Pan-cancer metabolic landscapes: A multi-omics view

Yiguang Bai[1,2#*] Qiaoling Chen[3,4#] Yuan Li[5,6*]

[1]Department of Orthopedics, Beijing Anzhen Nanchong Hospital, Capital Medical University (Nanchong Central Hospital), Nanchong, China

[2]Department of Orthopedics, Nanchong Central Hospital, The Second Clinical Medical College of North Sichuan Medical College, Nanchong, China

[3]Department of Oncology, Beijing Anzhen Nanchong Hospital, Capital Medical University (Nanchong Central Hospital), Nanchong, China

[4]Department of Oncology, Nanchong Central Hospital, The Second Clinical Medical College of North Sichuan Medical College, Nanchong, China

[5]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Lund University, Lund, 223 87, Sweden

[6]Department of Immunotechnology, Lund University, Medicon Village, Lund, 22387, Sweden

[#]Those authors contribute equally to the paper.

[*]Corresponding author: yuan.li@immun.lth.se +46(0)735854860, baiyiguang@163.com

Abstract

Metabolic reprogramming fuels cancer progression, but whether common metabolic patterns exist across tumor types remains elusive. To address this, we developed parseMetab, an R package that integrates large-scale proteomic, transcriptomic, and spatial transcriptomic datasets from 24 human cancers and 3,226 samples to map pan-cancer metabolic dysregulation. Strikingly, glycan biosynthesis pathways – especially those driving

fucosylation and sialylation – were consistently upregulated, while early sugar nucleotide precursors were suppressed. Nucleotide metabolism was broadly enhanced, uncovering conserved metabolic programs that enable tumor growth, adaptation, and immune evasion. Our results highlight universal metabolic vulnerabilities that may be therapeutically exploited across diverse cancer contexts.

Significance statement

While metabolic reprogramming is a known hallmark of cancer, identifying universal patterns across diverse tumor types remains a major challenge. We developed **parseMetab**, a computational framework to integrate large-scale proteomic and transcriptomic data from over 3,200 samples across 24 human cancers. Our analysis reveals a striking, conserved "glyco-switch": tumors consistently prioritize the production of complex, immune-evading surface sugars (fucosylation and sialylation) while depleting their own internal sugar precursor pools. Furthermore, we demonstrate that nucleotide metabolism is broadly hyperactivated across nearly all studied malignancies. By defining these ubiquitous metabolic signatures, this study uncovers conserved vulnerabilities that transcend specific cancer types, providing a map for developing broad-spectrum metabolic therapies to combat tumor growth and immune evasion.

Classification: Biological Sciences, Medical Science

Keywords

Pan-cancer, Metabolism, Multi-omics integration, Spatial, Glycosylation, Nucleotide

Introduction

Metabolic reprogramming is a hallmark of cancer, enabling malignant cells to proliferate rapidly, survive stress, and adapt to dynamic microenvironments [1-3]. Beyond altered energy production, cancer metabolism supports biosynthesis, redox balance, and signaling pathways that collectively drive tumor growth [4, 5]. While central metabolic pathways have been extensively studied, key processes such as glycan biosynthesis – which play critical roles in protein modification, cell communication, and immune interactions – have received comparatively limited pan-cancer attention [6]. This is notable given that glycan synthesis – a complex enzymatic process occurring within the endoplasmic reticulum and Golgi apparatus – adds regulatory layers that significantly influence tumor behavior [7, 8]. Although aberrant glycosylation is broadly recognized as a hallmark of tumorigenesis affecting essential cancer cell functions, its prevalence and variation across diverse tumor types remain incompletely characterized in systematic pan-cancer studies [9].

Many studies of cancer metabolism focus on individual tumor types [1-3]. Pan-cancer analyses, when performed, are frequently based on a single omics layer – most commonly transcriptomic data from The Cancer Genome Atlas (TCGA) [10-13]. Although multi-omics pan-cancer approaches have emerged, they often emphasize selected metabolic pathways rather than systematically characterizing the full metabolic landscape [14-17]. Here, we uniquely integrate independent proteomic, transcriptomic, and spatial transcriptomic dataset collections to construct a multi-dimensional pan-cancer metabolic landscape. By applying complementary computational approaches across diverse omics modalities and leveraging spatial data to distinguish tumor microenvironment compartments, we offer a more comprehensive and biologically contextualized view of tumor metabolism. This integrated framework facilitates the identification of shared metabolic programs and highlights pathways that may serve as broadly targetable vulnerabilities across cancer types.

Results and Discussion

To characterize pan-cancer metabolic reprogramming, tumor and matched non-tumor samples were analyzed across independent proteomics, bulk transcriptomics, and spatial transcriptomics datasets, encompassing 24 cancer types and 3,226 samples (1,610 tumor, 1,616 non-tumor) from multiple studies (Fig. 1A, Table S1). Metabolic task activities were first inferred using the CellFie framework on two proteomics datasets: the unified pan-cancer resource by Zhou et al. [18] and a curated proteomics collection by Hu et al. [19], identifying 176 and 138 active tasks, respectively, grouped into seven metabolic classes comprising 77-136 proteins (Fig. 1B; Table S1). Complementarily, Gene Set Variation Analysis (GSVA) [20] was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [21]. GSVA was applied to the aforementioned proteomics datasets, transcriptomic data from The Cancer Genome Atlas (TCGA) [22], and spatial transcriptomic pseudobulk RNA-seq data from SMTdb [23]. This analysis resolved 55–83 metabolic pathways across eleven classes, representing 1,454–1,705 genes (Fig. 1B, S1Q-S; Tables S1, S2).

Fig. 1.

Fig. S1.

Table S1.

Table S2.

