Metabolic Flux Correlations, Genetic Interactions, and Disease

BALAJI VEERAMANI and JOEL S. BADER

ABSTRACT

Many diseases are caused by failures of metabolic enzymes. These enzymes exist in the context of networks defined by the static topology of enzyme-metabolite interactions and by the reaction fluxes that are feasible at steady state. We use the local topology and the flux correlations to identify how failures in the metabolic network may lead to disease. First, using yeast as a model, we show that flux correlations are a powerful predictor of pairwise mutations that lead to cell death—more powerful, in fact, than computational models that directly estimate the effects of mutations on cell fitness. These flux correlations, which can exist between enzymes far-separated in the metabolic network, add information to the structural correlations evident from shared metabolites. Second, we show that flux correlations in human align with similarities in Mendelian phenotypes ascribed to known genes. These methods will be useful in predicting genetic interactions in model organisms and understanding the combinatorial effects of genetic variations in humans.

Key words: flux balance analysis, metabolic network, OMIM, synthetic lethal genetic interactions, systems biology.

1. INTRODUCTION

OVER 200 OF THE DISEASES LISTED IN THE Online Mendelian Inheritance in Man (OMIM) database are due to failures in metabolic enzymes (McKusick, 2007). Complex diseases such as diabetes may also arise through disruptions of metabolism. Yeast is a valuable model for metabolic disease due to evolutionary conservation of core pathways and availability of genetic tools, most significantly a knockout collection for facile tests of single (Giaever et al., 2002) and combinatorial gene deletion phenotypes.

Though no doubt far from complete, metabolic reconstructions are now sufficiently mature to provide testable computational predictions. Reconstructions are available for human (Duarte et al., 2007), yeast (Duarte et al., 2004), and a growing number of other organisms. These reconstructions are arguably more complete than those for signal transduction, gene regulation, or any other type of cellular network. Correlations between enzymes, based on the static network structure and shared metabolites, have recently been mapped to correlations in disease phenotypes (Lee et al., 2008).

Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland and High-Throughput Biology Center, Johns Hopkins School of Medicine, Baltimore, Maryland.

Flux balance analysis (FBA), based on the stoichiometric matrix, provides a quantitative, computationally efficient method for systems-level investigations of the effects of single and multiple perturbations of metabolic enzymes on the performance of the network (Price et al., 2004a). Most often, metabolic fitness has been assessed by the optimal value of a particular linear combination of fluxes over the FBA feasible space. This was used to determine lethal mutants in bacterial metabolic networks (Savinell et al., 1992).

More recently, fitness optimization was used to identify epistatic effects of multiple gene deletions in yeast: pairs of genes whose combined deletions give non-additive effects, often aggravating (Segre et al., 2005). Epistatic interactions were suggested to cross between metabolic modules, but were not compared systematically to genetic interactions known from experiment. Comparison with experimental data is feasible due to related high-throughput experimental methods of SGA (Tong et al., 2004), dSLAM (Pan et al., 2004), and EMap (Schuldiner et al., 2005), with over 14,585 interactions deposited in the BioGRID database (Stark et al., 2006).

An important limitation of a computational approach based on fitness is imperfect knowledge of what exactly a cell is trying to optimize. Fitness functions are typically based on perceived demands for metabolites required for building biomass, but do not necessarily represent the real environmental and temporal requirements of a cell. While the fitness function used for calculations is linear for efficiency, non-linear terms may also contribute. Thus, rather than attempting to identify the single optimal cellular state, it may incur less error to calculate averages over feasible regions (Almaas et al., 2004; Price et al., 2004b). Averaging over the feasible flux space estimates flux correlations for all pairs of enzymes. Although flux states represent reaction velocities that define metabolic rates, they are feasible steady-state equilibria rather than “dynamics” in the conventional sense of time-ordered trajectories or transients.

Our hypothesis is that flux correlations can provide evidence of metabolic correlations that are not evident from either fitness-based epistatic interactions or the static metabolic network topology. First, using the yeast model, we compare the abilities of computational methods based on topological correlations, epistasis, and flux correlations to predict experimentally observed genetic interactions in yeast. Next, we extend this work to human, identifying metabolic couplings between enzymes that are also linked by similar disease phenotypes. These results indicate that flux correlation structure inferred from computational FBA can reveal the metabolic functioning of model organisms and humans.

