ANALYSIS

Heterogeneity of tumor-induced gene expression changes in the human metabolic network

Jie Hu[1], Jason W Locasale[2], Jason H Bielas[3,4], Jacintha O'Sullivan[5], Kieran Sheahan[5], Lewis C Cantley[6,7], Matthew G Vander Heiden[8,9] & Dennis Vitkup[1,10]

Reprogramming of cellular metabolism is an emerging hallmark of neoplastic transformation. However, it is not known how the expression of metabolic genes in tumors differs from that in normal tissues, or whether different tumor types exhibit similar metabolic changes. Here we compare expression patterns of metabolic genes across 22 diverse types of human tumors. Overall, the metabolic gene expression program in tumors is similar to that in the corresponding normal tissues. Although expression changes of some metabolic pathways (e.g., upregulation of nucleotide biosynthesis and glycolysis) are frequently observed across tumors, expression changes of other pathways (e.g., oxidative phosphorylation) are very heterogeneous. Our analysis also suggests that the expression changes of some metabolic genes (e.g., isocitrate dehydrogenase and fumarate hydratase) may enhance or mimic the effects of recurrent mutations in tumors. On the level of individual biochemical reactions, many hundreds of metabolic isoenzymes show significant and tumor-specific expression changes. These isoenzymes are potential targets for anticancer therapy.

All tumors share a common phenotype of uncontrolled cell proliferation. To support the synthesis of biomass components and to generate the energy required for cellular growth, cancer cells have to reshape the regulatory and functional properties of their metabolic networks. Over 80 years ago, Otto Warburg identified a shift from oxidative to fermentative metabolism as a common physiological trait of tumor cells[1]. Despite this early insight into cancer metabolism,

the main focus of cancer research shifted toward the analysis of signaling, gene-regulatory and genetic perturbations in various tumors[2,3]. Recently, however, there has been a resurgence of interest in cancer metabolism[4–6]. An important factor contributing to this renaissance is the observation that many signaling pathways altered in cancer are key regulators of the human metabolic network[5]. In addition, the therapeutic potential of metabolic targets in cancer has been rediscovered[7,8].

Here we use a large compendium of gene expression profiles accumulated over the last decade[9,10] to comprehensively analyze tumor-induced changes in mRNA expression of human metabolic genes across 22 diverse cancer types. To minimize confounding metabolic adaptations that may arise from tissue culture conditions, we analyzed only microarray data obtained from biopsies of primary tumors. We compared gene expression in tumors and corresponding normal tissues at several conceptual levels of biochemical organization: at the global network level, at the level of individual biochemical pathways and at the level of single enzymatic reactions. The focus on the human metabolic network and the analysis of the large collection of tumor and normal samples allowed us to gain statistical power and establish significance for many expression patterns that, to our knowledge, have not been reported.

RESULTS
Global changes in metabolic gene expression
To understand metabolic gene expression in different cancers, we assembled a comprehensive collection of >2,500 microarray measurements spanning 22 different tumor types (Online Methods and **Supplementary Table 1**). We analyzed only expression data obtained using the most comprehensive human expression array platform (HG U133 Plus 2.0; **Supplementary Table 2**). Differences in expression of metabolic genes between tumor and corresponding normal tissues for the same tumor types obtained in independent studies and with different microarray platforms (**Supplementary Table 3**) were highly correlated (average Spearman's rank correlation coefficient = 0.63), confirming the generality of the observed expression patterns (**Supplementary Fig. 1**).

Using the assembled expression compendium, we first investigated the global shifts in metabolic gene expression between and within different cancers and their corresponding normal tissues. For this analysis we used 1,421 human genes assigned to metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[11]. Using two different measures of divergence between a

Figure 1 Global differences in metabolic gene expression between tumors and normal tissues. Colors represent distributions of the Euclidean expression distance (root mean squared deviation, r.m.s.d.) between different samples of identical normal tissues (Normal$_n$-Normal$_n$, magenta), different samples of identical tumors (Tumor$_n$-Tumor$_n$, cyan), tumors and corresponding normal tissues (Tumor$_n$-Normal$_n$, blue), different tumors (Tumor$_n$-Tumor$_m$, green) and different normal tissues (Normal$_n$-Normal$_m$, red). The distributions were binned for display purposes only. Inset summarizes the average distances between pairs of tissues as a percentage of the average distance between two different normal tissues.

pair of expression profiles[12], the Euclidean distance and the correlation distance (Online Methods), we compared global expression patterns between tumors and normal tissues (**Supplementary Table 2**). The influence of batch effects, which arise from variations in laboratory conditions and measurements, were estimated (Online Methods and **Supplementary Table 4**) and subtracted from expression distances between expression profiles measured in different studies.

Relative differences between the distributions are consistent for the two metrics of expression divergence (**Fig. 1** and **Supplementary Fig. 2**). The expression distance between tumors and corresponding normal tissues (Tumor$_n$-Normal$_n$) is significantly larger than the distance between different samples of the same normal tissues (Normal$_n$-Normal$_n$; Mann-Whitney U test, $P = 10^{-8}$; **Fig. 1**) or between different samples of the same tumors (Tumor$_n$-Tumor$_n$; $P = 4*10^{-7}$). The distance Tumor$_n$-Normal$_n$, however, is significantly smaller than the distance between different tumors (Tumor$_n$-Tumor$_m$; $P = 2*10^{-7}$), which in turn is significantly smaller than the distance between different normal tissues (Normal$_n$-Normal$_m$; $P < 2*10^{-16}$). The average expression distance between two different tumors is ~82% of the average distance between two different normal tissues, whereas the distance between a tumor and a corresponding normal tissue is ~63% of the distance between two different normal tissues. Consequently, although the metabolic expression patterns in different tumors become more similar than in corresponding normal tissues, the overall similarity between the metabolic expression programs of tumors and corresponding original tissues remains high.

