The Genetic Landscape of a Cell
Michael Costanzo, *et al.*
Science **327**, 425 (2010);
DOI: 10.1126/science.1180823

The Genetic Landscape of a Cell

Michael Costanzo,[1,2]* Anastasia Baryshnikova,[1,2]* Jeremy Bellay,[3] Yungil Kim,[3] Eric D. Spear,[4] Carolyn S. Sevier,[4] Huiming Ding,[1,2] Judice L.Y. Koh,[1,2] Kiana Toufighi,[1,2] Sara Mostafavi,[1,5] Jeany Prinz,[1,2] Robert P. St. Onge,[6] Benjamin VanderSluis,[3] Taras Makhnevych,[7] Franco J. Vizeacoumar,[1,2] Solmaz Alizadeh,[1,2] Sondra Bahr,[1,2] Renee L. Brost,[1,2] Yiqun Chen,[1,2] Murat Cokol,[8] Raamesh Deshpande,[3] Zhijian Li,[1,2] Zhen-Yuan Lin,[9] Wendy Liang,[1,2] Michaela Marback,[1,2] Jadine Paw,[1,2] Bryan-Joseph San Luis,[1,2] Ermira Shuteriqi,[1,2] Amy Hin Yan Tong,[1,2] Nydia van Dyk,[1,2] Iain M. Wallace,[1,2,10] Joseph A. Whitney,[1,5] Matthew T. Weirauch,[11] Guoqing Zhong,[1,2] Hongwei Zhu,[1,2] Walid A. Houry,[7] Michael Brudno,[1,5] Sasan Ragibizadeh,[12] Balázs Papp,[13] Csaba Pál,[13] Frederick P. Roth,[8] Guri Giaever,[2,10] Corey Nislow,[1,2] Olga G. Troyanskaya,[14] Howard Bussey,[15] Gary D. Bader,[1,2] Anne-Claude Gingras,[9] Quaid D. Morris,[1,2,5] Philip M. Kim,[1,2] Chris A. Kaiser,[4] Chad L. Myers,[3]† Brenda J. Andrews,[1,2]† Charles Boone[1,2]†

A genome-scale genetic interaction map was constructed by examining 5.4 million gene-gene pairs for synthetic genetic interactions, generating quantitative genetic interaction profiles for ~75% of all genes in the budding yeast, *Saccharomyces cerevisiae*. A network based on genetic interaction profiles reveals a functional map of the cell in which genes of similar biological processes cluster together in coherent subsets, and highly correlated profiles delineate specific pathways to define gene function. The global network identifies functional cross-connections between all bioprocesses, mapping a cellular wiring diagram of pleiotropy. Genetic interaction degree correlated with a number of different gene attributes, which may be informative about genetic network hubs in other organisms. We also demonstrate that extensive and unbiased mapping of the genetic landscape provides a key for interpretation of chemical-genetic interactions and drug target identification.

The relation between an organism's genotype and its phenotype are governed by myriad genetic interactions (*1*). Although a complex genetic landscape has long been anticipated (*2*), exploration of genetic interactions on a genome-wide level has been limited. Systematic deletion analysis in the budding yeast, *Saccharomyces cerevisiae*, demonstrates that the majority of its ~6000 genes are individually dispensable, with only a relatively

[1]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [2]Department of Molecular Genetics, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [3]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [5]Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada. [6]Department of Biochemistry, Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA. [7]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [8]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [9]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [10]Department of Pharmacy, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [11]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [12]S&P Robotics, Inc., 1181 Finch Avenue West, North York, Ontario M3J 2V8, Canada. [13]Institute of Biochemistry, Biological Research Center, H-6701 Szeged, Hungary. [14]Department of Computer Science, Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Princeton University, Princeton, NJ 08544, USA. [15]Biology Department, McGill University, Montreal, Quebec H3A 1B1, Canada.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: cmyers@cs.umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); charlie.boone@utoronto.ca (C.B.)

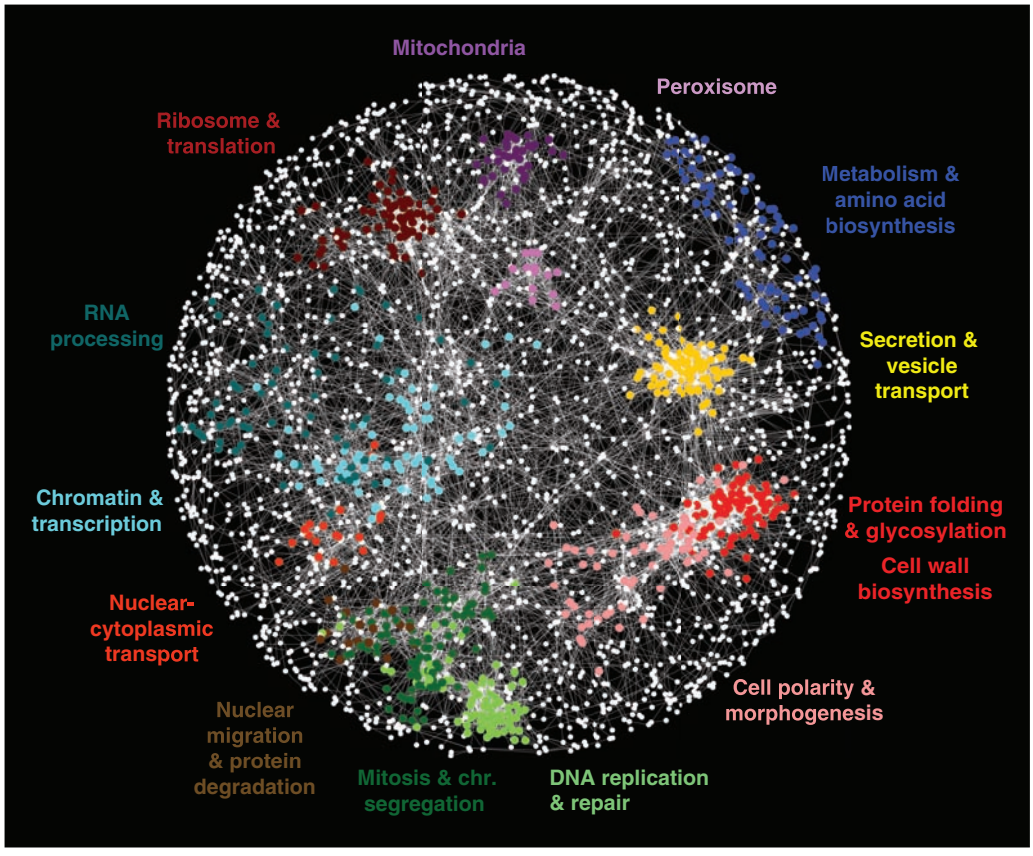
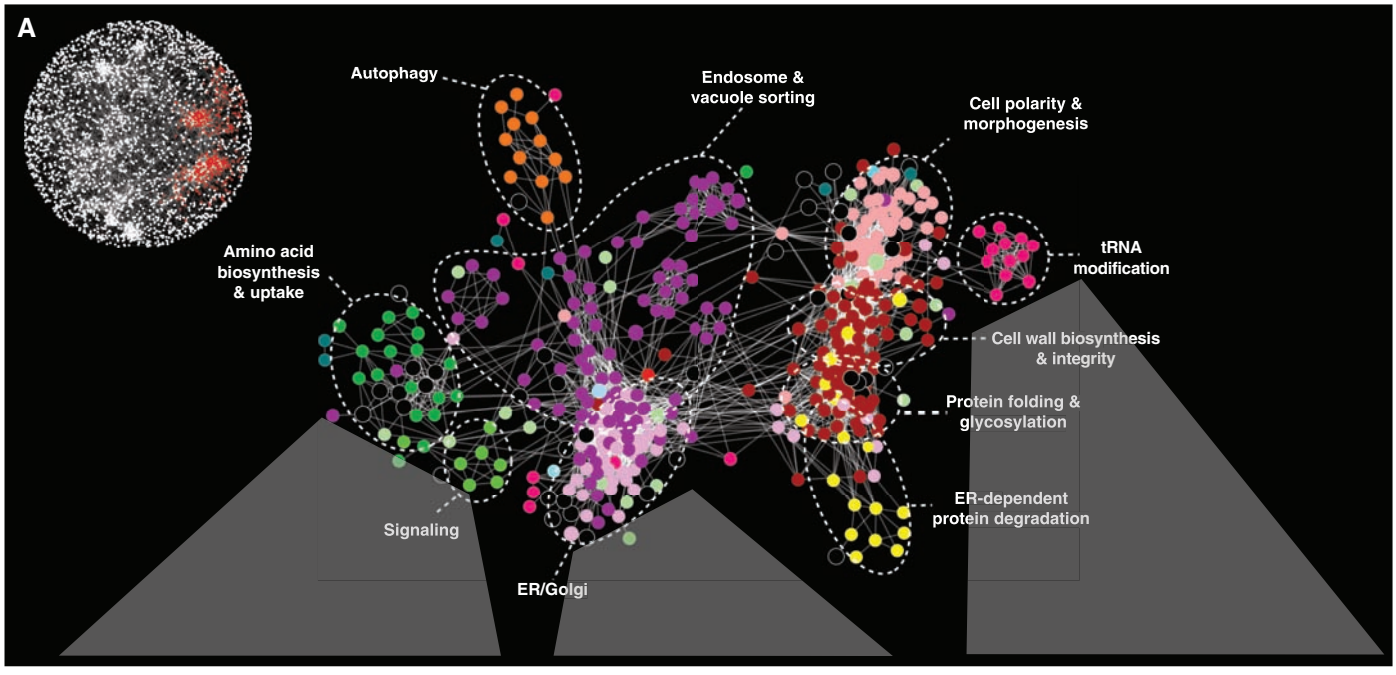
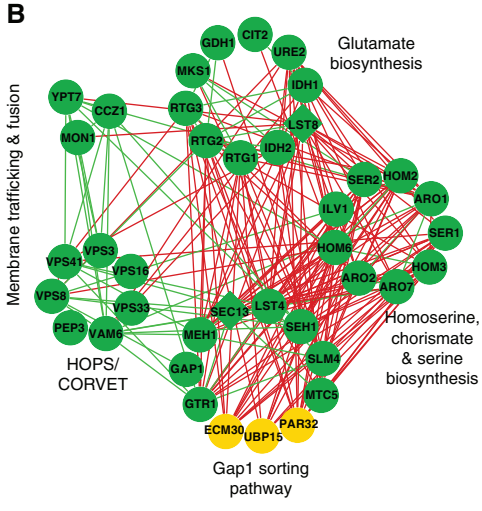
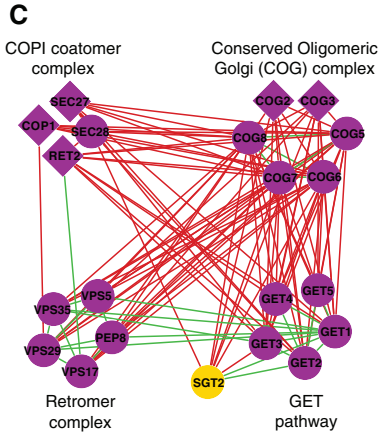
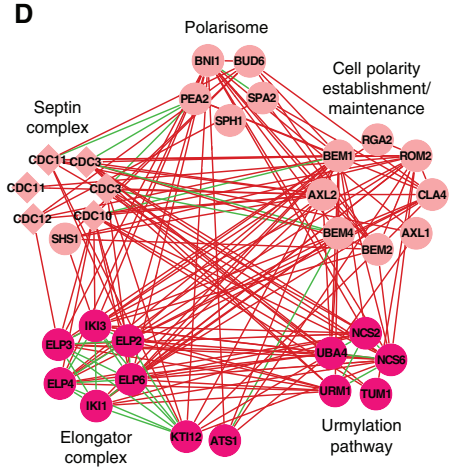


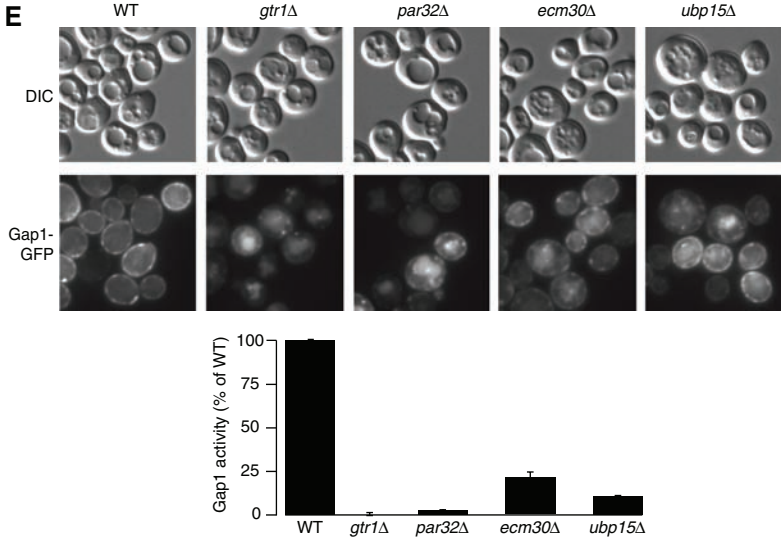
Fig. 1. A correlation-based network connecting genes with similar genetic interaction profiles. Genetic profile similarities were measured for all gene pairs by computing Pearson correlation coefficients (PCCs) from the complete genetic interaction matrix. Gene pairs whose profile similarity exceeded a PCC > 0.2 threshold were connected in the network and laid out using an edge-weighted, spring-embedded, network layout algorithm (*7*, *8*). Genes sharing similar patterns of genetic interactions are proximal to each other; less-similar genes are positioned farther apart. Colored regions indicate sets of genes enriched for GO biological processes summarized by the indicated terms.

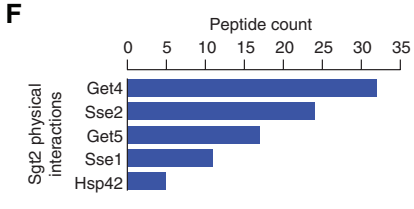


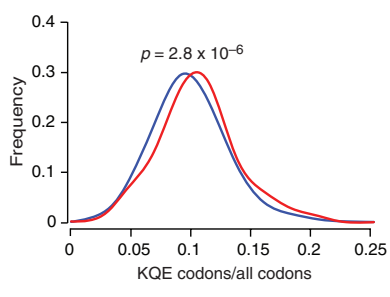












small subset (~20%) required for viability (*1*), which suggests the evolution of extensive buffering against genetic perturbations (*3*). Genome-scale screens for genetic interactions that affect the fitness of a cell or organism can chart the genetic network underlying functional redundancy (*1*). In particular, synthetic genetic array (SGA) methodology (*4*) enables the systematic mapping of synthetic lethal genetic interactions through an automated form of genetic analysis that produces high-density arrays of double mutants (*5*). Here, we report construction of a functionally unbiased genetic interaction map for a eukaryotic cell.

Genome-scale, quantitative analysis of genetic interactions. We consider a digenic interaction as a double mutant that shows a significant deviation in fitness compared with the expected multiplicative effect of combining two single mutants (*6*). Negative interactions refer to a more severe fitness defect than expected, with the extreme case being synthetic lethality; positive interactions refer to double mutants with a less severe fitness defect than expected. To quantitatively score genetic interactions in large-scale SGA screens, we developed a model to estimate fitness defects directly from double-mutant colony sizes (*7*, *8*) (fig. S1A). We screened 1712 *S. cerevisiae* query genes, including 334 conditional or hypomorphic alleles of essential genes, for a total of ~5.4 million gene pairs spanning all biological processes (fig. S1, B and C) (*7*, *8*). These queries were selected randomly with respect to function; however, preference was given to mutants exhibiting fitness defects (*7*, *8*). Comparing fitness estimates of single mutants with their corresponding double-mutant phenotypes identified ~170,000 interactions, a threefold increase over all previously reported genetic interaction data (fig. S1, D and E). Our data captured ~35% of previously reported negative genetic interactions (*7*, *8*) (fig. S1D) and exhibited significant correlation (r = 0.89) (fig. S1F) with genetic interactions identified by high-resolution liquid growth profiles (*7*–*9*), which confirmed the accuracy of our measurements (fig. S1F). Thus, our approach enabled assembly of a quantitative fitness-based profile of genetic interactions on a genome-wide scale.

We determined false-negative and false-positive rates at a defined confidence threshold ($|\varepsilon| > 0.08$, $P < 0.05$) (fig. S2A) (*7*, *8*) and used this filtered data set for all analyses. Data evaluation, by several different measures (*7*, *8*), indicated that interactions that corresponded to specific confidence levels were functionally informative (fig. S2, B and C). In particular, enrichment for Gene Ontology (GO) coannotated gene pairs was correlated with the significance and magnitude of genetic interaction (fig. S2B), as well as with genetic profile similarity (fig. S2C) (*7*, *8*). Notably, we found about twice as many negative interactions as positive genetic interactions (fig. S1B). Moreover, negative genetic interactions tended to be more informative for identifying physical interactions and GO coannotated gene pairs than positive interactions (fig. S2C).

A functional map of the cell. Genes belonging to the same pathway or biological process tend to share similar profiles of genetic interactions (*5*). We exploited this property to construct a global network, grouping genes with similar interaction patterns together: Nodes in this network represent genes, and edges connect gene pairs that share common sets of genetic interactions or similar interaction profiles (Fig. 1). This network highlights genetic relations between diverse biological processes and the inherent functional organization of the cell. Genes displaying tightly correlated profiles form discernible clusters corresponding to distinct bioprocesses, and the relative distance between distinct clusters appears to reflect shared functionality (Fig. 1). For example, the role of the microtubule cytoskeleton in bridging nuclear chromosomal- and actin cytoskeleton–based functions is illustrated by the close proximity and relative positioning of clusters corresponding to genes annotated with roles in cell polarity and morphogenesis, mitosis and chromosome segregation, and DNA replication and repair (Fig. 1). Despite screening only ~30% of the genome as query genes, we recovered genetic interactions for ~75% of the genome because partial genetic interaction profiles were generated for nearly all nonessential genes in the genome. Our data were able to precisely predict known gene functions (GO biological process annotations), as well as or better than all other genome-scale data sets (fig. S2D), and assigned a substantial amount of unique functional information for the genes not captured by previous genetic interaction studies (fig. S2D).

Predicting function and relations. Although complex, the genetic interaction network contains functional information at multiple levels of resolution. The interrogation of the genetic map at higher resolution enabled the dissection of broad biological processes into distinct, yet interdependent, gene cohorts (Fig. 2) [supporting data file S8 (*8*)]. In even more detail, we can also visualize networks in which genes are connected by edges that correspond to genetic interactions directly. Indeed, gene clusters that are correlated by negative (red) and positive (green) genetic interactions reveal network organization reflecting biological pathways and/or protein complexes and their functional integration with one another (Fig. 2, B to D). The genetic interactions occurring between different pathways and complexes were often monochromatic, as predicted previously (*10*), such that they were composed almost exclusively of a single type of genetic interaction, either all negative or all positive.