Comparison of tumors with matched normal samples revealed widespread metabolic reprogramming across cancer types, with consistent patterns across independent datasets. Notably, Uterine Corpus Endometrial Carcinoma (UCEC) showed strong metabolic upregulation, potentially driven by frequent activation of the insulin-phosphoinositide 3

kinase (PI3K) signaling pathway and by estrogen signaling [24] (Fig. 1C, 1F, 1G and S2A). In contrast, Liver Hepatocellular Carcinoma (LIHC) consistently ranked among the most metabolically downregulated cancers (Fig. 1C, 1F, 1G and S2A), in line with a liver tissue metabolomics study showing that the majority of significantly altered metabolites were reduced in tumor tissue relative to matched adjacent liver [25]. These results highlight robust and reproducible pan-cancer metabolic alterations across diverse cohorts.

Metabolic reprogramming in cancer showed significant heterogeneity across metabolic classes. Notably, nucleotide and glycan metabolism were consistently upregulated across all four datasets, with stronger effects observed in (spatial) transcriptomics (Fig. 1D-E, Fig. S1T-U, Table S3). This pattern was further confirmed at the protein level in the two proteomics datasets using paired differential analysis (Fig. 1D, Table S3). Multi-omics integration using iNMF – restricted to glycan and nucleotide pathways – demonstrated a clear separation between malignant and normal tissues, while successfully mitigating batch effects across the integrated omics layers (Fig. 2A-B). Conversely, amino acid metabolism was among the most downregulated classes in three bulk omics comparisons (tumor vs. matched normal), a trend not seen in spatial omics comparing malignant and stromal spots (Fig. 1D-E, Fig. S1T-U).

Table S3.

Fig. 2.

The tumor glyco-switch: from simple sugars to complex capping

Among those broadly conserved alterations, glycan metabolism displayed a particularly coherent and biologically informative pattern. Tumors showed coordinated upregulation of pathways generating activated sugars for complex glycan modifications – including fucosylation and sialylation – accompanied by downregulation of common sugar nucleotide

precursors such as uridine diphosphate glucose (UDP-glucose) and guanosine diphosphate mannose (GDP-mannose) (Fig. 1F-G, S2A-B, S3A-B, Table S3). Consistently, perturbations in mannose metabolism have been shown to disrupt nucleotide sugar biosynthesis and impair cancer cell proliferation [26, 27]. This shift is consistent with cancer-associated remodeling of cell surface, secreted, and extracellular matrix glycoproteins, which enhances invasiveness, immune evasion, and signaling through structurally complex glycans [8, 28]. The reduced synthesis of early sugar precursors likely reflects a metabolic prioritization toward downstream glycosylation pathways that promote tumor survival and progression.

Fig. S2.

Fig. S3.

Biologically, increased CMP-N-acetylneuraminate and guanosine diphosphate L-fucose (GDP-L-fucose) synthesis facilitate the generation of sialylated and fucosylated glycans that modulate immune recognition and metastasis [29, 30]. Concurrently, the reduction in UDP-glucose and GDP-mannose production may indicate altered N-glycan core assembly or glycogen metabolism, further supporting a tumor-adapted metabolic state.

This robust, cross-cancer evidence – derived from a unified analytical framework – strengthens the concept that alterations in glycosylation pathways represent a broadly conserved metabolic hallmark of tumorigenesis. While previous studies have suggested glycosylation as a universal feature of malignant cells [31], our systematic analysis provides the first comprehensive, multi-dimensional confirmation of this phenomenon across diverse cancers, underscoring its potential as a target for diagnostic and therapeutic strategies [32]. Gene co-expression network analysis in glycan metabolism identified key hub genes – e.g. ALG3 (Alpha-1,3-mannosyltransferase), RPN2 (ribophorin II), and DDOST (Dolichyl-diphosphooligosaccharide – protein glycosyltransferase 48 kDa subunit) – that may drive this

metabolic rewiring and represent candidates for targeted cancer therapies (Fig. 2C-D**).** Notably, ALG3, an ER-localized glycosylation enzyme, is implicated in promoting tumor growth, invasion, and resistance to radio-, and immunotherapies [33-35]. Moreover, RPN2, an oncogenic subunit of the N-oligosaccharyltransferase complex, is aberrantly expressed in multiple human cancers, where it promotes tumor progression, metastasis, and drug resistance [36].

Pan-cancer upregulation of nucleotide metabolism: a systemic hallmark

Moreover, nucleotide metabolism – covering the synthesis, salvage, and catabolism of purine and pyrimidine nucleotides – is a fundamental metabolic vulnerability exploited by cancer cells to drive proliferation, evade immune surveillance, resist chemotherapy, and promote metastasis [37, 38]. Leveraging a comprehensive pan-cancer analytical framework, we systematically quantify and characterize the upregulation of nucleotide metabolic pathways across a wide spectrum of tumor types. Our findings reveal that this metabolic hallmark is not confined to isolated cancer subsets but represents a widespread feature observed in multiple malignancies, thereby expanding upon and unifying prior reports focused on individual tumor contexts [39].

Beyond the hallmark: the heterogeneous metabolic architecture of cancer

In contrast to the consistent upregulation observed in nucleotide and glycan metabolism, our analysis reveals that the dysregulation of other metabolic pathways exhibits substantial heterogeneity both between different cancer types and across independent dataset collections. This variability is likely driven by a complex interplay of factors, including tissue-specific metabolic programming, diverse oncogenic mutations, the influence of the tumor microenvironment, and technical differences in sample processing and data generation [3, 40].

