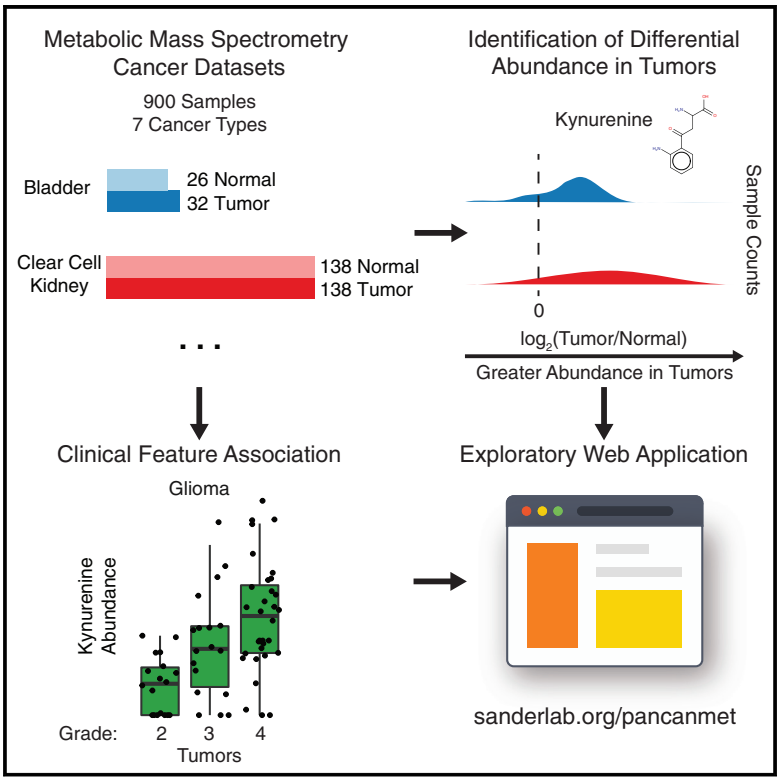
Article

Cell Systems

A Landscape of Metabolic Variation across Tumor Types

Graphical Abstract



Highlights

- Computational pipeline to integrate cancer metabolomics data across studies
- Free, open-source pipeline, data, and online visualization portal available
- Recurrent patterns of differentially abundant metabolites across cancer types
- Discovery of metabolites associated with aggressive disease across cancer types

Authors

Ed Reznik, Augustin Luna, Bülent Arman Aksoy, ..., Chad J. Creighton, A. Ari Hakimi, Chris Sander

Correspondence

reznike@mskcc.org (E.R.), sander.research@gmail.com (C.S.)

In Brief

Reznik et al. develop a computational approach for integrative analysis of metabolomics data and apply it to data from 900 tissue samples spanning seven different cancer types. They identify recurrent metabolic changes associated with tumor initiation and progression to aggressive disease.

Reznik et al., 2018, Cell Systems 6, 1–13
February 28, 2018
https://doi.org/10.1016/j.cels.2017.12.014

CellPress

Please cite this article in press as: Reznik et al., A Landscape of Metabolic Variation across Tumor Types, Cell Systems (2017), https://doi.org/10.1016/j.cels.2017.12.014

Cell Systems
Article

A Landscape of Metabolic Variation across Tumor Types

Ed Reznik,[1,2,14,15,*] Augustin Luna,[3,4,14] Bülent Arman Aksoy,[5] Eric Minwei Liu,[6,7] Konnor La,[1,8,9] Irina Ostrovnaya,[2] Chad J. Creighton,[10,11] A. Ari Hakimi,[12] and Chris Sander[3,13,*]

[1]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
[2]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
[3]cBio Center, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[4]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[5]Genetics and Genomics Department, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[6]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10065, USA
[7]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA
[8]Laboratory of Metabolic Regulation and Genetics, The Rockefeller University, New York, NY 10065, USA
[9]Tri-Institutional Training Program in Computational Biology & Medicine, New York, NY 10065, USA
[10]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA
[11]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA
[12]Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
[13]Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[14]These authors contributed equally
[15]Lead Contact
*Correspondence: reznike@mskcc.org (E.R.), sander.research@gmail.com (C.S.)
https://doi.org/10.1016/j.cels.2017.12.014

SUMMARY

Tumor metabolism is reorganized to support proliferation in the face of growth-related stress. Unlike the widespread profiling of changes to metabolic enzyme levels in cancer, comparatively less attention has been paid to the substrates/products of enzyme-catalyzed reactions, small-molecule metabolites. We developed an informatic pipeline to concurrently analyze metabolomics data from over 900 tissue samples spanning seven cancer types, revealing extensive heterogeneity in metabolic changes relative to normal tissue across cancers of different tissues of origin. Despite this heterogeneity, a number of metabolites were recurrently differentially abundant across many cancers, such as lactate and acyl-carnitine species. Through joint analysis of metabolomic data alongside clinical features of patient samples, we also identified a small number of metabolites, including several polyamines and kynurenine, which were associated with aggressive tumors across several tumor types. Our findings offer a glimpse onto common patterns of metabolic reprogramming across cancers, and the work serves as a large-scale resource accessible via a web application (http://www.sanderlab.org/pancanmet).

INTRODUCTION

Profiling the genomes, epigenomes, and proteomes of large cohorts of tumors has generated a detailed census of hallmark molecular alterations evident across many tumor types and likely to drive malignancy (Ciriello et al., 2013). These findings have contributed to a holistic understanding of cancer and have elucidated novel and potentially targetable vulnerabilities. In contrast, while altered metabolism is a well-accepted hallmark of cancer (Hanahan and Weinberg, 2011; Vander Heiden et al., 2009), the technical challenges of measuring the abundance of metabolites (e.g., using mass spectrometry) have limited the use of metabolomic profiling in studies of cancer tissues. Instead, much of the work in cancer metabolism has focused on examination of changes in the expression of metabolic enzymes and the identification of patterns of metabolic gene expression common to many cancer types (Gatto et al., 2014; Nilsson et al., 2014; Hu et al., 2013). In contrast, metabolomic studies of cancer have been largely limited to *in vitro* flux studies of central carbon metabolism and larger metabolomic studies focused on individual cancer types (Sullivan et al., 2016; Possemato et al., 2011; Mullen et al., 2012; Hensley et al., 2016; Lu et al., 2012, 2013), with notable exceptions (Mayers et al., 2016; Goveia et al., 2016).

Tumors grow and divide in the presence of stress imposed by proliferation and augmented by cytotoxic and targeted therapy. To survive, the activity of metabolic pathways is modulated to meet energetic demands, produce suitable levels of biosynthetic precursors, sustain redox potential, and maintain epigenetic integrity. As a whole, these metabolic alterations are implemented via changes in the levels of intracellular metabolites, enzymes, and transporters (DeBerardinis and Chandel, 2016; Pavlova and Thompson, 2016). As with genetic alterations, these metabolic alterations may differ substantially according to, among other factors, the tissue of origin and stage of disease. Furthermore, many functional and common genetic alterations in cancer involve metabolic enzymes (e.g., IDH, SDH, FH) or regulators of metabolism (e.g., MTOR, VHL) (Jalbert et al., 2017;

Haider et al., 2016; Grabiner et al., 2014). In contrast, in this study, we seek to derive some organizing principles for how tumors alter their metabolomes, the repertoire of intracellular, small-molecule metabolites constituting the cell.

Our work focuses on analysis of published mass spectrometry-based metabolomic profiling data of cancer tissues. Integration of mass spectrometry measurements from different labs is inherently difficult because of incomplete/inconsistent measurement conditions and normalization procedures between different labs. To overcome this limitation, we use meta-analysis, i.e., metabolomic data are analyzed separately within each dataset, and the results of the individual analyses are subsequently combined. This approach enables us to determine if different cancers exhibit common patterns of metabolic changes between tumor and normal tissues. By additionally integrating clinical features into our analysis (e.g., pathological grade of the tumor), we are able to examine the association between tissue metabolite levels and the clinically aggressive features of a tumor. Unlike other recent analyses, this approach produces a single merged dataset with standardized nomenclature. This dataset, the informatic pipeline used to produce it, and all analysis code are made publicly available at https://github.com/dfci/pancanmet_analysis for future work. An accompanying website (http://www.sanderlab.org/pancanmet/) is provided to allow readers to explore the dataset, including (1) univariate and bivariate change in metabolite levels and (2) clinical associations by metabolite and by study. The Supplemental Information also contains tables facilitating mapping metabolites to various database identifiers.

RESULTS

Assembly of a Cross-Cancer Compendium of Metabolomics Data

The first step of our analysis was to develop a computational framework to jointly analyze metabolomic data produced from different laboratories. We obtained published cancer tissue metabolomics data from 11 studies covering seven distinct cancer types (see Figures 1, S1, and Table S1: Merged Metabolomics). Data for all studies were collected by the original investigators using mass spectrometry. Three cancer types (breast, prostate, and pancreatic ductal adenocarcinoma) were represented by at least two different datasets, enabling us to evaluate the consistency of findings across different studies. In total, our dataset encompasses 928 tissue samples.

Usage of metabolomics data for the purpose of meta-analyses involves unique technical and informatic challenges. Unlike genomic sequencing, which frequently covers the full breadth of the exome or genome, metabolomics relies on the identification of a small (in this case, on the order of tens or hundreds) library of compounds. Each individual metabolomic study is likely to profile distinct assortments of these molecules (Cho et al., 2014). Therefore, molecules profiled in one study may partially overlap the set of molecules profiled in other studies, and inconsistent use of standard nomenclature for metabolites makes their "alignment" across studies challenging. For example, lactate is synonymously referred to as lactic acid and (S)-2-hydroxypropanoate and is reported alongside many other database identifiers (e.g., CHEBI, KEGG, HMDB, and PubChem IDs) (Kim et al., 2016; Wishart et al., 2007; Kanehisa et al., 2016).

Technical limitations and inconsistent use of common standards render infeasible the direct comparison of raw mass spectrometry-based data of metabolite levels from different laboratories. Nevertheless, there is a clear missed opportunity for analyzing such data jointly, e.g., for discovering common metabolic changes across cancer types. To overcome the technical issues associated with normalizing raw mass spectra from different labs, we elected to use pre-normalized, published data and to apply a series of tools to make comparison between them possible. Each of the datasets we used provided metabolite abundances normalized by the original investigators. In 3/11 studies, the acquired data were not imputed or were un-standardized (here, by un-standardized we specifically mean the data were on a scale unsuitable for visualization, *not* that the data were inappropriately normalized). For metabolomic data that were already imputed and standardized by the original investigators, we left the data unaltered. For the studies where this was not the case, we imputed any missing measurements for a compound with the minimum measured value of that compound. To simplify data visualization, we then divided all measurements of the compound by the median abundance of that compound within the study (for further details, see Figure 2, STAR Methods, and code at https://github.com/dfci/pancanmet_analysis). All downstream analysis of the processed metabolomics data (e.g., differential abundance tests) then relied on (1) the use of non-parametric statistical methods, which do not make assumptions on the underlying distribution of the data, and (2) the use of normal tissues as a "reference" by which to identify cancer-specific changes in metabolite abundance. Importantly, while our pipeline is robust to variation in the underlying distribution of the data, erroneous normalization by the original producers of the data will (obviously) lead to erroneous results from our pipeline.