Genetic clusters were used to predict function for uncharacterized genes on the basis of network connectivity (Fig. 2, A to D). Three genes, *PAR32*, *ECM30*, and *UBP15*, had interaction profiles similar to those of members of the Gap1-sorting module (Fig. 2B), and consistent with a role in this process, all three genes led to Gap1 sorting and transport defects when deleted (Fig. 2E). Additional experimental results (fig. S3) (*11*) suggest that Par32 may function in target of rapamycin (TOR)–dependent regulation of the Gln3, Gat1, Rtg1, and Rtg3 transcription factors (*12*), whereas Ecm30 forms a stoichiometric complex with the Ubp15 ubiquitin protease (*7*, *8*) that may modulate Gap1 localization, perhaps by controlling its ubiquitination state.

In another example, similar genetic interaction profiles suggested a strong functional relation between the GET pathway and the poorly characterized gene, *SGT2* (Fig. 2C). Consistent with a role in endoplasmic reticulum (ER)–dependent membrane targeting (*13*) or protein folding (*14*), we found that Sgt2 physically interacts with Get4, Get5, and heat shock 70

Fig. 2. Magnification of the functional map better resolves cellular processes. (**A**) A subnetwork corresponding to a region of the global map described in Fig. 1 is indicated in red (inset). Node color corresponds to a specific biological process: dark green, amino acid biosynthesis and uptake; light green, signaling; light purple, ER-Golgi; dark purple, endosome and vacuole sorting; yellow, ER-dependent protein degradation; red, protein folding and glycosylation, cell wall biosynthesis and integrity; fuchsia, tRNA modification; pink, cell polarity and morphogenesis; orange, autophagy; and black, uncharacterized. Individual genetic interactions contributing to genetic profiles revealed by (A) are illustrated for three specific subnetworks in (B) to (D). (**B** to **D**) Subsets of genes belonging to amino acid biosynthesis and uptake, ER-Golgi, and tRNA modification regions of the network were selected, and, in some cases, additional genes were included from the complete network shown in Fig. 1. Nodes are grouped according to profile similarity, and edges represent negative (red) and positive (green) genetic interactions ($|\varepsilon| > 0.08$, $P < 0.05$). Nonessential (circles) and essential (diamonds) genes are colored according to the biological process indicated in (A), and uncharacterized genes are depicted in yellow. (**E**) *PAR32*, *ECM30*, and *UBP15* are required for plasma membrane localization (micrographs) and activity (histogram) of the Gap1 amino acid permease. DIC, differential interference contrast; GFP, green fluorescent protein. (**F**) Sgt2 physically interacts with components of the GET pathway and members of the Hsp70 chaperone family. Proteins identified with high confidence as specific interactors for tandem affinity purification (TAP)–tagged Sgt2 (Sgt2-TAP) are shown in decreasing order of spectral counts. (**G**) Distribution of the Elp and Urm modified codon usage among synthetic sick or lethal interaction partners. The fraction of Elp and Urm modified codons (lysine, glutamine, and glutamic acid) relative to all codons was measured for all negative interactors with genes in the Elp or Urm complex (red) relative to the background usage of all genes (blue).

(Hsp70) protein family members (Fig. 2F), and, similarly to GET pathway mutants (*13*), deletion of *SGT2* results in mislocalization of the tail-anchored protein, Pex15 (fig. S4).

Deciphering complex regulatory relations from the global genetic network. Because the global genetic interaction map represents a broad functional survey, it should provide insights into the regulatory wiring diagram of the cell. For example, synthetic lethal interactions between genes encoding the elongator (Elp) complex and those of the urmylation (Urm) pathway suggested that the Urm pathway collaborates with the Elp complex in the modification of specific transfer RNAs (tRNAs) (*15*) (Fig. 2D). In addition to their synthetic lethal relation, Elp and Urm pathway genes shared highly similar genetic interaction profiles; notably, these interactions were enriched for cell polarity and secretion genes ($P < 10^{-3}$) (Fig. 2D), which reflects a specific cell polarity defect associated with Elp pathway mutants (*16*).

The elongator tRNA modification machinery has been postulated either to broadly affect the translation of a suite of mRNAs whose genes have cell polarity roles or to selectively influence the activity of a key polarity regulatory gene (*17*). We were intrigued by the finding that the subset of Elp-Urm negative interactors, as well as cell polarity and secretion genes, in general, encode proteins that are significantly enriched for the amino acids that charge Elp- and Urm-modified tRNAs (Fig. 2G) (*7*, *8*). These findings suggest that Elp and Urm pathways may be biased toward the regulation of a functionally specific subset of cellular proteins. *ELP1* is a highly conserved gene whose human ortholog, inhibitor of kappa light polypeptide gene enhancer in B cells, kinase complex–associated protein (IKBKAP), is associated with a neurological disorder, familial dysautonomia, which leads to disruption of cytoskeletal organization when mutated (*18*, *19*). Thus, it is possible that disease manifestation may involve impaired IKBKAP-dependent translation of a set of human genes belonging to a specific functional group.

Genetic network connectivity. Consistent with the degree distribution of other biological networks (*1*), the majority of genes have few interactions, whereas a small number are highly connected and serve as network hubs (Fig. 3A). We found subsets of genes that showed a strong bias in their interaction type. About 2% of array genes exhibited more than eight times as many negative interactions as positive ones, whereas a smaller set containing ~1% of all array genes showed four times as many positive as negative interactions (Fig. 3B). Genes displaying this behavior were functionally distinct. Specifically, a bias toward negative interaction was observed for genes required for normal progression of the cell division cycle ($P < 10^{-8}$), which highlights the central role of checkpoints in maintaining viability in dividing cells. Predominantly positive interactions were indicative of genes involved in translation, ribosomal RNA processing, and mRNA decay ($P < 10^{-5}$), which may suggest that defects in the translation machinery somehow mask phenotypes that would otherwise be expressed in normal cells.

Genetic interaction degree, fitness, multifunctionality, and pleiotropy. Genetic interaction hubs show a clear association with several fundamental physiological and evolutionary properties (Fig. 3C), which may be predictive of genetic interactions in other organisms. In particular, we uncovered a strong correlation between genetic interaction degree and single-mutant fitness (r = 0.73). Single mutants with increasingly severe fitness defects tended to exhibit an increased number of both negative and positive interactions (Fig. 3C and fig. S5, A and B) (*7*, *8*). This relation was also observed for essential genes where the average number of interactions involving a temperature-sensitive mutant allele was inversely proportional to allele fitness at a given semipermissive temperature (fig. S5B). The increased connectivity of genes with fitness defects when singly mutated was not due to nonspecific interactions derived from a generally compromised cell or experimental noise; interactions with these genes were found to overlap with known functional relationships just as frequently as other interactions (fig. S5C).

In addition to the correlation with single-mutant fitness defects, genetic interaction hubs showed a high degree of pleiotropy. Specifically, the number of genetic interactions for a particular hub was significantly correlated with the number

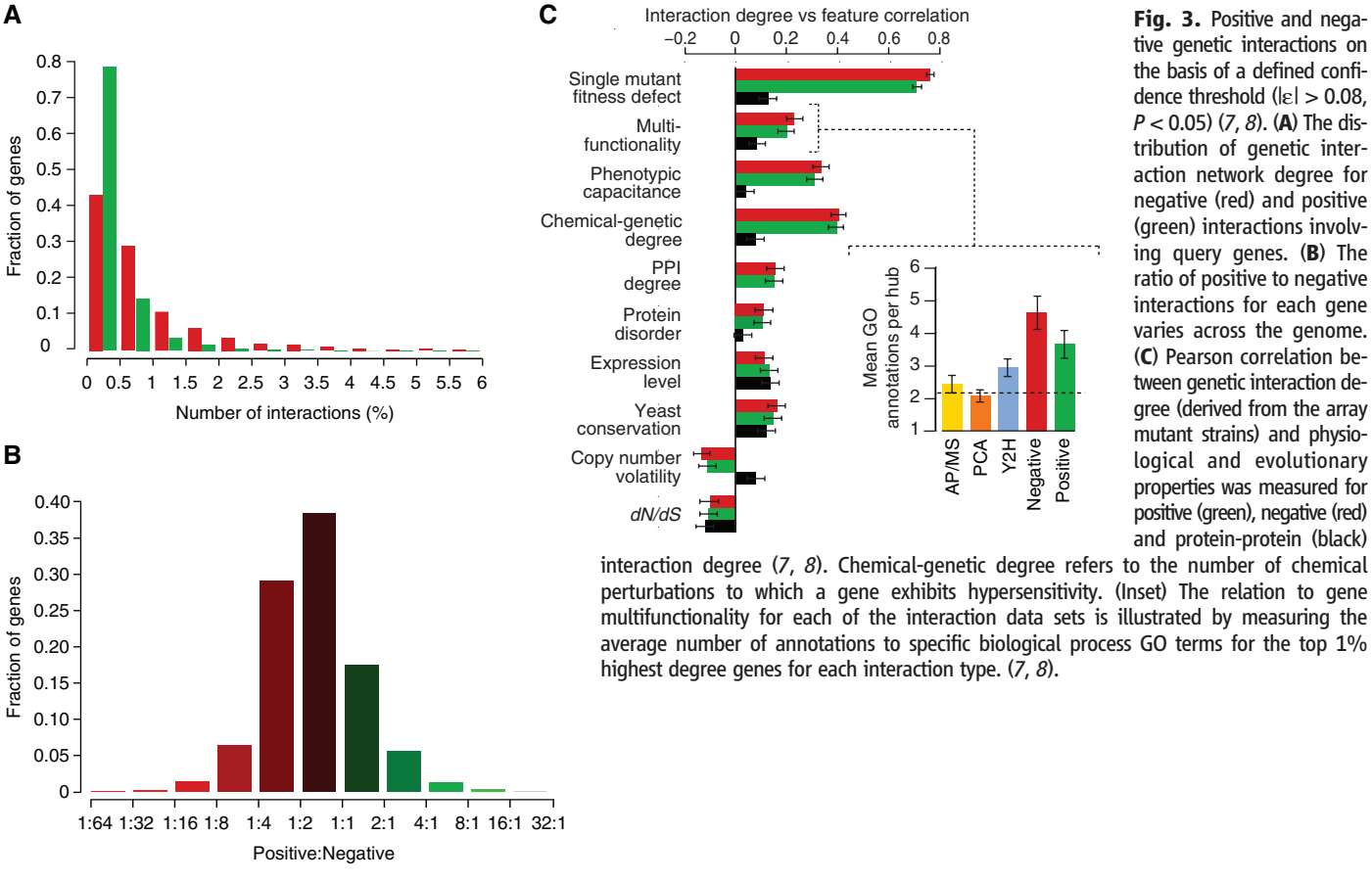


Fig. 3. Positive and negative genetic interactions on the basis of a defined confidence threshold ($|\varepsilon| > 0.08$, $P < 0.05$) (*7*, *8*). (**A**) The distribution of genetic interaction network degree for negative (red) and positive (green) interactions involving query genes. (**B**) The ratio of positive to negative interactions for each gene varies across the genome. (**C**) Pearson correlation between genetic interaction degree (derived from the array mutant strains) and physiological and evolutionary properties was measured for positive (green), negative (red) and protein-protein (black) interaction degree (*7*, *8*). Chemical-genetic degree refers to the number of chemical perturbations to which a gene exhibits hypersensitivity. (Inset) The relation to gene multifunctionality for each of the interaction data sets is illustrated by measuring the average number of annotations to specific biological process GO terms for the top 1% highest degree genes for each interaction type. (*7*, *8*).

of distinct annotated functions (multifunctionality) for that gene (Fig. 3C and inset). This connection between network hubs and pleiotropy was further reflected by the rich variation associated with hub mutant phenotypes and increased phenotypic capacitance, the number of different morphological phenotypes linked to a specific gene as defined quantitatively (Fig. 3C) (*20*). This relation suggests that genetic network hubs play key roles in the integration and execution of morphogenetic programs.

It is noteworthy that these correlations persisted after we controlled for fitness defects of single mutants (fig. S5D). Furthermore, these trends reveal characteristics that distinguish genetic network hubs from hubs on the physical interaction network (Fig. 3C). Notably, the correlation to both fitness and multifunctionality was several fold stronger for genetic interaction degree (Fig. 3C). This likely reflects the ability of genetic perturbation analysis to identify broad phenotypic connections that cannot be captured in networks subject to physical constraints and suggests that large-scale genetic interaction networks will be of wide utility for defining the functional wiring diagrams of cells and organisms.

Although there are several distinguishing characteristics of genetic interaction hubs, we measured a significant correlation ($r \sim 0.2$) between the genetic and physical interaction degree for any given gene (Fig. 3C). Similar to protein-protein interaction hubs (*21*–*23*), we found that genetic network hubs tend to be expressed at higher mRNA levels. In comparison with the whole-genome sequences of 23 different Ascomycota fungi species, we found that genetic interaction degree correlated positively with gene conservation and negatively with copy number volatility, which indicates that they tend to be lost or duplicated less frequently. Genes showing more genetic interactions evolved (*dN/dS*) more slowly than genes

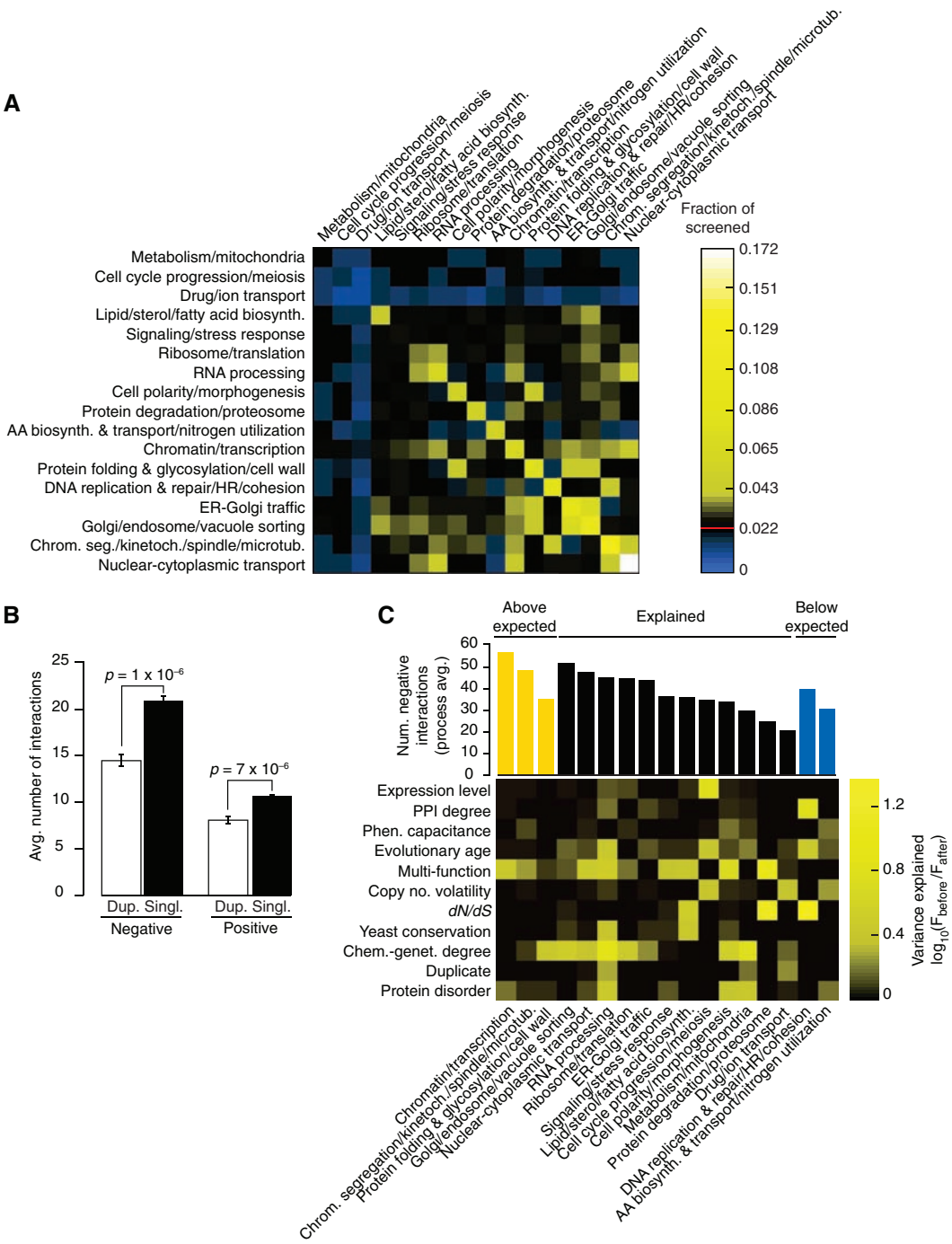


Fig. 4. (**A**) Frequency of synthetic lethal/sick (negative) genetic interactions within and across biological processes. The fraction of screened gene pairs exhibiting negative interactions was measured for 17 broadly defined functional gene sets (*7*, *8*). A color was assigned to each process-process element reflecting the fraction of interaction (blue, below the frequency of random pairs; black, statistically indistinguishable from the random background of interactions; and yellow, above the frequency of random pairs), with the diagonal representing within-process interactions. The red line in the color scale bar indicates random background. (**B**) Genetic interaction frequency of duplicate genes. T bars, SEM. (**C**) Gene-specific factors explaining the variation in number of negative interactions across biological processes. (Top) The average number of interactions across each process with the color indicating processes that have more interactions than expected (yellow, $P < 0.05$); processes whose interaction degree is explained by the factors indicated on the *y* axis; and those with fewer interactions than expected (blue, $P < 0.05$). The influence of each gene-specific factor in explaining the number of interactions observed was measured by plotting the ratio of *F* statistics of the bioprocess factor before and after incorporating the additional gene-specific factor. This ratio is indicated by the corresponding column in the heat map (*7*, *8*). (AA, amino acids; chrom. seg., chromosome segregation; HR, homologous recombination; kinetoch., kinetochore)

with few interactions (Fig. 3C), which suggests that genetic hubs generally tend to be evolutionarily constrained. However, a subset of genetic interaction hubs appears to behave differently. Despite their tendency to evolve faster (fig. S5F) (*24*), proteins with higher levels of native disorder tend to exhibit a large number of genetic interactions, which suggests that genes encoding disordered proteins may represent a distinct class of genetic interaction hub.

