

# The Metabolic Landscape of Tumors

E Reznik, A Luna, BA Aksoy, C Sander, others less worthy

<sup>1</sup> 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- $\rightarrow$  tumor transformation - Recurrent metabolic alterations - Clinical association with metabolites

## 1 Introduction

Tumors and normal tissue use metabolism to derive energy and cellular building blocks from the environment. Unlike normal tissues. However, to grow and divide in the face of increased oxidative and chemotherapeutic stress, cancers modulate the activity of metabolic pathways to (among other things) supplant ATP production, produce suitable levels of precursors for the production of membranes, proteins, and nucleic acids, and sustain appropriate levels of small-molecules necessary to maintain epigenetic integrity. As a whole, these metabolic alterations are implemented by changes in the levels of intracellular metabolites and enzymes.

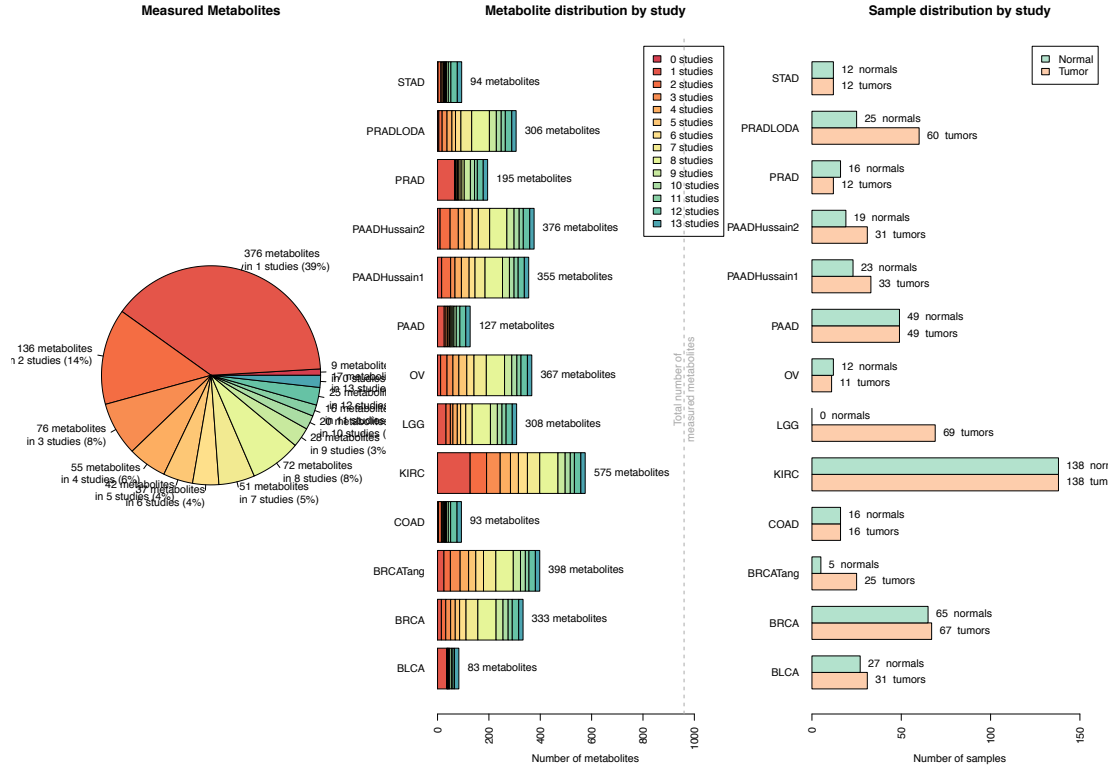
To date, the majority of large-scale work undertaken in studying the metabolism of cancer tissues has focused on the analysis of metabolic enzyme proteins, and the levels of their associated RNA transcripts [3, 6, 8]. While enzymes obviously constitute a critical component of the metabolic network, their levels shed limited light on how intracellular metabolite levels change in tumors. Because of the technical difficulty and comparatively high cost of measuring metabolite levels at a large scale, relatively few studies have been completed which measure metabolite levels in tumor tissue. The translatability of findings from metabolomics studies has been hampered by several challenges inherent to metabolomic data: 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.

Despite the challenges of working with metabolomics data, several pressing questions require...

In this work, we survey the metabolic landscape of tumors by integrating tumor metabolomic data from thirteen 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 thirteen datasets covering nine distinct cancer types (see Figure 1 and Supplementary Table XX). Data from all but two cancer types was collected using mass spectrometry (colorectal and stomach cancers was collected with NMR). Three cancer types (breast, prostate, and pancreatic [5]) were represented by at least 2 different datasets, enabling us to evaluate the consistency of findings across different studies.



**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: quantitative 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 for us, 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 central bioinformatic challenge to our meta-analysis was the identification of metabolites profiled across multiple studies. Unlike gene expression studies, 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 developed a bioinformatic pipeline which searched for synonymous identifiers of

each metabolite, and then used these synonyms to assemble a meta-dataset of all metabolomics data.

	Kidney Cancer						Breast Cancer					
	Normal Samples			Tumor Samples			Normal Samples			Tumor Samples		
Glucose	0.4	2.0	1.2	1.3	0.8	0.7	1.4	0.7	1.0	0.4	0.4	
Fructose	1.8	1.1	3.7	0.9	0.2	0.4	0.8	0.7	1.0	5.2	2.0	
Citrate	2.1	2.0	1.1	0.9	0.6	0.5						
Serine	0.5	0.1	0.6	1.4	2.0	4.0	2.5	3.2	0.1	0.1	1.0	

Direct comparisons **cannot** be made between measurements of different metabolites in the same study.

Direct comparisons **can** be made between measurements of the same metabolite across the same study.

Each entry corresponds to the ratio of (1) abundance of metabolite in a given sample, to (2) median abundance of metabolite across all samples in the same study. Missing data is imputed to lowest

**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

We began with a relatively simple question: how do the metabolic changes associated with transformation to cancer compare to the natural metabolic variation one might observe in normal tissue? Each tissue can be expected to have some characteristic level of fluctuation in metabolite levels. In turn, we asked if the similarity between a randomly selected pair of tumor and normal tissues is substantially greater than the similarity of two randomly chosen normal tissues. To estimate similarity, we calculated a metabolic distance between each pair of tissue samples in a study (see Methods).

