

Predicting selective drug targets in cancer through metabolic networks

Ori Folger¹, Livnat Jerby¹, Christian Frezza², Eyal Gottlieb², Eytan Ruppin^{1,3,5,*} and Tomer Shlomi^{4,5,*}

¹ The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel, ² Cancer Research UK, The Beatson Institute for Cancer Research, Glasgow, Scotland, UK, ³ The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel and ⁴ Computer Science Department, Technion, Israel Institute of Technology, Haifa, Israel

⁵ These authors contributed equally to this work

* Corresponding authors. E Ruppin or T Shlomi, School of Computer Sciences, School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. Tel.: + 97 23 640 6528; Fax: + 97 23 640 9357; E-mail: ruppin@post.tau.ac.il or tomer.shlomi@cs.technion.ac.il

Received 2.11.10; accepted 12.4.11

The interest in studying metabolic alterations in cancer and their potential role as novel targets for therapy has been rejuvenated in recent years. Here, we report the development of the first genome-scale network model of cancer metabolism, validated by correctly identifying genes essential for cellular proliferation in cancer cell lines. The model predicts 52 cytostatic drug targets, of which 40% are targeted by known, approved or experimental anticancer drugs, and the rest are new. It further predicts combinations of synthetic lethal drug targets, whose synergy is validated using available drug efficacy and gene expression measurements across the NCI-60 cancer cell line collection. Finally, potential selective treatments for specific cancers that depend on cancer type-specific downregulation of gene expression and somatic mutations are compiled.

Molecular Systems Biology 7: 501; published online 21 June 2011; doi:10.1038/msb.2011.35

Subject Categories: metabolic and regulatory networks; simulation and data analysis

Keywords: cancer; metabolic; metabolism; modeling; selectivity

Introduction

The interest in studying cancer metabolism has recently grown in light of the decreasing number of newly released anticancer drugs, the development of metabolite profiling technologies and metabolic network databases, and due to increased efforts to target metabolic diseases by the pharmaceutical industry. During tumor development, cancer cells modify their metabolism to meet the requirements of cellular proliferation, thus facilitating the uptake and conversion of nutrients into biomass (Vander Heiden *et al*, 2009). Numerous studies have shown similar metabolic alterations occurring across tumor cells, including changes in glucose metabolism that give rise to the Warburg effect, and an increase in biosynthetic activities (such as nucleotide, lipids and amino-acid synthesis), important for cellular proliferation (DeBerardinis *et al*, 2008; Tennant *et al*, 2009; Vander Heiden *et al*, 2009). The latter are regulated by several signaling pathways implicated in cell proliferation, in which cancer-associated mutations lead to accelerated growth (Christofk *et al*, 2008). The similarity in metabolic activity across cancer types is evident in experimental measurements of metabolic flux and from analyzing multiple high-throughput molecular data sources, including gene expression, metabolomics and phenotypic assays of cell line growth following shRNA-facilitated gene silencing (knockdown) (Supplementary Information). Metabolic commonalities across tumors have important practical implications, where drugs, such as antimetabolites that target

nucleotide biosynthesis are used to fight a variety of different malignancies.

The observation that many types of cancer cells adapt their metabolism to facilitate biomass formation to enable proliferation suggests that it may be possible to predict characteristic alterations in cancer metabolism via genome-scale computational modeling approaches that have been successfully used in the past to predict the metabolic state of fast-growing microorganisms (Price *et al*, 2004). In particular, flux balance analysis (FBA) is a constraint-based modeling (CBM) approach that is suitable for modeling cancer metabolism as it assumes that a cell is under selective pressure to increase its growth rate, and hence searches for metabolic flux distributions that produce essential biomass precursors at high rates (Price *et al*, 2003). A fundamental step toward large-scale modeling of human metabolism has been taken in recent studies (Duarte *et al*, 2007; Ma *et al*, 2007), which reconstructed a generic (non-tissue specific) human metabolic network based on an extensive evaluation of genomic and bibliomic data. The potential clinical utility of a human metabolic model has already been demonstrated, showing its ability to identify reactions causally related to hemolytic anemia (Duarte *et al*, 2007), to predict potential drug targets for treating hypercholesterolemia (Duarte *et al*, 2007), disease co-morbidity (Lee *et al*, 2008) and tissue specificity of disease genes (Shlomi *et al*, 2008), in identifying diagnostic biomarkers for inborn errors of metabolism (Shlomi *et al*, 2008). A recent study by Li *et al* (2010) have further utilized the

human metabolic model as a basis for the prediction of novel targets for known anticancer drugs (Li *et al*, 2010).

Results

A genome-scale model of cancer metabolism correctly predicts cancer growth-supporting genes

Here, we construct the first large-scale model of cancer metabolism that aims to capture the main metabolic functions common across many cancer types. To this end, we integrate the human metabolic model of Duarte *et al* (2007) with cancer gene expression data, utilizing a variant of our recent computational method for the automatic reconstruction of human tissue metabolic models, termed MBA (Model Building Algorithm) (Jerby *et al*, 2010) (see Materials and methods). In brief, we begin by assembling an initial core set of 197 metabolic enzyme-coding genes that are highly expressed across 90% of all cancer cell lines in the NCI-60 collection (Grever *et al*, 1992). Then, applying MBA, a minimal set of additional reactions from the human metabolic model that are needed to activate the reactions associated with this initial core set is added, obtaining a cancer metabolism model that is consistent, that is, each of the core reactions can potentially carry a non-zero metabolic flux, within a global flux distribution satisfying stoichiometric, mass balance and reaction directionality constraints. The minimal set of added reactions was chosen such that it enables the steady-state production of biomass compounds required for cellular proliferation (based on prior knowledge of their relative concentrations; see Materials and methods). This model reconstruction approach, accounting for both a core set of highly expressed genes along with a cellular proliferation constraint extends upon previous model reconstruction methods (Shlomi *et al*, 2008; Jerby *et al*, 2010). The resulting cancer metabolic model includes 772 reactions and 683 genes (see Materials and methods; Supplementary Table S8). Applying FBA to this model enables the prediction of the metabolic state of cancer cells (their proliferation rate and flux distribution) across different gene knockdowns and modeling the effects of drug applications on a large scale.

