



# GutUDB: A comprehensive multiomics database for intestinal diseases

Yi Bao<sup>1</sup> | Yaxin Chen<sup>2</sup> | Lizhu Lin<sup>3</sup> | Jingyi Li<sup>4</sup> | Xinli Liu<sup>1</sup> |  
Gang Wang<sup>5</sup> | Yueqi Li<sup>1</sup> | Yao Lin<sup>1</sup> | Yajing Chen<sup>6</sup> | Lijuan Zhou<sup>1</sup> |  
Yawen Qi<sup>2</sup> | Yufang Xie<sup>2</sup> | Zhenrui Lin<sup>7</sup> | Zhe Sun<sup>7</sup> | Yuwen Fan<sup>7</sup> |  
Jinjing Jiang<sup>7</sup> | Feiyu Zhang<sup>7</sup> | Hubin Chen<sup>1</sup> | Jiemei Chu<sup>1</sup> |  
Jiegang Huang<sup>1</sup> | Xuena Chen<sup>8</sup>  | Hao Liang<sup>1</sup> | Shuaiyi Liang<sup>9</sup> | Sanqi An<sup>1</sup> 

<sup>1</sup>Guangxi Key Laboratory of AIDS Prevention and Treatment & Guangxi Colleges and Universities Key Laboratory of Prevention and Control of Highly Prevalent Diseases, School of Public Health, Guangxi Medical University, Nanning, China

<sup>2</sup>Institute of Respiratory Health, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

<sup>3</sup>Department of Anesthesiology, The First People's Hospital of Qinzhou, Qinzhou, China

<sup>4</sup>Department of Pathology, Guangdong Second Provincial General Hospital, Guangzhou, China

<sup>5</sup>Precision Medicine Translational Research Center, West China Hospital, Sichuan University, Chengdu, China

<sup>6</sup>Department of Biochemistry and Molecular Biology, School of Basic Medicine, Guangxi Medical University, Nanning, China

<sup>7</sup>Life Sciences Institute, Guangxi Medical University, Nanning, China

<sup>8</sup>The Key Laboratory of Experimental Teratology, Ministry of Education, Department of Systems Biomedicine, School of Basic Medical Sciences, Shandong University, Jinan, China

<sup>9</sup>Department of Bioinformatics, Anjin Biotechnology Co., Ltd., Guangzhou, China

## Correspondence

Sanqi An and Hao Liang, Guangxi Key Laboratory of AIDS Prevention and Treatment & Guangxi Colleges and Universities Key Laboratory of Prevention and Control of Highly Prevalent Diseases, School of Public Health, Guangxi Medical University, Nanning 530021, China.

Email: [ansq@mail2.sysu.edu.cn](mailto:ansq@mail2.sysu.edu.cn) and [lianghao@gxmu.edu.cn](mailto:lianghao@gxmu.edu.cn)

Xuena Chen, The Key Laboratory of Experimental Teratology, Ministry of Education, Department of Systems Biomedicine, School of Basic Medical Sciences, Shandong University, Jinan 250012, China.

Email: [chenxuena2021@sdu.edu.cn](mailto:chenxuena2021@sdu.edu.cn)

Shuaiyi Liang, Department of Bioinformatics, Anjin Biotechnology Co., Ltd., Guangzhou 510000, China.

Email: [SLiang2@my.harrisburgu.edu](mailto:SLiang2@my.harrisburgu.edu)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 82160389; Guangxi Medical University Training Program for Distinguished Young Scholars to Sanqi An; Guangxi Science and Technology Base and Talent Project, Grant/Award Number: 2022AC19006

The intestine serves as the central organ in the digestive system for nutrient absorption and digestion processes [1]. In recent years, high-throughput sequencing technologies have significantly broadened our comprehension of

biological mechanisms from diverse aspects, including DNA-level processes, transcriptional dynamics, protein-related activities, and epigenetic modifications. These various omics sequencing data could provide a more

Yi Bao, Yaxin Chen, Lizhu Lin, and Jingyi Li contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *iMeta* published by John Wiley & Sons Australia, Ltd on behalf of *iMeta* Science.

systematic and better insight into intestinal diseases [2, 3]. However, the complexity of analytical techniques for sequencing data, coupled with the dispersed nature of data storage and the voluminous size of data sets, has impeded researchers to fully exploit these omics resources [4, 5]. Therefore, the establishment of a comprehensive database to integrate and analyze these sequencing data sets would be urgent and instrumental in dealing with these challenges.