We found that the metabolic similarity of tumors and normal tissues varied greatly, in a tissue-of-origin-dependent manner (Figure 3). Some cancer types (including clear-cell kidney and breast cancers) were highly dissimilar from their normal tissue counterparts. At the other end of the spectrum, we observed that prostate and pancreatic tumors displayed a surprisingly small amount of variation, compared to the natural variation inherent to normal prostate and pancreatic tissue. Importantly, this effect was evident all replicate studies of the same cancer type (*i.e.* both breast cancer studies showed increased variation, orange bars in Figure 3).

An orthogonal method for analyzing high-dimensional data, and one that is conventionally used as a

first-step during analysis, is to project it onto a highly informative, low-dimensional space using performed principal components analysis (PCA). Upon applying PCA to each dataset in our meta-analysis, we found that the majority of cancer types displayed a clear separation of tumor and normal samples. Surprisingly, however, we again found that prostate tumors and pancreatic tumors displayed poor to no separation from their respective normal samples using the first 2 principal components (see SI Figs XX-XX, evident in both prostate studies and all three of the pancreatic studies). Together with the results in Figure XX, these observations suggest that prostate and pancreatic tumors exhibit relatively small changes in the process of transformation, when compared to the natural metabolic variation evident in normal tissues.

The perplexing observation that prostate and pancreatic tumors showed relatively minor changes in metabolite abundances compared to normal tissue led us to investigate the metabolic alterations associated with tumors at higher resolution. We examined, on a study-by-study basis, which metabolites were differentially abundant between tumor and normal samples (Mann-Whitney U test, p-value  $\leq 0.05$ ). We found the number of differentially abundant metabolites varied drastically from one cancer type to the next (Figure XX). 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.

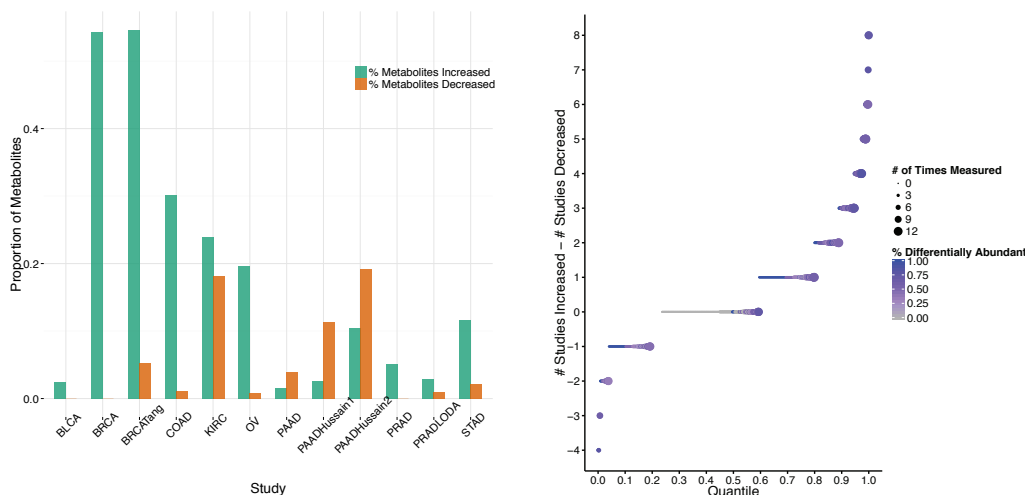
This analysis also revealed a cancer-type-dependent trend towards unequal proportions of metabolites which increased or decreased between tumor and normal tissues. In both breast cancer studies, we found that nearly all metabolites deemed differentially abundant were at higher levels in tumor tissue, 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 cancers. 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).

## 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).

To make greater sense of which metabolic processes may be recurrently up- or down-regulated in tumors, we used our aggregated differential abundance data to complete pathway analysis. For each

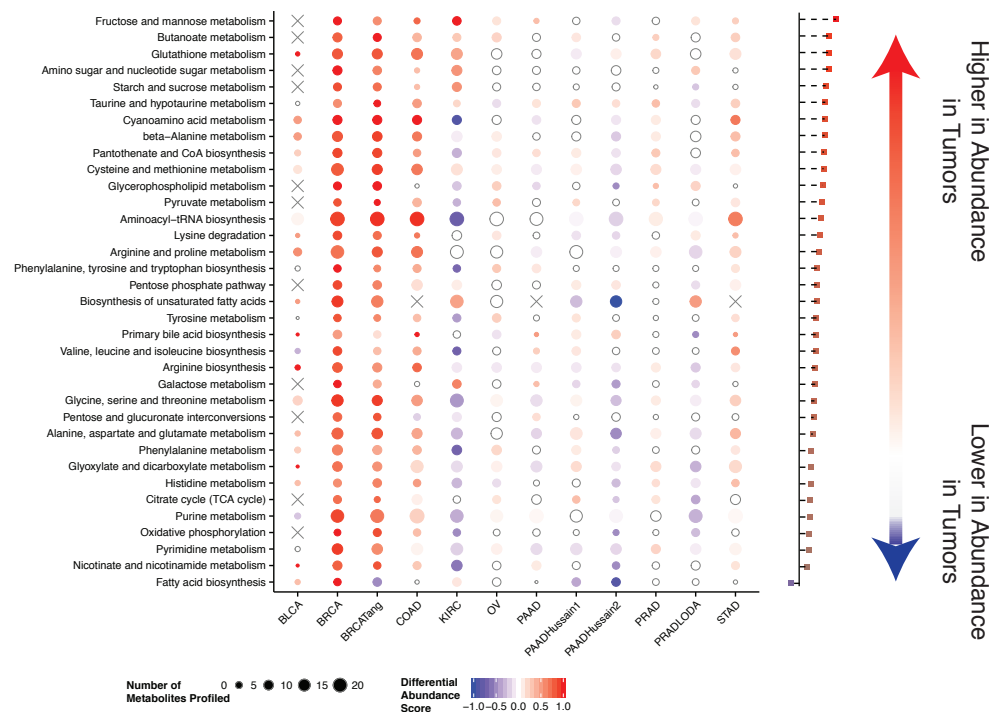


**Figure 3. Tumor/Normal Comparison.** (A) Comparison of metabolic alterations in tumors to natural metabolic variation in normal tissue. By comparing the similarity of (1) randomly chosen pairs of tumor/normal tissues, to (2) randomly chosen pairs of normal tissues, we assessed the magnitude of change 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.

study, we mapped metabolites onto KEGG pathways, and calculated an aggregate differential abundance score for each pathway (see Methods). We observed a tendency across cancers for increases in central carbon metabolism (including the KEGG pathways glycolysis/gluconeogenesis, fructose and mannose metabolism, and propanoate metabolism), and a general down-regulation of lipid and fatty acid pathways (such as glycerolipid metabolism and fatty acid biosynthesis). 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].