Expression changes of individual biochemical pathways

We next analyzed the expression changes associated with individual biochemical pathways defined in the KEGG database[11] (see **Supplementary Table 5** for pathway information and numbering).
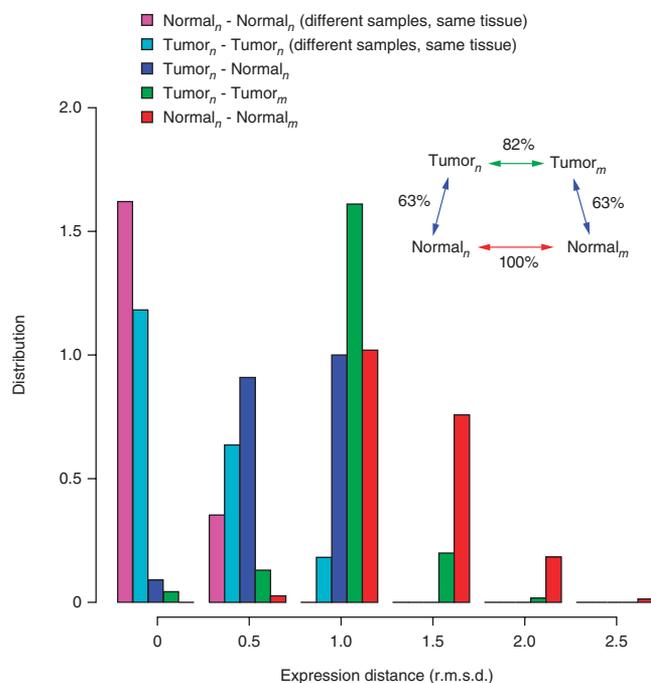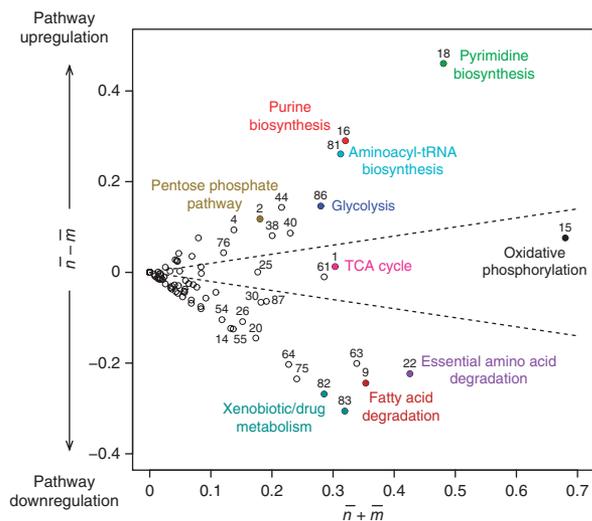
To identify the patterns of up- and downregulation for each metabolic pathway, we determined the significance of its expression changes in the tumor samples relative to the corresponding normal samples using the Wilcoxon signed-rank test, adjusted for multiple hypothesis testing. Based on this analysis, we calculated the average fraction of tumor samples in which each metabolic pathway was significantly (false-discovery rate (FDR)-corrected, $P < 0.05$) upregulated ($\bar{n}$) and downregulated ($\bar{m}$) across 22 cancer types (**Fig. 2**). The observed behaviors of the metabolic pathways highlighted in **Figure 2** were statistically significant when compared with the null distributions of $(\bar{n}+\bar{m})$ and $(|\bar{n}-\bar{m}|)$ values based on 1,000 random permutations of the expression data (**Supplementary Fig. 3**). The general expression patterns observed with the KEGG pathways were similar to results obtained using the BioCyc pathway definitions[13] (**Supplementary Fig. 4a** and **Supplementary Table 6**), suggesting the robustness of the results with respect to alternative pathway definitions.

As expected, pathways responsible for production of biomass components that are essential for cell division, such as pyrimidine and purine biosynthesis, are significantly upregulated in many tumor samples (**Fig. 2**). Along with these two pathways, glycolysis is also

Figure 2 Expression of individual metabolic pathways in tumors. The biochemical pathways defined in the KEGG database (see **Supplementary Table 5** for pathway numbering) are shown in the coordinates of $\bar{n}+\bar{m}$ and $\bar{n}-\bar{m}$, where $\bar{n}$ is the average fraction of tumor samples in which a pathway is significantly upregulated, and $\bar{m}$ is the average fraction in which a pathway is significantly downregulated. The averages $\bar{n}$ and $\bar{m}$ were calculated across all 22 tumors. The up- and downregulation significance was determined using Wilcoxon signed-rank test (FDR-corrected, $P < 0.05$, see **Supplementary Fig. 4b** for the same analysis with FDR = 0.2). The dashed lines demarcate the region where $\bar{n}-\bar{m}$ is <20% of $\bar{n}+\bar{m}$ and are shown for visualization purposes only. Metabolic pathways without significant expression changes are primarily clustered on the left of the figure. Pathways that are often significantly upregulated (high $\bar{n}$ values) occupy positions in the upper right corner, whereas pathways that are primarily downregulated (high $\bar{m}$ values) occupy positions in the lower right corner. Highly heterogeneous pathways that show, in different tumors, both significant up- and downregulation are clustered on the right near zero on the vertical axis.

significantly upregulated in many tumor samples, consistent with an enhanced glucose uptake frequently observed in tumors[14]. Among other pathways displaying frequent and significant upregulation are pathways related to protein synthesis (aminoacyl-tRNA biosynthesis) and glycoprotein biosynthesis (N-glycan biosynthesis). In tumor samples where the expression of the aforementioned pathways is not significantly changed, the overall metabolic gene expression was mostly either downregulated or not significantly changed. Specifically, this is the case for 72% of tumor samples with no significant change in glycolysis pathway expression, 84% of tumor samples with no change in purine biosynthesis pathway expression and 88% of tumor samples with no change in pyrimidine biosynthesis pathway expression.

In contrast to the biosynthesis pathways, pathways responsible for degradation of essential amino acids (valine, leucine and isoleucine degradation), cofactors (retinol metabolism) and fatty acids are frequently and significantly downregulated. Notably, two metabolic pathways that are also consistently downregulated across various tumors are xenobiotic and drug metabolism. These processes are responsible for detoxification and disposal of compounds foreign to the normal biochemistry of the cell. Several previous studies have shown that polymorphisms in cytochrome P450 genes correlate with cancer susceptibility in different types of cancer, including those of the lung, bladder and breast[15,16]. Although the decreased expression of xenobiotic pathways in cancer needs to be validated and its potential causes investigated, the observed downregulation may contribute to the increased sensitivity of cancer cells to chemotherapies.

The heterogeneous behavior of the oxidative phosphorylation and the TCA cycle pathways is also notable (**Fig. 2**). Oxidative phosphorylation shows the most heterogeneous behavior of all metabolic pathways considered. In brain, colon, kidney, pancreatic and thyroid cancers, genes involved in oxidative phosphorylation are significantly downregulated, whereas in breast, leukemia, lung, lymphoma and ovarian cancers, these genes are significantly upregulated (**Supplementary Table 7**). This pattern suggests that the role of oxidative phosphorylation is not universal for all tumors, but possibly reflects the adaptation of different cancers to tissue-specific physiological conditions such as hypoxia, nutrient availability or complement of genetic lesions driving a specific tumor type.

We also explored the heterogeneity of metabolic pathway expression across different samples of the same (or similar) tumor types. Such an analysis (**Supplementary Fig. 5**) showed that oxidative phosphorylation gene expression is not only heterogeneous between different tumor types, but also frequently varies between samples of the same tumor. This observation suggests that the activity of oxidative phosphorylation is influenced not only by the variability of environments across different tumor types, but also by the specific physiological conditions and/or genetic composition of individual tumors in each cancer patient. In contrast, other metabolic pathways showed similar expression patterns across different samples of the same tumor.

Correlation between metabolic pathways and signaling genes
We next investigated correlations between the expression of metabolic pathways and expression of signaling and regulatory genes frequently involved in tumorigenesis. Although several previous studies[17] demonstrated that correlated expression patterns usually cannot be equated with regulation causality, that is, one gene being regulated by the other, significant correlations could still reveal important functional relationships. To find significant expression correlations (Z-score > 2.0; **Supplementary Table 8**) between the 214 nonmetabolic genes annotated in the KEGG signaling and/or cancer pathways

(**Supplementary Table 9**) and the 87 metabolic pathways considered in our analysis, we used the context likelihood of relatedness (CLR) method[18]. The CLR algorithm uses mutual information—an information theory measure that quantifies statistical dependencies between two variables—to identify significant correlations between two expression patterns.