Distribution of genetic interactions by bioprocess. We assessed the distribution of genetic interactions across different cellular processes for both negative (Fig. 4A) and positive (fig. S6A) (*7*, *8*) interactions. The heat map identified functions enriched (yellow) or depleted (blue) for genetic interactions relative to the expected frequency of a random gene set. As expected, genes involved in similar biological processes were enriched for negative interactions; however, we also observed genetic interactions bridging bioprocesses (Fig. 4A). Specifically, genes involved in chromatin, transcription, ER-Golgi transport, and Golgi-endosome transport showed a significant number of interactions that bridge diverse functions, which suggests that many of these genes are interconnected or pleiotropic. These bioprocess-level findings concur with individual gene analyses, which indicated that genes involved in processes related to chromatin structure and transcription ($P < 10^{-14}$), as well as secretion and vesicle transport ($P < 10^{-9}$), were among the most highly connected genes in our network. The central role for chromatin- and transcription-related processes identified in the yeast genetic network is consistent with large-scale genetic network mapping in *Caenorhabditis elegans* (*25*), and the bridging function for secretory pathway genes emphasizes their role as communication conduits for the cell. In contrast to genetic interactions, protein-protein interactions connect relatively fewer bioprocesses, and thus, although highly informative of local pathway architecture, physical interactions fail to provide a complete picture of multifunctionality or interconnections between cellular processes (fig. S6A). Reduced interactions in particular gene sets, such as meiosis, drug or ion transport, and metabolism or mitochondrial genes (blue in Fig. 4A), may arise because some processes are more buffered than others and require more complex genetic analysis to uncover their interactions (*5*), whereas others may function only under certain environmental conditions (*26*).

Because variation was observed in the average number of genetic interactions for genes across different bioprocesses, we tested whether gene-specific properties (Fig. 3C) were predictive of this variation. For example, we found that gene duplicates exhibited fewer interactions when surveyed across the entire genome (Fig. 4B) (*7*, *8*), and therefore, we asked if bioprocesses with relatively few genetic interactions could be explained by specific factors, such as a high percentage of duplicated genes. An analysis of covariance (ANCOVA) (Fig. 4C) (*7*, *8*) showed that a linear model including the gene-specific properties predictive of genetic interaction hubs (Fig. 3C) was sufficient to explain the number of negative (12 out of 17) (Fig. 4C) and positive (13 out of 17) (fig. S6B) genetic interactions for the majority of bioprocesses. For example, the relatively few genetic interactions seen for genes with roles in drug and ion transport are explained by a combination of a high rate of gene duplication (~50 to 60%) and copy number volatility among genes annotated to this process. This is consistent with the tendency of genes encoding protein pumps to undergo numerous duplication events (*27*), which confirms that extensive redundancy associated with large gene families complicates the identification of digenic interactions. Three bioprocesses had significantly more negative interactions than predicted (Fig. 4C) ($P < 0.05$),

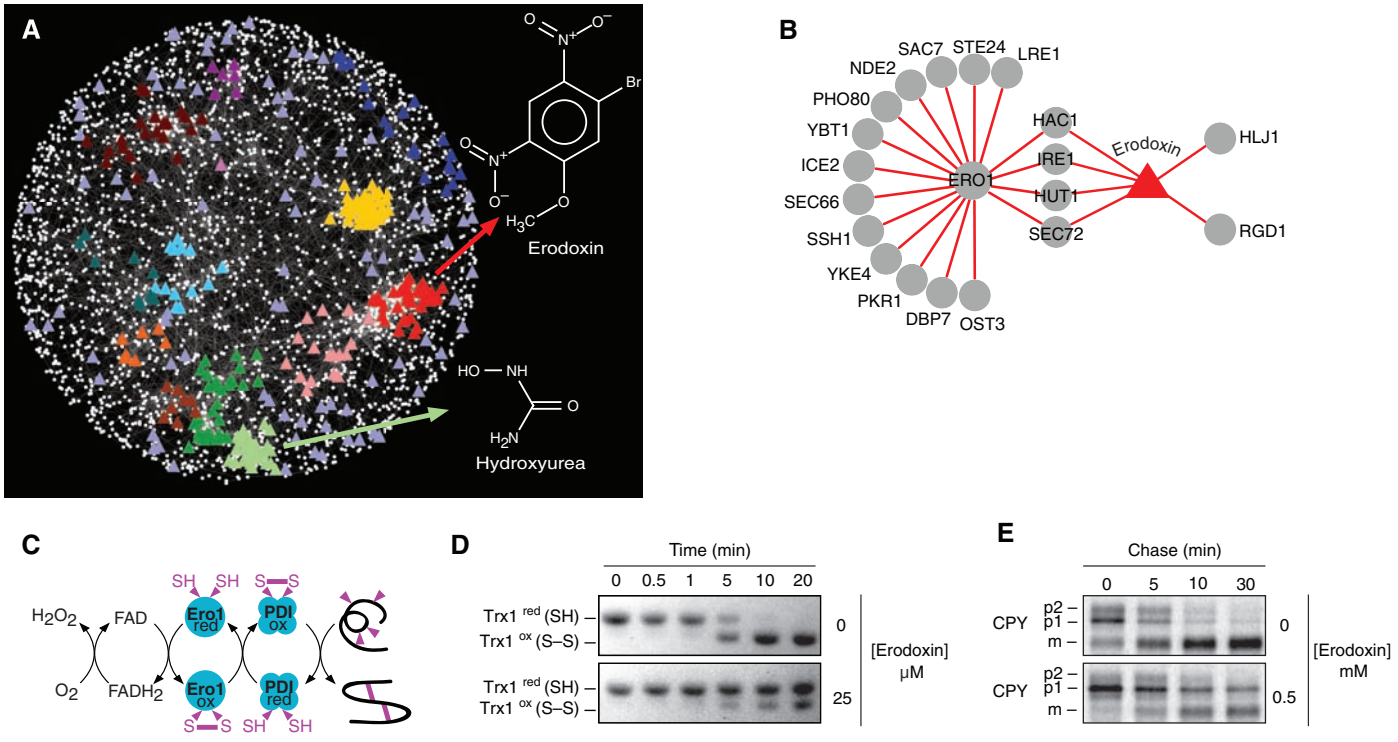


Fig. 5. (**A**) A chemical-genetic interaction map is shown in which colored triangles represent chemical compounds and white nodes correspond to genes. Compounds were positioned on the map by highlighting the gene node whose genetic interaction profile most closely resembles the chemical genetic profile of the compound derived from three sources (*7*, *8*). Compounds tightly correlated to genes positioned within functional clusters (Fig. 1) were colored accordingly to the color of the cluster as in Fig. 1. The chemical-genetic profile of hydroxyurea clustered with genes involved in DNA replication and repair, whereas that of erodoxin clustered with genes involved in protein folding, glycosylation, and cell wall biosynthesis. Compounds positioned outside functional clusters are colored light purple. (**B**) Network displaying overlap between *ERO1* negative genetic interactions and genes resulting in growth inhibition when deleted in the presence of erodoxin. (**C**) *ERO1*-dependent pathway for oxidative protein-folding pathway. (**D**) Erodoxin inhibits Ero1-dependent oxidation of Trx1 in vitro. (**E**) Erodoxin inhibits CPY processing to the vacuolar form in vivo. ER (p1), Golgi (p2), and vacuolar (m) forms of CPY are indicated.

including those that show functional enrichment for genetic interaction hubs (Fig. 4A). Conversely, DNA replication and repair and amino acid biosynthesis showed significantly fewer negative interactions than predicted ($P < 0.05$), which suggested that either more genetic interactions remain to be found for these genes under different environmental conditions or that these genes are more buffered and thus are inherently less connected on the digenic network.

Overlap between the genetic and the protein-protein interaction networks. We observed genetic interactions overlapping with 10 to 20% of protein-protein interaction pairs, depending on the physical interaction mapping methodology (fig. S7), which is significantly higher than expected randomly (~3%). Considering the global yeast physical interaction network as defined by affinity purification–mass spectrometry (*28*, *29*), yeast two-hybrid protocol (*30*), or protein-fragment complementation assay (PCA) (*31*), roughly an equivalent number of physical interactions overlapped with negative and positive genetic interaction pairs: ~7% of protein-protein interacting pairs shared a negative genetic interaction, whereas ~5% shared a positive interaction. Conversely, considering our genetic interaction network, only a small fraction of gene pairs that show a genetic interaction (0.4% negative and 0.5% positive) are also physically linked. These findings suggest that the vast majority of both positive and negative interactions occurs between, rather than within, complexes and pathways, connecting those that presumably work together or buffer one another, respectively.

Navigating from genetic to chemical-genetic interaction networks. The set of ~4700 viable yeast deletion mutants has been exposed to hundreds of different chemical compounds (*26*). We quantified the chemical-genetic degree for each gene by counting the number of chemical (environmental) perturbations for which the corresponding gene deletion mutant showed hypersensitivity. We found a significant correlation ($r = 0.4$, $P < 10^{-5}$) between genetic interaction and chemical-genetic degree (Fig. 3C). These observations suggest that hubs on a chemical-genetic network are predictive of hubs on the genetic interaction network and can be used to link environmental capacitance and genetic robustness. Furthermore, our data suggest that the same genes buffer the cell against both environmental and genetic insults. It is not known whether natural selection favors genetic robustness (*32*), but the positive correlation between genetic interaction degree and environmental capacitance suggests that genetic and environmental robustness may coevolve (*33*).

Because chemical perturbations mimic genetic perturbations, the genetic network should be useful for predicting the cellular targets of bioactive molecules (*34*). We identified genetic interaction profiles that are significantly correlated to a chemical-genetic profile of a particular compound (*7*, *8*, *26*, *34*) and showed that compounds often clustered to dense regions of the genetic network indicative of specific bioprocesses (Fig. 5A). For example, hydroxyurea, a compound that inhibits ribonucleotide reductase and blocks DNA synthesis, clusters with the gene cohort annotated with roles in DNA replication and repair (Fig. 5A). These results demonstrate that clustering of chemical-genetic and genetic interaction profiles complements haploinsufficiency profiling, which has the potential to identify drug targets directly (*26*). We used this network approach to examine the previously uncharacterized compound, 0428-0027, which we have subsequently named erodoxin (Fig. 5A). Erodoxin clustered with genes associated with protein folding, glycosylation, and cell wall biosynthesis functions (Fig. 5A) because the erodoxin chemical-genetic profile most closely resembled the genetic interaction profile of *ERO1* (Fig. 5B and fig. S8A), an essential gene involved in oxidative protein folding (Fig. 5C) (*35*). Two additional lines of evidence suggested that Ero1 is the target of erodoxin. First, *ero1*Δ/+ and *fad1*Δ/+ heterozygotes were the most hypersensitive mutants identified from haploinsufficiency profiling (fig. S8B) (*7*, *8*). Second, we found that erodoxin leads to inhibition of Trx1 oxidation (Fig. 5D) and delayed carboxy peptidase Y (CPY) processing (Fig. 5E), which suggests that it inhibits Ero1 activity both in vitro and in vivo.

Exploring the universe of genetic interactions. Unbiased, systematic, and quantitative analysis of digenic loss-of-function perturbations assigns a rich phenotypic profile to each gene and enables construction of a functional map of the cell, organizing genes and higher-order bioprocesses according to their related roles (Fig. 1). The functional connections defined by genetic interactions complement the information derived from networks based upon physical interactions, which links previously uncharacterized genes to specific pathways and complexes and reveals connections between pathways and complexes. The global mapping of genetic networks is becoming feasible in more complex cells and metazoans because of the growing availability of whole-genome sequences and large-scale sets of gene-knockdown reagents (*1*). Although negative genetic interactions can be conserved from yeast to worms and from yeast to human cells, the extent to which individual genetic interactions are conserved over large evolutionary distances remains unclear (*1*). The conservation of the genetic map may also occur at various levels of resolution. For example, overall network topology (Fig. 1) and properties (Fig. 3C) may be more highly conserved than particular genetic interactions because they reflect the fundamental architecture of the cell. The ability to integrate genetic and chemical-genetic perturbation data offers the potential to link bioactive compounds to their targets (Fig. 5), to identify genetic interaction hubs through chemical perturbations (Fig. 3C), to design synthetic lethal therapies for targeting genetically defined tumors (*36*), and to understand the mechanistic basis of drug synergy (*37*). Finally, genetic interaction maps provide a model for understanding the link between genotype and phenotype and for outlining the general principles of complex genetic interaction networks, which play a key role in governing inherited phenotypes, including human disease (*3*).

References and Notes

1. S. J. Dixon, M. Costanzo, A. Baryshnikova, B. Andrews, C. Boone, *Annu. Rev. Genet.* **43**, 601 (2009).
2. C. H. Waddington, *The Strategy of the Gene* (Allen & Unwin, London, 1957).
3. L. Hartwell, *Science* **303**, 774 (2004).
4. A. H. Tong *et al.*, *Science* **294**, 2364 (2001).
5. A. H. Tong *et al.*, *Science* **303**, 808 (2004).
6. R. Mani, R. P. St. Onge, J. L. Hartman 4th, G. Giaever, F. P. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3461 (2008).
7. Materials and methods are available as supporting material on *Science* Online.
8. See supplementary information at http://drygin.ccbr.utoronto.ca/~costanzo2009.
9. R. P. St Onge *et al.*, *Nat. Genet.* **39**, 199 (2007).
10. D. Segrè, A. Deluna, G. M. Church, R. Kishony, *Nat. Genet.* **37**, 77 (2005).
11. A. Huber *et al.*, *Genes Dev.* **23**, 1929 (2009).
12. E. J. Chen, C. A. Kaiser, *J. Cell Biol.* **161**, 333 (2003).
13. M. C. Jonikas *et al.*, *Science* **323**, 1693 (2009).
14. M. B. Metzger, S. Michaelis, *Mol. Biol. Cell* **20**, 1006 (2009).
15. S. Leidel *et al.*, *Nature* **458**, 228 (2009).
16. P. B. Rahl, C. Z. Chen, R. N. Collins, *Mol. Cell* **17**, 841 (2005).
17. A. Esberg, B. Huang, M. J. Johansson, A. S. Byström, *Mol. Cell* **24**, 139 (2006).
18. T. Naumanen, L. D. Johansen, E. T. Coffey, T. Kallunki, *Cell Adh. Migr.* **2**, 236 (2008).
19. L. D. Johansen *et al.*, *J. Cell Sci.* **121**, 854 (2008).
20. S. F. Levy, M. L. Siegal, A. Levchenko, *PLoS Biol.* **6**, e264 (2008).
21. P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, *Science* **314**, 1938 (2006).
22. H. B. Fraser, D. P. Wall, A. E. Hirsh, *BMC Evol. Biol.* **3**, 11 (2003).
23. C. Pál, B. Papp, L. D. Hurst, *Genetics* **158**, 927 (2001).
24. P. M. Kim, A. Sboner, Y. Xia, M. Gerstein, *Mol. Syst. Biol.* **4**, 179 (2008).
25. B. Lehner, C. Crombie, J. Tischler, A. Fortunato, A. G. Fraser, *Nat. Genet.* **38**, 896 (2006).
26. M. E. Hillenmeyer *et al.*, *Science* **320**, 362 (2008).
27. M. J. Dunham *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16144 (2002).
28. A. C. Gavin *et al.*, *Nature* **440**, 631 (2006).
29. N. J. Krogan *et al.*, *Nature* **440**, 637 (2006).
30. H. Yu *et al.*, *Science* **322**, 104 (2008).
31. K. Tarassov *et al.*, *Science* **320**, 1465 (2008).
32. J. A. de Visser *et al.*, *Evolution* **57**, 1959 (2003).
33. C. D. Meiklejohn, D. L. Hartl, *Trends Ecol. Evol.* **17**, 468 (2002).
34. A. B. Parsons *et al.*, *Nat. Biotechnol.* **22**, 62 (2004).
35. C. S. Sevier *et al.*, *Cell* **129**, 333 (2007).
36. P. C. Fong *et al.*, *N. Engl. J. Med.* **361**, 123 (2009).
37. J. Lehár, B. R. Stockwell, G. Giaever, C. Nislow, *Nat. Chem. Biol.* **4**, 674 (2008).

38. We thank S. Dixon, T. Hughes, P. Jorgensen, and M. Tyers for critical comments. Supported by Genome Canada through the Ontario Genomics Institute (2004-OGI-3-01) and the Canadian Institutes of Health Research (GSP-41567) (C.B., B.A.), the University of Minnesota Biomedical Informatics and Computational Biology program (J.B., R.D.), and a seed grant from the Minnesota Supercomputing Institute (J.B., B.V.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/327/5964/425/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
References

20 August 2009; accepted 12 November 2009
10.1126/science.1180823

Supporting Online Material for

The Genetic Landscape of a Cell

Michael Costanzo, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D. Spear, Carolyn S. Sevier, Huiming Ding, Judice L.Y. Koh, Kiana Toufighi, Sara Mostafavi, Jeany Prinz, Robert P. St.Onge, Benjamin VanderSluis, Taras Makhnevych, Franco J. Vizeacoumar, Solmaz Alizadeh, Sondra Bahr, Renee L. Brost, Yiqun Chen, Murat Cokol, Raamesh Deshpande, Zhijian Li, Zhen-Yuan Lin, Wendy Liang, Michaela Marback, Jadine Paw, Bryan-Joseph San Luis, Ermira Shuteriqi, Amy Hin Yan Tong, Nydia van Dyk, Iain M. Wallace, Joseph A. Whitney, Matthew T. Weirauch, Guoqing Zhong, Hongwei Zhu, Walid A. Houry, Michael Brudno, Sasan Ragibizadeh, Balázs Papp, Csaba Pál, Frederick P. Roth, Guri Giaever, Corey Nislow, Olga G. Troyanskaya, Howard Bussey, Gary D. Bader, Anne-Claude Gingras, Quaid D. Morris, Philip M. Kim, Chris A. Kaiser, Chad L. Myers,* Brenda J. Andrews,* Charles Boone*

*To whom correspondence should be addressed. E-mail: cmyers@cs.umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); charlie.boone@utoronto.ca (C.B.)

Published 22 January 2010, *Science* **327**, 425 (2010)
DOI: 10.1126/science.1180823

This PDF file includes

Other Supporting Online Material for this manuscript includes the following:
(available at http://drygin.ccbr.utoronto.ca/~costanzo2009)

Table of Contents

SGA genetic interaction dataset

All data files can be downloaded from:
http://drygin.ccbr.utoronto.ca/~costanzo2009/

General description of the Synthetic Genetic Array (SGA) interaction score

To derive quantitative genetic interactions, we modeled colony size as a multiplicative combination of the double mutant fitness, time, and experimental factors. Specifically, for a double mutant carrying mutations of genes *i* and *j*, colony size C_{ij} can be expressed as $C_{ij} = f_{ij} \cdot t \cdot s_{ij} \cdot e$, where f_{ij} is the double mutant fitness, t is time, s_{ij} is the combination of all systematic factors, and e is log-normally distributed random noise. The double mutant fitness f_{ij} can be further expressed as $f_{ij} = f_i f_j + \varepsilon_{ij}$, where f_i and f_j are the fitnesses of the two single mutants and ε_{ij} is a quantitative measure of the genetic interaction between them. Note that this reflects the typical assumption of a multiplicative "null model" (*S1*) and ε_{ij} measures deviations from that null model.