## 5 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

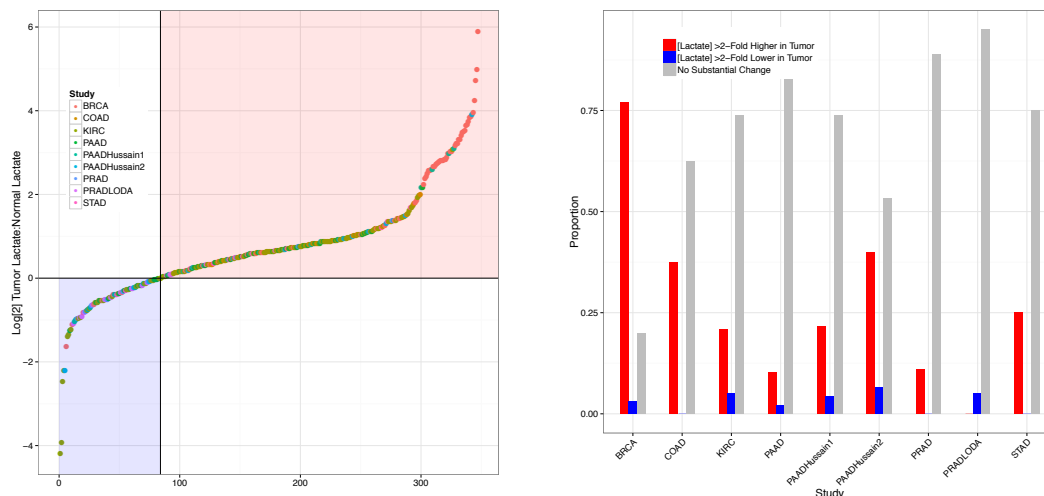


**Figure 4. Recurrent metabolic alterations.** (A) Heatmap of metabolites which are differentially abundant across at least 8 different cancer studies. (B) Pathway analysis **Write more.**

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 [15]. 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. 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



**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)

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.

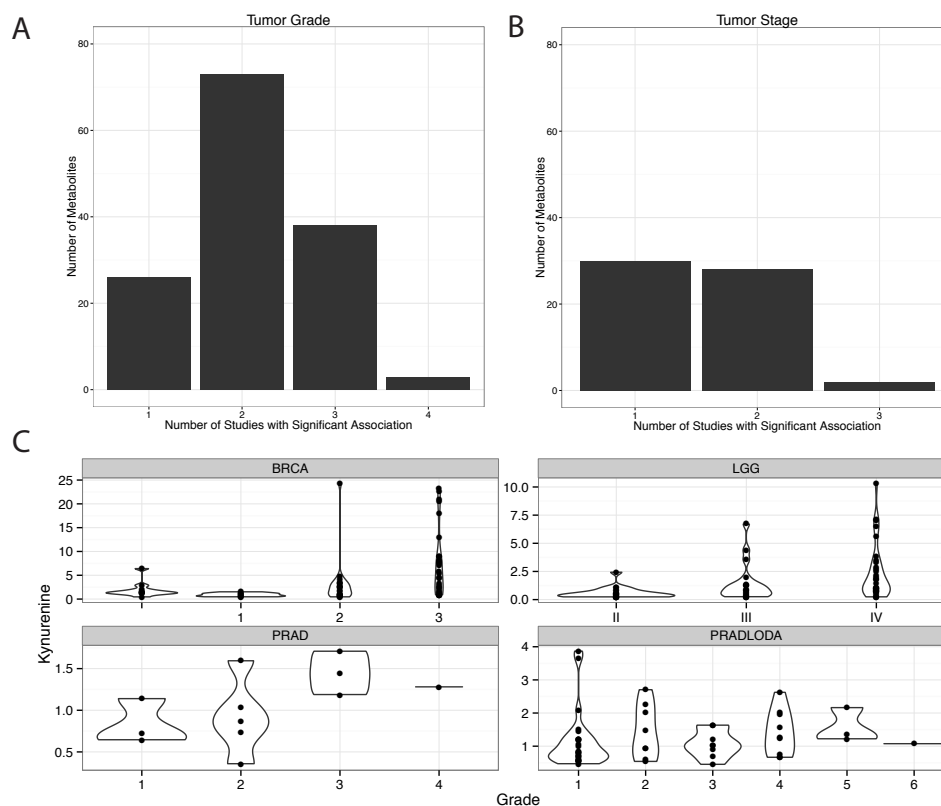
## 6 Metabolic Indicators of Tumor Progression

The processes, including genetic and signaling events, which drive the progression of cancer to more aggressive stages are distinct from those which initiate 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 appearance of tumor cells. In contrast, tumor stage describes the severity of a tumor based on its size, infiltration 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 analysis 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 tryptophan by two groups of enzymes: tryptophan dioxygenases and indoleamine 2,3-dioxygenases. 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).

## 7 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 (dimensionless) 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.

Our analysis presented here is a first step towards a more holistic understanding of cancer metabolism. While detailed analyses of metabolic flux in cancer models have investigated central carbon metabolism, the remainder of the several thousand reactions and metabolites constituting human metabolism remain to be adequately investigated. In contrast to flux studies, which provide quantitative assessments of the activity of each metabolic reaction under study, the data we have analyzed here constitutes static snapshots of the abundance of metabolites in a tumor. Deriving estimates of metabolic flux from these estimates is non-trivial, but may be amenable via methods like flux balance analysis [?, 11]. The compiled data from our study, as well as the results of all analyses, are provided in the Supplementary Data files. We expect that the data will serve as a useful tool for the cancer community at large for validating observations, and for future inquiries into the metabolic alterations underlying transformation and tumor progression.

