The Metabolic Landscape of Tumors

E Reznik, A Luna, BA Aksoy, C Sander, others less worthy
[1] Computational Biology Center, Sloan-Kettering Institute, New York NY
[*] E-mail: reznike@mskcc.org

Abstract

- Assembled 13 studies, 9 cancer types, 1K samples, benchmark for future meta-analyses? - Examined intrinsic variation associated with normal–¿ tumor transformation - Recurrent metabolic alterations - Clinical association with metabolites

1 Introduction

Tumors must grow and divide in the face of stress imposed by indefinite proliferation and augmented by cytotoxic and targeted therapy. To do so, cancer cells modulate the activity of metabolic pathways to supplant ATP production, produce suitable levels of biosynthetic precursors, sustain redox potential and maintain epigentic integrity. As a whole, these metabolic alterations are implemented via changes in the levels of intracellular metabolites, enzymes, and transporters.

One approach to studying the underlying biology of cancer has been to profile the genomes, epigenomes, and proteomes of large cohorts of tumors and benign tissues. Statistical identification of recurrent molecular features, across tumors of distinct tissues of origin, has elucidated the key pathways and processes driving cancer. In contrast, while the deregulation of metabolism is a well-accepted hallmark of cancer, surveys of metabolic alterations in tumors have remained limited in scope. Restricted by the availability of adequate data, signatures of recurrent metabolic alterations have only been investigated through analysis of the transcriptome. To our knowledge, no comparable metabolomic investigations have been completed.

Metabolomic studies of cancer tissues to date have been hampered by several technical and informatic challenges: incomplete coverage of the metabolome (typically fewer than 600 distinct metabolites profiled per study), inconsistent use of standard nomenclature (*i.e.* KEGG, HMDB IDs), and relative (rather than absolute) quantification of metabolite abundances. Overcoming these challenges has enabled At a basic level, we seek to understand whether the abundances of metabolites are dysregulated in a common manner across different cancer types.

In this work, we survey the metabolic landscape of tumors by integrating tumor metabolomic data from eleven studies examining nine different cancer types. We find that the extent to which tumors metabolically differ from normal tissues is highly dependent on cell-of-origin: for example, prostate tumors show comparatively small changes compared to normal prostate tissue. A detailed examination of these metabolic alterations leads to the observation that two metabolites, lactate and kynurenine, are nearly always increased in abundance in the cancer types under examination. Finally, by integrating our analysis with clinical information on cancer progression, we identify a handful of metabolites which show recurrent correlation with tumor stage and grade. Metabolomic data and analysis results are made publically available for the cancer research community

2 Assembly of a Cross-Cancer Compendium of Metabolomics Data

We obtained published cancer tissue metabolomics data from eleven datasets covering seven distinct cancer types (see Figure 1 and Supplementary Table 1). Data for all studies was collected using mass spectrometry. Three cancer types (breast, prostate, and pancreatic [**?**]) were represented by at least 2 different datasets, enabling us to evaluate the consistency of findings across different studies. In total, our dataset encompasses 935 distinct small molecule compounds across 928 tissue samples.

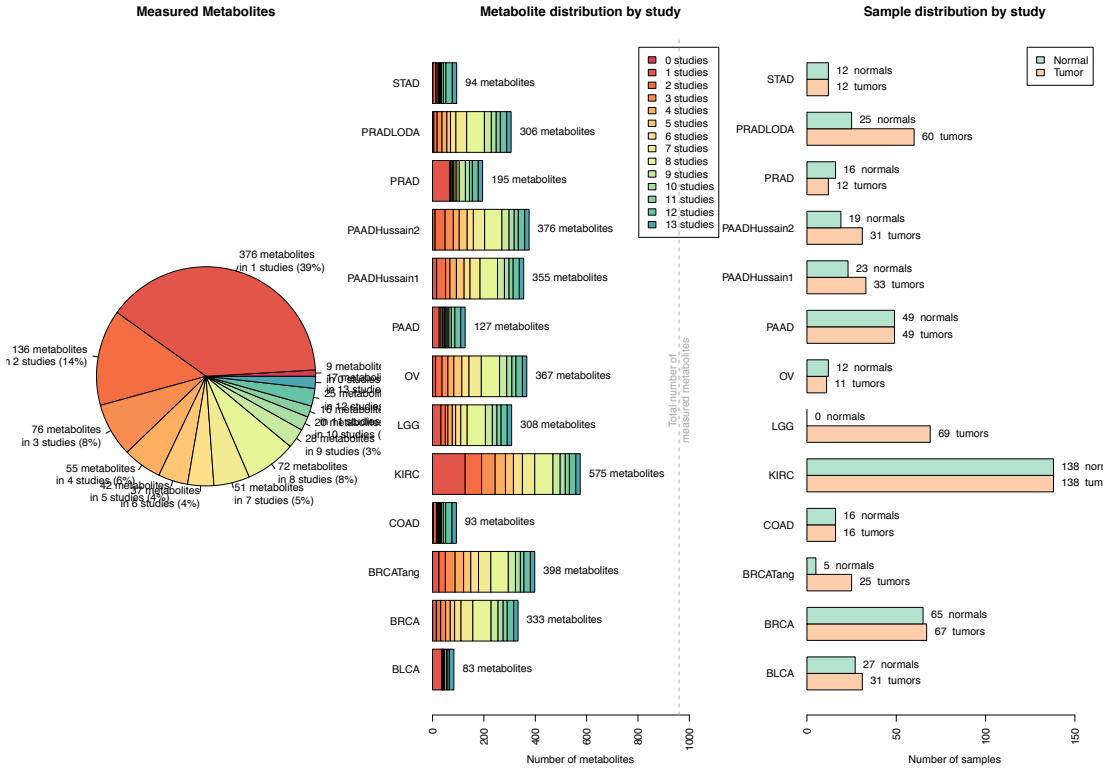

Figure 1. Metabolomics data analyzed in this study. Data from 13 distinct metabolomics studies, examining 9 cancer types, were aggregated. Due to incomplete coverage of the metabolome, many metabolites were profiled in a small proportion of studies. The number of tumor/normal samples varied from study to study, and all but one study (gliomas) contained normal samples.

To complete a meta-analysis, we implemented a data standardization pipeline (Figure 2) addressing two independent problems: quantatitative standardization of metabolomic measurements, and bioinformatic alignment of metabolites profiled across many studies (see Methods for detailed description). Data generated from mass spectrometry studies were in general reported as relative quantifications of metabolites (*i.e.* peak intensities), rather than absolute measures of concentration. making direct comparison of the same metabolite across studies infeasible (see Figure 2). Furthermore, across all studies, a number of metabolites were at sufficiently low abundance so as to fall outside the sensitivity of the measuring instrument, and were often imputed. Making this data amenable to quantitative analysis was essential, because it contained useful information indicating that the concentration of a metabolite was low (compared to samples where the concentration was within the quantifiable limits). To enable a fair comparison of metabolomics data across different studies, we implemented a common data imputation and standardization pipeline.

In addition to data standardization, a bioinformatic challenge to our meta-analysis was the identification of metabolites profiled across multiple studies. Unlike other high-throughput technologies which can measure the abundance of all relevant species in a sample (*e.g.* RNA sequencing), metabolomic profiling samples only a fraction of all compounds in the metabolite. More importantly, metabolites are referred to by different synonymous names (*e.g.* lactate, lactic acid, (S)-2-Hydroxypropanoate, etc.), and are often reported alongside a variety of identifiers (*e.g.* KEGG IDs, HMDB IDs, Pubchem IDs). To address this

issue, we searched for synonymous identifiers of each metabolite using the Chemical Translation Service (CITE). These identifers were then used to assemble a meta-dataset of all metabolomics data, "aligning" metabolites sharing common identifiers. Manual inspection of the aligned dataset confirmed that our method correctly matched metabolites across different studies. All code for data standardization and alignment is provided at github (URL).