The product of our computational pipeline is a merged metabolomics dataset containing data and samples from all studies and additional metadata (Figure S1, Table S1 Merged Metabolomics) including mappings to identifiers of several key databases (i.e., CHEBI, KEGG, HMDB, and PubChem IDs) found in Table S1: Merged IDs. Because ambiguous nomenclature from each individual dataset is resolved, metabolites that are measured across many studies are aligned and can be compared with meta-analysis. Access to the study data is available for interactive data exploration at the project website http://www.sanderlab.org/pancanmet. On this website, users may explore the data by visualizing univariate and bivariate relationships, changes in abundance between tumor and normal samples across user-selected sets of metabolites, associations between the available clinical data and metabolite levels, and links between drug compounds that target enzymatic processes used to synthesize metabolites of interest.

Analysis of Metabolic Variation in Tumors and Normal Tissue

We first sought to understand the extent to which each cohort of tumors differed from normal tissue. In analogy to prior studies of genetic changes across cancers, we expected that some tumor types may have relatively few metabolic changes compared with normal tissues, while others may have many.

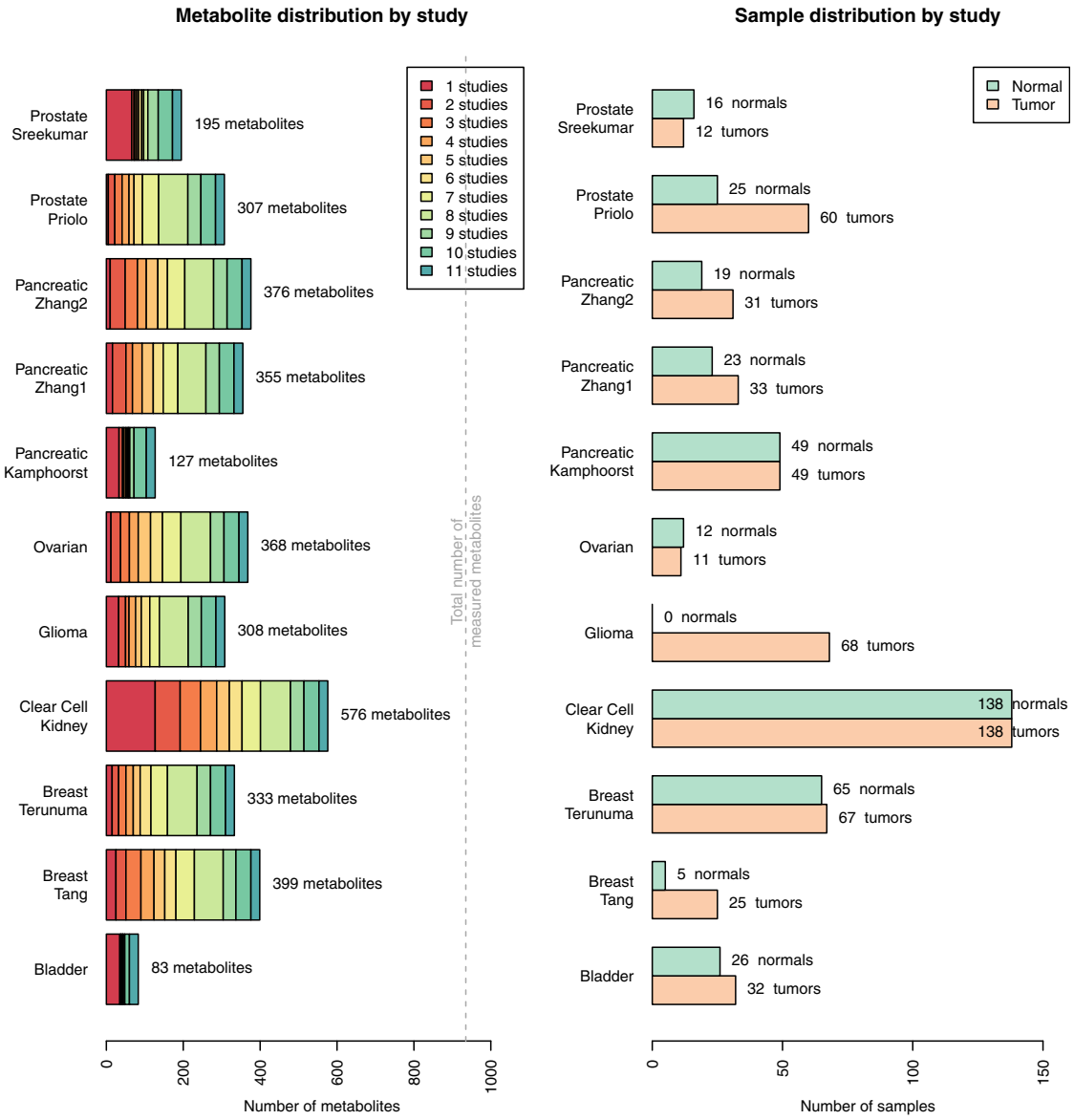


Figure 1. Metabolomics Data (928 Samples, 7 Cancer Types, 11 Studies) Analyzed in This Study

Data from 11 distinct metabolomics studies 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. All but one study (glioma) contained normal samples. In some, but not all cases, tumor/adjacent normal tissue samples were collected from the same patient (i.e., "matched" or "paired" samples, terms used interchangeably). These matched data are indicated in Table S2.

Therefore, on a study-by-study basis, we examined which metabolites were differentially abundant between tumor and normal samples (Mann-Whitney U test, Benjamini-Hochberg (BH)-corrected $p < 0.05$). For the three cancer types for which we had replicate studies, changes in abundance across the studies were in good agreement (Figure S2). To further corroborate whether our differential abundance findings (and the data in general) reflect true biological differences or are artifacts of technical noise, we took advantage of detailed knowledge on metabolic network structure. We asked whether the levels of metabolites that are adjacent in the metabolic network (i.e., are substrates/products for a common enzyme) are more likely to be correlated to each other than non-adjacent metabolites. We expected that, if technical noise in the data was not excessive, we would see an enrichment for high positive correlations in adjacent metabolite pairs. Applying this analysis to each dataset in our study, we found that in all but one dataset (Bladder), adjacent metabolites were significantly more correlated than non-adjacent metabolites (Figure S2, Supplemental Information Table S1: Covariation_Results).

The proportion of differentially abundant metabolites varied substantially from one cancer type to the next (Figure 3A). When using the multiple-hypothesis-adjusted Mann-Whitney p value as the only criterion for significance (i.e., ignoring the

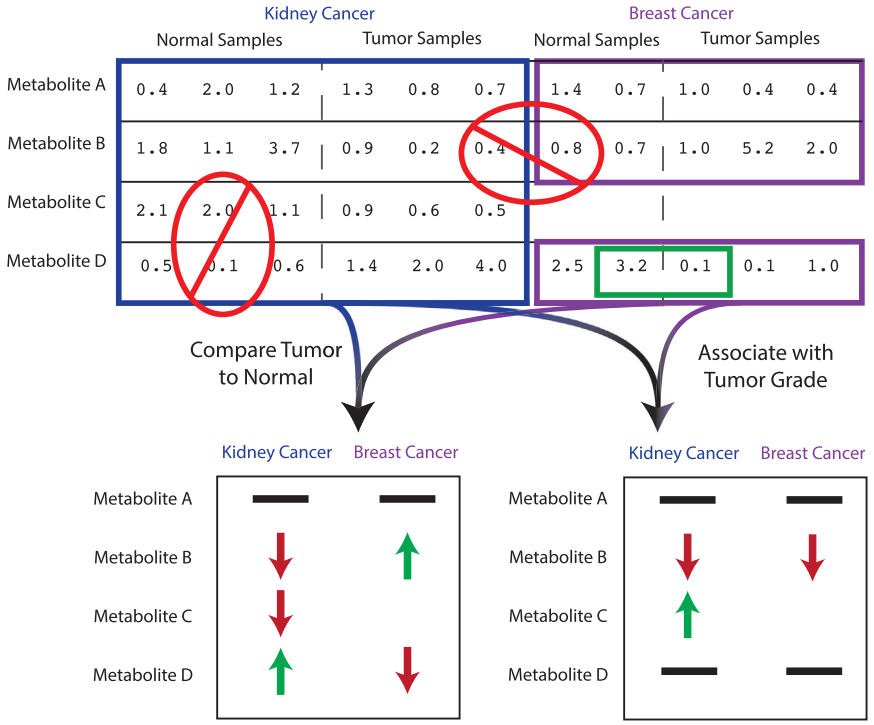


Figure 2. Workflow for the Aggregation and Comparison of Metabolite Profiles across Studies

Metabolomics data collected in different laboratories are difficult to directly compare. Our approach is to analyze data from each metabolomics dataset separately (e.g., comparing tumor with normal samples) and then aggregate these results across cancer types. Typically, for each metabolite, abundance is reported relative to the median of all measurements of a metabolite within a study. Comparison between different metabolites profiled in the same study is not directly possible (leftmost red oval). Furthermore, comparison between the same metabolite profiled across different studies is not possible (topmost red oval). Throughout our analysis, we 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). Correlative analysis with variables of interest (e.g., clinical data) is also feasible.

magnitude of change), over 60% of metabolites in both breast cancer studies and the kidney cancer study were differentially abundant. In contrast, pancreatic cancers had characteristically fewer differentially abundant metabolites (<20% in 2/3 studies, Table S1: Differential Abundance Summary). We tested whether the large variation in propensity for differential abundance arose due to sample-size effects, and we did not find evidence to support this hypothesis (Figure S3).

This analysis revealed a cancer-type-dependent trend toward unequal proportions of metabolites that 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 with normal tissue. Similar results were reported in the original publications for these two datasets (Terunuma et al., 2014; Tang et al., 2014). While biases were evident in other studies (e.g., differentially abundant metabolites in pancreatic tumors were more frequently depleted than accumulated in tumors), the effect was particularly striking for breast cancer. Although it is not possible for us to determine the source of this effect, we speculate it could arise from the disproportionate extent of data imputation for metabolites in normal breast tissue, compared with the extent of imputation in tumor tissue (Figure S1). Because the effect is replicated in both breast cancer studies, it is possible that metabolites are broadly at higher concentrations in breast tumor compared with benign breast tissue.

