

# OrganoidDB: a comprehensive organoid database for the multi-perspective exploration of bulk and single-cell transcriptomic profiles of organoids

Qinfeng Ma<sup>†</sup>, Haodong Tao<sup>†</sup>, Qiang Li, Zhaoyu Zhai, Xuelu Zhang, Zhewei Lin, Ni Kuang and Jianbo Pan<sup>\*</sup>

Center for Novel Target and Therapeutic Intervention, Institute of Life Sciences, Chongqing Medical University, Chongqing 400016, China

Received July 23, 2022; Revised September 24, 2022; Editorial Decision October 06, 2022; Accepted October 11, 2022

## ABSTRACT

Organoids, three-dimensional *in vitro* tissue cultures derived from pluripotent (embryonic or induced) or adult stem cells, are promising models for the study of human processes and structures, disease onset and preclinical drug development. An increasing amount of omics data has been generated for organoid studies. Here, we introduce OrganoidDB ([http://www.inbirg.com/organoid\\_db/](http://www.inbirg.com/organoid_db/)), a comprehensive resource for the multi-perspective exploration of the transcriptomes of organoids. The current release of OrganoidDB includes curated bulk and single-cell transcriptome profiles of 16 218 organoid samples from both human and mouse. Other types of samples, such as primary tissue and cell line samples, are also integrated to enable comparisons with organoids. OrganoidDB enables queries of gene expression under different modes, e.g. across different organoid types, between different organoids from different sources or protocols, between organoids and other sample types, across different development stages, and via correlation analysis. Datasets and organoid samples can also be browsed for detailed information, including organoid information, differentially expressed genes, enriched pathways and single-cell clustering. OrganoidDB will facilitate a better understanding of organoids and help improve organoid culture protocols to yield organoids that are highly similar to living organs in terms of composition, architecture and function.

## INTRODUCTION

Organoids are three-dimensional *in vitro* tissue cultures that are derived from pluripotent (embryonic or induced)

or adult stem cells or even patient-specific tissue samples (1). Organoids can be directed to differentiate into a variety of cell types with organ characteristics and have histological organization similar to and sometimes indistinguishable from that of real organs. Furthermore, their formation recapitulates the self-organizing process of organ growth (2). Therefore, organoids are more promising models than two-dimensional cell lines for the *in vitro* study of human processes and structures, disease onset and preclinical drug development. Moreover, compared to animal models, organoids can be generated with higher efficiency and speed and provide a more accurate representation of human tissues (3). However, certain caveats in the current applications of organoids remain. For example, organoids may not fully represent all cell types in a primary tissue and typically lack immune and vasculature cells (4).

To fully realize the potential of organoid technology in the field of disease research and drug development, the key problem that urgently needs to be solved is to verify that organoids can faithfully reflect the biological process in the human body, which involves cell composition, differentiation, and states and the standard definitions of responses to external stimuli, and describe these patterns. In addition, developing a reliable culture protocol for organoids that can more accurately simulate the *in vivo* environment will also become a future research direction. Therefore, comparing organoids with corresponding human tissues using high-throughput sequencing technologies, especially single-cell sequencing, will provide a powerful method for assessing the quality of organoids (5). The ongoing Organoid Cell Atlas project was proposed in 2021, dedicated to exploring the single-cell characterization of organoids (6). Moreover, an increasing amount of transcriptomic data has been generated for organoid studies (7). However, to the best of our knowledge, there is currently no public organoid-related database or a centralized data portal of transcriptome profiles from public organoid studies.