## 8 Methods

### 8.1 Data Imputation and Standardization

For data which was already imputed and standardized (including the BLCA, KIRC, BRCA, BRCA\_Tang, OV, PAAD\_Hussain1, PAAD\_Hussain2, PRAD\_LODA 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.

### 8.2 Metabolite ID Mapping

Augustin please fill in

### 8.3 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-parametric Mann-Whitney U-tests.

### 8.4 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.

## 8.5 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.

## 8.6 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 [16]. P-values from the chi-squared test were corrected using the Benjamini-Hochberg procedure.

## 8.7 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 [16], and then transformed to two-sided combined p-value.

## References

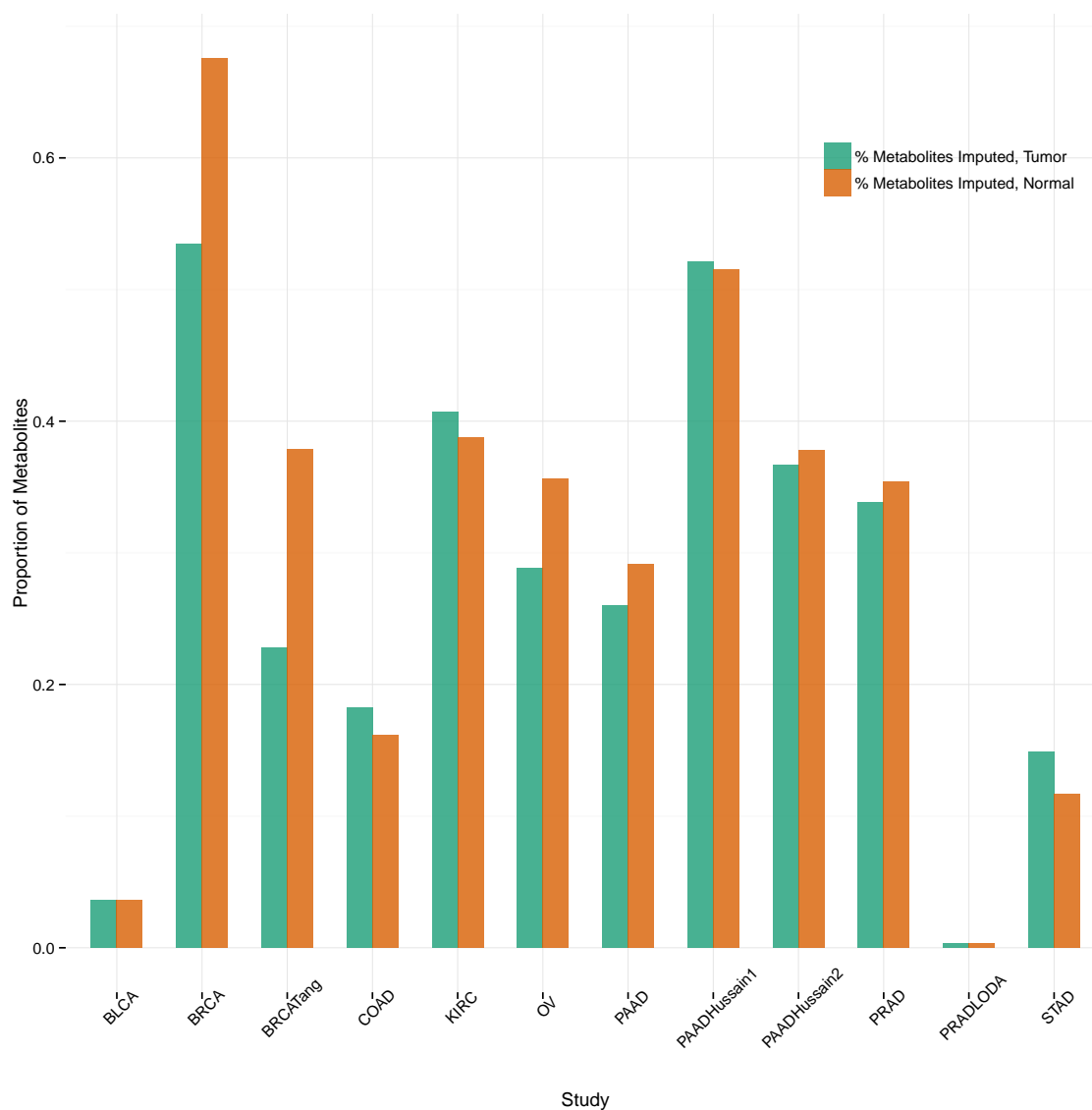
1. Prakash Chinnaiyan, Elizabeth Kensicki, Gregory Bloom, Antony Prabhu, Bhaswati Sarcar, Soumen Kahali, Steven Eschrich, Xiaotao Qu, Peter Forsyth, and Robert Gillies. The metabolomic signature of malignant glioma reflects accelerated anabolic metabolism. *Cancer research*, 72(22):5878–88, November 2012.