Accurate measurements of quantitative genetic interactions are dependent on the correct estimation of single and double mutant fitness. We found that observed yeast mutant colony sizes can be severely affected by systematic experimental biases that alter a colony's growth independently of its fitness. We identified and corrected five major systematic effects that contribute the majority of the observed variance in colony size. These systematic effects included positional effects, spatial effects, nutrient competition effects, and screen batch effects.

To derive accurate estimates of single mutant fitness, we applied our correction method to a set of control SGA screens, where the queries carried a mutation in a neutral genomic locus. The obtained single mutant fitnesses (f_i and f_j) were combined with the double mutant fitnesses (f_{ij}) estimated from regular SGA experiments to derive genetic interaction measures as $\varepsilon = f_{ij} - f_i \cdot f_j$. A statistical confidence measure (p-value) was assigned to each interaction based on a combination of the observed variation of each double mutant across four experimental replicates and estimates of the background log-normal error distributions for the corresponding query and array mutants. Estimates of the error distribution for each array and query mutant were estimated separately from the set of all other double mutants carrying the corresponding array or query mutation, respectively.

General information about the dataset

The SGA genetic interaction dataset is composed of 1712 queries crossed to 3885 array strains. Of 1712 queries, 1378 are deletion mutants of non-essential genes and 334 are essential gene alleles (214 temperature-sensitive and 120 DAmP alleles). TS and DAmP queries are indicated by "_tsq" and "_damp" suffixes, respectively. For a subset of essential genes, several different TS alleles were tested, as indicated by a unique number following the "_tsq" suffix. The set of array strains originally contained 4293 non-essential deletion mutants, but 408 mutants were removed for quality control reasons or due to incompatibility with SGA technology. Approximately 645,000 individual gene pairs were filtered from the 1712 x 3885 genetic interaction matrix for the same reasons.

The resulting dataset contains raw genetic interaction scores for ~6 million double mutants.
The SGA genetic interaction dataset is available for download in various text formats (see Data Files S1-S5) or through a web-based database, DRYGIN (http://drygin.ccbr.utoronto.ca, (*S2*)). SGA data can also be explored through the SGAExpress browser (http://www.ailab.si/sgaexpress).

Materials and Methods

SGA query strain construction and screening

SGA query strain construction and screening were conducted as described in (*S3*). 1712 query mutants were screened against an array of 3885 non-essential deletion mutants. Some deletion array mutants were removed for quality control reasons or due to incompatibility with SGA technology. All screens were conducted a single time with 4 replicate colonies per double mutant, except for 211 query mutants that were screened twice (8 replicate colonies).

Biological process annotations

Biological process categories, described in Fig. 4, were derived functional enrichment of network clusters identified in Fig. 1 and further refined by examining cluster enrichment of network sub-regions (Fig. 2A, Data File S8). Genes were assigned to one or more of these functional categories on the basis of examination of their Gene Ontology annotations (www.geneontology.org) and gene function descriptions available from the *Saccharomyces* Genome Database (SGD, www.yeastgenome.org). Specific annotations are provided in Data File S6. A subset of uncharacterized or multi-functional genes that were difficult to assign to defined functional groups were not assigned to any functional category.

False positive and false negative rate estimation

Sensitivity is defined as the fraction of known interactions identified by an assay.

$$sensitivity = \frac{TP}{TP + FN}$$

We estimated SGA sensitivity for detecting true genetic interactions by evaluating previously published interactions from Biogrid (www.thebiogrid.org, downloaded on May 4th, 2009). For negative genetic interactions, we considered Biogrid interactions annotated as phenotypic enhancement, synthetic growth defect and synthetic lethality. For positive genetic interactions, we considered Biogrid interactions annotated as phenotypic suppression and synthetic rescue. By systematically varying SGA genetic interaction score (ε) and p-value cutoffs, we calculated the number of true positives (*TP*) as the number of Biogrid interactions with SGA genetic interaction scores and p-values more extreme than the chosen cutoff (Fig. S2A, sensitivity).

Precision is the fraction of true interactions in the set of all identified hits.

$$precision = \frac{TP}{TP + FP}$$

We estimated precision of our positive and negative genetic interactions by comparing the frequency of overlap between replicate AB-BA screened pairs using the following strategy. Given a square matrix of SGA genetic interaction scores where every gene *A* is present both as query (*AB*) and as array (*BA*), the total number of interactions at a certain cutoff is the sum of interactions called by *AB* pairs only, *BA* pairs only and those called by both. The interactions called by both *AB* and *BA* ($N_{overlap}$) are expected to be a mixture of true positives and false positives. The true positive portion of $N_{overlap}$ depends on the

total number of interactions called by *AB* (N_{AB}), the probability of them being true positives (precision *p*) and the probability of them not being missed by *BA* (sensitivity = 1 – *FNR*). The false positive portion of $N_{overlap}$ depends on the total number of interactions called by *AB* (N_{AB}), the probability of them being false ($1 - p$) and the probability of them being incorrectly called positive by *BA* (specificity = 1 – *TNR*).
Overall, the number of interactions called by both AB and BA is given by:

$$N_{overlap} = N_{overlap\_true} + N_{overlap\_false} = p \cdot N_{AB} \cdot (1 - FNR) + (1 - p) \cdot N_{AB} \cdot (1 - TNR)$$

Given that $N_{overlap}$, N_{AB} and *FNR* are known, and 1 – *TNR* can be at most equal to the total density of interactions in the matrix (adopting the conservative assumption that all the interactions reported are false), precision can be calculated as:

$$precision\ (p) = \frac{N_{overlap} - N_{AB} \cdot (1 - TNR)}{N_{AB} \cdot (1 - FNR) - N_{AB} \cdot (1 - TNR)}$$

Sensitivity and precision of SGA genetic interaction scores at a range of cutoffs are reported in Table S6.

Constructing the functional map of the cell

To generate the network shown in Fig. 1, genetic interaction profile similarities were measured for all query and array gene pairs by computing Pearson correlation coefficients (PCC) from the complete genetic interaction matrix. Correlation coefficients of gene pairs screened both as queries and as arrays were averaged. Gene pairs whose profile similarity exceeded a PCC > 0.2 threshold were connected in the network, and an edge-weighted spring-embedded layout, implemented in Cytoscape (*S4*), was applied to determine node position. Genes sharing similar patterns of genetic interactions located proximal to each other in two-dimensional space, while less-similar genes were positioned further apart.

Functional enrichment of genetic interactions

Enrichment for protein-protein interactions (E_{ppi}, and analogously E_{GO} for Gene Ontology co-annotated gene pairs) shown in Fig. S2B was calculated as follows:

$$E_{ppi} = \frac{N_{cutoff,ppi} / N_{cutoff}}{N_{ppi} / N}$$

where:

N = total number of pairs tested
N_{ppi} = total number of protein-protein interactions among the pairs tested
N_{cutoff} = number of pairs above the cutoff
$N_{cutoff,ppi}$ = number of protein-protein interactions among the pairs above the cutoff

The protein-protein interaction standard was downloaded from Biogrid (www.thebiogrid.org) on May 4th, 2009. The functional co-annotation standard was derived from the biological process aspect of the Gene Ontology, as described in (*S5*).

Recall of published genetic interactions

The analysis presented in Fig. S1D was performed as follows. The published set of genetic interactions was downloaded from Biogrid (www.thebiogrid.org, on May 4th, 2009), and restricted to gene pairs assayed in our study. We determined the fraction of

interactions identified in this study ($p < 0.05$, systematic ε thresholds) that overlap with the Biogrid set and compared it to random expectation. Both negative and positive SGA genetic interactions (ε) were significantly enriched for known genetic interactions. Extreme negative ($\varepsilon < -0.8$) and positive ($\varepsilon > 0.25$) interactions were approximately 140- and 80-fold enriched for previously identified interactions, respectively.

Comparison of SGA genetic interaction scores with high-resolution liquid growth profiles

A recent study analyzed 26 genes, conferring resistance to the DNA damaging agent methyl methanesulfonate (MMS), by accurately measuring growth rates of all 26 single and 325 double deletion mutants in liquid medium (*S6*). Of the 325 gene pairs tested in (*S6*), 239 were covered by our SGA dataset. We compared quantitative genetic interaction scores derived from liquid growth profiles to those derived from colony size measurements for corresponding pairs, and observed strong correlation ($r = 0.89$) on both strong and modest interactions.

Comparative analysis of function prediction from genetic interaction profiles

Datasets

The datasets used in the function prediction analysis are provided in Table S3. All genetic and protein interaction datasets except two (Wilmes 2008 (*S7*) and Fiedler 2009 (*S8*)) were downloaded from the BIOGRID database (www.thebiogrid.org, version 2.0.45). Wilmes 2008 and Fiedler 2009 were directly downloaded from the corresponding online Supplements. Protein localization and gene expression datasets were downloaded from the corresponding online Supplements.

The input to the GeneMANIA algorithm is one or more functional association networks (for details, see (*S9*)). As such, for profiled-based datasets (gene expression and localization datasets), we constructed functional association networks using Pearson correlation coefficient (PCC). For protein and genetic interaction datasets, we used both an undirected binary interaction network and correlation-derived networks using PCC. The correlations were calculated on the quantitative scores when possible (Wilmes 2008, Fiedler 2009, and this study). All correlation-based networks were sparsified by setting to zero any interaction that is not among the top 100 highest interactions for either gene.

Comparison of genomic datasets in gene function prediction

We compared the performance of high-throughput datasets derived from nine publications with that in the current study, in predicting gene function using the GeneMANIA algorithm (Fig. S2D, i-ii). To do so, we used all functional association networks derived from each publication (see above) in turn to predict 1,885 GO biological process categories with 3-300 annotations (www.geneontology.org, downloaded in July 2009). For predicting each GO category, GeneMANIA combines the input networks in a context-specific manner. The performance in predicting each category was recorded as the average precision, across all recalls, and precision at 10% and 20% recalls. To compare across general genomic technologies (Figure S2A, i), we

grouped datasets by technology type (AP/MS, gene expression, Y2H, protein localization) (Table S3) and counted the number of distinct GO terms among the 1885 that were predicted at 70% precision or higher at 10% recall.

Leave-One-Out Analysis

We used the GeneMANIA algorithm to assess the amount of unique information that each genetic interaction dataset contributes to the existing knowledge about functional relationships between genes (Fig. S2D, iii, right). In particular, we recorded the average precision in predicting each of 1,885 gene sets, each corresponding to a GO category, in cross-validation, using eight genetic interaction datasets. We then removed each dataset in turn and recorded the resulting difference in precision. In this way, we quantified the average loss in precision in re-producing the known functional relationships in the absence of each genetic interaction dataset.

Identifying sub-networks from the global map

The correlation-based network (Fig. 1) was clustered using Markov Clustering Algorithm (MCL; inflation = 1.4) (*S10*), as well as GINECA, a newly developed approach for identifying partially overlapping clusters (see Materials & Methods, GINECA clustering algorithm). Clusters identified in both methods were aggregated into eight sub-networks on the basis of their relatedness, measured by the degree of cluster overlap for GINECA and by average network proximity after 25 network layout iterations for MCL. Eighty functionally unrelated genes were removed from the network because their similarities may be due to experimental artifacts, such as residual systematic effects that are enhanced in the absence of strong genetic interactions.

GINECA clustering algorithm

GINECA (Genetic Interaction Network Enumerative Clustering Algorithm) is a scalable, semi-supervised graph clustering algorithm for identifying overlapping clusters of functionally-related genes from dense, weighted molecular interaction networks. The core of GINECA is a set of matrix operations, simulating density-based searches of local maxima in a 2-dimensional data space, applied recursively to extract sub-matrices from a large dense matrix representation of weighted interaction scores between the genes. This iterative approach enables some genes to have overlapping memberships in more than one functional cluster, since a gene can be involved in multiple biological processes. Through a constraint-based semi-supervised clustering approach, GINECA exploits known functional assignments of the genes, such as the Gene Ontology (GO) annotations, to guide the clustering algorithm towards more meaningful data partitioning, thus generating clusters of highly connected genes that are also functionally enriched. One or more clusters meeting statistical functional significance criteria are pruned from the network at each iteration, with the algorithm converging when all functionally enriched clusters are extracted, or when the remaining network is loosely connected.

Gap1 permease localization and activity

Strains deleted for *PAR32, ECM30* and *UBP15* and expressing *GAP1-GFP* were grown in SD ammonium medium and Gap1 localization was visualized as described previously

(*S11*). Wild type (WT) cells and *gtr1*Δ mutant cells were included as positive and negative controls, respectively. Gap1 permease activity for the indicated strains (histogram) was measured by 14C-citrulline uptake assays as described previously (*S12*). Data shown are the average of four independent experiments.

Pex15 localization

A pAG-EGFP-*PEX15* plasmid was constructed by sub-cloning the *PEX15* ORF from the Yeast FLEXGene collection (*S13*) into the Gateway(R) donor vector pDONR201 and subsequent cloning into an N-terminal GFP fusion expression vector (pAG416GAL-GFP) as described elsewhere (*S14*). The EGFP-*PEX15* plasmid was transformed into WT, *sgt2*Δ, *get4*Δ, *pex19*Δ and *sgt2*Δ *pex19*Δ strains. Cell cultures were grown at 30°C in SD-HIS media to early log phase, harvested by centrifugation. A 1.5 µl suspension was spotted onto a glass slide and images were captured using Quorum WaveFX Spinning Disc Confocal System.

Elp/Urm amino acid usage analysis

The three codons that pair with the URM/ELP modified tRNAs are AAA, GAA and CAA. For every ORF in the yeast genome, we calculated the relative frequency of the three codons by dividing their occurrence by the total number of codons for any amino acid in the sequence. The relative frequencies of these codons in URM- and ELP-negative interacting genes were compared to a background distribution composed of all the genes tested in our study, excluding the URM/ELP interactors. Significance was evaluated using t-test and Wilcoxon rank sum statistics.

Identification of Sgt2 and Ubp15 physical interactors

Tandem Affinity Purification

TAP-tagged strains (Open Biosystems, Huntsville, AL) were grown to OD_{600} ~0.7 in Synthetic Complete media lacking histidine (1 L, Sigma-Aldrich, St. Louis, MO). Cells were collected by centrifugation and lysed (vortexing with glass-beads), and the lysate cleared by centrifugation, followed by filtering (0.45 µm filters). Purification and identification of TAP-tagged proteins and their interacting partners was performed as described in (*S15*), except that human IgG cross-linked to tosyl-activated magnetic beads (Dynal) were employed instead of IgG-sepharose. Following elution with 25 mM EGTA in 50 mM ammonium bicarbonate (pH 8), 500 ng of sequencing-grade trypsin (Sigma Trypsin Singles, Sigma-Aldrich, St. Louis, MO) was added, and the samples were incubated at 37°C overnight. Digested samples were analyzed by nano-scale liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS), as described in (*S16*). Briefly, 4 cm of 3.5 µm Zorbax C18 (Agilent, Santa Clara, CA) in a fritted 150 µm ID column was used as a pre-column, in-line with a 10 cm 3.5 µm Zorbax C18 75 µm ID analytical column. A 100 min gradient of 0–80% acetonitrile (with 0.1% formic acid) was delivered in a split-flow manner by an Agilent 1100 Capillary HPLC. Data were acquired on a ThermoFisher LTQ linear ion trap (ThermoFisher, Pittsburg, PA), using a data-dependent method consisting of 1 MS scan followed by 4 MS/MS scans. After peptide ions were selected for MS/MS, they were placed on a dynamic exclusion list for

30 sec. Four wash gradients followed each sample analysis gradient, with sample loop injections of isopropanol, 2% TFA, acetonitrile and 0.1% formic acid, respectively. Wash gradients were 45 min, and consisted of 3 oscillations from 30% to 60% acetonitrile. Following the wash steps, 50 fmol of digested BSA protein was injected and analyzed with a 30 min gradient prior to the next sample analysis to monitor the presence of carry-over peptides.

Protein identification and initial data analysis

LTQ data files were converted from binary *.raw files to Mascot generic files (*.mgf) and searched using Mascot 2.2.1 database search tool (*S17*) against the *S. cerevisiae* complement of RefSeq release 21 containing both forward and reverse protein entries (total 11,668 entries). Variable modifications were deamidation of asparagine and glutamine, oxidation of methionine and 1 missed trypsin cleavage. Mass accuracy requirements for searches were set to 3 Da on the parent, and 0.6 Da on product ions. Peptide scores above 35 corresponded to a protein false discovery rate of < 4%, determined using a target-decoy strategy (*S18*).

Data filtering and interpretation

A list of likely contaminants in TAP samples was generated by compiling results from 62 unrelated TAP purification experiments. Proteins detected in 10% or more of these 62 samples were placed on a contaminant list listed in Table S1. All other proteins identified with 2 or greater unique peptides were considered for subsequent analysis, and are listed in Table S2. To further validate the specificity of the detected interactions, we compared spectral counts for each protein detected in the Sgt2-TAP and Ubp15-TAP experiments with spectral counts detected across 19 unrelated experiments. In particular, we analyzed the occurrence of detection of given hits across these 19 unrelated samples, but also the maximal spectral counts with which this hit was detected across these samples. This allowed us to calculate an enrichment ratio for each of the observed hits in the Sgt2-TAP or Ubp15-TAP samples. For example, Sse1 (detected with 11 peptides in Sgt2-TAP) was detected in two out of 19 unrelated samples, with a maximal spectral count of 2 in these samples, indicating high enrichment in the Sgt2-TAP sample. Only those peptides detected with a number of spectra ≥ 5 and 5-fold "enrichment" are reported in the main portion of the text.

Ero1 in vitro activity and CPY processing assays

Ero1 *in vitro* activity was monitored in the absence or presence of 25 μM Erodoxin by examining oxidation of reduced Trx1. Proteins were resolved by non-reducing SDS-PAGE and Trx1 was visualized as described previously (*S19*). *In vivo* activity of Ero1 was assessed by carboxypeptidase Y processing. 0.5 mM Erodoxin was added to wild-type yeast cells in minimal medium, cells were incubated for 10 min at 30°C, pulse labeled for 7 min, and chased for 0-30 min. Carboxypeptidase Y (CPY) was immunoprecipitated and resolved by SDS-PAGE as described in (*S20*)

Identification of genetic interaction hubs and monochromatic genes

We measured the number of positive and negative interactions for all 3885 non-essential array deletion mutants at the intermediate cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). Genetic interaction hubs were selected as the top 10% highest connected genes. Genes with a bias towards positive interactions were selected by finding genes with at least 30 total interactions and positive to negative ratio greater than 1, which is twice the background ratio. Genes with a bias towards negative interactions were selected by finding genes with at least 30 total interactions and positive to negative ratio lower than 0.25, which is one-half of the background ratio. Both of these sets consisted of approximately 130 genes.