It is also possible that tumor and normal tissues differ in the inherent variability of metabolite levels, without differential abundance. To test this possibility, we compared the median absolute deviation (MAD; a non-parametric measure of the variability of a collection of measurements) of a metabolite for tumor and normal samples in each study. To control for potential biases in the relative numbers of tumor and normal samples, only data with matched tumor and adjacent normal samples from the same patient were used. The distribution of MAD ratios over all metabolites for a given cancer type indicates whether metabolite levels were inherently more variable in tumors (distribution shifted to the right of zero) or less variable in tumors (distribution shifted to the left) compared with normal tissue (Figure 3C). We observed significant differences in variability of metabolite levels across tumor types: metabolite levels in breast and kidney tumors were significantly more variable than in their normal tissue counterparts (Mann-Whitney $p < 0.05$). In contrast, in 2/3 of pancreatic tumor datasets, tumors were significantly less variable than their normal tissue counterparts. Together with the results above, these results reflect major differences in both absolute metabolite levels and metabolic variability associated with each individual cancer type.

Common Patterns of Metabolic Alterations across Cancers

Cancers share common genetic alterations despite inherent physiological and molecular differences in their cell of origin. Describing these commonalities across cancer types has expanded our understanding of the operation of molecular processes and the coordination of disparate cellular pathways (Lawrence et al., 2014; Zhang et al., 2017). Furthermore, it has changed our perspective on how patients are treated and recruited to clinical trials (Conley and Doroshow, 2014). Previously, because of a lack of data, it has not been possible to assess whether metabolic alterations, manifested through shifts in the abundances of metabolites from tumors to normal tissue, also exhibited common patterns across different cancer types. With the dataset organized for this work, we can begin to address this question.

Unlike the prior analysis, we now focused on analyzing common metabolomic changes across cancers of different tissues

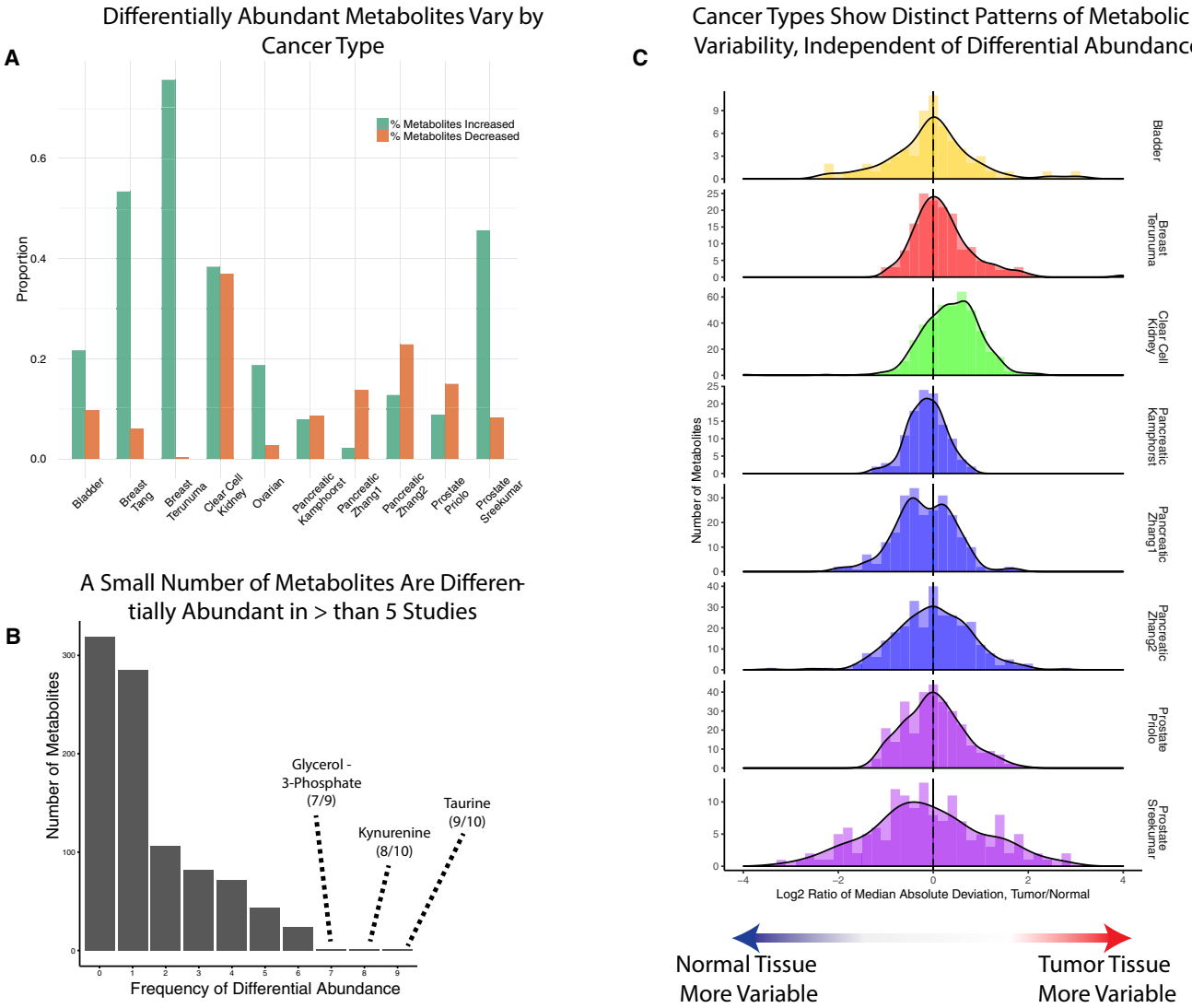


Figure 3. Differential Metabolite Abundances in Tumor versus Normal across Cancer Types

(A and B) In total, 10/11 studies contained adjacent normal tissue for comparison and were included in this analysis (in total, 482 tumor samples and 378 normal samples). Analysis was done without regard to whether tumor/normal samples were paired. (A) Proportion of metabolites in each study that were differentially abundant in tumor versus normal samples (BH-corrected $p < 0.05$). (B) The frequency of differential abundance across all metabolites. Most metabolites are rarely differentially abundant, in part because they may be measured in only a small number of studies. A small number of metabolites are often differentially abundant. (C) Variability of metabolite levels for each study. MAD ratio distributions show the inherent variability of metabolite levels; distributions shifted to the right of zero show a greater degree of variability in tumors compared with normal tissue, while those distributions shifted to the left show less. A total of 343 pairs of tumor/normal samples from eight studies (Bladder, Breast Terunuma, Breast Tang, Kidney, Pancreatic Kamphorst, Pancreatic Zhang1, Pancreatic Zhang2, Prostate Priolo, and Prostate Sreekumar) were used.

of origin. We identified 27 metabolites differentially abundant (tumor versus normal tissue) in at least 6 studies (BH-corrected $p < 0.05$, Figures 3B and 4, Table S1: Differential Abundance Summary). Among these, a single metabolite, taurine, was differentially abundant in 9 of the 10 studies in which it was measured (increased in tumors in 6 studies, decreased in tumors in 3 studies, Figure 3B). Only a handful of metabolites were uniformly elevated in 5 or more studies (with no findings of significant depletion in any tumor type): four acyl-carnitine species (acetyl-, butyryl-, hexanoyl-, and octanoyl-), carnitine itself, beta-alanine, 5,6-dihydrouracil, citrulline, kynurenine, and lactate (Table S1: Differential Abundance Summary). Comparatively fewer metabolites had a tendency toward recurrent depletion in cancers. Four metabolites, including three medium-chain fatty acids (pelargonate, caprate, and laurate), were depleted in 4 different studies. These results were in good agreement with a recent meta-analysis (see Figure S4) (Goveia et al., 2016) (e.g., recurrent elevation of lactate and taurine, depletion of myo-inositol, and a general elevation [rather than depletion] of metabolites in tumors compared with normal tissue). Interestingly, Goveia and colleagues (Goveia et al., 2016) observed substantially more frequent elevation of various amino acids

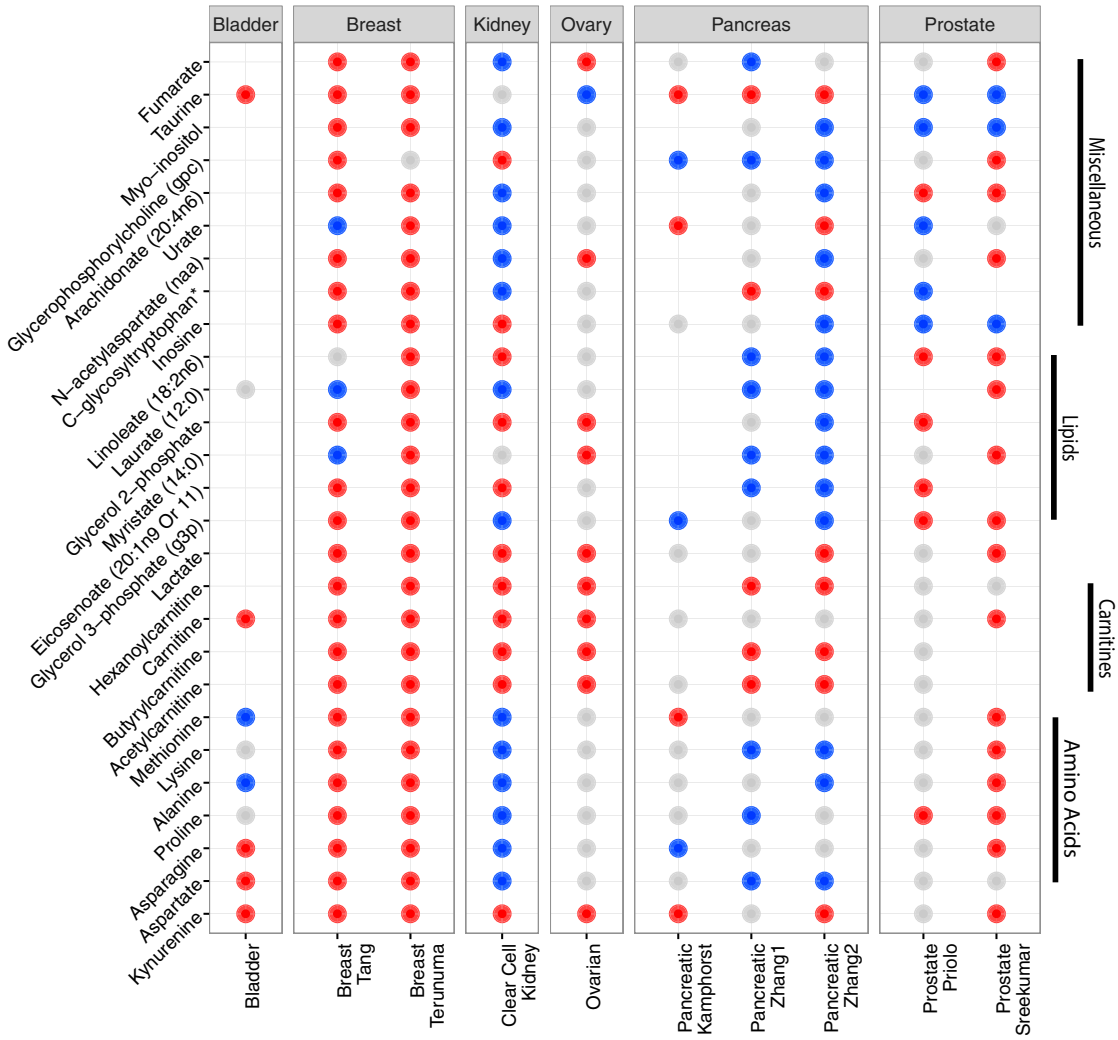


Figure 4. Differential Abundance across All Cancer Types

Blue circles indicate metabolites depleted in tumors relative to normal tissue, red circles indicate metabolites increased in tumors relative to normal tissue. Gray circles indicate metabolites with no statistically significant change in abundance. Only metabolites differentially abundant in six or more studies are displayed.

