

Friday, August 17, 2018

Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape

Running title: Modeling the metabolism of cancer

Cankut Cubuk¹, Marta R. Hidalgo¹, Alicia Amadoz², Miguel A. Pujana³, Francesca Mateo³, Carmen Herranz³, Jose Carbonell-Caballero⁴, Joaquin Dopazo^{1,5,6*}

1 Clinical Bioinformatics Area. Fundación Progreso y Salud (FPS). CDCA, Hospital Virgen del Rocío. 41013. Sevilla. Spain

2 Igenomix S.L. 46980 Valencia, Spain.

3 ProCURE. Catalan Institute of Oncology. Bellvitge Institute for Biomedical Research (IDIBELL). L'Hospitalet del Llobregat. 08908 Barcelona. Spain

4 Centre for Genomic Regulation, 08003, Barcelona, Spain.

5 Functional Genomics Node, INB-ELIXIR-es, FPS, Hospital Virgen del Rocío, Sevilla, 41013, Spain

6 Bioinformatics in Rare Diseases (BiER). Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER). FPS. Hospital Virgen del Rocío. 41013. Sevilla. Spain;

*Corresponding author: Joaquin Dopazo, Clinical Bioinformatics Area. Fundación Progreso y Salud (FPS). CDCA, Hospital Virgen del Rocío. 41013. Sevilla. Spain.
Email: joaquin.dopazo@juntadeandalucia.es; Tel. +34 677910687

Conflicts of interest: none

Abstract

Metabolic reprogramming plays an important role in cancer development and progression and is a well-established hallmark of cancer. Despite its inherent complexity, cellular metabolism can be decomposed into functional modules that represent fundamental metabolic processes. Here we performed a pan-cancer study involving 9428 samples from 25 cancer types to reveal metabolic modules whose individual or coordinated activity predict cancer type and outcome, in turn highlighting novel therapeutic opportunities. Integration of gene expression levels into metabolic modules suggests that the activity of specific modules differs between cancers and the corresponding tissues of origin. Some modules may cooperate, as indicated by the positive correlation of their activity across a range of tumors. The activity of many metabolic modules was significantly associated with prognosis at a stronger magnitude than any of their constituent genes. Thus, modules may be classified as tumor suppressors and onco-modules according to their potential impact on cancer progression. Using this modeling framework, we also propose novel potential therapeutic targets that constitute alternative ways of treating cancer by inhibiting their reprogrammed metabolism. Collectively, this study provides an extensive resource of predicted cancer metabolic profiles and dependencies.

Introduction

Changes in metabolism were first linked to cancer almost one century ago (1), particularly from the observation of enhanced aerobic glycolysis (also known as the Warburg effect) (2). It is well known that the common proliferative phenotype of cancer cells require intensive support for the biosynthesis of cellular components and generation of energy, which overall are accomplished by reprogramming metabolism (2,3). Actually, it has long been known that key signaling pathways that are altered in cancer are important regulators of metabolism (4). In fact, in addition to the Warburg effect, other alterations in the synthesis of nucleotides, amino acids and lipids (5), mutations in metabolic genes and accumulations of key metabolites (6) have been reported. Consequently, reprogramming of cellular metabolism is a recognized essential feature for cancer development and progression (7), and is therefore a recognized neoplastic hallmark (7,8). This observation, along with the discovery of the therapeutic potential of metabolic targets in cancer (9), has sparked a growing interest in cancer metabolism (3,4). Recent studies show that genes involved in metabolic pathways shows a remarkable heterogeneity across various cancer types (8), which suggests that personalized therapies are likely to be successful if the context of the intervention is accurately depicted. In this context, synthetic lethality, defined as combined molecular perturbations with a drastic effect on cell viability, but with no individual effect, offers a promising range of potential therapeutic interventions based on cancer metabolic dependencies (10).

Despite the inherent complexity of the metabolism, various approaches based on the integration and modeling of functional genomic data have allowed the reconstruction of genome-scale metabolic networks (11). Indeed, recent studies have demonstrated that complex phenotypes or outcomes such as patient survival (12,13) and drug activity (14) are better predicted by the inferred activity of pathways, than by the activity of their constituent genes and/or proteins. Constraint-Based Models (CBMs), which use maps of metabolic networks in combination with gene activity inferred from transcriptomic profiles, have been applied to decipher relationships between various aspects of the cellular metabolism and phenotypes (15). In fact, CBMs have enabled the analysis of metabolism in different scenarios at an unprecedented level of detail (16). However, CBMs have several analytical and modeling limitations, such as the dependence of their

solutions on initial conditions or the sometimes unrealistic and arbitrary nature of some of the assumptions and difficulties of convergence on unique solutions (17)).

Recently, realistic computational models of signal propagation have been successfully applied to predict complex phenotypes using estimates of signaling pathway activities inferred from gene expression data (13,14). In addition, such models provide important information about disease mechanisms and mode of action (MoA) of drugs (13). Here we generalize the application of this approach to describe the metabolic profiles and dependencies across 14 cancer types. Our study reveals common and specific metabolic modules that influence patient survival. We also identify metabolic dependencies based on targeted molecular predictions that point to novel beneficial therapeutic interventions.

An interactive and intuitive web has been developed to explore the data predictions and perform further metabolic modeling base on users' data and hypotheses (<http://metabolizer.babelomics.org/>).

Material and methods

Data

Human samples and data processing

RNA-seq counts for a total of 9428 samples, 8319 corresponding to cancer and 649 to healthy reference tissue, belonging to 25 cancer types, (see Table S1), as well as their subtype stratification, were downloaded from The International Cancer Genome Consortium (ICGC) repository (https://dcc.icgc.org/releases/release_20/Projects). The trimmed mean of M-values (TMM) normalization method (18) was used for gene expression normalization.

Data on somatic mutation in genes from the modules was taken from the CDG cancer portal (<https://portal.gdc.cancer.gov/files/995c0111-d90b-4140-bee7-3845436c3b42>), Expression data on responses to drugs were taken from GSE25066, GSE50948, GSE5462 datasets downloaded from GEO. Probes mapping in more than one gene were discarded. The median value of the probes mapping on a gene was used as the expression value for this gene. Microarray data normalization and background correction was done using RMA method implemented in the *affy* Bioconductor package (<https://bioconductor.org/packages/release/bioc/html/affy.html>). Normalized samples

were log-transformed and a truncation by quantile 0.99 was applied. The COMBAT method (19) was used for batch effect correction. Finally, data were re-scaled between 0 and 1.

Clinical data

Clinical data were available through the cBIOportal (<http://www.cbioportal.org/>). These data included individual survival information that was used for survival analysis.

Cell line expression values and cell line survival data

A total of 212 cell lines were used in this study (Table S2). Gene expression data were taken from the Cancer Cell Line encyclopedia (<https://portals.broadinstitute.org/ccle/>). Cell survival measurements after gene KD were taken from the Project Achilles 2.4.3 (<https://portals.broadinstitute.org/achilles/datasets/5/download>). Survival validation data were taken from the new release 2.20.2 of the project Achilles (<https://portals.broadinstitute.org/achilles/datasets/15/download>).

Modelling framework

A model of the activity of a pathway requires of: i) a description of the current map of relationships between the proteins that make up the pathway and ii) a function that relates the activity of the pathway to those of the constituent proteins.

Module definition

Pathway modules (20) are used to describe the map of interactions between the proteins that ultimately carry out the main metabolic transformations in the cell. Pathway modules are modular sequential metabolic reactions that consist of conserved functional units of enzyme complexes and metabolic sub-pathways, representing a summary of the known human metabolism (20). A total of 86 modules were downloaded from the KEGG MODULE (<http://www.genome.jp/kegg/module.html>) database (Table S3). Since seven of them ended in more than one metabolic product, these were decomposed into two sub-modules, resulting in a final number of 95 modules. The modules comprise a total of 446 reactions and 553 genes. Specifically, the metabolic pathways were downloaded as KGML files and were parsed to extract detailed information about the metabolites, genes and reactions. Reactions and metabolites in the module are represented as nodes that are connected by the edges of the graph in a way that

describes the sequence of reactions that transform simple metabolites into complex metabolites, or vice versa (see Figure S1).

Module activity estimation

Once a module has been formally described by a graph, its activity level can be described as a function of the activities of all the reaction nodes (composed of one or several isoenzymes or enzymatic complexes (21)) and the presence of all the intermediate metabolites. Reaction node activity can be estimated from the presence of the corresponding proteins. Since direct measurements of protein levels are commonly difficult to obtain, the presence of the corresponding mRNA expression level within the context of the module is widely used as a proxy (13,22,23). Normalized gene expression values are used for this purpose, as explained above. To estimate the potential activity of the reaction node two scenarios are considered (24): isoenzymes, where the activity is produced if at least one of them is present ($\text{Expression}_{\text{Isoenzyme1}} \text{ OR } \text{Expression}_{\text{Isoenzyme2}} \text{ OR } \dots$) and enzymatic complexes, where the activity occurs only if all the enzymes are present ($\text{Expression}_{\text{EnzymeA}} \text{ AND } \text{Expression}_{\text{EnzymeB}} \text{ AND } \dots$). For example, in Figure S1, the last reaction transforming *isocitrate* into *2-oxoglutarate* is catalysed by either an enzymatic complex or two alternative isoenzymes, represented as “(R01899 AND R00268) OR R00267 OR R00709”, which may be estimated from the normalized gene expression values of the mRNAs corresponding to proteins R01899, R00268, R00267 and R00709 as:

$$n = \max\{\min\{E_{R01899}, E_{R00268}\}, E_{R00267}, E_{R00709}\}$$

where n is the activity of the node and E_p is the 90th percentile of normalized expression of the gene corresponding to the enzyme p . This approach is very similar to computing the Gene-Protein-Reaction (GPR) rules in metabolic networks (24). It is worth noting that some enzymes can participate in more than one node (even in different modules), and thereby contribute to different reactions.