<sup>\*</sup>To whom correspondence should be addressed. Tel: +86 23 684 80209; Fax: +86 23 684 80209; Email: [panjianbo@cqmu.edu.cn](mailto:panjianbo@cqmu.edu.cn)

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Here, we present OrganoidDB (<http://www.inbirg.com/organoid.db/>), a comprehensive transcriptome data (microarray, bulk RNA-seq and single-cell RNA-seq) resource for organoids. In brief, we manually collected and annotated transcriptome profiles of human and mouse organoid samples from GEO and ArrayExpress. We also integrated other types of samples, including primary tissues, cell lines and xenografts, to enable comparisons with organoids. We then cleaned and analysed the raw data through a standard pipeline and presented the integrated data on our user-friendly web application. OrganoidDB enables users to search for organoid transcriptome expression data at both the bulk and single-cell levels under six different modes, i.e. ‘General’, ‘Organoid Comparison’, ‘Organoid Development’, ‘Organoid Specificity’, ‘Tissue Specificity’ and ‘Correlation Analysis’. OrganoidDB also allows users to browse organoid samples and datasets with corresponding organoid culture information, differentially expressed genes (DEGs), enriched functions/pathways, differential cell type markers, and single-cell analysis results such as cell clustering and marker gene results when available. OrganoidDB could not only facilitate a better understanding of organoids but also help improve organoid culture protocols to yield organoids that fully recapitulate the structure of the modelled organs.

## MATERIALS AND METHODS

### Workflow of OrganoidDB

The workflow of OrganoidDB is presented in Figure 1. The data and methods employed to construct the database are described below.

### Data collection and curation

We retrieved high-throughput omics data related to human and mouse organoid studies from GEO (8) and ArrayExpress (9) by searching the keyword ‘organoid’. After careful manual collection, transcriptomic data (microarray, bulk RNA-seq and single-cell RNA-seq) from these studies were included in our database. The metadata, including organoid type, organoid source, and culture protocol, among other variables, were collected for annotation of the samples. It should be noted that organoid sample types are tagged with a single type identity, such as brain, cerebral cortex, duodenum, intestine, etc. Type tags are noninclusive such that searching for brain organoids will not include cerebral cortex organoids and vice versa. We also integrated other types of samples, including samples of primary tissues, cell lines and xenografts, to enable comparisons to be made with organoids. Then, for the microarray data, we used the GEOquery R package (10) to obtain the platform information and the expression profile data of the corresponding samples. For the RNA-seq data, we obtained the download links for the fastq.gz files of these samples in the ENA database (11) and used IBM Aspera software (<https://www.ibm.com/products/aspera>) to perform a high-speed download. For samples without available fastq.gz files, we downloaded the processed expression data from the GEO database. In addition, to investigate gene specificity in different tissues, we downloaded bulk RNA-seq and single-cell

RNA-seq data from the GTEx (12) database, the HPA (13) database and two published studies (14,15) for human and mouse tissues to complement our database.

### Data processing pipeline

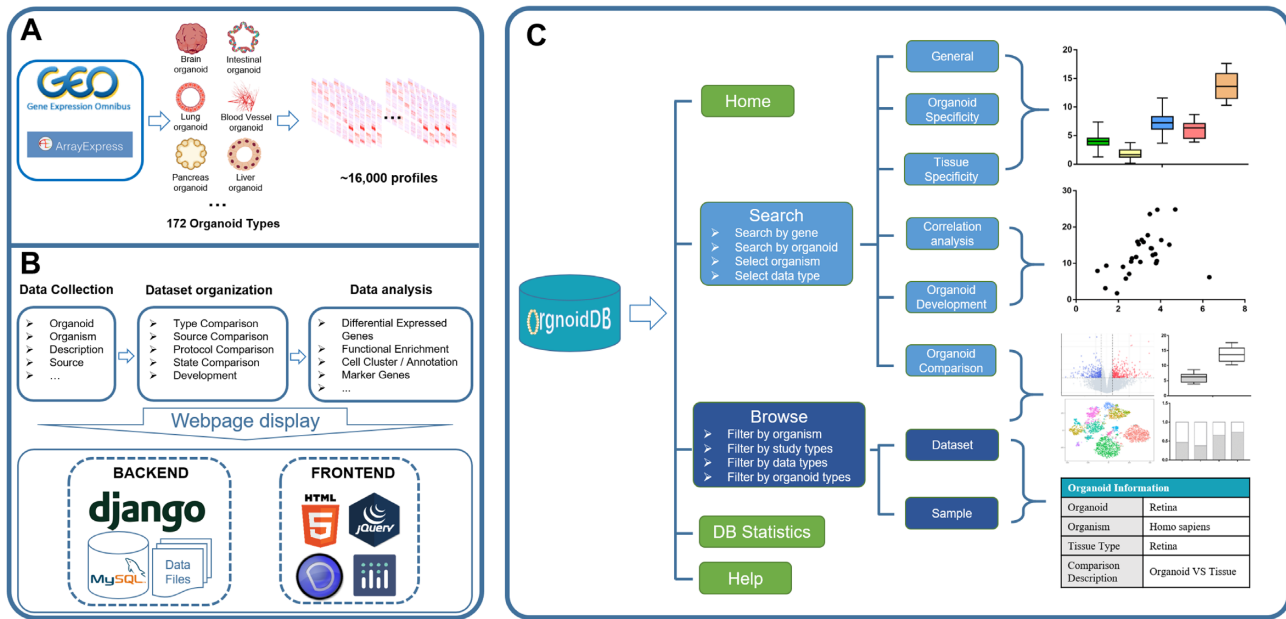
We have built a series of uniform pipelines to process these data. For microarray data, we used the GEOquery R package to obtain the expression matrix of the samples from GEO, and the platform annotation file (GPL) was also obtained. According to the GPL files, the probes were uniformly converted to Entrez IDs, and for multiple probes that were mapped to a single Entrez ID, the average value was used to represent the expression of this gene.