Functional enrichment analysis of genetic interaction hubs and monochromatic genes

Genes were evaluated for functional enrichment using the standard approach described in (*S21*) based on the hyper-geometric distribution with the Westfall and Young step-down procedure for multiple hypothesis correction (*S22*). Enrichment analysis was based on annotations downloaded from the *Saccharomyces* Genome Database (www.yeastgenome.org) on July 9th, 2009.

Correlation analysis of genetic and physical interaction degree

To characterize the properties of genetic interaction hubs, we assessed the relationship between genetic interaction degree and several other gene/protein-level features. We measured the number of positive and negative interactions for all of the 3885 non-essential array deletion mutants at an intermediate confidence cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). For each of the quantitative features described below, we measured the Pearson correlation coefficient (PCC) between both positive and negative degree across the 3885 array genes (Fig. 3C, main text). We also computed correlation of all features with protein-protein interaction degree using the union of AP/MS (*S23, 24*), two-hybrid (*S25*), and PCA-based (*S26*) datasets.

Because single mutant fitness is a major correlate of each of the features described below, we repeated the correlation analysis, controlling for the effect of single mutant fitness, by measuring partial correlation (*S27*). Correlation between genetic interaction degree and each of the features remains significant, as does the distinction between genetic protein-protein interaction hubs (Fig. S5D).

Description of gene/protein-level features analyzed

Single mutant fitness defect

Single mutant fitnesses for all non-essential deletion mutants as well as temperature-sensitive or hypomorphic alleles screened in this study was derived from mutant colony size data as described in Materials & Methods, SGA genetic interaction dataset. For correlation analysis, the fitness defect, $1 - f_i$ for a single mutant fitness f_i, was used.

Multi-functionality

A quantitative standard for gene multi-functionality was derived from annotations to "biological process" terms of the Gene Ontology. Specifically, the total number of annotations across a set of functionally distinct GO terms described in (*S5*) was used as a multi-functionality index.

Phenotypic capacitance

The phenotypic capacitance was used directly from the Levy & Siegal study (*S28*) and captures variability across a range of morphological phenotypes upon deletion of each of the non-essential genes.

Chemical-genetic degree

A measure of chemical-genetic degree was derived from (*S29*), which measured the sensitivity of all non-essential deletion mutants to a library of drugs as well as a variety of environmental conditions. Specifically, we summed the number of drug and environmental sensitivities for a given knockout in the homozygous dataset that met a p-value < 0.05 cutoff.

PPI degree

The protein-protein interaction degree is the total number of interactions in the union of four high-throughput physical interaction datasets (*S23-26*). These capture interactions derived from AP/MS (*S23, 24*), Y2H (*S25*), and PCA-derived physical interactions (*S26*). Interactions from (*S23, 24*) were taken from BioGRID, interactions from (*S25*) were taken from the reported Y2H-Union dataset, and interactions from (*S26*) were defined using the 97.7% confidence cutoff as recommended

Protein disorder

The protein disorder measure is the percent of unstructured residues as predicted by the Disopred2 software (*S30*) and reported in (*S31*). For correlation and ANCOVA analyses, this percentage was used directly as a correlate. For categorical analysis, we adopt the binning suggested in the original study: genes reporting scores lower than 0.1 are considered to be structured, scores between 0.1 and 0.3 to be moderately unstructured and scores above 0.3 to be unstructured.

Expression level

Expression level measurements representing the average number of mRNA copies of each transcript per cell were taken from (*S32*).

Yeast conservation

Yeast conservation is the number of species that possess an ortholog of a given gene, when considering 23 different species of Ascomycota fungi. This measure was first described in (*S33*) (called "persistence"), and ortholog data was downloaded from www.broadinstitute.org/regev/orthogroups/. The 23 species are an expanded set of the 17 species described in (*S33*) with the additions of *S. octosporus*, *S. japonicus*, *L. elongosporus*, *C. parasilosis*, *C. tropicalis*, and *C. guilliermondii*.

Copy number volatility

Volatility measures the frequency of gain (including duplication) or loss events across 23 species of Ascomycota fungi from (*S33*). The 23 species are those provided above in the description of Yeast conservation.

dN/dS ratios

We computed the average *dN/dS* ratio for *S. cerevisiae* in comparison to the *sensu strictu* yeast species (*S. paradoxus*, *S. bayanus* and *S. mikatae*). Sequences were aligned using MUSCLE (*S34*) and *dN/dS* ratios were computed using PAML (*S35*).

Evolutionary age

As an approximate measure of evolutionary age we used a reconstruction of the ancestral species as described in (*S36*).

Gene duplicates

The list of duplicate pairs consists of those identified as the result of the whole genome duplication event as reconciled from several sources in (*S37*). Additionally, any pair of genes fulfilling established similarity requirements (*S38*) was reasoned to be a duplicate pair resulting from a small scale duplication event. Specifically, the gene pair must have a sufficient sequence similarity score (FASTA Blast, E = 10) and sufficient protein alignment length (> 80% of the longer protein). The pair must also have an amino acid level identity of at least 30% for proteins with aligned regions longer than 150 a.a. and for shorter proteins the identity must exceed $0.01n + 4.8L - 0.32^{(1 + \exp(-L/1000))}$, where L is the aligned length and $n = 6$ (*S38, 39*). Pairs from the whole genome duplication event (WGD) were combined with pairs determined through sequence alone (SSD).

Analysis of variation in number of genetic interactions across bioprocesses

Analysis of covariance (ANCOVA) was used to investigate the basis for variation in the number of total genetic interactions across biological processes. We measured the number of positive and negative interactions for all of the 3885 non-essential array deletion mutants at an intermediate cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). ANCOVA was used to measure the significance of the biological process of a gene in determining its degree of connectivity in the genetic interaction network in the presence of other gene-specific properties including the single mutant fitness defect and other properties examined in Fig. 3C. Specifically, we fit a linear model separately for the number of negative and positive genetic interactions, including each of the 17 broadly defined biological processes (see Materials & Methods, Biological process annotations) as independent variables along with the complete set of gene-specific features listed above (see Materials & Methods, Correlation analysis of genetic and physical interactions degree). We specifically tested each bioprocess term to assess whether it had a significant effect on genetic interaction degree in the presence of the other factors.

Results of ANCOVA on negative genetic interaction degree

For both negative and positive genetic interaction degree, the majority of the bioprocess categories were not significant on the basis of our ANCOVA analysis (i.e., p-value > 0.05), suggesting the number of genetic interactions in these neighborhoods were largely explained by properties at the gene level. However, a subset of the bioprocesses (5 of 17) had a significantly more or less negative interactions than can be explained by the other features: more than explained: chromatin/transcription, chromosome segregation, and glycosylation; less than explained: DNA replication/repair, and amino acid biosynthesis. The complete results of the ANCOVA analysis of negative interaction degree are presented in Table S4.

Results of ANCOVA on positive genetic interaction degree

For positive interactions, 4 of 17 neighborhoods have significantly more/less interactions (p-value < 0.05) than can be explained by the single gene features. Bioprocesses with more interactions than could be explained by gene-specific features were chromatin/transcription, nuclear-cytoplasmic transport, and lipid/sterol/fatty acid biosynthesis, while DNA replication/repair/HR/cohesion was the only process with significantly fewer than could be explained. The complete results of the ANCOVA analysis of positive interaction degree are presented in Table S5.

Assessing the contribution of gene-specific factors to interaction degree variation across bioprocesses

We measured the contribution of each of the gene-specific features in "explaining away" variance observed at the bioprocess level. Specifically, we ran ANCOVA analysis on the following two models for each functional neighborhood:

$$Model\ 1: \quad degree(g_i) = \mu + \alpha SMF(g_i) + \beta\delta_{bioprocess\,g_i}$$

$$Model\ 2: \quad degree(g_i) = \mu + \alpha\, SMF(g_i) + \beta\delta_{bioprocess\,g_i} + \gamma\, factorX(g_i)$$

where $SMF(g_i)$ is the single mutant fitness of the current gene, $\delta_{\text{bioprocess}\,g_i}$ is the binary categorical variable for the corresponding bioprocess and $factorX(g_i)$ is the factor of interest. We measured the influence of *factorX* on the observed variation in degree for each bioprocess as follows:

$$\log\left(\frac{F_{\text{model}\,1}}{F_{\text{model}\,2}}\right)$$

where $F_{\text{model}\,i}$ is the F-statistic corresponding to the bioprocess variable in each model. Intuitively, this reflects the degree to which the bioprocess contribution to genetic interaction degree is "explained away" by the particular feature being evaluated. High quantities here indicate that much of the variation in degree for that bioprocess is explained by the factor of interest. The higher this score, the higher the contribution of that variable. We note that this measure can actually result in negative influence of some factors, indicating that adding that particular feature actually *increases* the variance attributed to that particular bioprocess, not *explains away*.

Chemical genomic analysis

Chemical genomic assays were performed as described previously (*S29*). The obtained fitness defect scores (log_2 control/treatment, Data file S7) were mean normalized by gene and by experiment. This data, combined with the dataset described in (*S29*), was used for comparison to the SGA dataset.

Supplementary Tables

Table S1. List of the likely contaminant proteins

List of likely contaminant proteins, detected in >10% of the 62 TAP-MS analyses. NCBI gene ID and protein ID are listed for each protein, alongside their frequency of detection in TAP-MS analyses. These proteins were removed from the primary list of interactors obtained for Sgt2-TAP and Ubp15-TAP.

Hit Gene Name	Hit Gene ID	Hit Protein ID	Frequ-ency		Hit Gene Name	Hit Gene ID	Hit Protein ID	Frequ-ency
ALB	280717	30794280	80.6		RPL35A	851336	6320010	40.3
ATP3	852327	6319513	40.3		RPL36B	855826	6325006	16.1
CDC19	851193	6319279	19.4		RPL38	851035	6323357	45.2
CMD1	852406	6319585	41.9		RPL43B	853557	6322554	19.4
DED1	854379	6324778	27.4		RPL4A	852319	6319505	45.2
DEF1	853811	6322796	74.2		RPL5	855972	6325126	12.9
EFT2	851993	6320593	12.9		RPL6A	854902	6323567	16.1
HHF1	852294	6319481	14.5		RPL6B	851169	6323481	37.1
HSC82	855224	6323840	21		RPL7A	852804	6321362	32.3
HSP60	850963	6323288	11.3		RPL8A	856352	6321754	14.5
ICS2	852454	41629678	67.7		RPL8B	850682	6322984	11.3
ILV2	855135	6323755	12.9		RPL9A	852730	6321291	22.6
INO80	852728	6321289	16.1		RPP0	851052	6323371	35.5
ISM1	856067	6325217	43.5		RPS0A	853128	6321653	48.4
KAR2	853418	6322426	17.7		RPS11B	852337	6319522	16.1
KRT13	3860	4504911	11.3		RPS12	854551	6324945	12.9
KRT14	3861	15431310	45.2		RPS13	851636	6320269	19.4
KRT2	3849	47132620	62.9		RPS14B	853248	6322270	80.6
KRT4	3851	109255249	19.4		RPS15	854117	6324533	32.3
KRT5	3852	119395754	29		RPS16B	851476	6320120	41.9
KRT6A	3853	5031839	19.4		RPS17B	852058	6320655	27.4
LIP2	850940	6323268	71		RPS18A	852061	6320658	69.4
LSM12	856521	6321913	77.4		RPS19B	855414	6324027	43.5
MYO1	856418	6321812	53.2		RPS1A	851162	6323474	51.6
NOP58	854487	6324886	11.3		RPS1B	854939	6323577	48.4
PBP1	853089	6321617	77.4		RPS2	852754	6321315	19.4
PBP4	851507	6320150	77.4		RPS20	856371	6321772	11.3
RPL10	850764	6323104	21		RPS22A	853249	6322271	27.4
RPL11B	852976	6321522	58.1		RPS23A	853015	6321556	27.4
RPL12B	852026	6320625	53.2		RPS24A	856805	6320918	16.1
RPL13A	851477	6320121	59.7		RPS25A	852911	6321464	12.9
RPL14B	856388	6321786	21		RPS26B	856868	6320978	16.1
RPL15A	850716	6323057	17.7		RPS3	855543	6324151	74.2
RPL16A	854673	6322058	21		RPS4B	856610	6321997	24.2
RPL16B	855655	6324260	24.2		RPS5	853587	6322583	61.3

RPL17A	853674	6322668	16.1		RPS6B	852479	6319658	54.8
RPL17B	853262	6322284	22.6		RPS7A	854263	6324670	64.5
RPL18B	855415	6324028	30.6		RPS7B	855628	6324233	67.7
RPL19B	852254	6319444	33.9		RPS8A	852206	6319399	21
RPL20B	854489	6324888	12.9		RPS9B	852487	6319666	43.5
RPL21A	852489	6319668	33.9		SCM4	852940	6321486	17.7
RPL22A	850750	6323090	29		SIK1	850894	6323226	22.6
RPL23A	852191	6319384	54.8		SRO9	850320	37362625	19.4
RPL24A	852852	6321407	21		SSA1	851259	6319314	62.9
RPL25	853993	6324445	59.7		SSA2	850636	6323004	74.2
RPL26A	851058	6323376	74.2		SSB1	851369	6319972	62.9
RPL27B	852082	6320679	17.7		SSB2	855512	6324120	29
RPL28	852775	6321335	12.9		TDH3	853106	6321631	67.7
RPL2B	854794	6322171	24.2		TEF2	852415	6319594	54.8
RPL3	854229	6324637	33.9		THI7	850938	6323266	21
RPL30	852853	6321408	24.2		TUP1	850445	6319926	11.3
RPL31A	851484	6320128	48.4		UTP22	852982	6321527	11.3
RPL32	852185	6319378	16.1		VMA2	852424	6319603	19.4
RPL33A	855960	6325114	12.9		VPS72	852096	37362638	11.3
RPL33B	854409	6324808	11.3					

Table S2. Detailed mass spectrometric evidence

Proteins detected by 2 or more unique peptides and not identified as likely contaminants (Table S1) are listed, alongside MS scores. "Hit Score" is the score from the Mascot search engine, "n specs" is the total number of spectra (or peptides), "n peps" is the number of unique peptides, "coverage" is the percentage of the amino acid sequence observed. "Max specs", "occurrence" and "enrichment" refer to the comparison of the individual runs (Sgt2-TAP, Ubp15-TAP) with 19 unrelated TAP-MS runs performed in the same time-frame. "Max specs" lists the maximum number of spectra assigned to the given hit across the unrelated 19 baits, "occurrence" refers to the number of times the hit was detected across these 19 purifications. "Enrichment" refers to the fold increase in spectral counts in the test purification as compared to the "max specs" across the 19 control purifications. Highlighted in green are proteins specifically enriched with the isolated bait and for which 5 spectra or more were identified.

Bait Gene Name	Hit Gene Name	Hit Gene ID	Hit Score	n specs	n peps	Cove-rage	Max specs	Occur-rence	Enrich-ment
SGT2(C-TAP)	SGT2	854168	1437	158	18	69.9	0	0	BAIT
SGT2(C-TAP)	YOR164C	854335	855	32	14	59.6	0	0	∞
SGT2(C-TAP)	SSE2	852467	1087	24	20	34.3	0	0	∞
SGT2(C-TAP)	MDY2	854038	388	17	7	50.5	0	0	∞
SGT2(C-TAP)	SSE1	855998	519	11	6	19	2	2	9.5
SGT2(C-TAP)	TIF2	853303	209	6	4	16.2	3	1	2
SGT2(C-TAP)	ENO1	853169	245	5	4	14.6	2	1	2.5
SGT2(C-TAP)	HSP42	851751	137	5	3	9.3	0	0	∞
SGT2(C-TAP)	TUB2	850506	192	4	4	13.3	2	2	2
SGT2(C-TAP)	MKT1	855639	184	3	3	5.3	0	0	∞
SGT2(C-TAP)	PMA1	852876	133	3	3	3.8	10	3	0.3
SGT2(C-TAP)	ADH1	854068	79	2	2	6	0	0	∞
SGT2(C-TAP)	ATP1	852177	109	2	2	4.6	0	0	∞
SGT2(C-TAP)	GPM1	853705	89	2	2	12.6	2	2	1
SGT2(C-TAP)	HSP26	852364	87	2	2	12.1	0	0	∞
SGT2(C-TAP)	RPO21	851415	111	2	2	1.7	0	0	∞
SGT2(C-TAP)	URA2	853311	101	2	2	1.4	2	1	1
UBP15(C-TAP)	UBP15	855350	4434	796	60	65.3	11	1	BAIT
UBP15(C-TAP)	ECM30	851156	3226	203	45	61.6	0	0	∞
UBP15(C-TAP)	DIG1	856058	511	13	8	26.8	4	7	3.25
UBP15(C-TAP)	STE12	856484	538	13	9	18.3	5	2	2.6
UBP15(C-TAP)	YMR160W	855195	264	5	5	8.1	0	0	∞
UBP15(C-TAP)	UME6	851788	202	4	4	10.9	0	0	∞
UBP15(C-TAP)	DIG2	852091	179	3	3	16.7	0	0	∞
UBP15(C-TAP)	CDC55	852685	132	2	2	5.3	0	0	∞
UBP15(C-TAP)	LSB3	850580	120	2	2	8	0	0	∞
UBP15(C-TAP)	RPL40A	854658	91	2	2	14.1	0	0	∞
UBP15(C-TAP)	SSD1	851887	186	2	2	2.9	0	0	∞
UBP15(C-TAP)	YGR250C	853165	74	2	2	3.8	4	3	0.5

Table S3. List of datasets used for comparative analysis of function prediction

Data type	Dataset name	Reference
AP/MS	Gavin 2006	(*S24*)
	Krogan 2006	(*S23*)
	Ho 2002	(*S41*)
Y2H	Yu 2008	(*S25*) (CCSB-YI1)
	Y2H Union	(*S25*) (union of CCSB-YI1, Ito-core and Uetz-screen)
PCA	Tarassov 2008	(*S26*)
Other PPI	Miller 2005	(*S42*)
Gene expression	Chua 2006	(*S43*) (union of deletion and overexpression datasets)
	Spellman 1998	(*S44*)
	Gasch 2000	(*S45*)
	Hughes 2000	(*S46*)
	Roberts 2000	(*S47*)
	Yvert 2003	(*S48*)
Protein localization	Huh 2003	(*S49*)
Genetic interactions	Tong 2004	(*S50*)
	Schuldiner 2005	(*S51*)
	Pan 2006	(*S52*)
	Collins 2007	(*S53*)
	Lin 2008	(*S54*)
	Wilmes 2008	(*S7*)
	Fiedler 2009	(*S8*)
Phosphorylation	Ptacek 2005	(*S55*)
Chemical genomics	McClellan 2007	(*S56*)