To validate the cancer network model generated, we analyzed it via FBA to predict growth-supporting genes whose knockdown would significantly reduce cellular proliferation rate (see Materials and methods), that is, we knocked down all genes one at a time, and recorded the resulting simulated growth rate in the model. Overall, we obtain a set of 199 growth-supporting genes, which are reassuringly ranked as highly essential based on shRNA gene silencing data (Luo *et al*, 2008) (Kolmogorov–Smirnov (KS) P -value=0.0045; Supplementary Figure S1). The reference shRNA data set consists of a list of genes, ranked according to the survival rate of 12 cancer cell lines after these genes are knocked down, thus denoting the genes' experimentally measured contribution to cancer growth. As a further validation, we compared the predicted set of growth-supporting genes with a set of genes in which cancer somatic mutations have been reported in the literature (Futreal *et al*, 2004), finding a high level of enrichment (hypergeometric P -value of 0.002). For comparison, we evaluated

the predictive performance of the gene expression data and the human metabolic model separately, finding that either source is a worse predictor of both shRNA gene essentiality (P -values are 0.075 and 0.029, respectively) and cancer mutations (P -values are 0.0077 and 0.025, respectively) compared with the generic cancer model (Materials and methods; Supplementary Information).

While we described above the reconstruction of a *generic* cancer metabolic model, capturing major metabolic alterations that are common across cancer types, a similar computational approach can be utilized to develop more fine-tuned metabolic models of specific cancer types. To demonstrate this potential, we applied our cancer model building approach to reconstruct a model of non-small cell lung cancer metabolism utilizing multiple gene expression data sets (Supplementary Information). The growth-supporting genes predicted by the lung cancer model are ranked as highly essential based on shRNA gene silencing data measured in this cell line (KS P -value=0.025), outperforming the predictions made by the generic cancer model for this cell line (Supplementary Information).

Predicting cytostatic anticancer targets

To identify which of the above-identified growth-supporting genes may be considered viable anticancer targets, we further aimed to predict whether their knockdown is expected to be toxic to non-dividing cells or damage the proliferation of normal cells. To simulate the effect of these gene knockdowns on normal, non-dividing cells, we predicted the potential damage to ATP production (utilizing the entire human metabolic network), a vital biochemical function that must be preserved in every cell. Based on these predictions, we define a *cytostatic score* for each gene, representing the extent to which its knockdown reduces cancer growth compared with its effect on ATP production in healthy cells (see Materials and methods) (with a *cytostatic score* of 1 denoting a non-toxic target that completely eliminates cancer growth without affecting ATP production in healthy cells). The resulting distribution of *cytostatic scores* is bimodal; out of the 199 genes that are predicted to be growth supporting in the cancer model, 52 have a high cytostatic score (above 0.9), and the remaining 147 genes have a low score (below 0.6; Figure 1A; Supplementary Table S2; Supplementary Information). As an additional method to predict the effect of these knockdowns on the metabolism of healthy, non-dividing cells, we tested the effects of knockdowns in a model of liver metabolism (Jerby *et al*, 2010) (specifically on normal urea secretion, glycogenesis, glycogenolysis, gluconeogenesis and bilirubin uptake), ruling out one of these drug targets (CMPK1) as potentially damaging normal liver uptake of bilirubin (see Materials and methods; Supplementary Table S2). Notably, in the absence of detailed metabolic networks for a wide range of different human tissues, it is currently impossible to strictly rule out that the predicted targets would not damage metabolic functions of other healthy tissues. As a final screening step, we studied how gene knockdowns would affect proliferation of healthy cells. To this end, we applied a standard FBA analysis on the entire human metabolic network model, aiming to identify growth-supporting genes as done for the cancer model. We found that

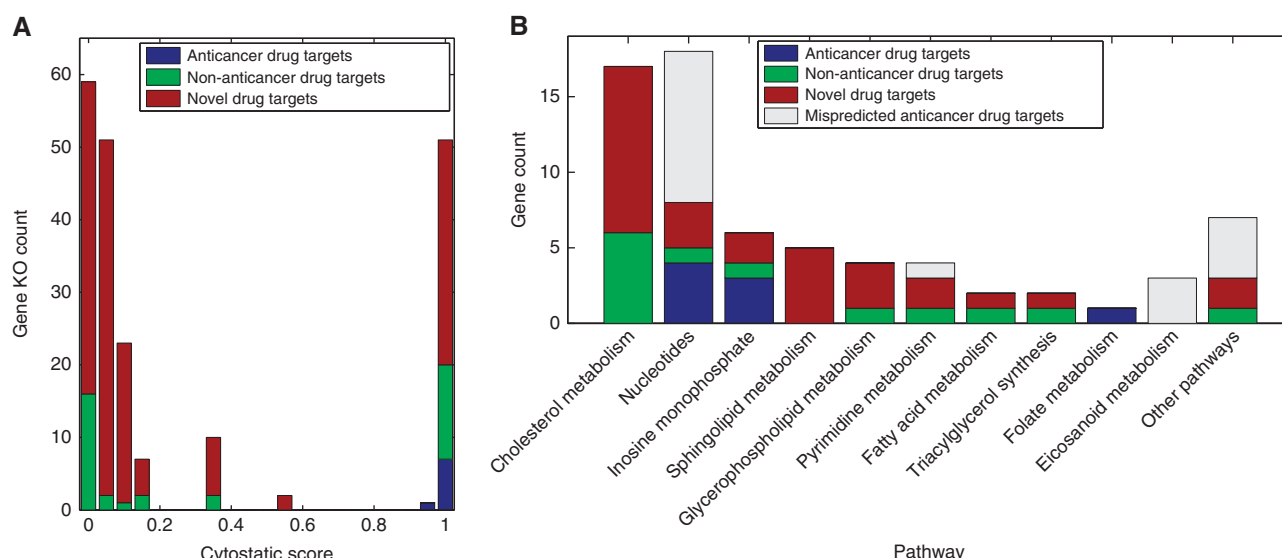


Figure 1 Cancer selectivity and pathway association of predicted growth-supporting genes. **(A)** Distribution of selectivity scores for the set of 199 predicted growth-supporting genes. **(B)** Pathway association of the highly cytostatic growth-supporting genes (cytostatic score > 0.9), showing for each pathway the number of predicted genes that are known targets of current anticancer drugs, the number of known targets of drugs that are currently used for non-cancer indications and entirely novel gene targets, that is, genes without any currently known drugs that target them. For each pathway, the number of missed predictions, that is, known anticancer drug targets that are not predicted to be highly selective, is also shown.

the knockdown of 49 out of the 52 high cytostatic scoring targets is likely to also damage proliferation of normal cells (Supplementary Table S2), suggesting that the targeting of these genes would cause similar side effects to current cytostatic drugs (Partridge *et al*, 2001).

