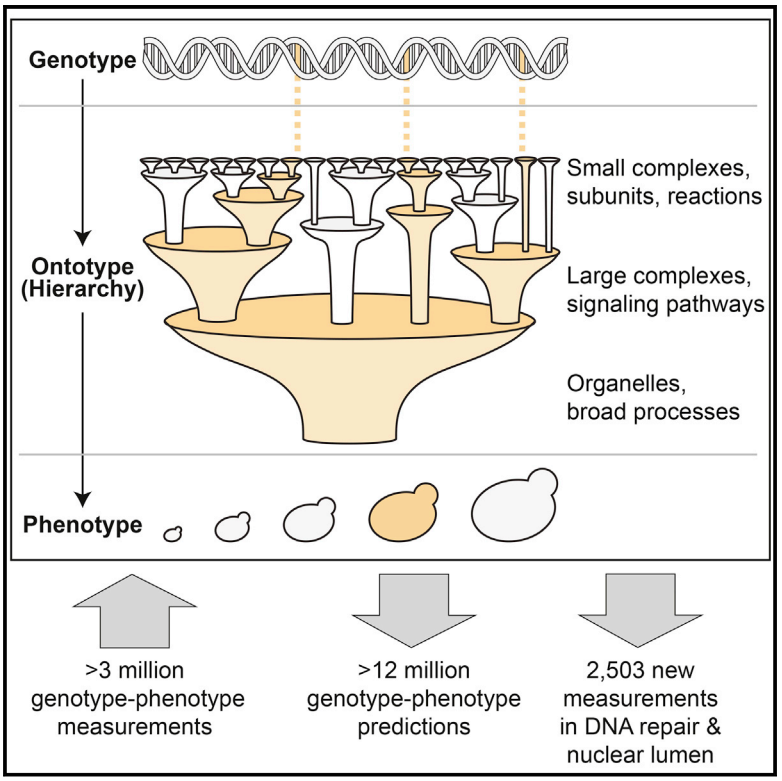
Article

Cell Systems

Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems

Graphical Abstract



Highlights

- Strategy for genotype-phenotype prediction using a deep hierarchy of cell systems
- Structure of model seeded from the GO hierarchy or directly from data
- Convergence of genetic variation up the hierarchy enabling functional prediction
- Striking ability to translate combinatorial yeast genotypes to growth rate

Authors

Michael Ku Yu, Michael Kramer, Janusz Dutkowski, ..., Nevan Krogan, Roded Sharan, Trey Ideker

Correspondence

tideker@ucsd.edu

In Brief

The Gene Ontology has become the definitive description of hierarchical cell structure. Matching this hierarchy to functional states enables an exciting new model for interpreting complex genotypes.

Yu et al., 2016, Cell Systems *2*, 77–88
February 24, 2016
http://dx.doi.org/10.1016/j.cels.2016.02.003

Cell Systems
Article

Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems

Michael Ku Yu,[1,2] Michael Kramer,[2,3] Janusz Dutkowski,[2,4] Rohith Srivas,[2,5] Katherine Licon,[2] Jason F. Kreisberg,[2] Cherie T. Ng,[6] Nevan Krogan,[7] Roded Sharan,[8] and Trey Ideker[2,*]
[1]Bioinformatics and Systems Biology Program
[2]Department of Medicine
[3]Biomedical Sciences Program
University of California, San Diego, La Jolla, CA 92093, USA
[4]Data4Cure, La Jolla, CA 92037, USA
[5]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[6]aTyr Pharmaceuticals, San Diego, CA 92121, USA
[7]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94143, USA
[8]Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel
*Correspondence: tideker@ucsd.edu
http://dx.doi.org/10.1016/j.cels.2016.02.003

SUMMARY

Accurately translating genotype to phenotype requires accounting for the functional impact of genetic variation at many biological scales. Here, we present a strategy for genotype-phenotype reasoning based on existing knowledge of cellular subsystems. These subsystems and their hierarchical organization are defined by the Gene Ontology or a complementary ontology inferred directly from previously published datasets. Guided by the ontology's hierarchical structure, we organize genotype data into an "ontotype," that is, a hierarchy of perturbations representing the effects of genetic variation at multiple cellular scales. The ontotype is then interpreted using logical rules generated by machine learning to predict phenotype. This approach substantially outperforms previous non-hierarchical methods for translating yeast genotype to cell growth phenotype, and it accurately predicts the growth outcomes of two new screens of 2,503 double gene knockouts affecting DNA repair or nuclear lumen. Ontotypes also generalize to larger knockout combinations, setting the stage for interpreting the complex genetics of disease.

INTRODUCTION

A central problem in genetics is to understand how different variations in DNA sequence, dispersed across a multitude of genes, can nonetheless elicit similar phenotypes (Waddington, 1942). In recent years, it has been repeatedly observed that different genetic drivers of a trait can be recognized by their aggregation in networks of pairwise protein or gene interactions (Califano et al., 2012; Greene et al., 2015; Hanahan and Weinberg, 2011; Kim and Przytycka, 2012; Ramanan et al., 2012; Wang et al., 2010). Rather than associating genotype with phenotype directly, variations in genotype are first mapped onto knowledge of gene networks; affected subnetworks are then statistically associated with phenotype. This approach can greatly increase our power to identify relevant associations between genotype and phenotype. This principle of "network-based" or "pathway-based" association (Califano et al., 2012) is now being applied to effectively map the genetics underlying complex phenotypes, including cancer and other common diseases (Hofree et al., 2013; Lee et al., 2011; Leiserson et al., 2014; Ng et al., 2012; Pe'er and Hacohen, 2011; Skafidas et al., 2014; Sullivan, 2012; Willsey et al., 2013).

In these studies, network knowledge is represented as a set of genes and pairwise gene interactions. In reality, however, genotype is transmitted to phenotype not only through gene-gene interactions but through a rich hierarchy of biological subsystems at multiple scales: genotypic variations in nucleotides (1 nm scale) give rise to functional changes in proteins (1–10 nm), which in turn affect protein complexes (10–100 nm), cellular processes (100 nm), organelles (1 μm), and, ultimately, phenotypic behaviors of cells (1–10 μm), tissues (100 μm to 100 mm), and complex organisms (>1 m). What has been less well studied in genotype-phenotype association is how to leverage our extensive pre-existing knowledge across these scales or how to identify the scales most relevant to a set of genetic variants (Deisboeck et al., 2011; Eissing et al., 2011; Walpole et al., 2013).

In many fields, knowledge across scales is modeled by ontologies: a factorization of prior knowledge about the world into a hierarchy of increasingly specific concepts (Brachman and Levesque, 2004). For instance, intelligent systems like Apple's Siri and IBM's Watson carry out logical reasoning using a large collection of world knowledge represented by ontologies (Carvunis and Ideker, 2014). In molecular and cellular biology, extensive knowledge of the hierarchy of subsystems in a cell has been represented by the Gene Ontology (GO), a community standard reference database that documents interrelationships among thousands of intracellular components, processes, and functions in a large hierarchy of terms (Gene Ontology Consortium, 2015). Thus far, genotype-phenotype association methods have sometimes used prior knowledge in the GO by flattening the term hierarchy to a network, in which pairwise interactions

connect genes annotated with the same GO term (Pesquita et al., 2009). This flattening, however, may discard important information about the rich hierarchy of biological systems connecting genotype to phenotype. Moreover, a hierarchical model is highly complementary, and in some ways orthogonal, to flat networks: the GO is primarily concerned with "deep" connectivity up and down a hierarchy of cellular processes spanning dozens of scales, whereas network models typically focus on horizontal flow of signaling, transcriptional, or metabolic information among genes or reactions at the same scale (Lee et al., 2010, 2011). Another advantage of the GO is that it is continuously improved by a very large community of dozens of curators and editors, who update GO from new knowledge published in thousands of peer-reviewed papers each year (Balakrishnan et al., 2013; Huntley et al., 2014). To complement this process of manual curation, recently we and others have shown that a large hierarchy of cellular systems can be systematically assembled directly from analysis of genome-wide datasets, including molecular interactions and gene expression profiles; we call this assembly NeXO (Dutkowski et al., 2013; Gligorijević et al., 2014; Kramer et al., 2014). This "data-driven" ontology closely resembles, and in some cases greatly revises and expands, the literature-curated GO.

Here we report a general approach for using deep hierarchical knowledge of the cell, represented by an ontology, to translate genotype to phenotype. This approach recursively aggregates the effects of genetic variation upward through the hierarchy; in this way, genetic variants comprising genotype are converted to effects on the cell subsystems affected by those variants. We call the set of all such effects an "ontotype," representing variation at intermediate scales between nanoscopic changes in genes and macroscopic changes in phenotype.

Here, we focus on yeast genetic interactions, in which the deletion of two or more genes results in an unexpectedly slow or fast cellular growth phenotype. Genetic interactions have previously been screened systematically using synthetic genetic arrays in yeast (Costanzo et al., 2010); these experiments comprise ~3 million different genetic backgrounds and are among the largest genotype-phenotype compendia in existence. We integrate these data with the GO to produce a multi-scale computational model, the functionalized ontology. The model accurately predicts growth phenotypes of 2,503 previously untested double-deletion genotypes, and it is also capable of predicting the phenotypes that result from larger combinations of gene disruptions. Similar predictive power is achieved by substituting the GO with NeXO, our data-driven ontology of cellular systems. In aggregate, this work suggests a strategy for building hierarchical models of the cell whose structure and function are learned completely from data.

RESULTS

Association between Genetic Interactions and Hierarchical Relations among Cellular Systems

As preparation for modeling, we identified patterns by which genetic interactions are associated with, and thus biologically explained by, the structure of gene ontologies. We observed that sets of genes assigned to the same GO term tended to be highly enriched for genetic interactions ($p < 10^{-5}$), for both positive genetic interactions (double gene disruptions with better than expected growth, e.g., epistasis) and negative genetic interactions (double gene disruptions with worse than expected growth, e.g., synthetic lethality) (Figure 1A). Such interaction enrichment within GO terms occurred over a wide range of term sizes—the number of genes annotated to a term—suggesting that genetic interactions emerge from both broad and specific cellular mechanisms at multiple scales.

Because of the hierarchical structure of the cell, genetic interactions among genes annotated to a term can potentially be re-interpreted as interactions between the genes of different terms at a lower scale in the GO. For example, the "parent" term "microtubule-associated complex" displays strong within-term interaction enrichment, which factors into strong between-term interaction enrichment across two of its "children" terms, kinesin and dynactin (Figure 1B). We found that such hierarchical relationships were widespread in the GO: approximately half of within-term enrichments could be factored into between-term enrichments among their descendants (Figure 1C). Occurrences of interactions within or between biological pathways have been previously investigated as separate biological interpretations (Bandyopadhyay et al., 2008; Bellay et al., 2011; Collins et al., 2010; Kelley and Ideker, 2005; Leiserson et al., 2011; Ma et al., 2008; Qi et al., 2008; Ulitsky et al., 2008). Here, both types of explanations can be applied to the same interaction, as they are related hierarchically within the unified structure of the cell. Overall, approximately 40,000 interactions were involved in 1,661 within- or between-term enrichments, representing a 24:1 compression of information (Figure 1D). Thus, the GO integrates genetic interactions in an overarching hierarchy capturing multiple scales of cell biology. As one moves upward in this hierarchy, separate disruptions to multiple systems converge to multiple disruptions to a single system, with the scale of this transition indicated naturally by the hierarchical structure.

The Ontotype: An Intermediate between Genotype and Phenotype

Guided by this concordance between the GO hierarchy and genetic interactions, we developed a general system for ontology-based translation of genotype to phenotype that involves three general steps. First, the genotype is described according to convention by the set of genes that have been disrupted relative to wild-type (e.g., *bΔdΔ*; Figure 2A). These disruptions are propagated recursively up the ontology, such that every term is assigned the disrupted genes annotated to that term plus all of those assigned to its children. For example, because the gene *KIP1* encodes a subunit of the kinesin complex (Figure 1B), its deletion in a *kip1Δ* strain propagates upward in the ontology to affect the parent term "kinesin complex" and continues to propagate upward to affect ancestor terms at higher scales such as "microtubule associated complex" and "cytoskeleton."

