

RESEARCH ARTICLE

Open Access



# Inferring cancer dependencies on metabolic genes from large-scale genetic screens

Shoval Lagziel<sup>1</sup>, Won Dong Lee<sup>2</sup> and Tomer Shlomi<sup>1,2,3\*</sup>

## Abstract

**Background:** Cancer cells reprogram their metabolism to survive and propagate. Thus, targeting metabolic rewiring in tumors is a promising therapeutic strategy. Genome-wide RNAi and CRISPR screens are powerful tools for identifying genes essential for cancer cell proliferation and survival. Integrating loss-of-function genetic screens with genomics and transcriptomics datasets reveals molecular mechanisms that underlie cancer cell dependence on specific genes; though explaining cell line-specific essentiality of metabolic genes was recently shown to be especially challenging.

**Results:** We find that variability in tissue culture medium between cell lines in a genetic screen is a major confounding factor affecting cell line-specific essentiality of metabolic genes—while, quite surprisingly, not being previously accounted for. Additionally, we find that altered expression level of a metabolic gene in a certain cell line is less indicative of its essentiality than for other genes. However, cell line-specific essentiality of metabolic genes is significantly correlated with changes in the expression of neighboring enzymes in the metabolic network. Utilizing a machine learning method that accounts for tissue culture media and functional association between neighboring enzymes, we generated predictive models for cancer cell line-specific dependence on 162 metabolic genes (representing a ~ 2.2-fold increase compared to previous studies). The generated predictive models reveal numerous novel associations between molecular features and cell line-specific dependency on metabolic genes. Specifically, we demonstrate how cancer cell dependence on one-carbon metabolic enzymes is explained based on cancer lineage, oncogenic mutations, and RNA expression of neighboring enzymes.

**Conclusions:** Considering culture media as well as accounting for molecular features of functionally related metabolic enzymes in a metabolic network significantly improves our understanding of cancer cell line-specific dependence on metabolic genes. We expect our approach and predictive models of metabolic gene essentiality to be a useful tool for investigating metabolic abnormalities in cancer.

**Keywords:** Cancer metabolism, Gene-silencing screens, RNAi, CRISPR, Tumor microenvironment, Tissue culture medium, Metabolic networks

## Background

Metabolic reprogramming is an emerging hallmark of cancer [1]. Evidence for alterations in metabolic activity in malignant cells goes back almost a century ago to the discovery of the “Warburg effect” [2]. The recent resurgence of interest in the field of cancer metabolism involves numerous findings of metabolic alterations in a variety of

pathways in cancer cells [3]. These metabolic adaptations are directly regulated by growth-promoting signaling pathways, in which somatic mutations drive tumorigenesis [4]. Multiple drugs targeting metabolic enzymes have been used in clinics for decades, and many promising agents are under clinical trials [5–7]. Major ongoing research is aimed to identify tumor-specific reliance on metabolic enzymes that are limiting for cancer progression and to exploit these liabilities to treat cancer [5].

Genome-scale RNAi and CRISPR-based gene silencing-screens are powerful tools to identify essential

\* Correspondence: [tomersh@cs.technion.ac.il](mailto:tomersh@cs.technion.ac.il)

<sup>1</sup>Faculty of Computer Science, Technion, Haifa, Israel

<sup>2</sup>Faculty of Biology, Technion, Haifa, Israel

Full list of author information is available at the end of the article



genes for cancer cell proliferation and survival [8–13]. Recent pan-cancer loss-of-function screens of hundreds of cancer cell lines have identified novel oncogenes involved in cell signaling pathways. Integration of measured gene essentiality profiles with genomic features (DNA mutations, copy number variations, and RNA expression) was shown to provide a mechanistic understanding and predictive models for tumor-specific gene essentiality [14–16].

Explaining cell line dependence specifically on metabolic enzyme-coding genes appears to be particularly challenging. For instance, in a recent large-scale RNAi screen by Tsherniak et al. [14], the number of generated predictive models for cell line-specific dependence on metabolic genes was significantly lower (by ~30%) than for non-metabolic genes (hypergeometric  $p$  value  $< 10^{-4}$ ). In the genetic screen by McDonald et al. [15], a similar difficulty to predict cell line-specific dependence on metabolic genes was shown. A notable example of cell line-specific dependency on a metabolic enzyme-coding gene that is explainable based on gene expression is that of TYMS (thymidylate synthase) [15], where TYMS was found to be essential in cell lines where its expression level is low (hence, cells may be limited for thymidylate needed for DNA biosynthesis) while the expression of TYMP (thymidine phosphorylase) is high (further depleting thymidylate pools).

Here, we aimed to generate predictive models for cell line-specific dependence on metabolic genes based on genomic features, providing insight into the underlying molecular mechanisms. Considering that cellular metabolism is highly dependent on the availability of metabolic nutrients in the environment [17–19], we further accounted for the variability in tissue culture medium used for different cell lines in the genetic screens—a confounding factor that appears to have a major effect on cell line-specific essentiality of metabolic genes, thought quite surprisingly, has not been previously accounted for. We show that considering culture media as well as accounting for the functional association between metabolic enzymes via a metabolic network model enables generating predictive models for cancer cell line-specific dependence on a ~2.2-fold higher number of metabolic genes than in previous studies.