(e.g., serine, arginine) than we do here, which may in part be due to their broader coverage of cancer types.

Among the metabolites exhibiting recurrent differential abundance, several have been previously implicated in the development or progression of cancer. Perhaps, the best studied example is lactate, the terminal product of the high-flux, energy- and precursor-producing pathway of glycolysis and the metabolite originally described by Otto Warburg as elevated in cancerous tissues (Vander Heiden et al., 2009). In contrast to the changes in lactate, which are likely associated with a fundamental rearrangement of energy metabolism in cancer cells, the elevation of kynurenine in many cancer cells is likely not a reflection of changes to energetic demands. Instead, kynurenine is a derivative of tryptophan, which is well known to have immunosuppressive properties through pro-apoptotic effects when secreted into the extracellular milieu, and its elevation suggests recurrent changes to the tumor microenvironment across many cancers (Platten et al., 2012; Belladonna et al., 2007; Chen and Guillemin, 2009).

Pathway-Scale Metabolic Alterations

An alternative approach to understanding the broad patterns of metabolic alterations is through analysis of pathways into which metabolites may be organized. For each study, we mapped metabolites onto KEGG pathways (Table S1: Merged IDs) and calculated an aggregate differential abundance score for each pathway (see STAR Methods, Figure S5, Table S1: All Pathway). In total, 373 metabolites were identified as belonging to any given KEGG pathway, and from a collection of 246 KEGG pathways 179 (73%) were represented by at least one metabolite. In addition, by jointly comparing KEGG and DrugBank databases as stored in the Pathway Commons database using the paxtoolsr R package (Luna et al., 2016), we were able to identify 70 enzymes (involved in 35 KEGG pathways) producing or

consuming substrates in the current study that are targetable by 57 approved or investigational compounds present in DrugBank (Table S1: kegg_drugbank_results) (Luna et al., 2016; Cerami et al., 2011; Law et al., 2014). While a large number of pathways are represented minimally, the overall coverage of KEGG remains low with only 373 out of 3,330 (11%) KEGG metabolites being represented. Because of low coverage across pathways and with many metabolites missing from each pathway (i.e., a high degree of missing data and low statistical power), we did not assign statistical significance to pathway-level differential abundance scores (in line with prior work; Hakimi et al., 2016).

We restricted our analysis to well-represented metabolic pathways for which at least five metabolites in the pathway were profiled across six or more studies for a total of 34 pathways. Pathways associated with sugar metabolism (e.g., fructose/mannose metabolism, glycolysis/gluconeogenesis, and amino sugar and nucleotide sugar metabolism) were generally elevated. Many tumor types also had increases in glutathione metabolism, which produces the critical cellular antioxidant reduced glutathione (Singh et al., 2012). A single pathway, fatty acid biosynthesis, had weak, but recurrent depletion of its constituent metabolites across cancers and mirrored the increase of various acyl-carnitines across many cancer types (Figure 3). Acyl-carnitines are generated by oxidation of fatty acids and are used to refuel the tricarboxylic acid (TCA) cycle with acetyl-CoA. Our data suggest that the general elevation of acyl-carnitines and the concurrent depletion of medium-chain fatty acids in tumors may reflect an increased reliance on fatty acid catabolism to replenish the TCA cycle in tumor cells.

A central problem in metabolism research is understanding the connection between metabolite levels and the fluxes of reactions involving those metabolites as substrates/products. In general, flux through a metabolic reaction is dependent (in a complex, non-linear fashion) on (1) the levels of substrates/products of the reaction, (2) the levels of the enzyme catalyzing the reaction, and (3) the network context (i.e., the thermodynamics and kinetics of the reactions surrounding the flux of interest in the metabolic network). For example, it is not necessarily true that an increase in the levels of a substrate will increase flux through the corresponding substrate-utilizing enzyme (e.g., if transcription of the enzyme is downregulated). Due to this complexity, the inference of fluxes from metabolite levels remains unresolved and has been the focus of numerous studies from our group and others (Reznik et al., 2013; Gerosa et al., 2015; Reaves et al., 2013). Despite these challenges, drawing inferences on changes in flux from the pattern of metabolite changes is instructive for understanding high-level changes in metabolism and critical for generating experimentally testable hypotheses regarding metabolic flux.

With the above caveats in mind, we began to dissect the patterns of metabolite changes in glycolysis and the TCA cycle, which in addition to producing NADH/$FADH_2$/ATP generate high volumes of metabolic building blocks for biosynthesis (Figure 5). We first noted that while lactate levels are commonly elevated across most cancers, the remaining metabolites in glycolysis were in general not elevated. To evaluate if the changes in metabolite abundances were synchronized across samples, we calculated the correlation between metabolites in glycolysis and lactate. We observed that levels of metabolites in upper glycolysis (e.g., glucose 6-phosphate [G6P] and fructose 6-phosphate [F6P]) and 3-phosphoglycerate were frequently correlated to levels of lactate across most cancer types. In contrast, 2-phosphoglycerate and phosphoenolpyruvate were commonly not positively correlated or negatively correlated to levels of lactate. Taken together, the (1) prevalence of coordinated changes in metabolites from upper glycolysis with lactate, (2) increase in lactate levels in 6/9 studies, and (3) increases in levels of phosphorylated glucose (G6P) in 4/8 studies suggest that increases in lactate levels may be attributed to increases in glucose uptake and phosphorylation and increases in glycolytic flux and excretion of lactate. As others have pointed out (Vander Heiden et al., 2009), such an increase in glycolytic flux may provide additional carbon (via overflow into peripheral biosynthetic pathways) needed for growth. Interestingly, one cancer type (clear cell kidney) had a strong (Spearman rho -0.6, $p < 10^{-14}$) anti-correlation between levels of citrate and lactate. Clear cell kidney tumors are known to be highly HIF-activated, which can induce phosphorylation and inactivation of pyruvate dehydrogenase via the action of pyruvate dehydrogenase kinase and lead to preferential shunting of pyruvate into lactate rather than into the TCA cycle.

Unexpected patterns related to tissue of origin emerged when contrasting the pattern of metabolomic changes in central carbon metabolism across cancer types. For example, both prostate cancer studies support a drop in citrate levels in prostate tumors compared with normal tissue, a feature that is not evident in any other cancer type. This is in agreement with prior research showing that while normal prostate glandular epithelium secretes large amounts of citrate, depletion of zinc (an inhibitor of mitochondrial aconitase) relieves a "metabolic bottleneck" and leads to an increase in the oxidation of citrate/aconitate to isocitrate (Singh et al., 2006). Similarly, clear cell kidney tumors are the only cancer type to show an increase in metabolites upstream of SDH in the TCA cycle (succinate, citrate, aconitate) and depletion of metabolites downstream of SDH (fumarate and malate). These tumors are known to be highly depleted of mtDNA and RNA (Reznik et al., 2016, 2017), which suppresses the activity of the mitochondrial respiratory chain and consequently complex II/SDH. We propose that the sudden shift in metabolite patterns at SDH in clear cell kidney tumors may therefore arise from a bottleneck associated with suppression of the mitochondrial respiratory chain.

In some ways, the depiction of metabolomic changes in Figure 5A may be misleading; while an increase in the abundance of a metabolite may be evident in aggregate across all samples, there may remain individual tumors that are depleted of that metabolite. To better understand the nature of elevated lactate levels in tumors, we leveraged a unique part of our data: 343 matched pairs of tumor/adjacent normal tissue samples (i.e., each such pair was derived from the same patient). While both paired and unpaired analyses are informative, a paired analysis normalizes for potential patient-to-patient variation (independent of tumor-to-normal variation) in the levels of a metabolite. Despite elevation of lactate in 6/9 studies, 82/319 (26%) paired tumor/normal samples contained lower levels of lactate than their matched normal tissue counterpart. While most cancer types were enriched for tumors with increased levels of lactate, prostate tumors were the outliers: the majority of prostate tumors had reduced levels of lactate relative to paired