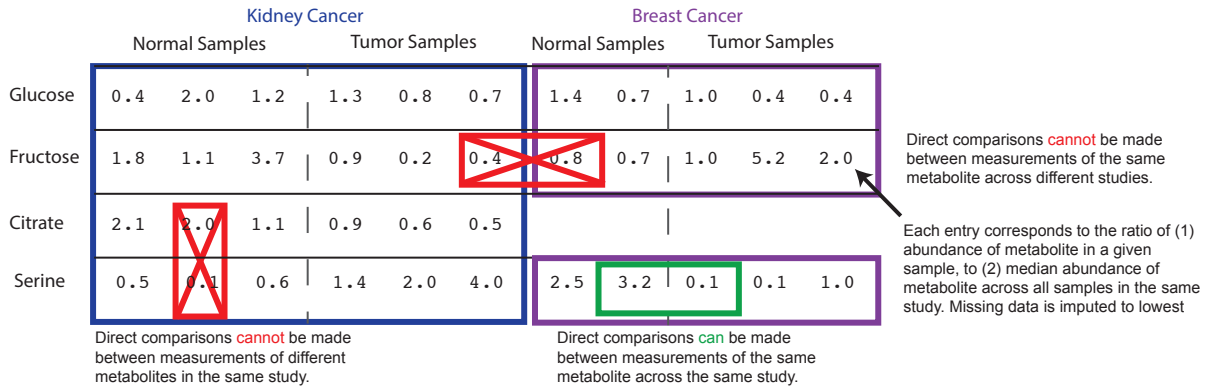

Figure 2. Metabolomics data aggregated from different studies cannot be directly compared. Typically, for each metabolite, abundance is reported relative to the median of all measurements of a metabolite within a study. Comparisons between different metabolites profiled in the same study is not directly possible (leftmost red box). Furthermore, comparisons between the same metabolite profiled across different studies is not possible (topmost red box). Throughout our analysis, we will frequently examine the change in abundance of a single metabolite across different subsets of tissue samples (*e.g.* tumor/normal samples) from the same study (green box).

3 Analysis of Metabolic Variation in Tumors and Normal Tissue

Large sequencing efforts have revealed the remarkable diversity of mechanisms which can drive the transformation of normal tissue to tumor. In some cases (*e.g.* leukemias and thyroid cancers), tumors are characterized by a handful of genetic mutations, while in others (*e.g.* melanomas), hundreds or thousands of mutations may collude to drive the tumor [**?**]. In analogy to these studies, we sought to understand whether the differences between tumor/normal tissue varied substantially across different tumor types. Therefore, on a study-by-study basis, we examined which metabolites were differentially abundant between tumor and normal samples (Mann-Whitney U test, BH-corrected p-value ¡ 0.05). We found the number of differentially abundant metabolites varied drastically from one cancer type to the next (Figure 3A). For example, in studies of breast and clear-cell kidney tumors, we found that greater than 40 percent of metabolites showed at least a 2-fold change in abundance, whereas two independent studies of prostate

cancer found that fewer than 6 percent of metabolites were differentially abundant. Importantly, because the number of tumor/normal samples were comparable for these two cancer types, the differences in differential abundance were unlikely to be statistical artifacts arising from differences in sample size.

This analysis also revealed a cancer-type-dependent trend towards unequal proportions of metabolites which increased or decreased between tumor and normal tissues. In both breast cancer studies, we found that nearly all metabolites deemed differentially abundant were at higher levels in tumor tissue, compared to normal tissue. While similar biases were evident in other studies (*.e.g.* differentially abundant metabolites in pancreatic tumors tended to be at lower concentration in tumors), the effect was particularly striking for breast cancer. Although it is not possible for us to determine with certainty the source of this effect, we speculate it could arise from the disproportionate extent of data imputation for metabolites in normal breast tissue, compared to the extent of imputation in tumor tissue (Figure S1). Because the effect is apparent in both breast cancer studies, it may be possible that metabolites are broadly at higher concentrations in breast tumor compared to benign breast tissue.

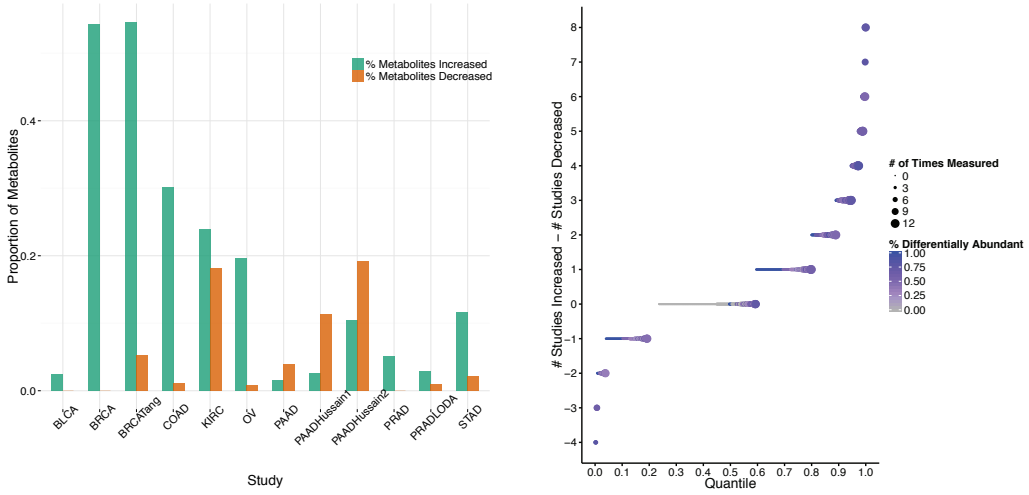

Figure 3. Tumor/Normal Comparison. (A) Comparison of metabolic alterations in tumors to natural metabolic variation in normal tissue. By comparing the similarity of (1) randomly chosen pairs of tumor/normal tissues, to (2) randomly chosen pairs of normal tissues, we assessed the magnitude of change associated with cancerous transformation. Metabolic alterations in prostate and pancreatic tumors are small when compared with natural metabolic variation in these tissues. (B) The number of differentially abundant metabolites between tumor and normal tissues across each study varies by cancer type. Breast tumors contain the most differentially abundant metabolites, while prostate tumors contain the least.

4 Common Patterns of Metabolic Alterations Across Cancers

We were particularly interested in identifying metabolites which showed consistent patterns of increased/decreased abundance across many tumor types. The existence of such metabolites could be indicative of a cancer-type-agnostic signature of transformation. By aggregating the results of our differential abundance screen, we identified 113 metabolites differentially abundant in at least 5 studies (Supplementary Table XX). Among these, a single metabolite, taurine, was differentially abundant in ten of the twelve studies in which it was measured. Two TCA cycle metabolites (fumarate and malate), and a number of amino acids, including aspartate, asparagine, arginine, methionine, and proline, were differentially abundant in 8 studies. However, for all of these metabolites, the direction of change (*i.e.* higher or lower in tumor, relative to normal) was not consistent across cancer types.