Here, we present the Gut Universe Database (GutUDB), a high-quality and comprehensive multiomics database about intestinal diseases. It provides a user-friendly platform for the comprehensive collection of eight types of omics data, including epigenomics, genomics, transcriptomics, spatial omics, single-cell omics, proteomics, metabolomics, and microbiomics data, spanning 56 distinct intestinal diseases across six various species. GutUDB offers a thorough analysis of intestinal diseases, presenting omics data through various informative charts. We also highlight the therapeutic targets for both chemical and traditional medicine suitable to intestinal diseases, along with the associated therapeutic outcomes involving probiotics. GutUDB will play a pivotal role in identifying diagnostic targets for intestinal diseases and unveiling the molecular mechanisms underlying the progression of these conditions.

## RESULTS

### Overview of GutUDB

To date, GutUDB has accumulated approximately 9 million generated profiles from eight types of omics data sets, which encompasses 56 intestinal diseases across six species: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Macaca mulatta*, *Danio rerio*, and *Sus scrofa*. Overall, GutUDB incorporated 58,970 genes derived from 11 subtypes of intestinal tissue or 63 intestinal cell lines, and identified various potential clinical therapeutics, including chemical drugs, traditional medicine, and probiotic agents. To assist users to easily elucidate and interrogate the intricate gene-disease-omics network, four core functionalities—Browse, Query, Visualization, and Download—were integrated into GutUDB (Figure 1A).

### Disease-gene associations across distinct intestinal diseases

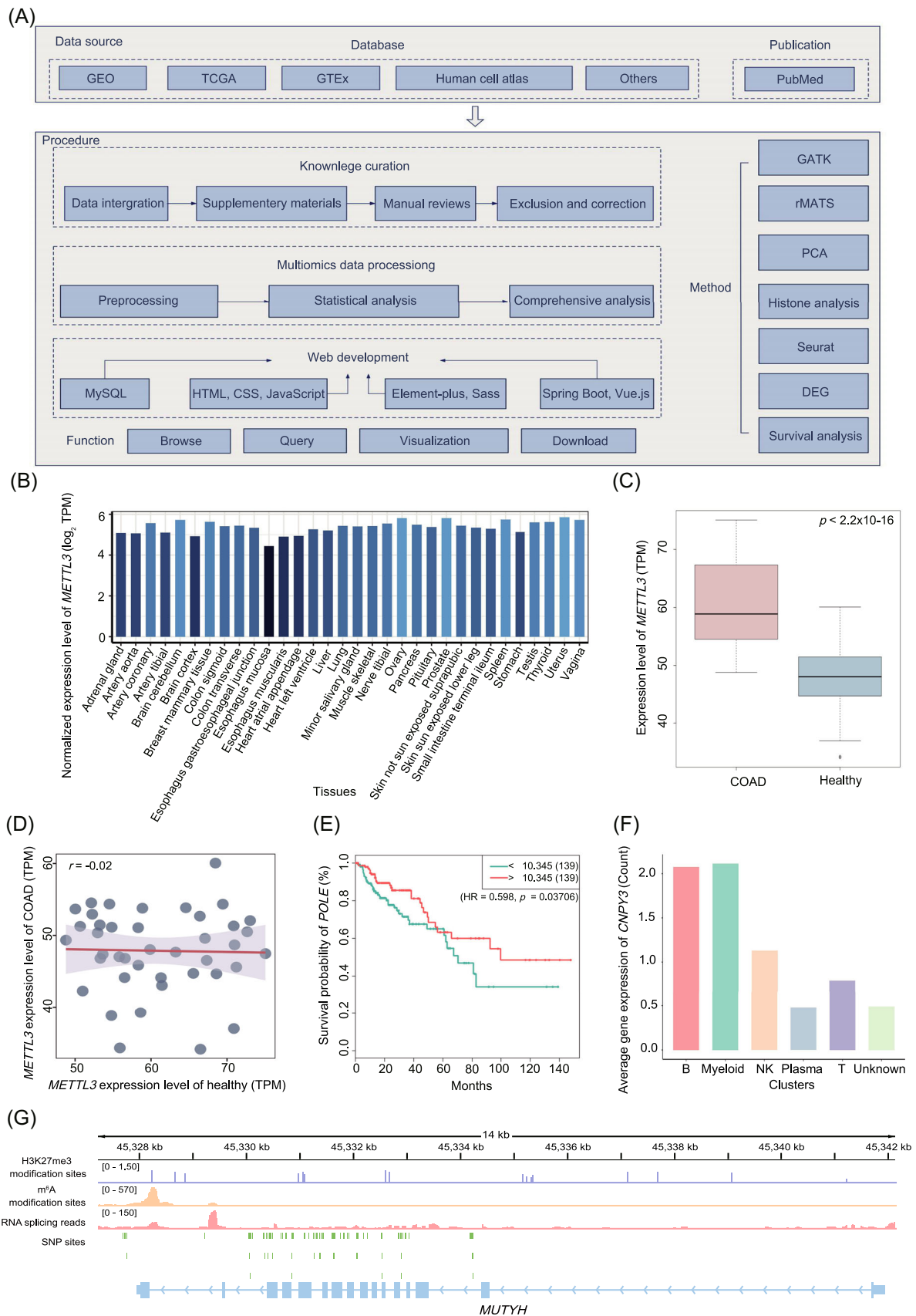
In total, 260,790 disease-gene associations were embedded in GutUDB. The main intestine-related diseases, including colon cancer, colorectal cancer (CRC), bleeding of the colon, rectal cancer, constipation,