2. Miranda Y Fong, Jonathan McDunn, and Sham S Kakar. Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. *PloS one*, 6(5):e19963, January 2011.
3. Francesco Gatto, Intawat Nookaew, and Jens Nielsen. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1319196111–, February 2014.
4. Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–74, March 2011.
5. Akiyoshi Hirayama, Kenjiro Kami, Masahiro Sugimoto, Maki Sugawara, Naoko Toki, Hiroko Onozuka, Taira Kinoshita, Norio Saito, Atsushi Ochiai, Masaru Tomita, Hiroyasu Esumi, and Tomoyoshi Soga. Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer research*, 69(11):4918–25, June 2009.
6. Jie Hu, Jason W Locasale, Jason H Bielas, Jacintha O’Sullivan, Kieran Sheahan, Lewis C Cantley, Matthew G Vander Heiden, and Dennis Vitkup. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nature biotechnology*, 31(6):522–9, June 2013.
7. Jurre J Kamphorst, Michel Nofal, Cosimo Commisso, Sean R Hackett, Wenyun Lu, Elda Grabocka, Matthew G Vander Heiden, George Miller, Jeffrey A Drebin, Dafna Bar-Sagi, Craig B Thompson, and Joshua D Rabinowitz. Human pancreatic cancer tumors are nutrient poor and tumor cells actively scavenge extracellular protein. *Cancer research*, 75(3):544–53, February 2015.
8. Roland Nilsson, Mohit Jain, Nikhil Madhusudhan, Nina Gustafsson Sheppard, Laura Strittmatter, Caroline Kampf, Jenny Huang, Anna Asplund, and Vamsi K Mootha. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nature communications*, 5:3128, January 2014.
9. Carmen Priolo, Saumyadipta Pyne, Joshua Rose, Erzsébet Ravasz Regan, Giorgia Zadra, Cornelia Photopoulos, Stefano Cacciatore, Denise Schultz, Natalia Scaglia, Jonathan McDunn, Angelo M De Marzo, and Massimo Loda. AKT1 and MYC induce distinctive metabolic fingerprints in human prostate cancer. *Cancer research*, 74(24):7198–204, December 2014.
10. Nagireddy Putluri, Ali Shojaie, Vihas T Vasu, Shaiju K Vareed, Srilatha Nalluri, Vasanta Putluri, Gagan Singh Thangjam, Katrin Panzitt, Christopher T Tallman, Charles Butler, Theodore R Sana, Steven M Fischer, Gabriel Sica, Daniel J Brat, Huidong Shi, Ganesh S Palapattu, Yair Lotan, Alon Z Weizer, Martha K Terris, Shahrokh F Shariat, George Michailidis, and Arun Sreekumar. Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer research*, 71(24):7376–86, December 2011.
11. Ed Reznik, Pankaj Mehta, and Daniel Segrè. Flux Imbalance Analysis and the Sensitivity of Cellular Growth to Changes in Metabolite Pools. *PLoS Computational Biology*, 9(8):e1003195, August 2013.
12. Arun Sreekumar, Laila M Poisson, Thekkelnaycke M Rajendiran, Amjad P Khan, Qi Cao, Jindan Yu, Bharathi Laxman, Rohit Mehra, Robert J Lonigro, Yong Li, Mukesh K Nyati, Aarif Ahsan, Shanker Kalyana-Sundaram, Bo Han, Xuhong Cao, Jaeman Byun, Gilbert S Omenn, Debashis Ghosh, Subramaniam Pennathur, Danny C Alexander, Alvin Berger, Jeffrey R Shuster, John T Wei, Sooryanarayana Varambally, Christopher Beecher, and Arul M Chinnaiyan. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, 457(7231):910–4, February 2009.

13. Xiaohu Tang, Chao-Chieh Lin, Ivan Spasojevic, Edwin S Iversen, Jen-Tsan Chi, and Jeffrey R Marks. A joint analysis of metabolomics and genetics of breast cancer. *Breast cancer research : BCR*, 16(4):415, January 2014.
14. Atsushi Terunuma, Nagireddy Putluri, Prachi Mishra, Ewy A Mathé, Tiffany H Dorsey, Ming Yi, Tiffany A Wallace, Haleem J Issaq, Ming Zhou, J Keith Killian, Holly S Stevenson, Edward D Karoly, King Chan, Susmita Samanta, DaRue Prieto, Tiffany Y T Hsu, Sarah J Kurley, Vasanta Putluri, Rajni Sonavane, Daniel C Edelman, Jacob Wulff, Adrienne M Starks, Yinneng Yang, Rick A Kittles, Harry G Yfantis, Dong H Lee, Olga B Ioffe, Rachel Schiff, Robert M Stephens, Paul S Meltzer, Timothy D Veenstra, Thomas F Westbrook, Arun Sreekumar, and Stefan Ambs. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *The Journal of clinical investigation*, 124(1):398–412, January 2014.
15. Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N. Y.)*, 324(5930):1029–33, May 2009.
16. M C Whitlock. Combining probability from independent tests: the weighted Z-method is superior to Fisher’s approach. *Journal of evolutionary biology*, 18(5):1368–73, September 2005.
17. Geng Zhang, Peijun He, Hanson Tan, Anuradha Budhu, Jochen Gaedcke, B Michael Ghadimi, Thomas Ried, Harris G Yfantis, Dong H Lee, Anirban Maitra, Nader Hanna, H Richard Alexander, and S Perwez Hussain. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 19(18):4983–93, September 2013.

Study Name	Cancer Type	Data Type	Reference
BLCA	Bladder	Mass Spectrometry	[10]
BRCA	Breast	Mass Spectrometry	[14]
BRCATang	Breast	Mass Spectrometry	[13]
COAD	Colorectal	NMR	[5]
KIRC	Clear-cell Renal Cell	Mass Spectrometry	[?]
LGG	Glioma	Mass Spectrometry	[1]
OV	Ovarian	Mass Spectrometry	[2]
PAAD	Pancreas	Mass Spectrometry	[7]
PAADHussain1	Pancreas	Mass Spectrometry	[17]
PAADHussain2	Pancreas	Mass Spectrometry	[17]
PRAD	Prostate	Mass Spectrometry	[12]
PRADLODA	Prostate	Mass Spectrometry	[9]
STAD	Stomach	NMR	[5]