The CLR analysis revealed several notable relationships. The oxidative phosphorylation pathway has high mutual information with the hypoxia-inducible factor (HIF1A) and its negative regulator RBX1. Notably, the oxidative phosphorylation expression is negatively correlated with the expression of HIF1A and is positively correlated with the expression of RBX1. The observed negative correlation between oxidative phosphorylation and HIF1A suggests that the heterogeneity in the expression of oxidative phosphorylation genes (**Fig. 2**) is likely to be influenced by tumor oxygen availability. In addition, the mutual information between HIF1A and glycolysis is not high, probably because although HIF1A is involved in the upregulation of glycolysis specifically under hypoxia, many tumors show upregulated expression of oxidative phosphorylation and may not be hypoxic. In contrast, there is significant mutual information between glycolysis and CDC42, a gene essential for cell cycle progression. Glycolysis is also strongly correlated with expression of RAS and genes from the MAPK pathway, which have been previously implicated in promoting aerobic glycolysis[19]. Apart from glycolysis, CDC42 expression also has high mutual information with other pathways essential for fast cellular growth (such as purine, pyrimidine and amino acid biosynthesis). On the other hand, CDC42 expression is not correlated with the expression of oxidative phosphorylation, suggesting that in fast-growing tumor cells glucose fermentation dominates.

Principal component analysis of pathway expression changes
Individual metabolic pathways do not function in isolation. In contrast, they display highly correlated and interdependent patterns of gene expression. Therefore, we used principal component analysis (PCA)[20] to better understand the joint behavior of metabolic pathways in cancer (**Table 1**; see **Supplementary Tables 10** and **11** for similar analysis based on alternative methods for calculating principal components). To reduce noise associated with heterogeneous expression of individual pathways, we considered the expression changes in the space of nine meta-pathways, such as glycolysis and nucleotide biosynthesis, that represent major metabolic processes. Combined, the first three principal components were able to capture ~85% of the expression variance of the meta-pathways.

The first principal component accounts for ~62% of the variance in the meta-pathway expression changes. As all pathway weights for this component have the same sign and similar values, it represents an approximately uniform shift in the overall expression of metabolic genes. The projection of cancer samples onto the plane defined by the first and second principal components (**Supplementary Fig. 6**) shows that tumors of the digestive system (colon, kidney and liver) have a high positive shift along this component, suggesting an overall decrease in metabolic gene expression. In contrast, several other tumors (e.g., cervix and lymphomas) show an overall increase in metabolic gene expression. It is likely that the observed shifts along the first principal component reflect, at least to some extent, the loss of specific metabolic functions required by the corresponding normal tissues, and the switch to a metabolic program primarily focused on growth and proliferation. This may account for the overall decrease in expression of metabolic genes observed in tumors of the gastrointestinal system that normally have high metabolic gene expression unique to these differentiated tissues.

Table 1 Principal component analysis (PCA) of gene expression in major metabolic processes.

Variables (meta-pathways)	Number of genes involved in each meta-pathway	Weights in the 1st component	Weights in the 2nd component	Weights in the 3rd component
Oxidative phosphorylation	135	−0.40	0.21	0.67
Glycolysis	24	−0.33	0.65	−0.01
Citric acid cycle	21	−0.50	−0.10	0.02
Amino acids biosynthesis	70	−0.27	−0.11	−0.17
Fatty acids and lipids biosynthesis	66	−0.28	0.05	0.09
Nucleotides and nucleosides biosynthesis	54	−0.35	0.34	−0.67
Amino acids degradation	123	−0.31	−0.40	−0.15
Fatty acids and lipids degradation	80	−0.22	−0.35	0.17
Nucleotides and nucleosides degradation	9	−0.26	−0.34	−0.04
Proportion of variance explained by each component	–	0.62	0.16	0.07

The PCA was performed using the average expression changes of genes forming nine meta-pathways representing major biochemical processes. The pathway weights indicate the relative contribution of each meta-pathway to the principal components; weights with identical signs indicate correlated contributions of pathways to a component, while weights with opposite signs indicate negatively correlated contributions. The table show the PCA weights of each meta-pathway for the first three principal components. The first three principal components explain ~62%, ~16% and ~7% of variance in expression data, respectively.

Shifts along the second component, explaining ~16% of the expression variance, involve a change in the expression of glycolysis and nucleotide biosynthesis with a concomitant opposite change in the expression of three catabolic pathways. Because an increased rate of nucleotide biosynthesis is especially important during ribosome biogenesis and chromosomal duplication, our results suggest that rapidly dividing cells appear to rely more on glycolysis than cells growing slowly. Oxidative phosphorylation is also associated with the second component, although with a substantially smaller weight than glycolysis (0.21 versus 0.65). Consequently, along this component glycolysis occurs concurrently with oxidative phosphorylation. Shifts along the third principal component, which explains ~7% of the variance, primarily involve a strong change in the expression of oxidative phosphorylation with a concomitant opposite change in nucleotide biosynthesis. Therefore, a strong upregulation of oxidative phosphorylation along this component is likely to be associated with slower growth rates.

Expression changes of biochemical reactions and isoenzymes
Next, we focused on expression changes associated with individual biochemical reactions, which form the most basic level in hierarchical organization of the human metabolic network. We used 2,307 reactions, each associated with at least one known enzyme in a model of human metabolism[21]. In the human metabolic network and in the networks of other organisms[22], a given biochemical reaction is frequently catalyzed by several different isoenzymes. Isoenzymes may be encoded by separate genes or arise from alternative splicing of the same gene. In the network model we used[21], ~30% of metabolic reactions contain at least two known isoenzymes, and this percentage is even higher (~40%) for the reactions of central carbon metabolism. Different kinetic and regulatory properties of isoenzymes are often fine-tuned to meet specific metabolic requirements of various human tissues[22]. Owing to metabolic demands and constraints different from those of native tissues, it is likely that tumors might preferentially express isoenzymes that facilitate survival and uncontrolled proliferation[23,24].

The heterogeneity of isoenzyme expression across tumors is apparent from the analysis of central metabolism (**Fig. 3** and **Supplementary Table 12**). Although genes encoding glycolytic enzymes are frequently upregulated in tumors, some isoenzymes are downregulated in specific cancers. Gene expression is significantly (Wilcoxon signed-rank test, $P < 0.05$) increased for key enzymes of the pentose phosphate pathway, including both the oxidative and non-oxidative branches. Lactate dehydrogenase is also strongly upregulated, consistent with the high level of lactate production observed in many tumors[25].

Frequent downregulation of the pyruvate dehydrogenase complex likely contributes to a decreased flux of pyruvate into the TCA cycle observed in many tumors. The enzymes essential for purine and pyrimidine synthesis, as well as the glutathione synthetase[26], are strongly upregulated. Although glutaminase, an enzyme important for the TCA cycle anaplerosis, is generally downregulated, it has been demonstrated that this enzyme is strongly upregulated post-transcriptionally by the MYC-mediated suppression of miR-23a/b[27]. Notably, a recent study[28] suggested that an alternative route for the glutamine-to-glutamate transformation, perhaps through nucleotide biosynthesis amidotransferases, may play an important role in the glutamine-dependent anaplerosis. This hypothesis is consistent with a strong upregulation of the corresponding enzymes (PPAT and CAD) across tumors observed in our analysis.

We investigated changes in relative isoenzyme expression using the Kullback-Leibler (KL) divergence; the KL divergence is an information theory measure used to quantify the difference between two probability distributions. For each biochemical reaction the KL divergence was used to measure shifts in the distribution of isoenzyme expression between tumors and corresponding normal tissues. This analysis demonstrated that, on average, the relative expression patterns of isoenzymes are about two times more similar for different samples of identical normal tissues than for different samples of identical tumors (**Fig. 4a**). But more importantly, both of these distances are significantly smaller than the average distance between isoenzyme expression patterns in tumors and corresponding normal tissues (Mann-Whitney U test, $P < 2*10^{-16}$). This suggests that for many biochemical reactions, neoplastic transformation leads to a significant shift in the relative expression of isoenzymes.