Table S4. ANCOVA analysis of negative genetic interaction degree

Source	Sum Sq. Error	d.f.	Mean Sq. Error	F-statistic	P-value	Coefficient
Single mutant fitness	750703.86	1	750703.86	1515.87	0.00	-281.65
Expression level	1891.59	1	1891.59	3.82	0.05	-0.22
PPI degree	1769.83	1	1769.83	3.57	0.06	0.09
Phenotypic capacitance	10546.64	1	10546.64	21.30	0.00	8.88
Evolutionary age	1976.93	3	658.98	1.33	0.26	-0.53, 0.42, -1.88, 1.99
Multi-function	4257.01	1	4257.01	8.60	0.00	0.76
Copy number volatility	214.45	1	214.45	0.43	0.51	-0.02
dN/dS	3165.72	1	3165.72	6.39	0.01	-17.83
Local conservation	3954.93	1	3954.93	7.99	0.00	0.38
Chemical-genetic degree	90079.40	1	90079.40	181.89	0.00	0.24
Duplicate	8036.69	1	8036.69	16.23	0.00	-2.23
Protein disorder	3574.10	1	3574.10	7.22	0.01	7.66
Protein folding & glycosylation, cell wall biogenesis & integrity	2204.75	1	2204.75	4.45	0.03	2.65
Cell polarity & morphogenesis	685.87	1	685.87	1.38	0.24	-1.32
Drug/ion-transport	436.95	1	436.95	0.88	0.35	-1.02
Metabolism & mitochondria	127.07	1	127.07	0.26	0.61	-0.36
G1/S and G2/M cell cycle progression/meiosis	13.92	1	13.92	0.03	0.87	-0.26
ER-Golgi traffic	902.42	1	902.42	1.82	0.18	2.48
Golgi-endosome-vacuole sorting	413.35	1	413.35	0.83	0.36	1.09
Ribosome & translation	1323.88	1	1323.88	2.67	0.10	-2.03
Nuclear-cytoplasmic transport	289.54	1	289.54	0.58	0.44	1.81
Lipid, sterol, fatty acid biosynthesis	766.90	1	766.90	1.55	0.21	1.65
Signaling & stress response	1134.58	1	1134.58	2.29	0.13	1.45
Chromatin/transcription	5529.19	1	5529.19	11.16	0.00	3.39
Protein degradation, proteosome	104.47	1	104.47	0.21	0.65	-0.73
RNA processing	158.11	1	158.11	0.32	0.57	-0.82
DNA replication & repair, HR, cohesion	4819.22	1	4819.22	9.73	0.00	-3.85
Chromosome segregation, kinetochore, spindle, microtubule	3361.23	1	3361.23	6.79	0.01	3.67
Amino acid biosynthesis & transport, nitrogen utilization	4010.91	1	4010.91	8.10	0.00	-3.91

Table S5. ANCOVA analysis of positive genetic interaction degree

Source	Sum Sq. Error	d.f.	Mean Sq. Error	F-statistic	P-value	Coefficient
Single mutant fitness	152579.20	1	152579.20	847.10	0.00	-126.92
Expression level	0.11	1	0.11	0.00	0.98	0.00
PPI degree	1037.66	1	1037.66	5.76	0.02	0.07
Phenotypic capacitance	4715.66	1	4715.66	26.18	0.00	5.95
Evolutionary age	921.19	3	307.06	1.70	0.16	-0.71, 0.82, -0.78, 0.67
Multi-function	1344.34	1	1344.34	7.46	0.01	0.43
Copy number volatility	195.80	1	195.80	1.09	0.30	-0.02
dN/dS	1567.34	1	1567.34	8.70	0.00	-12.34
Local conservation	76.71	1	76.71	0.43	0.51	0.05
Chemical-genetic degree	33176.10	1	33176.10	184.19	0.00	0.15
Duplicate	708.04	1	708.04	3.93	0.05	-0.66
Protein disorder	913.28	1	913.28	5.07	0.02	3.86
Protein folding & glycosylation, cell wall biogenesis & integrity	14.43	1	14.43	0.08	0.78	0.21
Cell polarity & morphogenesis	659.89	1	659.89	3.66	0.06	-1.29
Drug/ion-transport	266.19	1	266.19	1.48	0.22	-0.79
Metabolism & mitochondria	12.48	1	12.48	0.07	0.79	-0.11
G1/S and G2/M cell cycle progression/meiosis	2.86	1	2.86	0.02	0.90	-0.12
ER-Golgi traffic	22.62	1	22.62	0.13	0.72	0.40
Golgi-endosome-vacuole sorting	419.50	1	419.50	2.33	0.13	1.10
Ribosome & translation	17.69	1	17.69	0.10	0.75	-0.23
Nuclear-cytoplasmic transport	1544.63	1	1544.63	8.58	0.00	4.17
Lipid, sterol, fatty acid biosynthesis	1009.86	1	1009.86	5.61	0.02	1.92
Signaling & stress response	492.89	1	492.89	2.74	0.10	0.95
Chromatin/transcription	1120.68	1	1120.68	6.22	0.01	1.52
Protein degradation, proteosome	18.64	1	18.64	0.10	0.75	0.30
RNA processing	248.22	1	248.22	1.38	0.24	1.02
DNA replication & repair, HR, cohesion	1961.88	1	1961.88	10.89	0.00	-2.47
Chromosome segregation, kinetochore, spindle, microtubule	0.88	1	0.88	0.00	0.94	-0.06
Amino acid biosynthesis & transport, nitrogen utilization	337.17	1	337.17	1.87	0.17	-1.13

Table S6. Sensitivity and precision of SGA genetic interaction scores

The number of interactions was calculated after processing reciprocal interactions as described in Materials & Methods, Data files S3-S5.

Cutoff	Negative interactions			Positive interactions		
	Num. interact.	Sensitivity	Precision	Num.interact.	Sensitivity	Precision
Lenient $p < 0.05$	366,085	0.41	0.27	323,935	0.26	0.20
Intermediate $\|\varepsilon\| > 0.08$, $p < 0.05$	108,417	0.35	0.63	59,887	0.18	0.59
Stringent $\varepsilon < -0.12$, $p < 0.05$ $\varepsilon > 0.16$, $p < 0.05$	58,508	0.28	0.89	6,185	0.05	~1

Supplementary Figures

Fig. S1. The SGA dataset

(**A**) Development of the SGA genetic interaction score, which provides a measure of the genetic interaction strength (ε) and a p-value reflecting the confidence in interaction reproducibility (see Material & Methods). (**B**) Distribution of genetic interaction scores on linear (i) and logarithmic (ii) scales. As expected, most double mutant pairs show no genetic interaction (low ε values, black in ii), while fewer pairs exhibit negative (red in ii) and positive (green in ii) interactions. Negative interactions are approximately 2-fold more prominent than positive interactions. (**C**) Functional coverage evaluated on the basis of the extent to which query mutant strains screened in SGA covered 10 broad functional categories. The 10 groups were defined using a Bayesian framework for data integration as described in (*S57*). Approximately 30% of all genes were screened genome-wide by SGA, which translates to at least 35% coverage of each partially overlapping functional category (light blue). Dark blue bars indicate the proportion of essential genes screened in each category. (**D**) Recall of published genetic interactions. Both negative (i) and positive (ii) SGA genetic interactions (ε) are significantly enriched for known genetic interactions. Extreme negative ($\varepsilon < -0.8$) and positive ($\varepsilon > 0.25$) interactions are approximately 140- and 80-fold enriched for previously identified interactions, respectively. (**E**) Dataset size comparison. We compared the total number of negative (red) and positive (green) genetic interactions at our intermediate and stringent cutoffs (see Materials & Methods) to those reported by the 10 largest genetic interaction datasets available from Biogrid. The intermediate confidence SGA dataset represents a 3-fold increase over the number of reported genetic interactions. (**F**) Comparison of SGA genetic interaction score with high-resolution liquid growth profiles (see Materials & Methods). We find strong agreement between scores for the same gene pairs screened on solid vs. liquid medium ($r = 0.89$). This agreement was observed for both extreme and modest genetic interactions.

Fig. S2. Functional evaluation of genetic interaction scores

(**A**) False positive and false negative rates. Sensitivity (black line) and precision (blue line) (see Materials & Methods) were systematically measured at different SGA genetic interaction score (ε) cutoffs (x-axis) and two p-value cutoffs ($p < 0.5$, dashed line, and $p < 0.05$, solid line). (iii) The same analysis was repeated for a subset of 211 replicate pairs of screens. We observed a 20–25% increase in precision for interactions derived from SGA screens with two replicates. Thus, we anticipate that additional rounds of screening will result in further improvements in precision and sensitivity. (**B**) Enrichment analysis on individual SGA genetic interactions. Enrichment was assessed systematically at different SGA genetic interaction score (ε) cutoffs as described (see Materials & Methods). Gene pairs exhibiting extreme negative interactions are significantly enriched for physically interacting pairs (blue line, >85-fold enrichment at $\varepsilon = -0.9$) as well as Gene Ontology (GO) co-annotated gene pairs (black line, >16-fold enrichment at $\varepsilon = -0.9$). Extreme positive interactions are similarly enriched for protein-protein interactions (>9-fold enrichment at $\varepsilon = 0.4$) and co-annotated gene pairs (~8-fold enrichment at

ε = 0.4). (**C**) Enrichment analysis on SGA genetic interaction profile similarities. Genetic profile similarity was determined by computing Pearson correlation coefficients for gene pairs using their complete genetic interaction profiles (negative and positive interactions, black line), only negative interactions (red line) and only positive interactions (green line). Precision and recall against protein-protein interaction (i) and GO co-annotation (ii) standards were calculated as described previously (*S5*). Similarity over positive interactions alone is less informative than similarity over negative interactions alone indicating a higher incidence of noise or less functionally informative interactions profiles in the positive interaction dataset. However, optimal predictive power is obtained using complete profiles based on both negative and positive interactions. (**D**) Function prediction capacity of the genetic interaction dataset. (i) We compared the performance of four data types (genetic interactions, PPIs, gene expression and cellular localization) in predicting gene function as determined by the GeneMANIA algorithm (*S9*) (see Materials & Methods). Despite covering only 30% of genes in the genome, the SGA dataset alone shows function prediction capacity similar to the combination of two AP/MS studies. (ii) Comparison of individual high-throughput datasets in predicting gene function (see Materials & Methods). The genetic interactions reported in this study provide the second best predictive power. (iii) Comparison of high-throughput genetic interaction studies using the leave-one-out approach (see Materials & Methods). Removing the genetic interactions from the current study results in the greatest loss in average precision.

Fig. S3. par32Δ mutants exhibit altered Tor-regulated responses in the presence of ammonia

Wild-type (BY4741) and *par32*Δ cells transformed with pMRT2 (*MET17*, *HIS3*, *LEU2*) and either (**A**) pMS29 (p*GAP1-lacZ*) or (**B**) pCK272 (p*CIT2-lacZ*) were cultured in SD ammonia medium at 24°C and assayed for β-galactosidase activity as described in (*S58*). Data shown is the average and standard deviation for three to six assays. Loss of *PAR32* results in reduced *GAP1* and increased *CIT2* reporter gene activity when grown in the presence of ammonia. These results suggest that Par32, which is phosphorylated in response to rapamycin (www.yeastgenome.org), may function in Tor-mediated regulation of Gln3, Gat1 and Rtg1/3.

Fig. S4. Sgt2 physical interactions and Pex15 localization

(**A**) Similar to GET pathway mutants, *SGT2* is required for proper Pex15 localization (see Materials & Methods). (**B**) Pex15 mis-localization to mitochondria is exacerbated in a strain lacking both *SGT2* and *PEX19*.

Fig. S5. Correlation between genetic interaction network degree and single mutant fitness defects

(**A**) Genetic interaction degree vs. single mutant fitness trend. (Left) The fraction of screened pairs that have interactions (either positive, negative, or combined) among the 3885 array genes at the intermediate cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). The genes are binned into seven groups on the basis of their single mutant fitness. (Right) Degree vs. single mutant fitness trend for query strains. The fraction of screened pairs that have interactions

among the 1712 query genes across the 3885 arrays at the intermediate cutoff. Moderately sick query mutants tend to show more interactions similar to the trend for array strains, but mutants below ~0.6 fitness show a decrease in both the number of positive and negative genetic interactions identified in SGA screens. This asymmetry is a side effect of the increased experimental noise associated with the screening of sick query mutants, which results in a loss of interaction sensitivity for extremely sick query mutants. (**B**) Genetic interaction degree vs. single mutant fitness trend for different types of genetic perturbations. The fractions of interactions among the 1712 query genes across 3885 array genes are separated by the type of genetic perturbation used (either deletion, temperature-sensitive allele (TS), or DAmP allele). The number of interactions for essential genes correlates with allele fitness measured at semi-permissive temperature (26^{o}C), which is similar to the trend for deletion mutants. A decrease in the number of interactions for very sick queries is a result of increased experimental noise as described in (A). (**C**) Functional relatedness of genetic interaction partners of mutants grouped by their single mutant fitness. To rule out the possibility that the increased number of genetic interactions identified for mutants with severe fitness defects was the result of increased experimental noise, we evaluated the functional coherence among each gene's interaction partners. Specifically, the utility of genetic interaction scores (ε) for predicting functional associations between genes was assessed on the basis of a gold standard derived from the biological process aspect of the Gene Ontology, as described in (*S5*). The area under an ROC curve (AUC) was measured using the Wilcoxon-Mann-Whitney statistic for each gene. Per-gene AUC scores were averaged over each set of genes, grouped by their fitness. On average, genetic interaction partners of sicker mutants tend to be more functionally related than interaction partners of healthy single mutants. (i) shows results of this analysis on array genes' interactions and (ii) shows the analogous experiment for query genes' interactions**.** (**D**) (left) Correlation of various features with genetic interaction negative degree, genetic interaction positive degree, and PPI degree controlling for the linear effect of single mutant fitness, using partial correlation analysis, as described in Materials & Methods, Correlation analysis of genetic and physical interaction degree. (**E**) Correlation between PPI degree and gene-specific features. Pearson correlation coefficients were computed between the panel of gene-specific features with degree as measured by three different physical interaction technologies. Data sets used for the various PPI networks are described in Materials & Methods. All interaction degree is measured on a set of interactions at the intermediate cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). 95% confidence intervals on correlation coefficients are indicated for all plots. (**F**) Analysis of evolutionary rates for disordered and non-disordered genetic interaction hubs. Negative genetic interactions were used to define a set of hub genes (> 95th percentile) and non-hubs genes (< 50th percentile) and for each group, the average *dN*/d*S* was measured for structurally disordered and ordered proteins. The disordered and ordered proteins were defined as the > 95th percentile and < 50th percentile based on the disorder measure reported in (*S31*). Disordered genetic interaction hubs tend to evolve significantly faster than other genes (both hubs and non-hubs).

Fig. S6. Positive genetic interaction frequency within and between bioprocesses

(**A**) The fraction of screened gene pairs exhibiting interactions was measured for 17 broadly defined functional neighborhoods (see Materials & Methods, Biological process annotations). The color of each process-process element reflects the fraction of positive genetic or AP/MS-derived interactions (blue: below frequency of random pairs; black: random background of interactions; yellow: above frequency of random pairs), with the diagonal representing within-process interactions. Functional neighborhoods for which the frequency of interactions could not be statistically distinguished from background ($p < 0.05$) are shown as black. (**B**) Gene-specific factors explaining the variation in number of positive genetic interactions across biological processes. ANCOVA analysis was used to assess the significance of gene-specific factors at predicting the number of genetic interactions observed for a specific biological process (see Materials & Methods). Single mutant fitness defect was always included as a factor in the analysis. The top panel plots the average number of positive genetic interactions across each process with the color indicating processes that have more interactions than expected ($p < 0.05$), processes whose positive interaction degree is explained by the factors indicated on the y-axis, and those with fewer interactions than expected ($p < 0.05$). For each process, the proportion of variance explained by each of the gene-specific factors is indicated by the corresponding column in the heat map. Four of 17 neighborhoods have significantly more/less interactions ($p < 0.05$) than can be explained by the single gene features. Bioprocesses with more interactions than could be explained by gene-specific features were chromatin/transcription, nuclear-cytoplasmic transport, and lipid/sterol/fatty acid biosynthesis, while DNA replication/repair/HR/cohesion was the only process with significantly fewer than could be explained.

Fig. S7. Evaluation of the overlap of the genetic and physical genetic interaction networks

The coverage of protein-protein interactions by SGA-derived genetic interactions. We measured the direct overlap between the genetic interactions reported in this study and physical interactions reported by four high-throughput physical interaction datasets based on three different technologies: AP/MS (*S23, 24*), Y2H (*S25*) and PCA (*S26*). For physical interactions identified by each of the three physical interaction technologies, we measured the fraction of pairs that also showed a genetic interaction (both positive and negative) at the intermediate cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). For comparison, the fraction of pairs identified by other protein-protein interaction technologies (excluding the dataset of interest) was also measured. The results illustrate that the coverage of protein-protein interactions by genetic interactions often exceeds the overlap between two independent physical interaction studies. Only PPIs that were screened in this study were considered.

Fig. S8. Homozygous and heterozygous profiling of Erodoxin

(**A**) Homozygous deletion profiling. Scatter plot visualization of results from homozygous profiling of Erodoxin. The y-axis represents the sensitivity of each homozygous deletion strain when treated for 5 generations in the presence of drug. Strains exhibiting severe sensitivity to Erodoxin are highlighted. Experiments were

conducted as described previously (*S29*). **(**B) Heterozygous deletion profiling. Scatter plot visualization of results from heterozygous profiling of Erodoxin. The y-axis represents the sensitivity of each heterozygous essential gene deletion strain when treated for 5 generations in the presence of drug. Strains exhibiting severe sensitivity to Erodoxin are highlighted. *ERO1* encodes an enzyme participating in disulfide bond formation by recharging protein disulfide isomerase (PDI) into its oxidized state (*S19*). *FAD1* encodes a cofactor that functions in the Ero1 reaction cycle (*S59*). Experiments were conducted as described previously (*S29*).