2. RESULTS

2.1. Definition of metabolic couplings

Four measures of metabolic coupling are considered: metabolite sharing, reaction sharing, epistatic interaction, and flux correlation. Metabolite sharing scores and reaction sharing scores are calculated directly from the static structure of the metabolic reconstructions for yeast (Becker et al., 2007; Duarte et al., 2004) and human (Duarte et al., 2007) with more recent updates. For each enzyme, its total number of metabolites (or reactions) was tabulated from the stoichiometry matrix, and for each pair the shared number was calculated. The score was calculated as the log-likelihood of a Bayesian model for enriched sharing relative to a null model of random sharing. The metabolite sharing method only considers whether an enzyme is connected to a metabolite, not the stoichiometry. Metabolites and reactions assigned to different compartments were considered different. High-frequency metabolites such as water, protons, and inorganic phosphate were retained.

Epistatic interactions were taken from published tables (Segre et al., 2005) that used FBA to calculate optimal fitness under a condition termed the “nominal simulated setting.” This condition corresponds to an *in silico* $\Delta his3\Delta leu2\Delta rip1$ strain grown on glucose and oxygen-limited minimal medium containing nitrogen, phosphate, sulfate, threonine, histidine, leucine, and uracil; this setting performed best in comparison with experimentally observed genetic interactions.

Samples of reaction fluxes were obtained by sampling methods provided with the COBRA toolbox (Becker et al., 2007). We transformed reaction fluxes to enzyme fluxes with a method that aggregated fluxes for multiple reactions catalyzed by a single enzyme and that apportioned fluxes when a single reaction was catalyzed by multiple enzymes. Equilibrium flux correlations were then calculated for all pairs of enzymes, with three independent runs used to assess convergence.

2.2. Distribution of metabolic couplings for enzyme pairs with genetic interactions

Enzyme pairs with experimentally observed genetic interactions have distinct patterns of metabolic couplings when compared with pairs that do not interact (Fig. 1). The experimentally observed genetic interactions considered here are all deleterious interactions reported as synthetic lethality or synthetic growth defects between yeast genes in BioGRID (Stark et al., 2006), abbreviated as synthetic lethal (SL). There are 68 SL pairs where each gene is part of the yeast metabolic reconstruction, and 17,323 non-synthetic lethal (NSL) pairs of genes that occur in both the metabolic and the SL network.

Despite experimental observations of a combined fitness defect for SL pairs, most of the corresponding epistasis scores are 0 (Fig. 1A). A small number of SL pairs have large negative epistasis scores, indicating successful predictions; no SL pairs have positive epistasis scores. Metabolite sharing scores are shifted to higher values for SL pairs relative to NSL pairs (Fig. 1B), indicating the synthetic lethality is enhanced for enzymes that share several metabolites. Reaction sharing scores have a similar pattern (results not shown).

Flux analysis shows enrichment of SL pairs for high correlation coefficients (Fig. 1C). High correlations are expected for enzymes that share reactions. Additionally, however, several SL pairs are also observed to have negative correlations. These may represent alternative pathways for generating a metabolite whose maximum flux is limited, generating a constraint that introduces a negative correlation.

The histograms in Figure 1 suggest that flux correlations and metabolic scores drawn from SL and NSL pairs may have different distributions. Quantitative tests of the corresponding null hypothesis were performed using two-sided tests, both parametric (*t*-test) and non-parametric (Wilcoxon rank sum tests). SL and NSL pairs have significantly different absolute flux-flux correlations (p -value $< 2.2 \times 10^{-16}$ for both *t*-test and Wilcoxon), metabolite-sharing scores (*t*-test p -value = 7.6×10^{-10} , Wilcoxon p -value = 4.2×10^{-13}), and reaction-sharing scores (*t*-test p -value = 1.3×10^{-6} , Wilcoxon p -value = 3.1×10^{-15}). Differences between scaled epistasis scores are insignificant based on *t*-tests (p -value = 0.24), but are significant based on the non-parametric Wilcoxon test (p -value = 0.00028).

The 68 SL pairs in the metabolic network group into several major themes. First, there are 35 pairs that catalyze at least one reaction in common. Destroying these parallel enzymatic routes between shared sets of metabolites causes lethality. Second, there are more than 12 distant reaction pairs that can be considered alternate pathway branches. These include alternate branches connecting key metabolites, such as TAL1 (transaldolase) and ZWF1 (glucose-6-phosphate dehydrogenase) in pentose phosphate metabolism, and alternate routes to obtain key metabolites by import versus synthesis, such as the the SL interaction between TAT2 (tryptophan permease) and TRP1 (tryptophan biosynthesis). Third, 14 pairs do not catalyze

FIG. 1. Pairs of yeast genes identified experimentally as synthetic lethal (SL) or not known to have a synthetic lethal interaction (NSL) are binned according to metabolic couplings estimated from scaled epistasis (A), topological correlation based on shared metabolites (B), and flux correlations (C). Only genes existing in the yeast reconstruction and having at least one SL interaction are included. Note the existence of SL pairs with negative epistatic interactions in panel A, the enrichment of SL pairs with high metabolite sharing scores in panel B, and the existence of SL pairs with both positive and negative flux correlations in panel C.

a reaction in common, but share metabolites. These pairs appear as serial steps in a single pathway. A possible explanation for lethality is that other enzymes are also able to catalyze each reaction, but at lower efficiency. Single hits are viable, but double hits are not.