As in the case of protein activities, metabolite measurements are typically not simultaneously available with gene expression measurements and therefore the metabolite nodes have been excluded here from the calculations (equivalent to setting all their values to 1, assuming that all of them are present). However, if metabolite measurements were available, they could easily be accommodated in this modelling

framework by normalising them and using the corresponding values in the graph. Once node activities have been estimated, the contribution of each node to the integrity of the chain of reactions leading from simple to complex metabolites can be computed using a recursive method. Starting with a value of 1, in the node corresponding to the simplest metabolite, the activity of the subsequent nodes is calculated by the formula:

$$S_i = n_i \cdot \left(1 - \prod_{s_a \in A_i} (1 - s_a) \right)$$

where n_i is the activity of the current node n , A is the total number of edges arriving at the node that account for the flux of metabolites produced in other nodes with activity values s_a , and S_i is the flux of the current reaction node.

The net value of integrity of the whole circuit of reactions represented in the module is summarized by the value of activity propagated until the last node that carries out the last reaction that produces the resulting metabolite of the chain of reactions.

Differential module activity estimation

Activity values for the modules calculated in this way are dimensionless values that, like normalized gene expression values, make sense in a comparative context that makes it possible to decide whether the activity of a given module exhibits a significant variation or not. The Wilcoxon test is used to assess the significance of the observed changes of module activity when samples of two conditions are compared.

Since many modules were simultaneously tested, the popular FDR method (25) was used to correct for multiple testing effects.

Survival analysis

Kaplan-Meier (K-M) curves were used to relate module activity to patient survival in the different cancers. The value of the activity estimated for each module in each individual was used to assess its relationship with individual patient survival. Calculations were carried out using the function *survdiff* in the *survival* R package (<https://cran.r-project.org/web/packages/survival/>).

Cox regression analysis (26) was used to relate combined module activity to survival in the different cancers. Calculations were carried out using the *coxph* function in the

survival R package (<https://cran.r-project.org/web/packages/survival/>). A stepwise algorithm implemented in the *step* function of the *stats* R package (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/step.html>), was used to add or remove modules according to the significance of their contributions to explain survival in the multiple regression model. The *step* function uses the Akaike Information Criterion (AIC) to select the best model by iteratively adding and removing variables. Finally, the method yields a list of modules whose combination is significantly related to survival. Adjustment for multiple testing was made by the FDR method (25). For patient stratification based on predicted module activity, high and low module activity groups were defined using the 80% and 20% percentiles, respectively.

Module essentiality

Simulation of the effect of gene knock-downs on module activity

Given a set of gene expression values (wt expression), the activity of the modules was estimated as described above (wt activity). Then, the knocked down gene(s) expression value(s) were set to 0.001 (KD expression) and the activity of the modules was recalculated again (KD activity). The log-fold-change in module activities was then calculated from the comparison of KD and wt module activity profiles as

$$\text{Log-fold-change} = \log(\text{KD module activity}) - \log(\text{wt module activity})$$

Relationship between module activity and cell survival

To estimate module activity essentiality cell lines were grouped by cancer type. For each cancer type the impact of gene KDs on the activity of the modules was calculated as described above. Then, a Spearman correlation coefficient between log-fold-change values and cell survival, as described in the Project Achilles was calculated. Lower Achilles scores indicate higher mortality and, consequently, essentiality of the KD gene. Positive correlations indicate essentiality in module activity (the less activity the lower the Achilles index) in this particular cancer type.

Validation of the essentiality predictions

Independent dataset validation

In order to check the correspondences between gene expression data and protein expression data we have used the reverse phase protein array (RPPA) assays available in the ICGC repository.

Data on cancer dependencies, which include estimates of cell viabilities after gene KD, from the Project Achilles 2.20.2 was used to check the validity of the predictions made with the Project Achilles 2.4.3 (TableS2 and references therein). It was expected that the inhibition of an onco-module would reduce the viability of cancer cells. Conversely, the inhibition of a tumor suppressor module should result in greater cell survival. In order to detect these increases or decreases, the Project Achilles 2.20.2 cell viability scores observed in the cell line in which an effect of KD on cell survival was predicted were compared with the scores reported for the other cell lines (background score). Increases or decreases in the mean values were taken as evidences of predicted effects on cell viability.

Experimental validation

The shRNAs targeting *UPBI* were purchased from the MISSION library (Sigma Aldrich), catalog SHCLNG-NM_016327. Lentivirus was produced and transduced following standard protocols and cell cultures were selected with puromycin for 72 hours before cell seeding for evaluation of proliferation/viability by methylthiazol tetrazolium (MTT)-based assays (Sigma-Aldrich). The data corresponds to sextuplicates and were replicated in different assays. *UPBI* expression was detected with the Human Protein Atlas HPA000728 antibody (Sigma-Aldrich) and gene expression measured with primers 5'-TCGACCTAAACCTCTGCCAG-3' and 5'-TAAGCCTGCCACACTTGCTA-3', using PPP1CA as control.

Results

Data pre-processing

RNA-seq counts for 25 cancer types, totaling 9428 samples (Table S1) were downloaded from The International Cancer Genome Consortium (ICGC) repository. Principal component analysis (PCA) was used to detect possible batch effects. The results are shown by plotting samples with respect to disease status (Figure S2A and S2B), sequencing center (Figure S2C and S2D) and project (Figure S2E and S2F). An appreciable technical batch effect due to sequencing center was found (Figure S2 C) and

this was corrected by application of the COMBAT (19) method (Figure S2 D). Samples were normalized and preprocessed as explained in Methods.

Concordance between gene expression, protein expression and module activity

Since the predictions of the proposed method are based on gene expression values, data from protein expression measurements in reverse reverse-phase protein array (RPPA) were used to validate them. Excluding biochemical modifications, 13 genes encoding for enzymes contained in KEGG pathways were included in RPPAs; thus, 11 of these genes showed positive expression correlations (Spearman's rank correlation coefficients, $p < 0.05$) with their corresponding protein measures in at least one cancer type (Table S4A). Only two genes, *GAPDH* and *ACACB*, were negatively correlated with their RPPA measures in some cancer types, which is a similar proportion to that revealed by the analysis of genome-wide gene and protein expression profiles (27). Importantly, previous pan- and specific-cancer analyses have shown common and robust positive correlations between gene expression and RPPA measures (28). In addition, in seven of the ten modules including an enzyme that was measured by RPPAs, a positive correlation between predicted module activity and enzyme expression was found in at least three cancer types (Table S4B). Moreover, analysis of 201 proteins included in the RPPAs identified positive gene-protein expression correlations in 176 (88%) instances across any cancer type (Table S5). To further assess the validity of the predictions, metabolomic data from breast and kidney cancer were then analyzed (Table S6 and references therein). We used the balance between the initial and final metabolite fold-changes as an indication of activation (relative increase in the final metabolite with respect to the initial one) or inactivation (relative decrease in the final metabolite with respect to the initial one). With the premise that both the initial and final metabolites of each module were measured in these studies, we found two datasets where three modules could be evaluated, and, for these settings, all of our five predictions proved to be correct (Table S6 and references therein). Therefore, the predictions from our study are generally transferable to the protein expression and metabolic activity levels.

Pan-cancer metabolic activity profiles

For this differential module activity analysis we used 14 cancer types in which at least a 5% of healthy reference samples were available (totaling 6299 cancer samples and 687 healthy samples). For each cancer type, the activity of the modules was calculated for

all tumors and for all healthy tissue samples as described in Methods. Briefly, gene expression profiles were converted into metabolic module activity profiles by applying a formula that takes into account the chain of metabolic reactions required to complete the transformation of simple into complex metabolites in each module. Next, the Wilcoxon test was used to assess differences between conditions. Figure 1 shows the significant activations and deactivations of modules in tumors with respect to the corresponding healthy tissue.

Subtype-specific metabolic profiles

Since ICGC provides subtype annotations for six cancer types, it is straightforward to produce a similar analysis of differential module activity at the subtype level. For each of the six cancer types analyzed, Table S7 recapitulates the modules that present significant differential activity in the subtypes of BLCA (Table S7A and S7B), BRCA (Table S7C), COAD (Table S7D and S7E), LUAD (Table S7F), PRAD (Table S7G) and STAD (Table S7H) with respect to the corresponding normal tissue (in red) and those whose differential activity is specifically detected only in one of the subtypes (in blue). In some cancer subtypes no subtype-specific activities were detected. In many cases this was due to the small number of samples available.

Metabolic modules may be altered by oncogenic mutations

Many cancer drivers are known to promote metabolic reprogramming in cancer. To test our predictions in this context, we analyzed the impact of relatively frequent oncogenic mutations in well-established drivers linked to metabolic reprogramming; *AKT1* and *PIK3CA* in BRCA, and *IDH1* in GBM (Table S8A and S8B). Consistent with a major role linked to metabolism, mutations in *PIK3CA* caused significant changes in the predicted values of many metabolic modules, with coherent changes in seven of the modules also being significantly altered by *AKT1* mutations (Table S8C). Among the altered modules in *PIK3CA* mutants, some of the findings were consistent with current knowledge. The largest predicted metabolic activation in *PIK3CA* mutants is found to be the M00027 module of *GABA shunt* (end metabolite succinate), which is consistent with data of reprogrammed glutamine metabolism in this setting (Table S8C and references therein). In contrast, the activity of the M00034_1 module of *methionine salvage* is predicted to be higher in *PIK3CA* wild-type tumors, but interestingly this pathway becomes activated as a mechanism of resistance to *PI3K* inhibitors (Table S8C

and references therein). The *IDH1* mutations in GBM caused fewer module alterations, probably in part because only seven mutated samples were included in the analysis. Nonetheless, the largest impact upon *IDH1* mutations is predicted to be activation of *proline biosynthesis* (M00015), which links to metabolites downstream of *IDH1* activity (Table 5A and references therein). Another predicted effect was the downregulation of *glycosphingolipid biosynthesis* (M00071), which is consistent with the major demand of citrate towards the substrate of the reaction catalyzed by *IDH1*. In turn, the major activation corresponds to components down-stream of its activity that is related to *proline biosynthesis* (M00015) (Table S8C). Thus, the predictions from this study may also support the identification of specific, cancer driver-linked, metabolic reprogramming and/or vulnerabilities.