diarrhea, ileus, inflammatory bowel disease, and small intestine cancer (Supporting Information: Figure S1A), shared common genes such as *STK11*, *CFTR*, *BMPRIA*, *SMAD4*, *NOTCH1*, *PKD1*, *MLH1*, *MSH2*, *APC*, and *MEFV* (Supporting Information: Figure S1B) [6–11]. Specifically, *STK11* is a well-known etiological factor for Peutz-Jeghers syndrome. *SMAD4* and *BMPRIA* are reported to be associated with Juvenile Polyposis Syndrome [12]. These findings illustrated the strong relationship between intestinal pathologies and these genes.

To further explore specific genes related to intestinal diseases, users can easily obtain detailed information on the homepage using gene symbols. For example, *METTL3*, upon clicking the “Search” icon, the results exhibited the details of this gene, such as genome location and functional characteristics (e.g., RNA-binding protein or transcription factor) (Supporting Information: Figure S1C), the expression levels and patterns in different tissues across eight omics levels (Figure 1B). We observed a frequency of copy number variation (CNV) deletion of only 0.04 for *METTL3* in CRC, while there was a significantly high RNA expression of this gene. We concluded that there is a weak correlation between CNV RNA expression on *METTL3*, similar correlation analyses can also be conducted for other epigenetic and proteomics data (Supporting Information: Figure S1D). Additionally, GutUDB specifically offered information on gene sets with differential expression and survival prognosis in colon adenocarcinoma (COAD) patients (Figure 1C–E).

### Four distant models: Disease, treatment, species, and omics data

GutUDB was primarily divided into four major modules—SPECIES, DISEASES, OMICS, and THERAPY, facilitating users to access and browse corresponding details upon clicking each icon on the homepage. In the “THERAPY” module, GutUDB curated 21,984 drug-disease interactions, including 6281 chemical compounds, 393 traditional medicinal herbs, and 22 probiotics (Supporting Information: Figure S1E). Among these associations, cisplatin, as a chemotherapy drug for colon cancer patients, could infiltrate tumor cells, induce DNA damage, and ultimately lead to cell death [13]. It was demonstrated cisplatin had strong connectivity in the drug-disease-gene network, highlighting the reliability and accessibility of the information stored in GutUDB. In the “DISEASES” module, users can take deeper insight into



**FIGURE 1** (See caption on next page).

genes of interest associated with various intestinal disorders across different omics levels (Supporting Information: Figure S1F). In the “SPECIES” module, users can browse all genes and select species they are interested in (Supporting Information: Figure S1G). All results were presented in a tabular format and facilitated users to efficiently retrieve and filter them by inputting keywords (e.g., a gene symbol or a specific type of intestinal disease) or by clicking on terms (e.g., omics level or hot genes) at the top of the current page.