Second, every term is assigned a functional state, representing the aggregate impact of gene disruptions on the activity of the component or process that term represents. Although it is possible to envisage many ways one might compute this functional impact, as proof of principle, we explored a simple and parameter-free computation, the number of disrupted genes associated with the term. This general approach is iterated across all terms; we call the profile of states across all terms

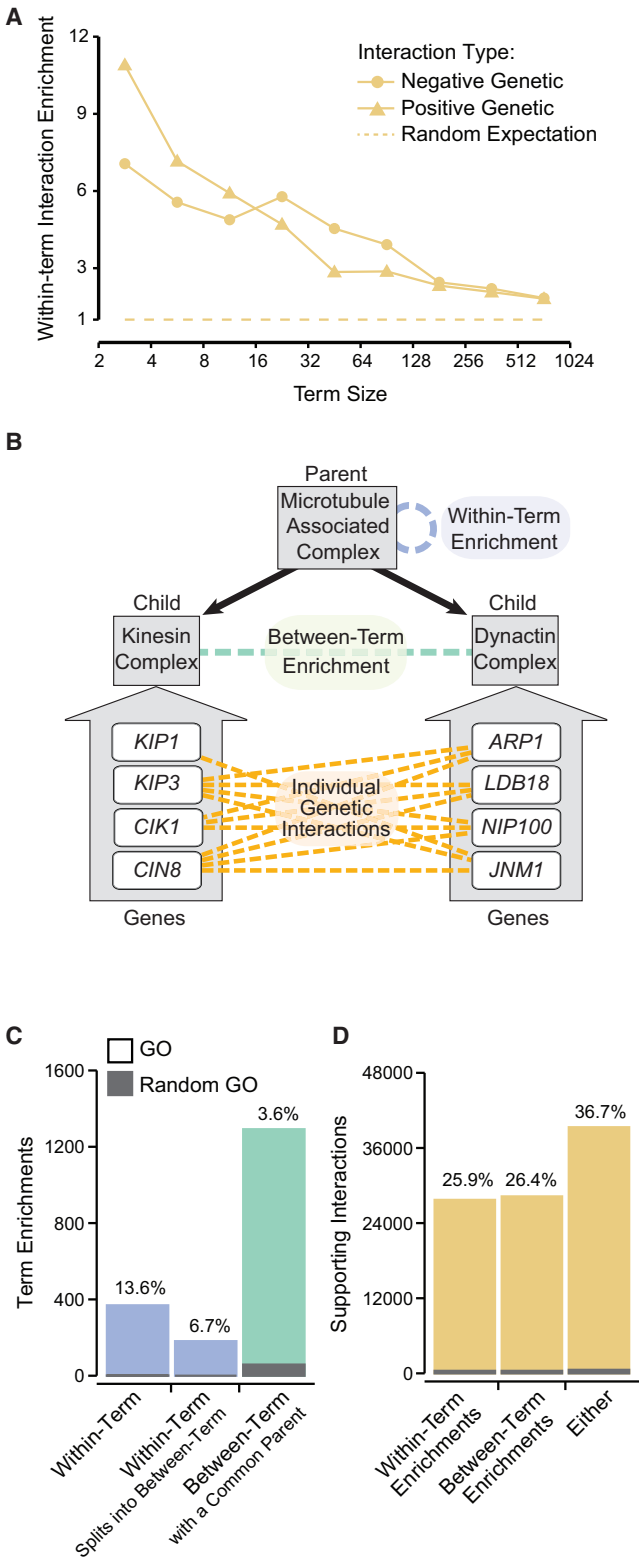


Figure 1. Patterns of Genetic Interaction Reflect the Hierarchical Structure of the GO

(A) Enrichment for negative (circle) or positive (triangle) genetic interactions among genes annotated to the same GO term as a function of term size, measured by the number of genes annotated to that term or its descendants. Enrichment is normalized as the fold change over expected for randomized GO annotations.

(B) Genetic interactions are propagated up the GO hierarchy to support "between-term enrichment" between the dynactin and kinesin complexes and "within-term enrichment" within the parent "microtubule associated complex."

(C) Number of within-term and between-term enrichments highlighted by current genetic interaction data. Approximately half of within-term enrichments can be factored into one or more between-term enrichments that occur lower in the GO hierarchy. Percentages are calculated with respect to the total possible tests for within-term (2,719) and between-term (36,210) enrichments.

(D) Number of genetic interactions involved in a within-term, between-term, or either type of enrichment. Percentages are calculated with respect to the total number of genetic interactions (107,133). The expected numbers of enrichments (C) and supporting interactions (D) were also calculated over randomized GO annotations (dark gray bars).

the "ontotype." In this way, the ontotype provides a complete picture of cell function and spans scales between genotype and phenotype. Whereas genotype describes the states of genes, and phenotype describes the states of observable traits, ontotype describes the states of all known biological objects. Many of these objects exist at scales bigger than genes but too small to be classically observable by eye, such as protein complexes and other subcellular structures, or too diffuse, such as signaling pathways (Figure 2A). In its most general definition, ontotype encompasses both genotype and phenotype, with genes and observable traits positioned at lower and higher levels of the hierarchy of objects encoding life.

A Functionalized Gene Ontology Integrating Cell Structure and Functional Prediction

Third, once genotypes are transformed to ontotypes, a supervised learning approach based on the technique of random forests regression (Breiman, 2001) is used to learn rules by which term states predict phenotypes. Rules are organized as a collection, or "forest," of decision trees (Experimental Procedures), with a typical decision tree describing a series of logical true-or-false tests to evaluate the states of several terms (e.g., T4, T5, and T7 in Figure 2A). Making decisions on the states of terms rather than nucleotide variants or genes enables machine learning across a range of scales, so that different genotypes converging on similar ontotypes (e.g., $a\Delta d\Delta$ and $b\Delta d\Delta$ in Figure 2B) can yield the same phenotype. Decision tree logic was trained to predict quantitative genetic interaction scores from ~3 million tests for pairwise genetic interactions (Costanzo et al., 2010) (Experimental Procedures). This hierarchical structure of the ontology, when coupled to the decision logic described above, forms a "functionalized" ontology, that is, a computational cell model that defines both the sub-structures of the cell and how these sub-structures hierarchically translate genotype to phenotype.

Separate functionalized ontologies were trained using either the GO curated from the *Saccharomyces* literature (Cherry et al., 2012) (F_{GO}) or a data-driven ontology assembled from *Saccharomyces* datasets using the method of network-extracted ontologies (Dutkowski et al., 2013; Kramer et al., 2014) (F_{NeXO}). Whereas the GO represents knowledge of published cell biology, application of NeXO yielded an ontology whose hierarchy of cell systems was learned directly from publicly available data,

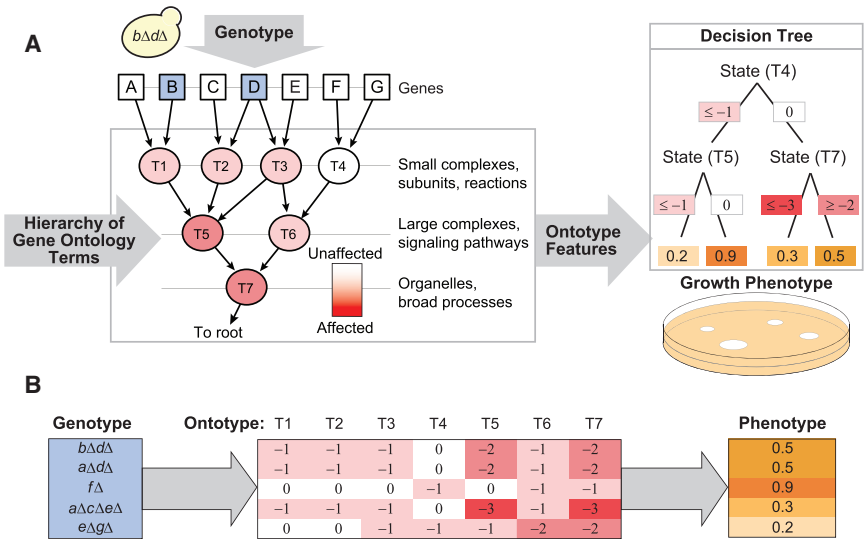


Figure 2. The Ontotype Method of Translating Genotype to Phenotype

(A) The relationship between genotypic and phenotypic variation is modeled through an intermediate "ontotype," defined as the profile of states corresponding to the effect of genotype on each cellular component, biological process, and molecular function represented as a term in GO. To generate an ontotype, perturbations to genes are propagated hierarchically through the ontology, altering term states. A random forest regresses to predict a phenotype using the ontotype as features. An example decision tree from the forest is shown.

(B) Example genotype, ontotype, and phenotype associations from the ontology in (A). Different genotypes (e.g., *b*Δ*d*Δ and *a*Δ*d*Δ) give rise to similar or identical phenotypes by influencing similar or identical combinations of terms.

including protein-protein interactions, gene expression profiles, and protein sequence properties, but excluding any prior information about genetic interactions (datasets taken from the YeastNet v3 study; Kim et al., 2014). NeXO (4,805 terms) was tuned so that the resulting ontology was approximately similar in size to the GO (5,125 terms). Alignment of these two ontologies revealed 1,614 significantly overlapping terms. Thus, NeXO represents a distinct hierarchy of cellular systems that provides an alternative to the hierarchy maintained by GO curators.

Quantitative Assessment of Performance for Genotype-Phenotype Translation

F_{GO} accurately predicted growth phenotypes across a range of genetic interaction scores (Figures 3A and 3B). The correlation between predicted and measured scores was highly significant (Figure 3C; Pearson's $r = 0.35$, $p < 2.2 \times 10^{-16}$) and reduced substantially when a randomized version of the ontology was used ($r = 0.04$); the maximum achievable correlation, as previously determined by experimental genetic interaction replicates (Baryshnikova et al., 2010), was $r = 0.67$. Progressively removing either small or large terms from the model degraded the correlation (Figures 3D and 3E), indicating that all scales in the hierarchy aid in prediction. F_{NeXO} achieved nearly the same correlation (Figure 3C; $r = 0.32$) and was also sensitive to randomization ($r = 0.03$).

Both functionalized ontologies compared favorably with non-hierarchical approaches for predicting genetic interactions (Boucher and Jenna, 2013; Lehner, 2013). We evaluated three state-of-the-art methods: flux balance analysis (FBA), which uses a mechanistic model of yeast metabolic pathways to simulate the impact of gene deletions on cell growth (Szappanos et al., 2011); guilt by association (GBA), which predicts the phenotype of pairwise gene deletions on the basis of the phenotypes of their network neighbors (Lee et al., 2010); and the multi-network multi-classifier (MNMC), a "black box" supervised learning system that uses many different lines of experimental evidence as features to predict genetic interactions (Pandey et al., 2010) (Experimental Procedures). In comparison with all of these approaches, the functionalized ontologies achieved substantially greater correlation between predicted and measured interaction scores (Figure 3C) as well as better trade-offs in precision versus recall (Figure 3F) in 4-fold cross-validation. We also assessed prediction performance in a challenging validation scenario in which the training set of genotypes does not disrupt any genes in the test set (Park and Marcotte, 2012) (Supplemental Experimental Procedures). In this scenario, any genotype-phenotype logic that applies to individual genes is no longer generalizable; for example, promiscuous genes with a high degree of genetic interactions (Gillis and Pavlidis, 2012; Mackay, 2014) could be used to explain training data but not test data. Despite this challenge, F_{GO} still outperformed predictions made with a randomized GO or with the non-hierarchical methods (Figure S1).

We found that the accuracy of growth phenotype prediction depends significantly on the degree to which cellular systems have been characterized in the GO. F_{GO} was especially accurate at modeling genotypes for which the disrupted genes are well characterized by GO annotations; conversely, it was far less able to model genotypes for which the genes are poorly characterized (Figure S2). Moreover, many genes that are poorly characterized in the GO are better characterized in NeXO, such that genotypes involving these genes lead to better phenotypic predictions by F_{NeXO} than by F_{GO} (Figures S2A–S2C). These differences demonstrate the utility of data-driven ontologies for translating genotype to phenotype, especially in species that are lacking in GO curation but have "omics" datasets from which a gene ontology can nonetheless be built.

Finally, we investigated whether hierarchical features (i.e., the ontotype) were essential or if equally good predictions could be made from "flat" features derived from the same ontologies. The GO was flattened by computing the semantic similarity (Resnik, 1995), which scores every pair of genes by their functional relatedness in GO. As a non-hierarchical representation of NeXO, we directly considered the data on which it had been based: pairwise gene-gene similarities derived from different types of experimental evidence in YeastNet. Use of these flat datasets derived