## Results

### Metabolic gene essentiality depends on tissue culture medium

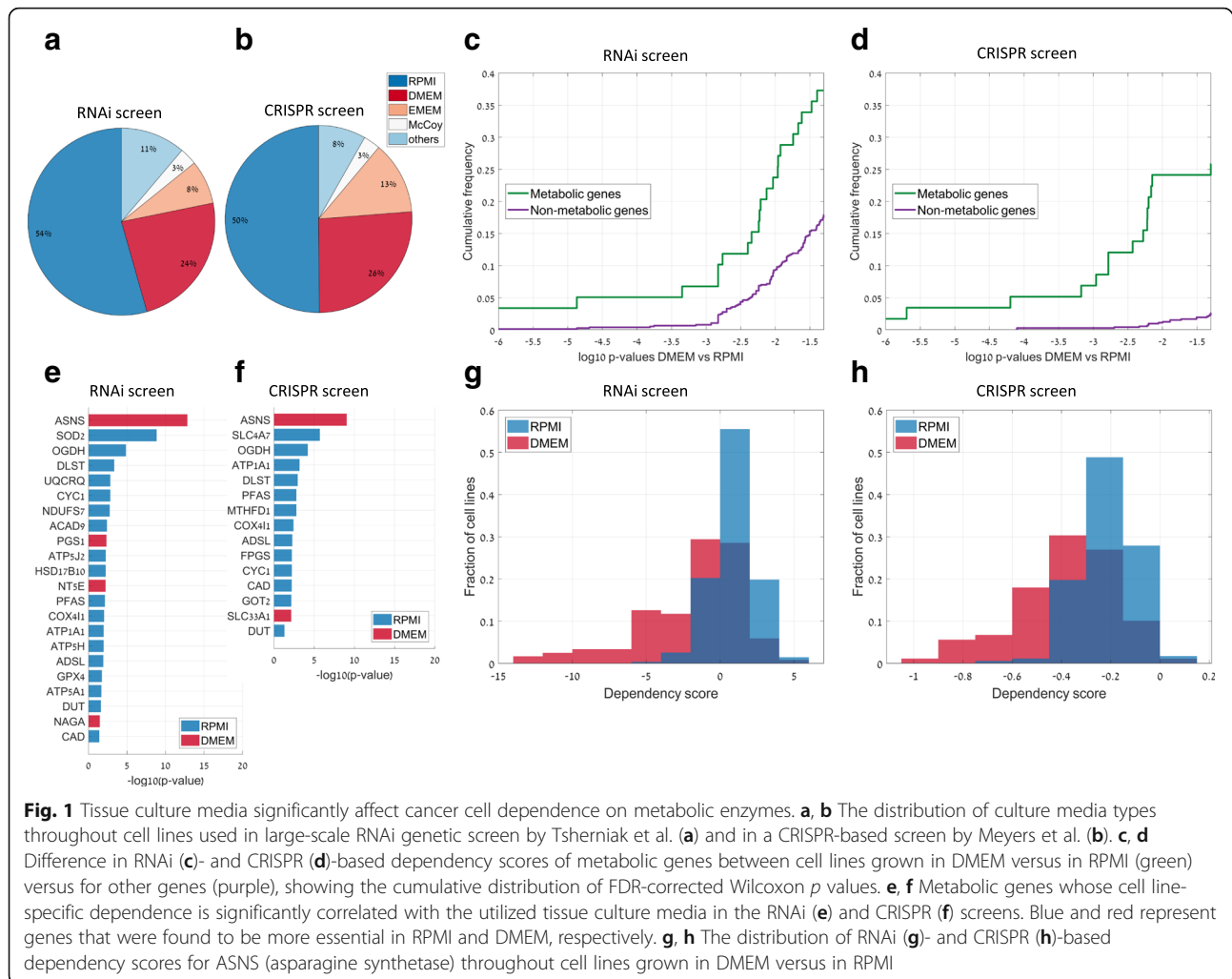
Considering that metabolic activity is highly dependent on nutrient availability, we hypothesized that variability in culture media used for different cell lines in genetic screens may have a profound effect on cellular dependence on specific metabolic genes. To test this hypothesis, we utilized data from recent large-scale RNAi and CRISPR screens (covering 501 and 341 cell lines, respectively, as part of the

Cancer Dependency Map) [14, 16]. Analyzing the reported information on tissue culture media in both screens showed that multiple media were used for different cell lines within each study (Fig. 1a, b). The two most commonly used media in both studies were RPMI and DMEM, in which  $> 50\%$  and  $> 20\%$  of the cell lines were cultured, respectively. RPMI and DMEM differ in terms of their composition of inorganic salts, amino acids, vitamins, and metabolic nutrients. For instance, RPMI includes several non-essential amino acids such as proline, aspartate, and asparagine which are absent from DMEM [20].

We compared the distribution of gene dependency scores of metabolic genes for cell lines grown in DMEM and RPMI (considering genes included in the metabolic network model Recon 2.2 [21]). We focused on metabolic genes whose dependency scores showed major variation throughout the cell lines, considering those whose dependency score in at least one cell line is lower by more than six standard deviations from the mean of each gene as in [14] (a low dependency score of a gene in a certain cell line reflects increased essentiality). We find that the RNAi-based dependency scores of 37% (22/59) of the analyzed metabolic genes are significantly associated with the utilized culture media (FDR-corrected Wilcoxon  $p$  value  $< 0.05$ ) versus 18% of the non-metabolic genes (135/754, Fig. 1c). Similarly, we find that utilized culture media are significantly correlated with CRISPR-based dependency scores for a markedly higher percent of the metabolic genes (26%, 15/58 genes) versus other genes (3%, 19/716 genes; Fig. 1d). The metabolic gene that showed the strongest dependence on culture media in both RNAi and CRISPR screens was ASNS (asparagine synthase, Fig. 1e, f) and found to be significantly more essential in cells grown in DMEM (Fig. 1g, h; FDR-corrected Wilcoxon  $p$  value  $< 10^{-12}$  and  $10^{-9}$  for the RNAi and CRISPR screens, respectively, comparing the distribution of dependency scores in cell lines grown in DMEM and RPMI). This is readily explainable by the lack of asparagine in DMEM, rendering ASNS essential for de novo asparagine biosynthesis under this culture media. Furthermore, ASNS was found to be significantly more important in cell lines grown in DMEM than those grown with other asparagine containing media, McCoy (FDR-corrected Wilcoxon  $p$  value  $< 10^{-3}$  for both RNAi and CRISPR screens), and Hams F12 ( $p$  value  $< 10^{-6}$  for both screens).

Overall, we found a significantly stronger association between metabolic gene dependency scores and culture media versus that of non-metabolic genes and media (Wilcoxon  $p$  value  $< 10^{-2}$  for the RNAi screen and  $p$  value  $< 10^{-4}$  for the CRISPR screen, comparing the distribution of FDR-corrected Wilcoxon  $p$  values computed above for metabolic versus non-metabolic genes; Fig. 1b).