The 52 targets with high cancer cytostatic scores contain 8 out of 24 known targets of the 14 FDA-approved metabolic anticancer drugs found in DrugBank (Wishart *et al*, 2008) (Supplementary Tables S1, S2 and S5), representing a highly significant enrichment (hypergeometric P -value $< 6 \times 10^{-7}$), testifying to predictive power of the model. Interestingly, the model misses the prediction of the knockdown effects on proliferation of 16 known drug targets. These include four genes involved in the metabolism of signaling molecules (e.g. aminoglutethimide, which targets a hormone-signaling mechanism), whose effect on cellular growth is outside the scope of the model; seven additional genes that are targeted by antimetabolites (Mercaptopurine, Thioguanine and Capecitabine) having multiple metabolic targets (Supplementary Table S5). Finally, four additional genes are correctly predicted as part of synthetic lethal simulations described below, such that overall, out of all the metabolically active drug targets currently known, only one missed prediction remains unaccounted for. Moving from the gene to the drug level and simulating drug treatments by concurrently knocking down multiple genes targeted by the same drug, we successfully identify all three antimetabolites as having a high cytostatic score, correctly predicting 11 (78%) out of all 14 FDA-approved metabolic anticancer drugs (Supplementary Table S5).

Interestingly, the set of 52 predicted highly cytostatic genes contains 13 additional genes that are targeted by existing non-cancer drugs. Remarkably, all of these predicted drugs but one are currently under experimental testing for cancer therapy

(Supplementary Table S2): eight of these genes are targeted by drugs already approved by the FDA for the treatment of various non-cancer diseases (e.g. antimalarial treatment, hypercholesterolemia, obesity, etc.) and five genes are targeted by experimental drugs. The remaining 31 (out of the original 52) cytostatic genes are currently without any known drug that currently targets them, and form interesting candidates for potential drug targeting. Notably, these novel drug targets span a much broader set of potential anticancer target pathways; yet, reassuringly many of them participate in the same metabolic pathways as the known and experimental anticancer drug targets (Figure 1B; Supplementary Table S2). For example, we predict five highly selective drug targets in sphingolipids metabolism (a subclass of fatty acid metabolism; currently not targeted by any FDA-approved drugs), whose targeting was recently shown to represent a promising new approach to treat pancreatic cancers (Guillermet-Guibert *et al*, 2009).

Predicting synthetic lethal gene targets

The concept of synthetic lethality promises to have a central role in the future selection of anticancer drug targets (Kaelin, 2005). Synthetic lethal genes are common in metabolic networks due to their highly robust nature (resulting from backup mechanisms such as isozymes and alternative pathways; Stelling *et al*, 2002), and their identification via FBA was successfully demonstrated in microbial networks (Deutscher *et al*, 2006; Harrison *et al*, 2007). To study synthetic lethal drug targets, we systematically simulated all double gene knockdowns in the cancer model. We assigned each gene pair with a *synergy score*, reflecting the additional drop in proliferation rate below the minimal rate achieved by its individual single knockdowns (in accordance with the commonly

used 'highest single agent' synergy model; Berenbaum, 1989). The analysis resulted in 342 synthetic lethal gene pairs with a synergy score >0.5 .

As a validation for the plausibility of the predicted synergistic gene pairs, we find that they are significantly enriched with genetic interactions between the corresponding yeast orthologs (hypergeometric P -value=0.028; see Materials and methods), considering a large-scale data set of genetic interactions in yeast (Costanzo *et al*, 2010). Notably, the significance of the enrichment is quite remarkable, considering that the conservation level of genetic interactions between distant species is considered to be rather low (as demonstrated, e.g., between *Saccharomyces cerevisiae* and *Caenorhabditis elegans*; Tischler *et al*, 2008). As a second validation, we investigated whether drugs that target a single gene (participating in a predicted synthetic lethal pair) indeed have higher efficacy in cell lines in which the synergistic gene is lowly expressed. To this end, we analyzed together growth inhibition measurements for 11 metabolic drugs for which such data exists (Lee *et al*, 2007) and gene expression measurements over all NCI-60 cell lines (Grever *et al*, 1992) (Figure 2A). We find that drugs that target one of the genes in a predicted pair indeed exert a significantly stronger inhibitory effect on proliferation in the cell lines in which their paired target gene has a low expression level (P -value=0.02), supporting our predictions (Figure 2B).

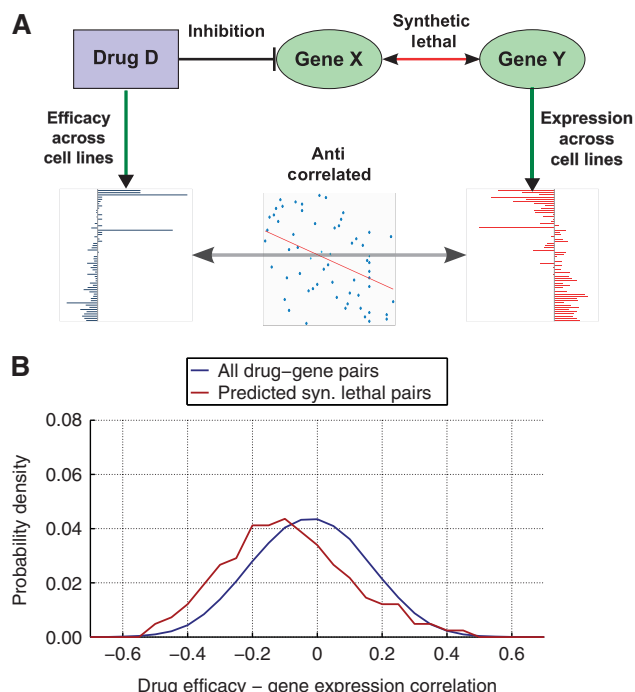


Figure 2 The relation between drugs' efficacy and the expression of their paired target genes. **(A)** A schematic illustration of the expected relation between an efficacy of a drug and the expression of genes that have a synthetic lethal interaction with the drug's target. **(B)** The distribution of efficacy-expression correlation values for the set of predicted synthetic lethal drug-gene pairs, and for a background distribution of random pairs. The predicted synthetic lethal pairs show a significant anticorrelation compared with the random pairs (Wilcoxon P -value=0.02).