Tissue origin profoundly shapes the metabolic phenotype of tumors, as distinct organs possess unique baseline metabolic demands and enzyme expression profiles that tumors may retain or reprogram selectively [41]. Furthermore, oncogenic drivers such as MYC, KRAS (Kristen Rat Sarcoma viral oncogene), and TP53 (tumor protein p53) mutations modulate metabolic pathways differently across cancer types, resulting in context-dependent metabolic adaptations [2]. The tumor microenvironment, including nutrient availability, hypoxia, and stromal interactions, further sculpts metabolic reprogramming in a spatially and temporally dynamic manner [42]. Technical heterogeneity across datasets – including differences in sample collection, sequencing platforms, and bioinformatic pipelines – can also contribute to observed variability, underscoring the need for careful data integration and validation [43]. This pronounced metabolic heterogeneity challenges the identification of universal metabolic vulnerabilities beyond the consistently conserved nucleotide and glycan pathways, emphasizing the value of integrative multi-cohort analyses to comprehensively capture metabolic diversity and enhance the robustness and predictive power of metabolic biomarkers.

Conclusion

Our integrative analysis across proteomics, bulk transcriptomics, and spatial transcriptomics reveals conserved pan-cancer metabolic reprogramming, with coordinated upregulation of nucleotide and glycan metabolism and key hubs such as ALG3 and RPN2. These results highlight broadly conserved metabolic vulnerabilities and the importance of glycosylation in tumor progression. Nevertheless, limitations including tumor purity and the observational nature of multi-omics data underscore the need for future validation using direct metabolomics/glycomics measurements and functional experiments.

Methods

To study pan-cancer metabolic reprogramming, we compared tumor and matched non-tumor samples across multi-omics datasets, integrating two proteomics resources (Hu2025, Zhou2020) [18-19] alongside the TCGA transcriptomics cohort [22] and SMTdb spatial transcriptomics collection [23]. For bulk datasets, tumor samples were compared to matched non-tumor tissues, while for spatial transcriptomics data, comparisons were conducted between malignant spots and stromal spots within the same tissue section. For each STMdb tissue section, raw RNA counts from all malignant and stromal spots were aggregated separately to generate pseudobulk RNA-seq profiles. The majority of analyzed samples were matched tumor-normal pairs from the same patients, enabling robust intra-patient comparisons across diverse cancer contexts (Table S1).

We used the Zhou et al. dataset in its preprocessed form as provided by the original publication [18], while the Hu et al. collection was curated and uniformly reprocessed by the authors [19]. Transcriptomic data for multiple cancer types were obtained from TCGA [22] using TCGAbiolinks v2.34.1. Raw gene-level counts generated using Spliced Transcripts Alignment to a Reference (STAR) were retrieved, including only Primary Tumor and Solid Tissue Normal samples with paired data per subject. Genes were filtered to retain those with counts > 5 in at least 15% of samples per group or at least two samples when 15% was < 2. Data were normalized using the limma-voom pipeline (limma v3.62.1[44], edgeR v4.4.0 [45]), yielding log2-transformed counts per million (logCPM). Spatial transcriptomics data for multiple cancer types were obtained from SMTdb [23]. For each cancer type, expression data per spot and spatial annotations (malignant, stromal, boundary) were retrieved from the SMTdb website. We included only cancer types with at least three slices containing malignant spots and three with stromal spots. For each slice, raw counts from malignant and stromal spots were aggregated separately into pseudobulk profiles, typically yielding one malignant and one stromal sample per slice. Genes were filtered using the same criteria as

TCGA data. Pseudobulk data were normalized with edgeR (v4.4.0) using TMMwsp normalization, which accounts for sample composition bias and variable expression across distinct cell populations, resulting in logCPM values.

Metabolic task scoring using CellFie

To interpret proteomics data across 16 cancer types from the Hu2025 and Zhou2020 datasets, proteins were grouped into metabolic tasks, and task activities were quantified using CellFie on the GenePattern webserver [46]. Metabolic tasks represent small reaction modules converting specific substrates to products, reflecting metabolic reprogramming that supports tumor growth and therapy resistance [2,3].

Proteins were mapped to Entrez Gene IDs using org.Hs.eg.db v3.20.0 and AnnotationDbi v1.68.0. The Zhou2020 dataset was log□-transformed prior to analysis, while the Hu2025 dataset was exponentiated to avoid negative values. Feature activity was determined via local thresholding (Local.Threshold.Type = "minmaxmean"), with thresholds set as feature means constrained between the 25th and 75th percentiles of overall abundance.

All Zhou2020 samples were analyzed together, whereas Hu2025 datasets were submitted individually. CellFie categorized metabolic tasks into seven classes (lipid, energy, nucleotide, amino acid, glycan, carbohydrate, and vitamin & cofactor metabolism). To align with Kyoto Encyclopedia of Genes and Genomes (KEGG) classifications, several tasks originally assigned to carbohydrate metabolism were reassigned to glycan metabolism: Synthesis of lactose, UDP-glucose synthesis, UDP-galactose synthesis, UDP-glucuronate synthesis, GDP-L-fucose synthesis, GDP-mannose synthesis, UDP-N-acetyl D-galactosamine synthesis, CMP-N-acetylneuraminate synthesis, N-Acetylglucosamine synthesis, Glucuronate synthesis (via inositol), Glucuronate synthesis (via UDP-glucose).

Metabolic enrichment scoring using KEGG-based GSVA

To infer metabolic activity from the two aforementioned proteomics, as well as the TCGA transcriptomic and SMTdb pseudobulk RNA-seq data, metabolic gene sets/pathways and their classes were extracted from KEGG using KEGGREST v1.46.0 [47]. Gene Set Variation Analysis (GSVA) was performed on the metabolic gene sets (size 5–500) to calculate pathway-level enrichment scores reflecting relative metabolic activity per sample. GSVA, a non-parametric unsupervised method, estimates variation in pathway activity across samples by integrating gene expression into pathway-centric scores, enabling sample-wise comparisons. Analysis was conducted using GSVA v2.0.7 [20].