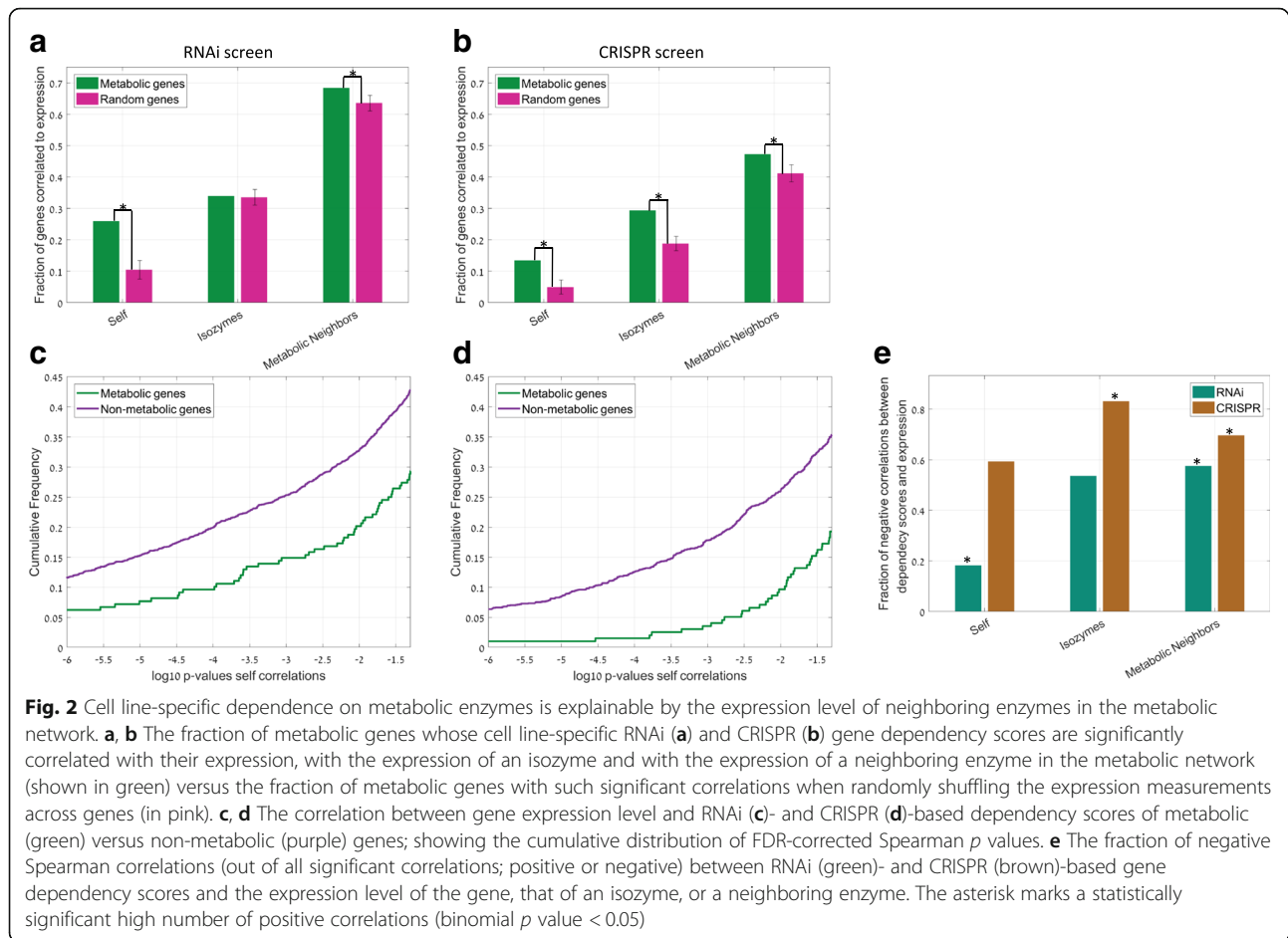
We next examined whether the identified correlation between metabolic gene dependency scores and culture media can be explained by variation in cell culture type



(i.e., adherent versus suspension) or cancer lineage. Notably, DMEM was initially developed for culturing adherent cells [22] while RPMI was introduced for suspension cells [23]. Indeed, we found that  $\sim 90\%$  and  $\sim 30\%$  of the adherent and suspension cell lines from the genetic screens were grown in DMEM and RPMI, respectively. However, the correlation between metabolic gene dependency scores and media type remained highly significant when controlling for the cell culture type; for  $\sim 90\%$  of the metabolic genes whose RNAi- or CRISPR-based dependency score correlate with media type, the correlation remains significant when controlling for cell culture type (FDR-corrected Spearman partial correlation  $p$  value  $< 0.05$ ). Likewise, the correlation between gene dependency scores and media type remained significant for 64% of the cases when controlling for cancer lineage (see the “Methods” section; FDR-corrected  $p$  value  $< 0.05$ ). Taken together, our results suggest that culture media is an important confounding factor affecting metabolic gene essentiality in RNAi- and CRISPR-based screens.

### Metabolic gene essentiality depends on the expression level of neighboring enzymes in the metabolic network

We explored whether cell line-specific dependence on metabolic genes is explainable based on altered RNA expression level. We find that RNAi- and CRISPR-based dependency scores of 26% and 14% of the metabolic genes, respectively, are significantly correlated with their expression level across cell lines (FDR-corrected Spearman correlation  $p$  value  $< 0.05$ ; with the  $r$  value having an either positive or negative sign). This is significantly higher than expected by chance, when considering correlations between gene dependency scores and the expression of randomly chosen genes ( $p$  value  $< 0.05$ , Fig. 2a, b). However, the fraction of metabolic genes whose cell line-specific dependence is predictable by their expression level is significantly lower (by  $\sim 40\%$ ) than for non-metabolic genes (Wilcoxon  $p$  value  $< 10^{-4}$  for the RNAi screen and  $p$  value  $< 10^{-7}$  for the CRISPR screen; comparing the distribution of FDR-corrected Wilcoxon  $p$  values computed for metabolic versus



non-metabolic genes; Fig. 2c, d). Interestingly, RNAi-based gene dependency scores and gene expressions are negatively correlated for a significantly small fraction of the metabolic genes (in  $\sim 20\%$  of the cases; binomial  $p$  value  $< 10^{-5}$ , assuming a probability of 0.5 for a positive correlation; Fig. 2e); that is, a negative correlation representing high essentiality in cell lines in which their expression level is high. With CRISPR, on the other hand, metabolic gene dependency scores were negatively correlated with the gene expression in  $\sim 60\%$  of the cases. These results suggest that RNAi-based knockdown is especially effective in cell lines where the expression level of the target metabolic gene is already low, while CRISPR-based gene knockout affects cell viability regardless of the gene expression level.

Considering that metabolic flux typically depends on the collective activity of multiple interconnected enzymes (rather than the expression of a single enzyme), we hypothesized that cell line dependence on a specific metabolic gene might be explained by altered expression of neighboring enzymes in the metabolic network. To test this hypothesis, we computed the Spearman correlation between the RNAi- or CRISPR-based metabolic gene dependency scores and

expression levels of isozymes and neighboring enzymes in the metabolic network (utilizing Recon 2.02) [21]. Specifically, we consider *neighboring enzymes* as those producing or consuming a shared metabolite, excluding those sharing highly prevalent currency metabolites such as energy or redox co-factors (see the “Methods” section). We find that the RNAi- and CRISPR-based dependency scores of a significantly high fraction of the metabolic genes are correlated with the expression level of at least one isozyme (for CRISPR screen) and with the expression of at least one neighboring enzyme (for both RNAi and CRISPR screens,  $p$  value  $< 0.05$ ; Fig. 2a, b). Accordingly, we find higher Spearman correlations between gene dependency scores and the expression level of isozymes versus with other genes ( $p$  value  $< 10^{-2}$  for the RNAi screen and  $p$  value  $< 10^{-26}$  for the CRISPR screen) and with neighboring enzymes versus with other genes ( $p$  value  $< 10^{-6}$  for the RNAi screen and  $p$  value  $< 10^{-26}$  for the CRISPR screen; Wilcoxon test comparing the two distributions of absolute Spearman  $r$  values). A significantly high negative correlation between both RNAi- and CRISPR-based gene dependency scores and the expression levels of neighboring enzymes (binomial  $p$  value  $< 10^{-5}$ ) represent increased

reliance on metabolic genes in cell lines having induced expression level of interconnected enzymes.