The targeting of both synthetic lethal genes via combination therapy

To identify which of the synthetic lethal genes may be further considered for combinatorial drug therapies, we further predicted whether their joint knockdown is expected to be toxic for non-dividing cells or damage the proliferation of normal cells (as performed above for the single targets). We find that 133 gene pairs have high cytostatic score (>0.6 ; computed when they are both knocked down; Figure 3A; see Materials and methods), suggesting (within the screening limitations pointed above) that their double knockdown may not be toxic to non-dividing cells. Exploration of the metabolic pathways in which the predicted highly cytostatic gene pairs participate (Figure 3B; Supplementary Table S4) reveals that both the pentose-phosphate pathway (PPP) and pyruvate metabolism have a central role, with 45% of the gene pairs having at least one target in these pathways. PPP is involved in the production of both NADPH (necessary for lipid and nucleotide biosynthesis as well as protection against oxidative stress) and ribose 5-phosphate (a nucleotide precursor), making it an attractive target for cancer therapy, though currently there are still no inhibitors for this pathway in clinical trials. In pyruvate metabolism, pyruvate carboxylase (PC; participating in 13 pairs) is notable; it is involved in one of the two major pathways that contribute to anaplerosis (i.e. the replenishment of TCA cycle intermediates), the other pathway being glutaminolysis. Indeed, several previous attempts have already been made to target glutaminolysis (Rosenfeld and Roberts, 1981), further supporting the plausibility of targeting PC as well.

Predicting how the knockdown of the synthetic lethal targets would affect proliferation of normal cells, we find that in a marked contrast to the single target analysis, the knockdown of 99 of the pairs is not expected to adversely damage growth of normal cells (Supplementary Table S3). The fact that synthetic lethal genes are predicted to achieve selective targeting of cancer cells is in accordance with previous experimental findings (Lehár *et al*, 2009). For example, we predict that the double knockdown of glycine hydroxymethyltransferase (SHMT1) and alanine-glyoxylate transaminase (AGXT) to differentially inhibit proliferation of cancer versus normal cells. While these knockdowns would block glycine biosynthesis in cancer cells (from serine and glyoxylate, respectively; Supplementary Figure S2A), their knockdown in normal cells is predicted to spare glycine production through an alternative pathway involving choline degradation. The latter pathway was not included in the cancer model due to low expression level of its member genes (while choline utilization in the cancer model proceeds via choline kinase toward phosphocholine biosynthesis, in agreement with experimental data; Nakagami *et al*, 1999). Other examples include synthetic lethal genes that are predicted to differentially inhibit nutrient uptake rates in cancer versus normal cells, due to the exclusion of several lowly expressed membrane transporters in the cancer model (Supplementary Figure S2B and C).

Predicting the effect of the 133 double gene knockdowns on liver metabolism resulted in eight pairs whose targeting could

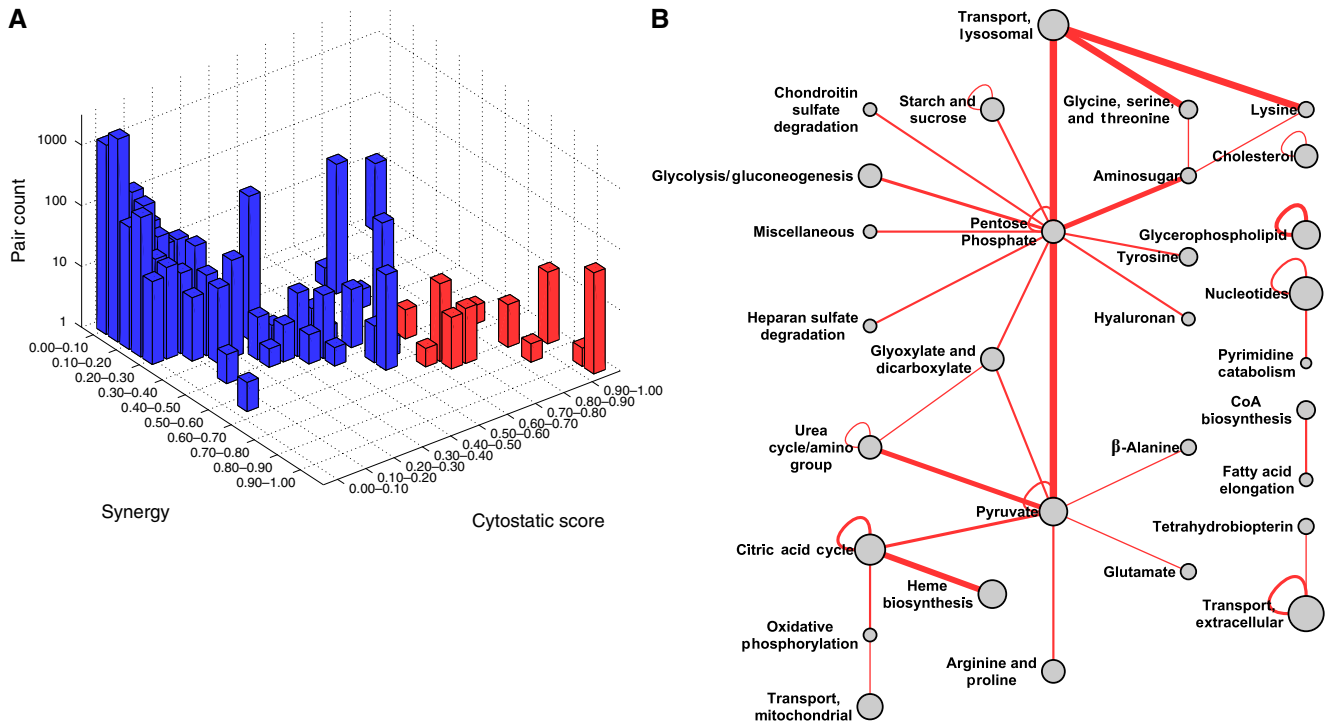


Figure 3 Cytostatic scores, synergy and pathway association of predicted target gene pairs. **(A)** The distribution of predicted drug target pairs across their synergy and selectivity scores. Highly synergistic and selective drug target pairs are colored red, while the remaining ones are colored blue. **(B)** A network of pathway associations of the predicted synergistic and highly selective drug target pairs (Supplementary Table S4). Node size represents the number of genes that participate in a pathway. Edge width represents the number of synergistic and selective pairs involving genes in the two connected pathways (its nodes).