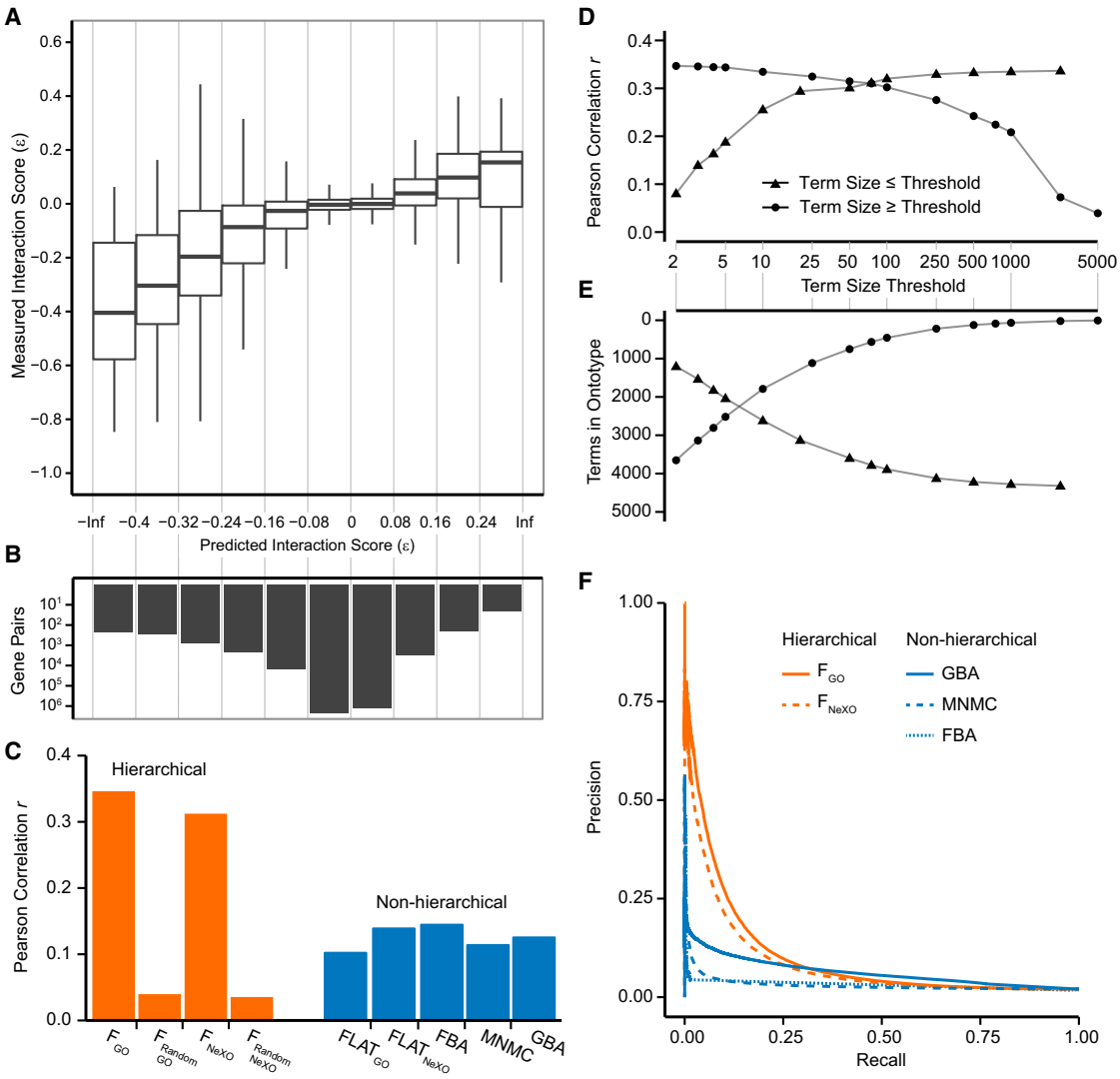


Figure 3. Genome-wide Prediction of Pairwise Genetic Interactions in Yeast

(A) Measured genetic interaction scores versus those predicted from ontotypes constructed from GO using 4-fold cross-validation. For each bin of predicted scores, box plots summarize the distribution of measured scores by its median (central horizontal line), interquartile range (box), and an additional 1.5σ (whiskers).

(B) Number of gene pairs in each bin of predicted scores.

(C) Method performance, as represented by the correlation of measured versus predicted interaction scores across gene pairs that meet an interaction significance criterion of $p < 0.05$ in Costanzo et al. (2010). Comparison is made with ontotypes constructed from a randomized gene ontology or NeXO and to previous non-hierarchical methods for predicting genetic interactions. FBA correlation is reported for the set of 104,826 gene pairs considered by this model and for which gene annotations are available in the GO. The ontotype correlations do not fluctuate greatly (<4%) whether computed over all gene pairs (shown) or the FBA gene pairs. See also Figures S1 and S2.

(D) Method performance when the ontotype is constructed from only GO terms that are no larger than (triangles) or no smaller than (circles) a size threshold.

(E) The number of GO terms that meet each size threshold criterion.

(F) Precision-recall curves for classification of negative genetic interactions.

from the two ontologies resulted in a substantial degradation in prediction performance ($FLAT_{GO}$ and $FLAT_{NeXO}$; Figure 3C), even though the same random-forests regression procedure was used as for the functionalized ontologies.

Simulating Growth Phenotypes for "New" Genotypes Not Yet Observed or Examined

We next used F_{GO} to simulate growth for all 12,512,503 pairwise deletions of non-essential yeast genes, 73% of which had not yet been tested in the laboratory (Figure 4A; Data S1). A total of 41,605 genetic interactions were predicted. These predictions were concentrated within and between particular terms and term pairs (Figures 4A and 4B), covering a total of 1,367 unique terms and indicating where in the ontology the logic of F_{GO} takes place. For example, F_{GO} predicted many genetic interactions within "oxidative phosphorylation" (Figure 4C), with negative interactions linking the sub-systems of electron and proton transport and positive

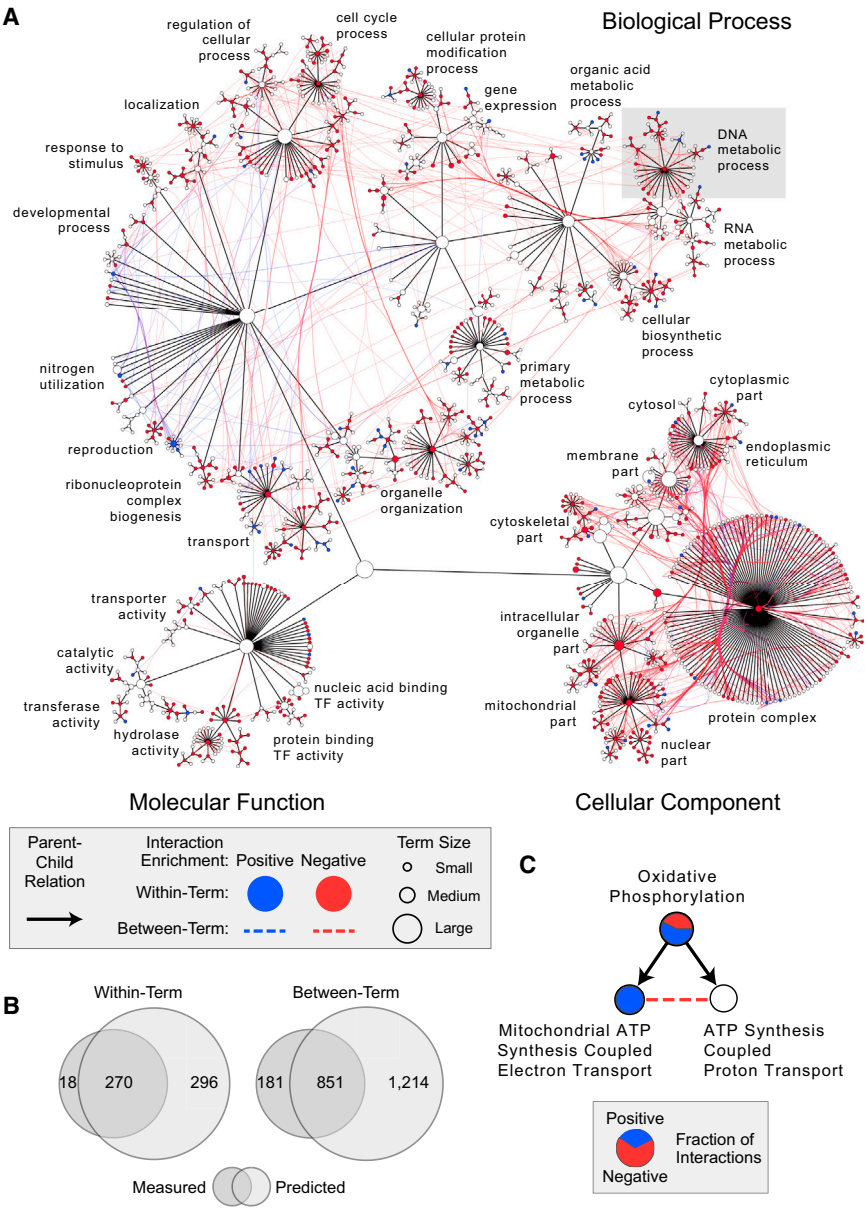


Figure 4. The Functionalized GO

(A) Visualization of F_{GO} structure and function. Terms and hierarchical parent-child relations are represented by nodes and black edges. Colored nodes and edges denote within- and between-term interaction enrichments, illustrating how terms and term combinations are used for prediction.

(B) Venn diagrams showing number of term enrichments identified for measured interactions, predicted interactions, or both.

(C) Example term "oxidative phosphorylation," which factors into the transport of electrons (left child) versus protons (right child). Although both positive and negative genetic interactions are predicted within the oxidative phosphorylation genes (represented by a pie with both blue and red slices), positive interactions segregate within electron transport (blue pie) while negative interactions segregate between electron and proton transport (dotted red edge).

See also Figure S3.

interactions segregating entirely within electron transport. These distinct patterns of positive and negative segregation were observed broadly across F_{GO} (Figure S3). Of particular interest were predicted interactions between 71 term pairs, as these terms were only distantly related in the GO (Table 1; Table S1; Supplemental Experimental Procedures). For example, all ten genes in "intron homing" had negative interactions with all four genes in the "Phosphatidylinositol-3-kinase complex," although neither these terms nor their parents shared any genes, and these terms were in entirely separate branches of the GO (biological process versus cellular component). Thus, F_{GO} makes predictions guided by, but not rigidly confined to, known hierarchical relations among cellular subsystems. The unexpected connections point to potential new cellular functions and functional relationships important for regulating cell growth.

Validation and Expansion of the Functionalized Ontology of DNA Repair and Nuclear Lumen

Key terms in F_{GO} were "DNA repair" and "nuclear lumen," which featured prominently in the decision tree logic, leading to a high concentration of predicted interactions (9.0 and 7.6 times the expected interaction density, respectively) according to particular patterns of disruption (Figures 5A and S4). Genetic perturbations within each term led to particularly accurate growth phenotypes in cross-validation, as the correlation between predicted interactions and those measured by Costanzo et al. (2010) was noticeably better for gene pairs in DNA repair or nuclear lumen (both r = 0.61) than for gene pairs in other terms (average r = 0.35; Figure S2G; Table S2). To test whether this performance generalized to new data, we experimentally measured growth phenotypes for 1,218 pairwise deletions of DNA repair genes and 1,600 pairwise deletions of nuclear lumen genes and scored these mutants for genetic interactions (Table S3; Supplemental Experimental Procedures). Of these, 1,345 mutants had also been scored previously by Costanzo et al. (2010). Surprisingly, we observed that the new measurements were better predicted by F_{GO} than by the previous measurements of those same genotypes (i.e., experimental replicates; Figure 5B). Such improvement suggests that functionalized ontologies may be able to reduce experimental noise by learning the overarching patterns of cellular subsystems that translate genotype to phenotype.

We next tested F_{GO}'s ability to generalize to unseen mutant genotypes. For this purpose we constructed a "limited" F_{GO}, trained only on those genotypes that had been tested earlier

Table 1. Top New Functional Relationships in F_{GO}

	Term A (Number of Genes)	Term B (Number of Genes)	Interactions/Total (%) Predicted	Measured	p Value[a]
Negative Interactions	intron homing (10)	phosphatidylinositol 3-kinase complex II (4)	40/40 (100.0%)	2/2 (100.0%)	6.74E-96
	negative regulation of chromatin silencing at silent mating-type cassette (8)	protein import into mitochondrial inner membrane (3)	24/24 (100.0%)	14/14 (100.0%)	3.56E-55
	pre-mRNA binding (5)	RNA pol II transcription coactivator activity in preinitiation complex assembly (3)	15/15 (100.0%)	2/2 (100.0%)	2.86E-32
	protein lipoylation (4)	carbon-oxygen lyase activity, acting on phosphates (3)	12/12 (100.0%)	9/9 (100.0%)	1.23E-24
	Swr1 complex (8)	U6 snRNP (3)	22/24 (91.7%)	6/6 (100.0%)	1.20E-47
	alpha-1,6-mannosyltransferase complex (6)	negative regulation of chromatin silencing involved in replicative cell aging (4)	21/24 (87.5%)	1/4 (25.0%)	3.08E-44
	tubulin complex assembly (5)	maintenance of DNA trinucleotide repeats (3)	13/15 (86.7%)	12/12 (100.0%)	3.67E-25
	inositol phosphate biosynthetic process (5)	minus-end-directed microtubule motor activity (3)	12/15 (80.0%)	5/6 (83.3%)	5.56E-22
	regulation of ARF GTPase activity (6)	phosphatidylinositol-3,5-bisphosphate 5-phosphatase activity (4)	19/24 (79.2%)	2/4 (50.0%)	7.92E-38
	positive regulation of RNA elongation from Pol I promoter (5)	HIR complex (4)	14/20 (70.0%)	8/10 (80.0%)	3.82E-25
Positive Interactions	negative regulation of chromatin silencing at silent mating-type cassette (8)	U6 snRNP (3)	19/24 (79.2%)	10/11 (90.9%)	7.92E-38
	tubulin complex assembly (5)	DNA-directed RNA polymerase I complex (4)	15/20 (75.0%)	8/8 (100.0%)	4.37E-28
	RSC complex (8)	inactivation of MAPK activity (4)	19/32 (59.4%)	3/4 (75.0%)	6.33E-34
	vacuolar proton-transporting V-type ATPase, V1 domain (8)	free ubiquitin chain polymerization (3)	14/24 (58.3%)	7/14 (50.0%)	1.91E-23
	alpha-1,6-mannosyltransferase complex (6)	dynactin complex (5)	16/30 (53.3%)	8/12 (66.7%)	1.14E-26
	vacuolar proton-transporting V-type ATPase, V0 domain (7)	AP-3 adaptor complex (4)	13/28 (46.4%)	7/16 (43.8%)	1.26E-19
	SLIK (SAGA-like) complex (14)	positive regulation of stress-activated MAPK cascade (3)	14/42 (33.3%)	12/21 (57.1%)	4.92E-19
	histone exchange (9)	minus-end-directed microtubule motor activity (3)	9/27 (33.3%)	7/22 (31.8%)	2.38E-10
	histone methyltransferase activity (H3-K4 specific) (7)	snoRNA transcription from an RNA polymerase II promoter (3)	7/21 (33.3%)	4/6 (66.7%)	7.33E-07
	glycerol transport (4)	transcription-coupled nucleotide-excision repair (4)	5/16 (31.3%)	1/4 (25.0%)	3.42E-03

See also Table S1.