#### Generating predictive models for cell line-specific dependence on metabolic genes accounting for culture medium and neighboring enzymes

Having found a significant correlation between metabolic gene essentiality and culture medium, and with the expression of functionally related enzymes in the metabolic network, we aimed to generate improved predictive models for cell line-specific dependence on metabolic genes. For each metabolic gene, we build a predictive model considering expression and copy number of functionally related metabolic genes (i.e., considering isozymes and neighboring enzymes in the metabolic network), somatic mutation of 125 known oncogenes/tumor suppressors [24], and culture media type. We utilized a bootstrap-aggregated random forest algorithm [25, 26] (similar to [14]; see the “Methods” section). The goodness of fit was evaluated based on the Pearson correlation between out-of-bag predicted and measured knockdown effects, estimating statistical significance via a random permutation test of dependency scores (see the “Methods” section).

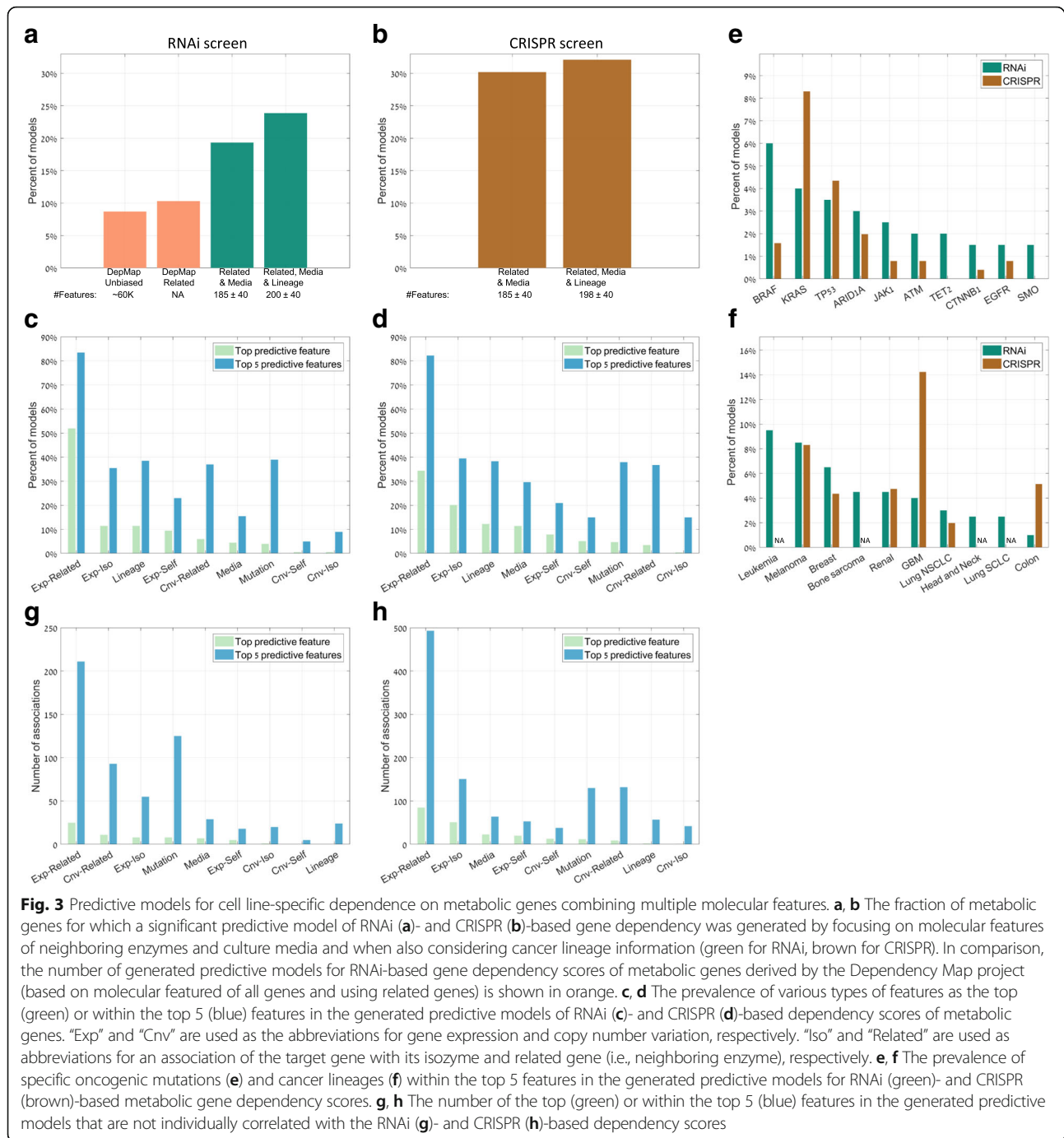
Our analysis resulted in significant predictive models for RNAi-based dependency scores for 19% of the analyzed metabolic genes (162 out of 838) (FDR-corrected  $p$  value  $< 0.05$ ; Fig. 3a), with a mean Spearman correlation between predicted and measured dependency scores of  $\sim 0.37$ . This represents  $\sim 2.2$ -fold higher number of metabolic genes with a significant predictive model than those previously found via an unbiased approach as part of the Dependency Map project [14]. Notably, Tsherniak et al. repeated their analysis focusing on molecular features of functionally related genes (protein-protein interaction [27] or members of a protein complex [28]) to reduce the feature space and avoid overfitting. However, this increased the number of metabolic genes for which significant predictive models were generated by only 11% (while increasing the total number of genes with a predictive model by up to 19%). Hence, while focusing on molecular features of functionally related genes is clearly useful to increase the predictive power of gene dependence, the functional association of metabolic genes is better captured by our neighboring enzymes measure. Further considering that cellular metabolism is highly dependent on cancer lineage [29, 30], accounting for the information on cancer lineage markedly increased the number of significant predictive models for metabolic gene dependency to 24% of the metabolic genes (Fig. 3a, Additional file 1). Similarly, our analysis resulted in significant predictive models for CRISPR-based dependency scores for 32% of the metabolic genes (Fig. 3b, Additional files 2, 3, 4 and 5).