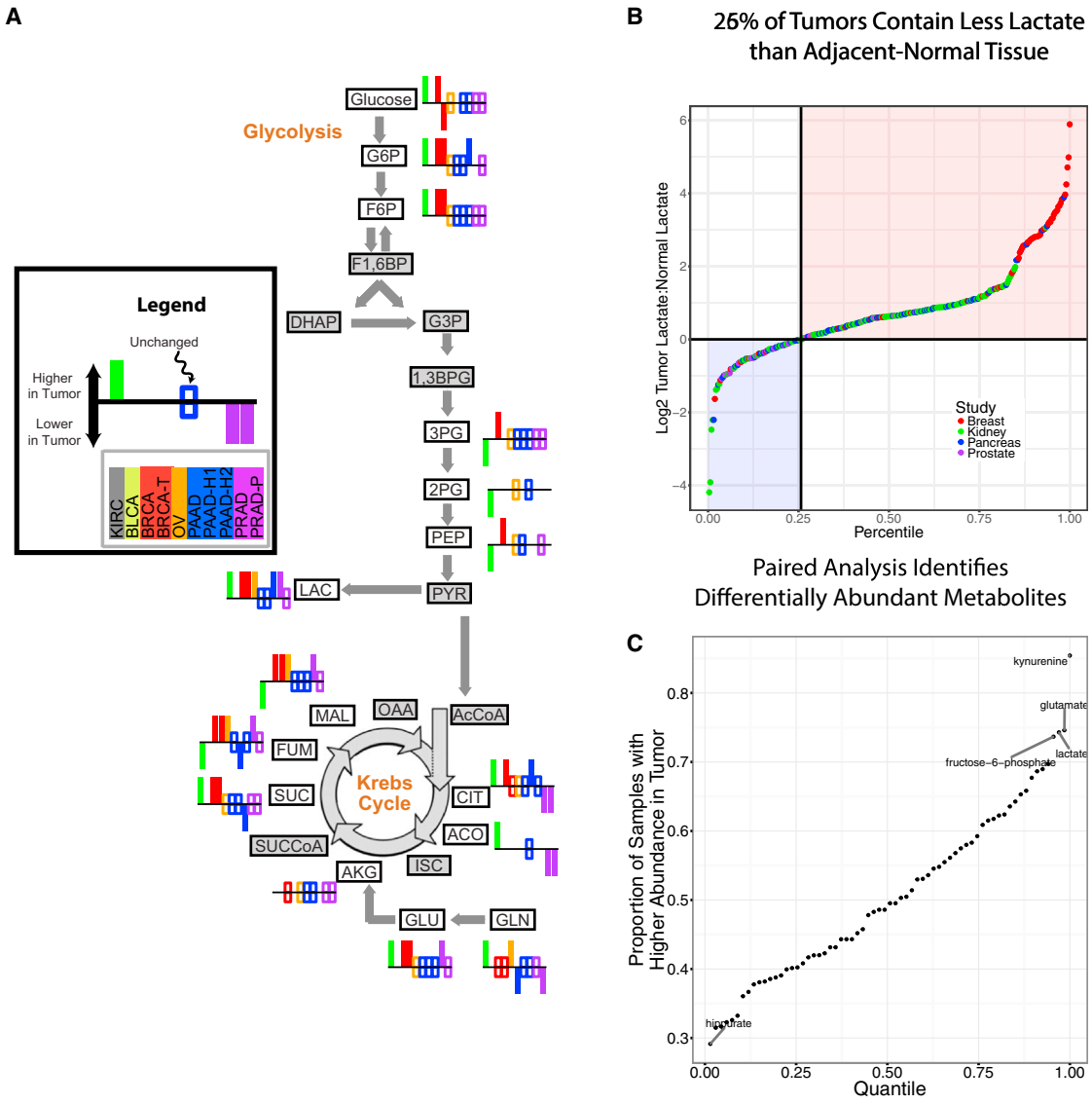


Figure 5. Central Carbon Metabolism across Cancer Types

(A) Differential abundance (tumor versus normal) in glycolysis and the TCA cycle across many cancer types. Bars indicate whether the metabolite was higher/lower/unchanged in tumor samples compared with normal tissue; the absence of a bar means the metabolite was not measured. BRCA-T, breast data from (Tang et al., 2014), PAAD-H1/2, pancreatic data from (Zhang et al., 2013), PRAD-P, prostate data from (Priolo et al., 2014). Data from 10/11 studies (glioma excluded for no normal samples, in total, 482 tumor samples and 378 normal samples).

(B) Sorted $\log_2$ ratio of lactate levels in matched tumor/normal samples from five cancer types. Approximately 26% of samples have less lactate in tumors compared with normal tissue. In both (A) and (B), analysis was done without regard to whether tumor/normal samples were paired.

(C) Frequency of elevation for each metabolite across all matched tumor/normal pairs (343 pairs in total). Kynurenine is elevated in 85% of matched samples.

adjacent normal tissue. Nevertheless, all studies had >10% of samples with depleted lactate levels relative to normal tissue (e.g., 32/138 ≈ 23% of kidney cancer pairs). This heterogeneity in lactate levels compared with normal tissue may partially explain the cancer-type-specific utility of FDG-PET imaging, which relies on elevated glucose uptake (and potentially an increased rate of lactate production) to visualize tumors (Menzel et al., 2013). Interestingly, when repeating the paired analysis above on all metabolites, we found that lactate was the third most frequently elevated metabolite, behind glutamate and kynurenine (Figure 5C).

Metabolic Indicators of Tumor Progression

Tumors become clinically aggressive by acquiring genetic, physiological, and morphological features that enable them to invade foreign tissues and metastasize. Although clinical data were sparsely reported in the studies we examined, five studies reported data on clinical stage or grade, with a greater number of data points available for tumor grade (Table S1: ClinFeatures). Importantly, while tumor grading systems vary from one cancer type to the next, the principle underlying grading (examining the abnormal appearance of a cell) is common across all cancer types: higher grade tumors are more aggressive. Therefore, we

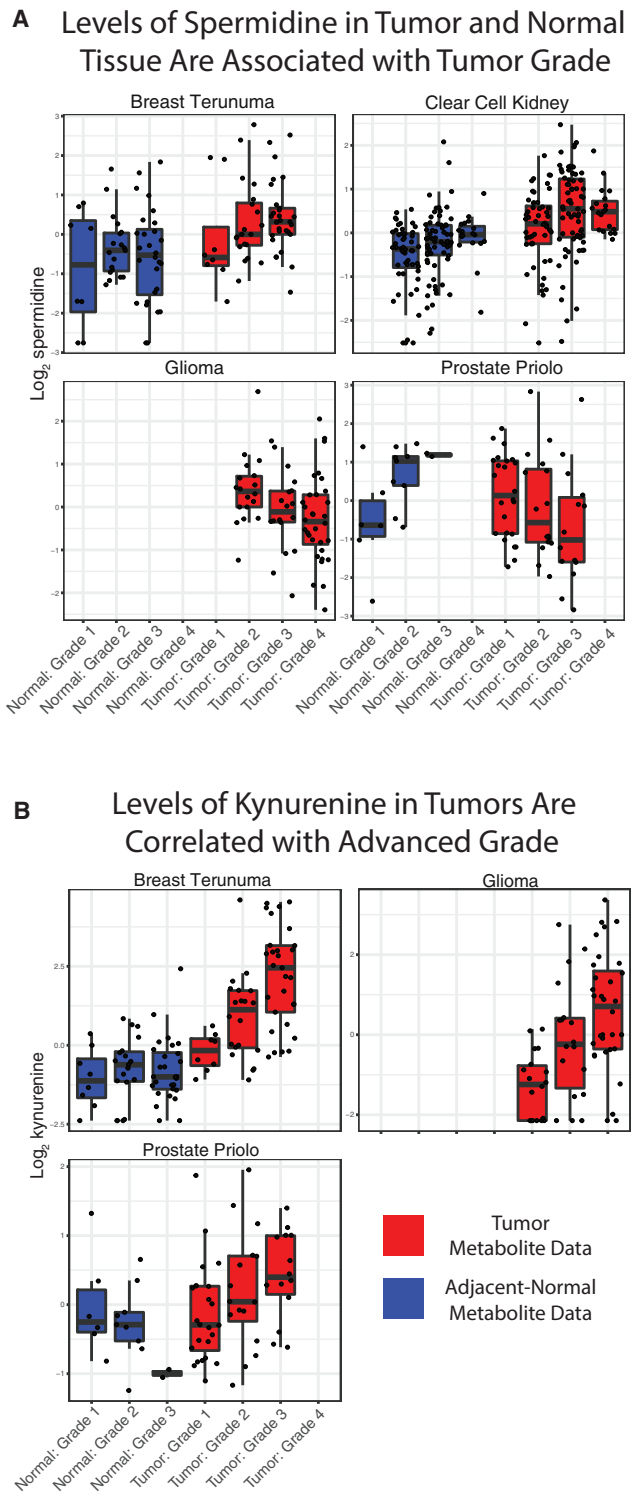


Figure 6. Metabolic Correlation with Tumor Progression

A total of 638 of the tissue samples had some type of associated clinical data (392 tumors and 246 normal tissues from seven studies, including Breast Terunuma, Breast Tang, Kidney, Glioma, Ovarian, Prostate Priolo, and Prostate Sreekumar). Analysis was done without regard to whether tumor/normal samples were paired. Meta-analysis identified 174 metabolites whose abundance in tumors was significantly associated with tumor grade in one or more metabolomics studies.

(A) Spermidine, a polyamine involved in cell proliferation, had a significant association with tumor grade across four different cancer types (breast, kidney, brain, and prostate cancer).

(B) Association of kynurenine levels with tumor grade across several cancers; kynurenine is part of the tryptophan degradation pathway with pro-apoptotic effects on immune cells.

examined our data for metabolic signals associated with progression of tumors to higher grade, a histological measure of the extent of abnormal appearance of tumor cells. Because of the decreased statistical power associated with looking for associations between metabolite levels and subsets of our tumor data, we adopted a meta-analysis approach. We identified metabolites that had recurrent increases/decreases in metabolite levels with increasing tumor grade across several cancer types. As before, our meta-analysis approach relies strictly on non-parametric methods (that do not make assumptions about the underlying distribution of the data) and is suitable for the metabolomics data used here. Furthermore, we designed our analysis to account for the frequency of imputed data for each metabolite under consideration (see detailed description in STAR Methods).

In total, we found 174 tumor tissue metabolites whose abundances were significantly correlated to tumor grade (BH-adjusted meta $p < 0.05$, Table S2). Filtering these *post hoc* results further to extract metabolites significantly associated with clinical features across many tumor types, we found four metabolites (spermidine, uracil, 3-aminoisobutyrate, and 2-hydroxyglutarate) associated with tumor grade in at least four studies and 57 metabolites associated with tumor grade in three studies. However, the number of metabolites significantly associated with grade in each study (*post hoc*) was correlated to the number of samples in the study with clinical data, suggesting that this analysis may be biased toward studies with larger sample sizes (Figure S3). Notably, spermidine had a significant association with tumor grade across four different cancer types (breast, kidney, brain, and prostate cancer) (Figure 6A). Spermidine is a polyamine, a class of molecules essential for eukaryotic cell growth as well as tumorigenesis (Mandal et al., 2013). Furthermore, a related polyamine (spermine) has been previously reported as a potential biomarker for prostate cancer aggressiveness (Giskeødegård et al., 2013). In many cases the direction of association for a single metabolite often varied between cancer types. For example, 2-hydroxyglutarate (2-HG) was associated with increased grade in breast and kidney tumors, but decreased grade in glioma and prostate tumors. We speculate that differences in genetic background, as well as in enantiomeric identity, may explain the discordant associations between 2-HG and tumor grade. Specifically, in gliomas, the D-enantiomer of 2-HG is produced by neomorphic *IDH* mutations, which are themselves associated with a better prognosis. In contrast, the more potent (in terms of inhibitory capacity) L-enantiomer of 2-HG, whose production is induced by hypoxia (Intlekofer et al., 2017), is likely the enantiomer produced by HIF-driven clear-cell renal cell carcinomas (Shim et al., 2014). To evaluate whether differences in the stereoisomeric identity of 2-HG actually account for the differences in association with grade, further enantiomer-specific metabolite profiling will be required.

We observed that kynurenine levels in tumor samples were associated with increased tumor grade in three studies (Figure 6B, Breast Terunuma, Glioma, and Prostate Sreekumar, uncorrected $p < 0.05$, meta $p < 10^{-4}$) and were trending toward significance in the other prostate study (uncorrected $p = 0.097$). Kynurenine is a metabolic by-product of the degradation of tryptophan by two groups of enzymes: tryptophan dioxygenases and indoleamine 2,3-dioxygenases (Platten et al., 2012). Binding of kynurenine to aryl hydrocarbon receptors suppresses the activity of T effector cells, as well as indirectly activating regulatory pro-tumorigenic T cells (Nguyen et al., 2014). Kynurenine was particularly unique in our meta-analysis because it was found to be elevated (compared with normal tissue) in 8/10 studies. In these studies, it was more than double the abundance of that found in normal tissue; thus, kynurenine seems to be not only a recurrent differentiating feature of tumor tissue, but also a potential marker of disease progression in cancer (Chuang et al., 2014; Lim et al., 2017; Heng et al., 2016).