potentially damage various liver functions (Supplementary Table S3). Hence, as noted above for single targets, the current lack of genome-scale metabolic models for various tissues prevents the prediction of other potential sources of cellular damage that may be caused by these knockdowns—and it would be reasonable to assume that many of these pairs would be actually labeled as toxic as new tissue models are developed.

The targeting of a gene whose synthetic lethal partner is inactivated in specific cancer types leads to selective treatments

The specific targeting of a gene participating in a synergistic pair is especially appealing in tumors in which its interacting gene is specifically inactivated. The targeting of such a gene solely is likely to selectively damage the tumor, without affecting the function of healthy tissues in which the interacting gene is not inactivated and hence can still back up for the targeted gene. Notably, this strategy can achieve therapeutic selectivity without requiring the ability to computationally predict the effect of double knockdowns on healthy tissues. The inactivation of one of the interacting genes specifically in cancer cells may result from cancer type-specific regulation of metabolic activity, or from the dysfunction of metabolic enzymes due to cancer somatic mutations. Next, we utilized genomic and transcriptomic data to infer gene inactivation across an array of cancers, which leads to the

identification of cancer type-specific targets based on the predicted synergistic gene pairs.

Germline mutations in two metabolic enzymes, succinate dehydrogenase (SDH) and fumarate hydratase (FH) have been previously found to be causally implicated in oncogenesis (Futreal *et al*, 2004). In both cases, the malignancy was characterized by somatic loss of the corresponding wild-type allele, resulting in complete loss of enzymatic activity. Our model predicts SDH to be synthetic lethal with PC, showing that either one of these enzymes is required to satisfy TCA anaplerotic demands essential for the synthesis of aspartate (that is used to further synthesize arginine and pyrimidine nucleotides) from oxaloacetate (OAA) (Figure 4A). OAA can be produced either by PC from pyruvate or through glutaminolysis that produces α -ketoglutarate (aKG), which is converted to OAA through several TCA cycle reactions. SDH dysfunction blocks this anaplerotic flow from glutaminolysis toward OAA. Hence, inhibition of PC is predicted to specifically inhibit proliferation of SDH-deficient cancer cells, without damaging normal cells. Several experimental drugs that target PC are currently available and may potentially be used for treating SDH-deficient cancers such as paragangliomas and pheochromocytomas (Frezza *et al*, 2011). A similar argument regarding synthetic lethality between SDH and PC also holds for FH and PC, as FH deficiency is also expected to block metabolic flux toward OAA. Our model predicts that FH forms synthetic lethality with nine additional genes (Supplementary Table S3), four of which already have drugs that target them. These pairs can hence be potentially used to target specific