Differential analysis of metabolic activities between cancer and normal samples

To understand how cancer cells reprogram their metabolic processes to support rapid growth, survival, invasion, and therapy resistance, we performed univariate differential analysis using linear models implemented in limma v3.62.2. Each metabolic task/pathway was tested individually for differences between tumor and matched normal samples. In this analysis, each cancer type was analyzed separately using the Zhou2020, Hu2025, Zhou2020_KEGG, Hu2025_KEGG, TCGA, and SMTdb dataset collections.

Differential analysis was only performed on cancer types with both cancer and normal samples, and there were at least two Cancer-Normal pairs, paired differential analysis was conducted using limma v.3.62.1 on each cancer individually. The correlation between paired tumor and matched normal samples from the same patient was modeled using the duplicateCorrelation() function, allowing for appropriate handling of intra-patient dependencies. Differential testing was performed using limma's empirical Bayes (EB) moderation, which stabilizes variance estimates and improves statistical power, particularly in datasets with limited sample sizes. To further increase robustness, the EB procedure was

carried out with robust = TRUE to reduce the influence of potential outliers and trend = TRUE to account for mean–variance relationships in the metabolic activity data. For the Zhou2020 and Hu2025 proteomics data, standard limma differential analysis was performed whereever there was not enough paired data. In the Zhou2020 and Hu2025 datasets, CellFie task metabolic activity scores were log2 transformed before limma analysis, and only metabolic tasks with at least two quantified scores in each sample group (normal and tumor) were included in limma analysis. For the twelve cancer types in the Hu2025 data, we also controlled for batch effects when multiple batches were present within a single cancer type. Benjamini-Hochberg correction was applied to control the false discovery rate.

Differential analysis of metabolic-related protein abundance between cancer and normal samples

CellFie was originally designed for bulk RNA-seq with extensive gene coverage; thus, some metabolic tasks in proteomics data may have insufficient protein measurements for stable activity scores. To validate CellFie-derived patterns, we performed paired differential expression analysis for each metabolic class at the protein level using the same limma-based approach. This analysis included only paired samples and proteins quantified in at least two samples per group (normal and tumor).

Gene co-expression network analysis

Gene co-expression network analysis was performed to uncover coordinated regulation and identify key hub genes driving metabolic rewiring in cancer. For each KEGG metabolic class, pairwise Spearman correlations among member genes were computed across TCGA samples within each cancer type and averaged to generate a co-expression matrix. Gene connectivity was calculated as the sum of absolute mean correlations excluding self-correlations, with the top 1% most connected genes (minimum three per class) defined as hubs. Edges with

absolute mean correlation > 0.6 were retained to construct undirected weighted networks using igraph v2.0.3.

Multi-omics integration

To integrate the two proteomics datasets with the TCGA cohort, we employed the iNMF method from rliger v1.0.0 [48] for a joint analysis of KEGG-derived metabolic activity scores. The analysis was restricted to pathways associated with glycan and nucleotide metabolism.

All analyses were conducted in R v4.4.3. We developed the **parseMetab** R package (v1.0.4, (https://github.com/Maj18/parseMetab, https://doi.org/10.5281/zenodo.19153922) to streamline the workflow, enabling inference and analysis of metabolic activity from omics data, differential and network analyses, and cohort-level visualizations. All computations were performed within a Docker container (Docker Hub ID: yuanli202004/cancer:v2.0.2) to ensure reproducibility and consistent computational environments. All R code required to reproduce the analyses is available at https://github.com/Maj18/cancer.

Acknowledgements

We thank Lund university library for supporting open access publications. The authors thank Carl Brunius from NBIS for valuable comments and suggestions on an earlier draft of this manuscript.

Funding

This work is supported by the Chinese Government Scholarship (Grant no. 202108515040), Sichuan Healthcare Commission Science and Technology Project (Grant no. 24QNMP006), Nanchong City-School Science and Technology Cooperation Special Project (Grant no. 22SXQT0096) and 2023 North Sichuan Medical College Research and Development Fund

Project (CBY23-ZDA11). Yuan Li is financially supported by the SciLifeLab & Wallenberg Data Driven Life Science Program, Knut and Alice Wallenberg Foundation (KAW 2020.0239, KAW 2017.0003). The funders had no role in study design, data analysis, decision to publish, or preparation of the manuscript.

Ethics approval and consent to participate

This study involved secondary analysis of publicly available, de-identified proteomic, transcriptomics and spatial transcriptomics datasets. No new human participants or animal subjects were recruited, and therefore ethics approval and consent to participate were not required.

Consent for publication

Not applicable.

Availability of data and materials

The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s)).

Competing interests

The authors declare no competing interests.

Authors' contributions

Y.L. conceived and designed the study; Y.L., Y.B. and Q.C. analyzed the data, prepared figures and/or tables. Y.L., Y.B. and Q.C. wrote the paper. All authors read and approved the final manuscript.

References

1. L.K. Boroughs, R.J. DeBerardinis. Metabolic pathways promoting cancer cell survival and growth. *Nat Cell Biol*. 17, 351-9 (2015).

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

3. N.N. Pavlova, J. Zhu, C.B. Thompson. The hallmarks of cancer metabolism: Still emerging. *Cell Metab*.34, 355-77 (2022).

4. I. Martínez-Reyes, N.S. Chandel. Cancer metabolism: looking forward. *Nat Rev Cancer*. 21, 669-80 (2021).

5. J. Yang, C. Shay, N.F. Saba, Y. Teng. Cancer metabolism and carcinogenesis. *Exp hematol oncol*. 13, 10 (2024).

6. S. Dong, T. Li. Metabolic reprogramming in cancer: signaling pathways and therapeutic targets. *Mol Cancer*. https://doi.org/10.1186/s12943-026-02582-0 (2026).