Cooperation between metabolic modules

Metabolic modules do not function in isolation, but rather display highly correlated (positive or negative) patterns of activity that influence cancer development and/or progression (8,23). However, how these correlations vary from normal tissue to cancer is poorly understood. Our results document a variable proportion of modules, ranging from 5.3% (in LIHC) to 26.9% (in READ), that are significantly positively correlated in normal tissue but not in the corresponding tumor. This proportion is smaller for negative correlations, ranging from 1.1% (in LUSC) to 10.6% (in BLCA). 10-35% of the activity of metabolic modules is uncoupled when normal and cancer metabolic activities are compared (Figure 2). Figure S3 represents in detail the modules whose activities are correlated in normal and/or cancer tissue and those in which the significance or direction of the correlation change.

Modules associated with cancer outcome

Modules differentially activated between cancer and the corresponding normal tissue may highlight metabolic processes that are required for cancer development and/or progression. The availability of patient survival data in 21 cancer types (See Table S1) allows the identification of modules in which changes in activity are significantly associated with the progression of each cancer type.

Table S9 portrays the modules whose change in activity is significantly associated with poorer patient survival in at least one cancer type. Since the number of deceased patients and, in general, data on mortality follow-ups is limited in the ICGC repository,

significant results were obtained for only a few modules. In particular, kidney (KIRC), liver (LIHC) and glioma (LGG) cancer types featured a remarkable number of modules influencing cancer outcome.

Moreover, following from the observation of correlated modules, the impact on survival may be further determined by combinations of their metabolic activities. Thus, we applied Cox multiple regression analysis (26) to find the combination of module activities that best accounted for patient survival. Table S10 shows the combinations of module activities significantly related to survival in various cancer types.

Previous results have shown that predicted activities of single or combined metabolic modules are associated with differences in cancer outcome, further emphasizing their fundamental role in cancer progression. In addition, we observed that the magnitude of their effect on survival was greater in some instances than for any of the individual activities of the genes that comprise a given module, which provides additional evidence that modules are real entities that contribute as whole units to cell functioning (see Table S11).

Modules associated with differential cancer therapeutic responses

Since we observed that metabolic activities influence cancer progression, they may also modify therapeutic responses. To examine this possibility, we analyzed gene expression data from three breast cancer studies involving different therapeutic settings: response to neoadjuvant taxane-anthracycline chemotherapy, response to neoadjuvant herceptin (NOAH trial), and early response to letrozole (Table S12A and references therein). Several metabolic modules were identified as potential modifiers of response to neoadjuvant taxane-anthracycline chemotherapy, including *glycolysis* (M00002), which has been extensively linked to chemoresistance and is thought to be a target for impairing this process (Table S12A and references therein). The analysis of the data from neoadjuvant herceptin identified a single significant metabolic module, *catecholamine biosynthesis* (M00042), a process that has been linked to resistance to anti-HER2 therapy (Table S12B and references therein). Evaluation of the metabolic modules that change significantly in response to letrozole identified several candidates, including *serine biosynthesis* (M00020) and *methionine degradation* (M00035_1), both of which processes have been linked to resistance to another endocrine therapeutic approach, based on tamoxifen (Table S12C and references therein). Therefore,

metabolic alterations predicted by this study may represent novel therapeutic opportunities for boosting the clinical benefit of standard treatments.

Essentiality and module activity

The availability of basal gene expression data from 212 cell lines of the Cancer Cell Line encyclopedia, along with the release of the results of Project Achilles (Table 2 and references therein), which assessed the consequences of individual silencing of thousands of genes across many cancer cell lines, allows the influence of predicted metabolic module activities on cancer robustness to be evaluated. The effect of every gene expression KD on the activity of the corresponding module was calculated as the log-fold-change between the estimated activity using cell line gene expression values and the activity estimated by assigning a very low expression value (see Methods) to the KD gene. Subsequently, the correlations of the log-fold-change values with the Achilles score, which accounts for cell viability, were calculated. Given that different cancer types display specific patterns of differential module activations, essentiality in modules is also expected to be specific to particular cancer types. Therefore, cell lines were grouped by cancer type to obtain the correlations between module activity and cell viability. Considering only significant correlations (FDR adjusted p-value < 0.05) with a correlation coefficient > 0.5 (or < -0.5) obtained from at least eight data points (cell lines x KD genes), a total of 20 modules in 12 cancer types showed significant positive (Table 1 and Figure 3A) or negative correlations (Table 1 and Figure 3B).

Validation of the gene essentiality predictions

We used a recently published study on cancer dependencies (Table 2 and references therein) that provides extra data on cell survival after massive gene KD. The comparison of cell survival in the cancer types predicted with respect to survival in cancers validated 48 of the 77 predictions (62%), along with three less conclusive validations, which would result in a 66% validation rate, covering 24 of the 28 modules predicted to affect cell viability (see Table 1 and, in more detail Table S13). This is an excellent proportion of validations, especially if we consider that the method used for validation can fail to detect real KD effects when the KD also markedly affects background survival.

Actually, independent experiments can confirm inconclusive validations of predicted inhibitions of essential modules. An interesting example is the small molecule, CBR

5884, which inhibits *PHGDH* causing selective toxicity in breast cancer cell lines by inhibiting serine biosynthesis (29), as predicted (see Table 1 and Table S13).

Finally, to further validate of our predictions (Table 1), the impairment of cell proliferation upon depletion of *UPBI* gene expression was assessed in two models of gastric cancer (AGS and MKN45 cell lines). This gene encodes an enzyme (β -ureidopropionase) that catalyzes the final step in the pyrimidine degradation pathway, which in turn is required for epithelial-mesenchymal transition (30). Thus, two short hairpin shRNA sequences directed to *UPBI* caused a significant decrease in proliferation of the two gastric cancer cell lines (AGS and MKN45), as predicted by the model. Conversely, the inhibition in a colon adenocarcinoma cell line (SW480), predicted as non-essential by our model did not result in a significant difference in growing (Figure 4), providing a negative control validation. Although additional experiments may be warranted to confirm cancer vulnerability or resistance based on predicted metabolic activities, these results can be considered independent validations that reinforce the predictions made by the model proposed (Table 1 and Table S13).

Therapeutic targeting of metabolic modules

Onco-modules are effective candidates for treating cancer (individually or in combinations), but interventions that activate some tumor suppressor metabolic modules may also offer useful therapeutic strategies. Table S14 lists 137 potential interventions with known drugs that are likely to affect cancer cell viability according to the predictions of the model proposed here.

Discussion

Although the role of metabolism in cancer has long been known (1,2), the results presented here provide a more detailed, mechanistic view documenting the relevance of specific metabolic module activities in cancer. This study is based on the integration of gene expression data into metabolic pathway modules and, therefore, may be limited by the lack of correlation between gene and protein expression, and with metabolic activities, in some instances; however, evaluation of RPPA data and module activity predictions, in accordance with independent studies (28), indicates that gene expression measures are generally a valuable proxy for protein expression and activities.

As expected, the production of nucleotides and their precursors (CDP and GTP) shows recurrent significant activation in all cancer types when compared with the corresponding reference tissues (Figure 1). Other pervasively activated modules include well-known cancer metabolic dependencies, like *cholesterol biosynthesis* (M00101), which is consistent with its essential role in cell membranes and as a precursor of steroid hormones (31), and *proline biosynthesis* (M00015), which is essential in many carcinogenesis settings (32). In addition, the predicted overexpression of *L-cystathionine* and *L-cysteine* across many cancer types may reflect a defect in *S-adenosyl-L-methionine*, which in turn is consistent with common DNA hypomethylation in cancer cells (33). On the other hand, this study reveals metabolites whose production is significantly reduced in several cancers types. For example, the well-known Warburg effect, i. e., the preference of cancer cells for anaerobic over aerobic metabolism is apparent in modules such as *Citrate cycle, second carbon oxidation* (M00011). It is also known that many human tumors do not express *ASS1* (34), one of the key enzymes of the *Urea cycle* (M00029) module, which is systematically downregulated in almost all cancer types.

Specific observations also support the relevance of the predicted metabolic activities. Examples of cancer metabolic specificities are: upregulation of *leucine* (M00036) and *catecholamine* metabolism (M00042) in prostate (35) and colorectal (36) cancer, respectively, and downregulation of *glycosaminoglycan* (M00058) and *polyamine biosynthesis* (M00134) in liver (37) and breast (38) cancer, respectively. In turn, this study highlights less explored metabolic associations, such as downregulation of the *pentose phosphate cycle* (M00004) in head and neck cancer, or *accumulation of cysteine* (M00338_1) in several cancer types, which may indicate a link to altered metabolism of reactive oxygen species. Collectively, the results of this study depict biologically relevant metabolic profiles throughout human cancer and provide many novel hypotheses about metabolic alterations in the disease.

Metabolic modules are also relevant for establishing the molecular basis that differentiates between cancer subtypes. Table S7 provides a detailed survey of differential and common metabolic module activities when cancer subtypes are compared. Although a detailed description of the findings is beyond the scope of this manuscript it is worthwhile highlighting some observations, such as the significant specific reduction of the activity of the module *C21-Steroid hormone biosynthesis, progesterone* (M00109) in basal-like breast cancer subtype (the only non-hormone

dependent form of the disease) (39). Experts in specific cancer types can use Table S7 to identify relevant subtype-specific module activities that can be exploited for therapeutic purposes.

Some of the modules that display different behaviors in cancer are expected to have a direct effect on patient survival. In spite of the limited patient survival data in the ICGC repository, Table S9 demonstrates that a remarkable number of modules are associated with poorer patient survival. Specifically, a high level of activity of the *pentose phosphate* module was found to be significantly associated with poor survival in five cancer types (see Table S9 and K-M plots in Figure 5A). This observation is consistent with the role of this module in the biosynthesis of nucleotides and NADPH, which is known to play a key role in facilitating cancer cells to cope with anabolic demands and to fight oxidative stress (40).