The normal tissue and microenvironment surrounding the tumor play a key role in the development of cancers (Chang et al., 2015; Place et al., 2011; Quail and Joyce, 2013). It is therefore possible that the metabolic demands of a tumor can drive a physiological change in the metabolism of nearby normal tissue. Therefore, we repeated our meta-analysis above, focusing on the association between metabolite levels in the adjacent normal tissue and the pathological grade of the matched tumor. We identified four metabolites whose abundance in normal tissue was associated with tumor grade: putrescine, galactose, and two fatty acids (margarate and stearate) (uncorrected individual study $p < 0.05$ in at least two studies, BH-corrected meta $p < 0.1$). For putrescine, another polyamine, increased levels in normal tissue were associated with higher grade in both kidney and prostate cancers. Notably, 2-difluoromethylornithine is an enzymatic inhibitor of ornithine decarboxylase, an enzyme necessary for the production of both spermidine and putrescine (Casero and Marton, 2007), and there is some evidence that it may be useful in the treatment of a variety of cancers (colorectal and melanoma) as well as a part of combinatorial therapy for brain cancers (Gerner and Meyskens, 2004). Interestingly, clinical associations in normal tissue were sometimes, but not always, mirrored in the analogous tumor tissue (Figure S6A). While many metabolites (e.g., choline) had common patterns of association in kidney tumor and normal tissue (and less frequently in breast and prostate tissues), other metabolites had opposing associations (e.g., spermidine in prostate tumors/normal tissues, Figure 6A).

What is a plausible interpretation of the connection between normal tissue metabolism and aggressive tumor features? One explanation is that the nearby tissue changes physiologically in the course of tumor progression, for example, in response to cytotoxic therapies. Another, more intriguing, possibility is that there may some interaction between tumor cells and the surrounding normal tissue. To evaluate if this may be the case, we again used metabolite data on 343 matched tumor/normal pairs from the same patient and evaluated (study by study) whether metabolite levels in tumor and normal tissue were correlated. Of the 205 metabolites measured in at least 5/8 studies with paired samples, we identified eight metabolites that were recurrently correlated between tumor and normal tissues in at least five studies (Table S1: Pairscores). Of these, a number were metabolites that circulate physiologically through the body (e.g., 1,5-anhydroglucitol, 3-hydroxybutyrate, urea). Thus, our data are suggestive of many correlations between tumor and normal metabolite levels arising because of broad changes to the physiology of the patient. Interestingly, however, alanine was significantly correlated between tumor and normal tissues in the Pancreatic Kamphorst study and this correlation approached statistical significance in a second pancreatic study (Pancreatic Zhang1). This possibly supports the recent report of shuttling of alanine from nearby pancreatic stellate cells to tumor cells (Sousa et al., 2016).

DISCUSSION

In spite of the diversity of genetic alterations across cancers, it is possible that the development of cancer may place common demands on cellular metabolism, irrespective of the tissue of origin. In this study, we offer a glimpse into the common patterns and distinguishing features of metabolic reprogramming across cancers.

Our approach to analyzing a compendium of small-molecule data derived from mass spectrometry has inherent assumptions, which we addressed to the extent possible. For example, we relied on the normalization and standardization procedures undertaken by the original researchers of a study. While all the data in our study were pre-normalized, three studies contained data that were not imputed or standardized; in those cases, we applied a simple imputation/standardization pipeline. We also implemented a bioinformatic pipeline to align metabolites identified using disparate nomenclature, the code repository for which is publicly available on GitHub (https://github.com/dfci/pancanmet_analysis). Finally, to compare data across studies following unknown and potentially various statistical/distributional properties, we relied heavily on non-parametric statistics that make no assumptions on underlying distributions and either (1) used changes in abundance relative to normal tissue as the unit of comparison or (2) non-parametrically associated changes in metabolite levels with clinical data. We expect that this pipeline, through collaboration with others in the metabolomics community, can be expanded and improved to more fully address systematic biases and differences in data acquisition.

Our findings illustrate that cancers of widely different types share some common patterns of changes to metabolite levels. A subset of metabolites, including acyl-carnitines, lactate, kynurenine, and taurine, were recurrently higher in abundance across many tumor types (compared with normal tissue). Comparatively fewer metabolites, including some fatty acids, were recurrently depleted across these same cancer types. These findings echo those from a recent meta-analysis of published cancer metabolomics data (Goveia et al., 2016) (which did not directly analyze the metabolomics data, but aggregated results from individual studies into a meta-analysis). Importantly, the metabolites displaying clear, consistent shifts in abundance across cancers are in the minority. The majority of metabolites, both in central carbon metabolism and in secondary metabolic pathways, had heterogeneous changes relative to normal tissues across tumor types.

We also found that the metabolism of tumors is linked to their clinical presentation and, in particular, to their grade. While it should come as no surprise that tumor metabolism necessarily accommodates and supports the development of aggressive disease and metastasis, it was intriguing to find that some metabolites were indicative of aggressive disease across several cancer types. Our analysis also indicated that the metabolism of adjacent normal tissue, which is physiologically linked to the tumor by physical location and shared vasculature, is remodeled in patients with high-grade tumors. One potential explanation for the association between metabolite levels in tumor/normal tissue and tumor grade is due to the variation in the degree of infiltrating immune or stromal cells between tumors of different grades. Such infiltration could also, for example, enrich or dilute the abundance of metabolites specific to either tumor or immune/stromal cells. Because one study (Tang et al., 2014) used The Cancer Genome Atlas-profiled tissue samples, we were able to recover estimates of stromal and immune infiltration for 23 samples from this study and investigate if infiltration was correlated with metabolite levels in this study. Interestingly, while no metabolites reached statistical significance upon multiple-hypothesis correction, several metabolites of high interest (e.g., NADH) were close to significance (Figure S6). Given the relatively small sample size of this dataset (23 samples), we hypothesize that improved statistical power may indeed yield metabolites indicative of, or enriched in, various components of the tumor microenvironment.

In some tumor types and, in particular, in clear-cell renal cell carcinomas, many metabolites had changes in abundance in normal tissue which were associated with the grade of the patient's tumor. Much remains to be understood of this phenomenon, including in particular the possibility that tumor and normal tissue may be shuttling nutrients between each other in support of tumor cell proliferation (Sousa et al., 2016). Importantly, observing (or failing to observe) a correlation between steady-state metabolite levels in tumor/normal tissues is not sufficient evidence to conclude nutrient shuttling between the two tissues. Instead, the precise mechanism of shuttling could lead to different kinds of correlations between tumor/normal metabolites; for example, a positive correlation may arise if excess production of a metabolite in the normal tissue is shuttled to the tumor passively by blood vessels. Similarly, a negative correlation could arise if a fixed amount of metabolite produced in the normal tissue is actively transported to the tumor, e.g., by overexpression of a nutrient exporter in the normal tissue. Thus, tumor/normal metabolite correlations should be interpreted with care and would make an interesting subject for future investigation.

One outcome of our work is a merged, standardized metabolomics dataset, which can be used by others in the future, with an accompanying website (http://www.sanderlab.org/pancanmet/) allowing readers to explore (1) metabolite data stratified by metadata (e.g., HMDB classifications) and by study, (2) univariate and bivariate changes by metabolite and by study, and (3) clinical associations by metabolite and by study. The analysis in this paper has focused on non-parametric, univariate analysis of changes in metabolite levels between tumor and normal samples. Yet, other analyses that respect the limitations of the data (Figure 2) are possible. One example is a correlation analysis, akin to those proposed in several publications using gene expression data (Reznik and Sander, 2015). It is possible that, by examination of changes in covariation patterns between metabolites between different subsets of samples (e.g., tumor/normal, high-grade/low-grade tumors), the identification of putatively "rewired" metabolic pathways may be possible. Expanding covariation analyses to include other data types (e.g., transcriptomics, proteomics) may yield further insights.

Finally, given the proliferation of clinical genomic sequencing of cancer samples, it seems natural to propose that connections be drawn between genetic alterations and the metabolic landscape of tumors. Several prominent examples now exist of molecular alterations that manifest with, among other things, distinct metabolic phenotypes: *KRAS* mutations in pancreatic adenocarcinomas, which induce macropinocytotic scavenging of extracellular nutrients (Kamphorst et al., 2015), and hotspot *IDH1* and *IDH2* mutations, which produce 2-HG and induce DNA hypermethylation via inhibition of α-ketoglutarate-dependent DNA demethylases (Lu et al., 2012; Xu et al., 2011). Furthermore, recent work has demonstrated that metabolic changes arising from a single common genetic lesion are tissue specific and can vary substantially from one cancer type to the next (Mayers et al., 2016). Thus, it is likely that other mutation-associated metabolic effects, perhaps subtle but nevertheless impacting tumor viability, remain to be uncovered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Imputation and Standardization
  - Metabolite ID Mapping
  - Differential Abundance Tests
  - Pathway Differential Abundance Score
  - Calculation of Metabolic Variation (MAD Score)
  - Clinical Features
  - Association with Tumor Grade
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and two tables and can be found with this article online at https://doi.org/10.1016/j.cels.2017.12.014.

ACKNOWLEDGMENTS

We would like to thank each of the original investigators and their teams who contributed their data to this work and kindly offered us their feedback, including but not limited to James Hsieh (kidney), Stefan Ambs (breast), Jeffrey R. Marks (breast), Max Loda (prostate), Arun Sreekumar (bladder), Josh Rabinowitz (pancreatic), Prakash Chinnaiyan (glioma), Arul Chinnaiyan (prostate), S. Perwez Hussain (pancreatic), and Sham Kakar (ovarian). We also thank Niki Schultz for useful discussions. The work was supported by a Ruth L. Kirschstein National Research Service Award (grant F32 CA192901 to A.L.) and the National Resource for Network Biology (NRNB) from the National Institute of General Medical Sciences (NIGMS) (grant P41 GM103504 to C.S.).

C.J.C. was supported by NIH grant CA125123. E.R., A.A.H., and C.S. were supported by the National Cancer Institute (P30-CA008748). E.R. was supported by the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

AUTHOR CONTRIBUTIONS

E.R., A.L., and C.S. conceived the project. E.R., A.L., B.A.A., E.M.L., K.L., I.O., C.J.C., A.A.H., and C.S. contributed to analysis. E.R., A.L., and C.S. wrote the manuscript. All authors approved the manuscript.

Received: April 21, 2017
Revised: August 21, 2017
Accepted: December 17, 2017
Published: January 24, 2018

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Li et al. (2014); Yoshihara et al. (2013).