7. M. Hu, Y. Lan, A. Lu, X. Ma, L. Zhang. Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future. *Prog Mol Biol Transl Sci*. 162, 1-27 (2019).

8. A. Purushothaman, M. Mohajeri, T.P. Lele. The role of glycans in the mechanobiology of cancer. *J Biol Chem*. 299, 102935 (2023).

9. V. Varadharaj, W. Petersen, S.K. Batra, M.P. Ponnusamy. Sugar symphony: glycosylation in cancer metabolism and stemness. *Trends Cell Biol*. 35, 412-25 (2025).

10. F. Gatto, R. Ferreira, J. Nielsen. Pan-cancer analysis of the metabolic reaction network. *Metab Eng*.57, 51-62 (2020).

11. S.R. Rosario, M.D. Long, H.C. Affronti, A.M. Rowsam, K.H. Eng, D.J. Smiraglia. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat Comm*. 9, 5330 (2018).

12. Z .Xiao, Z .Dai, J.W. Locasale. Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat Commun*. 10, 3763 (2019).

13. J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 45, 1113-1120 (2013).

14. Y. Hao, D. Li, Y. Xu, J. Ouyang, Y. Wang, Y. Zhang, et al. Investigation of lipid metabolism dysregulation and the effects on immune microenvironments in pan-cancer using multiple omics data. *BMC bioinformatics*. 20, 195 (2019).

15. J. Zou, K. Du, S. Li, L. Lu, J. Mei, W. Lin, et al. Glutamine metabolism regulators associated with cancer development and the tumor microenvironment: A pan-cancer multi-omics analysis. *Genes*. 12, 1305 (2021).

16. G. Bi, Y. Bian, J. Liang, J. Yin, R. Li, M. Zhao, et al. Pan-cancer characterization of metabolism-related biomarkers identifies potential therapeutic targets. *J Transl Med*. 19, 219 (2021).

17. S. Liu, Y. Yuan, J. He, Y. Zhou, Y. Wang, X. Ye, et al. Integrated multi-omics analysis reveals a glycolytic signature that predicts pan-cancer immune checkpoint

inhibitor response and LDHA as a combinatorial target in fumarate hydratase-deficient renal cell carcinoma. *Front Immunol*. 16, 1666121 (2025).

18. Y. Zhou, T.M. Lih, J. Pan, N. Höti, M. Dong, L. Cao, et al. Proteomic signatures of 16 major types of human cancer reveal universal and cancer-type-specific proteins for the identification of potential therapeutic targets. *J Hematol Oncol*. 13, 170 (2020).

19. G.S. Hu, Z.Z. Zheng, Y.H. He, D.C. Wang, R.C. Nie, W. Liu. Integrated Analysis of Proteome and Transcriptome Profiling Reveals Pan-Cancer-Associated Pathways and Molecular Biomarkers. *Mol Cell Proteomics*. 27 (2025).

20. S. Hänzelmann, R. Castelo, J. Guinney. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics*. 14, 7 (2013).

21. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 40, D109-14 (2012).

22. J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 45, 1113-1120 (2013).

23. W. Zhou, Q. Yang, J. Guo, S. Li, M. Su, F. Leng, et al. SMTdb: A Comprehensive Spatial Meta-Transcriptome Resource in Cancer. *Mol Biol Evol*. 42, msaf263 (2025).

24. S. Kyo, K. Nakayama. Endometrial cancer as a metabolic disease with dysregulated PI3K signaling: Shedding light on novel therapeutic strategies. *Int J Mol Sci*. 21, 6073 (2020).

25. A. Ferrarini, C. Di Poto, S. He, C. Tu, R.S. Varghese, A. Kara Balla, et al. Metabolomic analysis of liver tissues for characterization of hepatocellular carcinoma. *J Proteome Res*. 18, 3067-76 (2019).

26. Y. Mizote, T. Suzuki, A. Hirayama, M. Nishida, T. Hiratsuka, Y. Imagawa, et al. Metabolic clogging of mannose triggers dNTP loss and genomic instability in human cancer cells. *Glycobiology*. 33, 971–972 (2023).

27. J.H. Ding, Y. Xiao, F. Yang, X.Q. Song, Y. Xu, X.H. Ding, et al. Guanosine diphosphate–mannose suppresses homologous recombination repair and potentiates antitumor immunity in triple-negative breast cancer. *Sci Transl Med*. 16, eadg7740 (2024).

28. S.S. Pinho, C.A. Reis. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 15, 540-55 (2015).

29. E. Rodriguez. Tumor glycosylation: a main player in the modulation of immune responses. *Eur J Immunol*. 55, e202751318 (2025).

30. W. Zhu, Y. Zhou, L. Guo, S. Feng. Biological function of sialic acid and sialylation in human health and disease. *Cell Death Discov*. 10, 415 (2024).

31. S.L. Bellis, C.A. Reis, A. Varki, R. Kannagi, P. Stanley. Glycosylation Changes in Cancer. In: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, et al., editors. Essentials of Glycobiology [Internet]. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Chapter 47 (2022).

32. H. Wang, D.J. Mooney. Metabolic glycan labelling for cancer-targeted therapy. *Nat Chem*. 12, 1102-14 (2020).

33. M.L. De Leoz, L.J. Young, H.J. An, S.R. Kronewitter, J. Kim, S. Miyamoto, et al. High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics*. 10, M110-002717 (2011).