**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.

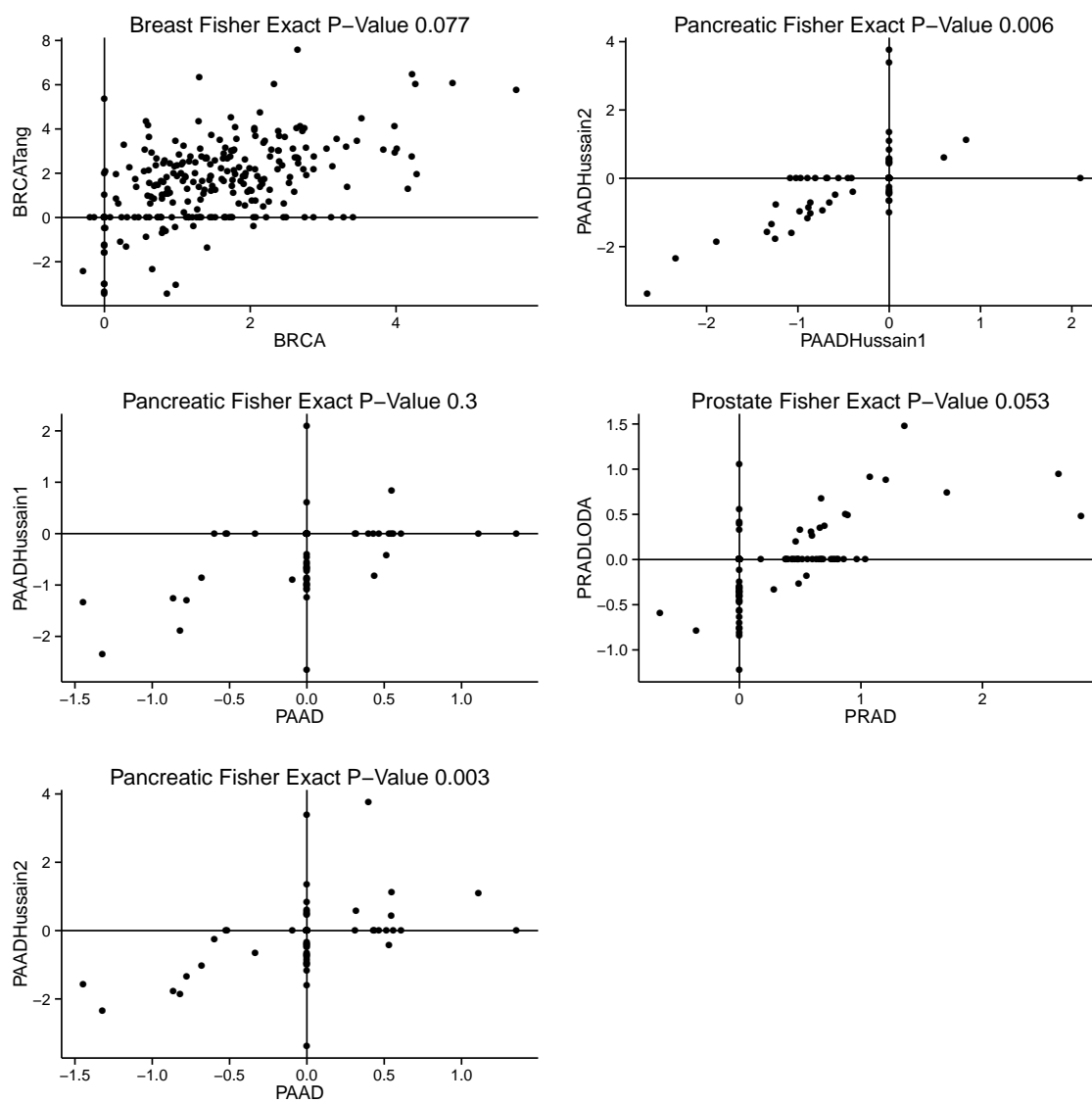
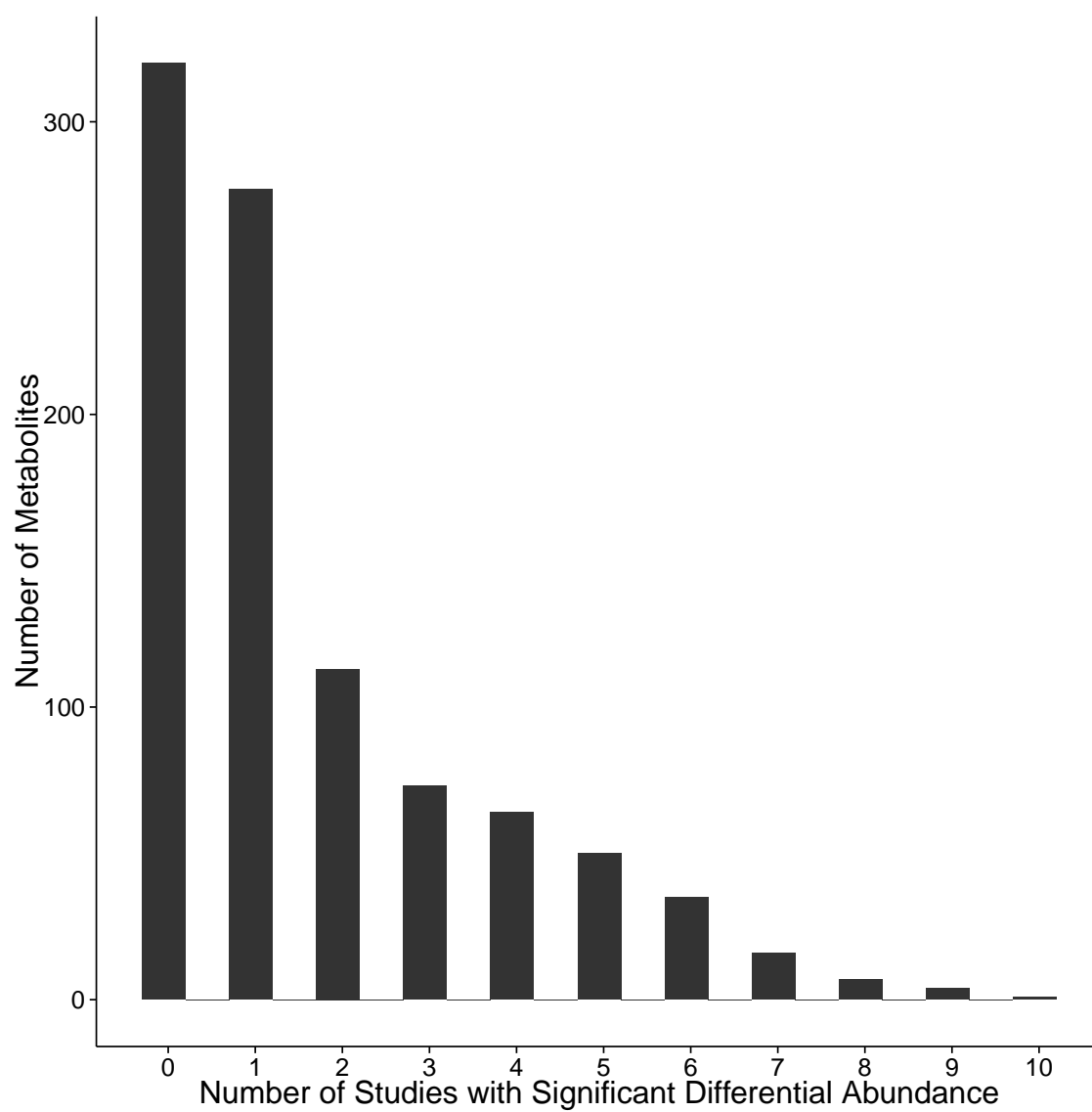


Figure S2. Insert caption.