To identify metabolic alterations characteristic of transformation, we focused on metabolites which showed recurrent, consistent changes in abundance across tumor types. Two metabolites, lactate and kynurenine, were observed to increase in abundance in 8 tumor types (and decrease in abundance in zero). Three metabolites showed recurrent depletion in four distinct studies: caprate, pelargonate, and guanidoacetic acid. Both caprate and pelargonate are medium chain fatty acids, with chain lengths of ten and nine, respectively, while guanidoacetate is an intermediary metabolite of the TCA cycle. Interestingly, laurate, a medium chain fatty acid of carbon chain length 12, also showed a tendency towards recurrent downregulation (decreased in abundance in tumors in 4 studies, increased in abundance in one).

5 Pathway Analysis

To make greater sense of which metabolic processes may be recurrently up- or down-regulated in tumors, we aggregated the differential abundance results across KEGG metabolic pathways. For each study, we mapped metabolites onto KEGG pathways, and calculated an aggregate differential abundance score for each pathway (see Methods). We restricted our analysis to pathways for which at least 5 metabolites in the pathway were profiled across 6 or more studies, leaving us with XX total pathways.

We observed a tendency across cancers for increases in metabolites in sugar metabolism (including fructose and mannose metabolism, amino sugar and nucleotide sugar metabolism). Many tumor types also showed increases in glutathione metabolism, which produces the cell's primary antioxidant (reduced glutathione, GSH). Interestingly, we found that only a single pathway, fatty acid biosynthesis, showed recurrent recurrent depletion of its constituent metabolites across cancers. Interestingly, we observed that while metabolites in the TCA cycle and oxidative phosphorylation were frequently differentially abundant, their direction of change (*i.e.* higher or lower in tumors) was mixed, echoing a similar finding in a pan-cancer analysis of metabolic gene expression data [**?**].

6 Many Tumors Do Not Exhibit the Warburg Effect

The observation that many tumors accumulate high levels of lactate, compared to normal tissue, is commonly referred to as the Warburg effect. The Warburg effect is characterized by suppression of mitochondrial oxidative phosphorylation in favor of cytosolic aerobic glycolysis, which results in the excess production and excretion of lactate. Diverting glycolytic flux towards lactate can shift cellular metabolism towards a more inefficient but also more rapid rate of ATP production, when compared to oxidative phosphorylation [**?**]. The consequences of elevated aerobic glycolysis are often exploited in clinical imaging using fluorescent glucose for the identification of metastastic lesions [**?**].

To determine how prevalent aerobic glycolysis/the Warburg effect was in our data, we restricted our analysis to tumor samples for which matched normal tissue from the same patient was also available (347 pairs of tumor/normal tissue across 9 different cancer studies). Interestingly, we found that 84/347 (24%) of tumor samples contained lower levels of lactate than their matched normal tissue counterpart.

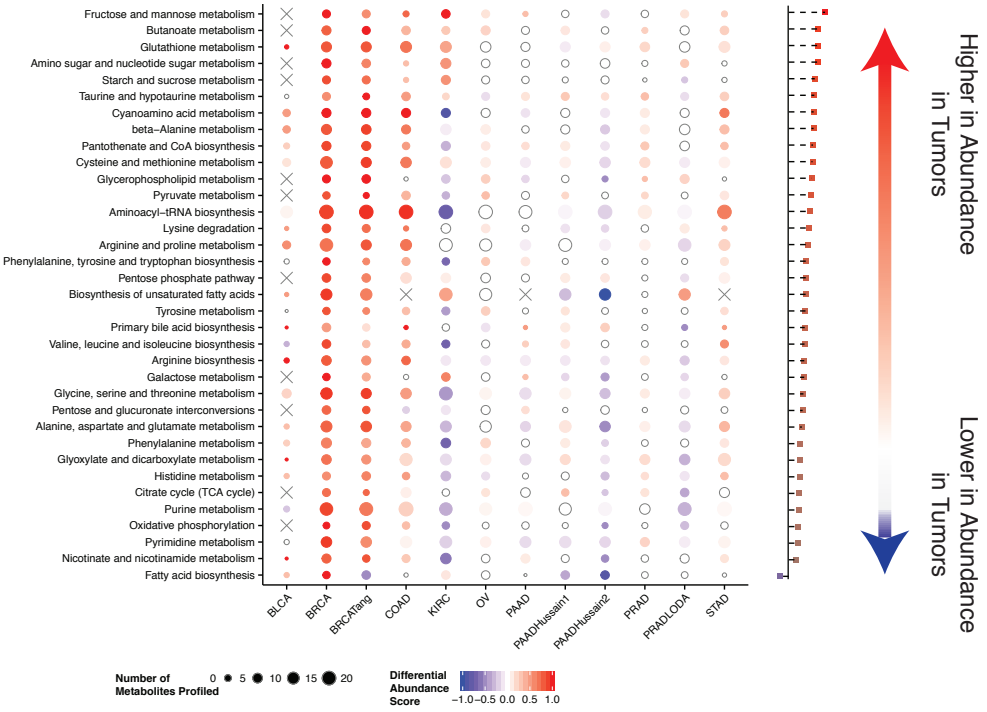

Figure 4. Recurrent metabolic alterations. (A) Heatmap of metabolites which are differentially abundant across at least 8 different cancer studies. (B) Pathway analysis Write more.

While most studies were enriched for tumors with increased levels of lactate, we again observed prostate tumors to be the outliers: the majority of prostate tumors showed reduced levels of lactate relative to paired adjacent normal tissue. We further confirmed whether the apparent depletion of lactate in prostate tumors (relative to normal tissue) did not arise from abberantly high levels of lactate in normal tissue (SI Figure S6.

Increased levels of lactate in tumor tissue are suggestive of an increase in glycolytic flux. In search of evidence to support this hypothesis, we calculated the correlation coefficient (and associated p-value) of every metabolite with levels of lactate on a study-by-study basis. Surprisingly, we observed that among glycolytic metabolites, only glucose-6-phosphate,fructose-6-phosphate, and pyruvate were positively correlated with levels of lactate in more than one study. Instead, the group of metabolites most positively correlated with lactate were urea, S-adenosylhomocysteine, and a large subset of amino acids including all three branched chain amino acids (leucine, isoleucine, valine) and two amino acids with aromatic side chains (tryptophan and phenylalanine). Taken together, these results suggest that while increased lactate levels are an indicator of increased glycolytic flux, they point further to alterations in more peripheral pathways as well.

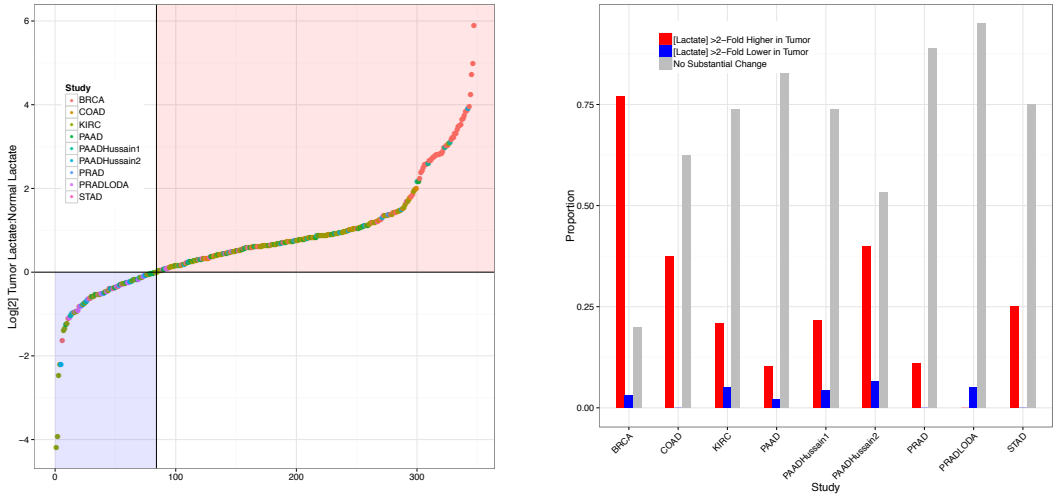

Figure 5. Many tumors do not display the Warburg effect. (A) A quarter of tumors contained less lactate than matched normal tissue from the same patient. (B)