For bulk RNA-seq data, the downloaded fastq.gz files were first checked for md5 values to ensure that the data were error-free during transmission. Then, standard strict read quality control (QC) was performed by fastp (v0.23.1) (16). Subsequently, these clean data were mapped to the reference genome (human: GCF\_000001405.39.GRCh38.p13; mouse: GCF\_000001635.26.GRCm38.p6) by using HISAT2 (v2.2.1) (17) to generate SAM files. After that, we used the featureCounts tool of subread (v2.0.1) (18) software to count the reads aligned to each gene. The read counts were normalized using the transcript per million (TPM) method to enable comparison of expression values between samples.

The single-cell RNA-seq data were checked and subjected to QC as described above. Different pipelines were developed for these data that were obtained from 4 different construction protocols, namely, 10X Genomics, Smart-seq2, Drop-seq and CEL-seq2 (19,20). For data from 10X Genomics, we used Cell Ranger (v6.0.2) and its built-in reference genomes (human: refdata-gex-GRCh38-2020-A; mouse: refdata-gex-mm10-2020-A) to identify the barcode of each cell as well as the unique molecular identifier (UMI) and construct a matrix of UMI counts for each sample. For data from Smart-seq2, the HISAT2 + featureCounts pipeline was used to quantify gene expression in each cell. For both the Drop-seq and CEL-seq2 datasets, STAR (v2.7.10a) (21) software was used for alignment, and the data were demultiplexed to construct a matrix of UMI counts for each sample using the published pipelines dropseqRunner (<https://github.com/aselewa/dropseqRunner>) and celseq2 (22), respectively. After constructing the expression matrix, a series of analyses were performed using Seurat (v4.1.1) (23), such as TSNE and UMAP clustering and identification of the marker genes for different clusters or groups. We used SCTransform (v0.3.3) (24) to normalize the count matrix and Harmony (v0.1.0) (25) to integrate the samples. The annotation of cell types was performed by scType (26). All analyses were based on default parameters used to build a relatively standard analysis pipeline.

### Differential expression analysis

Comparing the differences between organoids and other types of samples is a common task of OrganoidDB, and screening for DEGs allows inference of the molecular genetic characteristics of specific organoids. Differential analysis of microarray data was performed based on limma



**Figure 1.** Overview of OrganoidDB. (A) Data were manually collected from GEO and ArrayExpress. (B) The data are annotated, integrated and analysed and presented to users via a web application interface; the application backend is built on Django and a hybrid MySQL/datafile data storage system, while the frontend is built on HTML5 and jquery libraries. (C) Functions including search for genes and browse for datasets/samples are provided in OrganoidDB. Various interactive plots are also created for data visualization.