[a]Calculated with respect to predicted interactions. Bonferroni corrected for family wise error rate.

(Costanzo et al., 2010) but not by our new screens. This limited F_{GO} achieved high sensitivity versus specificity (Figure 5C) and precision versus recall (Figure 5D) in predicting the new interactions measured for DNA repair and nuclear lumen genes. Given this validation, we combined the genetic interaction scores from both new screens with previous data (Costanzo et al., 2010) and re-trained the ontotype decision logic on this more complete dataset. The structure of this improved F_{GO}, with the accompanying ontotype-phenotype logic, is available online on the Network Data Exchange (http://goo.gl/cYIXWJ; UUID: 01b46d52-c3a5-11e5-8fbc-06603eb7f303; Pratt et al., 2015) and as a Cytoscape file in Data S2.

Toward More Complex Genotypes

Although the ontotype had been trained using double-deletion genotypes, we hypothesized that once trained, it might be capable of predictions for genotypes involving mutations to larger numbers of genes. Although few studies have examined three-way or higher order genetic interactions, a recent study (Haber et al., 2013) showed proof of principle for a three-way gene deletion methodology, representing one of the few systematic screens for triple mutants to date. This work reported that deletion of *CAC1* in combination with any gene in the HIR complex (*HIR1*, *HIR2*, *HIR3*, and *HPC2*) results in a synthetic growth defect (negative genetic interaction); however, the

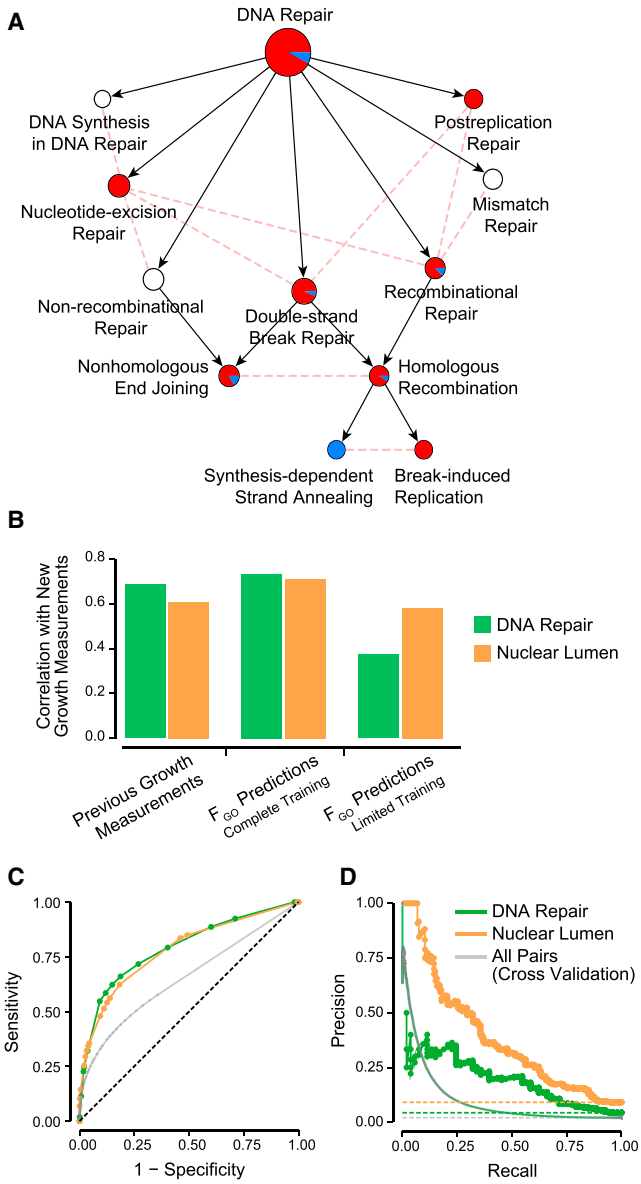


Figure 5. Elucidating the Genetic Logic of DNA Repair and the Nuclear Lumen

(A) DNA repair has a rich structure of predicted genetic interactions among specific repair processes. Coloring and visual style of panels follow the convention of Figure 4. See also Figure S4.

(B–D) Yeast growth was experimentally measured for double gene deletion strains in which both genes are involved in either DNA repair (green) or nuclear lumen (orange). See also Tables S2 and S3. (B) The new measurements are correlated with previous data by Costanzo et al. (2010) as well as predictions of a F_{GO} trained with all previous data or predictions of a "limited" F_{GO} trained with all previous data excluding genotypes tested in the new screen. In all cases, correlation is computed among the genotypes tested by both the new screen and Costanzo et al. (2010). Among all genotypes in the new screen, we calculated receiver-operating (C) and precision-recall curves (D) for predicting negative genetic interactions in DNA repair and the nuclear lumen using the limited F_{GO}. The corresponding curves across all gene pairs in the previous screen are reproduced for comparison (gray; see Figure 3F).

additional deletion of a third gene, *ASF1*, suppresses this phenotype. Consistent with these findings, F_{GO} predicted both the synthetic sickness of the double mutants and phenotypic suppression by the triple mutant (Figure 6A). Visual inspection of the model (Figure 6B) implicated decision logic on the basis of the functional activities of two related processes, DNA replication-independent nucleosome assembly and nucleosome organization. Deleting a single gene in DNA replication-independent nucleosome assembly leads to a state in which the deletion of another gene functioning elsewhere in nucleosome organization causes synthetic sickness. In contrast, the triple mutants include deletion of two genes in DNA replication-independent nucleosome assembly (*asf1*⊿HIRΔ), leading to a neutral phenotype. This effect probably occurs because the double mutant impairs growth to such an extent that additional perturbations have no detectable effect. Indeed, whereas *CAC1* is primarily involved in regulating DNA replication, *ASF1* and the HIR complex have been linked to other chromatin-related processes, including transcriptional elongation (Formosa et al., 2002; Schwabish and Struhl, 2006) and mRNA export (Pamblanco et al., 2014). This triple-mutant case study illustrates the complexity of logic in interpreting genetic interactions, underscoring the utility of a knowledge representation and reasoning system for unraveling such combinatorial genetic effects.

DISCUSSION

Many years of work by the Gene Ontology Consortium have established an extensive description of cell structure spanning a hierarchy of biological scales. Here, we have shown that the ontology structure can also be used functionally for interpretation of genetic variants to make phenotypic predictions. The ability to systematically map and then integrate these two aspects, structure and function, outlines a general strategy for development of computational cell models. First, a knowledge base of the cell's hierarchical structure is acquired, either through literature curation (GO) or data-driven methods (NeXO). In a second step, mathematical relations are learned by algorithms that translate how the functional states of these subsystems—the ontotype—give rise to a phenotype of interest. Thus, cell structure is determined from physical information derived from literature or systematic data, and cell function is learned from genetic data such as synthetic lethal interactions and genome-wide association studies.

Functionalized ontologies substantially outperformed previous phenotypic predictors (Figures 3C and 3F), a notable finding given the simplicity of the ontotype and its use as the sole feature set for learning. We believe this success follows from several key aspects of implementation. First and most important, the utility of hierarchical organization in genotype-phenotype translation cannot be overstated. Indeed, the functionalized ontologies also outperformed predictors based on non-hierarchical versions of the same information (Figure 3C) or truncated versions of the ontology (Figures 3D and 3E). From the perspective of the ontology, all mutations or variants in a genotype coalesce to the same cellular module, provided one looks at a high enough level (Figure 1B). A genotype may include some mutations that map to the same gene, others to

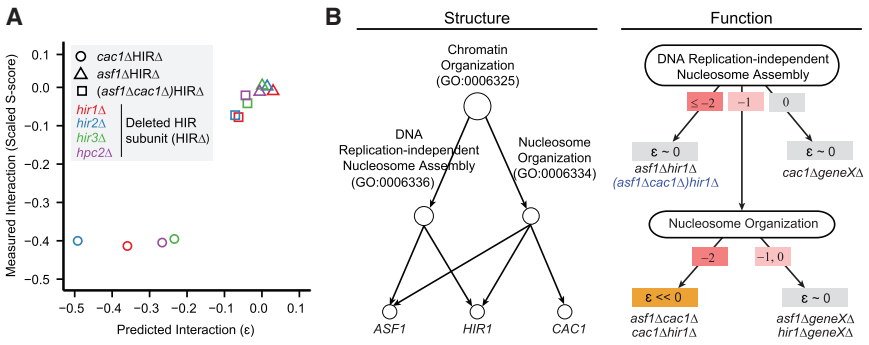


Figure 6. Prediction of Triple Mutants

(A) Measured versus predicted interaction scores for genotypes involving pairwise and three-way deletions involving *ASF1*, *CAC1*, and genes in the HIR complex (*HIR1*, *HIR2*, *HIR3*, and *HPC2*) (Haber et al., 2013).

(B) Relevant GO structure (left) and corresponding functional decision tree (right) for predicting the two- and three-way interactions in (A). At left, arrows represent parent-child relations and gene annotations in GO. At right, arrows represent decisions based on ontotype: numbers on arrows are term states; arrows point to predicted interaction scores (ε).

the same protein complex, still others to the same broad process or organelle, with all mutations falling within the highest scale represented by the cell itself. Propagating mutations upward through terms of increasing scale enables subsequent selection of the "right" scale for accurate prediction. In this regard, F_{GO} sheds light on previous, partially discrepant, studies of genetic interaction networks. Some analyses have found that negative genetic interactions tend to connect between complementary modules, whereas positive interactions tend to occur within a single module (Bandyopadhyay et al., 2008; Collins et al., 2010; Kelley and Ideker, 2005; Leiserson et al., 2011; Ma et al., 2008; Qi et al., 2008; Ulitsky et al., 2008); a more recent report identified dense patterns of both positive and negative interactions between modules (Bellay et al., 2011). Analysis of F_{GO} suggests that both interpretations can be correct, depending on the scale of the module(s) within the cellular hierarchy.

The second factor in the success of functionalized ontologies is the sustained efforts of biologists at large. The GO is a rich resource of cellular knowledge that is both broad, in its extensive coverage of cell biology, and deep, in its resolution of cell subsystems across many different scales. Although not perfect, this knowledge is continuously refined, updated, and expanded by the sustained efforts of a global community. Given the staggering complexity of the cell, such a collaborative approach incorporating diverse expertise and tools may be instrumental in establishing robust and complete prior knowledge for computational cell modeling. Previously, cellular modeling efforts have typically involved independent curation within a single laboratory or institute.

A last factor that may have worked in our favor is the fact that functionalized ontologies balance rigid modeling constraints imposed by prior knowledge with flexible statistical learning guided by experiments. Computing the ontotype requires no parameters and instead leverages the topology of the ontology. Logical rules for predicting phenotype are based on the ontotype, but their functional form (i.e., which terms are used and how their states are combined) is learned from data. In contrast, many previous efforts in mechanistic modeling (e.g., see Cahan et al., 2014; Carrera et al., 2014; Deutscher et al., 2006; Karr et al., 2012; Lerman et al., 2012; Machado et al., 2011; O'Brien et al., 2013; Orth et al., 2010; Segrè et al., 2005; Szappanos et al., 2011; Szczurek et al., 2009; Takahashi et al., 2003; Tomita et al., 1999) have been driven by low-level prior knowledge in the form of biophysical equations. Although naturally conferring a mechanistic explanation when correct, these equations have a known challenge that they are often of preset form and have sensitive parameters (Apgar et al., 2010; Ashyraliyev et al., 2009; Gutenkunst et al., 2007), such that achieving accurate predictions within one dataset risks overfitting.

Extending Functional Ontologies beyond Current Limits

F_{GO} based its predictions principally on 1,367 terms, spread across various biological processes, cellular components, and molecular functions (Figure 4A). Although this coverage of cell biology is substantial (27% of the yeast GO), one might wonder whether it should be more complete. First, some term logic is likely missed because those terms are not frequently disrupted in the current set of genotypes. For example, genes annotated to 783 GO terms were never disrupted in any genotype tested (Costanzo et al., 2010). Second, some biological processes are likely not required for the phenotype tested—growth of cells in rich media—but instead may drive a wide variety of other phenotypes (Dowell et al., 2010; Hillenmeyer et al., 2008; Ideker and Krogan, 2012; Lee et al., 2014). Third, important processes or components may not yet have been curated in the GO, and some existing terms might have errors in gene annotations or relations to other terms. Such false-positive and false-negative information could obscure a term's utility in prediction. We expect that testing additional genotypes, phenotypes, and environmental conditions will increase the functional coverage of terms and enhance F_{GO} with new and more robust logic.