7 Metabolic Indicators of Tumor Progression

The processes, including genetic and signaling events, which drive the progression of cancer to more aggressive stages are distinct from those which intitiate the tumor itself [**?**]. In light of this, we examined our data for metabolic signals associated with progression of tumors to higher stage and more aggressive grade. Tumor grade is a histological measure of the extent of abnormal apperance of tumor cells. In contrast, tumor stage describes the severity of a tumor based on its size, infilitration of lymph nodes, and metastatic status. Among the 12 cancer studies which we collected in our dataset, XX had associated clinical data on tumor stage or grade. We used statistical meta-analysis techniques to identify metabolites which showed consistent changes (*i.e.* consistent increase/decrease in metabolite levels with increasing tumor stage) across several cancer types. Our analyis accounted for the frequency of imputed data (see detailed description in Methods).

In total, we found 140 metabolites whose abundance were significantly correlated to tumor grade, and 60 metabolites with abundances significantly correlated to tumor stage. Filtering these results further to extract metabolites significantly associated with clinical features across many tumor types, we found 2 metabolites, erythronate and cytidine 5'-diphosphocholine, significantly associated to tumor stage in breast kidney, and ovarian cancers. We found 14 metabolites associated to tumor grade in at least 3 studies, including several amino acids (asparagine, proline, phenylalanine, leucine), 3 pyrimidines and their derivatives (thymine, uracil, and 5,6-dihydrouracil), and kynurenine.

Among the most interesting findings was the observation that kynurenine levels were increased in tumors versus normal tissues. Kynurenine is a metabolic byproduct of the degradation of trytophan by two groups of enzymes: tryptophan dioxygenases and indoleamine 2,3-dixoygenases. Binding of kynurenine to aryl hydrocarbon receptors (AHRs) and suppress the activity of T-effector cells, as well

as indirectly activating regulatory pro-tumorigenic T cells. Kynurenine was unique in our study because it was found to be elevated in the majority of tissues, and it was also found to be positively correlated to tumor grade in 4 different studies (2 different prostate cancer studies, as well as breast cancers and gliomas). Together, these findings point to a critical role for kynurenine in the metabolism of tumors.

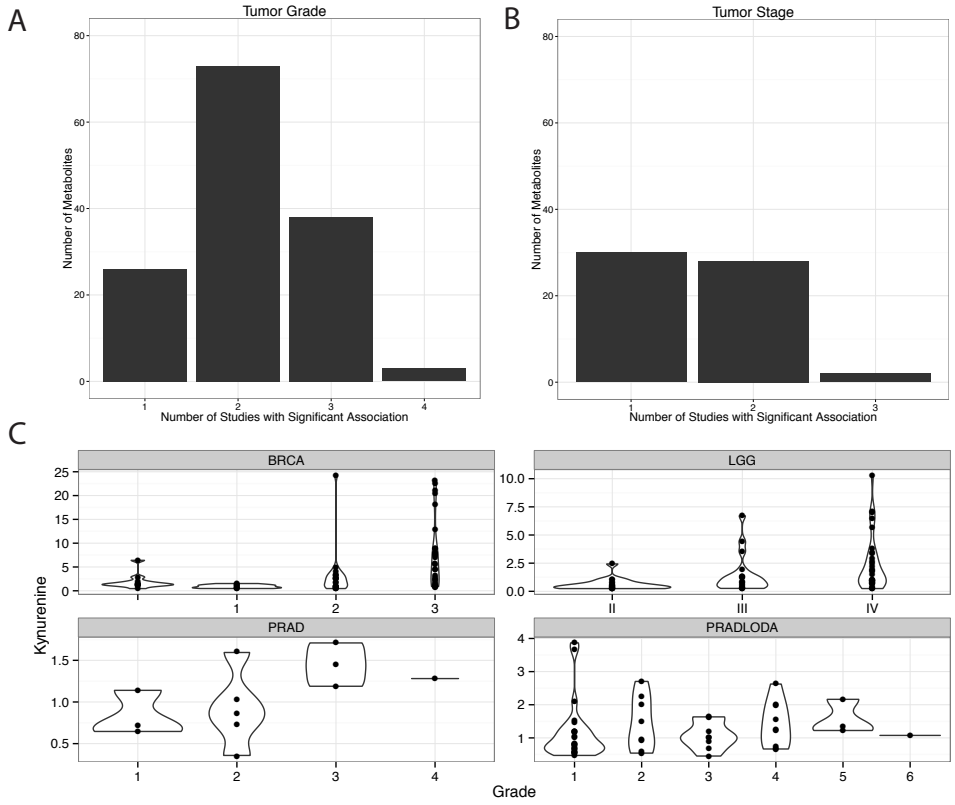

Figure 6. Metabolic correlation with tumor progression. (A) and (B) Relatively few metabolites correlate with tumor stage and grade across multiple cancer types. (C) Kynurenine is significantly associated with tumor grade across three different cancer types (breast, gliomas, and prostate cancers.

8 Discussion

Tumors must bear the metabolic challenges faced by the tissue they arise, from the frequent osmotic stress inherent to the filtration performed by kidney cells to the xenobiotic stress endured by cells of the liver. Thus, it should come as no surprise that tumors vary substantially in how their metabolic phenotype. In this study, we examined in detail the extent of this variation across nine cancer types. While we observed that some tumors (*e.g.* of the prostate) appear to show comparatively minor metabolic alterations compared to other cancer types, we also found metabolic alterations common to the majority of cancer types (*e.g.* the increase in abundance of lactate and kynurenine).

Our approach to the assembly and alignment of several metabolomics studies relied on several computational pipelines which we expect to serve as a benchmark for future work. As described in detail earlier and in the Methods, the majority of data compiled here was reported in terms of (dimensinoa-

less) relative abundances, rather than absolute concentrations. To make data across studies comparable, we implemented a computational pipeline which assured, where possible, common standards for data imputation and standardization. We also implemented a bioinformatic pipeline to "align" metabolites identified by distinct metabolite identifiers, the code for which is included as a supplementary file (SI File XX). To compare results across studies, we relied heavily on non-parametric statistics, and used changes relative to normal tissue as the fundamental unit of comparison.

8.1 Caveats

Sev - Data quality: imputation, relative abundances - Variability in sample prep

8.2 Future Directions

The results of our analysis are intended to be a starting point for future analysis integrating metabolic characterization of tumors into our understanding of tumor biology. Below we describe several promising challenges whose resolution would add to our understanding of the contribution of metabolism to oncogenesis.