A fourth distinct category are five SL pairs that could be considered to share reactions or metabolites, except that they are localized to different compartments (and hence not considered the same for sharing calculations). This category is interesting because it includes three of the five examples of large negative flux correlations: TPI1 (cytosol) with GUT2 (mitochondrion), which convert glyceraldehyde phosphate; PSD1 (mitochondrion) and PSD2 (vacuolar membrane), both phosphatidylserine decarboxylases; and GPD1 (cytoplasm) with GPD2 (mitochondrion), both glycerol 3-phosphate dehydrogenases. The negative flux correlation suggests that the cell chooses to run cytoplasmic pathways or equivalent mitochondrial pathways, but not both.

Finally, SL interactions CHO1 (phosphatidylserine synthesis) and both TRP1 and TRP2 (tryptophan biosynthesis) had no obvious relationship based on metabolism. The genetic interaction is due to the need for phosphatidylserine for efficient tryptophan uptake (Nakamura et al., 2000).

2.3. Predicting genetic interactions with metabolic couplings

Even though the metabolic couplings do not provide a clear separation between SL and NSL pairs, their information can still be used to predict genetic interactions. Rank-order lists were generated based on reaction and metabolite sharing scores (descending order), epistasis (ascending order), and flux correlations (descending order of absolute value). Performance was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC) plot of the true positive rate versus false positive rate, and the maximum F -score along the precision-recall curve. (The F -score is the harmonic mean of precision and recall.) Known positives corresponded to SL pairs. Known negatives for this assessment were restricted to NSL pairs where at least one gene had at least three SL partners to exclude pairs that might not have been tested experimentally.

According to the AUC metric, the flux correlation provides the best performance (Fig. 2A). Predictions based on flux correlation, metabolite sharing, and reaction sharing all out-perform the epistasis score. The

FIG. 2. Synthetic lethal (SL) gene pairs were predicted by flux correlations, metabolite sharing scores, reaction sharing scores, and epistasis scores, with performance assessed by a receiver operating characteristic (ROC) plot of the true-positive rate versus false-positive rate, with a summary statistic given by the area under the curve (AUC) (A), and a precision-recall plot (recall is synonymous with true-positive rate), with a summary statistic given by the maximum F -score (harmonic mean of precision and recall) (B). Results are also shown for a logistic regression predictor using the flux correlation and the metabolite sharing score, selected using stepwise regression starting with all four predictors.

three best performers rise rapidly on the ROC plot to a true-positive rate (recall) of 0.5–0.6, with a very small false-positive rate. From this point, the flux correlation rises higher than the metabolite sharing score, which in turn performs better than the reaction sharing score.

Greater detail of the initial high precision, low recall region is provided by the precision-recall curve (Fig. 2B). The reaction sharing score is able to achieve the highest precision of nearly 60% at about 50% recall, much better than the 40% precision of the flux correlation and the metabolite sharing score at the same recall. The performance of the reaction sharing score and metabolite sharing score drop rapidly beyond this point, however, while the flux correlation degrades less. Consequently, the *F*-score, a balanced measure of precision and recall performance, is very similar for the reaction score (0.53) and the flux correlation (0.52).

If these measures of metabolic coupling provide independent evidence, a combined classifier may provide better performance. We therefore developed a logistic regression predictor by starting with a full model including all four measures, then using backwards stepwise regression to identify a parsimonious model based on the Bayes Information Criterion (BIC). This procedure can be interpreted as identifying a subset of explanatory variables providing non-redundant information according to the BIC. The objective function optimized by the BIC is not identical with either the ROC/AUC or the PR/*F*-score, however, and significance based on BIC does not necessarily translate to a sizable difference in AUC or *F*-score. The final model included only the flux correlation and the metabolite sharing score. Metabolite sharing may carry more information than the reaction sharing because shared reactions imply shared metabolites, but shared metabolites may indicate a serial pathway relationship in addition to shared reactions.

The combined predictor has the same AUC and *F*-score as the flux correlation (logistic regression does not attempt to optimize either measure). The precision-recall plot shows that the combined predictor has better initial precision than the flux correlation, but then drops below the flux correlation ranking for recall above 65%.