Complex traits arise from a landscape of genetic variants and mutations, in which it is often challenging to interpret the effects of individual genes because of many multi-gene interactions (Kim and Przytycka, 2012; Zuk et al., 2012). Toward meeting this challenge, we have shown that gene ontologies can be transformed into multi-scale models capable of general genotype-phenotype reasoning. Although based on simple rules of propagation, the model substantially outperforms previous methods for predicting cellular growth phenotypes, whether based on mechanistic modeling of pathways or "black-box" machine learning methods. It also generalizes in ways that previous predictors are incapable of doing, including the ability to analyze genotypes of arbitrary complexity. These advances are important steps toward building intelligent systems that can one day interpret the complex genetics underlying human health and disease.

In moving forward, special consideration should be given to the mathematical functions that govern each term state. Here,

we found success with a surprisingly straightforward and parameter-free function that counts the disrupted genes assigned to a term and its sub-terms. More generally, this function might be tailored to each term according to specific knowledge about the inner workings of that cellular component or process. Defining the mathematical relationships between genes within a cellular process has been the focus of ''bottom-up'' systems biology (Bruggeman and Westerhoff, 2007; Chen et al., 2010). In contrast, defining the broad organization of genes into cellular processes has been the domain of ''top-down'' systems biology. With its hierarchy of terms and functions spanning many different biological scales, a functionalized ontology may offer a means to bridge this long-standing divide.

EXPERIMENTAL PROCEDURES

Genetic Interaction Data

Experimental genetic interaction scores for >6 million double mutants in yeast, measured using synthetic genetic arrays (SGAs) (Costanzo et al., 2010) (1,711 queries × 3,885 arrays), were downloaded from http://drygin.ccbr.utoronto.ca/~costanzo2009/. Double gene deletion mutants affecting DNA repair and the nuclear lumen were generated on solid agar media using SGA technology as previously described (Collins et al., 2010; Tong and Boone, 2006). See also Supplemental Experimental Procedures.

Preparation of Ontologies

We used all three branches of the GO (biological process, cellular component, and molecular function) by joining them under an artificial root. We removed annotations with the evidence code ''inferred by genetic interaction'' (IGI) to avoid potential circularity in predicting genetic interactions. We also removed terms that were not annotated with any yeast genes or were redundant with respect to their children terms to construct a gene ontology relevant to yeast (Table S4), following a previously described procedure (Dutkowski et al., 2013; http://mhk7.github.io/alignOntology/).

To construct NeXO (Table S5), we integrated the YeastNet v3 networks (Kim et al., 2014), spanning 68 experimental studies across eight data types excluding genetic interactions, into a single network, and then applied the method of Clique Extracted Ontology (Kramer et al., 2014; http://mhk7.github.io/clixo_0.3/). Code for constructing ontotypes is available at https://github.com/michaelkyu/ontotype. See also Supplemental Experimental Procedures.

Random Forests Regression

Random forests (Breiman, 2001) were used to regress genetic interaction scores ε_{ab}, as described in Results. Because of the very large size of the ontotype feature matrix, we optimized the random forests library from the Python scikit-learn package (Pedregosa et al., 2011); the modified code is available at https://github.com/michaelkyu/scikit-learn-fasterRF. Although trees grown at approximately 29% (GO) or 37% (NeXO) of the maximal depth did improve performance slightly (<0.02 gain in correlation; Figure S5), we chose to grow trees to maximal depth because it is unclear how significant this gain is and whether it would reproducible in different random partitions of the data for cross-validation or in different genotype-phenotype datasets. See also Supplemental Experimental Procedures.

Comparison of Methods for Predicting Genetic Interactions

We updated the MNMC method because the original implementation (Pandey et al., 2010) was trained on a set of literature-curated synthetic lethal interactions that was much smaller in size than the set of genetic interactions considered in our study, and because the set of features used by the method to score each gene pair had been updated since the 2010 publication. To train MNMC, we calculated five base features that were identified in the original MNMC as among the most informative for predicting synthetic lethality of a gene pair. This updated MNMC outperformed the original MNMC (Figure S6); this performance difference may have been due to calculating more recent versions of the base features. See also Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, five tables, and six data files and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.02.003.

AUTHOR CONTRIBUTIONS

M.K.Y., J.D., M.K., R. Sharan, and T.I. designed the study and developed the conceptual ideas. M.K. constructed NeXO. M.K.Y. implemented all other computational methods and analysis. M.K.Y. and T.I. wrote the manuscript with input from the other authors. R. Srivas and K.L. performed the DNA repair and nuclear lumen interaction screens.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful discussion and comments from Hannes Braberg, Anne-Ruxandra Carvunis, Manolis Kellis, Benjamin Kellman, Jianzhu Ma, Jenhan Tao, Alex Thomas, members of the Ideker laboratory, and the anonymous referees. This work was funded by the National Institute of General Medical Sciences (P41-GM103504, P50-GM085764) and the National Institute of Environmental Health Sciences (R01-ES014811). M.Y. received first-year support from the University of California, San Diego, Graduate Training Program in Bioinformatics (T32-GM008806). R. Sharan was supported by a research grant from the Israel Science Foundation (grant 241/11). M.K. was supported by the National Human Genome Research Institute (F30-HG007618) and the University of California, San Diego, Medical Scientist Training Program (T32-GM007198). R. Srivas is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2187-14). J.D. is a co-founder, CEO, and shareholder of Data4Cure, Inc.

Received: September 18, 2015
Revised: December 17, 2015
Accepted: February 1, 2016
Published: February 24, 2016

REFERENCES

Apgar, J.F., Witmer, D.K., White, F.M., and Tidor, B. (2010). Sloppy models, parameter uncertainty, and the role of experimental design. Mol. Biosyst. *6*, 1890–1900.

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J.A., and Blom, J.G. (2009). Systems biology: parameter estimation for biochemical models. FEBS J. *276*, 886–902.

Balakrishnan, R., Harris, M.A., Huntley, R., Van Auken, K., and Cherry, J.M. (2013). A guide to best practices for Gene Ontology (GO) manual annotation. Database (Oxford) *2013*, bat054.

Bandyopadhyay, S., Kelley, R., Krogan, N.J., and Ideker, T. (2008). Functional maps of protein complexes from quantitative genetic interaction data. PLoS Comput. Biol. *4*, e1000065.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat. Methods *7*, 1017–1024.

Bellay, J., Atluri, G., Sing, T.L., Toufighi, K., Costanzo, M., Ribeiro, P.S.M., Pandey, G., Baller, J., VanderSluis, B., Michaut, M., et al. (2011). Putting genetic interactions in context through a global modular decomposition. Genome Res. *21*, 1375–1387.

Boucher, B., and Jenna, S. (2013). Genetic interaction networks: better understand to better predict. Front. Genet. *4*, 290.

Brachman, R.J., and Levesque, H.J. (2004). Knowledge Representation and Reasoning (San Francisco: Morgan Kaufmann).

Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32.

Bruggeman, F.J., and Westerhoff, H.V. (2007). The nature of systems biology. Trends Microbiol. *15*, 45–50.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat. Genet. *44*, 841–847.

Carrera, J., Estrela, R., Luo, J., Rai, N., Tsoukalas, A., and Tagkopoulos, I. (2014). An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of Escherichia coli. Mol. Syst. Biol. *10*, 735.

Carvunis, A.-R., and Ideker, T. (2014). Siri of the cell: what biology could learn from the iPhone. Cell *157*, 534–538.

Chen, W.W., Niepel, M., and Sorger, P.K. (2010). Classic and contemporary approaches to modeling biochemical reactions. Genes Dev. *24*, 1861–1875.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700–D705.

Collins, S.R., Roguev, A., and Krogan, N.J. (2010). Quantitative genetic interaction mapping using the E-MAP approach. Methods Enzymol. *470*, 205–231.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Deisboeck, T.S., Wang, Z., Macklin, P., and Cristini, V. (2011). Multiscale cancer modeling. Annu. Rev. Biomed. Eng. *13*, 127–155.

Deutscher, D., Meilijson, I., Kupiec, M., and Ruppin, E. (2006). Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat. Genet. *38*, 993–998.

Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., et al. (2010). Genotype to phenotype: a complex problem. Science *328*, 469.

Dutkowski, J., Kramer, M., Surma, M.A., Balakrishnan, R., Cherry, J.M., Krogan, N.J., and Ideker, T. (2013). A gene ontology inferred from molecular networks. Nat. Biotechnol. *31*, 38–45.

Eissing, T., Kuepfer, L., Becker, C., Block, M., Coboeken, K., Gaub, T., Goerlitz, L., Jaeger, J., Loosen, R., Ludewig, B., et al. (2011). A computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and molecular reaction networks. Front. Physiol. *2*, 4.

Formosa, T., Ruone, S., Adams, M.D., Olsen, A.E., Eriksson, P., Yu, Y., Rhoades, A.R., Kaufman, P.D., and Stillman, D.J. (2002). Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. Genetics *162*, 1557–1571.

Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056.

Gillis, J., and Pavlidis, P. (2012). "Guilt by association" is the exception rather than the rule in gene networks. PLoS Comput. Biol. *8*, e1002444.

Gligorijević, V., Janjić, V., and Pržulj, N. (2014). Integration of molecular network data reconstructs Gene Ontology. Bioinformatics *30*, i594–i600.

Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. *47*, 569–576.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P. (2007). Universally sloppy parameter sensitivities in systems biology models. PLoS Comput. Biol. *3*, 1871–1878.

Haber, J.E., Braberg, H., Wu, Q., Alexander, R., Haase, J., Ryan, C., Lipkin-Moore, Z., Franks-Skiba, K.E., Johnson, T., Shales, M., et al. (2013). Systematic triple-mutant analysis uncovers functional connectivity between pathways involved in chromosome regulation. Cell Rep. *3*, 2168–2178.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. Nat. Methods *10*, 1108–1115.

Huntley, R.P., Sawford, T., Martin, M.J., and O'Donovan, C. (2014). Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt. Gigascience *3*, 4.

Ideker, T., and Krogan, N.J. (2012). Differential network biology. Mol. Syst. Biol. *8*, 565.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell *150*, 389–401.

Kelley, R., and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. Nat. Biotechnol. *23*, 561–566.

Kim, Y.-A., and Przytycka, T.M. (2012). Bridging the Gap between Genotype and Phenotype via Network Approaches. Front. Genet. *3*, 227.

Kim, H., Shin, J., Kim, E., Kim, H., Hwang, S., Shim, J.E., and Lee, I. (2014). YeastNet v3: a public database of data-specific and integrated functional gene networks for Saccharomyces cerevisiae. Nucleic Acids Res. *42*, D731–D736.

Kramer, M., Dutkowski, J., Yu, M., Bafna, V., and Ideker, T. (2014). Inferring gene ontologies from pairwise similarity data. Bioinformatics *30*, i34–i42.

Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010). Predicting genetic modifier loci using functional gene networks. Genome Res. *20*, 1143–1153.

Lee, I., Blom, U.M., Wang, P.I., Shim, J.E., and Marcotte, E.M. (2011). Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res. *21*, 1109–1121.

Lee, A.Y., Onge, R.P.S., Proctor, M.J., Wallace, I.M., Nile, A.H., Spagnuolo, P.A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M.K., et al. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science *344*, 208–211.

Lehner, B. (2013). Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. *14*, 168–178.

Leiserson, M.D.M., Tatar, D., Cowen, L.J., and Hescott, B.J. (2011). Inferring mechanisms of compensation from E-MAP and SGA data using local search algorithms for max cut. J. Comput. Biol. *18*, 1399–1409.

Leiserson, M.D.M., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al. (2014). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat. Genet. *47*, 106–114.

Lerman, J.A., Hyduke, D.R., Latif, H., Portnoy, V.A., Lewis, N.E., Orth, J.D., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Zengler, K., and Palsson, B.O. (2012). In silico method for modelling metabolism and gene product expression at genome scale. Nat. Commun. *3*, 929.

Ma, X., Tarone, A.M., and Li, W. (2008). Mapping genetically compensatory pathways from synthetic lethal interactions in yeast. PLoS One *3*, e1922.

Machado, D., Costa, R.S., Rocha, M., Ferreira, E.C., Tidor, B., and Rocha, I. (2011). Modeling formalisms in Systems Biology. AMB Express *1*, 45.

Mackay, T.F.C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat. Rev. Genet. *15*, 22–33.

Ng, S., Collisson, E.A., Sokolov, A., Goldstein, T., Gonzalez-Perez, A., Lopez-Bigas, N., Benz, C., Haussler, D., and Stuart, J.M. (2012). PARADIGM-SHIFT predicts the function of mutations in multiple cancers using pathway impact analysis. Bioinformatics *28*, i640–i646.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., and Palsson, B.Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. *9*, 693.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. *28*, 245–248.

Pamblanco, M., Oliete-Calvo, P., García-Oliver, E., Luz Valero, M., Sanchez del Pino, M.M., and Rodríguez-Navarro, S. (2014). Unveiling novel interactions of histone chaperone Asf1 linked to TREX-2 factors Sus1 and Thp1. Nucleus *5*, 247–259.

Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010). An integrative multi-network and multi-classifier approach to predict genetic interactions. PLoS Comput. Biol. *6*, 6.

Park, Y., and Marcotte, E.M. (2012). Flaws in evaluation schemes for pair-input computational predictions. Nat. Methods *9*, 1134–1136.

Pe'er, D., and Hacohen, N. (2011). Principles and strategies for developing network models in cancer. Cell *144*, 864–873.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Pesquita, C., Faria, D., Falcão, A.O., Lord, P., and Couto, F.M. (2009). Semantic similarity in biomedical ontologies. PLoS Comput. Biol. *5*, e1000443.

Pratt, D., Chen, J., Welker, D., Rivas, R., Pillich, R., Rynkov, V., Ono, K., Miello, C., Hicks, L., Szalma, S., et al. (2015). NDEx, the Network Data Exchange. Cell Syst. *1*, 302–305.

Qi, Y., Suhail, Y., Lin, Y.Y., Boeke, J.D., and Bader, J.S. (2008). Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. Genome Res. *18*, 1991–2004.

Ramanan, V.K., Shen, L., Moore, J.H., and Saykin, A.J. (2012). Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet. *28*, 323–332.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In IJCAI '95: Proceedings of the 14th International Joint Conference on Artificial Intelligence (San Francisco: Morgan Kaufmann), pp. 448–453.

Schwabish, M.A., and Struhl, K. (2006). Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol. Cell *22*, 415–422.

Segrè, D., Deluna, A., Church, G.M., and Kishony, R. (2005). Modular epistasis in yeast metabolism. Nat. Genet. *37*, 77–83.

Skafidas, E., Testa, R., Zantomio, D., Chana, G., Everall, I.P., and Pantelis, C. (2014). Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. Mol. Psychiatry *19*, 504–510.

Sullivan, P.F. (2012). Puzzling over schizophrenia: schizophrenia as a pathway disease. Nat. Med. *18*, 210–211.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat. Genet. *43*, 656–662.

Szczurek, E., Gat-Viks, I., Tiuryn, J., and Vingron, M. (2009). Elucidating regulatory mechanisms downstream of a signaling pathway using informative experiments. Mol. Syst. Biol. *5*, 287.

Takahashi, K., Ishikawa, N., Sadamoto, Y., Sasamoto, H., Ohta, S., Shiozawa, A., Miyoshi, F., Naito, Y., Nakayama, Y., and Tomita, M. (2003). E-Cell 2: multi-platform E-Cell simulation system. Bioinformatics *19*, 1727–1729.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C.A., 3rd (1999). E-CELL: software environment for whole-cell simulation. Bioinformatics *15*, 72–84.

Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol. Biol. *313*, 171–192.

Ulitsky, I., Shlomi, T., Kupiec, M., and Shamir, R. (2008). From E-MAPs to module maps: dissecting quantitative genetic interactions using physical interactions. Mol. Syst. Biol. *4*, 209.

Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. Nature *150*, 563–565.

Walpole, J., Papin, J.A., and Peirce, S.M. (2013). Multiscale computational models of complex biological systems. Annu. Rev. Biomed. Eng. *15*, 137–154.

Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nat. Rev. Genet. *11*, 843–854.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. U S A *109*, 1193–1198.

Cell Systems, Volume *2*

Supplemental Information

Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems

Michael Ku Yu, Michael Kramer, Janusz Dutkowski, Rohith Srivas, Katherine Licon, Jason F. Kreisberg, Cherie T. Ng, Nevan Krogan, Roded Sharan, and Trey Ideker

Supplemental Figures and Legends

Figure S1. Prediction of pairwise genetic interactions under a stringent cross-validation setup, related to Figure 3. Gene pairs were partitioned such that genes represented in the training set are absent in the corresponding test set (**Supplemental Experimental Procedures**). Correlation was calculated across gene pairs that meet an interaction significance criterion of $p<0.05$ in Costanzo et al.

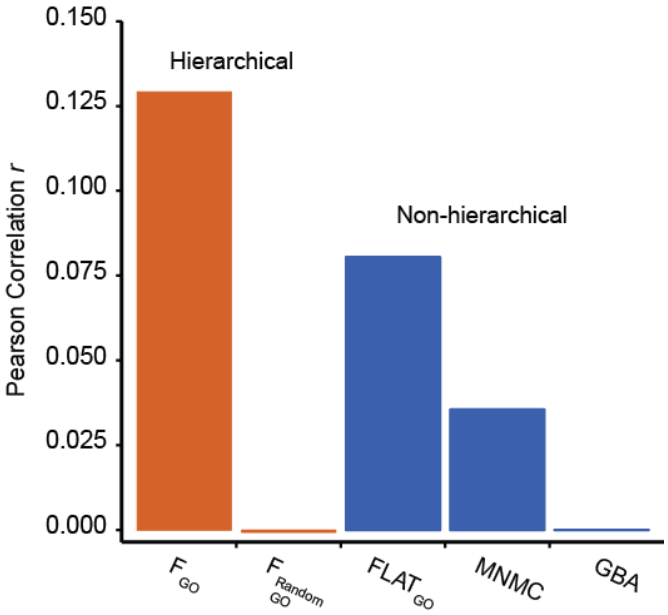


Figure S2. Fuller characterization of genes in an ontology increases prediction accuracy, related to Figure 3. The extent of characterization of a gene in a gene ontology was measured by its Annotation Information Content (AnnIC, **Supplemental Experimental Procedures**). **(A)** Distribution of AnnIC in GO and NeXO across the set of genes annotated to both. **(B)** Distribution of differences between AnnICs computed in NeXO versus GO. **(C,E)** Prediction performance across gene pairs where the GO AnnIC of at least one of the genes is at most a maximum level **(C)** or where the GO AnnICs of both genes are at least a minimum level **(E)**. **(D,F)** The number of gene pairs evaluated in **(C)** and **(E)**. **(G)** Scatterplot of GO terms, comparing the mean AnnIC of genes within the term (x-axis) to the ability to predict within-term genetic interactions (y-axis). Prediction performance is assessed by the correlation between measured and predicted genetic interaction scores across gene pairs that met the interaction significance criterion of $p<0.05$ in Costanzo et al. Black line indicates the relationship $y=0.0086x+0.000091$, fit by a linear regression ($r=0.39$) where each term is weighted by the number of evaluated gene pairs in the term (area of circles in scatterplot). Only terms with at least 25 and at most 10^5 evaluated gene pairs are shown. The terms "DNA repair" and "nuclear lumen", explored further in the paper, are indicated.

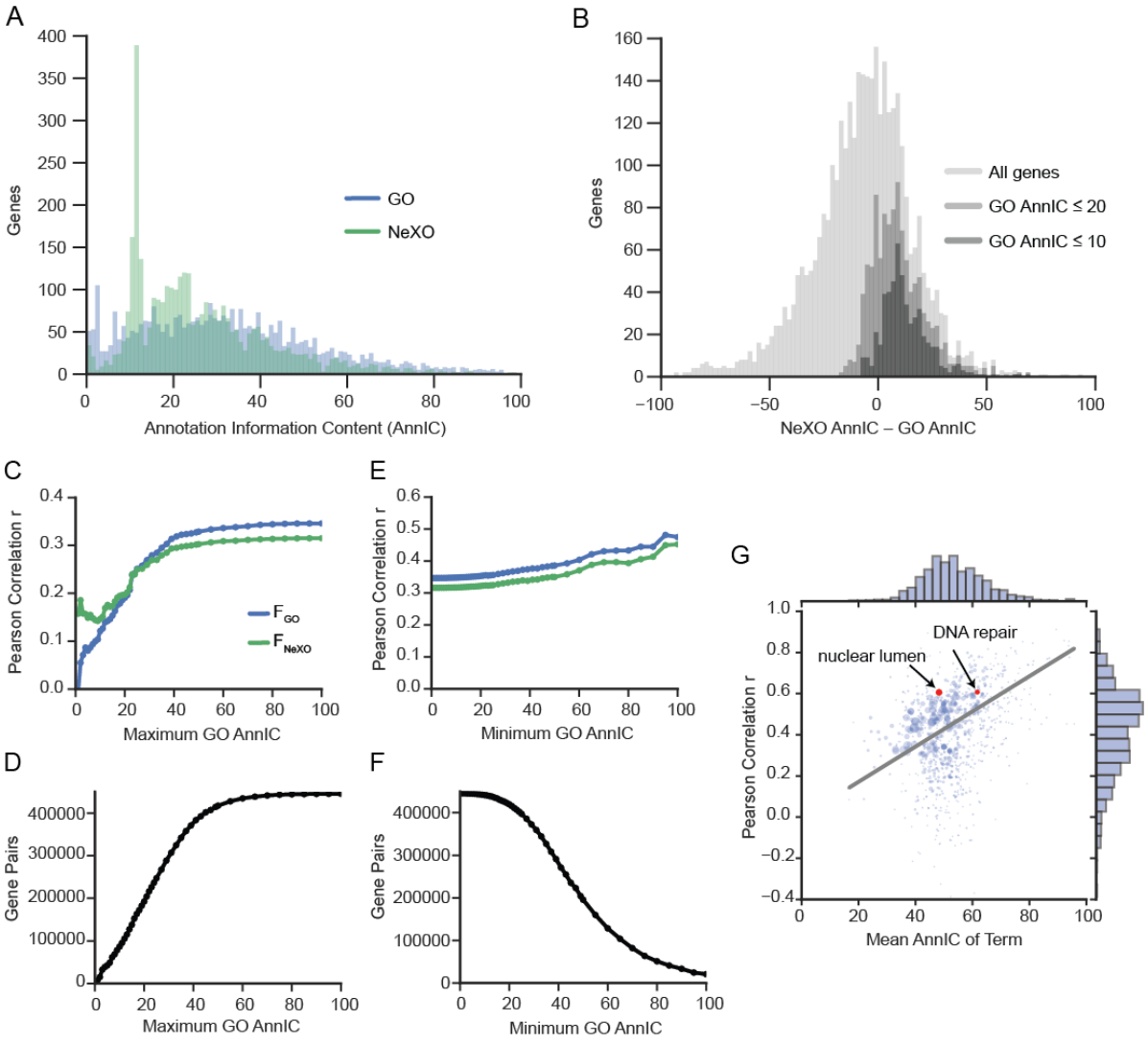


Figure S3: Segregation of positive and negative interactions across F_{GO}, related to Figure 4. For every term enriched for within-term genetic interactions, we calculate the fraction of positive versus negative interactions that continue to fall within one of its children. The distribution of this fraction is shown for all such terms, separately for positive (blue) and negative (red) interactions. Positive interactions falling within a term tend to remain within-term at lower hierarchical levels, whereas negative interactions are more likely to split their genes between two children.

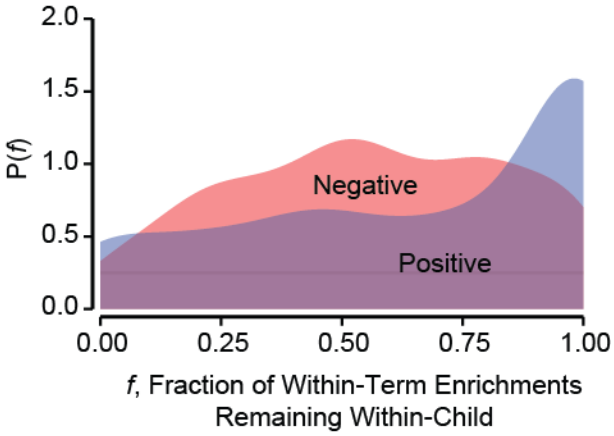


Figure S4: Genetic logic of the nuclear lumen, related to Figure 5. The nuclear lumen contains a rich hierarchy of cellular spaces, sub-components, and protein complexes interwoven by genetic interactions. Coloring and visual style of panels follow the convention of previous figures.