### Browsing spatial omics and single-cell omics data of intestinal diseases

Recently, spatial omics and single-cell sequencing methods have significantly advanced the study of cellular heterogeneity, immune regulation, and molecular mechanisms in intestinal diseases. The spatial omics data in GutUDB revealed intricate spatial expression maps delineating tumor-specific genes across diverse samples, along with annotation maps. Upon selecting a gene, users could peruse gene-related information of spatial omics data in GutUDB. Furthermore, we also provided details about sample, tissue type, biotechnology, and the gene's expression profiles of spatial omics data in GutUDB. The homepage of gene expression showed the uniform manifold approximation and projection plots of single-cell RNA-seq, and each sample's detail page contained the expression maps of the genes in different cells (Figure 1F). Besides, single-cell gene expression, single-cell alternative polyadenylation, single-cell alternative splicing, and single-cell proteomics data related to intestinal diseases were also embedded in GutUDB, facilitating users to comprehensively understand regulatory mechanisms from different dimensions such as RNA and proteins at the single-cell level.

### Interactive visualization of bulk multiomics profiles related to intestinal diseases

To facilitate the integration and analysis of diverse data sets and data types, we have combined diverse omics data types with the complex interplay among DNA, RNA, proteins, and other aspects of genetic turbulence. Users can assess a specific omics type through “Omics” option in the navigation bar. Eight different types of bulk omics were integrated into GutUDB for interactive visualization.

Epigenetics plays a crucial role in the development and progression of intestinal diseases, and researchers have screened and utilized epigenetic molecules as diagnosis and prognosis biomarkers in clinical trials [14]. In GutUDB, three mainly epigenomics were embedded, including DNA methylation, histone modification (H3K27me3, H3K27ac, H3K36me3, H3K4me1, H3K4me3, and H3K9me3), and chromosome structure. Among them, the transcriptional and posttranscriptional regulatory mechanisms involving RNA m<sup>6</sup>A modification and alternative splicing, are the focal points of current research. Here we demonstrate the histone modification status, specifically the H3K27me3 modification, using *MUTYH* as an example (Figure 1G).

Based on miCLIP-seq, 1908 RNA m<sup>6</sup>A modification sites were included in GutUDB. As for genomics, alterations can influence RNA modification and result in gene transcript level alteration, which may consequently affect the protein expression levels. Additionally, six types of alternative splicing were also incorporated into the GutUDB database, including exon skipping, alternative 5' splice site, alternative 3' splice site, mutually exclusive exons, and retained introns [15]. Here, we use *MUTYH* as an example to demonstrate the status of alternative splicing and m<sup>6</sup>A modification in CRC (Figure 1G).

For genomics, GutUDB currently includes 72,248 single-nucleotide polymorphisms (SNPs), 54,131 CNVs,