As the AUC and *F*-score measures are quite close for the top-performing methods, we used bootstrap replicates to determine the reproducibility of the ranking of methods. Sampling of SL and NSL pairs with replacement was used to generate 100 bootstrap replicates, with SL and NSL counts identical to the original data. For each bootstrap replicate, the performance of each method was ranked from best (rank 1) to worst (rank 5). For AUC, flux correlation was top-ranked 60% of the time, and logistic regression was top-ranked 40% of the time. These two methods were in first and second place for all but 4% of the bootstrap replicates, indicating near equivalence in AUC and dominance over the other methods.

For *F*-score, flux correlations and logistic regression were each top-ranked 30% of the time, with reaction sharing top-ranked 40% of the time. These three methods took the top three places for all but 5% of the bootstrap replicates. These results suggest that reaction sharing is best for enriching known positives among the top-ranked predictions, while flux correlation (possibly as part of a logistic regression combined predictor) performs better over the entire range of predictions.

2.4. Interpreting disease models through metabolic coupling

Metabolic coupling through flux analysis can reveal connections between diseases, including diseases whose symptoms have no overt connection to metabolism. An example is Lesch Nyhan syndrome (LNS, OMIM 300322), characterized by neurological disorders and self-mutilation. This disease has been mapped to mutations in HPRT1, hypoxanthine guanine phosphoribosyl transferase. Partial deficiencies of this enzyme cause Kelley Seegmiller syndrome (OMIM 300323), which causes gouty arthritis and renal failure. HPRT is involved in purine synthesis and salvage. Yeast is a relevant model for this disease because of strong evolutionary conservation of purine metabolism pathways.

A view of genes metabolically coupled to purine synthesis and salvage in yeast shows that HPT1 is isolated from most of the pathway (Fig. 3A). A closer inspection of its local neighborhood (Fig. 3B) shows that its connections to purine metabolism are entirely through PRS1-5, 5-phospho-ribosyl-1(α)-pyrophosphate (PRPP) synthetase. In murine studies of HPRT1-deficient cells, these closely connected genes are over-expressed to compensate for HPRT1 deficiency (Song et al., 2007). Defects in human PRPS1 cause Charcot-Marie-Tooth disease (CMTX5, OMIM 311070) and Arts syndrome (OMIM 301835), whose phenotypes notably include peripheral neuropathy and other neurological impairments. Super-activity of

FIG. 3. Yeast metabolic network based on flux correlations. **(A)** Metabolic connections are depicted for HPT1 (yellow), up-regulated genes in HPRT-deficient cells (green), other yeast genes in the purine salvage and biosynthesis pathways (pink) and genes with correlated metabolic properties in other pathways (gray). Thicker lines denote stronger flux correlations. Solid lines denote sharing of metabolites; dashed lines indicate no significant sharing of metabolites. Red lines indicate synthetic lethality of the yeast knockouts. **(B)** A zoomed-in view of the neighborhood around HPT1 shows that the strongest correlations of this gene are with PRS1-5, phosphoribosylpyrophosphate (PRPP) synthetase. HPT1 and PRS1-5 share the single metabolite PRPP. Visualization was provided by Cytoscape (Shannon et al., 2003).

PRPS1 also causes neural defects (OMIM 300661), with early family studies undertaken by Nyhan et al. (1969). This example demonstrates a metabolic explanation for neurological phenotypes.

2.5. Flux correlations and human disease

For a more global view of disease and metabolism, we calculated flux correlations using a genome-scale reconstruction of the human metabolic network by Duarte et al. (2007). The conditions in the model were based on the minimum essential medium where hippocampal neurons and human embryonic cells (HEK) are grown (Banker et al., 1998). The reduced network under these conditions had 953 metabolites and 987 reactions.

Rather than classifying diseases by hand, we relied on fully automated text-mining procedures developed by others to establish correlations between disease descriptions from OMIM (van Driel et al., 2006; Wu et al., 2008). Disease correlations from MimMiner (van Driel et al., 2006) were mapped to genes, which in turn were mapped to enzymes in the metabolic network. A total of 16,547 gene pairs existed in both data sets.

For a global view, we considered a phenotypic correlation cut-off of 0.4, corresponding to 7% of the phenotypic pairs. These pairs were organized according to flux correlations and metabolite sharing scores (Fig. 4A). Many pairs have both high flux correlations and high metabolite sharing scores, representing the themes of shared reactions (parallel routes) and shared metabolites (serial connections). There are also pairs with high metabolite sharing but low flux correlation, and pairs with low metabolite sharing and high positive and negative flux correlation. When yeast SL pairs are plotted on the same axes, the distribution appears similar (Fig. 4B), suggesting that the metabolic themes observed in yeast synthetic lethality may apply more generally to human disease.