The human aldolase is an example of an enzyme with perturbed expression patterns in tumors (**Fig. 4b**). The enzyme has three main isoforms, A, B and C. Although aldolase A (ALDOA) is preferentially expressed in muscle cells, it is also strongly expressed in most other human tissues. Aldolase B (ALDOB) is preferentially expressed in the liver and aldolase C (ALDOC) in the brain. The expression analysis shows that the expression of ALDOA, relative to the other aldolase isoenzymes, is significantly (Online Methods; Mann-Whitney U test, $P < 0.05$) greater in tumors than in normal tissues. Notably, ALDOA is also highly expressed in developing embryos[29], and therefore may be particularly suitable for metabolic requirements during fast cell proliferation. Indeed, the k_{cat} of ALDOA is substantially higher than that of the other isoenzymes[30].

Another example of an enzyme with perturbed expression patterns is aconitase. Our analysis suggests that the citrate efflux from the TCA cycle is likely to be enhanced in cancers by frequent downregulation
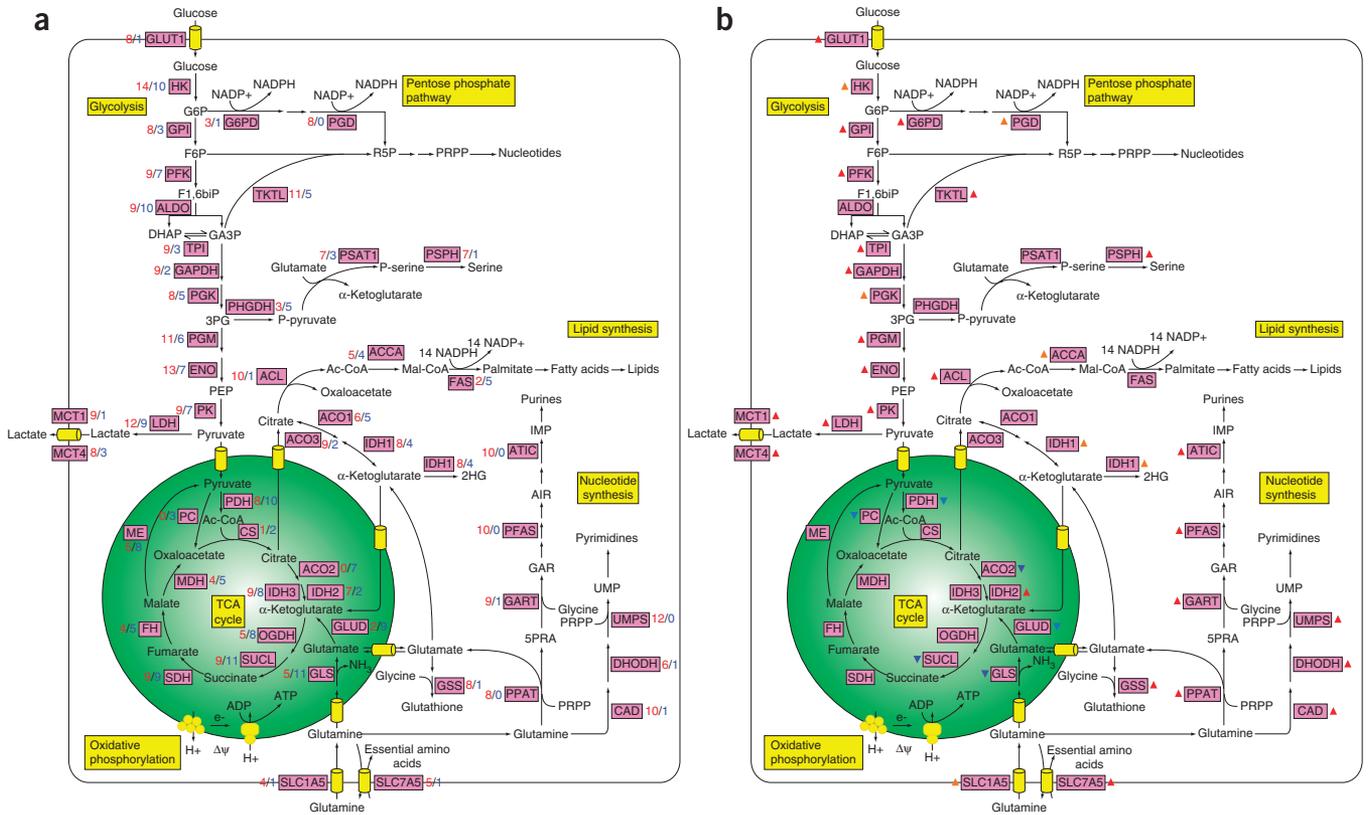
Figure 3 Tumor-induced mRNA expression changes for individual biochemical reactions in central metabolism. (**a**) Each metabolic reaction is marked with the number of tumors (out of 22 considered in our analysis) in which at least one isoenzyme catalyzing the corresponding reaction is significantly (FDR-corrected, $P < 0.05$) upregulated (red) and downregulated (blue). (**b**) Reactions that are significantly upregulated (red triangles) or downregulated (blue triangles) when all isoenzymes and members of the corresponding protein complexes are considered together across all tumors (deep red or deep blue, FDR-corrected, $P < 0.05$; orange or light blue, FDR-corrected, $P < 0.1$). If unmarked, no statistically significant change in mRNA expression was detected.

of the mitochondrial isoform of aconitase (ACO2) (**Fig. 3a,b**). Cytosolic citrate is used to generate acetyl-CoA, an important precursor required for many biosynthetic reactions involved in lipogenesis[31]. Inhibition of the mitochondrial aconitase in normal human tissues[32] and yeast[33] was previously shown to significantly increase the TCA cycle citrate efflux. The strong upregulation of the ATP citrate lyase ACL across tumors (**Fig. 3**) provides additional support for the idea that these changes promote fatty acid biosynthesis in tumors. A recent study showed that an important route for the synthesis of lipogenic

acetyl-CoA under hypoxia[34] is through reductive metabolism of α-ketoglutarate by cytosolic isocitrate dehydrogenase (IDH1) and cytosolic aconitases (ACO1/ACO3). This pathway is also supported by the observed expression patterns because in contrast to the mitochondrial aconitase, the cytoplasmic aconitases and IDH1 (see below) are frequently upregulated in specific cancers (**Fig. 3a**).

To identify specific isoenzymes with frequently perturbed expression profiles, we calculated, for each isoenzyme in every biochemical reaction, the number of tumors in which the fractional expression of