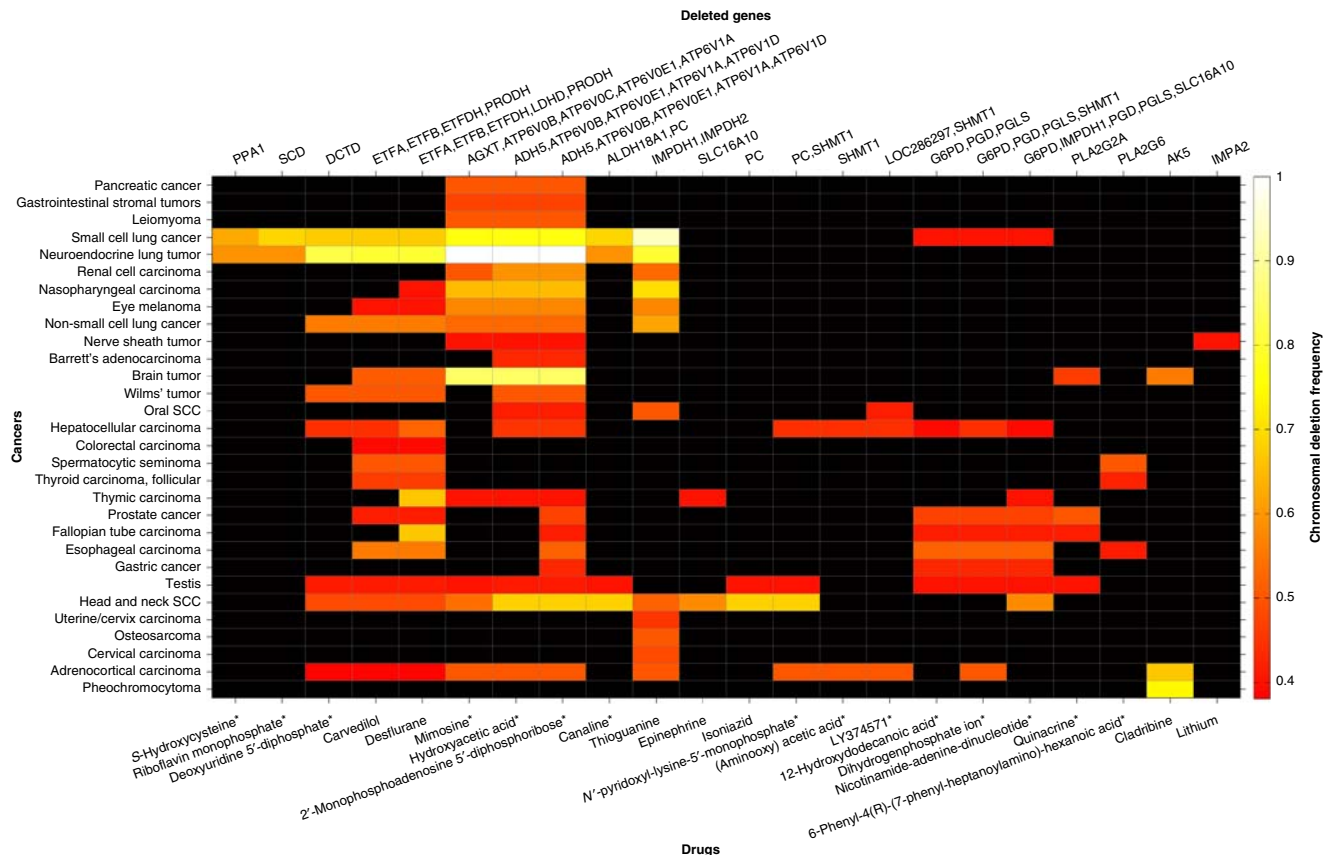


Figure 5 Drugs predicted to target specific cancer types based on chromosomal loss of synthetic lethal participant genes. Specific drugs may be effective in treating specific types of cancer, based on our predictions of selective synthetic lethal gene pairs. Cancer types that show a high frequency (in yellow and white) of chromosomal deletions of specific genes are susceptible to drugs inhibiting the genes' synthetic lethal complements. Experimental drugs are followed by an asterisk. In cases where multiple drugs target the same gene, only a single representative drug is shown here with the remaining drugs specified in Supplementary Table S9.

suggesting that the targeting either of its corresponding synthetic lethal partners should be useful for selectively inhibiting proliferation of these specific cancers. Notably, the experimental drug hydroxyacetic acid was previously shown to inhibit G6PD activity, suggesting its potential usage for treating the above-mentioned cancer types. Figure 5 shows the association of approved and experimental drugs with cancer types, in accordance with cancer-type chromosomal deletions of synthetic lethal genes (Supplementary Table S9).

Analyzing gene expression data of healthy human tissues taken from Su *et al* (2004), we have found that out of the 342 predicted synthetic lethal targets, 72 pairs have at least one gene with a significantly low expression level in a target cancer type and not in healthy tissues, spanning leukemia, melanoma, lung, colon, CNS, ovarian, renal, prostate and breast cancer (see Materials and methods). Taken together, these results testify to the promising utility of this approach for identifying new potential treatment targets for specific cancers.

Discussion

As a further experimental support for the plausibility of the predicted targets, a follow-up work by Frezza (under review)

studied the predicted synergy (described above; Figure 2) between the TCA cycle enzyme FH and the heme metabolism pathway. This synergy is of a particular interest both from a practical perspective, as FH loss-of-function is associated with Hereditary Leiomyomatosis and Renal-Cell Cancer (HLRCC) (Tomlinson *et al*, 2002) and from a basic science perspective as, currently, no mechanism that explains the capability of these cells to survive without FH (and hence without a functional TCA cycle) has been provided. Notably, the synergy with heme metabolism predicted by the cancer model is not evident based on the gene expression data measured under FH-deficient cells solely, and cannot be gleaned from analyzing the human metabolic model of Duarte *et al* (2007). In the study by Frezza (under review), newly characterized genetically modified kidney mouse cells carrying a conditionally targeted FH allele were used to confirm that targeting heme metabolism would render FH-deficient cells non-viable while not damaging the control wild-type FH-containing cells. This was found to result from a linear pathway involving the biosynthesis and degradation of heme, which enables FH-deficient cells to utilize the accumulated TCA cycle metabolites and permits partial mitochondrial NADH production. These results confirm the suggested targeting of heme metabolism to treat HLRCC patients, providing an interesting specific demonstration of

the utility of using our generic cancer model, and the usage of synthetic lethality as a mean for specifically targeting cancer cells. Future validation of predicted targets could involve specific shRNA silencing of predicted single and synergistic targets, across different cancer cell lines. Additionally, predicted repurposing and combinations of approved and experimental drugs can be easily tested across cell lines, measuring the reduction in proliferation rate following their treatment. Finally, of particular interest in this respect are ‘hubs’ participating in many predicted SLs, whose knockdown could have large functional effects.

Beyond the prediction of new potential drug treatments, the modeling approach presented here is expected to open up many additional exciting possibilities in the near future. First, data on gene regulation may be integrated with the stoichiometric model using existing approaches (Covert *et al*, 2004; Shlomi *et al*, 2007) to further enhance the model’s accuracy. Second, the forthcoming introduction of a new version of the human metabolic model may further increase the accuracy of the resulting cancer models. Third, the reconstruction of human tissue-specific models, such as the recently developed models for liver (Jerby *et al*, 2010) and kidney (Chang *et al*, 2010), can lead to better computational predictors of drug selectivity. A major current difficulty in the reconstruction of such models involves the definition of a metabolic ‘objective function’ for normal tissues. Fourth, future studies could probe the functional effects of the gradual knockdown of reactions, aiming to capture a richer repertoire of enzymatic inhibition and gene knockdown. Finally, individual cancer cell line models may lead to the prediction of specific cancer biomarkers, building on existing CBM methods (Shlomi *et al*, 2008). In summary, the model presented here lays down a fundamental computational counterpart for interpreting the rapidly accumulating proteomics (Bichsel *et al*, 2001) and metabolomics (Fan *et al*, 2009) data characterizing cancer metabolic alterations, and paves the way both for obtaining a systems level understanding of cancer metabolism and for designing new therapeutic means that selectively target them.

