A Multimodal Atlas of Tumor Metabolism Reveals the Architecture of Gene-Metabolite Co-regulation

Elisa Benedetti¹,², Eric Minwei Liu³*, Cerise Tang¹,³*, Fengshen Kuo⁴, Mustafa Buyukozkan¹,², Tricia Park³, Jinsung Park⁵, Fabian Correa⁵, A Ari Hakimi⁴, Andrew Intlekofer⁵, Jan Krumsiek¹,²,⁶,#, Ed Reznik¹,³,⁷,#

¹Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY
²Institute of Computational Biomedicine, Weill Cornell Medicine, New York, NY
³Computational Oncology Service, Memorial Sloan Kettering Cancer Center, New York, NY
⁴Department of Surgery, Urology Service, Memorial Sloan Kettering Cancer Center, New York, NY
⁵Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY
⁶Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY
⁷Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY

* Equal contribution

Correspondence to:
Jan Krumsiek (jak2043@med.cornell.edu)
Ed Reznik (reznike@mskcc.org)
**Abstract**

Tumor metabolism is controlled by coordinated changes in metabolite abundance and gene expression, but simultaneous quantification of metabolites and transcripts in primary tissue is rare. To overcome this limitation and study gene-metabolite coregulation in cancer, we assembled the Cancer Atlas of Metabolic Profiles (cAMP) of metabolomic and transcriptomic data from 988 tumor/normal specimens spanning 11 cancer types. Meta-analysis of the cAMP revealed two classes of Gene-Metabolite Interactions (GMIs) that transcended cancer types. The first corresponded to a small number of gene-metabolite pairs engaged in direct enzyme-substrate interactions, identifying putative metabolite-pool-size-controlling genes. A second class of GMIs represented a small number of hub metabolites, including quinolinate and NAD+, which correlated to many genes specifically expressed on immune cell populations. These results provide evidence that gene-metabolite coregulation in human tissue arises, in part, from both mechanistic interactions between genes/metabolites, and from metabolic remodeling in specific immune microenvironments.
Introduction

Coordinated changes of genetically encoded metabolic enzymes and transporters, and the metabolites they act on, underpin diverse cancer-associated phenomena, including tumorigenesis\(^1\), pluripotency\(^2,3\), the onset of drug resistance\(^4-6\), and the modulation of immune responses\(^7-11\). However, in spite of the high value of joint profiling of metabolites and gene expression/protein levels, previous large-scale studies of tumor metabolism have overwhelmingly focused on the analysis of gene expression data\(^12\). Conversely, the few instances of multimodal metabolomic and transcriptomic profiling of human tumor specimens have largely been performed in disparate, unrelated studies by a multitude of research teams\(^13-22\). Integration of metabolomic datasets produced in different patient cohorts is challenging due to technical batch effects and the semi-quantitative nature of untargeted metabolomic data (reported in arbitrary units of relative abundance). Thus, both the scarcity of multimodal metabolomic/transcriptomic data from tissue specimens and the challenges of harmonizing available datasets fundamentally impede the discovery of recurrent, coordinated changes in metabolic gene expression and metabolite abundance across cancers.

Tumors from diverse cancer types differ in their cell-type composition, vascularization, and other factors ultimately influencing metabolism. Yet, they share a convergent set of metabolic alterations\(^23-27\). For example, several meta-analyses of the tumor metabolic transcriptome have identified recurrent up-regulation of genes in one-carbon metabolism and oxidative phosphorylation across cancer types\(^26,28,29\). Analogously, meta-analyses of metabolomics data have demonstrated that numerous central carbon metabolites (e.g., lactate) and effector metabolites (e.g., kynurenine) are at higher abundance in tumor tissue compared to normal tissue across many cancer types\(^23,30\). These studies have illustrated the power of meta-analysis for distilling highly recurrent metabolic phenotypes from heterogeneous data but have left unresolved the question of how transcription of metabolic genes and metabolite abundance are coordinated and ultimately shape tumor physiology.

To systematically investigate gene-metabolite coregulation in cancer, we assembled, harmonized, and integratively analyzed metabolomics and transcriptomics profiles from 988 primary tumor and matched adjacent normal tissue collected in 15 independent studies covering 11 cancer types. The preprocessed and harmonized data constitute the Cancer Atlas of Metabolic Profiles (cAMP), representing what is to our knowledge the largest harmonized dataset of multimodal metabolomic and transcriptomic data on primary tumor specimens. The cAMP is publicly available for download and can be interactively explored at https://rezniklab.shinyapps.io/cAMP-shiny-app/. Leveraging the diversity of diseases in our dataset, we designed a concordance-based statistical meta-analysis approach to discover instances of Gene-Metabolite Interactions (GMIs) that transcended cancer type. This revealed two distinct classes of GMIs: First, we identified a small number of strong interactions between enzymes and metabolites involved in the same or subsequent reactions (“proximal” GMIs), suggesting that these enzymes are the primary determinants of their respective metabolite pool sizes. A second group of GMIs consisted of a small number of metabolites broadly correlated to large numbers of genes. Interestingly, this second class of GMIs were enriched for genes specifically expressed in immune cells, and for metabolites related to nicotinamide adenine
dinucleotide (NAD+), a pleiotropic metabolite which acts both as a central cofactor in metabolism\(^{31}\) and as a signaling molecule influencing cell identity\(^{32}\). Taken together, these findings suggest that gene-metabolite co-regulation in tumors is controlled, in part, by two complementary phenomena: the expression of enzymes with strong control over metabolite pool size, and the presence of specific cell populations in the tumor microenvironment with characteristic metabolomic profiles.

**Results**

**A Cancer Atlas of Metabolic Profiles (cAMP) reveals the architecture of gene-metabolite coregulation across cancers**

Since metabolomic profiling has so far been excluded from large multi-modal tumor profiling projects (e.g., in the TCGA\(^{33}\)), there is no unified resource of metabolomic/transcriptomic data in the cancer research field. However, several groups have independently produced and released matched metabolomic/transcriptomic data in diverse cancer types\(^{13–22}\). We combined these datasets with several in-house studies to create a comprehensive collection of 988 samples (764 tumor sample and 224 adjacent normal samples) across 11 different cancer types, covering 15 datasets, which we called the Cancer Atlas of Metabolic Profiles (cAMP) (Table 1 and Figure 1A). The overall collection includes a total of more than 40,000 unique transcripts and almost 2,500 unique metabolites. To maximize comparability across these heterogeneous studies, we applied a unified workflow to process RNA expression from microarray and RNA sequencing data, harmonize metabolite names and annotations, and standardize data normalization and preprocessing (see Methods). The cAMP represents an unprecedented resource to investigate the coregulation of metabolite levels and gene expression at scale across diverse lineages of human cancers and normal tissues.