Figure 4 Cancer-induced changes in relative isoenzyme expression. (**a**) The Kullback-Leibler (KL) divergence was used to characterize differences in the relative expression of isoenzymes for all biochemical reactions with multiple isoenzymes. Colors represent distributions of the KL divergence in isoenzyme expression between different samples of identical normal tissues (Normal$_n$-Normal$_n$), different samples of identical tumors (Tumor$_n$-Tumor$_n$), and tumors and corresponding normal tissues (Tumor$_n$-Normal$_n$). Inset summarizes the average KL divergences as a percentage of the average KL divergence between different samples of identical normal tissues. (**b**) Relative expression of the aldolase isoenzymes for kidney, liver, stomach, brain (glioblastoma multiforme) tumors and the corresponding normal tissues.
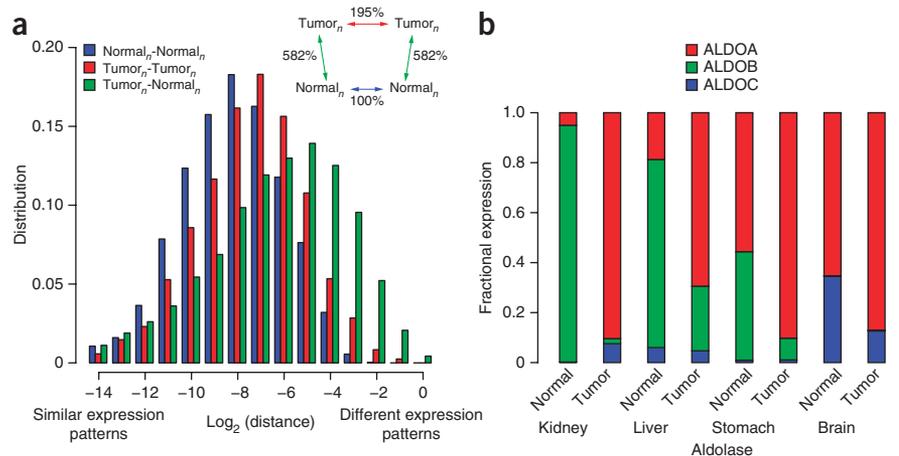
Figure 5 Concentration changes for metabolites of the TCA cycle in colon cancer. The metabolite data, obtained from ten colon cancer patients, contained matched normal and tumor samples. Every point in the figure represents the $\log_2$ ratio of tumor-to-normal concentration change for a single patient. The P-values above double arrows (in black) indicate the Wilcoxon signed-rank test significance of changes between consecutive metabolites. The P-values below metabolite names at the top of the figure (in colors) indicate the Wilcoxon signed-rank test significance of changes between matched normal and tumor samples. The horizontal dashed lines indicate medians of concentration changes for each metabolite. The inset shows the measured metabolites in the context of the TCA cycle.
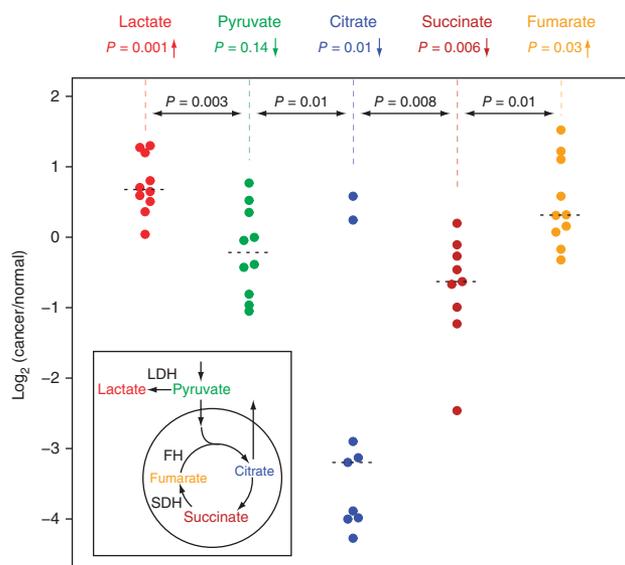
the isoenzyme among all isoenzymes catalyzing the same reaction is significantly upregulated. After correcting for multiple hypothesis testing, 919 isoenzymes were relatively upregulated in at least one tumor type, and 322 were upregulated in >25% of the 22 tumor types considered in our analysis (**Supplementary Table 13**).

Expression of enzymes with recurrent mutations in tumors

We next investigated expression changes for metabolic genes with known tumor-associated mutations. Recent sequencing studies have identified recurrent mutations in several genes associated with the TCA cycle[35,36]. Heterozygous somatic mutations in two isoenzymes of isocitrate dehydrogenase (cytoplasmic IDH1 and mitochondrial IDH2) are frequently detected in gliomas and acute myeloid leukemia (AML). These gain-of-function mutations affect the IDH active site and make it possible for the mutated enzymes to catalyze the conversion of α-ketoglutarate to 2-hydroxyglutarate, which has been proposed to promote cancer development[37]. Our analysis reveals that IDH1 and IDH2 isoenzymes are frequently upregulated in cancers (**Fig. 3b**), but the expression of the other isocitrate dehydrogenase isoenzyme IDH3 (not commonly mutated in tumors) is not significantly perturbed. This focused analysis of IDH expression across cancers (**Supplementary Table 14**) demonstrated that the upregulation of IDH1 in the three brain cancers and AML is among the six strongest of all considered tumor types. Recent sequencing efforts[38] also demonstrated the presence of similar IDH active site mutations in peripheral T-cell lymphoma, another tumor in our study with significant upregulation of IDH expression (**Supplementary Table 14**).

Germline and somatic loss-of-function mutations in fumarate hydratase (FH) and three subunits of succinate dehydrogenase (SDHB, SDHC, SDHD) are also observed in several tumors including renal cell carcinoma[36,39]. These deleterious mutations lead to the accumulation of the metabolites fumarate and succinate that regulate HIF protein levels and chromatin state to influence tumor growth[40,41]. We found that the SDH subunits and FH are strongly downregulated specifically in renal cell carcinoma (**Supplementary Tables 15 and 16**). The only cancer in our analysis with a more significant downregulation is colorectal cancer, in which decreased expression of SDH was reported previously[42]. Although somatic mutations in SDH and FH have not been observed in colorectal cancer[43,44], the significant decrease in their expression, similar to deleterious mutations in other tumors, is likely to cause mitochondrial efflux of the tumor-promoting TCA cycle intermediates and contribute to tumorigenesis.

Validation of TCA cycle metabolite changes in colon cancer

To confirm our computational prediction about TCA cycle intermediates in colon cancer, we measured and analyzed concentrations of specific metabolites from ten colon cancer patients (**Fig. 5** and **Supplementary Table 17**). The metabolite levels were obtained using gas chromatography/ mass spectrometry (GC/MS) or liquid chromatography (LC)/MS and contained matched tumor and normal samples from each patient.

Consistent with significant downregulation of oxidative phosphorylation pathway genes (Wilcoxon signed-rank test, $P = 10^{-9}$, **Supplementary Table 7**) and downregulation of the pyruvate dehydrogenase complex ($P = 0.02$) that controls the majority of glucose carbon flux into the TCA cycle, there is a significant decrease in the citrate concentration (Wilcoxon signed-rank test, $P = 0.01$) in tumor samples and a concomitant increase in the lactate concentration ($P = 0.001$). Despite a large decrease in the citrate concentration (average decrease ~65%, median ~90%), the average concentration of a downstream metabolite succinate is only 33% lower than normal and the average concentration of fumarate is >50% higher than in normal samples ($P = 0.03$). This pattern of concentration changes is consistent with the significant downregulation of the FH and SDH enzymes we observed in colon cancer expression profiles. Such a downregulation should lead to a significant increase, relative to the available citrate, in the concentration of their substrates fumarate and succinate. Notably, it was previously demonstrated[40] that even a small increase in fumarate concentration is enough to stabilize HIF1A by inhibition of the α-ketoglutarate-dioxygenases regulating its degradation[41]. The average increase in fumarate concentration (~50%) was about half of the amount observed previously for bi-allelic deletions of the FH enzyme (~90%)[40]. For four patients, fumarate concentration was >50% higher in tumor samples than in matched normal samples, and for three patients it was >100% higher. Consequently, the expression changes we observed should mimic the effects of cancer-associated heterozygous FH mutations in a substantial fraction of colon cancer patients.

DISCUSSION

It has now been recognized that reprogramming of cellular metabolism is essential for tumorigenesis[2]. An overarching conclusion of our study is that cancer-induced changes in the expression of metabolic genes are very heterogeneous across different tumor types, that is, there is no uniform metabolic transformation associated with all tumors. We observe heterogeneous behavior at all levels of biochemical organization, from global expression patterns to metabolic pathways to individual reactions and corresponding isoenzymes. The heterogeneous behavior of cancer metabolism is reminiscent of the high variability between tumors in terms of genetic and expression changes in signaling and regulatory pathways[3].