Materials and methods

Reconstructing a human cancer metabolic model

To reconstruct a cancer metabolic model in humans, we collected gene expression data from all cancer cell lines in the NCI-60 collection (Grever *et al*, 1992), a collection of microarray experiments in 59 cancer cell lines of various types performed in a controlled and consistent manner with a single platform and within a single laboratory, and assembled a core set of 197 cancer genes that are highly expressed (with intensity above 200) in >90% of cell line measurements (while the results of our analysis remain highly robust to different definitions of core gene sets; Supplementary Information). Assuming that reactions associated with these highly expressed cancer genes are likely to be activated in cancer cells, we applied the following approach: our goal is to extract from the human metabolic model of Duarte *et al* (2007), the most parsimonious, consistent model, that includes this core reactions set and enables cellular proliferation. A metabolic network model is considered consistent if it enables the activation all of its reactions—that is for each reaction there exist a feasible flux distribution in the model (satisfying stoichiometric mass balance and directionality constraints) in which this reaction is carrying a non-zero flux. Bearing this definition, we employ a greedy search heuristic approach that is based on iteratively pruning reactions from the human metabolic model in a random order, while maintaining the consistency of the

pruned model with regard to its core reactions set (starting from a generic version of the human metabolic model that is consistent). In each pruning step, a reaction is removed from the cancer model being built only if its removal does not prevent the activation of the reactions in the core reactions set. Since the reactions’ scanning order may affect the resulting model, the algorithm is executed with different, random pruning orders (1000 in our implementation) to construct multiple candidate models. The fraction of models that contain a certain reaction reflects the latter’s confidence score, which indicates the extent to which it should be included in the final cancer model. Hence, to construct the final cancer model, the candidate models are aggregated, starting from core reactions set and iteratively adding reactions ordered by their confidence scores, until a final, minimal but consistent with the core reactions, model is obtained. A similar approach has been concurrently used to develop a model of healthy liver metabolism (Jerby *et al*, 2010) (Supplementary Information). Notably, blocked reactions from the model of Duarte *et al* were removed before the application of the MBA algorithm.

To predict gene contribution to biomass production, we added a growth reaction to the resulting model, representing the steady-state consumption of biomass compounds required for cellular proliferation. The stoichiometric coefficients of the growth reaction represent the relative molecular concentrations of 42 essential metabolites, including nucleotides, deoxynucleotides, amino acids, lipids, etc. in human tissues. These relative concentrations are calculated based on data regarding mass composition of liver and muscle tissues (Supplementary Table S6). A sensitivity analysis shows that the prediction performance of the cancer model is highly insensitive to the specific definition of biomass composition (Supplementary Information). In all simulations, we assume a standard RPMI-1640 medium, in accordance with the medium used in our reference shRNA experimental data set (Luo *et al*, 2008; Supplementary Table S7). The biomass production rate predicted by the cancer model is 40% lower than that predicted by the human network model, reflecting that the two models are indeed functionally different. Notably, the generic human model does not represent a concrete cell-type (but rather a collection of reactions that take place within different cell types), and hence its predicted optimal biomass production rate does not accurately represent a rate that is achievable by a specific cell type. By definition, the maximal biomass production rate in the cancer model cannot be higher than that achievable in the generic human model as the cancer model consists of a subset of the reactions of generic model.

FBA was used to simulate the effects of gene knockdowns on the biomass production rates of the proliferating cells (Förster *et al*, 2003). Genes whose knockdown reduces the maximal biomass production rate in >1% were considered to be growth supporting. Notably, the results presented in the paper are insensitive to the specific choice of threshold for defining growth reduction, as the majority of the knockouts (99.13%) result in either no reduction of growth or its complete elimination (Supplementary Information).

Validating predicted growth-supporting genes based on shRNA data

Luo *et al* (2008) provide a ranking of gene probes that are targeted by shRNA, based on the effect that their knockdown exerts on cancer cellular proliferation. To check whether a set of predicted growth-supporting genes are ranked as highly essential for cancer proliferation, we computed a one-sided KS statistic, comparing the ranking distribution of the predicted gene set probes with the ranking distribution of the remaining probes. Significance is assessed by comparing the derived KS statistic with similar statistics computed for 10 000 randomly selected gene sets of the same size (in accordance with the statistical method underlying Gene Set Enrichment Analysis statistic; Subramanian *et al*, 2005).

Computing cytostatic scores for single and double drug target predictions

A *cytostatic score* of a gene represents the extent to which its knockdown obliterates cancer growth compared with its effect on

ATP production in the normal, human metabolic model. The gene knockdown effect on growth rate is computed by applying FBA on the cancer model, denoting by WT_{growth} and KO_{growth} the growth rate before and after the knockdown, respectively. The gene knockdown effect on ATP production is computed by applying FBA on the human metabolic model, measuring the reduction in maximal achievable ATP production rate, denoting by WT_{atp} and KO_{atp} the maximal ATP production rate before and after the knockdown, respectively. The cytostatic score is defined as $(KO_{\text{atp}}/WT_{\text{atp}}) \times (1 - KO_{\text{growth}}/WT_{\text{growth}})$. A cytostatic score of 1 denotes a highly selective knockdown that completely eliminates cancer growth without affecting ATP production of healthy cells, and a score of 0 denotes a highly non-selective knockdown. The cytostatic score for synergistic gene pairs is computed in an analogous manner, by simulating the effect of double knockdowns on growth and ATP production rates.

Predicting and validating synergistic drug targets

We define a *synergy score* for a gene pair as the additional drop in growth rate below the minimal rate achieved by the single knockdowns. Specifically, denoting by KO_A , KO_B and KO_{AB} , the growth rates following the knockout of gene A, gene B and the joint knockout of genes A and B, respectively, the synergy score is defined as: $KO_{AB}/\min(KO_A, KO_B)$ (in accordance with the commonly used ‘highest single agent’ synergy model; Berenbaum, 1989). The validation of the predicted synergistic gene targets via yeast orthologs was based on an orthology mapping by Berglund *et al* (2007). Among the yeast genes that participate in the high-throughput genetic interaction screen (Costanzo *et al*, 2010), we identified 75 genes that are orthologous to human genes that participate in predicted synergistic targets.

The correlation between drug efficacy and gene expression was computed as following: for each gene pair in which one gene is targeted by drug D and the other one is denoted G, we computed the Spearman correlation between the effective growth inhibition concentrations (GI-50) of D with the expression levels of G across the 60 cell lines. The *P*-value was computed by a Wilcoxon test, comparing the distribution of Spearman correlations for all predicted synergetic drug-gene pairs, with the distribution of Spearman correlations for a large random sample of drug-gene pairs.

To identify synergistic gene pairs that consist of genes with significantly low expression levels specifically in cancer cells, we obtained gene expression data from 75 healthy human tissues (Su *et al*, 2004) and compared it with the NCI-60 cancer cell line expression data (Grever *et al*, 1992). First, we normalized the tissue expression data set to the same mean and s.d. as the NCI-60 cancer cell line expression data. Then, we applied Wilcoxon test to identify genes with significantly low expression level in at least one out of eight cancer types represented in NCI-60, followed by false discovery rate correction for multiple testing.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We are grateful to Martin Kupiec and Eithan Galun for providing very valuable comments on the manuscript. ER and TS are supported by grants from the Israel Science Foundation (ISF) and the Israeli Cancer Research Foundation (ICRF). EG and CF are supported by Cancer Research UK (CRUK) and CF is supported by a Long-Term Fellowship from the European Molecular Biology Organization (EMBO).