(v3.52.1) (27), while bulk RNA-seq data were analysed by using DESeq2 (v1.36.0) (28). For the RNA-seq datasets lacking biological replicates, analysis was performed by manually setting the biological coefficient of variation (bcv, human:  $bcv = 0.4$ ; mouse:  $bcv = 0.1$ ) using edgeR (v3.38.1) (29). The significance of each DEG was characterized by Student's *t* test, and *P*-values were adjusted based on the Benjamini–Hochberg (BH) multiple testing correction method. For datasets with larger sample sizes, such as single-cell RNA-seq datasets and datasets from the organoid specificity module, the Wilcoxon rank-sum test was used to characterize significance, and the false discovery rate (FDR) method was used to adjust the *P*-values.

### Enrichment analysis

OrganoidDB also supports functional enrichment analysis of DEGs, including GO analysis, KEGG pathway analysis, gene set enrichment analysis (GSEA) and cell marker enrichment. For functional enrichment analysis, we downloaded annotation files from GO (30) and KEGG (31) to build local libraries, and for cell marker enrichment analysis, we integrated two databases, CellMarker (32) and PanglaoDB (33), to build local libraries. All enrichment analyses were performed based on the most significant 2000 DEGs (adjusted *P*-values < 0.05), except for GSEA, which was based on the ranking of fold changes of all expressed genes. The enrichment analysis was performed by using hypergeometric tests to calculate the *P*-values, which were adjusted using the BH multiple testing correction method. These analyses and the visualization of the results were performed on clusterProfiler (v4.0) (34).

### Database implementation

OrganoidDB is hosted with NGINX and uWSGI in a CentOS environment on the Alibaba cloud. The database is built on a Django stack with a hybrid MySQL/filesystem data storage system. The front end was rendered with HTML5, bootstrap, and jquery libraries. Interactive data tables are rendered with the datatables.js library. All interactive plots are rendered with the Plotly.js library.

## CONTENTS AND FEATURES OF ORGANOIDDB

### Database overview

Currently, OrganoidDB provides transcriptome data from 16 218 organoid samples, including 12 911 human organoid samples and 3,307 mouse organoid samples (Table 1 and DB Statistics page). These samples cover 172 different organoid types, including intestinal organoids, brain organoids, lung organoids and so on. The samples are organized into 1,069 datasets, including 145 single-cell RNA-seq datasets, in five major categories—‘type comparison’, ‘state comparison’, ‘protocol comparison’, ‘source comparison’ and ‘organoid development’. These five categories describe biological contexts as follows: ‘type comparison’ compares organoid samples with other sample types, such as tissue, primary cells, and xenografts; ‘state comparison’ compares organoids of different states, such as disease versus normal state, proliferative vs. quiescent state, and adult versus fetal state; ‘protocol comparison’ compares organoids grown under different growth conditions, such as differentiation medium vs. expansion medium, co-culture versus monoculture, and Matrigel versus collagen and ‘source comparison’ compares organoids grown

**Table 1.** Statistics of OrganoidDB

Function/term	Human bulk	Human single-cell	Mouse bulk	Mouse single-cell
General	722 S <sup>a</sup>	449 529 C	322 S	221 261 C
Organoid specificity	975 S	172 149 C	332 S	67 773 C
Tissue specificity	17 382 S	566 108 C	72 S	53 760 C
Organoid development	39 D	9 D	8 D	2 D
Organoid comparison	630 D	115 D	247 D	19 D
type comparison	420 D	68 D	188 D	14 D
state comparison	123 D	21 D	9 D	0 D
protocol comparison	71 D	14 D	42 D	4 D
source comparison	16 D	12 D	8 D	1 D
Organoid samples	7 243 S	5 668 S	2 953 S	354 S
Organoid types	119 T	66 T	67 T	19 T

<sup>a</sup> 'S', 'C', 'D' and 'T' indicate 'samples', 'cells', 'datasets' and 'types', respectively.

from different sources, such as adult stem cells (ASCs) versus pluripotent stem cells (PSCs), induced pluripotent stem cells (iPSCs) versus embryonic stem cells (ESCs) and patient-derived organoids (PDOs) versus patient-derived xenograft organoids (PDXOs). Finally, 'organoid development' calculates the correlation for each gene between gene expression values and organoid developmental time points.