We interrogated the cAMP dataset for recurrent covariation between genes and metabolites across datasets and cancers. Such gene-metabolite coregulation could emerge via numerous mechanisms including direct metabolic interactions, e.g., via expression changes of a rate-limiting enzyme, or by the accumulation/depletion of metabolites as part of a broader phenotype, such as a cytotoxic immune response\(^{34–36}\). While each cancer type is likely to demonstrate its own unique pattern of transcriptomic and metabolomic changes, the cAMP enables discovery of metabolomic/transcriptomic covariation that transcends diseases. To identify cancer-type-agnostic metabolite/transcript covariation across cAMP datasets in a statistically principled manner, we developed a concordance-based meta-analysis approach (Figure 1B, Methods). Concordance is a non-parametric measure of correlation, similar to Kendall’s Tau, which enables the identification of consistently positive or negative gene-metabolite associations across datasets\(^{37}\). We focused on the 276 metabolites that were quantified in more than half of the tumor datasets (at least 8 of 15 tumor datasets) and the 16,082 genes that were quantified in all 15 studies. Out of all possible gene-metabolite pairs (276 metabolites \(\times\) 16,082 genes = 4,438,632 pairs), a total of 22,619 pairs (0.51\%) were significantly correlated after multiple testing correction at 0.01 FDR (Figure 2A). This included 269 metabolites (~97\%) and 7,987 genes (~50\%) participating in at least one significant association (Table S1), which we refer to as Gene-Metabolite Interactions (GMIs).

A small number of strong GMIs correspond to direct enzyme-substrate interactions
Among the statistically significant GMIs (Figure 2A), we noted that the strongest positively correlated GMI (IDO1-kynurenine, adjusted p-value: 4.82x10^{-35}) and the two strongest negatively correlated GMIs, GDA-guanine (adjusted p-value: 2.31x10^{-20}) and CD38-nicotinamide ribonucleotide (NMN) (adjusted p-value: 3.90x10^{-16}) corresponded to “proximal” metabolic interactions, which are interactions between an enzyme and its direct or nearly direct substrate/product. For example, IDO1 catalyzes the catabolism of tryptophan to N-formylkynurenine, which is subsequently metabolized to kynurenine, and both CD38 and GDA directly degrade the metabolites guanine and NMN, respectively. We confirmed that statistical significance for these 3 GMIs was likely driven by several cancer types rather than a single dataset with very strong associations (Figure S1).

The direct biochemical relationship between the three GMIs above raised the possibility that functional proximity between enzymes and their substrates/products might underlie a large fraction of GMIs. To test this, we systematically computed the distance between all gene-metabolite pairs using the highly curated Human-1 metabolic network model (see Methods). Although statistically significant GMIs were enriched for proximal interactions relative to non-significant gene-metabolite pairs (odds ratio: 1.42, Fisher’s exact test p-value: 3.09 x10^{-15}), proximal interactions themselves constituted only a small fraction of the total ensemble of statistically significant GMIs (2.5%, 565/22,619) (Figure 2B). Thus, while several of the strongest GMIs arose from proximal interactions, gene-metabolite proximity was a weak determinant of the full GMI landscape.

To further investigate the above observations about the strength of specific proximal GMIs and the overall relationship between metabolic proximity and GMIs, we investigated the relative strength of different GMIs affecting a common metabolite. We focused on the 22 metabolites whose strongest GMI was proximal, covering diverse molecules involved in nucleotide metabolism (guanine, cytidine), cofactor metabolism (NAD+), redox metabolism (cystine, GSSG), and other pathways (Figure 2C). Interestingly, we found that for 8 out of 22 metabolites (kynurenine, guanine, NMN, GSSG, tryptophan, GPC, cystine, and cytidine), a large gap existed between the most significant GMI and the second-highest correlating transcript. This gap suggested that the pool size of these metabolites was strongly controlled, in a lineage-agnostic manner, by a single, dominant gene. Consequently, we hypothesized that targeted genetic knockdown of dominant GMIs for these 8 metabolites would have a higher likelihood of producing significant changes in pool size than for other metabolites with multiple, comparably strong GMIs, each of which might control the pool size of the metabolite (Figure 2C, bold).

We sought to functionally validate a subset of these predicted, metabolite pool-controlling genes. First, we investigated the association between IDO1 and two metabolites, kynurenine and tryptophan. IDO1 converts tryptophan to N-formylkynurenine, which is subsequently catabolized to kynurenine by AFMID. We observed that kynurenine levels were strongly associated with IDO1 (Figure 2D) but not AFMID (which acts directly to produce kynurenine, Figure 2E) expression across the cAMP, consistent with previously published results from the Cancer Cell Line Encyclopedia data. To experimentally test the hypothesis that disruption of IDO1 impacts both tryptophan and kynurenine pool sizes, we used CRISPR-Cas9 mediated knockdown with sgRNAs
targeted against IDO1 human colorectal carcinoma HCT116 cells. These experiments corroborated earlier data indicating that knockdown of IDO1 depleted kynurenine pools (Figure 2F and Figure 2G). However, while the association between IDO1 and kynurenine has been widely described in the literature, our analysis indicated that IDO1 is also expected to be pool-determining for tryptophan, an amino acid involved in numerous other reactions in the cell, most obviously the synthesis of proteins (Figure 2H). In another experimental validation, we observed that knockdown of IDO1 was sufficient to increase tryptophan levels, indicating that the pool size of this highly connected proteinogenic amino acid could be perturbed in part through disruption of IDO1 activity (Figure 2I). Second, in support of a proximal GMI between GGT1-GSSG (Figure 2J), we reanalyzed existing metabolomic data from a functional knockdown of GGT1 vs. control in human embryonic kidney HEK293T cells. This data confirmed that knockdown of GGT1 was associated with an increase in GSSG levels with respect to mock control (p-value = 2.90x10^{-2}), suggesting that GGT1 is a pool-determining consumer of GSSG (Figure 2K). Taken together, these data demonstrate that lineage-transcending GMIs discovered through pathway-based analysis of the cAMP represents examples of genes exerting strong control over metabolite pool sizes.