Next, we explored which type of molecular feature contributes the most to explaining metabolic gene dependency by performing permutation-based feature importance test on the derived predictive models from RNAi and CRISPR data (see the “Methods” section). We find that the expression level of neighboring enzymes is the top predictive feature of gene dependency in both RNAi ( $\sim 50\%$ ) and CRISPR (34%) screens (Fig. 3c, d). Surprisingly, the expression level of the gene itself is the top predictive feature of RNAi- and CRISPR-based dependency scores in only  $\sim 10\%$  and 8% of the metabolic genes, respectively, emphasizing the importance of analyzing the expression of interconnected enzymes in the metabolic network. Media type is the top predictive feature for another  $\sim 5\%$  and  $\sim 11\%$  of the metabolic genes with the RNAi and CRISPR screens, respectively. In comparison, oncogenic mutations were found to be the top predictive feature of RNAi- and CRISPR-based gene dependency for less than 5% of the metabolic genes (though oncogenic mutations have important predictive value when combined with other molecular features; among the top 5 predictive features of up to 40% of the metabolic genes).

Among the known oncogenic mutations, BRAF and KRAS mutations are the most frequently observed alterations within the top 5 predictive features of the RNAi-based (6%) and CRISPR-based (8%) metabolic gene dependency (Fig. 3e), followed by TP53 ( $\sim 4\%$ ) for both screens. This is in agreement with major metabolic reprogramming known to be regulated by oncogenic mutations in these genes [31–34]. Among cancer lineages, leukemia is the most frequent cancer type within the top 5 predictive features of the RNAi screen (10%), followed by melanoma in both screens (8%, Fig. 3f).

Integrating multiple molecular and cancer lineage features within a predictive model for gene dependency reveals numerous important predictive features that cannot otherwise be inferred when analyzing each feature individually (Fig. 3g, h). For example, for 18 metabolic genes, we find that their gene expression level is an important predictive feature of RNAi-based cell line dependence (within the top 5 predictive features), while no significant correlation is found between their expression and dependency score across all cell lines. We find that the top predictive feature of RNAi-based cell line dependence for DAD1 (involved in *N*-glycan synthesis) is its expression level, while no significant correlation is found between DAD1 expression level and its dependency score across all cell lines (Spearman  $r = 0.03$ ,  $p$  value = 0.57). Considering that the mutation status of NOTCH1 is also within the top 5 predictive features of DAD1, we find a significant correlation between DAD1 expression and its dependency score when focusing on NOTCH1-mutated cell lines (Spearman  $r = -0.35$ ,  $p$

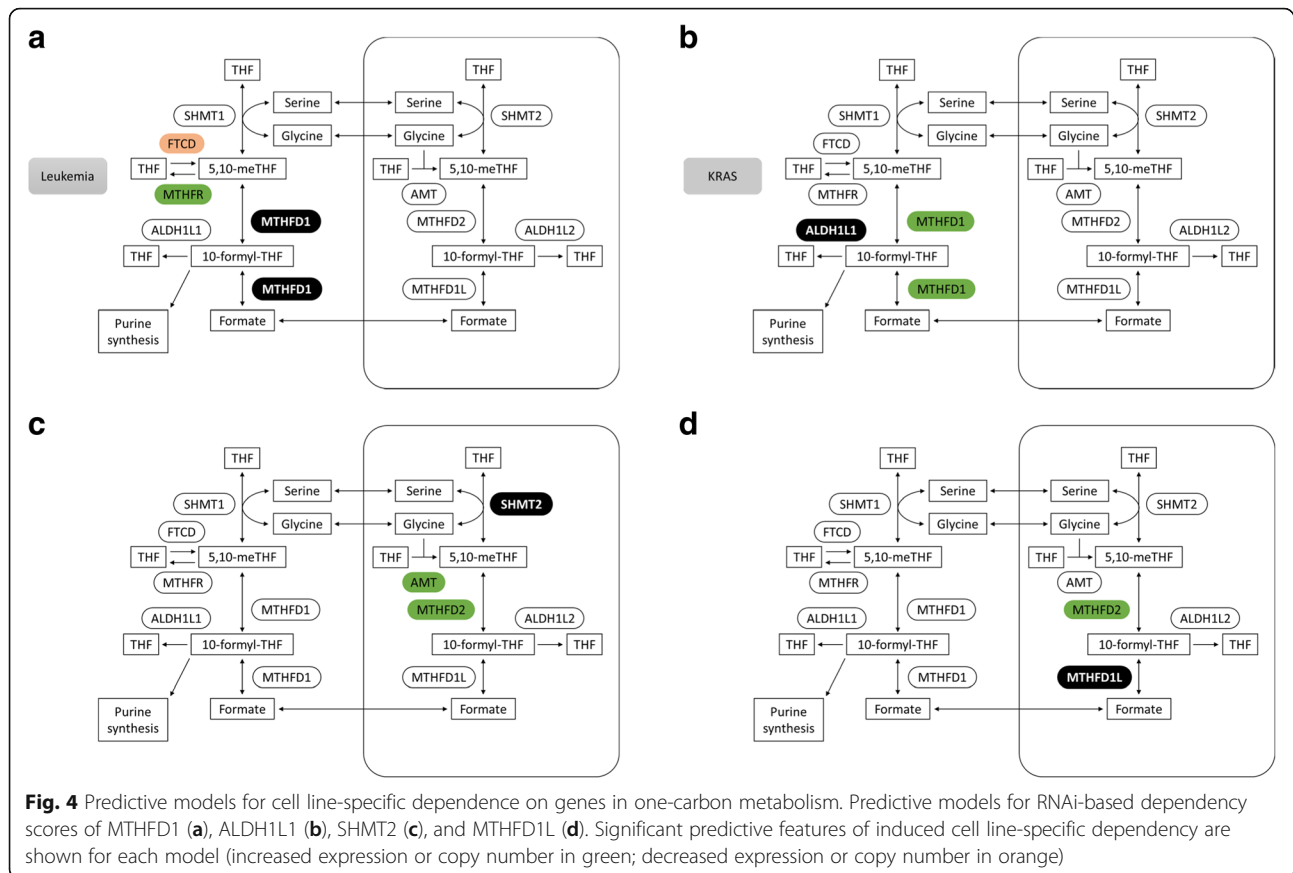




value < 0.05). This is consistent with the known important role of glycosylation in Notch signaling [35], suggesting that in cancer cell driven by NOTCH1 mutation, high DAD1 expression level indicates high dependence on its function.