**FIGURE 1** Gut Universe Database (GutUDB) contents and features. (A) The process of data collection and processing, as well as the construction procedure of the database website, incorporating four core functions (browsing, querying, visualizing, and downloading). (B) The expression levels of the *METTL3* in different tissues. The data derived from GTEx. (C) Boxplot displaying gene expression difference between COAD and healthy tissues.  $p < 0.05$  indicated that the expression of *METTL3* was significantly different between the two groups. (D) Scatter plot showing the correlation analysis of *METTL3* expression between COAD group and healthy group.  $r = -0.02$  indicated that *METTL3* expression almost no correlation between the two groups. (E) Survival analysis curve of *POLE* associated with COAD. (F) The expression levels of *CNPY3* in various cells and the presentation of expression maps in scRNA-seq data. (G) Tracks displaying the read coverage of H3k27me3, the m<sup>6</sup>A sites as well as RNA-seq and SNP data in colorectal cancer on the *MUTYH*. B, B lymphocyte; COAD, colon adenocarcinoma; CSS, cascading style sheets; DEG, differential expression analysis; GEO, Gene Expression Omnibus; GTAK, Genome Analysis Toolkit; GTEx, Genotype-Tissue Expression project; HR, hazard ratio; HTML, HyperText Markup Language; m<sup>6</sup>A, N<sup>6</sup>-methyladenosine; MySQL, My Structured Query Language; NK, natural killer cell; Plasma, Plasma cell; PCA, principal component analysis; r, correlation coefficient; Sass, Syntactically Awesome Stylesheets; SNP, single-nucleotide polymorphism; T, T lymphocyte; TCGA, The Cancer Genome Atlas; TPM, Transcripts Per Million.

1097 structural variations of chromosome (SVs), and 92,888 mutated genes related to different intestinal diseases, details were established on the “Statistics” page. Notably, each genomic alternation was accompanied by its frequency in various populations as well as data set source. Here, we present the SNP sites on the *MUTYH* gene in CRC (Figure 1G).

For transcriptomics, 137 RNA-seq data sets related to 7.9 million transcriptional profiles with distinct gene expression patterns under various conditions or in different tissues were collected (Supporting Information: Figure SIH). Besides, noncoding RNAs associated with intestinal diseases were separately displayed as panels, including 62 circular RNAs, 182 long noncoding RNAs, and 58 microRNAs.

For proteomic and metabolomic profiles, aberrations significantly have impact on the pathophysiology of intestinal diseases [16, 17]. In the GutUDB, the proteomics data provided information of protein characteristics including functional domains, active sites, and post-translational modifications under artificial regulation. Moreover, data from metabolomics and microbiomics demonstrated 2764 relationships between gut microbes and metabolites in GutUDB.

Therefore, GutUDB enabled researchers to conduct a deep study into the complex interconnections of genetic modifications from multiple biological levels and insights, thus unveiling the sophisticated processes underlying gene regulation.

## DISCUSSION AND FUTURE DIRECTION

In this study, the main challenge was the standardization of integrating sequencing data across different platforms, such as RNA-seq and microarray data. Admittedly, there are variations in sequencing platforms and instruments. Our database GutUDB mainly utilizes RNA-seq data for bulk transcriptomics due to it can still be standardized and compared effectively. Integrating diverse omics data, including transcriptomics, can illuminate pathogenic changes and significantly enhance our understanding of disease diagnosis, mechanisms, and treatment strategies [18]. GutUDB will definitely serve as a comprehensive resource for a wide range of users, such as clinicians specializing in gastroenterology, researchers in academic and scientific institutions, educators and students from universities and anyone with an interest in intestinal research and clinical applications. Users can easily get access to an extensive array of omics data related to intestinal diseases directly through GutUDB, without exhaustive searches for disease-specific information across various databases, such as noncoding RNA databases, spatial omics repositories, and microbiome databases.

With the rapid accumulation of diverse sequencing data, we are responsible for frequently updating GutUDB, and proudly announce that the GutUDB is dedicated to serving as an open-access resource for the global community to advance the study of intestinal diseases. We also plan to integrate interactive models into GutUDB within the next 1–2 years, facilitating real-time communication between users on the platform. We will also integrate a more comprehensive collection of pathological slides on intestinal diseases, radiomics of intestinal diseases, and cohorts of individuals with intestinal diseases into GutUDB. All in all, these procedures will ensure GutUDB obtains the newest and latest extensive multiomics database for intestinal diseases.