**Metabolic and transcriptomic changes are not coregulated at a pathway level**

Despite the interesting findings related to proximal GMIs, the vast majority of GMIs in our datasets did not represent proximal pathway reactions. In total, only 2.5% (565/22,619) of GMIs were proximal, whereas 97.5% of statistically significant GMIs represented distant, non-proximal interactions beyond obvious enzyme-substrate metabolic relationships (Figure 2B). One possible implication of such non-local covariation is that genes and metabolites in the same metabolic pathway would show asynchronous and uncorrelated changes across different groups of samples, such as tumors and normal tissues. To investigate this hypothesis, we studied the consistency of transcriptional and metabolic differences in tumor versus adjacent normal tissue across cancer types. To this end, we performed differential analysis of metabolite and transcript levels between tumor and normal tissues in the 7 cAMP datasets where both tissues were available (Table 1) and aggregated the results into 85 KEGG metabolic pathways. Out of these, we considered 63 pathways with at least one metabolite or gene measured in at least 5 of the 7 cAMP datasets (Table S2).

For each KEGG pathway, we evaluated (using a differential abundance “DA” score, see Methods) whether metabolites and transcripts showed synchronous accumulation or depletion patterns in tumors relative to normal tissues. Nevertheless, pathways were biased towards asynchronous changes (276/441, 63%), where increases in metabolite levels coincided with decreases in transcript levels, and vice versa (Figure 3A). Only one pathway (Histidine metabolism) demonstrated fully synchronous changes in all datasets, whereas a few others demonstrated uniformly asynchronous changes (e.g., Primary Bile Acid Biosynthesis). We also assessed whether there was a correlation between the extent of metabolomic versus transcriptomic disruption regardless of the direction (using a differential fraction “DF” score, see Methods). A minority of pathways (9/63) demonstrated significant associations (nominal p-value < 0.05) between RNA and metabolite DF scores (Figure 3B, see Figure 3C as an example). Interestingly, these 9 pathways belonged to just two KEGG pathway classes (Table S3): Amino
acid metabolism and Carbohydrate metabolism. Enrichment analysis indicated that the class of carbohydrate metabolism pathway (out of 8 total) was significantly over-represented relative to the others (Fisher’s exact test p-value: 2.36x10^{-2}). Thus, the majority of pathways showed no evidence of a correlation between metabolomic and transcriptomic disruption, emphasizing the implications of predominantly distally-acting GMIs and prompting the question of which biological phenomena produce these distal GMIs.

**Hub metabolites are enriched for associations with immune genes**

Having established that the majority of GMIs do not represent biochemically proximal interactions, we adopted a broader approach to identify the driving factors of metabolite-transcript correlations. First, we investigated the distribution of GMIs across metabolites and genes, observing that GMIs were strongly concentrated in a small number of metabolites (**Figure 4A**). The top three metabolites, quinolinate, nicotinamide mononucleotide (NMN), and 5'-methylthioadenosine (MTA) alone contributed to 17% (3,823/22,619) of all GMIs in our analysis, and the top ten covered 35% of the GMIs (8,048/22,619), far higher than the fraction covered by the top 10 genes (**Figure 4B**). That is, a small number of metabolites participated in an exceptionally high number of GMIs, acting as “hubs” for strong co-variation with gene expression. Interestingly, hub metabolites concentrated in certain metabolic pathways: Among the top 10 most correlated metabolites, we found several constituents of the NAD+ biosynthesis pathway (quinolinate, NMN, NAD+) and nucleotide metabolism (thymine, uracil, adenine).

To determine whether the genes correlated with a particular metabolite were enriched for specific cellular functions, we performed an unsupervised pathway enrichment analysis. For each metabolite with at least one GMI, we investigated whether specific pathways and processes were over-represented in the transcripts correlated with that metabolite. Overall, the analysis used 146 KEGG pathways (**Table S4**), covering the 85 metabolic processes in **Figure 3** as well as cellular processes (e.g., cell growth and death), signaling pathways, genetic processing pathways (e.g., transcription and translation), and organismal systems pathways (e.g., immune, endocrine and sensory system). 32 unique pathways were over-represented across 40 metabolites (adjusted p-value < 0.01, **Figure 4C**). Interestingly, only two of those 32 pathways represented metabolic processes: oxidative phosphorylation, for which the top 3 most associated metabolites were adenine, NMN, and FAD, and the TCA cycle, also associated with FAD. The remaining top-ranked pathways were exclusively non-metabolic. Notably, three metabolite hubs in particular, quinolinate, NMN, and NAD+, showed broad enrichment for immune-related cellular processes, including chemokine signaling, as well as B-cell and T-cell receptor signaling pathways. This indicates that these metabolites, which are involved in the NAD+ biosynthetic pathway, are correlated with the expression of a wide array of genes associated with the immune response.

**NAD+-related metabolites are associated with immune cell infiltration across cancers**

Human tumor tissues are heterogeneous compositions of various cell populations, including tumor cells, immune cells, and stromal cells. Bulk and single cell profiling technologies have revealed that a large subset of genes are exclusively expressed in immune cells or non-immune cell subpopulations (e.g., tumor cells). We reason that the correlation between metabolites in the NAD+ biosynthesis pathway and immune-related genes (**Figure 4C**) could therefore arise if
NAD+-related metabolites were at a characteristically higher or lower abundance in immune cells relative to non-immune cells. One implication of this hypothesis is that, while each cancer type might demonstrate its own unique metabolomic changes associated with immune infiltration, NAD+-related metabolites should be at elevated concentrations in highly immune-infiltrated tumors relative to lowly-infiltrated tumors across many different cancer types.

To determine if NAD+-related metabolites were associated with immune infiltration across tumors, we used single-sample gene set enrichment (ssGSEA) to compute a previously validated 141-gene RNA signature of overall immune cell infiltration (“ImmuneScore”) directly from bulk RNA sequencing data, and identified the individual metabolites correlated with this immune phenotype. Concordance between metabolite levels and the ImmuneScore signature was assessed across all samples in each cAMP dataset (Figure 5A). In general, covariation between specific metabolite pools and ImmuneScore expression was cancer-type specific. For example, of the 12 metabolites significantly associated with ImmuneScore in intrahepatic cholangiocarcinoma (ICC) and hepatocellular carcinoma (HCC) (representing the top 2 cancer types with the highest number of metabolites significantly associated to ImmuneScore), only 3 metabolites were consistently associated with ImmuneScore in both datasets (Figure 5B, Table S5). Notably, we did not observe a significant correlation (Spearman’s rank correlation p-value = 0.69) between the percentage of metabolites significantly associated with ImmuneScore and expression of the ImmuneScore signature itself (shown in Figure 5A, central panel).