REFERENCES

Belladonna, M.L., Puccetti, P., Orabona, C., Fallarino, F., Vacca, C., Volpi, C., Gizzi, S., Pallotta, M.T., Fioretti, M.C., and Grohmann, U. (2007). Immunosuppression via tryptophan catabolism: the role of kynurenine pathway enzymes. Transplantation *84* (1 Suppl), S17–S20.

Casero, R.A., Jr., and Marton, L.J. (2007). Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. Nat. Rev. Drug Discov. *6*, 373–390.

Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011). Pathway Commons, a web resource for bio- logical pathway data. Nucleic Acids Res. *39* (Database issue), D685–D690.

Chang, C.H., Qiu, J., O'Sullivan, D., Buck, M.D., Noguchi, T., Curtis, J.D., Chen, Q., Gindin, M., Gubin, M.M., van der Windt, G.J., et al. (2015). Metabolic competition in the tumor microen- vironment is a driver of cancer progression. Cell *162*, 1229–1241.

Chen, Y., and Guillemin, G.J. (2009). Kynurenine pathway metabolites in humans: disease and healthy States. Int. J. Tryptophan Res. *2*, 1–19.

Chinnaiyan, P., Kensicki, E., Bloom, G., Prabhu, A., Sarcar, B., Kahali, S., Eschrich, S., Qu, X., Forsyth, P., and Gillies, R. (2012). The metabolomic signature of malignant glioma reflects accelerated anabolic metabolism. Cancer Res. *72*, 5878–5888.

Cho, K., Mahieu, N.G., Johnson, S.L., and Patti, G.J. (2014). After the feature presentation: technologies bridging untargeted metabolomics and biology. Curr. Opin. Biotechnol. *28*, 143–148.

Chuang, S.C., Fanidi, A., Ueland, P.M., Relton, C., Midttun, O., Vollset, S.E., Gunter, M.J., Seckl, M.J., Travis, R.C., Wareham, N., et al. (2014). Circulating biomarkers of tryptophan and the kynurenine pathway and lung cancer risk. Cancer Epidemiol. Biomarkers Prev. *23*, 461–468.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. Nat. Genet. *45*, 1127–1133.

Conley, B.A., and Doroshow, J.H. (2014). Molecular analysis for therapy choice: NCI MATCH. Semin. Oncol. *41*, 297–299.

DeBerardinis, R.J., and Chandel, N.S. (2016). Fundamentals of cancer metabolism. Sci. Adv. *2*, e1600200.

Fong, M.Y., McDunn, J., and Kakar, S.S. (2011). Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. PLoS One *6*, e19963.

Gatto, F., Nookaew, I., and Nielsen, J. (2014). Chromosome 3p loss of heterozygosity is associ- ated with a unique metabolic network in clear cell renal carcinoma. Proc. Natl. Acad. Sci. USA *111*, E866–E875.

Gerner, E.W., and Meyskens, F.L., Jr. (2004). Polyamines and cancer: old molecules, new understanding. Nat. Rev. Cancer *4*, 781–792.

Gerosa, L., Haverkorn van Rijsewijk, B.R., Christodoulou, D., Kochanowski, K., Schmidt, T.S., Noor, E., and Sauer, U. (2015). Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. Cell Syst. *1*, 270–282.

Giskeødegård, G.F., Bertilsson, H., Selnæs, K.M., Wright, A.J., Bathen, T.F., Viset, T., Halgunset, J., Angelsen, A., Gribbestad, I.S., and Tessem, M.B. (2013). Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. PLoS One *8*, e62375.

Goveia, J., Pircher, A., Conradi, L.-C., Kalucka, J., Lagani, V., Dewerchin, M., Eelen, G., DeBerardinis, R.J., Wilson, I.D., and Carmeliet, P. (2016). Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities. EMBO Mol. Med. *8*, 1134–1142.

Grabiner, B.C., Nardi, V., Birsoy, K., Possemato, R., Shen, K., Sinha, S., Jordan, A., Beck, A.H., and Sabatini, D.M. (2014). A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. Cancer Discov. *4*, 554–563.

Haider, S., McIntyre, A., van Stiphout, R.G., Winchester, L.M., Wigfield, S., Harris, A.L., and Buffa, F.M. (2016). Genomic alterations underlie a pan-cancer metabolic shift associated with tumour hypoxia. Genome Biol. *17*, 140.

Hakimi, A., Reznik, E., Lee, C.H., Creighton, C.J., Brannon, A.R., Luna, A., Aksoy, B.A., Liu, E.M., Shen, R., Lee, W., et al. (2016). An integrated metabolic atlas of clear cell re- nal cell carcinoma. Cancer Cell *29*, 104–116.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Heng, B., Lim, C.K., Lovejoy, D.B., Bessede, A., Gluch, L., and Guillemin, G.J. (2016). Understanding the role of the kynurenine path- way in human breast cancer immunobiology. Oncotarget *7*, 6506–6520.

Hensley, C.T., Faubert, B., Yuan, Q., Lev-Cohain, N., Jin, E., Kim, J., Jiang, L., Ko, B., Skelton, R., Loudat, L., et al. (2016). Metabolic heterogeneity in human lung tu- mors. Cell *164*, 681–694.

Hu, J., Locasale, J.W., Bielas, J.H., O'Sullivan, J., Sheahan, K., Cantley, L.C., Vander Heiden, M.G., and Vitkup, D. (2013). Heterogeneity of tumor-induced gene expression changes in the human metabolic network. Nat. Biotechnol. *31*, 522–529.

Intlekofer, A.M., Wang, B., Liu, H., Shah, H., Carmona-Fontaine, C., Rustenburg, A.S., Salah, S., Gunner, M.R., Chodera, J.D., Cross, J.R., and Thompson, C.B. (2017). L-2-Hydroxyglutarate production arises from noncanonical enzyme function at acidic pH. Nat. Chem. Biol. *13*, 494–500.

Jalbert, L.E., Elkhaled, A., Phillips, J.J., Neill, E., Williams, A., Crane, J.C., Olson, M.P., Molinaro, A.M., Berger, M.S., Kurhanewicz, J., et al. (2017). Metabolic profiling of IDH mutation and malignant progression in infiltrating glioma. Sci. Rep. *7*, 44792.

Kamphorst, J.J., Nofal, M., Commisso, C., Hackett, S.R., Lu, W., Grabocka, E., Vander Heiden, M.G., Miller, G., Drebin, J.A., Bar-Sagi, D., et al. (2015). Human pancreatic cancer tumors are nu- trient poor and tumor cells actively scavenge extracellular protein. Cancer Res. *75*, 544–553.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a refer- ence resource for gene and protein annotation. Nucleic Acids Res. *44*, D457–D462.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al. (2016). PubChem substance and compound databases. Nucleic Acids Res. *44*, D1202–D1213.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. *42* (Database issue), D1091–D1097.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Li, B., Qiu, B., Lee, D.S., Walton, Z.E., Ochocki, J.D., Mathew, L.K., Mancuso, A., Gade, T.P., Keith, B., Nissim, I., and Simon, M.C. (2014). Fructose-1,6-bisphosphatase opposes renal carcinoma progression. Nature *513*, 251–255.

Lim, C.K., Bilgin, A., Lovejoy, D.B., Tan, V., Bustamante, S., Taylor, B.V., Bessede, A., Brew, B.J., and Guillemin, G.J. (2017). Kynurenine pathway

metabolomics predicts and provides mechanistic insight into multiple sclerosis progression. Sci. Rep. *7*, 41473.

Loughin, T.M. (2004). A systematic comparison of methods for combining p-values from inde- pendent tests. Comput. Stat. Data Anal. *47*, 467–485.

Lu, C., Venneti, S., Akalin, A., Fang, F., Ward, P.S., Dematteo, R.G., Intlekofer, A.M., Chen, C., Ye, J., Hameed, M., et al. (2013). Induction of sarcomas by mutant IDH2. Genes Dev. *27*, 1986–1998.

Lu, C., Ward, P.S., Kapoor, G.S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C.R., Khanin, R., Figueroa, M.E., Melnick, A., et al. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature *483*, 474–478.

Luna, A., Babur, Ö., Aksoy, B.A., Demir, E., and Sander, C. (2016). PaxtoolsR: pathway analysis in R using Pathway Commons. Bioinformatics *32*, 1262–1264.

Mandal, S., Mandal, A., Johansson, H.E., Orjalo, A.V., and Park, M.H. (2013). Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells. Proc. Natl. Acad. Sci. USA *110*, 2169–2174.

Mayers, J.R., Torrence, M.E., Danai, L.V., Papagiannakopoulos, T., Davidson, S.M., Bauer, M.R., Lau, A.N., Ji, B.W., Dixit, P.D., Hosios, A.M., et al. (2016). Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-driven cancers. Science *353*, 1161–1165.

Menzel, M.I., Farrell, E.V., Janich, M.A., Khegai, O., Wiesinger, F., Nekolla, S., Otto, A.M., Haase, A., Schulte, R.F., and Schwaiger, M. (2013). Multimodal assessment of in vivo metabolism with hyperpolarized [1-13C]MR spectroscopy and 18F-FDG PET imaging in hep- atocellular carcinoma tumor-bearing rats. J. Nucl. Med. *54*, 1113–1119.

Mullen, A.R., Wheaton, W.W., Jin, E.S., Chen, P.H., Sullivan, L.B., Cheng, T., Yang, Y., Linehan, W.M., Chandel, N.S., and DeBerardinis, R.J. (2012). Reductive carboxylation supports growth in tumour cells with defective mitochondria. Nature *481*, 385–388.

Nguyen, N.T., Nakahama, T., Le, D.H., Van Son, L., Chu, H.H., and Kishimoto, T. (2014). Aryl hydrocarbon receptor and kynurenine: recent advances in autoimmune disease research. Front. Immunol. *5*, 551.

Nilsson, R., Jain, M., Madhusudhan, N., Sheppard, N.G., Strittmatter, L., Kampf, C., Huang, J., Asplund, A., and Mootha, V.K. (2014). Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. Nat. Commun. *5*, 3128.

Pavlova, N.N., and Thompson, C.B. (2016). The emerging hallmarks of cancer metabolism. Cell Metab. *23*, 27–47.

Place, A.E., Jin Huh, S., and Polyak, K. (2011). The microenvironment in breast cancer progres- sion: biology and implications for treatment. Breast Cancer Res. *13*, 227.

Platten, M., Wick, W., and Van den Eynde, B.J. (2012). Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. Cancer Res. *72*, 5435–5440.

Possemato, R., Marks, K.M., Shaul, Y.D., Pacold, M.E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.K., Jang, H.G., Jha, A.K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature *476*, 346–350.

Priolo, C., Pyne, S., Rose, J., Regan, E.R., Zadra, G., Photopoulos, C., Cacciatore, S., Schultz, D., Scaglia, N., McDunn, J., et al. (2014). AKT1 and MYC induce distinctive metabolic finger- prints in human prostate cancer. Cancer Res. *74*, 7198–7204.