## AUTHOR CONTRIBUTIONS

Sanqi An, Shuaiyi Liang, HaoLiang, and Xuena Chen conceived and designed the study, revised the manuscript, and prepared it for publication. Yi Bao, Yaxin Chen, Lizhu Lin, and Jingyi Li carried out the implementation and supervision of the project, collected data, analyzed the results, built the website, interpreted the results, and wrote the manuscript. Gang Wang and Xinli Liu performed the data collection and analysis. Yueqi Li and Yao Lin assisted in database development and application. Yajing Chen and Lijuan Zhou organized and reviewed the omics data. Yawen Qi and Yufang Xie excluded and corrected the collected data. Zhenrui Lin, Zhe Sun, Yuwen Fan, Jinjing Jiang, and Feiyu Zhang collected and organized part of the omics data. Hubin Chen, Jiemei Chu, and Jiegang Huang conducted the search and integration of relevant literatures. All authors have read the final manuscript and approved it for publication.

## ACKNOWLEDGMENTS

The authors would like to thank Wanan Wang and Daowen Yang from Baidu Health Team for their contributions to the construction of the database website. This work was supported by the National Natural Science Foundation of China (82160389), the Guangxi Medical University Training Program for Distinguished Young Scholars to Sanqi An, and the Guangxi Science and Technology Base and Talent Project (2022AC19006).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

All data sources analyzed in this study are included in the Supporting Information (Table S1). GutUDB is freely available at <https://intestine.splicedb.net>. The code involved in this research and the data tables corresponding to the figures in the article have been uploaded to GitHub <https://github.com/Ansanqi/GutUDB>. Supporting Information

(methods, figures, tables, scripts, graphical abstract, slides, videos, Chinese translated version, and updated materials) may be found in the online DOI or iMeta Science <http://www.imeta.science/>.

## ETHICS STATEMENT

No animals or humans were involved in this study.