While metabolite concordance with ImmuneScore was generally heterogeneous across cAMP datasets, we observed that several NAD+-related metabolites, including quinolinate and NMN, were recurrently associated with high or low levels of immune infiltration in numerous disease contexts (Figure 5A). To systematically identify such lineage-agnostic metabolomic correlates of immune infiltration, we again applied concordance meta-analysis, identifying 23 metabolites significantly associated with ImmuneScore across datasets (adjusted p-value< 0.05, Figure 5C), with quinolinate and NMN being the strongest hits (Figure 5D). Interestingly, 4/14 significantly associated metabolites (quinolinate, NAD+, NMN, and kynurenine) were members of the NAD+ biosynthesis pathway. Consistent with this, we identified NAD+ metabolism as the sole pathway whose metabolites demonstrated significantly stronger association with immune infiltration than all other pathways (adjusted p-value: 0.04, Figure 5E).

The above analysis suggested that NAD+-related metabolites displayed a differential abundance profile in immune cells relative to non-immune cells, and that this effect produced a consequent accumulation of NAD+-related metabolites in immune-infiltrated tumors. Some support for this hypothesis can be found in previously published immunohistochemical data indicating that the abundance of quinolinate increases dramatically in diverse immune cell populations in response to Toll-like receptor 4 (TLR4) ligands such as lipopolysaccharide (LPS). To provide more evidence for this hypothesis, we compared our findings to a recently published study of metabolomic profiles of purified CD45- tumor cells and CD45+(CD8+ and CD4+) T cells from ovarian cancer tumors. In our data, NAD+ was negatively correlated with ImmuneScore in ovarian cancer, suggesting that it was at lower abundance in immune cells relative to non-immune cells (Figure 5F). Consistent with this, NAD+ was at significantly lower abundance in CD45+ T
cells than CD45- tumor cells in the dataset of purified cell populations (NAD+ Log₂FC = -1.22, p value = 1.7x10^-4) (Figure 5F). Together, these analyses suggest that the pool sizes of NAD+-related metabolites are at characteristically different abundance in immune cells relative to other cell types, and that this effect ultimately drives the association of quinolinate and other NAD+-related metabolites with immune infiltration in bulk tumor data.

**Kynurenine and histamine accumulate in the presence of specific immune cell populations**

Whereas in the previous analysis we investigated the association of metabolite levels with overall immune infiltration, we next turned to investigating the association of metabolite levels and specific immune cell populations (e.g., T cells, macrophages, and numerous other cell types), each with unique transcriptional phenotypes and immunologic functions. To investigate how these diverse immune cell populations contributed to the observed GMIs, we estimated the abundance of 23 immune cell types from bulk transcriptomics profiles using ssGSEA as shown previously, and computed their association to metabolite levels across cancers. We again focused on lineage-agnostic relationships by performing a concordance meta-analysis to calculate associations between metabolite levels and immune cell signatures across tumors from all cancer types. In total, 7.3% of all metabolite-signature pairs (466 out of 6,348 pairs) demonstrated statistically significant associations (adjusted p-value < 0.05, Figure 6A). Among these, quinolinate was positively associated with almost all immune cell populations (17/23), consistent with prior immunohistochemical data and suggesting that it accumulates in a variety of immune cell types in a cancer-type-agnostic manner.

Aside from quinolinate, the two strongest associations between immune cell type signatures and metabolite levels were two comparatively rare cell populations, namely mast cells-histamine and activated dendritic cells (aDCs)-kynurenine (Figure 6B). Mast cells are a myeloid cell population which, when stimulated, mediate the inflammatory process by synthesizing histamine from histidine using the enzyme HDC. We found both histamine and its related metabolite 1-methylhistamine were the two metabolites most significantly associated with the presence of mast cells (Figure 6C and Figure 6D), and that the association between histamine and mast cell abundance was likely driven by diverse cancer types in the cAMP (Figure S2). Histamine levels themselves were strongly associated with the expression of HDC across the cAMP (Figure 6E). Interestingly, the median variation of histamine across the tumor datasets in the cAMP was ~740-fold, implying that fluctuations in the abundance of an otherwise rare cell population were sufficient to produce large-magnitude changes in the abundance of a histamine in the bulk tumor.

In contrast to mast cells and their physiological role in producing histamine, dendritic cells are not known to be dedicated sources of kynurenine in the microenvironment, although single cell data indicates that IDO1 expression is strongly elevated in dendritic cells relative to other cell types (Figure S3). While the aDC signature contains IDO1 (and IDO1 participates in a strong GMI with kynurenine, Figure 2), an aDC signature without IDO1 preserved strong positive concordance with kynurenine (concordance = 0.59, p-value = 2.3x10^-5), confirming that this association was not solely dependent on IDO1 itself (Figure 6F). IDO1 and kynurenine corresponded to the strongest GMI in the cAMP, raising two opposing hypotheses on the mechanism underlying the
primary source of kynurenine in the tumor microenvironment: either bulk tumor expression of IDO1 is primarily determined by high dendritic-cell-specific expression, or alternatively, it is driven by a comparatively low expression in far more abundant non-dendritic-cell populations. Resolving this ambiguity would require single cell measurements of metabolite levels which have recently become technically feasible\(^5^3\). Together, the data above suggests that the presence of individual and comparatively rare cell populations is associated with shifts in the abundance of immunomodulatory metabolites in the tumor microenvironment.

**Discussion**

Metabolism is jointly controlled by genetically-encoded enzymes and small molecule metabolites. To study the interactions between genes and metabolites at scale, we assembled and harmonized a database of metabolomic and transcriptomic data from ~1,000 tumor and normal samples which we refer to as the Cancer Atlas of Metabolic Profiles (cAMP). Although large-scale multimodal measurements of metabolism have previously been produced in bacteria\(^5^4\) and yeast\(^5^5\), a comparable resource for human cancers had been missing. The cAMP is among the first such resources and represents a significant public database for studying the metabolism of complex human tissues and cancers. Our analysis of the cAMP demonstrates that large-scale studies of multimodal metabolic data can reveal fundamental principles of metabolic regulation both at the scale of individual metabolic reactions (Figure 2) as well as complex human tissues (Figure 5 and Figure 6).