Despite the growing number of studies characterizing metabolite abundances in human tissues, to our knowledge no methods exist for utilizing these data to infer what the rates at which metabolic reactions proceed, *i.e.* metabolic flux. In contrast, beginning nearly a decade ago and associated with the proliferation of gene expression studies, many methods have been developed to integrate transcript abundances with genome-scale metabolic models to predict metabolic flux. While several algorithms have been proposed to jointly integrate metabolomics and transcriptomics data into GSMMs [**?**, **?**], it is clear that the transition to a model fully parametrized exclusively with metabolomics data is a major challenge. Most of the methods proposed to date asssume *a priori* that the time-averaged change of metabolites either does not change, or is quantified directly. In contrast, metabolomics data on surgically resected tumors is limited to single time-point measurements. Furthermore, much of the effort invested in integrating metabolomics data into GSMMs has focused on prokaryotic organisms which lack intracellular compartments. Deconvolving the relative contributions of such compartments (*e.g.* the mitochondria and the cytosol) to metabolite levels will likely be critical to making accurate flux predictions.

While our analysis in this paper has focused on non-parametric, univariate analysis of changes in metabolite levels between tumor and normal samples, other analyses which respect the limitations of the data (Figure 2) are possible. One example, which we have not explored, is a correlation analysis, akin to those proposed in several publications using gene expression and metabolomics data [**?**, **?**]. It is possible that, by examination of changes in covariation patterns between metabolites between different subsets of samples (*e.g.* tumor/normal, high stage/low stage tumors), the identification of putatively "rewired" metabolic pathways may be possible. By coupling such correlation analyses with GSMMs and the possible ways flux may be routed through them (so-called elementary flux modes [**?**]), progress may be made towards flux prediction.

Finally, given the proliferation of clinical genomic sequencing of cancer samples, it seems obvious to propose that connections be drawn with molecular subtypes and the metabolic landscape of tumors. Several prominent examples now exist of molecular alterations which manifest with, among other things, distinct metabolic phenotypes: *KRAS* mutations in pancreatic adenocarcinomas induces macropinocytotic scavenging of extracellular nutrients, and hotspot *IDH1* and *IDH2* mutations produce 2-hydroxyglutarate, which induces DNA hypermethylation via inhibition of α-ketoglutarate-dependent DNA demethylases. It is likely that other effects, perhaps more subtle but nevertheless impacting tumor viability, remain to be uncovered.

9 Methods

9.1 Data Imputation and Standardization

For data which was already imputed and standardized (including the BLCA, KIRC, BRCA, BRCATang, OV, PAADHussain1, PAADHussain2, PRADLODA studies), we used the data as reported by the original authors. For five studies (COAD,LGG,PAAD,PRAD,and STAD) for which no imputation was completed, we applied the following imputation and standardization procedure. For each metabolite, imputed values were set equal to the minimum measured abundance of that metabolite. Then, all measurements of the metabolite were median normalized.

9.2 Metabolite ID Mapping

Augustin please fill in

9.3 Differential Abundance Tests

Differential abundance was calculated using the ratio of the average abundance of a metabolite in tumor tissue, to the average abundance of a metabolite in normal tissue. Statistical significance was assessed using non-paramatric Mann-Whitney U-tests.

9.4 Pathway Differential Abundance Score

The differential abundance score for a pathway was defined as

$$DA = \frac{I - D}{S}$$

where I is the number of measured metabolites in a pathway which increased in abundance relative to normal tissue, D is the number which decreased, and S is the total number of measured metabolites. A DA score of 1 indicates that all metabolites increased in abundance, whereas a score of -1 indicates that all metabolites decreased in abundance, relative to normal tissue.

9.5 Metabolomics Data Acquisition and Normalization

Metabolomics data from prior, published work was obtained either through the corresponding journal, or by contacting the corresponding author. The data for all studies except COAD and STAD was reported in relative abundances, *i.e.* the abundance of a metabolite i in sample j could only be compared to other values of metabolite i in different samples from the same study. For COAD and STAD, absolute abundances were reported.

Because metabolomics data does not necessarily obey a known distribution, we applied minimal normalization techniques in order to make data minimally comparable across all studies. For each metabolite in a study, we calculated the median abundance, and normalized by this abundance. Given the unknown distribution of the data in our study, we used only non-parametric statistical tests (which are independent of data distribution) to assess changes in metabolite abundance.

9.6 Correlation with Lactate Levels

Within each study which profiled lactate abundance, spearman correlations and associated p-values were calculated between the levels of all profiled metabolites and lactate. P-values across all studies were combined using Fisher's method and a chi-squared test was applied to determine statistical significance [**?**]. P-values from the chi-squared test were corrected using the Benjamini-Hochberg procedure.

9.7 Association with Stage and Grade

Different statistical tests were applied to identify correlation between metabolites and clinical features depending on the level of censoring in the data. If there is less than 20% censoring within the tested set, the Jonckheere-Terpstra test (a non-parametric test which uses permutations to calculate the p-value) was used to examine ordered differences among classes . If there are more than 20% but less than 80% censored data, an exact log-rank trend test is used for interval (left) censored data. Both tests assume a metabolite should increase/decrease monotonically with stage/grade. If there is more than 80% censoring no test is used, and the metabolite is ignored. To summarize the tests across cohorts I use Fishers method to combine p-values. Note that resulting combined p-value is parametric.

Since we are interested in finding associations of a common sign (*e.g.* consistently positive correlation across multiple studies), one-sided p-values are calculated first, aggregate into a combined p-value using Fisher's method [**?**], and then transformed to two-sided combined p-value.

Study Name	**Cancer Type**	Data Type	**Reference**
BLCA	Bladder	Mass Spectrometry	[**?**]
BRCA	Breast	Mass Spectrometry	[**?**]
BRCATang	Breast	Mass Spectrometry	[**?**]
COAD	Colorectal	NMR	[**?**]
KIRC	Clear-cell Renal Cell	Mass Spectrometry	[**?**]
LGG	Glioma	Mass Spectrometry	[**?**]
OV	Ovarian	Mass Spectrometry	[**?**]
PAAD	Pancreas	Mass Spectrometry	[**?**]
PAADHussain1	Pancreas	Mass Spectrometry	[**?**]
PAADHussain2	Pancreas	Mass Spectrometry	[**?**]
PRAD	Prostate	Mass Spectrometry	[**?**]
PRADLODA	Prostate	Mass Spectrometry	[**?**]
STAD	Stomach	NMR	[**?**]

Table S1. Summary of the published cancer tissue metabolomics studies examined in this work.