Author contributions: OF, LJ, ER, and TS conceived and designed the research. OF, ER and TS wrote the paper. OF and TS performed the analysis and the statistical computations. CF and EG designed and performed the follow-up experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Ariketh D, Nelson R, Vance JE (2008) Defining the importance of phosphatidylserine synthase-1 (PSS1): unexpected viability of PSS1-deficient mice. *J Biol Chem* **283**: 12888–12897
- Berenbaum MC (1989) What is synergy? *Pharmacol Rev* **41**: 93–141
- Berglund A, Sjolund E, Ostlund G, Sonnhammer E (2007) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* **36** (Database issue): D263–D266
- Bichsel VE, Liotta LA, Petricoin III EF (2001) Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer J* **7**: 69–78
- Chang R, Xie L, Bourne P, Palsson B, Dunbrack R (2010) Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol* **6**: e1000938
- Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC (2008) The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**: 230–233
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M *et al* (2010) The genetic landscape of a cell. *Science* **327**: 425–431
- Covert M, Knight E, Reed J, Herrgard M, Palsson B (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**: 92–96
- DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB (2008) The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab* **7**: 11–20
- Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat Genet* **38**: 993–998
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson B (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* **104**: 1777–1782
- Fan TW, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM (2009) Altered regulation of metabolic pathways in human lung cancer discerned by 13 C stable isotope-resolved metabolomics (SIRM). *Mol Cancer* **8**: 41
- Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA (2009) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* **38** (Database issue): D652–D657
- Frezza C, Pollard PJ, Gottlieb E (2011) Inborn and acquired metabolic defects in cancer. *J Mol Med* **89**: 213–220
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR (2004) A census of human cancer genes. *Nat Rev Cancer* **4**: 177–183
- Förster J, Famili I, Palsson B, Nielsen J (2003) Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*. *Omics A J Integr Biol* **7**: 193–202
- Grever MR, Schepartz SA, Chabner BA (1992) The National Cancer Institute: cancer drug discovery and development program. *Semin Oncol* **19**: 622–638
- Guillemet-Guibert J, Davenne L, Pchejetski D, Saint-Laurent N, Brizuela L, Guilbeau-Frugier C, Delisle MB, Cuvillier O, Susini C, Bousquet C (2009) Targeting the sphingolipid metabolism to defeat pancreatic cancer cell resistance to the chemotherapeutic gemcitabine drug. *Mol Cancer Ther* **8**: 809–820
- Harrison R, Papp B, Pal C, Oliver SG, Delneri D (2007) Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci USA* **104**: 2307–2312
- Huret J, Senon S, Bernheim A, Dessen P (2004) An Atlas on genes and chromosomes in oncology and haematology. *Cell Mol Biol* **50**: 805–807

- Jerby L, Shlomi T, Ruppin E (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol* **6**: 1
- Kaelin Jr WG (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer* **5**: 689–698
- Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási AL (2008) The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci USA* **105**: 9880–9885
- Lee JK, Havaleshko DM, Cho H, Weinstein JN, Kaldjian EP, Karpovich J, Grimshaw A, Theodorescu D (2007) A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc Natl Acad Sci* **104**: 13086
- Lehár J, Krueger AS, Avery W, Heilbut AM, Johansen LM, Price ER, Rickles RJ, Short 3rd GF, Staunton JE, Jin X, Lee MS, Zimmermann GR, Borisy AA (2009) Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat Biotechnol* **27**: 659–666
- Li L, Zhou X, Ching WK, Wang P (2010) Predicting enzyme targets for cancer drugs by profiling human metabolic reactions in NCI-60 cell lines. *BMC Bioinformatics* **11**: 501
- Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhi R, Weir BA, Mermel C, Barbie DA, Awad T, Zhou X, Nguyen T, Piqani B, Li C, Golub TR, Meyerson M, Hacohen N *et al* (2008) Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci USA* **105**: 20380–20385
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* **3**: 135
- Nakagami K, Uchida T, Ohwada S, Koibuchi Y, Suda Y, Sekine T, Morishita Y (1999) Increased choline kinase activity and elevated phosphocholine levels in human colon cancer. *Cancer Sci* **90**: 419–424
- Partridge AH, Burstein HJ, Winer EP (2001) Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J Natl Cancer Inst Monogr* **2001**: 135–142
- Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* **21**: 162–169
- Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2**: 886–897
- Rosenfeld H, Roberts J (1981) Enhancement of antitumor activity of glutamine antagonists 6-diazo-5-oxo-L-norleucine and acivicin in cell culture by glutaminase-asparaginase. *Cancer Res* **41**: 1324–1328
- Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* **26**: 1003–1010
- Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* **3**: 101
- Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**: 190–193
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* **101**: 6062
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**: 15545
- Tennant DA, Duran RV, Boulahbel H, Gottlieb E (2009) Metabolic transformation in cancer. *Carcinogenesis* **30**: 1269–1280
- Tischler J, Lehner B, Fraser AG (2008) Evolutionary plasticity of genetic interaction networks. *Nat Genet* **40**: 390–391
- Tomlinson IP, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, Leigh I, Gorman P, Lamlum H, Rahman S, Roylance RR, Olpin S, Bevan S, Barker K, Hearle N, Houlston RS, Kiuru M, Lehtonen R, Karhu A, Vilkkii S *et al* (2002) Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat Genet* **30**: 406–410
- Vance DE, Vance JE (2009) Physiological consequences of disruption of mammalian phospholipid biosynthetic genes. *J Lipid Res* **50**: S132
- Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**: 1029–1033
- Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**: D901–D906
- Yogev O, Yogev O, Singer E, Shaulian E, Goldberg M, Fox TD, Pines O (2010) Fumarate: a mitochondrial metabolic enzyme and a cytosolic/nuclear component of the DNA damage response. *PLoS Biol* **8**: e1000328



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.