The generated predictive models enable raising hypotheses regarding specific metabolic mechanisms that underlie cell line-specific dependence on metabolic genes. For example, we present predictive models generated for RNAi-based

dependence on genes involved in one-carbon metabolism. This metabolic system has long been a target for chemotherapy and has recently been shown to involve substantial metabolic rewiring that can be targeted for therapeutic purposes [36–41]. We find an increased dependency of leukemia cell lines on MTHFD1 (cytosolic methylene-tetrahydrofolate dehydrogenase) where the expression of MTHFR (methylene-THF reductase) that consumes 5,10-methylene-THF is high, and the expression of FTCD (formimidoyltransferase



cyclodeaminase) that produces 5,10-methylene-THF as part of histidine catabolism is low (Fig. 4a). This suggests that maintaining the high activity of MTHFD1 is required to overcome a potential depletion of its substrate 5,10-methylene-THF in these conditions. Increased dependence on ALDH1L1 (cytosolic formyltetrahydrofolate dehydrogenase) is found in KRAS-mutated cell lines where MTHFD1 copy number is high, potentially to facilitate the regeneration of THF from 10-formyl-THF (Fig. 4b). Increased dependency on SHMT2 (mitochondrial serine-hydroxymethyltransferase) is found in cell lines with increased expression of AMT (a subunit of the glycine-cleavage system; localized in mitochondria) and of MTHFD2 (mitochondrial methylenetetrahydrofolate dehydrogenase, Fig. 4c). This might occur in the face of a high demand rate for mitochondrial one-carbon units, requiring both serine and glycine catabolic activities. In cells where the expression of MTHFD2 is high, increased dependency on MTHFD1L might be needed to support mitochondrial formate production from 10-formyl-THF, facilitating the shuttling of one-carbon units to cytosol to support purine and pyrimidine biosynthesis [42, 43] (Fig. 4d).

## Discussion

A genome-scale genetic screen is a powerful approach for unbiased discovery of vulnerabilities in cancer cells that can

be targeted for therapeutic purposes. Since the development of RNAi- and CRISPR-based gene-silencing techniques, both approaches have been systematically applied for screening the gene knockdown and knockout effects in numerous cell lines. To identify genes whose perturbation would selectively affect cancer cells, gene-silencing profiles are typically integrated with tumor-specific mutations and alterations in gene expression to identify a specific molecular marker indicative of cancer cell dependence on a target gene. This is typically referred to as synthetic lethality or synthetic dosage lethality, where a perturbation of a single gene is viable but the perturbation of both genes is lethal to cells [44]. The identification of such synthetic lethal gene pairs provides a patient stratification strategy based on the mutation or expression status of one gene where the drug targeting of its partner gene enables selective eradication of cancer cells while minimizing toxic side effects. Previous studies suggest going beyond the search for synthetic lethal pairs, generating predictive models of gene essentiality in cancer cells based on multiple molecular features [14, 15]. However, as shown here, the ability to generate such predictive models for metabolic genes was significantly limited compared to other non-metabolic genes.

Analyzing genetic silencing screen data from recent RNAi and CRISPR screens, we found a confounding effect of

culture media on cell line-specific response to the silencing of especially metabolic genes. Our finding is consistent with recent studies, demonstrating the major effect of the tumor microenvironment in determining the cancer cellular metabolic phenotypes [17, 45–47]. In fact, several recent attempts have been made to develop culture media which would more closely mimic nutrient availability and concentration as in human plasma [20]. Our results, showing that the specific choice of media affects cell line dependence on metabolic genes, call for future genetic screens to maintain a similar culture media for different cell lines, optimally matching the true physiological conditions. Predictive models for cell line dependence on specific genes are not only important for potential usage as a patient stratification tool, but also useful for studying the molecular mechanisms that underlie cell line-specific metabolic activities. Hence, taking into account the culture media in our analysis, the generated predictive models are readily available for researchers interested in exploring cancer cell line dependence on specific metabolic genes. As shown, the integration of multiple molecular and cancer lineage features within a predictive model for gene dependency reveals numerous important predictive molecular features that cannot otherwise be inferred.

Another factor that complicates the finding of molecular features explaining cell line-specific gene dependence is the substantially high number of candidate predictive features: Specifically, the ability to identify synthetic lethality by analyzing the correlation between the dependency score of a certain gene and a large set of molecular feature suffers from multiple hypothesis testing, and the generation of predictive models via machine learning while considering a high number of features is complicated by overfitting. The incorporation of prior knowledge regarding feature relevance was previously shown highly useful for limiting the number of candidate predictive features and increasing statistical power [48]. Here, we show that a measure of functional relatedness of metabolic genes, defined based on connectivity in a metabolic network, is highly useful for generating predictive models for metabolic genes. Overall, utilizing this measure and also considering culture media composition, resulted in a 2.2-fold increase in the number of generated predictive models for metabolic genes compared to those derived as part of the Cancer Dependency Map [14].

Notably, while we were able to generate significant predictive models for cell line specific dependence on ~ 30% of the analyzed metabolic genes by integrating multiple molecular features, additional molecular characterization of the analyzed cell lines is expected to enable expanding this to other genes. Specifically, considering that metabolic activity is controlled beyond genomic and transcriptomic levels, additional high-throughput measurements of epigenetic modifications, proteomics, and metabolomics are expected to be highly useful in closing the gap.

## Conclusions

Culture media has a significant confounding effect on cell line-specific response to the silencing of metabolic genes, which should be taken into account when analyzing results from gene-silencing screens. Together with the media type, molecular features of functionally related enzymes in a metabolic network improve the prediction of cell line-specific dependence on metabolic genes. Overall, we expect our approach and predictive models of metabolic gene essentiality to be a useful tool for investigating metabolic abnormalities in cancer.

## Methods

### Pan-cancer gene essentiality screening data

The RNAi screen data analyzed in this study contains dependency scores for 17,098 genes over 501 cancer cell lines (released on July 2017) [14], while the CRISPR-Cas9 screen contains scores for 17,670 genes over 341 cancer cell lines (released on October 2017) [16].

### Statistical analysis of the correlation between gene dependency scores and culture media

To check whether cell culture type explains the correlation between media composition and dependency scores, we calculated for each gene the Spearman partial correlation between its RNAi dependency score across cell lines and a binary vector representing media composition in each cell line (DMEM or RPMI), controlling for binary vector representing cell culture type (suspension vs adherent). Derived  $p$  values were FDR-corrected via the method of Benjamini-Hochberg [49].

To test whether cancer cell lineage explains the correlation between media composition and dependency scores, we computed a similar Spearman partial correlation for each gene while controlling for the effect of an indicator variable representing each cancer lineage (having a value of one for cell lines of belonging to a specific lineage). We consider a correlation between gene dependency scores and media composition as not explainable based on cancer lineage information in case the above partial correlation remains significant when controlling for each cancer lineage (i.e., having an FDR-corrected  $p$  value < 0.05).

### Identifying neighboring enzymes in the human metabolic network

We consider a total of 1612 metabolic genes included in the genome-scale human metabolic network model Recon 2.02 [21] (omitting EGFR and PIK due to their central role in signaling pathways). The metabolic neighbors of a given metabolic enzyme are those that consume or produce substrate or product metabolites of this enzyme. Currency metabolites that participate in more than 1% of the reactions in the network (central energy or redox factors, etc.) were not considered to



define metabolic neighbors. For each gene, we considered no more than 50 neighboring enzymes and those connected through reactants participating in the minimal number of metabolic reactions in the network.