The homepage and DB statistics page give an overview of our samples and datasets. The help page provides a guide on OrganoidDB functionalities, while the search, browse, visualization functionalities, user feedback and data submission are described in the sections below.

### Search functions

On the search page, OrganoidDB allows users to perform multi-perspective exploration of the gene expression of organoids and other sample types under six modes at both the bulk and single-cell RNA levels in either human or mouse (Figure 2A). *General*. The 'general' mode allows users to search for an overview of gene expression across different organoids compared with primary tissues, cell line cultures, PSC cultures and spheroids. This module contains 1044 bulk RNA samples and 126 single-cell samples with ~670K cells (Table 1). For bulk RNA, searching for a gene will display box-plot comparison of its expression across the different sample types. For single-cell, the search result will display violin-plot comparison as well as a bar-plot of the percentage of cells a gene is expressed across the sample types. For more in-depth visualization, users can use the organoid comparison mode to navigate our comparison datasets or organoid/tissue specificity modes to identify organoid/tissue-specific genes.

*Organoid comparison*. The 'organoid comparison' mode compares gene expression in organoid samples with that in tissues, cell lines, and xenografts and compares organoids derived from different sources or different protocols. This mode provides users with access to four different categories of DEGs as described above, i.e. 'type comparison', 'state comparison', 'protocol comparison' and 'source comparison'. These four categories contain 690, 153, 131 and 37 datasets, respectively. Users can search for the gene of in-

terest to query datasets in which the gene is differentially expressed. OrganoidDB also provides comparison queries based on organoid types. Users can identify the DEGs of a certain organoid type under different conditions. In addition, users will be led to a detailed information page by clicking the dataset ID, as described below.

*Organoid development*. The 'organoid development' mode enables the identification of genes correlated with development by calculating Spearman's correlations as organoid cultures develop over time. There are 58 organoid development datasets in total. In this mode, users can identify development-correlated genes by searching for genes of interest across all our datasets or find genes with the highest Spearman's correlation in specific organoids.