Despite the heterogeneity between cancers, the metabolic expression changes associated with individual tumors are not random. On the contrary, many of the observed changes are reproducible in independent samples of the same tumors. We suggest three principles unifying the observed tumor-induced expression perturbations. (i) Tumors often retain a substantial imprint of the metabolic expression patterns present in the corresponding native tissues. This may be a consequence of similar local environments or indicate a relative rigidity of the metabolic expression program established in the original tissue. Such behavior is conceptually similar to the minimization of metabolic adjustments principle observed in microbial metabolism following genetic perturbations[45]. (ii) A large fraction of the variance in the expression of major biochemical processes can be rationalized in terms of several principal components, representing important expression modes for key metabolic processes. Although, in agreement with physiological studies[14,46], we do not observe universal up- or downregulation of genes associated with oxidative phosphorylation, the collective expression changes along the second and third principal components suggest that fast-growing tumor cells increasingly rely on glucose fermentation. (iii) We find that hundreds of isoenzymes show significant and tumor-specific expression changes. A substantial fraction of these changes are likely to be functionally important and, at least in some cases, mimic (as in the case of SDH and FH) or possibly enhance (as in the case of IDH) the effects of recurrent tumor-promoting genetic mutations.

Beyond contributing to the understanding of tumor-induced expression changes, we believe that our analysis has important implications for the development of anticancer therapeutics. Functionally important isoenzymes with cancer-specific expression changes can potentially serve as drug targets. The possibility of targeting specific isoenzymes, such as glutaminase 1 (GLS-1; ref. 8) and pyruvate kinase M2 (PKM2; ref. 7), has already been demonstrated, but our analysis suggests that many other potential targets may be pursued in a similar way. Owing to the tumor-specific nature of the observed expression patterns, such targeting will require a focused analysis and understanding of essential metabolic transformations in each specific cancer type.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS
J.H. performed computational research and data analysis, interpreted the results and wrote the manuscript. J.W.L. and M.G.V.H. interpreted the results and edited the manuscript. L.C.C. contributed to the interpretation of the results. J.O. and K.S. contributed tissue samples. J.H.B. designed and supervised experimental research and data analysis. D.V. designed the study, supervised the project, interpreted the results and wrote the manuscript. All authors read and approved the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Warburg, O., Posener, K. & Negelein, E. On the metabolism of carcinoma cells. *Biochem. Z.* **152**, 309–344 (1924).
2. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
3. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799 (2004).
4. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
5. DeBerardinis, R.J., Sayed, N., Ditsworth, D. & Thompson, C.B. Brick by brick: metabolism and tumor cell growth. *Curr. Opin. Genet. Dev.* **18**, 54–61 (2008).
6. Hsu, P.P. & Sabatini, D.M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707 (2008).
7. Anastasiou, D. *et al.* Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nat. Chem. Biol.* **8**, 839–847 (2012).
8. Le, A. *et al.* Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. *Cell Metab.* **15**, 110–121 (2012).
9. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.* **39**, D1005–D1010 (2011).
10. Parkinson, H. *et al.* ArrayExpress update–from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.* **37**, D868–D872 (2009).
11. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).
12. Glazko, G. & Mushegian, A. Measuring gene expression divergence: the distance to keep. *Biol. Direct* **5**, 51 (2010).
13. Romero, P. *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **6**, R2 (2005).
14. Koppenol, W.H., Bounds, P.L. & Dang, C.V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* **11**, 325–337 (2011).
15. Smith, C.A., Moss, J.E., Gough, A.C., Spurr, N.K. & Wolf, C.R. Molecular genetic analysis of the cytochrome P450-debrisoquine hydroxylase locus and association with cancer susceptibility. *Environ. Health Perspect.* **98**, 107–112 (1992).
16. Khedhaier, A. *et al.* Implication of xenobiotic metabolizing enzyme gene (CYP2E1, CYP2C19, CYP2D6, mEH and NAT2) polymorphisms in breast carcinoma. *BMC Cancer* **8**, 109 (2008).
17. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
18. Faith, J.J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
19. Gillies, R.J., Robey, I. & Gatenby, R.A. Causes and consequences of increased glucose metabolism of cancers. *J. Nucl. Med.* **49** (suppl. 2), 24S–42S (2008).
20. Jolliffe, I.T. *Principal Component Analysis*, 2nd edn. (Springer, New York, 2002).
21. Duarte, N.C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
22. Lehninger, A.L., Nelson, D.L. & Cox, M.M. *Lehninger Principles of Biochemistry*, 5th edn. (W.H. Freeman, New York, 2008).
23. Diaz-Ruiz, R., Uribe-Carvajal, S., Devin, A. & Rigoulet, M. Tumor cell energy metabolism and its common features with yeast metabolism. *Biochim. Biophys. Acta* **1796**, 252–265 (2009).
24. Christofk, H.R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
25. Walenta, S., Schroeder, T. & Mueller-Klieser, W. Lactate in solid malignant tumors: potential basis of a metabolic classification in clinical oncology. *Curr. Med. Chem.* **11**, 2195–2204 (2004).
26. Estrela, J.M., Ortega, A. & Obrador, E. Glutathione in cancer biology and therapy. *Crit. Rev. Clin. Lab. Sci.* **43**, 143–181 (2006).
27. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–765 (2009).
28. Cheng, T. *et al.* Pyruvate carboxylase is required for glutamine-independent growth of tumor cells. *Proc. Natl. Acad. Sci. USA* **108**, 8674–8679 (2011).
29. Asaka, M. *et al.* Alteration of aldolase isozymes in serum and tissues of patients with cancer and other diseases. *J. Clin. Lab. Anal.* **8**, 144–148 (1994).
30. Kusakabe, T., Motoki, K. & Hori, K. Human aldolase C: characterization of the recombinant enzyme expressed in *Escherichia coli*. *J. Biochem.* **115**, 1172–1177 (1994).
31. Swinnen, J.V., Brusselmans, K. & Verhoeven, G. Increased lipogenesis in cancer cells: new players, novel targets. *Curr. Opin. Clin. Nutr. Metab. Care* **9**, 358–365 (2006).
32. Costello, L.C. & Franklin, R.B. Testosterone and prolactin regulation of metabolic genes and citrate metabolism of prostate epithelial cells. *Horm. Metab. Res.* **34**, 417–424 (2002).
33. Evans, C.T., Scragg, A.H. & Ratledge, C. A comparative study of citrate efflux from mitochondria of oleaginous and non-oleaginous yeasts. *Eur. J. Biochem.* **130**, 195–204 (1983).
34. Metallo, C.M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2012).