34. X. Sun, Z. He, L. Guo, C. Wang, C. Lin, L. Ye. ALG3 contributes to stemness and radioresistance through regulating glycosylation of TGF-β receptor II in breast cancer. *J Exp Clin Cancer Res*. 40, 149 (2021).

35. P. Liu, C. Lin, Z. Liu, C. Zhu, Z. Lin, D. Xu, et al. Inhibition of ALG3 stimulates cancer cell immunogenic ferroptosis to potentiate immunotherapy. *Cell Mol Life Sci*. 79, 352 (2022).

36. Z. Han, Y. Wang, L. Han, C. Yang. RPN2 in cancer: An overview. *Gene*. 857, 147168 (2023).

37. N.J. Mullen, P.K. Singh. Nucleotide metabolism: a pan-cancer metabolic dependency. *Nat Rev Cancer*. 23, 275-94 (2023).

38. D.D. Shi, M.R. Savani, K.G. Abdullah, S.K. McBrayer. Emerging roles of nucleotide metabolism in cancer. *Trends Cancer*. 9, 624-35 (2023).

39. M.G. Vander Heiden, R.J. DeBerardinis. Understanding the intersections between metabolism and cancer biology. *Cell*. 168, 657-669 (2017).

40. P.P. Hsu, D.M. Sabatini. Cancer cell metabolism: Warburg and beyond. *Cell*. 134, 703-7 (2008).

41. J.R. Mayers, M.G. Vander Heiden. Nature and nurture: what determines tumor metabolic phenotypes? *Cancer Res*. 77, 2265-2269 (2017).

42. D.F. Quail, J.A. Joyce. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 19, 1423-37 (2013).

43. J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci signal*. 6, pl1 (2013).

44. M.E. Ritchie, B. Phipson, D.I. Wu, Y. Hu, C.W. Law, W. Shi, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43, e47 (2015).

45. M.D. Robinson, D.J. McCarthy, G.K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*. 26, 139-40 (2010).

46. A. Richelle, B.P. Kellman, A.T. Wenzel, A.W. Chiang, T. Reagan, J.M. Gutierrez, et al. Model-based assessment of mammalian cell metabolic functionalities using omics data. *Cell Rep Methods*. 1 (2021).

47. M. Kanehisa, S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28, 27-30 (2000).

48. J.D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, E.Z. Macosko. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*. 177, 1873-87 (. 2019).

Figures

Fig.1.pdf: Fig. 1. Pan-cancer metabolic dysregulation across independent pan-cancer datasets from distinct omics modalities.

(A) Sample sizes for the four independent pan-cancer datasets analyzed. **(B)** Overview of the number of identified metabolic tasks/pathways per metabolic class, in each dataset.

(C) Number of significantly up- or downregulated tasks/pathways per dataset: Hu2025 (adjusted P-value < 0.05), Zhou2020 (adj. P < 0.1), TCGA (adj. P < 0.05), and SMTdb (P < 0.1). Downregulated features are plotted as negative values. **(D/E)** Violin plots showing the distribution of moderated *t* statistics from task/pathway- or protein-level differential analyses for each metabolic class using either the proteomics (D) or the bulk/spatial transcriptomics (E) datasets. **(F/G)** Dot plots of task/ pathway -level dysregulation across cancer types in either Hu2025 proteomics (F) or TCGA transcriptomics (G) data. Here in this figure, the metabolic tasks are CellFie-based for the two proteomics datasets (Hu2025 & Zhou2020), the metabolic pathways are KEGG-based for the TCGA and SMTdb datasets. In (F) and (G), color represents moderated *t* statistics, and dot size reflects adjusted *p* values; only significant features (adj. P < 0.05) are shown.

Fig.2.pdf: Fig. 2. Multi-omics integration and gene co-expression networks for glycan and nucleotide metabolism.

(A-B) Integrative visualization of bulk proteomics (Hu2025 & Zhou2020) and bulk transcriptomics (TCGA) datasets. (C-D) Gene co-expression networks for glycan and nucleotide metabolism. Nodes represent genes, sized by normalized connectivity, with the top 1% most connected genes (≥3 per class) highlighted as hubs (red). Edges show absolute correlation > 0.6, width scaled by correlation strength; grey edges indicate positive correlations and blue edges indicate negative correlations. Networks were visualized using a Fruchterman-Reingold layout.

Additional files

Fig.S1.pdf: Fig. S1. Summary of metabolic activity and gene-level characteristics across pan-cancer datasets.

(A-D, K-L) Boxplots showing mean metabolic activity per task for each cancer type, with

one panel per dataset. (E-H, M-N) Boxplots showing mean metabolic activity per task for each metabolic class, with one panel per dataset. (I-J) Violin plots of mean protein expression for each metabolic class for the two proteomics datasets. (O-S) Boxplots of task/pathway size (number of genes per task) across metabolic classes for different datasets. (T-U) Violin plots showing the distribution of moderated *t* statistics from pathway-level differential analyses for each metabolic class in Hu2025_KEGG and Zhou2020_KEGG. (V-W) Circular barplots showing the number of metabolic genes for each cancer type within each dataset.

Fig.S2.pdf: Fig. S2. Pan-cancer metabolic dysregulation across independent pan-cancer datasets from two distinct omics modalities.

Dot plots of significant task-level dysregulation across cancer types in the Zhou2020 (A, adj. P < 0.1) and SMTdb (B, P.value < 0.1) data. Color represents moderated *t* statistics, and dot size reflects (adjusted) *p* values.

Fig.S3.pdf: Fig. S3. Pan-cancer metabolic dysregulation across independent pan-cancer datasets from two KEGG-based proteomics data. Dot plots of significant pathway-level dysregulation across cancer types in the Hu2025_KEGG (A, adj. P < 0.1) and Zhou_KEGG (B, P. adj. P < 0.05) data. Color represents moderated *t* statistics, and dot size reflects (adjusted) *p* values.