Our detailed statistical analysis of the cAMP revealed a multitude of gene-metabolite associations that transcended the tissue of origin. We focused on two specific types of gene-metabolite co-regulation. The first, induced by functional proximity, identified metabolic genes whose activation or inhibition is likely to have a significant effect on the direct substrate or product of the respective reaction (Figure 2). While metabolite pools are likely controlled by a large number of genes and other factors, our data-driven analysis identified a small subset of metabolites whose pool size was strongly associated to a single, proximal gene across tissue lineages. The GMIs identified in Figure 2B offer a data-informed, rational approach for modulating the pool sizes of their metabolite constituents. While each of the highlighted metabolites in Figure 2B may participate in numerous metabolic reactions, data from the cAMP specifically nominates single genes (e.g., GGT1 for GSSG, and IDO1 for tryptophan and kynurenine) as those targets whose perturbation is most likely to disrupt the corresponding pool size.

The second broad form of gene-metabolite coregulation, likely induced by cell-type-specific physiology, corresponded to metabolites whose abundance was associated with the presence of specific immune cells in the microenvironment. This class of GMIs was characterized by a small number of metabolites (enriched for NAD+-related molecules) that were significantly correlated with a large number of genes, and in particular immune-related genes. This apparent association between the abundance of specific immune cells in a tissue specimen and the levels of numerous NAD+-associated metabolites suggests that immune cells have evolved mechanisms to maintain the concentrations of these metabolites at characteristically different levels relative to other cell types. Because NAD+ is both the central mediator of redox poise in the cell and a cofactor for
numerous metabolic and non-metabolic reactions, understanding the mechanisms by which immune cells achieve differential abundance of NAD+-related metabolites, and the selective pressure to do so, can provide insights into the metabolic phenotypes underlying both cancer and other diseases involving dysfunctional immune responses\textsuperscript{56}. Given the growing use of spatial metabolomics technologies that can measure the metabolomic profiles of individual cells in heterogenous tissue slices\textsuperscript{57–60}, the full extent of cell-type-specific metabolomic alterations seems now in experimental reach.

Discoveries about the function of cancer genes have emerged from a combination of untargeted, population-scale genomic surveys\textsuperscript{61} and mechanistic experiments in specific disease and genetic backgrounds\textsuperscript{62}. Combining these approaches has proven transformative for the discovery of recurrent and large-effect-size alterations and prompted their characterization in model systems of human disease. In contrast, the field of cancer metabolism has primarily been driven by bottom-up experiments with modest support from large-scale (largely genetic or transcriptomic) datasets\textsuperscript{63–65}. The cAMP is a counterpoint to these efforts: By assembling and harmonizing in one database metabolomic and transcriptomic data from diverse diseases, the cAMP represents a unique opportunity for \textit{de novo} discovery of translationally relevant metabolic phenotypes in cancer. Expanding the scope of the cAMP to include other forms of data, including but not limited to genomic sequencing, epigenetic profiling, and proteomic measurements, holds the potential to reveal entirely new and highly recurrent metabolic phenomena in cancer.
Methods

Collection of Cancer Atlas of Metabolic Profiles (cAMP)

We combined 12 published datasets with 3 additional in-house datasets that profiled metabolite and gene expression from the same samples to create a comprehensive collection of 988 samples (764 tumor sample and 264 adjacent normal samples) across 11 different cancer types, covering 15 datasets, which we called the Cancer Atlas of Metabolic Profiles (cAMP). Details and references associated with these studies are provided in Table 1 and Figure 1A. Data are available for download under DOI: 10.5281/zenodo.7150252.

Gene Expression Data Processing Pipeline

Six of the cAMP datasets included gene expression data captured by the Affymetrix platform (GSE28735, GSE37751, GSE26193, GSE62452, GSE76297, and Cornell_PROSTATE). For these datasets, CEL files were downloaded from GEO or from their source within our respective institutions. Then, we applied the Robust Multichip Average (RMA) algorithm for background subtraction, quantile normalization, and summarization (via median-polish) by using the “rma” function implemented in the R oligo package (version 1.40.2). Each dataset's rma-normalized expression matrix was then used for downstream analysis. In the COAD dataset, gene expression data was captured by an Agilent custom array, and we downloaded the gene expression matrix deposited at the GEO repository (GSE89076) for further downstream analysis.

For cAMP datasets with RNA sequencing data, RNA-seq reads were aligned against human genome assembly hg19 by STAR 2-pass alignment (version 2.5.3a). QC metrics, such as general sequencing statistics, gene feature, and body coverage, were then calculated based on the alignment result through RSeQC (version 2.6.4). RNA-seq gene level count values were computed by using the R package GenomicAlignments (version 1.14.2) over aligned reads with UCSC KnownGene with hg19 as the base gene model. The Union counting mode was used and only mapped paired reads after alignment quality filtering were considered. Finally, gene level FPKM (Fragments Per Kilobase Million) and raw read count values were computed by the R package DESeq2 (version 1.18.1).

Across the 15 cAMP datasets, 16,082 out of ~40,000 distinct transcripts were profiled in all cohorts and used for analysis (Figure S4A).

Metabolomics Data Preprocessing

The metabolomics data for 3/15 datasets (GBM, LiCa1 and LiCa2) were provided preprocessed and were therefore used in their original form. OV data was already normalized and was only log2-transformed before analysis. For the other 11 metabolomics datasets in our study, we obtained the raw metabolomics data from the data owners. In this case, the processing pipeline was standardized across datasets and included batch correction via median scaling if multiple batches of data were present (only necessary for PRAD dataset as it was the only dataset produced in distinct batches) and probabilistic quotient normalization using either only normal samples if available, or all tumor samples if no normal sample was included, and only metabolites with less than 20% values missing to create the reference sample. After normalization, metabolite abundances were log2-transformed. Metabolites with more than 80% values missing were excluded from the analysis. For the remaining metabolites, missing values were imputed using
the minimum value recorded. Data preprocessing was performed using the R package maplet. Metabolite names and annotations were manually harmonized for consistency by matching HMDB\textsuperscript{74}, KEGG\textsuperscript{75} and Metabolon platform-specific peak IDs across datasets. The overlap of metabolites across datasets was heterogeneous (Figure S4B): Out of 2,411 unique molecules quantified across all datasets, fewer than 500 were measured in more than 5 datasets, and only three metabolites (gluconate, glucose, and glucose-3-phosphate) were quantified in all 15 datasets in tumor. This high variability can be attributed to a variety of technical and biological factors, including metabolite ionizability on the mass spectrometer and potential specificity of certain metabolites to distinct tissues.