Others have previously found that similar phenotypes are correlated with shared metabolites (Lee et al., 2008). Our results provide evidence through flux correlations for metabolic couplings between genes that do not share metabolites. As with yeast, the metabolite compartments were used for humans when calculating sharing, and many of the examples with high flux correlation and low sharing in fact represent similar reactions occurring in different cellular compartments.

Additionally, however, the flux correlation analysis reveals disease associations that would not be obvious from an analysis of the static topology. An example is provided by subtypes of porphyria caused by defects in different enzymes responsible for porphyrin metabolism (Fig. 5A). The enzymes UROS

Downloaded by 54.162.127.20 from www.liebertpub.com at 04/26/23. For personal use only.

Downloaded by 54.162.127.20 from www.liebertpub.com at 04/26/23. For personal use only.

Downloaded by 54.162.127.20 from www.liebertpub.com at 04/26/23. For personal use only.

Downloaded by 54.162.127.20 from www.liebertpub.com at 04/26/23. For personal use only.

phosphate synthetase, mitochondrial) has flux and phenotypic correlations with ASS (argininosuccinate synthetase, cytoplasmic) and ASL (argininosuccinate lyase, cytoplasmic), although it is several steps removed from both. Mutations in these enzymes cause aberrant levels of arginine and citrulline. A third example (not displayed) involves cholesterol metabolism genes EBP, DHCR7, and DHCR24. Mutations in these genes cause chondrodysplasia X2, Smith–Lemli–Opitz syndrome, and desmosterolosis (Waterham, 2002).

3. DISCUSSION

A natural analogy exists between synthetic lethal phenotypes in model organisms, where multiple gene perturbations cause death and diseases in humans. Using metabolic reconstructions, we have investigated both. Synthetic lethality with respect to protein pathways has often been interpreted in terms of two models: between-pathway (independent hits to parallel or partially redundant pathways cause death) and within-pathway (multiple hits to a single pathway cause death) (Kelley et al., 2005; Ulitsky et al., 2007; Ye et al., 2005). The same patterns are evident in metabolic pathways. Enzymes that catalyze the same reaction, either in the same cellular compartment or more generally in different compartments, correspond to the between-pathway model. Synthetic lethality is also observed for adjacent enzymes that share a metabolite within a serial pathway, the within-pathway model.

In addition to static correlations based on the topology of enzyme-metabolite interactions, we gain additional information from flux correlations. While most synthetic lethal pairs in yeast and most disease-associated pairs in human have positive flux correlations, as would be expected for coupled reactions in a serial pathway, negative correlations are also observed. In yeast, negative flux correlations are enriched for reactions that are partially redundant due to existence in both the mitochondria and the cytoplasm. In humans, diseases with similar phenotypes can be traced to defects in enzymes that share no metabolites but have high flux correlations.

A surprising result of this study is that a previous method to estimate epistasis directly from the FBA fitness function did not work as well as statistical correlations of feasible but sub-optimal fluxes. A possible interpretation is that the optimization function used in FBA is not actually what the cell is optimizing, and extracting the fitness from the optimal vertex in the polytope might amplify shortcomings in the optimization flux. The flux correlation method draws information uniformly from the feasible space and avoids errors due to an incorrect optimization condition.

A direct relationship between synthetic lethal pairs in yeast and disease-associated pairs in humans may be possible, but the results presented here do not directly address this point. Such relationships would be limited to gene pairs with readily identifiable orthologs, and thus most applicable to metabolism and other ancient conserved functions. Direct relationships would be difficult to identify for genes with expanded gene families along either lineage, or families extant in only one species.

These methods will become more powerful with continuing improvements in metabolic reconstructions for humans, model organism, and pathogens. While applications here are to Mendelian disease, the same methods should be valuable in interpreting data from genome-wide association studies for complex disorders, pharmacogenetic interactions between drugs and variants, and metabolic weak points of microbial pathogens.

4. METHODS

4.1. Metabolic reconstructions

The *in silico* yeast model was based on a fully compartmentalized genome-scale metabolic network reconstruction (Duarte et al., 2004), developed from an earlier model (Forster et al., 2003). The network was obtained from the database maintained by systems biology research group at University of California, San Diego. The file corresponding to *Saccharomyces cerevisiae*, “*Sc_iND750_GlcMM.xml*,” under glucose minimal media conditions was downloaded from gcrd.ucsd.edu/Downloads/Cobra_Toolbox (Becker et al., 2007). This model includes 1266 reactions, 1061 metabolites, and 750 genes.