The analysis of metabolic modules reveals their role as ultimate mechanistic entities whose activity is related to cancer cell fate. For example, the expression of *EHHADH* has recently been associated with poor prognosis of KIRC (41), but the corresponding module, *Malonate semialdehyde pathway (M00013)* better predicts outcome (see Table S11). In fact, out of the 69 metabolic modules associated with differences in survival, a total of 27 (40%) modules (see Table S11) showed a stronger effect (based on hazard ratio estimations) than any of their corresponding genes. Other modules are also significantly related to survival in other cancer types as LGG (Figure 5B), KIRC (Figure 5C) or KIRP (Figure 5D).

Moreover, in the same way that genes are co-regulated in higher-level entities (metabolic modules), the activations and deactivations of metabolic modules are not an independent process and, in fact, proper cell functionality seems to require a high degree of module activity coordination. Figure 2 illustrates how only a few core processes originally correlated in the normal tissues maintain the correlation in all cancer types. An example of this concordance is the positive coordination between *fumarate*, *succinyl-CoA*, and *urea*, which indicates the expected link between the *citric acid* and *urea* cycles (Figure S3). Unexpectedly, some modules uncorrelated in normal tissue emerge as being coupled in tumors (see Figure S3). Thus, according to cancer metabolic demands, bile acids (e.g. *cholic acid*, *M00104_1*) is positively correlated with *cholesterol* (*M00101*) and *triacylglycerol* (*M00089*). In turn, the negative correlation of the previous metabolites with a *glycosphingolipid* (*globoside*, *M00068*), which is linked to differentiation and antigenicity (42), is lost in cancer. Similarly, *cholesterol* is

positively correlated with *nucleotide sugar biosynthesis* (*M00554* and *M00632_1*), but another *glycosphingolipid* (*ganglioside*, *M00069*) is negatively correlated with this process only in normal tissue. Collectively, these results further highlight the complexity of metabolic reprogramming in cancer.

Available data on survival of cancer cell lines after extensive KD (Table 2 and references therein) allowed the model to be used to relate module activity to cell survival in cell lines. Positive correlations between module activity and cell viability (see Table 1) indicate that the corresponding module may play an essential role in the corresponding cancer type. Therefore, they can be classified as onco-modules. Such constitutively active modules include common cancer dependencies, like *nucleotide sugar biosynthesis*, necessary for cell proliferation, and *heparan sulfate degradation*, necessary for extracellular matrix biosynthesis and thereby, cancer progression and invasion (37). Conversely, tumor suppressor modules showed negative correlations with cell viability possibly indicating constraints in cancer development and/or progression. These modules include *bile acid biosynthesis* (*M00104*), which produces metabolites known to induce apoptosis and inhibit cancer cell proliferation (43) (Table 1 and Figure 3B). In addition, the study identifies modules with contrary effects depending on the tissue of origin, which probably indicate specific cancer dependencies. For example, *inosine monophosphate biosynthesis* is positively (bone) or negatively (prostate) correlated depending on whether there is also reduced or enhanced oxidative phosphorylation, respectively (44).

The detection of onco-modules and tumor suppressor modules was used to suggest previously unidentified potentially actionable genes (Table 1), because the model proposed predicted an effect of their KD on the activity of the corresponding modules. Recently published extra data on cell survival after massive gene KD (Table 2 and references therein) was used to validate the predictions made, confirming these for 62% of the genes (48 of the 77 predictions) included in 86% of the modules (24 of the 28), which constitutes a high rate of validation.

Given the level of accuracy of the predictions of the model of metabolic module activities the obvious subsequent step was to predict the effect of drugs, with known targets within modules, in order to shed light on their mechanisms of action (MoA). Actually, components of some metabolic modules are targeted by well-known clinical drugs, such as gemcitabine, which is approved for the treatment of several cancer types (See Table S14 and specifically DB00441 entry in DrugBank). This drug is a nucleoside

analog that impairs DNA synthesis by specifically inhibiting the production process of GTP, CDP, and their precursor metabolites. In addition, consistent with recent findings for different cancer types (32,45), targeting *proline* (M00015), and less frequently *serine* (M00020) metabolism, may be efficient strategies for cancer treatment.

Additional observations may extend the applications of cancer drugs. The predicted activation of *isoprenoid biosynthesis* (M00095) in breast cancer is consistent with a potentially protective role of simvastatin in progression of this cancer type (46). Following from this observation, predicted metabolic activities support similar applications in bladder and endometrial cancer (47). Furthermore, the use of *pamidronate*, which is currently applied to target bone metastasis in breast cancer and multiple myeloma, and targets *isoprenoid biosynthesis* (M00367) module, might also be applied to bladder and endometrial cancer (48). It is worth pointing out that other *bisphosphonates* show some benefit in these settings and in colorectal cancer (49), which was also predicted in this study. In addition, targeting accumulation of *L-cystathionine* (M00035_1) by *azacitidine*, which causes global DNA hypomethylation, may be useful in at least 10 cancer types. The study also supports drug repurposing, like the potential use of an approved drug for rheumatoid arthritis, *leflunomide* (which targets *UMP biosynthesis*) to treat several cancer types (50). Therefore, this study describes cancer metabolic dependencies that highlight novel therapeutic opportunities either by using current drugs or compounds, or by developing targeted approaches against essential gene products.

It is worth noting that a total of 16 commonly mutated genes from the COSMIC database (51) were present in 11 modules. Although it is likely that some of the samples used in this study contained any of these mutations, the information about the mutational status of the genes in the modules provided in the ICGC repository was scarce and so we could not include this information in the model. However, if this information were available, two scenarios could be considered by the model used here: i) activating mutation (e.g. a translocation to another constitutive promoter), which will be detected in the gene expression level itself, and ii) loss-of-function mutation, which can be simulated in the model by setting the gene expression value to 0 (an expressed non-functional gene is mechanistically equivalent to a non-expressed gene) (52,53).

Although Project Achilles has yielded abundant data, its results are far from exhaustive and, consequently, those obtained here can be considered an underestimate of the actual total number of modules that are essential in cancer.

Summarizing, changes in metabolic processes play a key role in cancer development and progression and this phenomenon is a recognized cancer hallmark (7). However, metabolic maps are complex and understanding the global implications in cancer of changes in activity of processes or components is challenging. Recently, the metabolic map has been decomposed into modules, which consist on sequential reactions representing a summary of fundamental metabolic processes (20). Here we have explored the usefulness of modules to understand cancer metabolic profiles and their relation with cancer outcome and treatment. A simple model is used to predict module activity from the expression levels of its gene components.

In a pan-cancer analysis, we demonstrate that the activity of certain modules change significantly between cancers and the corresponding tissues of origin. We also report changes in the correlated activity of modules. The activity of several modules is significantly associated with cancer prognosis and, moreover, these associations are stronger for the module than for any of their constituent genes. This finding strongly supports the notion that the effect on the phenotype arises from the coordinated activity of the genes in the module. Therefore, essentiality at the gene level would be a consequence of the impact of the activity of the corresponding gene product on the activity of the module. The associations with outcome and cell viability allow us to coin the concepts of tumor suppressor metabolic modules and onco-modules. Finally, using this modeling framework, we propose potential therapeutic targets to inhibit metabolic reprogramming in cancer.

Certainly, the metabolic modules used in this modelling framework describe only a limited (although representative) portion of the whole known map of human metabolism. Therefore, the model presented here provides mechanistic insights into cell metabolic activities that are significantly linked to complex phenotypes, such as cancer prognosis, but probably has limitations in the accurate prediction of the fate of specific metabolites or phenotypes not affected by the metabolites resulting from the 95 metabolic modules used in the model. More comprehensive models that encompass larger portions of the metabolism will, no doubt, increase the reliability of the predictions. We anticipate that the data and models produced will play an increasingly important role in personalized treatment (54).

Acknowledgments

This work is supported by grants BIO2014-57291-R and SAF2017-88908-R from the Spanish Ministry of Economy and Competitiveness (MINECO) to J. Dopazo, grant PI15/00854 from the FIS to M.A. Pujana, “Plataforma de Recursos Biomoleculares y Bioinformáticos” PT17/0009/0006 from the ISCIII, co-funded with European Regional Development Funds (ERDF) to J. Dopazo, FP7-PEOPLE-2012-ITN MLPM2012 (ref. 318861) from the EU FP7 to J. Dopazo and EU H2020-INFRADEV-1-2015-1 ELIXIR-EXCELERATE (ref. 676559) to J. Dopazo.

References

1. Warburg O. The metabolism of carcinoma cells. *J Cancer Res* **1925** 9:148-63
2. Carracedo A, Cantley LC, Pandolfi PP. Cancer metabolism: fatty acid oxidation in the limelight. *Nat Rev Cancer* **2013**;13:227-32
3. Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. *Cell* **2008**;134:703-7
4. Deberardinis RJ, Sayed N, Ditsworth D, Thompson CB. Brick by brick: metabolism and tumor cell growth. *Curr Opin Genet Dev* **2008**;18:54-61
5. Vander Heiden MG. Targeting cancer metabolism: a therapeutic window opens. *Nat Rev Drug Discov* **2011**;10:671-84
6. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **2009**;462:739-44
7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* **2011**;144:646-74
8. Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol* **2013**;31:522-9
9. Anastasiou D, Yu Y, Israelsen WJ, Jiang JK, Boxer MB, Hong BS, *et al.* Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nat Chem Biol* **2012**;8:839-47
10. Pfister SX, Markkanen E, Jiang Y, Sarkar S, Woodcock M, Orlando G, *et al.* Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. *Cancer Cell* **2015**;28:557-68
11. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, *et al.* A community-driven global reconstruction of human metabolism. *Nat Biotechnol* **2013**;31:419-25
12. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, *et al.* Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci Signal* **2015**;8:ra130
13. Hidalgo MR, Cubuk C, Amadoz A, Salavert F, Carbonell-Caballero J, Dopazo J. High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget* **2017**;8:5160-78