**Concordance meta-analysis**

To identify gene-metabolite pairs that were consistently associated across tumor types and cohorts, we used a stratified, weighted concordance model. Concordance is a non-parametric measure of correlation, similar to Kendall's Tau, that relies on the concept of concordant pairs\textsuperscript{37}.

Briefly, consider a pair of observations $i$ and $j$, for which both metabolite $m$ and gene $g$ have been measured. The pair is defined to be concordant if $\text{sign}(m_i - m_j) = \text{sign}(g_i - g_j)$, that is, if they have the same order in both samples, and discordant if they have opposite signs. Once pairwise concordance has been estimated for all pairs, the overall concordance $c$ is calculated as

$$c = \frac{\# \text{ concordant pairs}}{\# \text{ concordant pairs} + \# \text{ discordant pairs}}$$

To account for the multiple cohorts in our study, we performed a stratified concordance analysis, where pairwise comparisons are only calculated within datasets, but not across. A global concordance value is then estimated by counting the overall number of concordant pairs according to the formula above. Moreover, given the vast heterogeneity in sample sizes across studies, we downweighted each observation by the number of samples in the corresponding study (i.e., $1/\text{nsamples(dataset}_i)$), so that each dataset would contribute equally to the overall concordance. The concordance calculation was performed using the concordance function from the survival R package (v3.2-3)\textsuperscript{76}.

To make this quantity more intuitive, we further scaled the concordance range to values between -1 and 1 through $c_{\text{scaled}} = 2c - 1$, which is analogous to Somers' $D$\textsuperscript{77}. Furthermore, in the absence of ties, this value is also identical to Kendall's rank correlation coefficient $\tau$\textsuperscript{77}. In the figures presented in this paper, a scaled concordance of 0 indicates absence of association, a value greater than 0 indicates positive association, while a value less than 0 indicates negative association.

A z-score was computed as concordance minus 0.5 and divided by the square root of the variance, and the resulting value was used to derive a two-tailed $p$-value\textsuperscript{78,79}. $P$-values were corrected for multiple testing using the Benjamini-Hochberg method to control the false discovery rate\textsuperscript{80}. Pairs with adjusted $p$-values below 0.01 were considered significant.
Generation of *IDO1* Knockdown HCT116 Cell Line

sgRNAs (oligonucleotide sequences are indicated in Table S6) targeting *IDO1* as well as non-targeting control were cloned into lentiCRISPRv2 puro plasmid (Addgene, #98290). Lentiviral packaging vectors psPAX2 (Addgene, #12260), pMD2.G (Addgene, #12259), along with sgRNA expressing vector were transfected into HEK293T cells using PEI (polyethyleneimine) transfection reagent. 72 hours after transfection, supernatant containing lentivirus was harvested and filtered through a 0.45 µm pore size Whatman filter (Fisher Scientific) to remove cell debris. Target cells (HCT116, human colon cancer cell line, ATCC) were transduced with lentivirus using 8 µg/ml polybrene (Sigma). 72 hours after lentivirus transduction, 3 µg/ml puromycin (ThermoFisher) was added to cell culture media to select for virus-infected cells. 2 weeks after puromycin selection, target gene knock-out was confirmed by western blot. For induction of *IDO1* gene expression, cells were treated with 100 ng/ml IFN-γ for 48 hours.

Metabolomics Profiling on the *IDO1* Knockdown HCT116 Cell Line

Cells were plated in 6-well tissue culture plates at 100K cells/well density. 72 hours after cell seeding, metabolites were extracted and analyzed by liquid chromatography-mass spectrometry (LC-MS). For metabolite extraction, culture media was aspirated, and cells were washed once with ice-cold phosphate buffered saline (PBS). After PBS washing, 1 ml of ice-cold extraction solvent (methanol:water = 80:20) was added. After overnight incubation at -80°C, cells and extraction solvent were collected into 1.5ml Eppendorf tube using cell scraper. Samples were centrifuged at 20,000g for 20 min at 4°C. Supernatant (900 µl) was collected and dried in a vacuum evaporator (Genevac EZ-2 Elite).

For LC-MS, dried extracts were resuspended in 30 µl of 97:3 water:methanol containing 10 mM tributylamine and 15 mM acetic acid. Samples were vortexed, incubated on ice for 20 min, and clarified by centrifugation at 20,000g for 20 min at 4°C. LC-MS analysis used a Zorbax RRHD Extend-C18 column (150 mm x 2.1 mm, 1.8 µm particle size, Agilent Technologies). Solvent A was 10 mM tributylamine, 15 mM acetic acid in 97:3 water:methanol, and solvent B was 10 mM tributylamine and 15 mM acetic acid in 3:97 water: methanol, prepared according to the manufacturer’s instructions (MassHunter Metabolomics dMRM Database and Method, Agilent Technologies). LC separation was coupled to a 6470 triple quadrupole mass spectrometer (Agilent Technologies) which was operated in dynamic MRM scan type and negative ionization mode.

Tumor vs Normal Pathway Analysis

There were 7 cAMP datasets which had both tumor and normal samples available (BRCA1, COAD, GBM, PRAD, PDAC, ccRCC3 and ccRCC4). We applied differential gene expression tests between tumor and normal samples in each dataset using limma-voom (limma package version: 3.5.2). Genes with an FDR-adjusted p-value < 0.1 were considered significantly differentially expressed. Similarly, we also conducted differential metabolite abundance testing between tumor and normal samples using Wilcoxon rank sum tests. Metabolites with an FDR-adjusted p-value below 0.1 were considered significantly differentially abundant.
For each KEGG pathway, we calculated the differential fraction (“DF”) score and the differential abundance (DA) score for genes and metabolites separately:

**DF score** = \( \frac{\text{number of significantly up-regulated constituents} + \text{number of significantly down-regulated constituents}}{\text{number of measured constituents in a pathway}} \)

**DA score** = \( \frac{\text{number of significantly up-regulated constituents} - \text{number of significantly down-regulated constituents}}{\text{number of measured constituents in a pathway}} \)

Constituents are either genes or metabolites in the above formula.

Conceptually, the DF score captures the overall disruption of the constituents of a pathway, whereas the DA score captures the tendency for pathway constituents to increase or decrease in abundance relative to a reference (in this case, normal tissue) state.