**Figure S3.** Insert caption.



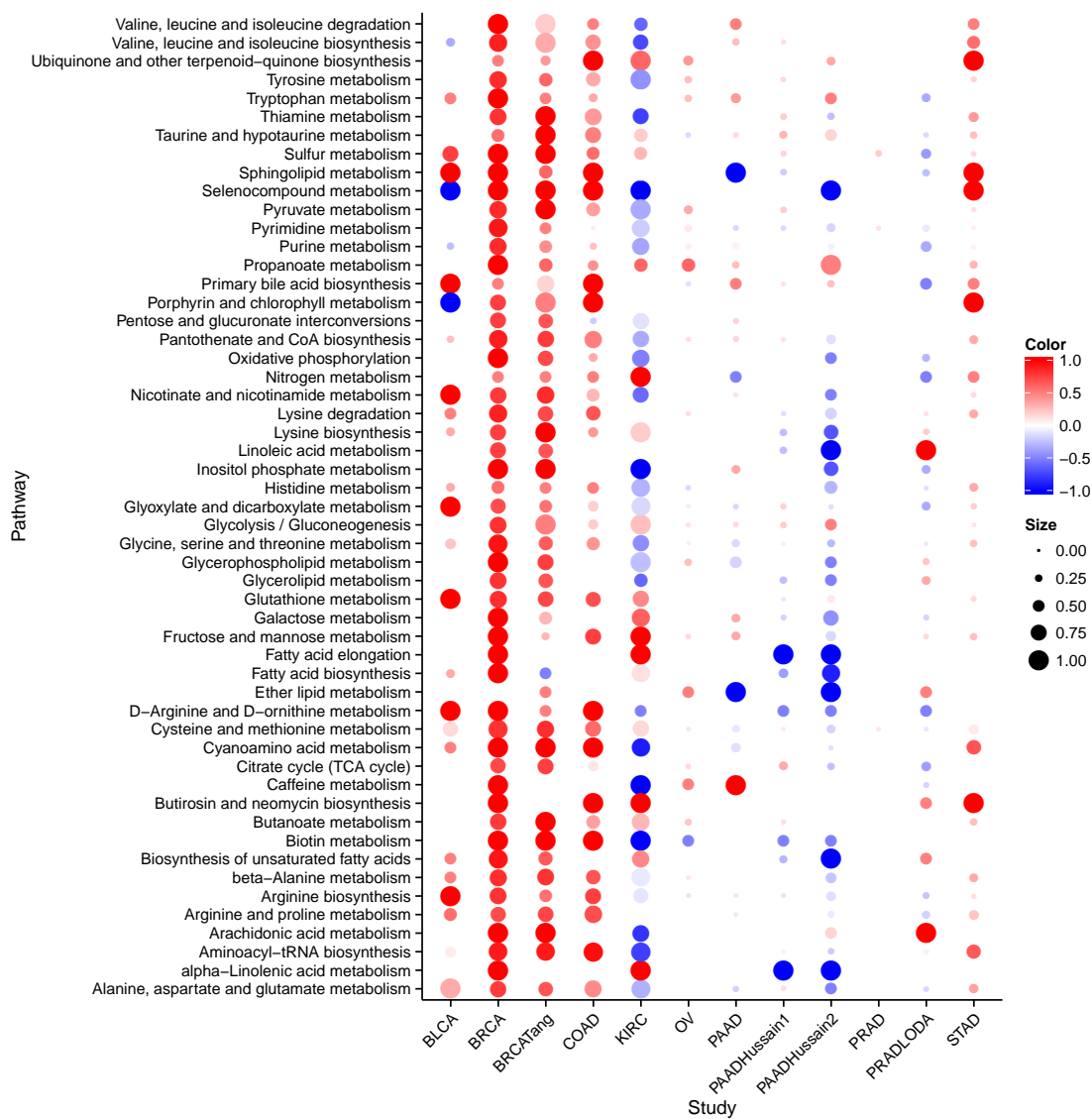


Figure S4. Insert caption.