The *in silico* human metabolic reconstruction was downloaded from the BIGG database, *bigg.ucsd.edu* (Duarte et al., 2007). The network we used has 2766 metabolites and 1905 gene loci. The reduced model after removing the blocked reactions had 953 metabolites and 987 reactions. These metabolic fluxes were sampled under conditions similar to minimum essential medium in which rat hippocampal and human embryonic cells are grown (HEK) (Banker et al., 1998). The *in silico* medium contained amino acids (L-arginine, L-cysteine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-threonine, L-tryptophan, L-tyrosine, and L-valine) and vitamins (choline, nicotinamide, nicotinate, pyridoxal, riboflavin, and thiamin).

These models are summarized by a stoichiometry matrix, $S(m, r)$, that indicates the number of metabolites m consumed or produced by reaction r , and a matrix $E(r, e)$ of 1's and 0's indicating whether enzyme e catalyzes reaction r . The models also provide a maximum flux for each reaction.

4.2. Flux sampling in the feasible space

Flux balance analysis uses stoichiometry, steady-state assumptions, and additional constraints to limit the space of feasible fluxes (Price et al., 2004a). Flux vectors may be sampled using a random walk algorithm, artificial centering hit-and-run (Kaufman et al., 1998), to generate vectors of fluxes, v , in the feasible space (Thiele et al., 2005). This algorithm takes advantage of the convexity of the feasible space by using known feasible flux vectors to generate feasible displacements. The sampling of the feasible space was carried out in MATLAB using a version of the function “*sampleCbModel*” in the COBRA Toolbox (Becker et al., 2007), modified to avoid loading the sampled fluxes into memory. In a standard preparation step, maximum and minimum feasible values of each flux were determined, and fluxes with a maximum magnitude of 10^{-6} or smaller were removed from the model. The reduced model obtained after removing reactions had a stoichiometric matrix with 470 metabolites and 571 reactions in yeast, and 953 metabolites and 987 reactions in humans. We used 5000 warm-up points to initialize the sampling.

Yeast sampling was run for a total of 5 million steps returning 20,000 points for calculating correlation between each pair of enzymes. The human metabolic network was run for a total of 41 million steps returning 200,000 points for calculation of correlation between enzymes. This well-defined protocol yields a series of flux vectors with components $v_r(n)$, where r labels the reaction ($r = 1..571$ for yeast and $1..987$ for human) and n labels the sampling step ($n = 1..20,000$ for yeast and $1..200,000$ for human). This entire protocol, from random sampling through to calculating correlations, was conducted three times for each network using different random seeds, with no evidence of non-ergodic sampling among the three runs. Final predictions used correlations averaged over the three runs.

4.3. Transforming reaction fluxes to enzyme fluxes

While the flux sampling provides samples of reaction fluxes, we are interested in the flux contributed by each enzyme. This is complicated because a single reaction may be catalyzed by multiple enzymes, and a single enzyme may catalyze multiple non-equivalent reactions. This section describes how reaction fluxes were transformed to enzyme fluxes.

The weighted enzyme flux of an enzyme e_i , denoted $f_{e_i}(n)$ for flux sample n , was defined as the sum of weighted flux of all the reactions catalyzed by e_i . Each individual flux in the sum was in turn weighted by its contribution to the production of various metabolites. The overall definition of the weighted enzyme flux of an enzyme e_i catalyzing the reactions r was

$$f_{e_i}(n) = \sum_r \left(\frac{\sum_m |S(m, r) v_r(n)|}{\sum_{r'} |S(m, r') v_{r'}(n)|} \right) |v_r(n)| G(r, e_i) \quad (1)$$

The index m runs over metabolites, and $S(m, r)$ is the standard stoichiometry matrix. The factor $G(r, e)$ is introduced to reflect the fraction of flux through reaction r due to enzyme e . It is obtained by normalizing the matrix $E(r, e)$, with $E(r, e) = 1$ if e catalyzes r and 0 otherwise, as $G(r, e) = E(r, e) / \sum_{e'} E(r, e')$.

The correlation between enzymes e_i and e_j (denoted by $R_{e_i e_j}$) over the samples $n = 1, 2, \dots$, was then calculated for each independent run of the sampling algorithm as

$$R_{e_i e_j} = \frac{\sum_n [f_{e_i}(n) - \mu_{f_{e_i}(n)}][f_{e_j}(n) - \mu_{f_{e_j}(n)}]}{\sqrt{\sum_n [f_{e_i}(n) - \mu_{f_{e_i}(n)}]^2} \sqrt{\sum_n [f_{e_j}(n) - \mu_{f_{e_j}(n)}]^2}} \quad (2)$$

The mean of enzyme flux e_i is denoted $\mu_{f_{e_i}(n)}$, calculated separately for each run.