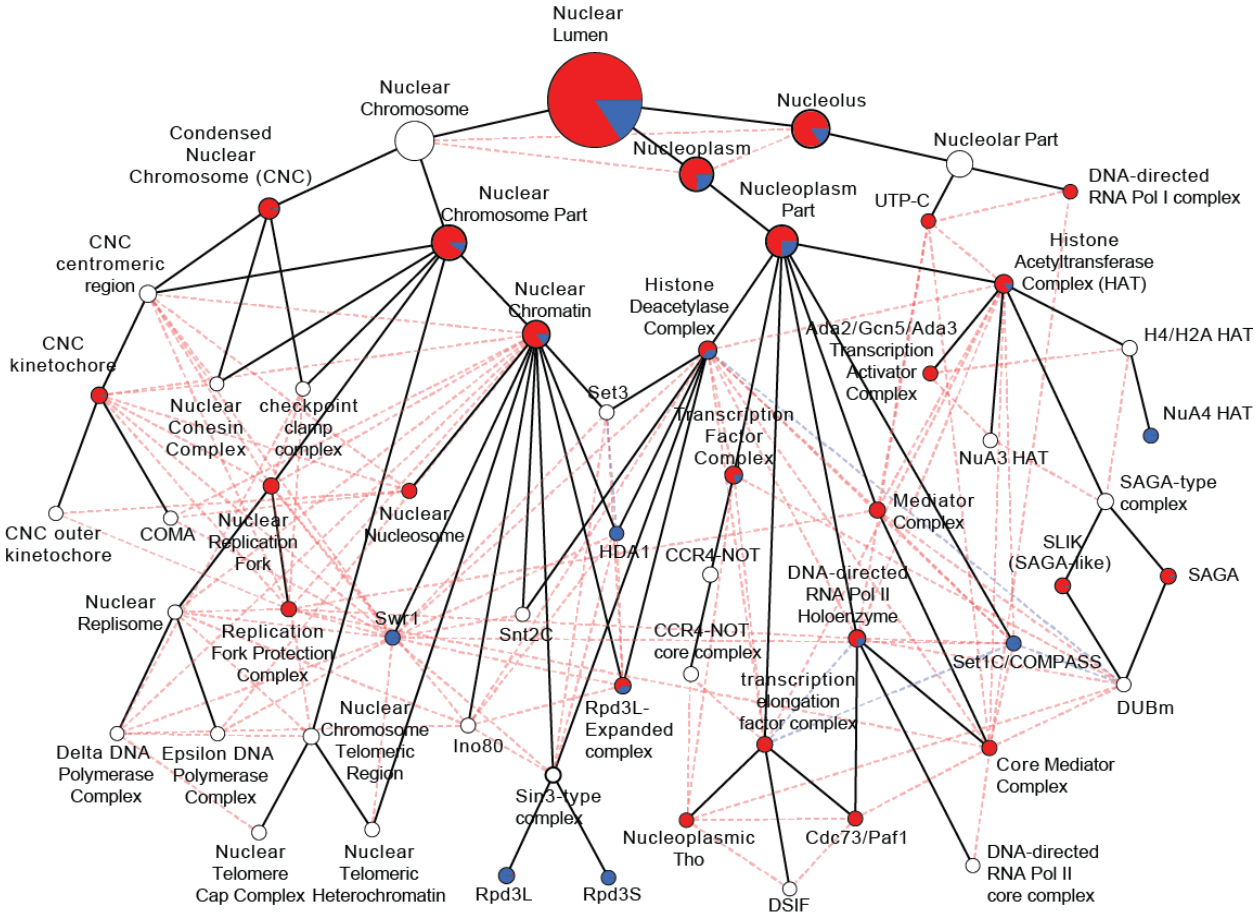


Figure S5: Prediction performance versus the depth of decision trees, related to Experimental Procedures. Decision trees in F_{GO} and F_{NeXO} were pruned to progressively shorter depths and evaluated across one of the cross-validation folds used in **Figure 3C**.

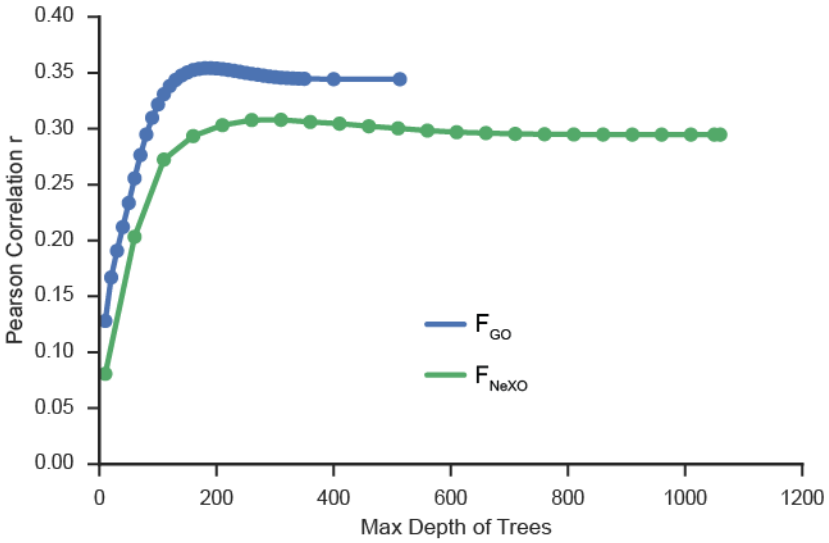
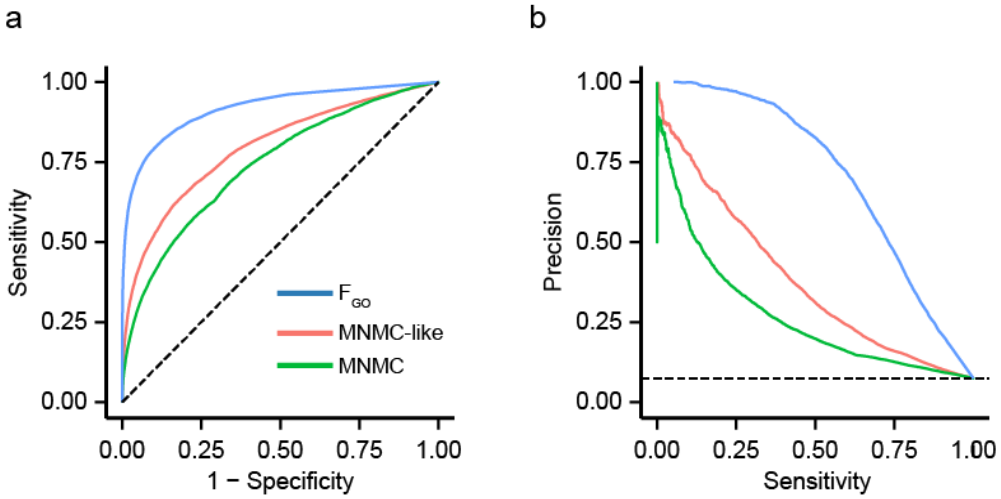


Figure S6: Prediction of synthetic lethal interactions curated in the *Saccharomyces* Genome Database, related to Experimental Procedures. 9,963 synthetic lethal and 124,312 non-synthetic lethal relations, as used in the original MNMC study (Pandey et al., 2010), were taken from the May 2008 version of the *Saccharomyces* Genome Database (SGD). Predictions were made in 10-fold cross validation using F_{GO}, the original MNMC, and a MNMC-like method we constructed based on updates to the original features with alternative classification techniques. **(A)** Receiver-operating and **(B)** Precision-recall curves. Performance of all approaches was far better on this 2008 SGD dataset than on the 2010 Costanzo et al. dataset explored in our main study (**Figure 3C,F**).



Supplemental Tables

Table S1: All 71 new functional relationships found in F_{GO}, related to Table 1.

Table S2: Prediction performance and gene characterization within GO terms, related to Figure 5.

Table S3: New genetic interaction scores for double mutants in DNA repair and the nuclear lumen, related to Figure 5.

Table S4: Gene Ontology structure and gene annotations, related to Experimental Procedures.

Table S5: NeXO structure and gene annotations, related to Experimental Procedures.

Supplemental Files

File S1: Genetic interaction scores predicted by F_{GO} for all pairs of non-essential genes, related to Figure 4.

File S2: Visualization of F_{GO} in Cytoscape format, related to Figure 4 and 5.

Supplemental Experimental Procedures

Previous genetic interaction data

Experimental genetic interaction scores for >6 million double mutants in yeast, measured using synthetic genetic arrays (Costanzo et al., 2010) (SGA, 1,711 queries × 3,885 arrays), were downloaded from http://drygin.ccbr.utoronto.ca/~costanzo2009/. We used only measurements where each gene is nonessential (any gene not listed as essential in the *Saccharomyces* Genome Deletion Project, http://www-sequence.stanford.edu/group/yeast_deletion_project/Essential_ORFs.txt) and is mutated as a deletion allele, excluding temperature-sensitive and DAmP alleles. Reciprocal tests for the same gene pair, i.e., query A × array B versus query B × array A for two different genes A and B, were merged as previously described (Costanzo et al., 2010). Analysis of these data is based on the principle that the vast majority of strains follow a multiplicative null model in which $f_{ab}=f_a f_b$, where f_{ab} is the measured growth of the double knockout and f_a and f_b are the corresponding single knockout measurements. The genetic interaction score is represented by the deviation $\epsilon_{ab}=f_{ab}-f_a f_b$ from the null model and assigned a *p*-value of significance p_{ab} using experimental replicates. Following the guidelines of the original data publication (Costanzo et al., 2010), we thresholded ϵ_{ab} and p_{ab} to categorize gene pairs as "negative" or "synthetic sick" interactions (ϵ_{ab}<-0.08 at p_{ab}<0.05), "positive" or "epistatic" interactions (ϵ_{ab}>0.08 at p_{ab}<0.05), or no significant interaction. For further analysis, we selected measurements for 3,331,172 pairs (69,639 negative and 37,494 positive across 3,686 unique genes) where both genes are annotated to a non-root term in GO, or 3,463,034 pairs (70,353 negative and 37,924 positive across 3,586 unique genes) where both genes are annotated in NeXO.

New genetic interaction mapping of DNA repair and the nuclear lumen

Double gene deletion mutants were generated on solid agar media using SGA technology as previously described (Collins et al., 2010; Tong and Boone, 2006). Briefly, double mutants were allowed to grow for 48 hours at 30°C and imaged using a Canon CCD camera. Colony sizes were quantified using the HT Colony Grid Analyzer (v. 1.1.7), after which genetic interaction scores (S-scores, **Supplemental Table S3**) were computed using the E-MAP toolbox (v. 2.0) (Collins et al., 2006). 55 genes in DNA repair and 81 genes in the nuclear lumen, of which 29 genes are in both terms, were knocked out among the 2,503 measured double mutants. For 315 double mutants, both genes were annotated to both terms. Negative interactions were called with an S-score threshold of –3.70 for gene pairs in DNA repair or –2.47 for gene pairs in the nuclear lumen. These thresholds were determined such that, for gene pairs tested both by these new screens and previously by Costanzo et al., the fraction of negative interactions based on the S-scores was the same as the fraction based on previous measurements (9.0% for DNA repair and 7.5% for the nuclear lumen). Interactions were deposited in the BioGRID database (Stark et al., 2006) and can also be accessed on NDEx (http://goo.gl/1yyxJT, UUID: 1515bf44-c398-11e5-8fbc-06603eb7f303, Pratt et al., 2015)

We combined all S-scores with the previous ϵ-scores by Costanzo et al. and retrained F_{GO} (**Supplemental File S2**). In this more complete dataset, we multiplied the S-scores that are <0 (or >0) by a factor of 0.02726 (or 0.04251) so that they would be on the same scale as the previous ϵ-scores. In particular, for the 1,345 gene pairs with both types of scores, the fraction of pairs with a rescaled S-score of <-0.08 (or >0.08) was the same as the fraction of negative (or positive) interactions based on the previous ϵ-scores. For these gene pairs, each type of score was listed as a separate data point for training F_{GO}.

Preparation of the Gene Ontology

The GO structure (OBO) and gene-to-term annotation (GOA) files were downloaded on December 19, 2011 (www.geneontology.org). Annotations with the evidence code "inferred by genetic interaction" (IGI) were removed to avoid potential circularity in predicting genetic interactions. The remaining annotations were propagated up the GO hierarchy in a transitive manner, i.e., if a gene is annotated to a term, then we also annotate the gene to all ancestral terms. We removed terms that were not annotated with any yeast genes or were redundant with respect to their children terms to construct a GO relevant to yeast, following a previously described procedure (Dutkowski et al., 2013, http://mhk7.github.io/alignOntology/). In total, 5,124 terms were drawn from all three branches of the Gene Ontology—Cellular Component (707), Biological Process (2,598), and Molecular Function (1,819). Of these, we retained 4,341 terms that were annotated to at least one gene in the genetic interaction data (Costanzo et al., 2010). The three branches were joined under an artificial root term to create a single unified ontology.

It is possible that some annotations were derived from genetic interaction information but were not labeled with the IGI code. Such mislabeling is a curation oversight, and identifying the problematic annotations would likely require a large-scale curation effort similar to that invested into constructing GO. However, while such mislabeled annotations might enable F_{GO} to make circular predictions, they would also enable circularity in $FLAT_{GO}$, in which

the pairwise gene similarities would also reflect the hidden genetic interaction information. The substantially worse performance of $FLAT_{GO}$ versus F_{GO} (**Figure 3C,F**) suggests that the degree of circularity is minimal.

Construction of NeXO: a data-driven ontology assembled from flat gene networks

Our aim was to systematically assemble NeXO from many diverse genome-sized datasets while, at the same time, guiding this assembly based on prior knowledge of a hierarchy of cell systems defined in GO. As input datasets, we used the YeastNet v3 networks (Kim et al., 2014), spanning 68 experimental studies across 8 data types excluding genetic interactions. We integrated these networks together by using them as features in a supervised learning of GO semantic similarity (Resnik, 1995) (random forests regression). Predictions were made "out of bag", i.e., the similarity of a gene pair was predicted based on information learned from other gene pairs. In effect, the random forest learns patterns in the networks which recapitulate information in GO. Hence, only relations in GO that can be systematically explained from data are included, any relations not justified by the data are excluded, and new relations not in GO are added when the network data support them. Next, we applied the method of Clique Extracted Ontology (CliXO) (Kramer et al., 2014, http://mhk7.github.io/clixo_0.3/), with parameters α=0.04 and β=0.5, on this network. CliXO recursively identifies densely-related groups of genes and hierarchically nests them to form NeXO. We identified the significantly aligned terms between NeXO and GO using a previously described ontology alignment procedure (Dutkowski et al., 2013, http://mhk7.github.io/alignOntology/).