35. Dang, L., Jin, S. & Su, S.M. IDH mutations in glioma and acute myeloid leukemia. *Trends Mol. Med.* **16**, 387–397 (2010).
36. Frezza, C., Pollard, P.J. & Gottlieb, E. Inborn and acquired metabolic defects in cancer. *J. Mol. Med.* **89**, 213–220 (2011).
37. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
38. Zou, Y. *et al.* IDH1 and IDH2 mutations are frequent in Chinese patients with acute myeloid leukemia but rare in other types of hematological disorders. *Biochem. Biophys. Res. Commun.* **402**, 378–383 (2010).
39. Isaacs, J.S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143–153 (2005).
40. Koivunen, P. *et al.* Inhibition of hypoxia-inducible factor (HIF) hydroxylases by citric acid cycle intermediates: possible links between cell metabolism and stabilization of HIF. *J. Biol. Chem.* **282**, 4524–4532 (2007).
41. Hewitson, K.S. *et al.* Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *J. Biol. Chem.* **282**, 3293–3301 (2007).
42. Habano, W. *et al.* Reduced expression and loss of heterozygosity of the SDHD gene in colorectal and gastric cancer. *Oncol. Rep.* **10**, 1375–1380 (2003).
43. Bass, A.J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A–TCF7L2 fusion. *Nat. Genet.* **43**, 964–968 (2011).
44. Wood, L.D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
45. Segre, D., Vitkup, D. & Church, G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* **99**, 15112–15117 (2002).
46. Zu, X.L. & Guppy, M. Cancer metabolism: facts, fantasy, and fiction. *Biochem. Biophys. Res. Commun.* **313**, 459–465 (2004).

ONLINE METHODS

Microarray expression data sets. Published gene expression data sets were assembled from the GEO[9] and ArrayExpress[10] databases (**Supplementary Table 1**). Unless specified otherwise, we analyzed only expression data obtained using the most comprehensive human expression array platform (HG U133 Plus 2.0; **Supplementary Table 2**). For calculations involving global network properties and comparisons of expression data between different studies (**Fig. 1**), samples from all data sets were processed together. For all other calculations, tumor and normal samples from the same study were processed together. The affyQCReport package from Bioconductor was used to search for poor quality chips. For GeneChip arrays that passed quality control (QC) checks, we used the GCRMA algorithm[47] from Bioconductor to perform quantile normalization and extract gene expression values on the log$_2$ scale.

Calculation of differential expression (DE) for metabolic genes. Separately for each data set, the Bioconductor method limma[48], which is based on a modified t-statistic, was used to analyze differences between tumor samples and corresponding normal samples. Using the method we calculated the differential expression for each metabolic gene on the log$_2$ scale. The differential expression P-values were adjusted for multiple hypothesis testing using Benjamini and Hochberg's method[49], controlling False Discovery Rate (FDR) at 5%.

Calculation of the global divergence between a pair of expression profiles. Two different measures of divergence between a pair of expression profiles were used in our study: (i) The Euclidean distance (root mean squared deviation),

$$r.m.s.d. = \sqrt{\sum_{i=1}^{n} [\text{Average}(\log_2 x_i) - \text{Average}(\log_2 y_i)]^2 / n}$$

where x_i and y_i are the expression of gene i over two expression profiles with p and q samples $(x^1, x^2,…,x^p)$, $(y^1, y^2,…,y^q)$, $n = 1421$ is the number of genes assigned to at least one metabolic pathway in the KEGG database. (ii) The correlation-based distance $d_{cor} = 1-r$ (Average $(\log_2 x)$, Average $(\log_2 y)$), where r is the Spearman's rank correlation coefficient between average log$_2$ expression values of corresponding genes in the two expression profiles.

When comparing data sets across different studies it is important to consider batch effects arising from variations in laboratory conditions and measurements. To explore and address batch effects, we collected a set of microarray expression data for the same tissues/tumor types from multiple independent studies (**Supplementary Table 4**). All samples in **Supplementary Table 4** were processed and normalized together. To estimate the influence of the batch effects, we calculated d_1, the average expression distance between tumors (Tumor$_n$) and corresponding normal tissues (Normal$_n$) measured in different studies, and d_2, the average expression distance between tumors (Tumor$_n$) and corresponding normal tissues (Normal$_n$) measured in the same studies. The difference (d_1-d_2) represents the average batch effect due to comparisons across different studies. To account for the batch effect the difference (d_1-d_2) was subtracted from all expression distances calculated between different studies.

Identification of metabolic pathways with significant expression changes. We used two different approaches to identify metabolic pathways with significant expression changes. The two approaches resulted in very similar results. In the first approach, which was used for all calculations presented in the paper, for each gene a, we calculated its expression change in tumor sample i relative to the corresponding normal samples, $\Delta E_a^i = \log_2 x_a^i - \text{Average}(\log_2 y_a)$, where x_a^i is the expression in tumor sample i, and y_a is the expression in the s corresponding normal samples $(y^1, y^2,…,y^s)$. Wilcoxon signed-rank test of ΔE for all genes within a metabolic pathway was then used to determine the significance of up- or downregulation of the pathway in that tumor sample. In the second approach, for each gene a, we calculated the z-score of its expression in tumor sample i relative to the distribution of its expression values in the s corresponding normal samples, $z_a^i = (\log_2 x_a^i - \text{Average}(\log_2 y_a))/\sigma(\log_2 y_a)$, where $\sigma(\log_2 y_a)$ is the s.d. Wilcoxon signed-rank test of z was then used to

determine the significance of up- or downregulation of each pathway in that tumor sample. The pathway expression heterogeneity based on the second approach is shown in **Supplementary Fig. 4c**.

Statistical significance of the observed metabolic pathway expression patterns. We used randomized expression data to assess the statistical significance of the reported pathway expression patterns (**Fig. 2**). To generate the null distributions for the $(\bar{n}+\bar{m})$ and $(|\bar{n}-\bar{m}|)$ values we used the real expression data and randomly permuted metabolic gene labels while preserving the pathway sizes. We then calculated the null distributions using the same procedure as the one applied to the real data. **Supplementary Figure 3** shows the null distributions for the 10 top-regulated pathways (pathways with highest $(\bar{n}+\bar{m})$ values in **Fig. 2**) based on 1,000 random permutations of the expression data.

Pathway expression heterogeneity across tumor samples of the same tumor type. To investigate the pathway expression heterogeneity across tumor samples of the same tumor type, we introduced a pathway-specific heterogeneity metric $H=|n-m|/(n+m)$, where n is the fraction of tumor samples of a certain tumor type in which the pathway is significantly upregulated, and m is the fraction of samples in which the pathway is significantly downregulated. According to this definition, for high H values the pathway shows consistent expression changes across different samples of the same tumor type, that is, the pathway is mostly upregulated or mostly downregulated. On the other hand, for small H values the pathway expression is variable, that is, in some tumor samples the pathway is significantly upregulated, while in other samples of the same tumor type the pathway is significantly downregulated. The distribution of H values across 22 tumor types or 16 tumor types of different tissue-of-origin for the 10 top-regulated pathways (pathways with highest $(n+m)$ values), is shown in **Supplementary Fig. 5**. The 16 tumors types of different tissue-of-origin were obtained from 22 tumor types by considering samples of the three types of brain cancers, the two types of breast cancers and the four types of lymphomas together, respectively.

Calculation of significant relationships between metabolic pathway expression and expression of non-metabolic cancer/signaling genes. The context likelihood of relatedness (CLR) method[18] is based on mutual information between expression profiles. CLR was used to identify significant relationships between 214 non-metabolic cancer/signaling genes annotated in the KEGG database and the 87 KEGG metabolic pathways (**Supplementary Table 5**). The set of 214 non-metabolic cancer/signaling genes was assembled using the following two criteria: (i) genes either from the 14 KEGG cancer or 25 KEGG signaling pathways (**Supplementary Table 9**), and (ii) not within any of the 87 KEGG metabolic pathways. For each gene a in each tumor sample i, the expression change ΔE_a^i was calculated. And the mutual information (MI) between each non-metabolic cancer/signaling gene i and each metabolic pathway j was calculated across all tumor samples in our study: $MI_{ij} = \text{MI}(\Delta E_i, \sum_{a}^{n_j} \Delta E_a / n_j)$, where n_j is the number of genes within the pathway j. All mutual information values were computed using 10 bins of ΔE; the calculated values were not sensitive to the exact number of bins used. The CLR interaction Z-score for each gene i and pathway j pair $Z_{ij} = \sqrt{(z_i^2 + z_j^2)/2}$ was calculated using (i) the z-score (z_i) of MI_{ij} relative to the distribution of $\{MI_{i,1}, MI_{i,2},…,MI_{i,87}\}$, and (ii) the z-score (z_j) of MI_{ij} relative to the distribution of $\{MI_{1,j}, MI_{2,j}…,MI_{214,j}\}$. In **Supplementary Table 8** we show the identified significant relationships (with z-score > 2.0) for each metabolic pathway.