Reported results are for the average of the three runs. Comparisons of means and correlations within and between runs did not reveal any significant deviations from the null hypothesis of ergodicity (results not shown). All reported results are stable across runs, indicating that the results have converged.

4.4. Reaction sharing score and metabolite sharing score

We used a Bayesian score based on a 2×2 contingency table for calculating the reaction and metabolite sharing between two enzymes. Let n , n_1 , n_2 , and n_{12} represent the total number of reactions (or metabolites) in the network, the number connected to enzyme 1, the number connected to enzyme 2, and the number connected to both.

Under the null hypothesis, the probability of connection to both is the product of p_1 , the probability of connection to 1, and p_2 , the probability of connection to 2. The likelihood of the null model is calculated as the probability of the data integrated over all possible values of the parameters,

$$\begin{aligned} \text{Lik}(\text{null}) &= \int_0^1 dp_1 \int_0^1 dp_2 [p_1 p_2]^{n_{12}} [(1-p_1)p_2]^{n_2-n_{12}} \times [p_1(1-p_2)]^{n_1-n_{12}} [(1-p_1)(1-p_2)]^{n-n_1-n_2+n_{12}} \\ &= [(n+1)^2 C(n, n_1) C(n, n_2)]^{-1}. \end{aligned} \quad (3)$$

The alternative model permits either enhancement or depletion of the probability of connection to both 1 and 2, introducing a third parameter p_{12} . The requirement that probabilities across the 2×2 table sum to 1 creates a standard multinomial likelihood,

$$\begin{aligned} \text{Lik}(\text{alt}) &= \int_0^1 dp_1 \int_0^{1-p_1} dp_2 \int_0^{1-p_1-p_2} dp_{12} [p_1 p_2]^{n_{12}} [(1-p_1)p_2]^{n_2-n_{12}} \\ &\quad \times [p_1(1-p_2)]^{n_1-n_{12}} [(1-p_1)(1-p_2)]^{n-n_1-n_2+n_{12}} \\ &= n_{12}!(n_1-n_{12})!(n_2-n_{12})!(n-n_1-n_2+n_{12})!/(n+3)!. \end{aligned} \quad (4)$$

The score is then calculated as the log likelihood ratio of alternative to null, which reduces to

$$\text{score} = \log \left[\frac{(n+1)C(n, n_2)}{(n+2)(n+3)C(n_1, n_{12})C(n-n_1, n_2-n_{12})} \right] \quad (5)$$

The combinatorial factor $C(n, k)$ is $n!/k!(n-k)!$, or a Beta function for non-integer arguments. This score is analogous to a two-sided test because it increases when n_{12} is either larger or smaller than the value $n_1 n_2 / n$ expected under the null hypothesis. For n_{12} smaller than the null expectation, we used $\text{score}(n_{12}) - |\text{score}(n_{12}) - \text{score}(n_1 n_2 / n)|$ to restrict attention to enrichment.

4.5. Synthetic lethality data sources

Synthetic lethality data for this study was obtained from the BioGRID database (version 2.0.38) (Stark et al., 2006). There are 35 synthetic lethal and 33 synthetic growth defect interactions in the BioGRID database with both genes present in the yeast metabolic network.

4.6. Epistasis data sources

Scaled epistasis values were taken directly from the fitness values of single gene and double gene knockouts from a previous study (Segre et al., 2005). Supplementary data tables containing fitness values of single and double gene knockouts were downloaded from kishony.med.harvard.edu/prism/index.html. For calculating AUC and *F*-score from epistasis scores alone, only pairs with epistasis score data were considered. For logistic regression, epistasis scores for pairs in the FBA model but absent from the published tables were set to 0.

4.7. Performance metrics

The ROC curves, AUC, *F*-score, and PR curves were generated in *R* using the ROCR package (Sing et al., 2005). Logistic regression used the *R* glm function, and stepwise regression used the step function with BIC.

4.8. Phenotypic correlation data sources

Phenotypic correlations were taken from MimMiner scores generated by automated text mining (van Driel et al., 2006) based on the OMIM database (McKusick, 2007). The MimMiner similarity score matrix was downloaded from www.cmbi.ru.nl/MimMiner/suppl.html. The OMIM morbid map, available at <ftp://ftp.ncbi.nih.gov/repository/OMIM/morbidmap>, was used to match gene entries in the metabolic reconstruction to OMIM entries.