14. Amadoz A, Sebastian-Leon P, Vidal E, Salavert F, Dopazo J. Using activation status of signaling pathways as mechanism-based biomarkers to predict drug sensitivity. *Scientific reports* **2015**;5:18494
15. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **2014**;15:107-20
16. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* **2011**;7:501
17. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell systems* **2017**
18. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **2010**;11:R25
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2007**;8:118-27
20. Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, Kanehisa M. Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J Chem Inf Model* **2013**;53:613-22
21. Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, *et al.* Improving metabolic flux predictions using absolute gene expression data. *BMC systems biology* **2012**;6:1
22. Sebastian-Leon P, Carbonell J, Salavert F, Sanchez R, Medina I, Dopazo J. Inferring the functional effect of gene expression changes in signaling pathways. *Nucleic Acids Res* **2013**;41:W213-7
23. Montaner D, Minguez P, Al-Shahrour F, Dopazo J. Gene set internal coherence in the context of functional profiling. *BMC Genomics* **2009**;10:197
24. Jensen PA, Lutz KA, Papin JA. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC systems biology* **2011**;5:147
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **1995**;57:289-300
26. Cox D. Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)* **1972**;34:187-220
27. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **2012**;13:227
28. Akbani R, Ng PKS, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, *et al.* A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nature communications* **2014**;5:3887
29. Mullarky E, Lucki NC, Zavareh RB, Anglin JL, Gomes AP, Nicolay BN, *et al.* Identification of a small molecule inhibitor of 3-phosphoglycerate dehydrogenase to target serine biosynthesis in cancers. *Proceedings of the National Academy of Sciences* **2016**;113:1778-83
30. Shaul YD, Freinkman E, Comb WC, Cantor JR, Tam WL, Thiru P, *et al.* Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **2014**;158:1094-109
31. Kuzu OF, Noory MA, Robertson GP. The role of cholesterol in cancer. *Cancer research* **2016**;76:2063-70

32. Sahu N, Cruz DD, Gao M, Sandoval W, Haverty PM, Liu J, *et al.* Proline Starvation Induces Unresolved ER Stress and Hinders mTORC1-Dependent Tumorigenesis. *Cell metabolism* **2016**;24:753-61
33. Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics* **2009**;1:239-59
34. Dillon BJ, Prieto VG, Curley SA, Ensor CM, Holtsberg FW, Bomalaski JS, *et al.* Incidence and distribution of argininosuccinate synthetase deficiency in human cancers. *Cancer* **2004**;100:826-33
35. Wang Q, Tiffen J, Bailey CG, Lehman ML, Ritchie W, Fazli L, *et al.* Targeting amino acid transport in metastatic castration-resistant prostate cancer: effects on cell cycle, cell growth, and tumor development. *Journal of the National Cancer Institute* **2013**;105:1463-73
36. Coelho M, Moz M, Correia G, Teixeira A, Medeiros R, Ribeiro L. Antiproliferative effects of β -blockers on human colorectal cancer cells. *Oncology reports* **2015**;33:2513-20
37. Sasisekharan R, Shriver Z, Venkataraman G, Narayanasami U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nature Reviews Cancer* **2002**;2:521-8
38. Huang Y, Keen JC, Pledge A, Marton LJ, Zhu T, Sukumar S, *et al.* Polyamine analogues down-regulate estrogen receptor α expression in human breast cancer cells. *Journal of Biological Chemistry* **2006**;281:19055-63
39. Fadare O, Tavassoli FA. Clinical and pathologic aspects of basal-like breast cancers. *Nature Clinical Practice Oncology* **2008**;5:149-59
40. Patra KC, Hay N. The pentose phosphate pathway and cancer. *Trends in biochemical sciences* **2014**;39:347-54
41. Dimitrieva S, Schlapbach R, Rehrauer H. Prognostic value of cross-omics screening for kidney clear cell renal cancer survival. *Biology direct* **2016**;11:68
42. Hakomori S-i. Glycosphingolipids as differentiation-dependent, tumor-associated markers and as regulators of cell proliferation. *Trends in biochemical sciences* **1984**;9:453-9
43. Martinez JD, Stratagoules ED, LaRue JM, Powell AA, Gause PR, Craven MT, *et al.* Different bile acids exhibit distinct biological effects: the tumor promoter deoxycholic acid induces apoptosis and the chemopreventive agent ursodeoxycholic acid inhibits cell proliferation. *Nutrition and Cancer* **1998**;31:111-8
44. Newman AC, Maddocks OD. One-carbon metabolism in cancer. *British journal of cancer* **2017**
45. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. *Trends in biochemical sciences* **2014**;39:191-8
46. Ahern TP, Lash TL, Damkier P, Christiansen PM, Cronin-Fenton DP. Statins and breast cancer prognosis: evidence and opportunities. *The lancet oncology* **2014**;15:e461-e8
47. Wang G, Cao R, Wang Y, Qian G, Dan HC, Jiang W, *et al.* Simvastatin induces cell cycle arrest and inhibits proliferation of bladder cancer cells via PPAR γ signalling pathway. *Scientific reports* **2016**;6
48. Coleman RE. Management of bone metastases. *The Oncologist* **2000**;5:463-70
49. Rennert G, Pinchev M, Rennert HS, Gruber SB. Use of bisphosphonates and reduced risk of colorectal cancer. *Journal of Clinical Oncology* **2011**;29:1146-50
50. Mathur D, Stratikopoulos E, Ozturk S, Steinbach N, Pegno S, Schoenfeld S, *et al.* PTEN regulates glutamine flux to pyrimidine synthesis and sensitivity to dihydroorotate dehydrogenase inhibition. *Cancer discovery* **2017**

51. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* **2015**;43:D805-D11
52. Hernansaiz-Ballesteros RD, Salavert F, Sebastian-Leon P, Aleman A, Medina I, Dopazo J. Assessing the impact of mutations found in next generation sequencing data over human signaling pathways. *Nucleic Acids Res* **2015**;43:W270-5
53. Salavert F, Hidalgo MR, Amadoz A, Cubuk C, Medina I, Crespo D, *et al.* Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. *Nucleic Acids Res* **2016**;44:W212-6
54. Dopazo J. Genomics and transcriptomics in drug discovery. *Drug Discov Today* **2014**;19:126-32

Figure Legends

Figure 1. Heatmap with the significant (FDR-adjusted p-value <0.05) changes in module activity when the 14 cancers analyzed were compared with the corresponding tissue of origin. Activity up-regulation is represented in red and downregulation in blue. The left-most column represents modules in which one or several gene products are targets of cancer drugs; the second column represents modules in which one or several gene products are targets of other types of drugs; the third column represents the general metabolic categories: carbohydrate (CH), amino acid (AA), lipid (LP) or nucleotide (NT).

Figure 2. Changes in correlations of module activities from the normal tissue (inner circle) to the corresponding cancer type (outer circle). The proportion of positive correlations in the activity of the modules is represented in red, while the proportion of negative correlations is represented in blue.

Figure 3. Correlation between increase in module activity, expressed as log-fold-change (X axis) and cell survival (Y axis) corresponding to gene KDs in A) *Heparan sulfate degradation* module and B) *Bile acid biosynthesis* module.

Figure 4. Graph showing relative cell proliferation upon *UPBI* expression depletion (two different MISSION shRNAs were used as detailed in the inset) or transduced with control vector pLKO.1. The asterisk indicates significant differences (Mann-Whitney test p-values < 0.01) and the range of reduction (%) of cell proliferation is also shown. The prediction of *UPBI* essentiality made by the model in lines AGS and MKN45 (stomach gastric adenocarcinoma) was confirmed by a relatively more sensitive behaviors, while *UPBI* does not seem to be relatively sensitive in SW480 (colon adenocarcinoma), as predicted by the model as well.

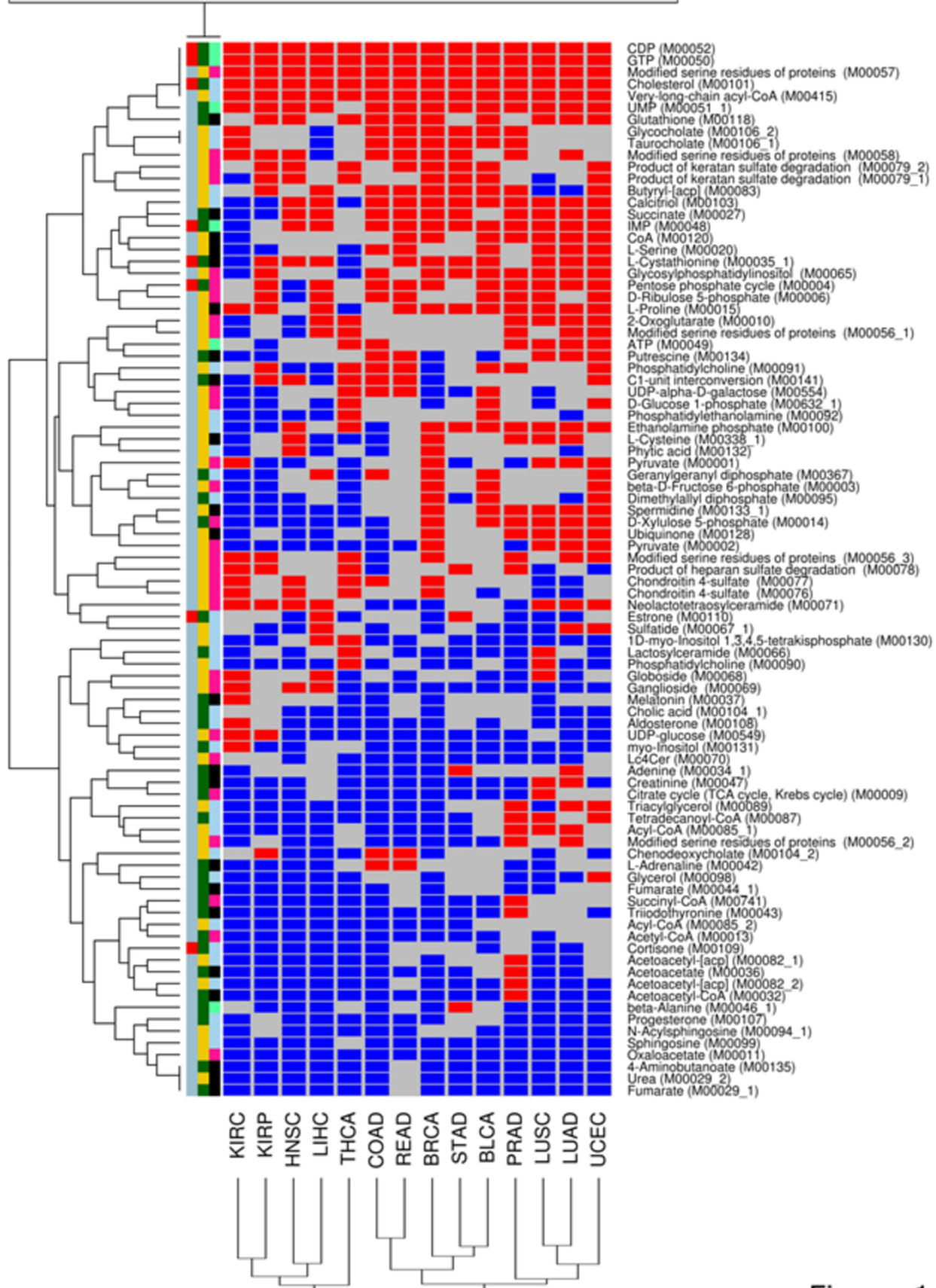
Figure 5. K-M plots showing the relationship between module activity and patient survival in different cancer types. High and low module activity groups were defined by patients in the 80% and 20% percentiles of module activity, respectively. The X axis shows time in months and the number of patients at risk in the high activity and low