Data Files

All data files can be downloaded from:
http://drygin.ccbr.utoronto.ca/~costanzo2009/

Data file S1. Raw data file

The file contains the complete SGA genetic interaction dataset in a tab-delimited format with 13 columns: Query ORF, Query gene name, Array ORF, Array gene name, Genetic interaction score (ε), Standard deviation, p-value, Query single mutant fitness (SMF), Query SMF standard deviation, Array SMF, Array SMF standard deviation, Double mutant fitness, Double mutant fitness standard deviation. No cutoff or filtering was applied.

Data file S2. Raw data matrix

The file contains the complete SGA genetic interaction matrix in Java Treeview format. No cutoff or filtering was applied.

Data file S3. Dataset at lenient cutoff

The file contains the SGA genetic interaction dataset with a lenient cutoff applied (p-value < 0.05). Reciprocal interactions, where query A was crossed to array B and query B was crossed to array A, were processed as follows. If genetic interaction scores for the double mutants AB and BA show opposite interaction signs (AB is positive and BA is negative, or vice-versa), both pairs were removed. If genetic interaction scores for AB and BA show the same interaction sign (both positive or both negative), the interaction with the lowest p-value was retained. The file is provided in a tab-delimited format with 7 columns: Query ORF, Query gene name, Array ORF, Array gene name, Genetic interaction score (ε), Standard deviation, p-value.

Data file S4. Dataset at intermediate cutoff

The file contains the SGA genetic interaction dataset with an intermediate cutoff applied ($|\varepsilon| > 0.08$, p-value < 0.05). Reciprocal interactions, where query A was crossed to array B and query B was crossed to array A, were processed as follows. If genetic interaction scores for the double mutants AB and BA show opposite interaction signs (AB is positive and BA is negative, or viceversa), both pairs were removed. If genetic interaction scores for AB and BA show the same interaction sign (both positive or both negative), the interaction with the lowest p-value was retained. The file is provided in a tab-delimited format with 7 columns: Query ORF, Query gene name, Array ORF, Array gene name, Genetic interaction score (ε), Standard deviation, p-value. All analyses were performed on this dataset unless otherwise indicated.

Data file S5. Dataset at stringent cutoff

The file contains the SGA genetic interaction dataset with a stringent cutoff applied ($\varepsilon < -0.12$, p-value < 0.05 or $\varepsilon > 0.16$, p-value < 0.05). Reciprocal interactions, where query A was crossed to array B and query B was crossed to array A, were processed as

follows. If genetic interaction scores for the double mutants AB and BA show opposite interaction signs (AB is positive and BA is negative, or viceversa), both pairs were removed. If genetic interaction scores for AB and BA show the same interaction sign (both positive or both negative), the interaction with the lowest p-value was retained. The file is provided in a tab-delimited format with 7 columns: Query ORF, Query gene name, Array ORF, Array gene name, Genetic interaction score (ε), Standard deviation, p-value.

Data file S6. Biological process annotations

Data file S7. Chemical genomics dataset

The file contains the fitness defect scores for 4933 non-essential homozygous and 1107 essential heterozygous deletion mutants in the presence of 11 chemical treatments. Each row represents a gene deletion strain, given by its ORF name. Each column represents a treatment experiment.

Data file S8. The Genetic Landscape of a Cell

A poster illustrating the global cell map based on genetic interaction profiles (as in Fig. 1) and eight sub-networks, each corresponding to a magnified region of the global map. Node color corresponds to a biological processes indicated in the legend.

Data file S9. List of query mutants screened in this study

The file contains the list of 1712 query mutants used in the study.

Data file S10. List of arrays mutants screened in this study

The file contains the list of 3885 array mutants used in the study.

Supplementary References

S1. R. Mani, R. P. St Onge, J. L. t. Hartman, G. Giaever, F. P. Roth, *Proc Natl Acad Sci U S A* **105**, 3461 (2008).
S2. J. L. Y. Koh *et al.*, *Nucleic Acids Res* **38**, (2009).
S3. A. H. Tong, C. Boone, *Methods Mol Biol* **313**, 171 (2006).
S4. P. Shannon *et al.*, *Genome Res* **13**, 2498 (2003).
S5. C. L. Myers, D. R. Barrett, M. A. Hibbs, C. Huttenhower, O. G. Troyanskaya, *BMC Genomics* **7**, 187 (2006).
S6. R. P. St Onge *et al.*, *Nat Genet* **39**, 199 (2007).
S7. G. M. Wilmes *et al.*, *Mol Cell* **32**, 735 (2008).
S8. D. Fiedler *et al.*, *Cell* **136**, 952 (2009).
S9. S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, Q. Morris, *Genome Biol* **9 Suppl 1**, S4 (2008).
S10. S. van Dongen, University of Utrecht (2000).
S11. A. L. Risinger, N. E. Cain, E. J. Chen, C. A. Kaiser, *Mol Biol Cell* **17**, 4411 (2006).
S12. K. J. Roberg, S. Bickel, N. Rowley, C. A. Kaiser, *Genetics* **147**, 1569 (1997).
S13. Y. Hu *et al.*, *Genome Res* **17**, 536 (2007).
S14. S. Alberti, A. D. Gitler, S. Lindquist, *Yeast* **24**, 913 (2007).
S15. A. C. Gingras *et al.*, *Mol Cell Proteomics* **4**, 1725 (2005).
S16. J. Peng, J. E. Elias, C. C. Thoreen, L. J. Licklider, S. P. Gygi, *J Proteome Res* **2**, 43 (2003).
S17. D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **20**, 3551 (1999).
S18. J. E. Elias, S. P. Gygi, *Nat Methods* **4**, 207 (2007).
S19. C. S. Sevier *et al.*, *Cell* **129**, 333 (2007).
S20. C. S. Sevier, C. A. Kaiser, *Mol Biol Cell* **17**, 2256 (2006).
S21. E. I. Boyle *et al.*, *Bioinformatics* **20**, 3710 (2004).
S22. P. H. Westfall, S. S. Young, *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*. (Wiley, New York, 1993).
S23. N. J. Krogan *et al.*, *Nature* **440**, 637 (2006).
S24. A. C. Gavin *et al.*, *Nature* **440**, 631 (2006).
S25. H. Yu *et al.*, *Science* **322**, 104 (2008).
S26. K. Tarassov *et al.*, *Science* **320**, 1465 (2008).
S27. M. G. Kendall, A. Stuart, J. K. Ord, S. F. Arnold, A. O'Hagan, *Kendall's advanced theory of statistics*. (Edward Arnold; Halsted Press, London, New York, ed. 6th, 1994), pp. v. in 4.
S28. S. F. Levy, M. L. Siegal, *PLoS Biol* **6**, e264 (2008).
S29. M. E. Hillenmeyer *et al.*, *Science* **320**, 362 (2008).
S30. J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, D. T. Jones, *Bioinformatics* **20**, 2138 (2004).
S31. J. Gsponer, M. E. Futschik, S. A. Teichmann, M. M. Babu, *Science* **322**, 1365 (2008).
S32. F. C. Holstege *et al.*, *Cell* **95**, 717 (1998).
S33. I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, *Nature* **449**, 54 (2007).

S34. R. C. Edgar, *BMC Bioinformatics* **5**, 113 (2004).
S35. Z. Yang, *Mol Biol Evol* **24**, 1586 (2007).
S36. V. Kunin, C. A. Ouzounis, *Bioinformatics* **19**, 1412 (2003).
S37. K. P. Byrne, K. H. Wolfe, *Genome Res* **15**, 1456 (2005).
S38. Z. Gu, A. Cavalcanti, F. C. Chen, P. Bouman, W. H. Li, *Mol Biol Evol* **19**, 256 (2002).
S39. B. Rost, *Protein Eng* **12**, 85 (1999).
S40. S. Pu, J. Wong, B. Turner, E. Cho, S. J. Wodak, *Nucleic Acids Res* **37**, 825 (2009).
S41. Y. Ho *et al.*, *Nature* **415**, 180 (2002).
S42. J. P. Miller *et al.*, *Proc Natl Acad Sci U S A* **102**, 12123 (2005).
S43. G. Chua *et al.*, *Proc Natl Acad Sci U S A* **103**, 12045 (2006).
S44. P. T. Spellman *et al.*, *Mol Biol Cell* **9**, 3273 (1998).
S45. A. P. Gasch *et al.*, *Mol Biol Cell* **11**, 4241 (2000).
S46. T. R. Hughes *et al.*, *Cell* **102**, 109 (2000).
S47. C. J. Roberts *et al.*, *Science* **287**, 873 (2000).
S48. G. Yvert *et al.*, *Nat Genet* **35**, 57 (2003).
S49. W. K. Huh *et al.*, *Nature* **425**, 686 (2003).
S50. A. H. Tong *et al.*, *Science* **303**, 808 (2004).
S51. M. Schuldiner *et al.*, *Cell* **123**, 507 (2005).
S52. X. Pan *et al.*, *Cell* **124**, 1069 (2006).
S53. S. R. Collins *et al.*, *Nature* **446**, 806 (2007).
S54. Y. Y. Lin *et al.*, *Genes Dev* **22**, 2062 (2008).
S55. J. Ptacek *et al.*, *Nature* **438**, 679 (2005).
S56. A. J. McClellan *et al.*, *Cell* **131**, 121 (2007).
S57. C. L. Myers *et al.*, *Genome Biol* **6**, R114 (2005).
S58. E. J. Chen, C. A. Kaiser, *J Cell Biol* **161**, 333 (2003).
S59. M. Wu, B. Repetto, D. M. Glerum, A. Tzagoloff, *Mol Cell Biol* **15**, 264 (1995).

Figure S1

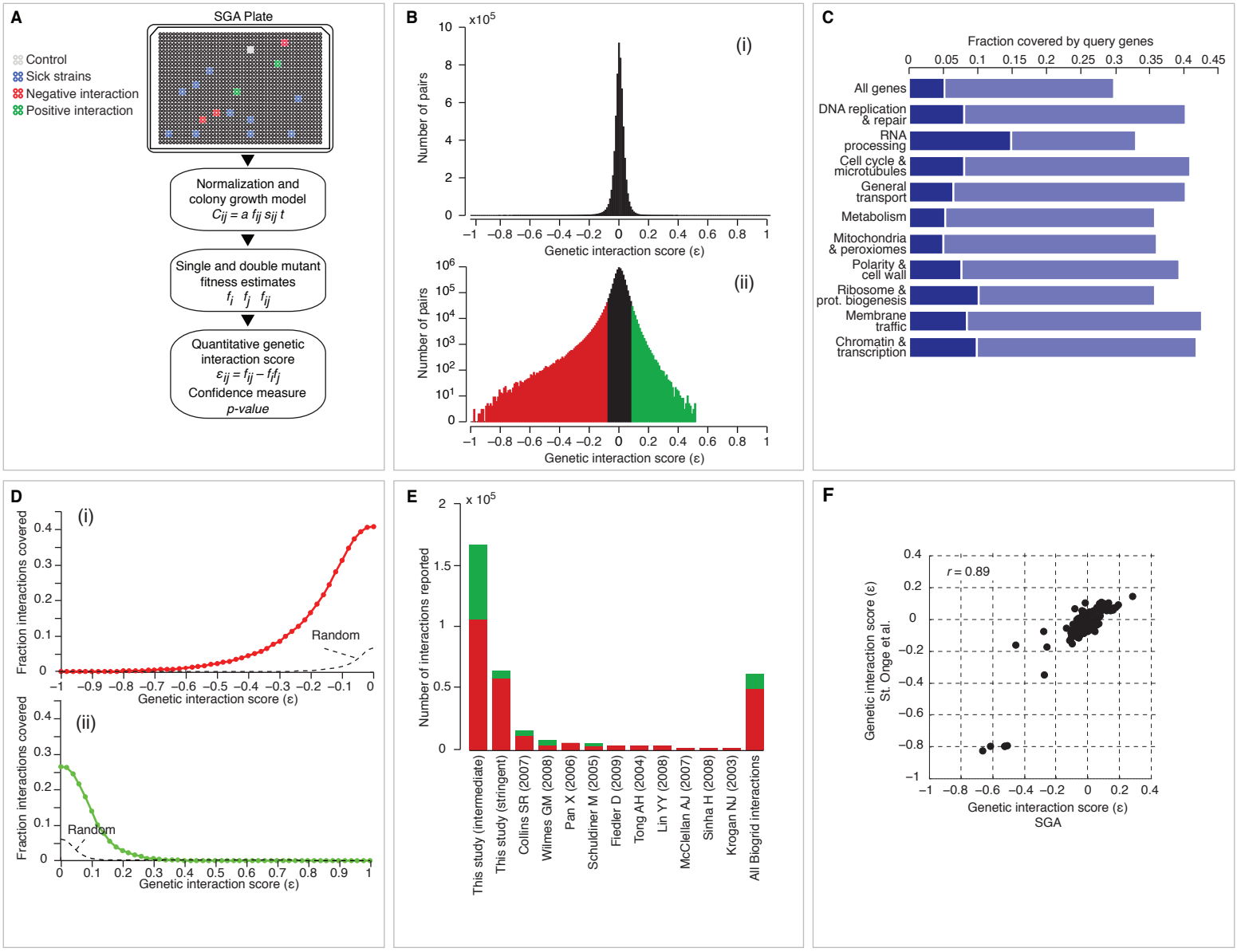

A
SGA Plate
Control
Sick strains
Negative interaction
Positive interaction
Normalization and colony growth model
$C_{ij} = a\, f_{ij}\, s_{ij}\, t$
Single and double mutant fitness estimates
$f_i \quad f_j \quad f_{ij}$
Quantitative genetic interaction score
$\varepsilon_{ij} = f_{ij} - f_i f_j$
Confidence measure
p-value
B
x10^5
(i)
Number of pairs
Genetic interaction score (ε)
(ii)
Number of pairs
Genetic interaction score (ε)
C
Fraction covered by query genes
All genes
DNA replication & repair
RNA processing
Cell cycle & microtubules
General transport
Metabolism
Mitochondria & peroxiomes
Polarity & cell wall
Ribosome & prot. biogenesis
Membrane traffic
Chromatin & transcription
D
(i)
Fraction interactions covered
Random
Genetic interaction score (ε)
(ii)
Fraction interactions covered
Random
Genetic interaction score (ε)
E
x 10^5
Number of interactions reported
This study (intermediate)
This study (stringent)
Collins SR (2007)
Wilmes GM (2008)
Pan X (2006)
Schuldiner M (2005)
Fiedler D (2009)
Tong AH (2004)
Lin YY (2008)
McClellan AJ (2007)
Sinha H (2008)
Krogan NJ (2003)
All Biogrid interactions
F
r = 0.89
Genetic interaction score (ε) St. Onge et al.
Genetic interaction score (ε) SGA

Figure S2

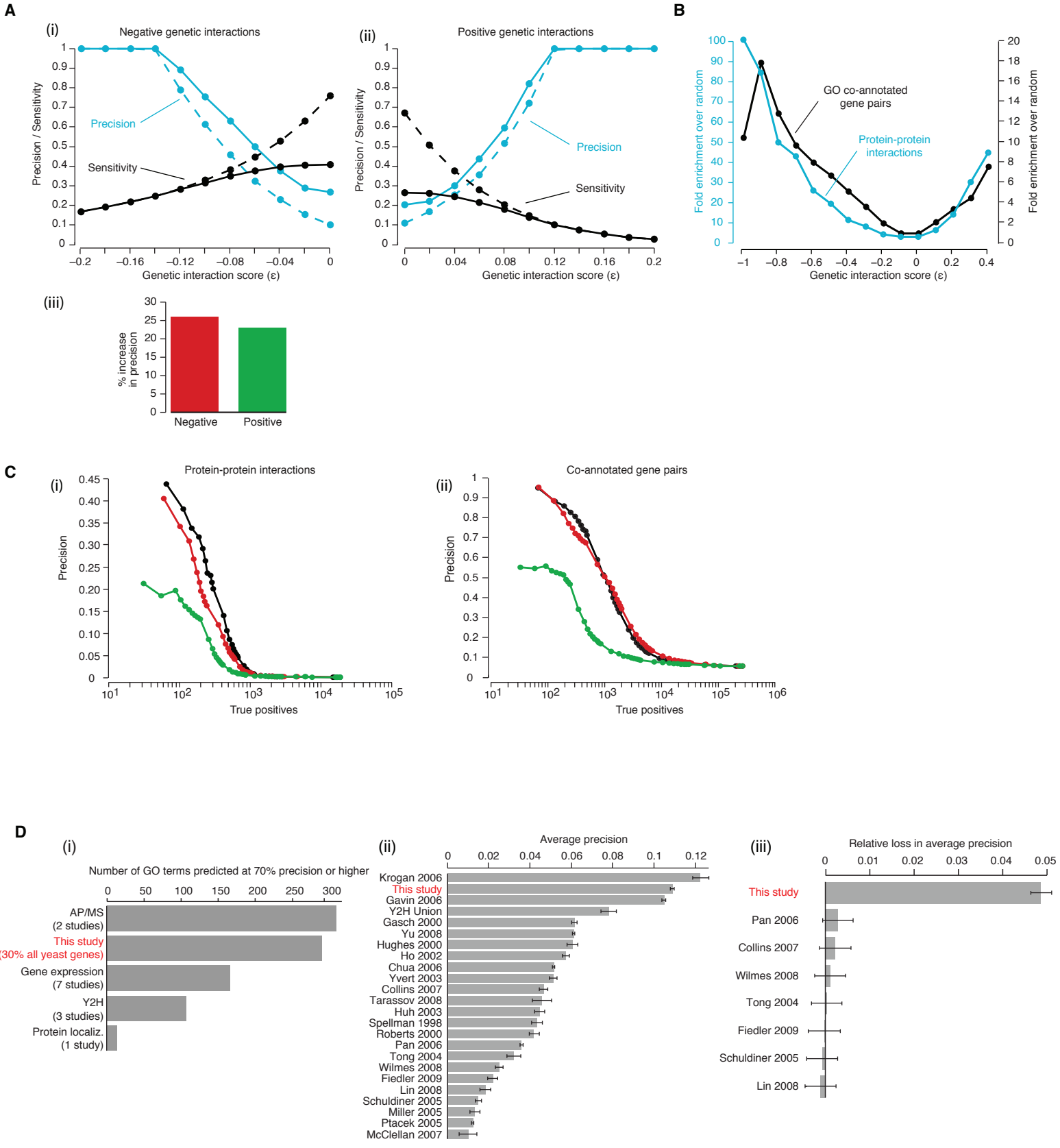

A
(i)
Negative genetic interactions
Precision / Sensitivity
Precision
Sensitivity
Genetic interaction score (ε)
(ii)
Positive genetic interactions
Precision / Sensitivity
Precision
Sensitivity
Genetic interaction score (ε)
(iii)
% increase in precision
Negative
Positive
B
Fold enrichment over random
GO co-annotated gene pairs
Protein-protein interactions
Genetic interaction score (ε)
Fold enrichment over random
C
(i)
Protein-protein interactions
Precision
True positives
(ii)
Co-annotated gene pairs
Precision
True positives
D
(i)
Number of GO terms predicted at 70% precision or higher
AP/MS (2 studies)
This study (30% all yeast genes)
Gene expression (7 studies)
Y2H (3 studies)
Protein localiz. (1 study)
(ii)
Average precision
Krogan 2006
This study
Gavin 2006
Y2H Union
Gasch 2000
Yu 2008
Hughes 2000
Ho 2002
Chua 2006
Yvert 2003
Collins 2007
Tarassov 2008
Huh 2003
Spellman 1998
Roberts 2000
Pan 2006
Tong 2004
Wilmes 2008
Fiedler 2009
Lin 2008
Schuldiner 2005
Miller 2005
Ptacek 2005
McClellan 2007
(iii)
Relative loss in average precision
This study
Pan 2006
Collins 2007
Wilmes 2008
Tong 2004
Fiedler 2009
Schuldiner 2005
Lin 2008