#### Pairwise correlations between gene dependency scores and expression of functionally related genes

We calculated the Spearman correlations between dependency scores of metabolic genes (focusing on those whose dependency score in at least one cell line is lower by more than four standard deviations from the mean of each gene) and the expression of 19,541 protein-coding genes derived from the CCLE [50]. Resulting  $p$  values were FDR-corrected using the method of Benjamini-Hochberg [49]. A similar analysis was performed using gene expression of isozymes or enzyme neighbors in the metabolic network. The dependency scores of a gene were considered to be explained by the expression levels of isozymes/metabolic neighbor in case a significant correlation (FDR-corrected  $p$  value  $< 0.05$ ) is found for at least a single isozyme/metabolic neighbor. In order to check the statistical significance of the fraction of metabolic genes having a significant correlation, we performed a random permutation test repeating the analysis above 1000 times while randomly shuffling the set of neighbors/isozymes for each enzyme.

#### Generation of predictive models of RNAi- and CRISPR-based gene dependency scores

We focus on metabolic genes from Recon 2.02 for which RNAi- or CRISPR-based dependency scores were available, and whose dependency scores showed considerable deviation throughout the cell lines (whose dependency score in at least one cell line is lower by more than two standard deviations from the mean of each gene). Molecular features, including RNA levels, copy number variation, and mutations in 125 oncogenes or tumor suppressors (non-synonymous point mutations, frameshift and splice mutations) [24], were derived from the CCLEs [50] and from CBioPortal [51, 52]. Information on media type and primary disease was derived from the screens data [14, 16]. Only cancer lineages having at least 10 cell lines in the screen were considered. To find predictive models, we use random forest learning [25, 26], generating an ensemble of 100 decision trees using MATLAB, with a min leaf size of 5. This method accommodates both continuous and categorical features, capturing nonlinear relationships and correcting for potential overfitting [53]. We perform out-of-bag prediction and record the Pearson correlation as the goodness of fit. We assessed the significance of the goodness of fit for each model by comparing its Pearson  $p$  value with a distribution of

$p$  values computed with randomly shuffled data (making 25K repetitions of selecting a gene by random, permuting its dependency scores, generating a predictive model, and computing the goodness of fit). To estimate the importance of the features used to learn the model, we calculated the out of bag permuted variable importance. To derive robust predictors, we selected the features with a positive feature importance for each model and regenerate that model by focusing on these features only (as in [14]). Extending the above analysis to include also copy number variation of oncogenes/tumor suppressors did not result in a higher number of significant predictive models for either RNAi or CRISPR screens.

#### Additional files

**Additional file 1:** RNAi models. Excel table containing statistically significant predictive models found using the RNAi screen data. (XLSX 32 kb)

**Additional file 2:** CRISPR models. Excel table containing statistically significant predictive models found using the CRISPR screen. (XLSX 39 kb)

**Additional file 3 Figure S1.**  $P$  value distributions calculated as part of the generation of predictive models for metabolic gene dependency scores using expression/copy number variation of related metabolic genes, genomic mutations, media information, and cancer lineage for RNAi (a, c, e) and CRISPR (b, d, f) screens. (a, b) Distributions of  $p$  values obtained using randomly shuffled data: Selecting a gene by random, randomly permuting its dependency scores throughout the cell lines, generating a predictive model of its shuffled dependency scores throughout cell lines, and computing the goodness of fit between the generated model and shuffled dependency scores (25K repetitions). (c, d) Distributions of Pearson  $p$  values assessing the goodness of fit between dependency scores and model predictions. (e, f) Distributions of empirical  $p$  values after FDR correction: For each predictive model, an empirical  $p$  value is computed based on the fraction of Pearson  $p$  values obtained with random data (shown in panels c and d) that are equal or lower to the  $p$  value computed with the original data (and correcting for multiple hypothesis testing using the method of Benjamini-Hochberg). The black lines in subfigures c and d denote the threshold of significance (based on the distribution of  $p$  values generated with shuffled data; shown in panels a and b), and the lines in subfigures e and f denote a threshold of 0.05 on the empirical  $p$  values. (PNG 485 kb)

**Additional file 4: Figure S2.** Percent of statistically significant predictive models identified in our analysis using expression/copy number variation of related metabolic genes, genomic mutations, and media information (green); the percent of statistically significant predictive models using the same set of features though without media information (orange); and the percent of significant predictive models when randomly shuffling the set of related metabolic genes (i.e., for a given gene having  $N$ -related genes,  $N$  genes were randomly selected), repeating the analysis 100 times (purple). The latter was significantly lower than the percent of genes with a significant predictive model when considering all features (green) and without media information (orange;  $p$  value  $< 0.05$ , marked with an asterisk). (PNG 196 kb)

**Additional file 5: Figure S3.** (a, b) The fraction of metabolic genes (whose dependency score in at least one cell line is lower by more than six standard deviations from the mean of each gene) for which a significant predictive model of RNAi (a)- and CRISPR (b)-based gene dependency was generated by focusing on molecular features of neighboring enzymes and culture media and when also considering cancer lineage information (green for RNAi, brown for CRISPR). In comparison, the fraction of predictive models for RNAi-based gene dependency scores derived by the Dependency Map project (based on

molecular featured of all genes and using functionally related genes) is shown in orange. (PNG 155 kb)

### Funding

TS was supported by a grant from the Israel Ministry of Science and Technology (MOST-DKFZ) #3-13167 and the European Research Council/ERC Grant Agreement #714738.

### Availability of data and materials

The datasets analyzed during the current study are available in the Cancer Dependency Map portal [54, 55]. Information on significant predictive models of cancer cell line dependence on metabolic genes based on RNAi and CRISPR data is provided in Additional files 1 and 2. The code used for this study is available as a GitHub repository at <https://github.com/shovall/MetabolicGeneDependencies> [56], <https://doi.org/10.5281/zenodo.2586665> [57].

### Authors' contributions

SL and TS designed the study. SL performed the computational analysis. All authors interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Faculty of Computer Science, Technion, Haifa, Israel. <sup>2</sup>Faculty of Biology, Technion, Haifa, Israel. <sup>3</sup>Lokey Center for Life Science and Engineering, Technion, Haifa, Israel.