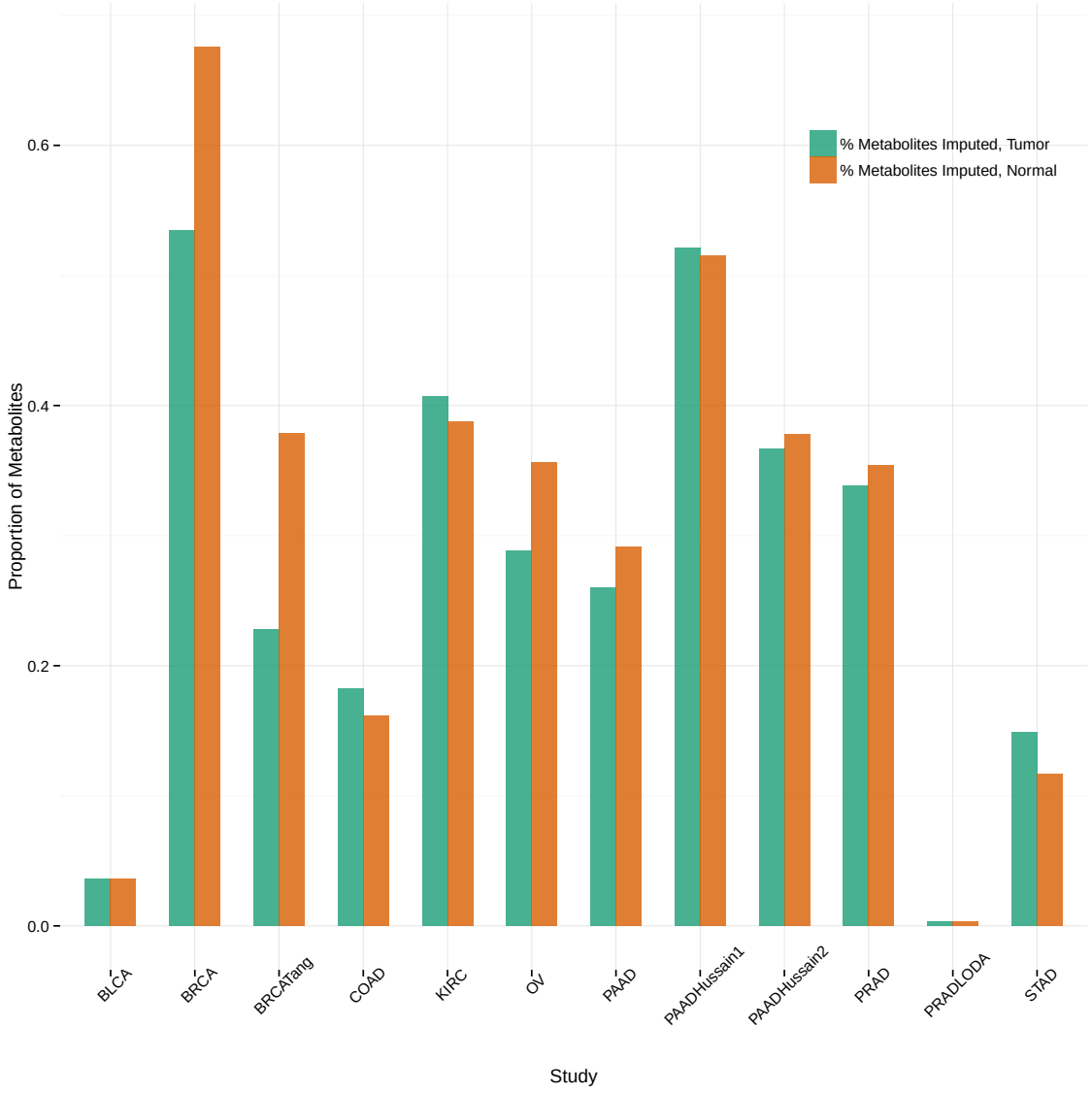

Figure S1. For each study, we calculated the proportion of metabolites with greater than 20% imputation in tumor samples. The calculation was repeated separately for normal samples. The two breast studies are exceptional because of the incongruent level of imputation between normal and tumor samples: far more metabolites are heavily imputed in normal breast samples, compared breast tumor samples.

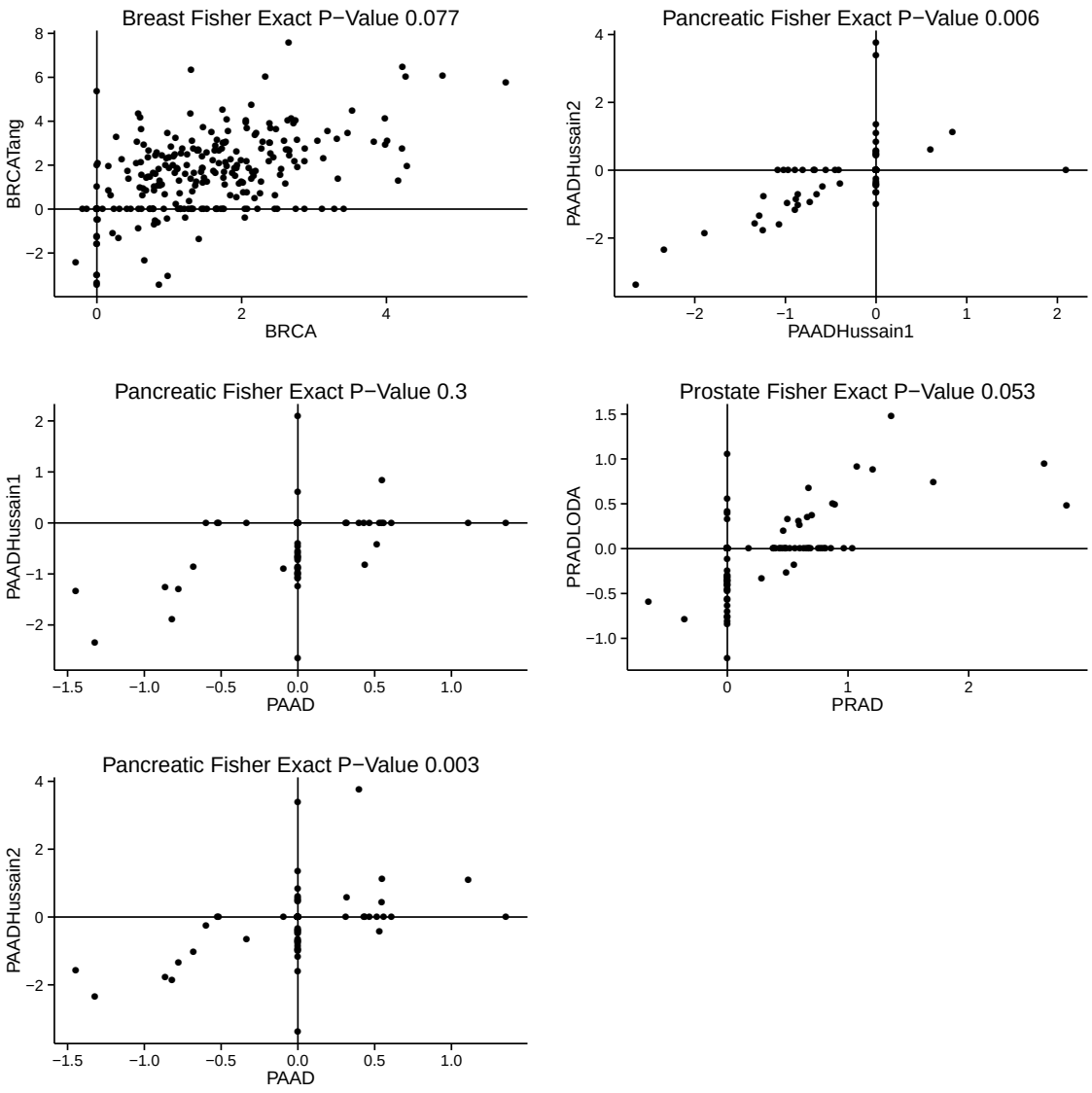

Figure S2. Insert caption.