**Gene-Metabolite Distance**
We used a genome-scale human metabolic model (“Human 1” model\(^{41}\)) covering the metabolic reaction information of transporters, enzymes and metabolites to calculate the distance between gene and metabolite pairs. We defined a “proximal” interaction by a gene/metabolite pair participating in the same reaction (shortest path distance of 1) or participating in subsequent reactions (shortest path distance of 2) (Figure S5).

**Gene Set Enrichment Analysis**
For the analysis in Figure 4, we ran a pathway enrichment analysis among all significant genes for each metabolite with at least one significant gene association. The enrichment test was performed using classical hypergeometric testing\(^{81}\). We tested the enrichment of 146 KEGG pathways. P-values were adjusted using the Benjamini-Hochberg method for controlling the false discovery rate\(^{80}\). Adjusted p-values < 0.01 were considered significant.

Results were then aggregated into a metabolite x pathway matrix and visualized as a heatmap, where metabolites and pathways were clustered based on \(\log_{10}(p\text{-value})\). Clustering was performed using the pheatmap function\(^{82}\) (pheatmap package v 1.0.12) with Ward linkage and Euclidean distance.

**Bulk Gene Expression Deconvolution Analysis**
To dissect the role of the immune compartment in the tumor microenvironment, we calculated the ImmuneScore through the estimate R package\(^{83}\). To calculate cell-type-specific infiltration patterns in Figure 6, we employed the ssGSEA\(^{46}\) for bulk gene expression deconvolution analysis. Signature gene lists of immune cell types as well as immune features were obtained from Bindea et al.\(^{51}\). Briefly, ssGSEA takes the sample FPKM expression values as the input and computes an enrichment score for a given gene list as compared to all the other genes in the sample transcriptome.
Acknowledgments

We thank each of the individual laboratories which contributed their data to the cAMP. The Reznik Lab acknowledges Payam Gammage for useful feedback and discussions.

Competing financial interests

JK holds equity in Chymia LLC and IP in PsyProtix and is cofounder of iollo. All other authors have no competing interests to declare.
Figure 1. Summary of the Cancer Atlas of Metabolic Profiles (cAMP).

(A) The cAMP integrates metabolomic and transcriptomic data from 15 datasets covering 11 different cancer types, covering both tumor and normal tissue. Cancer type abbreviations, BRCA1/BRCA2: breast cancer, COAD: colon adenocarcinoma, DLBCL: diffuse large B-cell lymphoma, GMB: glioblastoma, HürthleCC: Hürthle cell carcinoma, HCC: hepatocellular carcinoma, ICC: intrahepatic cholangiocarcinoma, OV: high-grade serous ovarian cancer, PDAC: pancreas adenocarcinoma, PRAD: prostate adenocarcinoma, ccRCC1/ccRCC2/ccRCC3/ccRCC4: clear cell renal carcinoma. (B) Overview of non-parametric concordance meta-analysis. For a given transcript (“T”) and metabolite (“M”) pair, all measurements from all datasets are considered and weighted according to sample size. The concordance is a non-parametric measure of bivariate association between T and M that can be applied in meta-analysis across multiple datasets.
Figure 2. Meta-analysis across the cAMP captures lineage-transcending gene-metabolite interactions.

(A) Volcano plot of the gene-metabolite interactions (GMIs) computed between the 16,082 genes present in all datasets, and the 276 metabolites present in at least 8 of our 15 tumor cohorts. The x-axis indicates the scaled concordance value, where values above 0 indicate positive association and values below 0 indicate negative association. The y-axis represents the corresponding $-\log_{10}$ FDR-adjusted p-value. The horizontal line indicates the significance cutoff of 0.01 FDR. Light gray dots indicate non-significant gene-metabolite pairs, and black dots indicate significant pairs. Three
top hits have been highlighted. (B) Statistically significant GMIs are enriched for proximal interactions, but proximal interactions nevertheless constitute a minority of all statistically significant GMIs. (C) Proximal GMI prioritization. GMIs for the 22 metabolites whose strongest GMI was proximal (distance less or equal 2). For each metabolite, we ranked genes by their statistical significance. Red and black dots indicate proximal and non-proximal genes significantly associated with the corresponding metabolite, respectively, while gray dots indicate genes with non-significant associations. Bold metabolites exhibit a large gap between the dominant GMI and all other GMIs for a metabolite. (D-E) Scatter plots of the association between kynurenine levels and two proximal genes (IDO1, Panel D; AFMID, Panel E). Metabolite abundances were scaled within each dataset to be displayed together. (F) CRISPR-CAS9 mediated knockdown of IDO1 depletes IDO1 protein levels in HCT116 cells. (G) Kynurenine levels are depleted upon knockdown of IDO1 in HCT116 cells. (H) Scatter plots of the association between tryptophan levels and IDO1 in the cAMP. (I) Tryptophan levels increase upon IDO1 knockdown. (J) Scatter plot of oxidized glutathione (GSSG) and GGT1 (K) Validation of the GSSG and GGT1 relationship based on the study from Priolo et al.42.
Figure 3. Tumor vs. normal changes in metabolite and transcript abundance are predominantly asynchronous.

(A) Heatmap of metabolite and transcript differential abundance ("DA") scores across datasets and pathways, capturing the tendency for metabolites and genes to accumulate or deplete in tumors relative to normal tissues. The size of the dots indicates the number of molecules...
measured in that pathway while the color represents the DA score. (B) Spearman correlation coefficients of the metabolite and transcript differential fraction (“DF”) scores in KEGG pathways. Red dots indicate nominal significance (p-value<0.05). A minority of pathways showed significant association between metabolomic and transcriptomic disruption across cAMP datasets. (C) Example of Spearman correlation calculation: Distribution of the metabolite (x-axis) and transcript (y-axis) DF scores across datasets for the Citrate cycle (TCA cycle) pathway.

Figure 4. GMIls are concentrated in hub metabolites associated with immune-related genes. (A) Distribution of significant GMIls across metabolites. The top 10 metabolites with the highest number of significant associations are labeled. (B) Distribution of significant GMIls across genes. The top 10 genes with the highest number of significant associations are labeled. (C) KEGG-
based gene pathway enrichment analysis results for metabolites with at least one GMI and at least one significantly enriched pathway. The heatmap color represents the strength of the enrichment as the negative log_{10}(p-value) of the pathway enrichment test. Cells colored in shades of red indicate pathways that were significant after multiple testing correction (FDR<0.01), and gray cells were insignificant associations. Individual pathways were classified based on the type of process they describe: Metabolism (lilac), immune (green), or other (cream).

