972 (2003).
3. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
4. Papp, B., Pál, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
5. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J. & Doyle, J. Robustness of cellular functions. *Cell* **118**, 675–685 (2004).
6. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
7. Hartman, J.L., IV, Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
8. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nat. Genet.* **24**, 355–361 (2000).
9. Nowak, M.A., Boerlijst, M.C., Cooke, J. & Maynard Smith, J. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
10. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
11. Tong, A.H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
12. Segrè, D., DeLuna, A., Church, G.M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2004).
13. Thiele, I., Vo, T.D., Price, N.D. & Palsson, B.Ø. An expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an *in silico* genome-scale characterization of single and double deletion mutants. *J. Bacteriol.* **187**, 5818–5830 (2005).
14. Elena, S.F. & Lenski, R.E. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398 (1997).
15. Klamt, S. & Gilles, E.D. Minimal cut sets in biochemical reaction networks. *Bioinformatics* **20**, 226–234 (2004).
16. Schuster, S., Fell, D. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332 (2000).
17. Förster, J., Famili, I., Palsson, B.Ø. & Nielsen, J. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae. OMICS* **7**, 193–202 (2003).
18. Förster, J., Famili, I., Fu, P., Palsson, B.Ø. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
19. Famili, I., Förster, J., Nielsen, J. & Palsson, B.Ø. *Saccharomyces cerevisiae* phenotypes can be predicted using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. USA* **100**, 13134–13139 (2003).
20. Kauffman, K.J., Prakash, P. & Edwards, J.S. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14**, 491–496 (2003).
21. Krylov, D.M., Wolf, Y.I., Rogozin, I.B. & Koonin, E.V. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* **13**, 2229–2235 (2003).
22. Wolf, Y.I., Carmel, L. & Koonin, E.V. Unifying measures of gene function and evolution. *Proc. Biol. Sci.* **273**, 1507–1515 (2006).
23. Hirsh, A.E. & Fraser, H.B. Protein dispensability and rate of evolution. *Nature* **411**, 1046–1049 (2001).
24. Papp, B., Pál, C. & Hurst, L.D. Genomic function (communication arising): rate of evolution and gene dispensability. *Nature* **421**, 496–497 (2003).
25. Nelson, D.L. & Cox, M.M. *Lehninger Principles of Biochemistry* 3rd edn. (Worth Publishers, New York, 2000).
26. Schaaff-Gerstenschläger, I., Mannhaupt, G., Vetter, I., Zimmermann, F.K. & Feldmann, H. TKL2, a second transketolase gene of *Saccharomyces cerevisiae* – cloning, sequence and deletion analysis of the gene. *Eur. J. Biochem.* **217**, 487–492 (1993).
27. Lord, P.W., Stevens, R., 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. Blank, L.M., Kuepfer, L. & Sauer, U. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* **6**, R49 (2005).
29. Meiklejohn, C.D. & Hartl, D.L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).
30. Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
31. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
32. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* **102**, 1974–1979 (2005).
33. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).