Randomization of ontologies

We constructed randomized ontologies by shuffling the names of genes, while preserving gene-to-term annotations and parent-child term relations. In particular, let $\{T_G\}$ refer to the set of terms annotated with gene G. If the names of genes A and B are swapped, then gene B becomes annotated to $\{T_A\}$ and gene A becomes annotated to $\{T_B\}$. This randomization procedure preserves global degree properties of the ontology, i.e., the number of genes, gene pairs, or any k-tuple of genes that span a single term, a pair of terms, or any k-tuple of terms. A total of 10^5, 10^3, and 3 randomized ontologies were used in **Figure 1A**, **Figure 1C,D**, and **Figure 3C**, respectively. For these figures, the original mapping of genotypes to phenotypes was used.

Random forests regression

Random forests (Breiman, 2001) from the Python scikit-learn package (Pedregosa et al., 2011) (development version 0.15) were used to regress genetic interaction scores ϵ_{ab}, as described in the **Results**. Due to the very large size of the ontotype feature matrix, e.g., 3,331,172 genotypes (rows) by 4,341 terms (columns) for GO, training a random forest on the ontotype required an immense amount of computation. To address this challenge, we optimized the scikit-learn source code for both memory consumption and runtime performance, as follows. First, the feature matrix was represented with 8-bit integers instead of the default, 32-bit floating points, such that it consumed ~14GB instead of 58GB of memory. Second, the matrix was stored in Fortran-style order, i.e., columns instead of rows were stored contiguously in memory. Since individual columns are accessed one at a time during training, this ordering improved memory cache hits and resulted in approximately a 1.5-fold decrease in runtime. Third, redundant arithmetic that was being recomputed to find optimal decision splits was eliminated, resulting in a nearly 5-fold decrease in runtime. Together, these optimizations allowed a decision tree in the random forest to be trained in approximately 2 instead of 12 CPU hours. The modified code is available on GitHub (https://github.com/michaelkyu/scikit-learn-fasterRF). Following similar thresholds for the measured interaction scores, we categorized gene pairs as being predicted to have a 'negative' interaction (predicted $\epsilon_{ab} < -0.08$), a 'positive' interaction (predicted $\epsilon_{ab} > 0.08$), or no interaction.

Each random forest consisted of 300 trees, where every tree was learned over a bootstrap sample of gene pairs. At every decision split in a tree, one-third of all unused features were considered for the optimal split, defined by the minimum squared error. Trees were grown to maximal depth, following the default behavior of popular random forests libraries (Liaw and Wiener, 2002; Pedregosa et al., 2011) and the original random forests study (Breiman, 2001). We tested whether pruning these trees to a shorter depth would improve performance. While trees at approximately 29% (GO) or 37% (NeXO) of the maximal depth did improve performance slightly (<0.02 gain in correlation, **Supplemental Figure S5**), we chose to keep the trees at maximal depth because it is unclear how significant this gain is and whether it would be reproducible in different random partitions of the data for cross validation or in different genotype-phenotype datasets.

We used F_{GO} to predict genetic interactions across all pairwise deletions of 5,003 genes that are nonessential and annotated to at least one non-root term in any of the GO branches (**Supplemental File S1**). Performance of these predictions was assessed against the measurements selected from Costanzo et al. using the technique of four-fold cross-validation, resulting in 17,807 negative and 3,246 positive predicted interactions. For the remaining pairs, predictions from the four forests across the cross-validation folds were averaged, resulting in an additional 17,888

negative and 2,666 positive predicted interactions. This averaging was also used to predict scores for the retested mutants using the F_{GO} with complete training in **Figure 5B** and for the triple knockouts involving genes in the HIRA complex.

Comparison of methods for predicting genetic interactions

We updated the MNMC method because the original implementation (Pandey et al., 2010) was trained on a set of literature-curated synthetic lethal interactions that was much smaller in size than the set of genetic interactions considered in our study, and because the set of features used by the method to score each gene pair had been updated since the 2010 publication. To train MNMC, we calculated five base features that, according to the original study, are among the most informative for predicting synthetic lethality of a gene pair. Three of these features corresponded to the semantic similarities of the gene pair in each branch of GO. Following the procedure in the original study, Tao semantic similarities (Tao et al., 2007) were calculated using an unprocessed GO structure and annotations (downloaded on Aug 5, 2012) instead of the processed GO for constructing ontotypes. A fourth feature was the binary co-membership of a gene pair in a protein-protein interaction (PPI) community. Twenty-six communities were identified using the network of curated protein-protein interactions in the BioGRID database (Stark et al., 2006) v3.1.91 and running the 'Graph.community_multilevel(return_levels=False)' method from the Python igraph library v0.6 (Nepusz, 2006). A fifth feature was the number of KEGG pathways (Kanehisa et al., 2014) shared by each gene pair. Annotations of genes to *S. cerevisiae* pathways in KEGG were retrieved on March 16, 2015 using the KEGG REST API (kegg.jp/kegg/rest). Lastly, these five base features were combined in pairs to create (5 choose 2) = 10 additional 'overlay' features as previously described in the original MNMC method. We trained a random forest using all 15 base and overlay features to classify the synthetic lethals evaluated in the original MNMC study. This updated MNMC outperformed the original MNMC (**Supplemental Figure S6**); this performance difference may have been due to calculating more recent versions of the base features. For **Figure 3C,F**, we trained a separate MNMC on the genetic interactions measurements by Costanzo et al.

We adapted a previous GBA method for predicting genetic interactions (Lee et al., 2010) and applied it on the data similarity network used to construct NeXO. Whereas the original GBA method considered binary genetic interactions and the network neighbors of only one gene in a gene pair, we consider continuous-valued genetic interactions and the neighbors of both genes. Let *S(a,b)* be the edge weight between genes *a* and *b* in the data similarity network. The GBA score between *a* and *b* is defined as

$$GBA(a,b) = \sqrt{GBA(a)GBA(b)}$$

$$GBA(x) = \sum_{y \in N(x)} S(x,y) * \epsilon_{xy}$$

where *N(x)* is the set of neighbors of gene *x* and ϵ_{xy} is the measured genetic interaction. Conceptually, *GBA(x)* is the contribution of *x*'s neighborhood, and *GBA(a,b)* is the geometric mean of the contributions of both *a*'s and *b*'s neighborhoods. We found that taking the geometric mean more accurately predicts genetic interactions than the arithmetic mean does. Using the data similarity network (Pearson's $r = 0.13$; **Figure 3C**) resulted in approximately similar performance as using the YeastNet v3 integrated network ($r = 0.14$).

The same random forest framework as the functionalized ontologies and MNMC was applied to train $FLAT_{NeXO}$, using the 68 input networks to NeXO as features, and $FLAT_{GO}$, using the Resnik semantic similarity (Resnik, 1995) between gene pairs in GO as the sole feature. In **Figure 3C,E**, F_{GO}, $F_{Random\ GO}$, $FLAT_{GO}$, and MNMC were trained over the same cross-validation partitioning of 3,331,172 gene pairs. Similarly, F_{NeXO}, $F_{Random\ NeXO}$, $FLAT_{NeXO}$, and GBA were trained over the same cross-validation of 3,463,034 pairs.

To generate the strict cross-validation setup in **Supplemental Figure S1**, the set of 3,686 genes in the processed GO were partitioned into 8 approximately equal-sized subsets $G_1, G_2 \ldots, G_8$. For every pair of subsets G_i and G_j, the set of all gene pairs where neither gene is in $G_i \cup G_j$ forms a training set, and the set of all gene pairs where both genes are in $G_i \cup G_j$ forms the corresponding test set. This partitioning scheme lets every gene pair be predicted in at least one test set. Gene pairs where both genes are in the same subset G_i are represented in seven test sets, and we take the average prediction across these sets.

FBA predictions were taken from Szappanos et al., 2011. Since the underlying metabolic model (Mo et al., 2009) has been iteratively refined by manual curation to be consistent with genetic and physical data, we were unable to evaluate these predictions in a cross-validation scenario.

Characterization of genes in a gene ontology

To score how well a gene has been characterized in a gene ontology, we introduce the concept of a gene's Annotation Information Content (AnnIC). Formally, the AnnIC of a gene g is defined as

$$AnnIC(g) = \sum_{t \in T_g} -\log_2 \frac{|t|}{N}$$

where T_g is the set of terms that contain g but are not ancestors of any other term that also contains g, $|t|$ is the size of term t, and N is the number of genes in the ontology. Conceptually, AnnIC(g) is high if g is well characterized by being annotated to many small terms, and it is low if g is poorly characterized by being annotated to only a few large terms.

Term enrichments

A term is a least common ancestor (LCA) of a pair of genes if both genes are assigned to it, and if it is not the ancestor of any other term shared by those genes. Note that a pair of genes may have multiple LCAs, since terms may have multiple parents. Let G define a set of gene pairs beneath a term or term pair, as follows:

$$G(T) = \{\{g_1, g_2\} \mid \text{Term } T \text{ is an LCA of } g_1 \text{ and } g_2\}$$
$$G(T_1, T_2) = \{\{g_1, g_2\} \mid \exists\, P, \text{a common parent of } T_1 \text{ and } T_2 \text{ that is also an LCA of } g_1 \text{ and } g_2\}.$$

A term T was identified as a within-term enrichment if the fraction of gene pairs in $G(T)$ that are interacting is at least twice the background rate of interactions, and the number of gene pairs in $G(T)$ that are interacting passes a hypergeometric test with $p<0.05$ using a Bonferroni correction for family-wise error rate. A pair of terms T_1 and T_2 was identified as a between-term enrichment according to a similar test on $G(T_1,T_2)$. We did not test for within-term enrichment when $G(T)=\emptyset$, or for between-term enrichment when $G(T_1,T_2)=\emptyset$, including the situation where T_1 and T_2 do not have a common parent. These situations were not counted in the total number of tests computed for Bonferroni correction. Note that it is possible for a parent term to have within-term enrichment even if none of its children have between-term enrichments, and vice versa. Enrichments were identified separately for negative and positive interactions. In **Figure 1C**, we counted the number of unique terms and unique term pairs that formed within-term and between-term enrichments for both types of interactions. Together, these enrichments involved 1,050 unique terms (~24% of the 4,341 GO terms that contained at least one gene among the measured genetic interactions).

In addition to between-term enrichments that share a common parent, we also tested enrichment among pairs of terms that were distantly related in GO, i.e., the genes contained by one of the term's parents did not overlap with those of the other term. We only considered pairs of terms that contained at least 12 gene pairs and for which each term contained at least 3 genes. Enrichment was based on the number and ratio of interactions according to a similar criteria for $G(T)$ and $G(T_1,T_2)$. Most of the resulting enrichments were partially redundant such that terms in different enrichments shared some genes. To remove this redundancy, we first sorted the enrichments in descending order according to the fraction of gene pairs that had an interaction, breaking ties by sorting according to the number of gene pairs. Using this sorted ordering, we greedily selected enrichments so that the genes contained by the terms in an enrichment were not already covered by a previously selected enrichment. This selection procedure resulted in a total of 71 enrichments (53 negative and 18 positive) (**Table 1, Supplemental Table S1**).

Visualization of the functionalized ontology

F_{GO} was visualized using Cytoscape version 3.2 (Saito et al., 2012). Only a spanning tree of the GO structure is shown in order to remove multi-parent relationships that complicate visualization. To form the spanning tree, the edges from each term to all but one of its parents were removed. The tree was laid out using the yFiles Circular layout tool, and edges representing between-term enrichments with a common parent were bundled together using the automatic edge bundling tool. The visualizations of the structure and logic of F_{GO} shown in **Figure 4A** (http://goo.gl/ViTztH, UUID: 2f0cd18a-c3a4-11e5-8fbc-06603eb7f303), **Figure 4C** (http://goo.gl/iWmvNk, UUID: 420e4e54-c3a2-11e5-8fbc-06603eb7f303), **Figure 5A** (http://goo.gl/mPkMVs, UUID: 555cd256-c3a3-11e5-8fbc-06603eb7f303), and **Supplemental Figure S4** (http://goo.gl/oQWh0A, UUID: 575ee3e8-c3a3-11e5-8fbc-06603eb7f303) can be accessed online on the Network Data Exchange (NDEx, Pratt et al., 2015) and as a Cytoscape file in **Supplemental File S2**.

Supplemental References

Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. *7*, R63.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–205.

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. R News *2*, 18–22.

Mo, M.L., Palsson, B.O., and Herrgård, M.J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst. Biol. *3*, 37.

Nepusz, G.C. and T. (2006). The igraph software package for complex network research. InterJournal *Complex Sy*, 1695.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., Lotia, S., Pico, A.R., Bader, G.D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. Nat. Methods *9*, 1069–1076.

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535–9.

Tao, Y., Sam, L., Li, J., Friedman, C., and Lussier, Y. a (2007). Information theory applied to the sparse gene ontology annotation network to predict novel gene function. Bioinformatics *23*, i529–38.