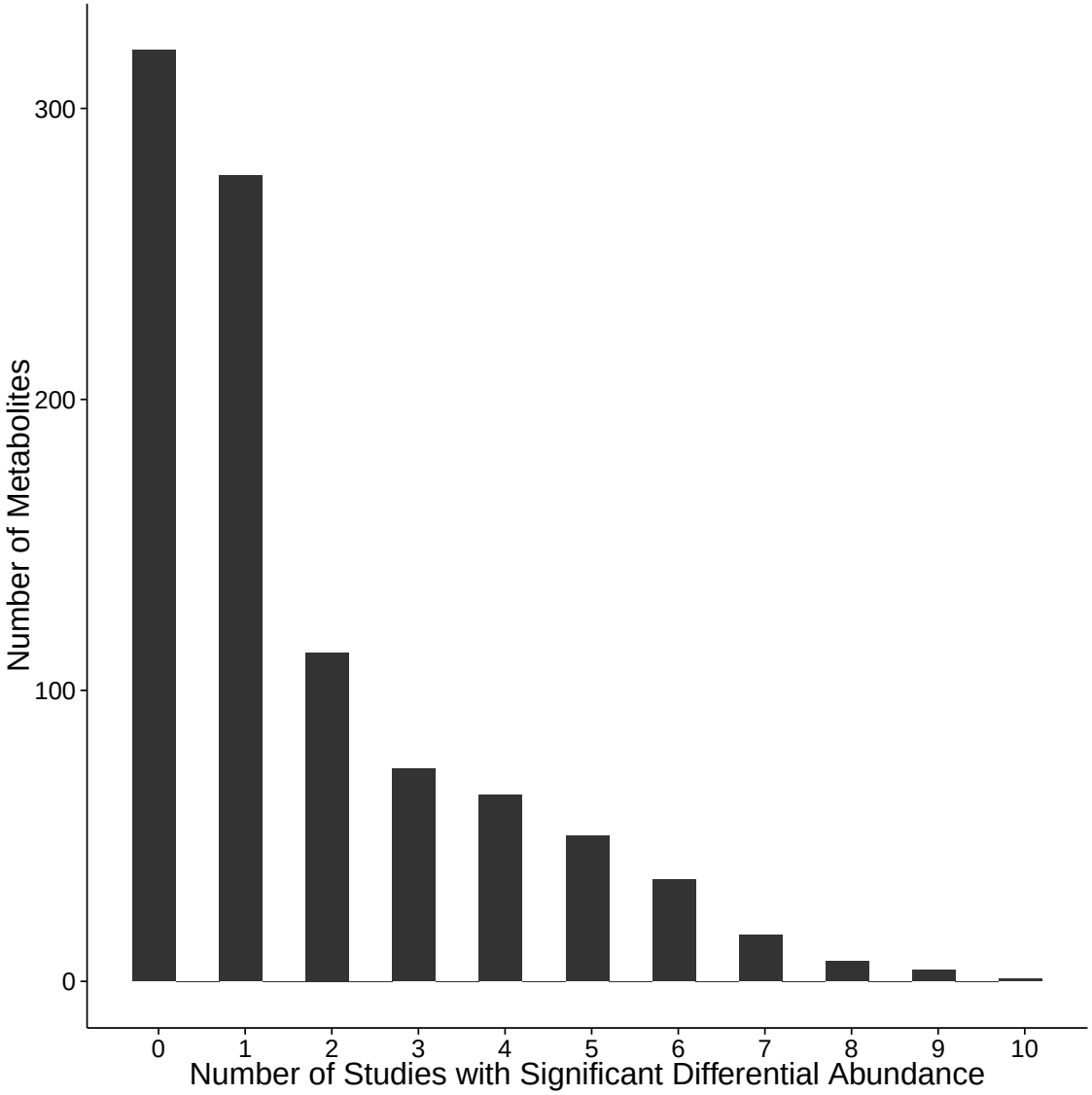

Figure S3. Insert caption.

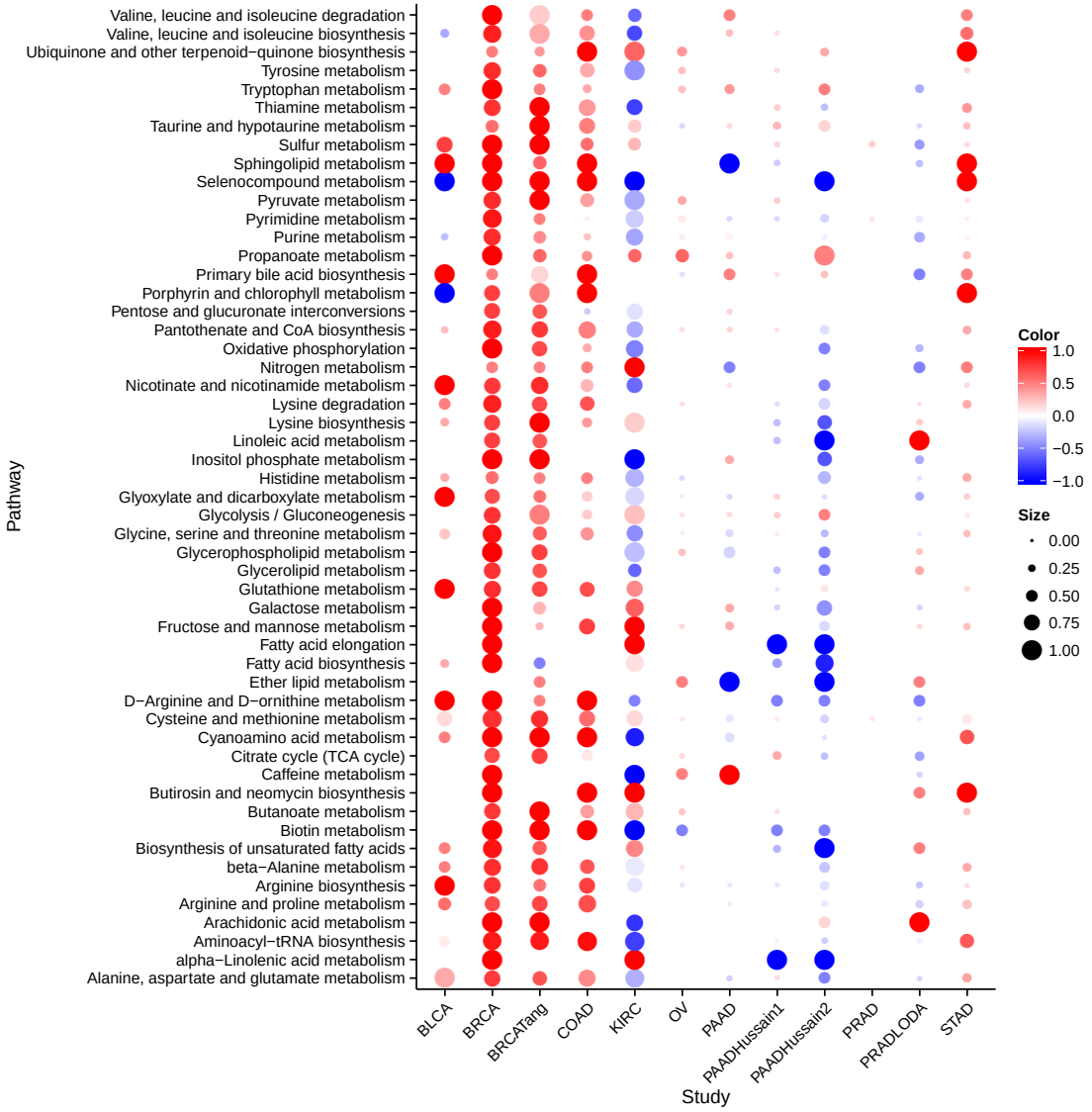

Figure S4. Insert caption.