Fig.S4.pdf: Fig. S4. Gene co-expression networks for different metabolic classes. Networks nodes represent genes, sized by normalized connectivity, with the top 1% most connected genes (≥3 per class) highlighted as hubs (red). Edges show absolute correlation > 0.6, width scaled by correlation strength; grey edges indicate positive correlations and blue edges indicate negative correlations. Networks were visualized using a Fruchterman–Reingold layout.

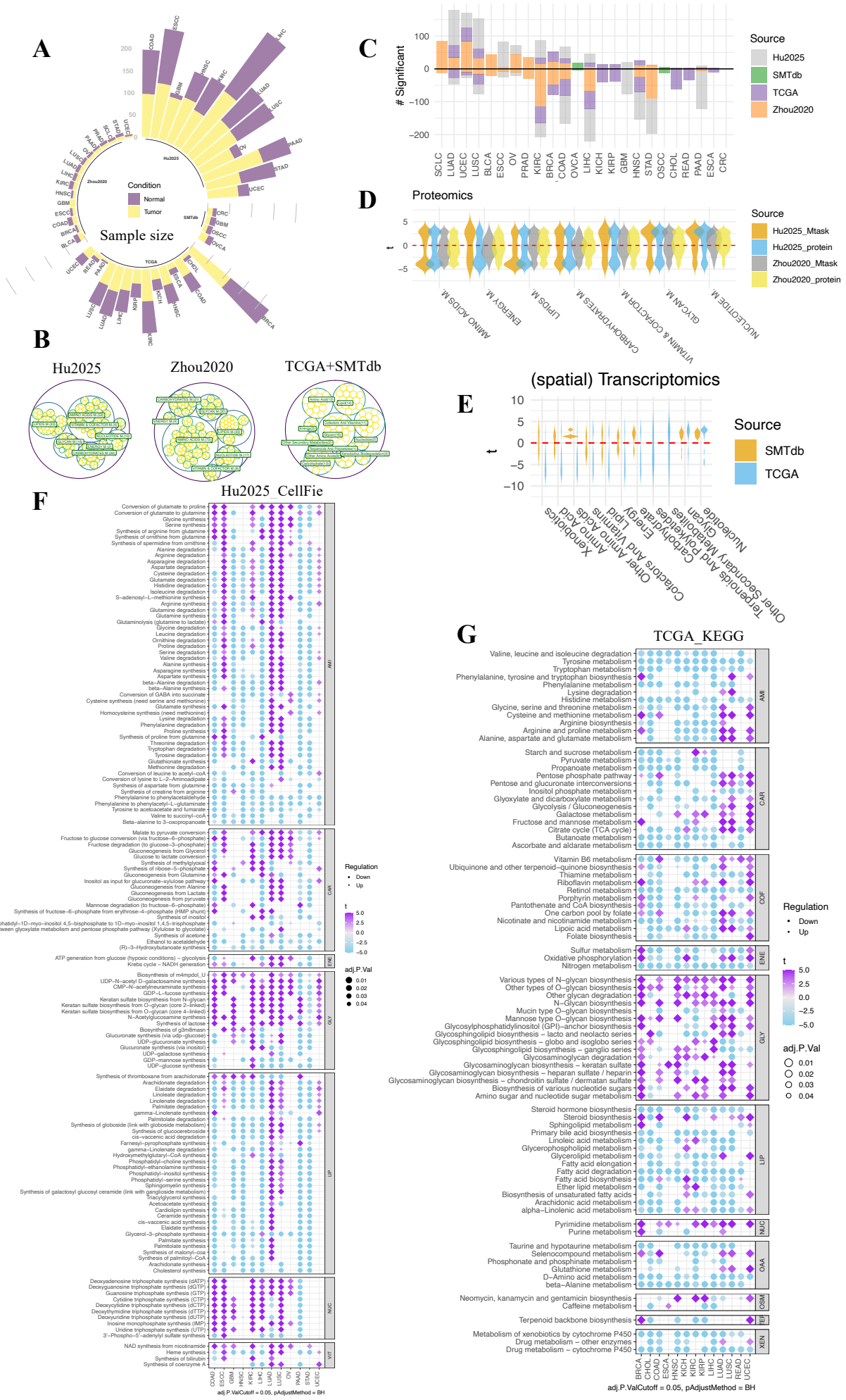
TableS1.xlsx: Table S1. Summary and sample information for pan-cancer datasets analyzed in this study.