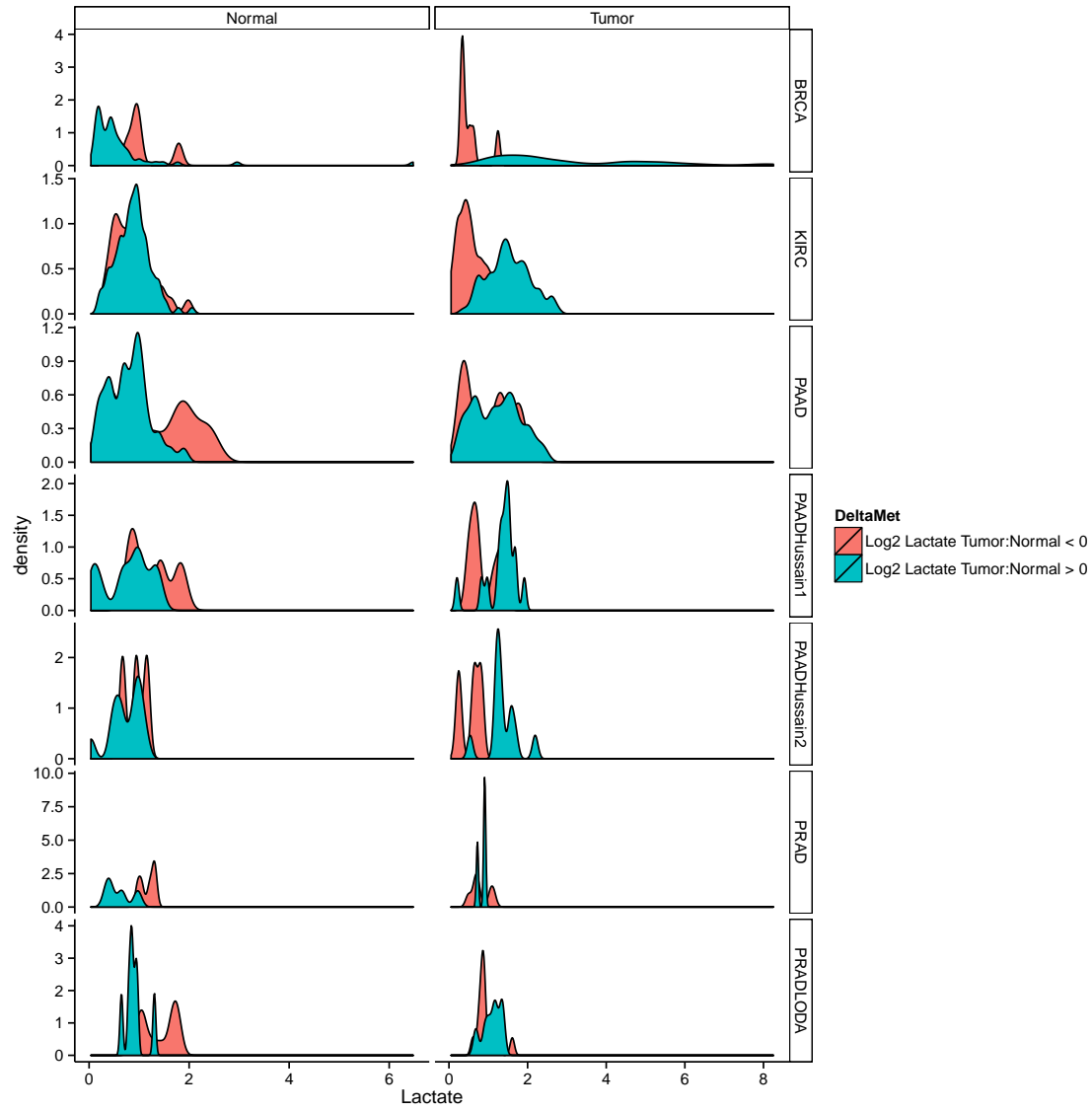


Figure S5. Insert caption.

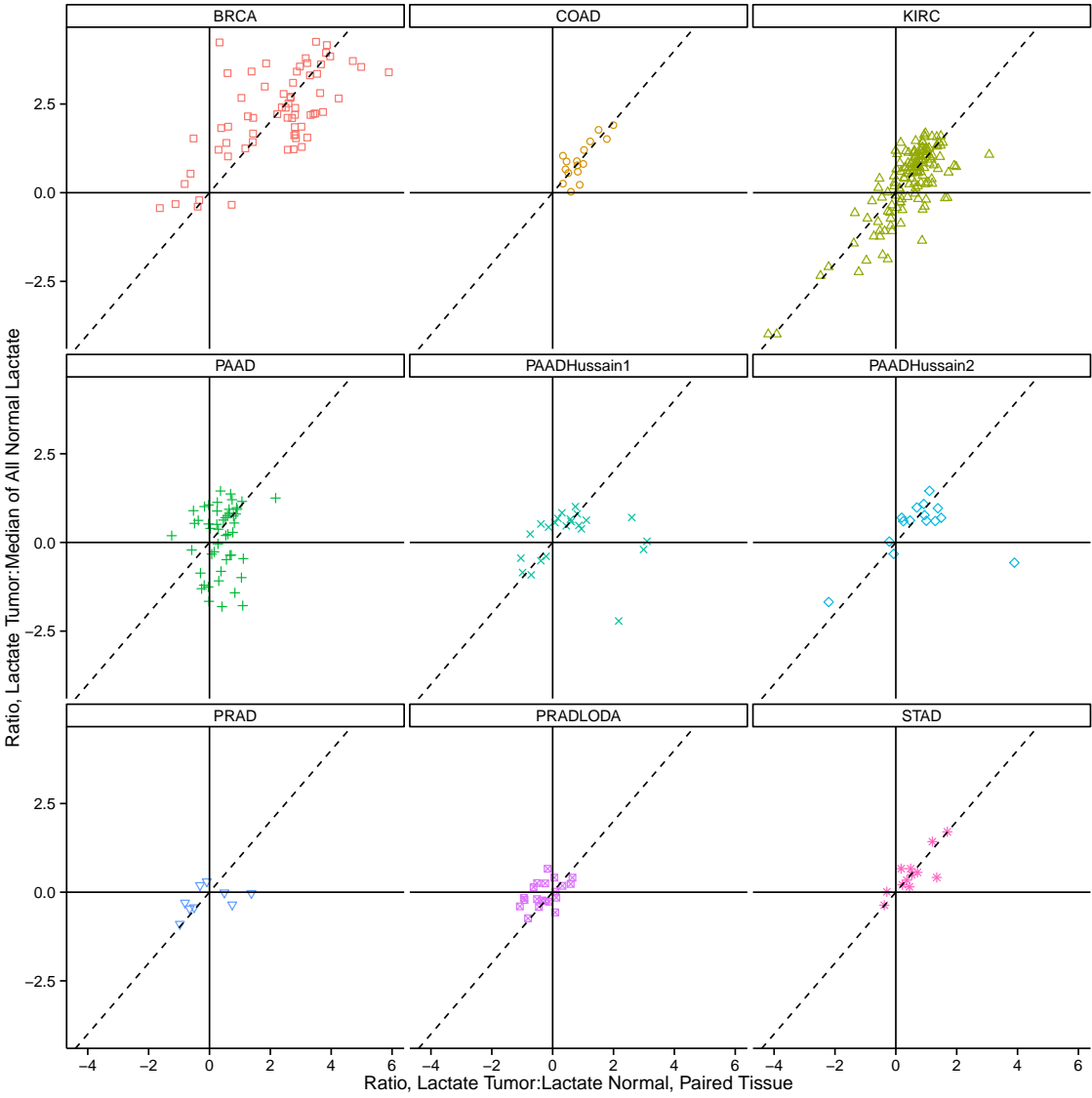


Figure S6. Insert caption.