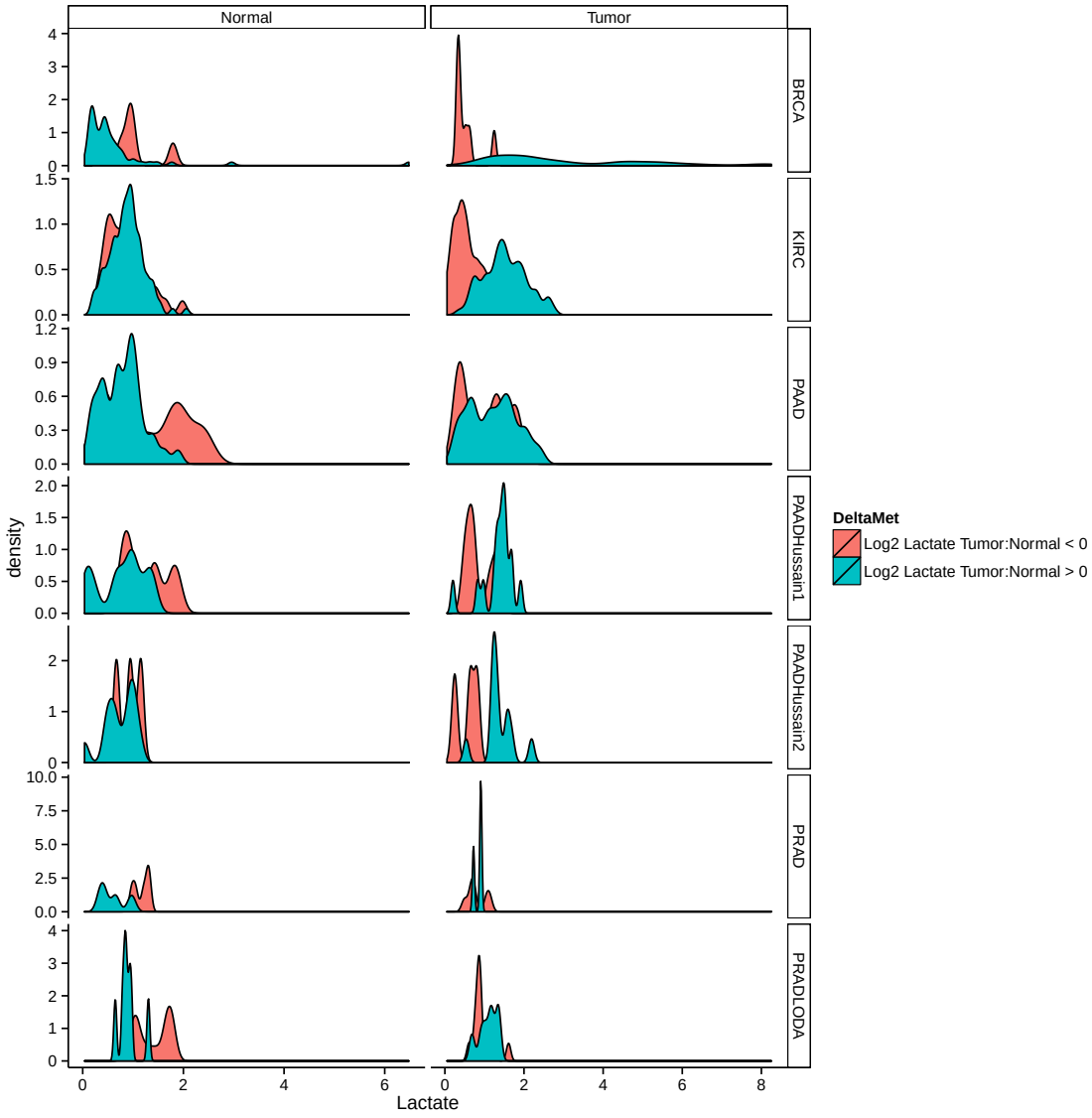

Figure S5. Insert caption.

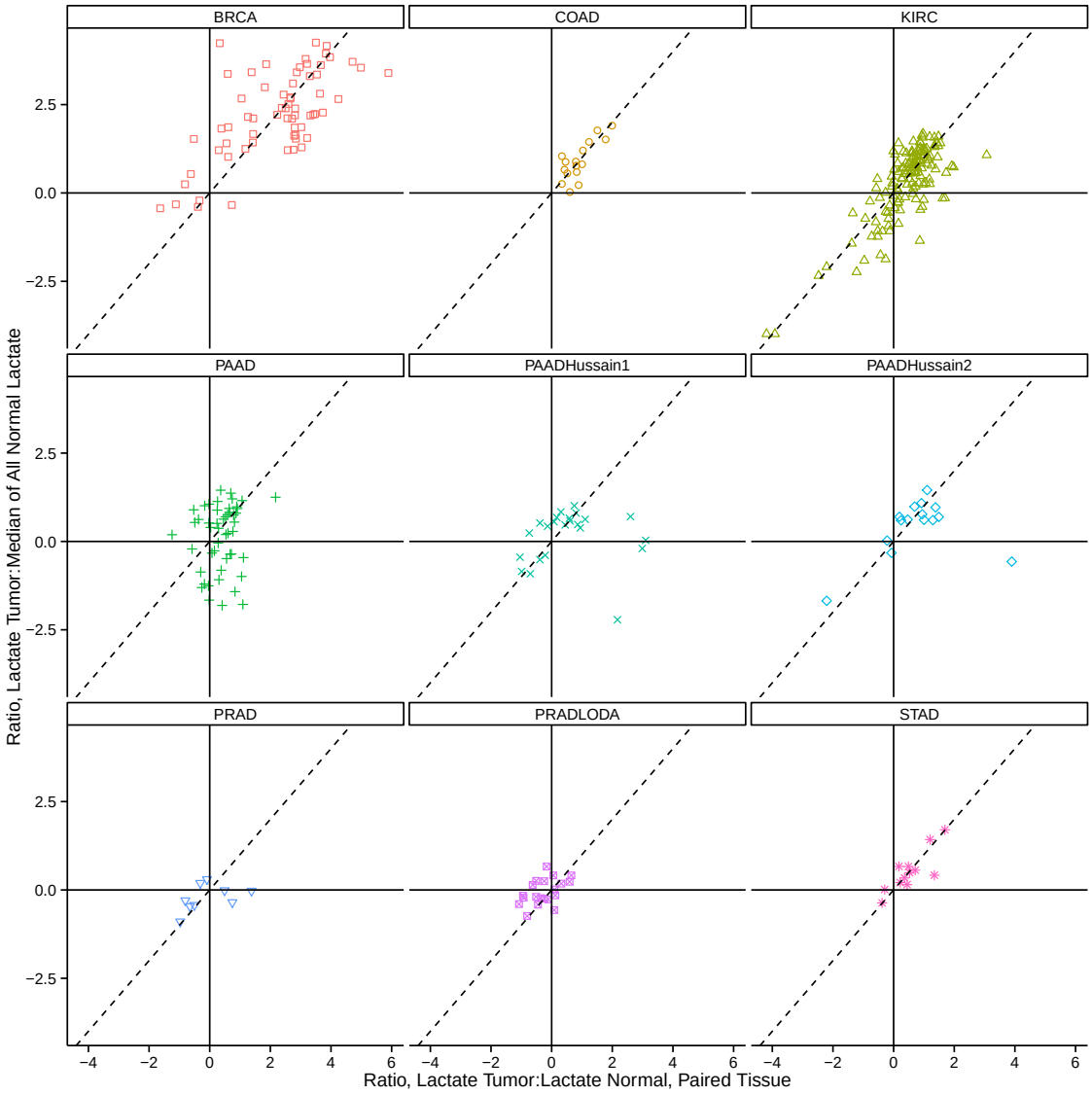

Figure S6. Insert caption.