## ORCID

Xuena Chen  <http://orcid.org/0009-0005-2065-9536>

Sanqi An  <http://orcid.org/0000-0002-3177-213X>

## REFERENCES

- Houtekamer, Ronja M., Mirjam C. van der Net, Madelon M. Maurice, and Martijn Goerich. 2022. "Mechanical Forces Directing Intestinal Form and Function." *Current Biology* 32: R791–805. <https://doi.org/10.1016/j.cub.2022.05.041>
- Hasin, Yehudit, Marcus Seldin, and Aldons Lusis. 2017. "Multi-Omics Approaches to Disease." *Genome Biology* 18: 83. <https://doi.org/10.1186/s13059-017-1215-1>
- Fiocchi, Claudio. 2023. "Omics and Multi-Omics in IBD: No Integration, No Breakthroughs." *International Journal of Molecular Sciences* 24: 14912. <https://doi.org/10.3390/ijms241914912>
- Kang, Mingon, Euseong Ko, and Tesfaye B. Mersha. 2022. "A Roadmap for Multi-Omics Data Integration Using Deep Learning." *Briefings in Bioinformatics* 23(1): bbab454. <https://doi.org/10.1093/bib/bbab454>
- Baysoy, Alev, Zhiliang Bai, Rahul Satija, and Rong Fan. 2023. "The Technological Landscape and Applications of Single-Cell Multi-Omics." *Nature Reviews Molecular Cell Biology* 24: 695–713. <https://doi.org/10.1038/s41580-023-00615-w>
- Bhattacharya, Rahul, Zachary Blankenheim, Patricia M. Scott, and Robert T. Cormier. 2022. "CFTR and Gastrointestinal Cancers: An Update." *Journal of Personalized Medicine* 12: 868. <https://doi.org/10.3390/jpm12060868>
- Kim, Hyun-A., Bon-Kyoung Koo, Ji-Hoon Cho, Yoon-Young Kim, Jinwoo Seong, Hee Jin Chang, Young Min Oh, et al. 2012. "Notch1 Counteracts WNT/ $\beta$ -catenin Signaling Through Chromatin Modification in Colorectal Cancer." *Journal of Clinical Investigation* 122: 3248–59. <https://doi.org/10.1172/jci61216>
- Nikonova, Anna S., Alexander Y. Deneka, Flaviane N. Silva, Shabnam Pirestani, Rossella Tricarico, Anna A. Kiseleva, Yan Zhou, et al. 2023. "Loss of Pkd1 Limits Susceptibility to Colitis and Colorectal Cancer." *Oncogenesis* 12: 40. <https://doi.org/10.1038/s41389-023-00486-y>
- Casey, Graham. 2005. "Conversion Analysis for Mutation Detection in MLH1 and MSH2 in Patients With Colorectal Cancer." *JAMA* 293: 799. <https://doi.org/10.1001/jama.293.7.799>
- Sharma, Deepika, Ankit Malik, Clifford S. Guy, Rajendra Karki, Peter Vogel, and Thirumala-Devi Kanneganti. 2018. "Pyrin Inflammasome Regulates Tight Junction Integrity to Restrict Colitis and Tumorigenesis." *Gastroenterology* 154: 948–964.e8. <https://doi.org/10.1053/j.gastro.2017.11.276>
- Michailidi, Christina, Stamatios Theocharis, Gerasimos Tsourouflis, Vasiliki Pletsas, Gregorios Kouraklis, Efstratios Patsouris, Athanasios G. Papavassiliou, and Constantinos Troungos. 2015. "Expression and Promoter Methylation Status of hMLH1, MGMT, APC, and CDH1 Genes in Patients With Colon Adenocarcinoma." *Experimental Biology and Medicine* 240: 1599–605. <https://doi.org/10.1177/1535370215583800>
- Jasperson, Kory W., Thérèse M. Tuohy, Deborah W. Neklason, and Randall W. Burt. 2010. "Hereditary and Familial Colon Cancer." *Gastroenterology* 138: 2044–58. <https://doi.org/10.1053/j.gastro.2010.01.054>
- Köberle, Beate, and Sarah Schoch. 2021. "Platinum Complexes in Colorectal Cancer and Other Solid Tumors." *Cancers* 13: 2073. <https://doi.org/10.3390/cancers13092073>
- Leclerc, Daniel, Nancy Lévesque, Yuanhang Cao, Liyuan Deng, Qing Wu, Jasmine Powell, Carmen Sapienza, and Rima Rozen. 2013. "Genes With Aberrant Expression in Murine Preneoplastic Intestine Show Epigenetic and Expression Changes in Normal Mucosa of Colon Cancer Patients." *Cancer Prevention Research* 6: 1171–81. <https://doi.org/10.1158/1940-6207.Capr-13-0198>
- An, Sanqi, Yueqi Li, Yao Lin, Jiemei Chu, Jinming Su, Qiuli Chen, Hailong Wang, et al. 2021. "Genome-Wide Profiling Reveals Alternative Polyadenylation of Innate Immune-Related mRNA in Patients With COVID-19." *Frontiers in Immunology* 12: 756288. <https://doi.org/10.3389/fimmu.2021.756288>
- Cai, Jie, Lulu Sun, and Frank J. Gonzalez. 2022. "Gut Microbiota-Derived Bile Acids in Intestinal Immunity, Inflammation, and Tumorigenesis." *Cell Host & Microbe* 30: 289–300. <https://doi.org/10.1016/j.chom.2022.02.004>
- Zhou, Ge-Yu-Jia, Dong-Yan Zhao, Teng-Fei Yin, Qian-Qian Wang, Yuan-Chen Zhou, and Shu-Kun Yao. 2023. "Proteomics-Based Identification of Proteins in Tumor-Derived Exosomes as Candidate Biomarkers for Colorectal Cancer." *World Journal of Gastrointestinal Oncology* 15: 1227–40. <https://doi.org/10.4251/wjgo.v15.i7.1227>
- Pan, Yi, Xiujuan Lei, and Yuchen Zhang. 2021. "Association Predictions of Genomics, Proteinomics, Transcriptomics, Microbiome, Metabolomics, Pathomics, Radiomics, Drug, Symptoms, Environment Factor, and Disease Networks: A Comprehensive Approach." *Medicinal Research Reviews* 42: 441–61. <https://doi.org/10.1002/med.21847>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1:** Web display of content from other sections.

**Table S1:** Statistical source data.

**How to cite this article:** Bao Yi, Yaxin Chen, Lizhu Lin, Jingyi Li, Xinli Liu, Gang Wang, Yueqi Li, et al. 2024. "GutUDB: A Comprehensive Multiomics Database for Intestinal Diseases." *iMeta* 3, e195. <https://doi.org/10.1002/imt2.195>