activity groups. A) *Pentose phosphate pathway* in LICH; B) *C5 isoprenoid biosynthesis* in LGG; C) *Propanoyl-CoA metabolism* in KIRC and D) *Guanine ribonucleotide biosynthesis* in KIRP.

Table 1. Essential modules. The first column contains the name of the module; the second column the genes knocked down in Project Achilles; the third column lists the other genes in the module, whose inhibition is predicted to cause inhibition of the corresponding module and therefore the same effect as the genes in the second column (codes are * confirmed effect, + inconclusive effect, - no information available for them); the fourth column states the correlation coefficient (r), whose positive and negative values respectively indicate an onco-module and a tumor suppressor module; the fifth column the p-value; the sixth column the cancer tissue from which the cell lines were derived; the seventh column lists the number of different cell lines derived from each tissue; the eighth column the metabolite class (LP: lipid, CH: carbohydrate, AA: amino acid and NT: nucleotide) and the last column the final metabolite of the module.

Module	KD Genes	Predicted KD	r	p-value	Tissue	Cell Lines	MC	Module End Metabolite
Bile acid biosynthesis	<i>AKR1D1 CYP8B1 SLC27A5</i>	<i>CYP27A1- AMACR- ACOX2* CYP7A1 HSD17B4* SCP2- HSD3B7* ACOT8*</i>	-0.883	0.003	Urinary tract	3	LP	C00695: Cholic acid
Dermatan sulfate degradation	<i>IDS ARSB HYAL1</i>	<i>IDUA- HYALA- SPAM1 HYAL3* HYAL2*</i>	-0.812	0	Bone	6	CH	G00872: Chondroitin 4-sulfate
C10-C20 isoprenoid biosynthesis	<i>IDI1 FDPS GGPS1</i>	<i>IDI2*</i>	-0.692	0.016	Stomach	4	LP	C00353: Geranylgeranyl diphosphate
Chondroitin sulfate degradation	<i>ARSB HYAL1</i>	<i>HYALA- SPAM1 HYAL3* HYAL2*</i>	-0.662	0.019	Bone	6	CH	G00872: Chondroitin 4-sulfate
Inosine monophosphate biosynthesis	<i>ATIC ADSL PAICS PFAS</i>	<i>PPAT GART</i>	-0.622	0.035	Prostate	3	NT	C00130: IMP
Serine biosynthesis	<i>PSAT1</i>	<i>PHGDH* PSPH*</i>	-0.61	0.03	Breast	13	AA	C00065: L-Serine
Leucine degradation	<i>DLD BCKDHA IVD BCAT1</i>	<i>BCKDHB HMGCL HMGCLL1* AUH MCCC1 MCCC2* DBT* BCAT2</i>	-0.601	0.043	Urinary tract	3	AA	C00164: Acetoacetate
beta-Oxidation	<i>ACAA1 HADHB EHHADH ECHS1</i>	<i>ACAA2* HADH HADHA</i>	-0.58	0	Esophagus	10	LP	C02593: Tetradecanoyl-CoA
Nucleotide sugar biosynthesis	<i>PGM1 HK2 HK3 UGP2</i>	<i>PGM2 HK1* HKDC1*</i>	-0.552	0.002	Skin	7	CH	C00029: UDP-glucose
Pentose phosphate pathway (Pentose phosphate cycle)	<i>RPE PGD PGLS</i>	<i>GPI* TKT* TKTL1* TKTL2* RPIA* RPEL1* G6PD* TALDO1</i>	-0.541	0	Breast	13	CH	C01172: beta-D-Glucose 6-phosphate
Sphingosine degradation	<i>SPHK1 SGPL1</i>	<i>SPHK2*</i>	-0.532	0.017	Esophagus	10	LP	C00346: Ethanolamine phosphate
Ceramide biosynthesis	<i>CERS5 DEGS2 DEGS1 SPTLC1 SPTLC2</i>	<i>CERS1* CERS3 CERS6* CERS2 CERS4* SPTLC3* KDSR*</i>	-0.523	0.045	Prostate	3	LP	C00195: N-Acylsphingosine
Melatonin biosynthesis	<i>AANAT</i>	<i>ASMT* DDC* TPH2* TPH1</i>	-0.515	0.044	Pancreas	16	AA	C01598: Melatonin
Inositol phosphate metabolism	<i>ITPK1 IPMK</i>	<i>IPPK</i>	-0.505	0.025	Kidney	10	LP	C01204: Phytic acid
Glycosphingolipid biosynthesis, ganglio series	<i>ST8SIA1 ST3GAL5</i>		-0.5	0	Haematopoietic	27	CH	G00118: Ganglioside (GT3)
C10-C20 isoprenoid biosynthesis.	<i>IDI1 FDPS GGPS1</i>	<i>IDI2*</i>	0.5	0.022	Skin	7	LP	C00353: Geranylgeranyl diphosphate
Heparan sulfate degradation	<i>IDS GNS</i>	<i>SGSH HPSE2* IDUA HGSNAT- NAGLU* GUSB*</i>	0.554	0	CNS	35	CH	G02632: glycan
Pyrimidine degradation	<i>DPYD DPYS</i>	<i>UPB1*</i>	0.574	0.035	Skin	7	NT	C00099: beta-Alanine
Pyrimidine degradation	<i>DPYD DPYS</i>	<i>UPB1*</i>	0.574	0.035	Skin	7	NT	C05145: 3-Aminoisobutyric acid
Conjugated bile acid biosynthesis	<i>SLC27A5 BAAT</i>		0.6	0.006	Kidney	10	LP	C05122: Taurocholate
Conjugated bile acid biosynthesis	<i>SLC27A5 BAAT</i>		0.6	0.006	Kidney	10	LP	C01921: Glycocholate
Methionine salvage pathway	<i>AD11 MRI1 SRM AMD1 MAT2B MAT1A</i>	<i>APIP* MTAP* MAT2A* ENOPH1*</i>	0.618	0.032	Soft tissue	2	AA	C00147: Adenine
Polyamine biosynthesis	<i>SRM AMD1</i>	<i>AZIN2* AGMAT*</i>	0.639	0	Haematopoietic	27	AA	C00315: Spermidine
Nucleotide sugar biosynthesis	<i>PGM1 HK2 HK3 UGP2</i>	<i>PGM2* HK1 HKDC1*</i>	0.641	0.025	Urinary tract	3	CH	C00029: UDP-glucose
Inosine monophosphate biosynthesis	<i>ATIC ADSL PAICS PFAS</i>	<i>PPAT GART</i>	0.677	0	Bone	6	NT	C00130: IMP
Dermatan sulfate degradation	<i>IDS ARSB HYAL1</i>	<i>IDUA1* HYALA- SPAM1* HYAL3 HYAL2*</i>	0.692	0	Esophagus	10	CH	G00872: Chondroitin 4-sulfate
Pyrimidine degradation	<i>DPYD DPYS</i>	<i>UPB1</i>	0.762	0.037	Stomach	4	NT	C00099: beta-Alanine
Pyrimidine degradation	<i>DPYD DPYS</i>	<i>UPB1</i>	0.762	0.037	Stomach	4	NT	C05145: 3-Aminoisobutyric acid

- Cancer specific drug
- Is a target of drug
- Carbohydrate metabolism
- ▲ Up Regulation $p \leq 0.05$
- Non cancer specific drug
- Is not a target of drug
- Amino acid metabolism
- ▲ Down Regulation $p \leq 0.05$
- or
- Lipid metabolism
- ▲ Non Significant $p > 0.05$
- Is not a target of drug
- Nucleotide metabolism



Downloaded from cancerres.aacrjournals.org on April 2, 2019. © 2018 American Association for Cancer Research. Figure 1