Putluri, N., Shojaie, A., Vasu, V.T., Vareed, S.K., Nalluri, S., Putluri, V., Thangjam, G.S., Panzitt, K., Tallman, C.T., Butler, C., et al. (2011). Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. Cancer Res. *71*, 7376–7386.

Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nat. Med. *19*, 1423–1437.

Reaves, M.L., Young, B.D., Hosios, A.M., Xu, Y.F., and Rabinowitz, J.D. (2013). Pyrimidine homeostasis is accomplished by directed overflow metabolism. Nature *500*, 237–241.

Reznik, E., Mehta, P., and Segrè, D. (2013). Flux imbalance analysis and the sensitivity of cellu- lar growth to changes in metabolite pools. PLoS Comput. Biol. *9*, e1003195.

Reznik, E., Miller, M.L., Şenbabaoğlu, Y., Riaz, N., Sarungbam, J., Tickoo, S.K., Al-Ahmadie, H.A., Lee, W., Seshan, V.E., Hakimi, A.A., and Sander, C. (2016). Mitochondrial DNA copy number variation across human cancers. Elife *5:e10769*.

Reznik, E., and Sander, C. (2015). Extensive decoupling of metabolic genes in cancer. PLoS Comput. Biol. *11*, e1004176.

Reznik, E., Wang, Q., La, K., Schultz, N., and Sander, C. (2017). Mitochondrial respiratory gene expression is suppressed in many cancers. Elife *6:e21592*.

Shim, E.H., Livi, C.B., Rakheja, D., Tan, J., Benson, D., Parekh, V., Kho, E.Y., Ghosh, A.P., Kirkman, R., Velu, S., et al. (2014). L-2-hydroxyglutarate: an epigenetic modifier and putative oncometabolite in renal cancer. Cancer Discov. *4*, 1290–1298.

Singh, K.K., Desouki, M.M., Franklin, R.B., and Costello, L.C. (2006). Mitochondrial aconi- tase and citrate metabolism in malignant and nonmalignant human prostate tissues. Mol. Cancer *5*, 14.

Singh, S., Khan, A.R., and Gupta, A.K. (2012). Role of glutathione in cancer pathophysiology and therapeutic interventions. J. Exp. Ther. Oncol. *9*, 303–316.

Sousa, C.M., Biancur, D.E., Wang, X., Halbrook, C.J., Sherman, M.H., Zhang, L., Kremer, D., Hwang, R.F., Witkiewicz, A.K., Ying, H., et al. (2016). Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. Nature *536*, 479–483.

Sreekumar, A., Poisson, L.M., Rajendiran, T.M., Khan, A.P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R.J., Li, Y., et al. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature *457*, 910–914.

Sullivan, L.B., Gui, D.Y., and Heiden, M.G.V. (2016). Altered metabolite levels in cancer: implications for tumour biology and cancer therapy. Nat. Rev. Cancer *61*, 680–693.

Tang, X., Lin, C.C., Spasojevic, I., Iversen, E.S., Chi, J.T., and Marks, J.R. (2014). A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res. *16*, 415.

Terunuma, A., Putluri, N., Mishra, P., Mathé, E.A., Dorsey, T.H., Yi, M., Wallace, T.A., Issaq, H.J., Zhou, M., Killian, J.K., et al. (2014). MYC-driven accumulation of 2- hydroxyglutarate is associated with breast cancer prognosis. J. Clin. Invest. *124*, 398–412.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the War- burg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., et al. (2007). HMDB: the human metabolome database. Nucleic Acids Res. *35* (Database issue), D521–D526.

Wohlgemuth, G., Haldiya, P.K., Willighagen, E., Kind, T., and Fiehn, O. (2010). The Chemical Translation Service–a web-based tool to improve standardization of metabolomic reports. Bioinformatics *26*, 2647–2648.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive in- hibitor of *a* -ketoglutarate-dependent dioxygenases. Cancer Cell *19*, 17–30.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

Zhang, G., He, P., Tan, H., Budhu, A., Gaedcke, J., Ghadimi, B.M., Ried, T., Yfantis, H.G., Lee, D.H., Maitra, A., et al. (2013). Integration of metabolomics and transcriptomics re- vealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. Clin. Cancer Res. *19*, 4983–4993.

Zhang, Y., Kwok-Shing Ng, P., Kucherlapati, M., Chen, F., Liu, Y., Tsang, Y.H., de Velasco, G., Jeong, K.J., Akbani, R., Hadjipanayis, A., et al. (2017). A Pan-Cancer Proteogenomic atlas of PI3K/AKT/mTOR pathway alterations. Cancer Cell *31*, 820–832.e3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Breast: Pre-normalized metabolomics data collected in tumor samples	Tang et al., 2014	N/A
Breast: Pre-normalized metabolomics data collected in tumor samples	Terunuma et al., 2014	N/A
Glioma: Pre-normalized metabolomics data collected in tumor samples	Chinnaiyan et al., 2012	N/A
Kidney: Pre-normalized metabolomics data collected in tumor samples	Hakimi et al., 2016	N/A
Ovary: Pre-normalized metabolomics data collected in tumor samples	Fong et al., 2011	N/A
Pancreatic: Pre-normalized metabolomics data collected in tumor samples (2-datasets)	Zhang et al., 2013	N/A
Prostate: Pre-normalized metabolomics data collected in tumor samples	Priolo et al., 2014	N/A
Bladder: Pre-normalized metabolomics data collected in tumor samples	Putluri et al., 2011	N/A
Pancreatic: Pre-normalized metabolomics data collected in tumor samples	Kamphorst et al., 2015	N/A
Prostate: Pre-normalized metabolomics data collected in tumor samples	Sreekumar et al., 2009	N/A
Human Metabolome Database: Metabolite classifications	http://www.hmdb.ca	Version 3.0
DrugBank: Drug compound classifications	http://drugbank.ca	Version 5.0.3
Software and Algorithms		
Data standardization, nomenclature alignment, and analysis pipeline	This paper, https://github.com/dfci/pancanmet_analysis	N/A
Chemical Translation Service	http://cts.fiehnlab.ucdavis.edu/	N/A
Pathway Commons Database accessed through paxtoolsr R package	https://www.bioconductor.org/packages/release/bioc/html/paxtoolsr.html	Pathway Commons: Version 8
Data exploration site	This paper, http://www.sanderlab.org/pancanmet	N/A
KEGG Pathways accessed through KEGGREST	https://www.bioconductor.org/packages/release/bioc/html/KEGGREST.html	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ed Reznik (reznike@mskcc.org).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Imputation and Standardization

All data used in this study was pre-normalized (*i.e.* no raw mass spectra were analyzed). For data which was already imputed and standardized, we used the data as reported by the original authors. This includes both breast studies (Terunuma et al., 2014; Tang et al., 2014), glioma (Chinnaiyan et al., 2012), kidney (Hakimi et al., 2016), ovary, (Fong et al., 2011), 2 of the pancreatic studies (Zhang et al., 2013), and one of the prostate studies (Priolo et al., 2014). For the bladder study (Putluri et al., 2011), data was obtained normalized after log transformation, with no further imputation required. For this study, normalized data was simply re-exponentiated. For two studies, PAAD (Kamphorst et al., 2015) and PRAD (Sreekumar et al., 2009), 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; additional statistics on imputed metabolites are shown in Figure S1. Then, all measurements of the metabolite were divided by the median abundance of that metabolite. The entire data import pipeline can be replicated using the code provided in the GitHub repository.

Because all subsequent statistical analyses were non-parametric, no assumptions were made on the underlying distribution of the data.

Metabolite ID Mapping

In addition to data standardization, a bioinformatic challenge in the current work 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 and database identifiers. To address this issue, we used the Chemical Translation Service (Wohlgemuth et al., 2010) to retrieve synonymous identifiers of each metabolite. These identifiers were then used to assemble a meta-dataset of all metabolomics data, thus "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 on GitHub (https://github.com/dfci/pancanmet_analysis).

These mapped identifiers were used in conjunction with the paxtoolsr R Bioconductor package (https://www.bioconductor.org/packages/release/bioc/html/paxtoolsr.html) to map identifiers onto the Pathway Commons Version 8 database, which includes both KEGG and DrugBank interactions to identify drug targetable pathways.

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. As with other analysis in the manuscript, statistical significance was assessed using non-parametric Mann-Whitney U-tests. P-values were adjusted for multiple-hypothesis testing using the Benjamini-Hochberg procedure.

Pathway Differential Abundance Score

The differential abundance (DA) score for a pathway was defined as DA = (I-D)/S, where *I* is the number of measured metabolites in a pathway that 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.

Calculation of Metabolic Variation (MAD Score)

Metabolic variation analysis was completed for each study with paired tumor/adjacent-normal samples from the same patient. For each study, we restricted analysis to such paired data and partitioned the metabolomics data into measurements from tumor and normal tissue samples. Then, for each metabolite m with measurements from n samples, $X = X_1, X_2, \ldots, X_n$, a normalized median absolute deviation (MAD) was calculated according to the formula

$$\mathrm{MAD}_X = \mathrm{median}(|\log_2(X_i/\mathrm{median}(X))|)$$

The MAD was calculated for each metabolite m separately for tumor (MAD^T) and normal (MAD^N) samples. The ratio of these two MAD scores for each metabolite m, MAD^r, was then calculated. The distribution of MAD^r scores over all metabolites indicates whether tumor samples show significantly different levels of variation than normal tissues (*i.e.* if the distribution of MAD^r scores is skewed to be significantly less than 0, this indicates that tumor samples exhibit less variation than normal samples). The results of the MAD analysis were invariant to the exclusion of highly-imputed metabolites.

Clinical Features

In general, clinical data was imported from the datasets reported by original investigators. For prostate tumors specifically (Prostate Priolo and Prostate Sreekumar studies), we converted Gleason score to tumor grade as follows (3+3 = Grade 1; 3+4,4+3 = Grade 2; 4+4,5+4 = Grade 3).

Association with Tumor Grade

Different statistical tests were applied to identify correlation between metabolites and clinical features depending on the level of censoring (i.e. imputation) in the data. If there was 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 was more than 20% but less than 80% censored data, an exact log-rank trend test was used with left censored data. Both tests assume a metabolite should increase/decrease monotonically with stage/grade. If there is more than 80% censoring, no test was used, and the metabolite was ignored. To summarize the tests across cohorts Fisher's method was used to combine p-values. Note that resulting combined p-value is parametric. Since, we were interested in finding associations of a common sign (*e.g.*

consistently positive correlation across multiple studies), one-sided p-values were calculated first, aggregated into a combined p-value using Fisher's method (Loughin, 2004), and then transformed to a two-sided combined p-value.

DATA AND SOFTWARE AVAILABILITY

All code associated with the analysis is available at https://github.com/dfci/pancanmet_analysis/. This code replicates both the merging of metabolomics data (in the import/ folder) and the analysis (in the analysis/) folder.

ADDITIONAL RESOURCES

Interactive visualization of the data and associated analysis is available at https://www.sanderlab.org/pancanmet.