Figure 5. The abundance of NAD+-related metabolites is shaped by immune infiltration in the tumor microenvironment.
(A) Left: Association of metabolite abundance with immune infiltration varies significantly across cancer types. The barplot indicates the fraction of metabolites significantly associated with the ImmuneScore signature in each dataset. Bar colors code for the sample size of the dataset.
(Middle) Boxplots indicate the expression range of the ImmuneScore signature within each dataset. Right: Plot of scaled concordance calculated between metabolites and ImmuneScore. Red dots indicate metabolites with a positive association with ImmuneScore, blue dots metabolites with negative associations, and gray dots indicate metabolites that were not
statistically significant. (B) Heatmaps of five metabolites, three of which are highly correlated to ImmuneScore in the ICC dataset and two which are highly correlated with ImmuneScore in the HCC dataset. Samples are sorted by increasing ImmuneScore. UDP- G = UDP-glucuronate, GPC-16:1 = 1-palmitoleoyl-GPC (16:1), GPC-18:2 = 1-linoleoyl-GPC (18:2). (C) Bar plot indicating the strength of association between metabolites and ImmuneScore from concordance meta-analysis. Bar length represents the \(-\log_{10}\) adjusted concordance meta-analysis p-value. Metabolites related to NAD+ metabolism are shown in black. (D) Scatterplots of the abundance of two NAD+-related metabolites, quinolinate (top) and NMN (bottom), versus ImmuneScore expression across all datasets. Metabolite abundances were scaled within each dataset. (E) Bar plot of FDR-adjusted p-values from a one-sided Wilcoxon test comparing the absolute concordance values of metabolites to ImmuneScore in a pathway compared to all other pathways. (F) Metabolomic measurements of purified populations of CD45- tumor cells and CD45+ T cells isolated from ovarian cancer tumors. NAD+ was negatively correlated to the ImmuneScore signature in our ovarian cancer dataset. NAD+ was similarly lower in abundance in CD4/8+ cells than CD45- (non-immune) cells in the dataset of purified cell populations.

Figure 6. A subset of metabolites associates with specific immune cell lineages. (A) Volcano plot of cell type signature-metabolite interactions. The rug plot at the bottom highlights the numerous statistically significant associations with quinolinate. (B) Manhattan plot of adjusted concordance p-values between metabolites and cell types. (C) Histamine associates with the abundance of mast cells across most datasets of the cAMP. (D) Bar plot of adjusted concordance p-values of metabolites correlated with the mast cell gene signature. Red dotted line indicates the...
significance cutoff of 0.05. (E) HDC expression strongly associates with mast cell abundance across the cAMP. (F) Kynurenine abundance associates with an activated dendritic cell signature across the cAMP.
**Tables**

Table 1. Dataset Overview.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Cancer Type</th>
<th>Reference</th>
<th>Samples (T / N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>Breast Cancer</td>
<td>Terunuma et al. 2014&lt;sup&gt;19&lt;/sup&gt;</td>
<td>61 / 47</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Breast Cancer</td>
<td>Tang et al. 2014&lt;sup&gt;20&lt;/sup&gt;</td>
<td>18 / --</td>
</tr>
<tr>
<td>COAD</td>
<td>Colon Adenocarcinoma</td>
<td>Satoh et al. 2017&lt;sup&gt;18&lt;/sup&gt;</td>
<td>37 / 39</td>
</tr>
<tr>
<td>DLBCL</td>
<td>Diffuse Large B-cell Lymphoma</td>
<td>Calvo-Vidal, Zamponi et al. 2021&lt;sup&gt;17&lt;/sup&gt;</td>
<td>62 / --</td>
</tr>
<tr>
<td>GBM</td>
<td>Glioblastoma</td>
<td>Wang et al. 2021&lt;sup&gt;16&lt;/sup&gt;</td>
<td>74 / 6</td>
</tr>
<tr>
<td>HürthleCC</td>
<td>Hürthle Cell Carcinoma</td>
<td>Ganly, Liu et al. 2022&lt;sup&gt;86&lt;/sup&gt;</td>
<td>28 / 3</td>
</tr>
<tr>
<td>HCC</td>
<td>Hepatocellular Carcinoma</td>
<td>Chaisaingmongkol et al. 2017&lt;sup&gt;15&lt;/sup&gt;</td>
<td>54 / --</td>
</tr>
<tr>
<td>ICC</td>
<td>Intrahepatic Cholangiocarcinoma</td>
<td>Chaisaeng Mongkol et al. 2017&lt;sup&gt;15&lt;/sup&gt;</td>
<td>86 / --</td>
</tr>
<tr>
<td>OV</td>
<td>High Grade Serous Ovarian Cancer</td>
<td>Gentric et al. 2019&lt;sup&gt;14&lt;/sup&gt;</td>
<td>45 / --</td>
</tr>
<tr>
<td>PDAC</td>
<td>Pancreas Adenocarcinoma</td>
<td>Zhang et al. 2013&lt;sup&gt;21&lt;/sup&gt;</td>
<td>27 / 12</td>
</tr>
<tr>
<td>PRAD</td>
<td>Prostate Adenocarcinoma</td>
<td>Penney et al. 2020&lt;sup&gt;13&lt;/sup&gt;</td>
<td>91 / 46</td>
</tr>
<tr>
<td>ccRCC1</td>
<td>Clear Cell Renal Carcinoma</td>
<td>Hakimi et al. 2016&lt;sup&gt;22&lt;/sup&gt;</td>
<td>32 / --</td>
</tr>
<tr>
<td>ccRCC2</td>
<td>Clear Cell Renal Carcinoma</td>
<td>in-house</td>
<td>30 / --</td>
</tr>
<tr>
<td>ccRCC3</td>
<td>Clear Cell Renal Carcinoma</td>
<td>in-house</td>
<td>67 / 47</td>
</tr>
<tr>
<td>ccRCC4</td>
<td>Clear Cell Renal Carcinoma</td>
<td>in-house</td>
<td>52 / 24</td>
</tr>
</tbody>
</table>
Supplementary Table Legends

Table S1
Table of statistically significant GMIs and associated statistics.

Table S2
Summary of differential abundance results for pathways considered in tumor vs. normal analysis.

Table S3
Association of DF gene/metabolite scores across pathways in tumor vs. Normal analysis.

Table S4
Pathways considered for the analysis in Figure 4.

Table S5
Comparison of metabolite correlated to ImmuneScore in the HCC and ICC datasets.

Table S6
Guide RNA sequences used for IDO1 knockout experiment.
References