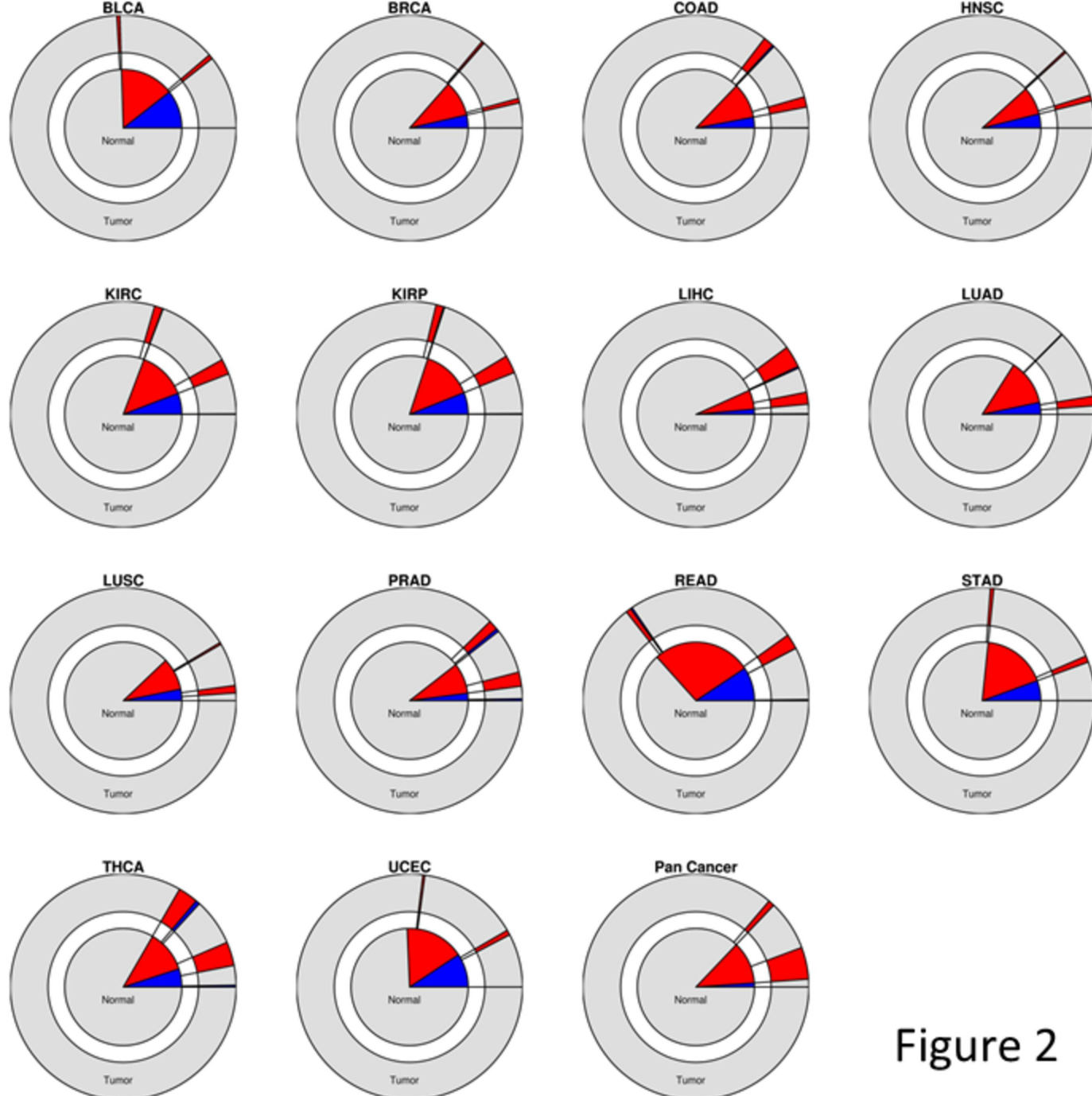
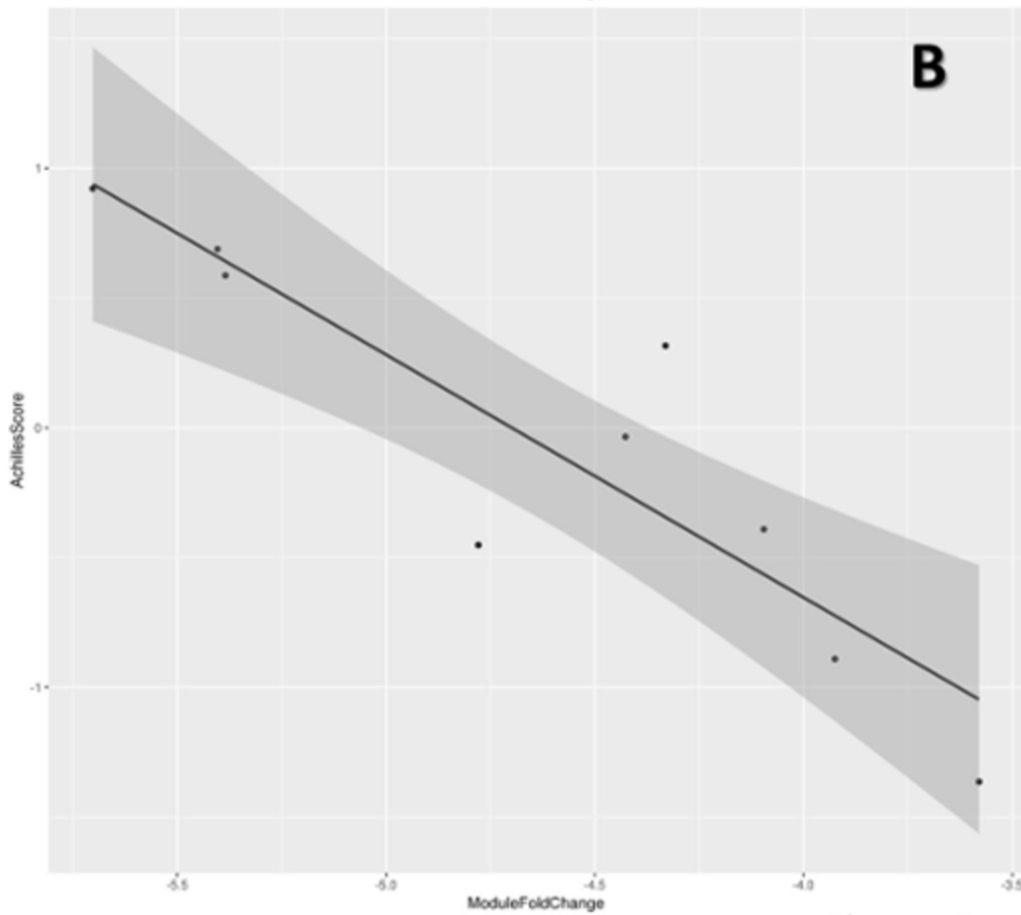
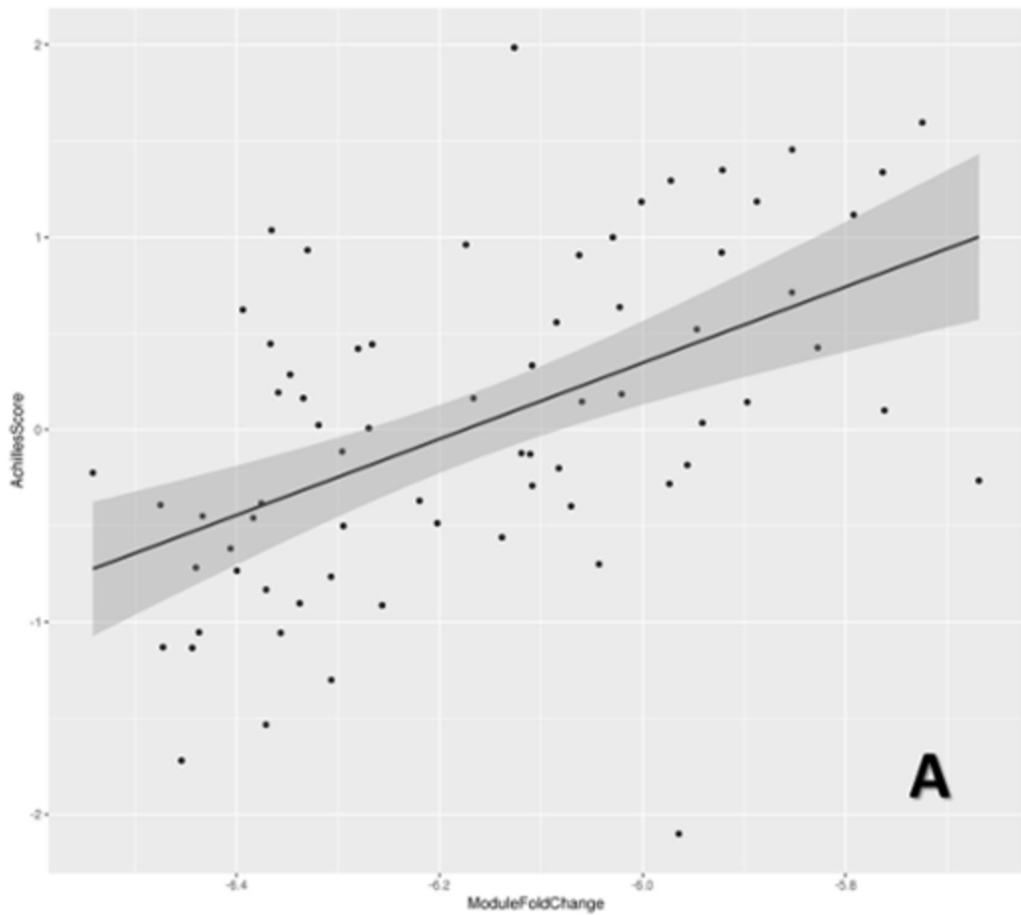
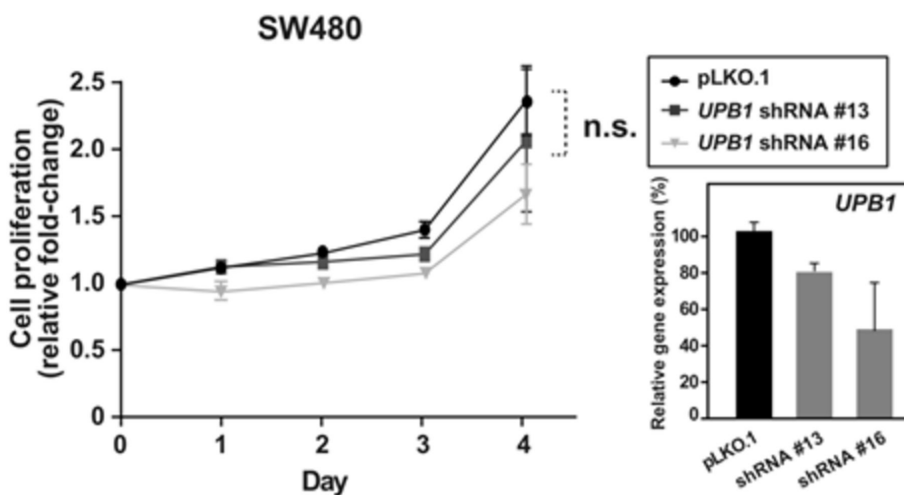
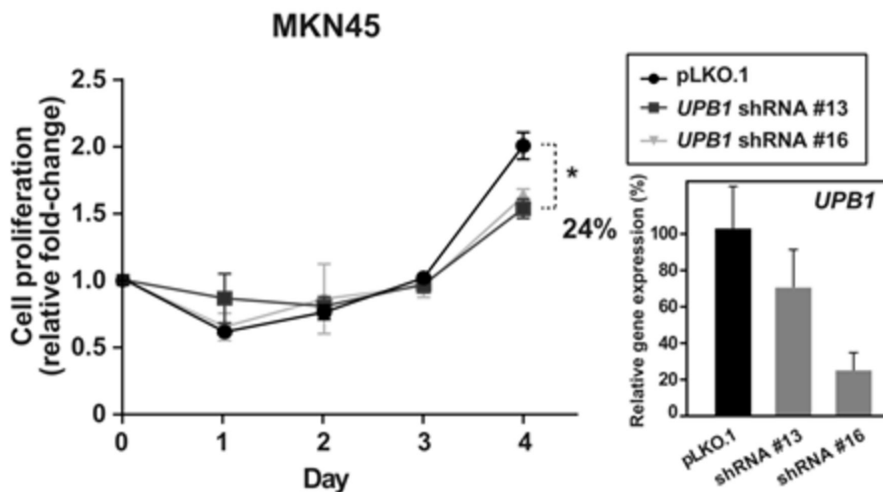
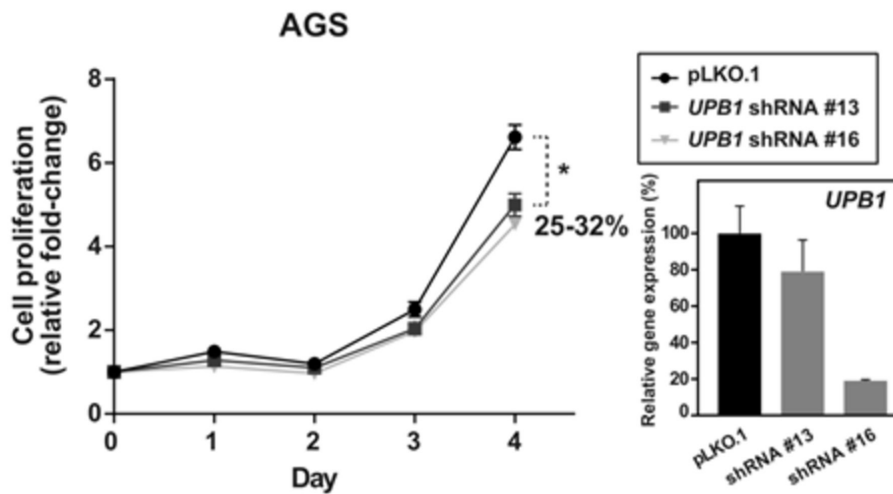
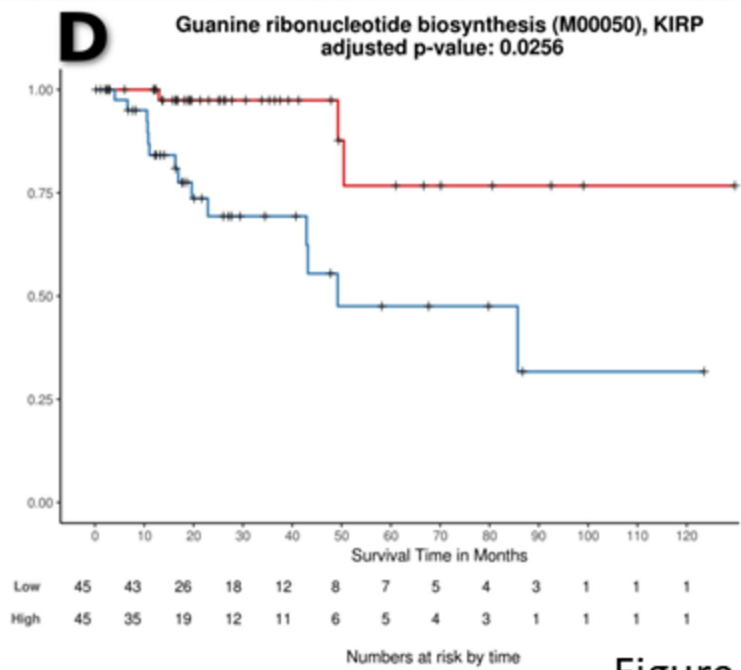
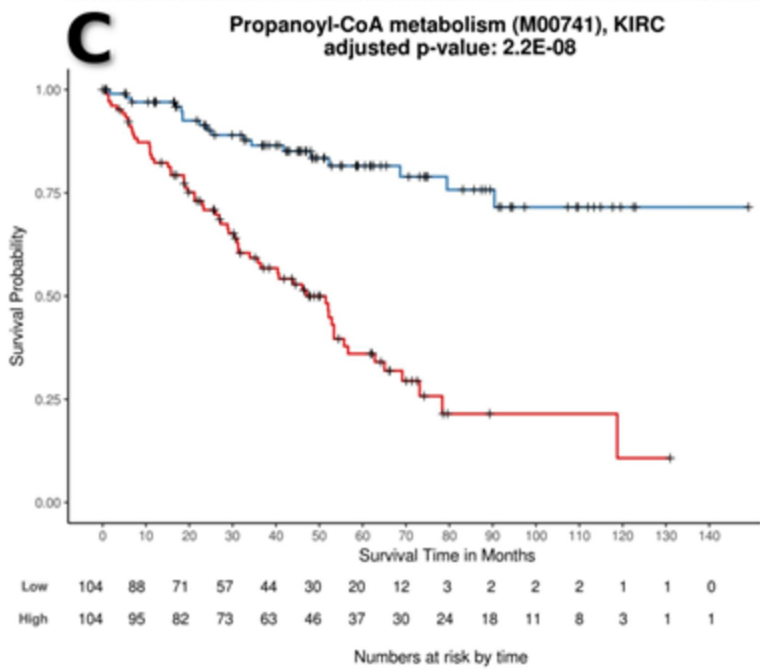
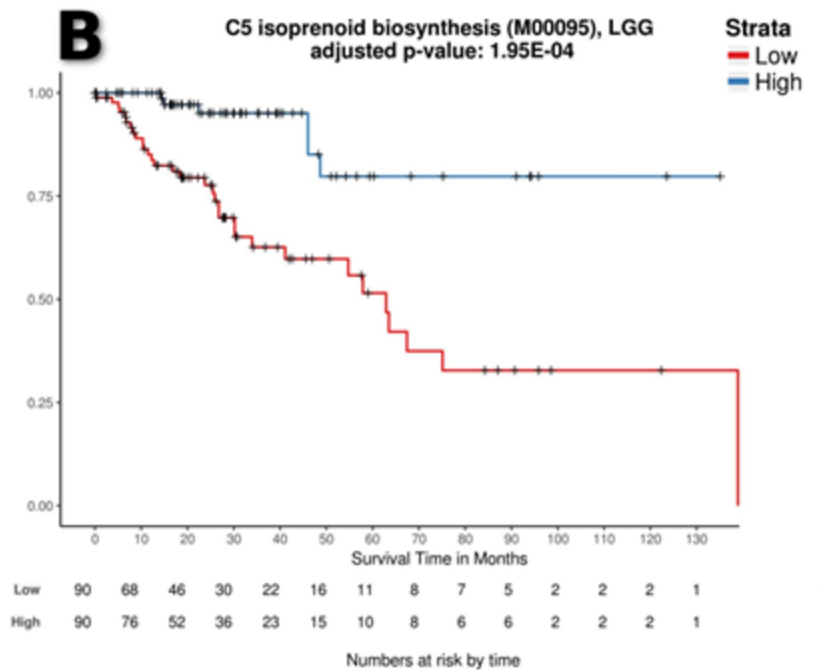
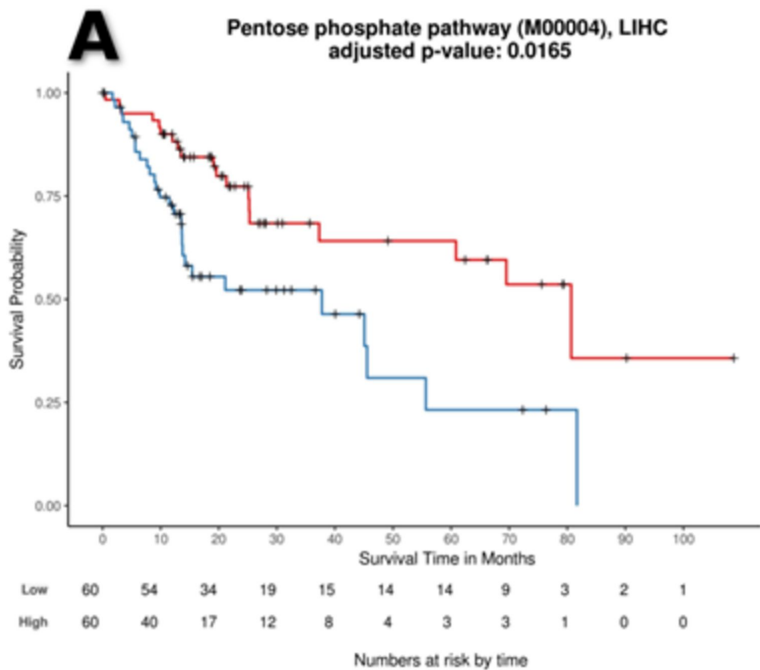


Figure 2







Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape

Cankut Cubuk, Marta R Hidalgo, Alicia Amadoz, et al.

Cancer Res Published OnlineFirst August 22, 2018.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-17-2705
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2018/11/03/0008-5472.CAN-17-2705.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/early/2018/08/22/0008-5472.CAN-17-2705 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.