Principal component analysis. The nine meta-pathways used for the principal component analysis were compiled by combining genes from corresponding metabolic pathways in the KEGG and BioCyc databases. To perform the principal component analysis (PCA), we calculated the p-by-q matrix D for tumor-to-normal expression changes of the meta-pathways, where $p = 466$ (the total number of tumor samples in our study) and $q = 9$ (the number of meta-pathways). We used two different approaches to calculate D. The two approaches resulted in very similar principal components. In the first approach, the

(i, j)-element of the matrix D_{ij} is the average gene-specific expression changes in tumor sample i across n_j genes within meta-pathway j: $D_{ij} = \sum_{a=1}^{n_j} \Delta E_a^i / n_j$.

In the second approach, D_{ij} is the average gene-specific z-scores: $D_{ij} = \sum_{a=1}^{n_j} z_a^i / n_j$. Principal components were then obtained using the covariance method, that is, we first centered the columns of D by subtracting the column means, and then calculated a covariance matrix based on D. The covariance matrix was then diagonalized and the eigenvectors and eigenvalues were calculated. The results of the PCA analysis based on the first approach are shown in **Table 1** and the results based on the second approach in **Supplementary Table 10**. We also explored the influence of genes shared between meta-pathways on the PCA results. The results obtained when all overlapping genes were excluded (**Supplementary Table 11**) were very similar to the results with all meta-pathway genes.

Human metabolic annotations used for isoenzyme expression analysis. A human metabolic network[21] was used for isoenzyme expression analysis. The network contains 1,496 genes, 2,712 compartment-specific metabolites, and 3,743 internal and exchange reactions. In the analysis we used 2,307 network reactions that are associated with at least one known metabolic gene. Proteins that are responsible for catalysis of identical reactions and are not members of the same complex were considered as isoenzymes. In total, the network contains 667 metabolic reactions with at least two isoenzymes.

Calculation of distances between isoenzyme expression patterns. The Kullback-Leibler (KL) divergence was used to quantify the changes in the relative expression of isoenzymes for pairs of expression profiles. For each sample, the fractional expression of a particular isoenzyme i was first calculated $f_i = x_i / \sum_i^n x_i$ (n is the number of isoenzymes catalyzing the reaction and x_i is the expression value of the isoenzyme i). The flexmix package in R was then used to estimate the KL divergence between the discrete distributions $\{m(f_1), m(f_2),\ldots m(f_n)\}$ and $\{g(f_1), g(f_2),\ldots g(f_n)\}$, where $m(f)$ and $g(f)$ are the averages of the two expression profiles over p and q samples $(x^1, x^2,\ldots,x^p)$, $(x^1, x^2,\ldots,x^q)$.

Identification of isoenzymes preferentially expressed in specific tumors. For each considered isoenzyme we used the nonparametric Mann-Whitney U test to determine the significance of its fractional expression changes in the tumor samples relative to the normal samples. Specifically, for an isoenzyme i we calculated its fractional expression among all isoenzymes associated with the same reaction: $f_i = x_i / \sum_i^n x_i$, where n is the number of isoenzymes catalyzing the same reaction, and x_i is the expression value of the isoenzyme i. We then used the Mann-Whitney U statistic to test the hypothesis that the distribution of f_i values for tumor samples associated with a particular cancer type has significantly larger mean than the distribution of f_i values for the corresponding normal samples. All the P-values were FDR-adjusted at 5% considering the total number of tested hypothesis, 22,704 (1,032 isoenzymes times 22 cancer types). The isoenzymes passing the significance threshold ($P < 0.05$) are reported in **Supplementary Table 13**. We confirmed the isoenzyme results using independently collected expression data from the TCGA consortium[50]; for the confirmation we used four tumor types from TCGA (glioblastoma multiforme, breast invasive carcinoma, colon adenocarcinoma, ovarian serous cystadenocarcinoma). Using the same tumor type 70% of the isoenzymes in

Supplementary Table 13 showed the same upregulation behavior in TCGA as in the data set analyzed in the paper.

Statistical significance and multiple hypothesis testing. For pathway and isoenzyme calculations involving multiple hypothesis testing, all the corresponding P-values were adjusted with the BH procedure[49] (using the multtest package in R) to control the false discovery rate (FDR) at 0.05. The FDR-corrected P-values were used to analyze statistical significances, and unless specified otherwise, significance was reported for the adjusted $P < 0.05$.

Quantitative metabolite profiling of TCA cycle intermediates in colon cancer. Sample collection and metabolite extraction. Tumors and surrounding grossly normal-appearing tissues were obtained from 10 colon cancer patients after surgical treatment. The excised tissues were immediately stored at −80 °C. Samples were extracted and prepared for analysis using Metabolon's standard solvent extraction method. The extracted samples were split into equal parts for analysis on the GC/MS and LC/MS platforms.

GC/MS. The samples destined for GC/MS analysis were re-dried under vacuum desiccation for a minimum of 24 h before being derivatized under dried nitrogen using bistrimethyl-silyl-triflouroacetamide (BSTFA). The GC column was 5% phenyl and the temperature ramp is from 40 to 300 °C in a 16-min period. Samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization.

LC/MS. The LC/MS portion of the platform was based on a Waters ACQUITY UPLC and a Thermo-Finnigan LTQ mass spectrometer, which consisted of an electrospray ionization (ESI) source and linear ion-trap (LIT) mass analyzer. The sample extract was split into two aliquots, dried, then reconstituted in acidic or basic LC-compatible solvents, each of which contained 11 or more injection standards at fixed concentrations. One aliquot was analyzed using acidic positive ion optimized conditions and the other using basic negative ion optimized conditions in two independent injections using separate dedicated columns. Extracts reconstituted in acidic conditions were gradient eluted using water and methanol both containing 0.1% formic acid, while the basic extracts, which also used water/methanol, contained 6.5 mM ammonium bicarbonate. The MS analysis alternated between MS and data-dependent MS2 scans using dynamic exclusion.

Data extraction and compound identification. The data extraction of the raw mass spec data files yielded information that could be loaded into a relational database and manipulated without resorting to BLOB manipulation. Once in the database the information was examined and appropriate QC limits were imposed. Peaks were identified using Metabolon's proprietary peak integration software, and component parts were stored in a separate and specifically designed complex data structure. TCA cycle intermediates were identified by comparison to library entries of purified standards. The combination of chromatographic properties and mass spectra gave an indication of a match to the specific compound or an isobaric entity. The collected metabolite data are presented in **Supplementary Table 17**.

47. Wu, Z. & Irizarry, R.A. Preprocessing of oligonucleotide array data. *Nat. Biotechnol.* **22**, 656–658, author reply 658 (2004).
48. Smyth, G.K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, e3 (2004).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.* **57**, 289–300 (1995).
50. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719–724 (2009).