Figure S3

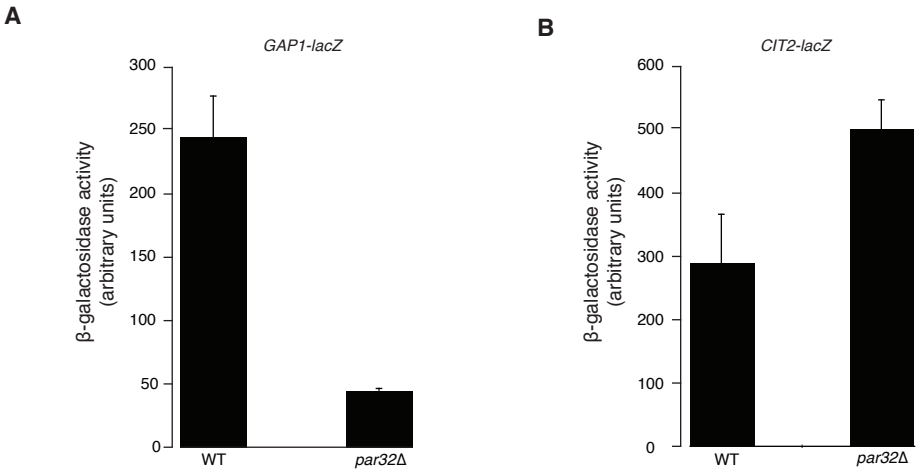
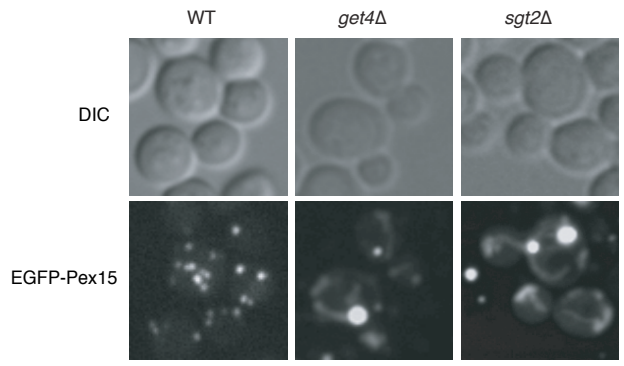

A
GAP1-lacZ
β-galactosidase activity (arbitrary units)
300
250
200
150
100
50
0
WT
par32Δ
B
CIT2-lacZ
β-galactosidase activity (arbitrary units)
600
500
400
300
200
100
0
WT
par32Δ

Figure S4

A



B

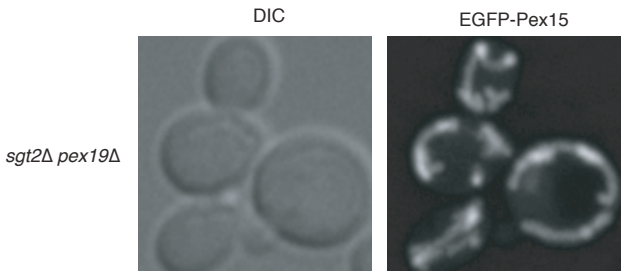


Figure S5

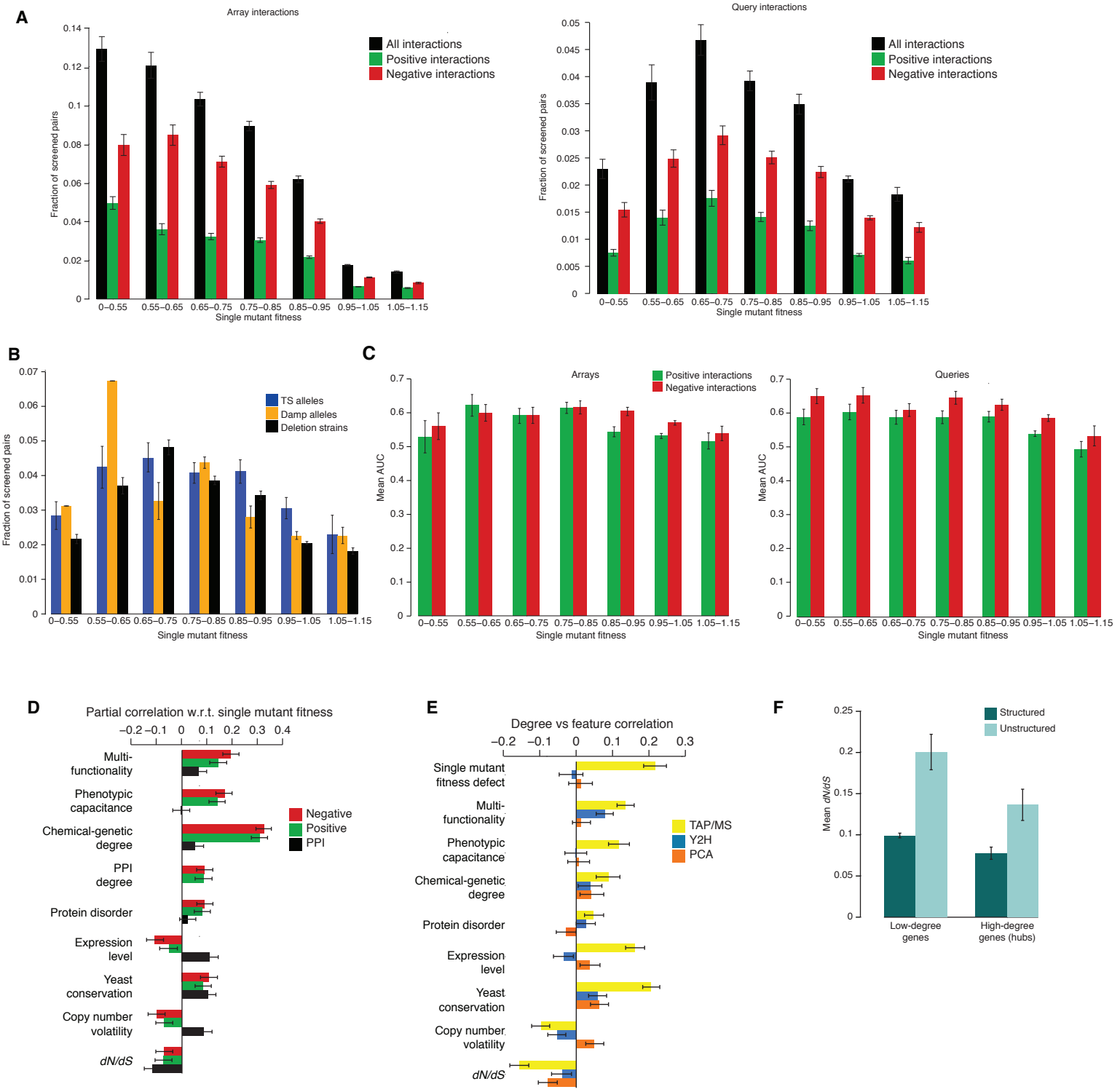

A
Array interactions
Fraction of screened pairs
All interactions
Positive interactions
Negative interactions
0–0.55
0.55–0.65
0.65–0.75
0.75–0.85
0.85–0.95
0.95–1.05
1.05–1.15
Single mutant fitness
Query interactions
B
TS alleles
Damp alleles
Deletion strains
C
Arrays
Queries
Mean AUC
D
Partial correlation w.r.t. single mutant fitness
Multi-functionality
Phenotypic capacitance
Chemical-genetic degree
PPI degree
Protein disorder
Expression level
Yeast conservation
Copy number volatility
dN/dS
E
Degree vs feature correlation
Single mutant fitness defect
TAP/MS
Y2H
PCA
F
Mean dN/dS
Structured
Unstructured
Low-degree genes
High-degree genes (hubs)

Figure S6

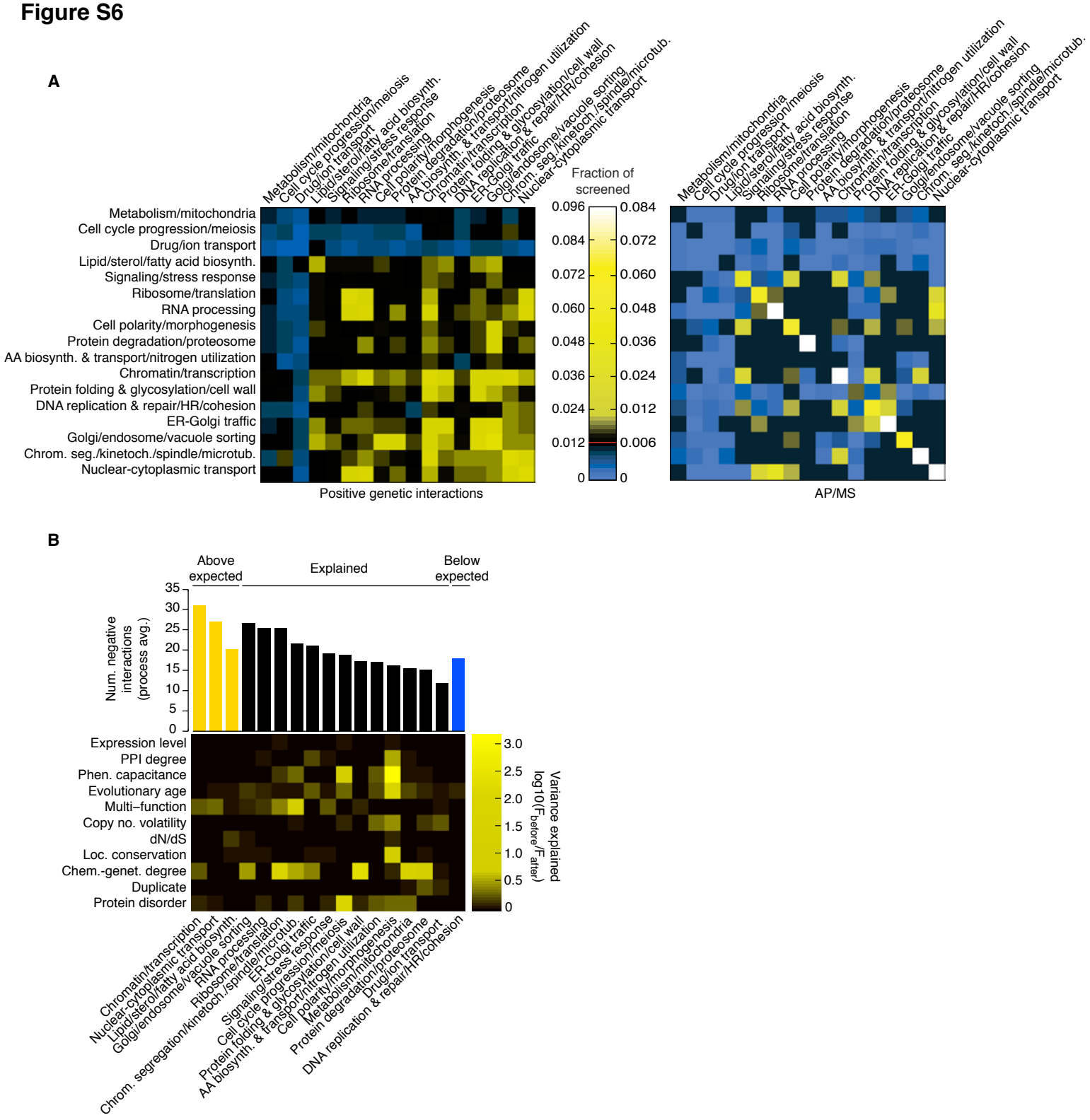

A
Metabolism/mitochondria
Cell cycle progression/meiosis
Drug/ion transport
Lipid/sterol/fatty acid biosynth.
Signaling/stress response
Ribosome/translation
RNA processing
Cell polarity/morphogenesis
Protein degradation/proteosome
AA biosynth. & transport/nitrogen utilization
Chromatin/transcription
Protein folding & glycosylation/cell wall
DNA replication & repair/HR/cohesion
ER-Golgi traffic
Golgi/endosome/vacuole sorting
Chrom. seg./kinetoch./spindle/microtub.
Nuclear-cytoplasmic transport
Fraction of screened
0.096
0.084
0.072
0.060
0.048
0.036
0.024
0.012
0
0.084
0.072
0.060
0.048
0.036
0.024
0.012
0.006
0
Positive genetic interactions
AP/MS
B
Above expected
Explained
Below expected
Num. negative interactions (process avg.)
35
30
25
20
15
10
5
0
Expression level
PPI degree
Phen. capacitance
Evolutionary age
Multi−function
Copy no. volatility
dN/dS
Loc. conservation
Chem.-genet. degree
Duplicate
Protein disorder
3.0
2.5
2.0
1.5
1.0
0.5
0
Variance explained log10(F_before/F_after)
Chromatin/transcription
Nuclear-cytoplasmic transport
Lipid/sterol/fatty acid biosynth.
Golgi/endosome/vacuole sorting
Chrom. segregation/kinetoch./spindle/microtub.
RNA processing
Ribosome/translation
ER-Golgi traffic
Signaling/stress response
Cell cycle progression/meiosis
Protein folding & glycosylation/cell wall
AA biosynth. & transport/nitrogen utilization
Cell polarity/morphogenesis
Metabolism/mitochondria
Protein degradation/proteosome
Drug/ion transport
DNA replication & repair/HR/cohesion

Figure S7

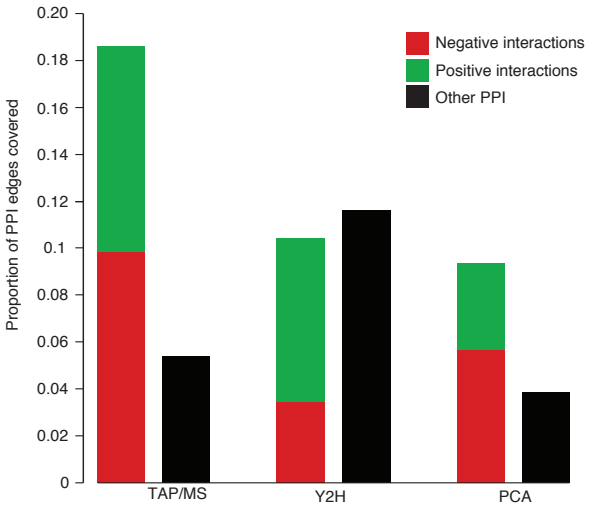
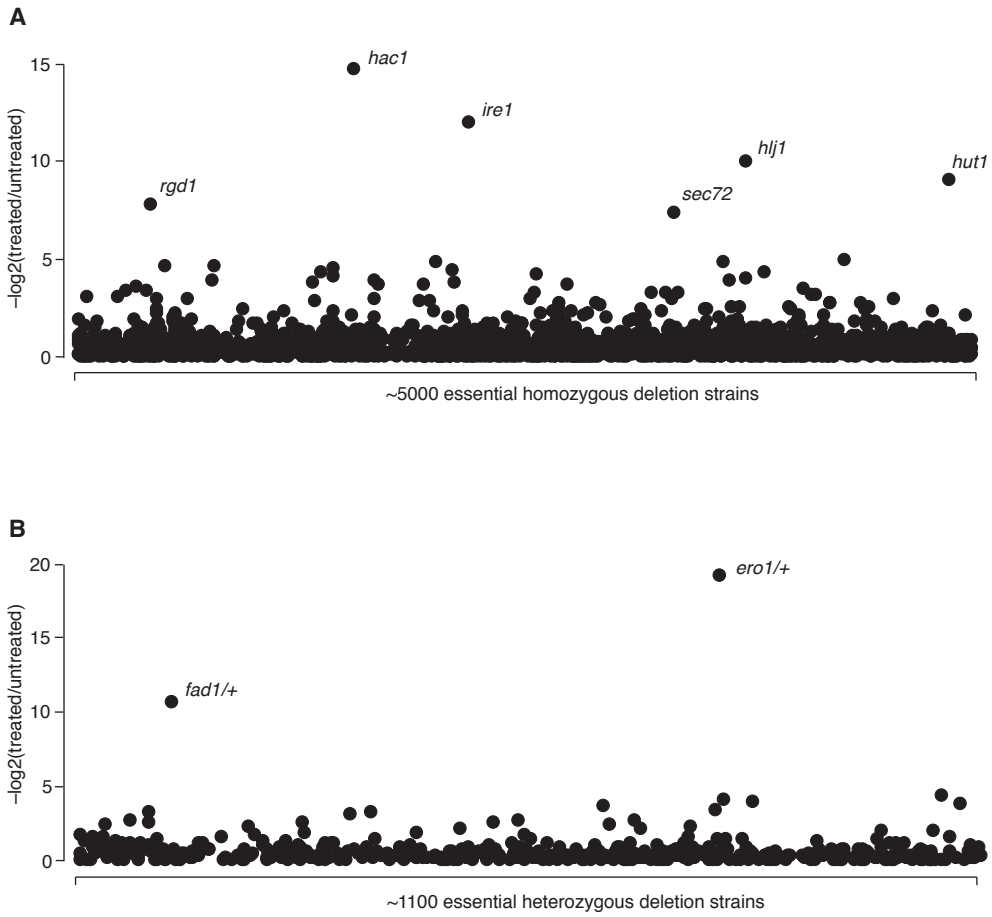

Negative interactions
Positive interactions
Other PPI
Proportion of PPI edges covered
0.20
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
TAP/MS
Y2H
PCA

Figure S8


A
15
10
5
0
–log2(treated/untreated)
hac1
ire1
hlj1
hut1
rgd1
sec72
~5000 essential homozygous deletion strains
B
20
15
10
5
0
–log2(treated/untreated)
ero1/+
fad1/+
~1100 essential heterozygous deletion strains