4.9. Data tables

Yeast and human data tables—including flux correlations, yeast SL and NSL pairs, human flux correlations, and human disease correlations—are freely available with no restrictions upon request to the authors and are posted at www.baderzone.org. Data tables are also available as Supplementary Information (see online Supplementary Material at www.liebertonline.com).

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF CAREER 0546446). J.S.B. acknowledges funding from the NIH and NSF. B.V. acknowledges support from the Center for Alternatives to Animal Testing. I.S.B. acknowledges helpful discussions with Nathan D. Price.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

- Almaas, E., Kovacs, B., Vicsek, T., et al. 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 427, 839–843.
- Banker, G., and Goslin, K., eds. 1998. *Culturing Nerve Cells*. MIT Press, Cambridge, MA.
- Becker, S.A., Feist, A.M., Mo, M.L., et al. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* 2, 727–738.
- Duarte, N.C., Becker, S.A., Jamshidi, N., et al. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* 104, 1777–1782.
- Duarte, N.C., Herrgard, M.J., and Palsson, B.O. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 14, 1298–1309.
- Forster, J., Famili, I., Fu, P., et al. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13, 244–253.

- Giaever, G., Chu, A.M., Ni, L., et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391.
- Kaufman, D.E., and Smith, R.L. 1998. Direction choice for accelerated convergence in hit-and-run sampling. *Oper. Res.* 46, 84–95.
- Kelley, R., and Ideker, T. 2005. Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* 23, 561–566.
- Lee, D.S., Park, J., Kay, K.A., et al. 2008. The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. USA* 105, 9880–9885.
- McKusick, V.A. 2007. Mendelian inheritance in man and its online version, OMIM. *Am. J. Hum. Genet.* 80, 588–604.
- Nakamura, H., Miura, K., Fukuda, Y., et al. 2000. Phosphatidylserine synthesis required for the maximal tryptophan transport activity in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 64, 167–172.
- Nyhan, W.L., James, J.A., Teberg, A.J., et al. 1969. A new disorder of purine metabolism with behavioral manifestations. *J. Pediatr.* 74, 20–27.
- Pan, X., Yuan, D.S., Xiang, D., et al. 2004. A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* 16, 487–496.
- Price, N.D., Reed, J.L., and Palsson, B.O. 2004a. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897.
- Price, N.D., Schellenberger, J., and Palsson, B.O. 2004b. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys. J.* 87, 2172–2186.
- Savinell, J.M., and Palsson, B.O. 1992. Optimal selection of metabolic fluxes for in vivo measurement. II. Application to *Escherichia coli* and hybridoma cell metabolism. *J. Theor. Biol.* 155, 215–242.
- Schuldiner, M., Collins, S.R., Thompson, N.J., et al. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123, 507–519.
- Segre, D., Deluna, A., Church, G.M., et al. 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* 37, 77–83.
- Shannon, P., Markiel, A., Ozier, O., et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Sing, T., Sander, O., Beerenwinkel, N., et al. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941.
- Song, S., and Friedmann, T. 2007. Tissue-specific aberrations of gene expression in HPRT-deficient mice: functional complexity in a monogenic disease? *Mol. Ther.* 15, 1432–1443.
- Stark, C., Breitkreutz, B.J., Reguly, T., et al. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539.
- Thiele, I., Price, N.D., Vo, T.D., et al. 2005. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J. Biol. Chem.* 280, 11683–11695.
- Tong, A.H., Lesage, G., Bader, G.D., et al. 2004. Global mapping of the yeast genetic interaction network. *Science* 303, 808–813.
- Ulitsky, I., and Shamir, R. 2007. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol. Syst. Biol.* 3, 104.
- van Driel, M.A., Bruggeman, J., Vriend, G., et al. 2006. A text-mining analysis of the human phenome. *Eur. J. Hum. Genet.* 14, 535–542.
- Waterham, H.R. 2002. Inherited disorders of cholesterol biosynthesis. *Clin. Genet.* 61, 393–403.
- Wu, X., Jiang, R., Zhang, M.Q., et al. 2008. Network-based global inference of human disease genes. *Mol. Syst. Biol.* 4, 189.
- Ye, P., Peyser, B.D., Pan, X., et al. 2005. Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol. Syst. Biol.* 1, 2005.0026.

Address reprint requests to:

Dr. Joel S. Bader
Department of Biomedical Engineering
Johns Hopkins University
Baltimore, MD 21218

E-mail: joel.bader@jhu.edu