Received: 30 November 2018 Accepted: 15 April 2019

Published online: 30 April 2019

### References

- Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even Warburg did not anticipate. *Cancer Cell*. 2012;21(3):297–308.
- Vander HMG, Cantley LC, Thompson CB, Mammalian P, Exhibit C, Metabolism A. Understanding the Warburg effect : cell proliferation. *Science*. 2009;324:1029.
- Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab*. 2016;23(1):27–47.
- Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. *Nat Rev Cancer*. 2011;11(2):85–95.
- Martinez-Outschoom UE, Peiris-Pagés M, Pestell RG, Sotgia F, Lisanti MP. Cancer metabolism: a therapeutic perspective. *Nat Rev Clin Oncol*. 2017;14(1):11–31.
- Luengo A, Gui DY, Vander Heiden MG. Targeting metabolism for cancer therapy. *Cell Chem Biol*. Elsevier Ltd. 2017;24:1161–80.
- Galluzzi L, Kepp O, Vander HMG, Kroemer G. Metabolic targets for cancer therapy. *Nat Rev Drug Discov*. 2013;12:829–46.
- Wing H, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Pnas*. 2011;108(30):12372–7.
- Marcotte R, Brown KR, Suarez F, Sayad A, Karamboulas K, Krzyzanowski PM, et al. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov*. 2012;2(2):172–89.
- Cowley GS, Weir BA, Vazquez F, Tamayo P, Scott JA, Rusin S, et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci Data*. 2014;1:1–12.
- Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*. 2015;163(6):1515–26.
- Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol*. 2016;34(9):933–41.
- Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep*. 2016;17(4):1193–205.
- Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell*. 2017;170(3):564–576.e16.
- McDonald ER, de Weck A, Schlabach MR, Billy E, Mavrikis KJ, Hoffman GR, et al. Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell*. 2017;170(3):577–92 e10.
- Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49(12):1779–84.
- Davidson SM, Papagiannakopoulos T, Olenchok BA, Heyman JE, Keibler MA, Luengo A, et al. Environment impacts the metabolic dependencies of Ras-driven non-small cell lung cancer. *Cell Metab*. 2016;23(3):517–28.
- Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. *Cancer Discov*. 2012;2(10):881–98.
- Hensley CT, Faubert B, Yuan Q, Lev-Cohain N, Jin E, Kim J, et al. Metabolic heterogeneity in human lung tumors. *Cell*. 2016;164(4):681–94.
- Cantor JR, Abu-Remaileh M, Kanarek N, Freinkman E, Gao X, Louissaint A, et al. Physiologic medium rewires cellular metabolism and reveals uric acid as an endogenous inhibitor of UMP synthase. *Cell*. 2017;169(2):258–272.e17.
- Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31(5):419–25.
- Dulbecco R, Freeman G. Plaque production by the polyoma virus. *Virology*. 1959;8(3):396–7.
- Moore GE, Gerner RE, Franklin HA. Culture of normal human leukocytes. *J Am Med Assoc*. 1967;199(8):519–24.
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546–58.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
- Breiman L. Bagging predictors. *Mach Learn*. 1996;24(2):123–40.
- Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;2016:1.30.1–1.30.33.
- Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes-2009. *Nucleic Acids Res*. 2009;38(SUPPL.1):497–501.
- Gaude E, Frezza C. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat Commun*. 2016;7:1–9.
- Yuneva MO, Fan TWM, Allen TD, Higashi RM, Ferraris DV, Tsukamoto T, et al. The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab*. 2012;15(2):157–70.
- Ying H, Kimmelman AC, Lyssiotis CA, Hua S, Chu GC, Fletcher-Sananikone E, et al. Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell*. 2012;149(3):656–70.
- Kerr EM, Gaude E, Turrell FK, Frezza C, Martins CP. Mutant Kras copy number defines metabolic reprogramming and therapeutic susceptibilities. *Nature*. 2016;531(7592):110–3.
- Haq R, Shoag J, Andreu-Perez P, Yokoyama S, Edelman H, Rowe GC, et al. Oncogenic BRAF regulates oxidative metabolism via PGC1 $\alpha$  and MITF. *Cancer Cell*. 2013;23(3):302–15.
- Liu J, Zhang C, Hu W, Feng Z. Tumor suppressor p53 and its mutants in cancer metabolism. *Cancer Lett*. 2015;356(2):197–203.
- Takeuchi H, Haltiwanger RS. Significance of glycosylation in Notch signaling. *Biochem Biophys Res Commun*. 2014;453(2):235–42.
- Jolivet J, Cowan KH, Curt GA, Clendeninn NJ, Chabner BA. The pharmacology and clinical use of methotrexate. *N Engl J Med*. 1983;309(18):1094–104.
- Ducker GS, Chen L, Morscher RJ, Teng X, Kang Y, Rabinowitz JD, et al. Reversal of cytosolic one-carbon flux compensates for loss of the mitochondrial folate pathway article reversal of cytosolic one-carbon flux compensates for loss of the mitochondrial folate pathway. *Cell Metab*. 2016; 23(6):1140–53.
- Ducker GS, Ghergurovich JM, Mainolfi N, Suri V, Jeong SK, Li SH. Human SHMT inhibitors reveal defective glycine import as a targetable metabolic vulnerability of diffuse large B-cell lymphoma. *Pnas*. 2017;114(43):11404–9.

39. Ducker GS, Rabinowitz JD. One-carbon metabolism in health and disease. *Cell Metab.* 2017;25(1):27–42.
40. Yang M, Vousden KH. Serine and one-carbon metabolism in cancer. *Nat Rev Cancer.* 2016;16(10):650–62.
41. Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat Rev Cancer.* 2013;13(8):572–83.
42. Minton DR, Nam M, McLaughlin DJ, Shin J, Bayraktar EC, Alvarez SW, et al. Serine catabolism by SHMT2 is required for proper mitochondrial translation initiation and maintenance of formylmethionyl-tRNAs. *Mol Cell.* 2018;69(4):648–663.e7.
43. Morscher RJ, Ducker GS, Li SHJ, Mayer JA, Gitai Z, Sperl W, et al. Mitochondrial translation requires folate-dependent tRNA methylation. *Nature.* 2018;554(7690):128–32.
44. O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. *Nat Rev Genet.* 2017;18(10):613–23.
45. Schug ZT, Vande Voorde J, Gottlieb E. The nurture of tumors can drive their metabolic phenotype. *Cell Metab.* 2016;23(3):391–2.
46. Elia I, Fendt S-M. In vivo cancer metabolism is defined by the nutrient microenvironment. *Transl Cancer Res.* 2016;5(5):S1284–7.
47. Muir A, Vander Heiden MG. The nutrient environment affects therapy. *Science.* 2018;360(6392):962–3.
48. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet.* 2015;16(6):321–32.
49. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57:289–300.
50. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012;483(7391):603–7.
51. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–4.
52. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal complementary data sources and analysis options. *Sci Signal.* 2014;6(269):1–20.
53. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. *Math Intell.* 2001;27(2):83–5.
54. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screen in cancer cells. <https://depmap.org/ceres/>. Accessed 14 Nov 2017.
55. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. <https://depmap.org/rnai/>. Accessed 4 Sept 2017.
56. Lagziel S, Lee WD, Shlomi T. Inferring cancer dependencies on metabolic genes from large-scale genetic screens: GitHub. <https://github.com/showall/MetabolicGeneDependencies>
57. Lagziel S, Lee WD, Shlomi T. Inferring cancer dependencies on metabolic genes from large-scale genetic screens: Zenodo. <https://doi.org/10.5281/zenodo.2586665>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