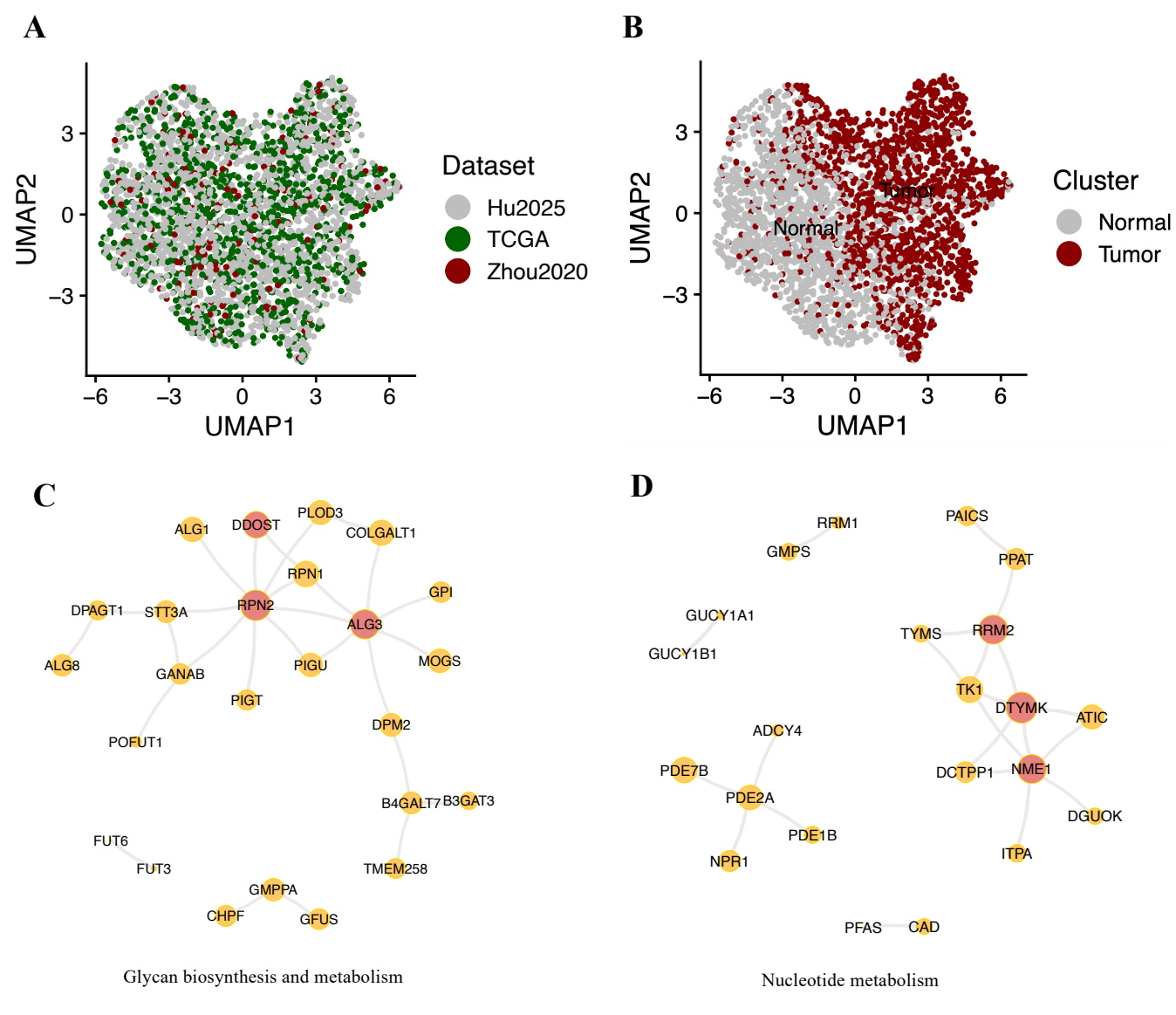
The summary sheet provides an overview of all included datasets and their corresponding omics modalities, cancer types, sample counts and identified task/pathway/protein/gene numbers. Additional sheets contain detailed sample-level information for each dataset, including sample identifiers, tissue type (tumor or normal), and associated clinical annotations.

TableS2.xlsx: Table S2. Metabolic activities inferred from proteomics, bulk transcriptomics, and spatial transcriptomics across pan-cancer datasets.

TableS3.xlsx: Table S3. Pan-cancer metabolic dysregulation results.

This Excel file contains multiple sheets showing differential analyses across cancer types and datasets. Task/Pathway-level sheets (MA_*)**:** Results for metabolic tasks or pathways for each dataset and cancer type. Protein-level sheets (Protein_*): Protein-level differential analysis within metabolic classes for each dataset. Sheet names indicate the level (task/pathway or protein), dataset, and cancer type/class for easy reference.


A
Condition
Normal
Tumor
Sample size
Hu2025
Zhou2020
SMTdb
TCGA
B
Hu2025
Zhou2020
TCGA+SMTdb
C
Significant
Source
Hu2025
SMTdb
TCGA
Zhou2020
D
Proteomics
Source
Hu2025_Mtask
Hu2025_protein
Zhou2020_Mtask
Zhou2020_protein
AMINO ACIDS M.
ENERGY M.
LIPIDS M.
CARBOHYDRATES M.
VITAMIN & COFACTOR M.
GLYCAN M.
NUCLEOTIDE M.
E
(spatial) Transcriptomics
Source
SMTdb
TCGA
Xenobiotics
Amino Acids
Vitamins
Lipid
Energy
Carbohydrate
Cofactors
Other Amino Acids
Terpenoids And Polyketides
Glycan
Nucleotide
Other Secondary Metabolites
F
Hu2025_CellFie
Regulation
Down
Up
adj.P.Val
adj.P.ValCutoff = 0.05, pAdjustMethod = BH
G
TCGA_KEGG
Regulation
Down
Up
adj.P.Val
adj.P.ValCutoff = 0.05, pAdjustMethod = BH


A
UMAP2
UMAP1
3
0
−3
−6
−3
0
3
6
Dataset
Hu2025
TCGA
Zhou2020
B
UMAP2
UMAP1
Normal
Tumor
Cluster
Normal
Tumor
C
ALG1
DDOST
PLOD3
COLGALT1
RPN1
GPI
DPAGT1
STT3A
RPN2
ALG3
ALG8
GANAB
PIGU
MOGS
PIGT
DPM2
POFUT1
B4GALT7
B3GAT3
FUT6
FUT3
TMEM258
GMPPA
CHPF
GFUS
Glycan biosynthesis and metabolism
D
RRM1
PAICS
GMPS
PPAT
GUCY1A1
TYMS
RRM2
GUCY1B1
TK1
DTYMK
ATIC
ADCY4
PDE7B
DCTPP1
NME1
PDE2A
DGUOK
PDE1B
ITPA
NPR1
PFAS
CAD
Nucleotide metabolism