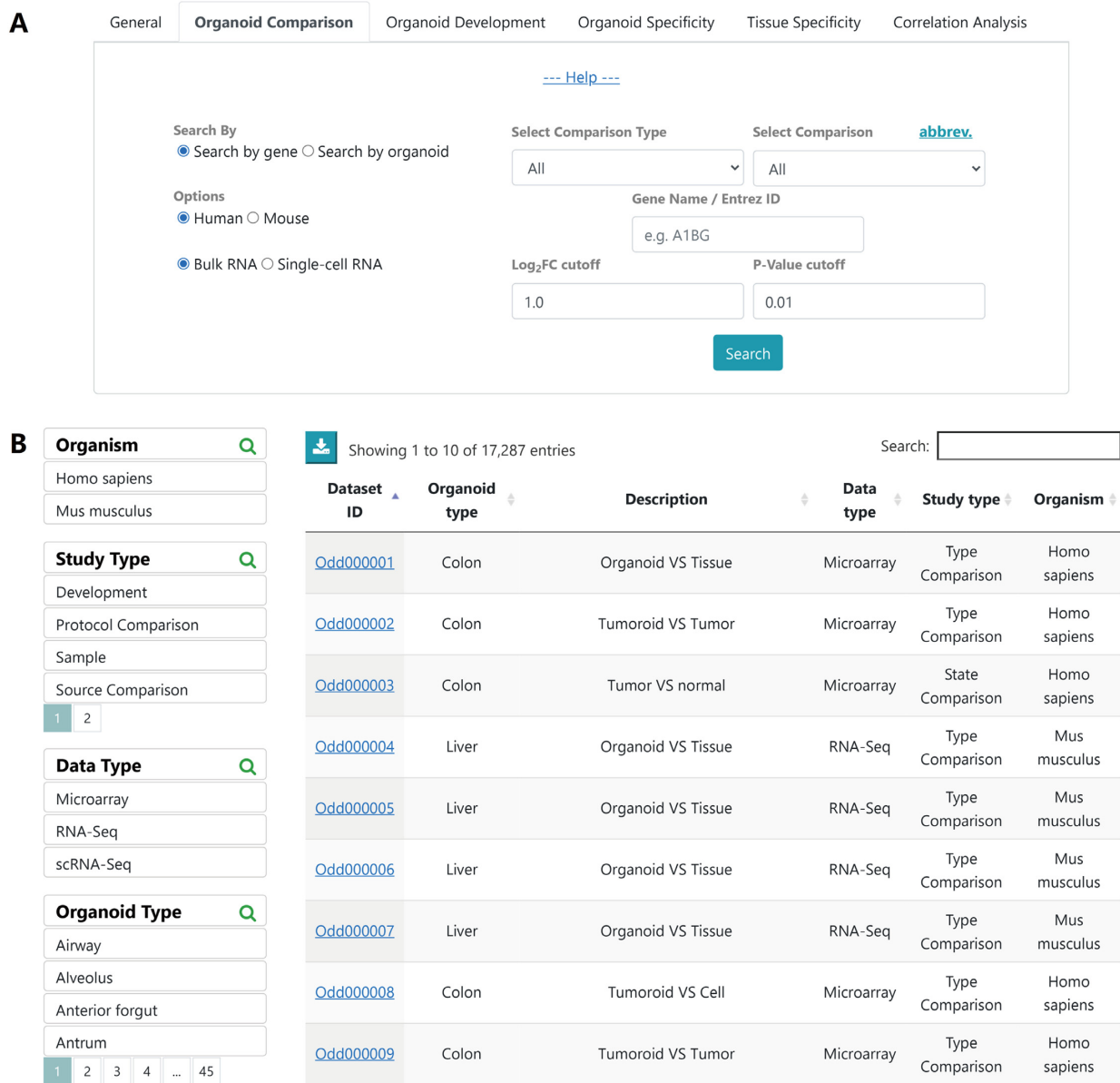
*Organoid specificity*. This mode will help identify organoid-specific genes in various organoid types. For bulk RNA-seq samples, we calculated specificity across 36 different human organoids and 15 different mouse organoids. For single-cell RNA-seq samples, we calculated specificity across 17 different human organoids and 11 different mouse organoids. Organoid specificity was measured through comparison between the given organoid type and the remaining organoid types using differential expression analysis. Fold change and *P*-values were used to estimate organoid specificity. Users can search for genes of interest across all organoid types to see if they have high specificity for certain organoids, or users can list all genes for an organoid type and filter them on the basis of fold change and *P*-values to identify the genes most specific to an organoid type. For bulk samples, box plots are shown to visualize differential expression across organoid types. For single-cell samples, bar plots are shown to visualize the percentage of cells with expression, and violin plots are shown to visualize the expression value across the organoid types.

*Tissue specificity*. This mode will help identify tissue-specific genes in various tissue types. For bulk RNA samples, we calculated specificity across 30 different human tissue types and 18 different mouse tissue types. For single-cell samples, we calculated specificity across 26 different human tissue types and 20 different mouse tissue types. Tissue specificity was measured through comparison between the given tissue type and the remaining tissue types using differential expression analysis. Fold change and *P*-values were used to estimate the tissue specificity. By providing tissue specificity as a reference, users can see if genes that have high specificity to organoid types also have high specificity in their corresponding tissue types. Similar to the organoid specificity module, box plots, bar plots and violin plots are shown to visualize differential expression.

*Correlation analysis*. Correlation analysis was used to study gene–gene expression correlations across organoid samples or tissue samples. Spearman's correlations between two genes were calculated based on gene expression across samples in the same dataset.

### Browsing interface

On the browse page (Figure 2B), users can browse through organoid datasets and samples using (i) filters such as organoid types, study types and comparison types. (ii) click-



**Figure 2.** Main functions of the OrganoidDB web content. (A) Search for organoid transcriptome expression data under six different modes at both the bulk and single-cell levels in human or mouse. (B) Browse for organoid datasets and samples by selecting the organism, data type, platform or tissue type.

ing the dataset ID will lead to a detailed information page described below.

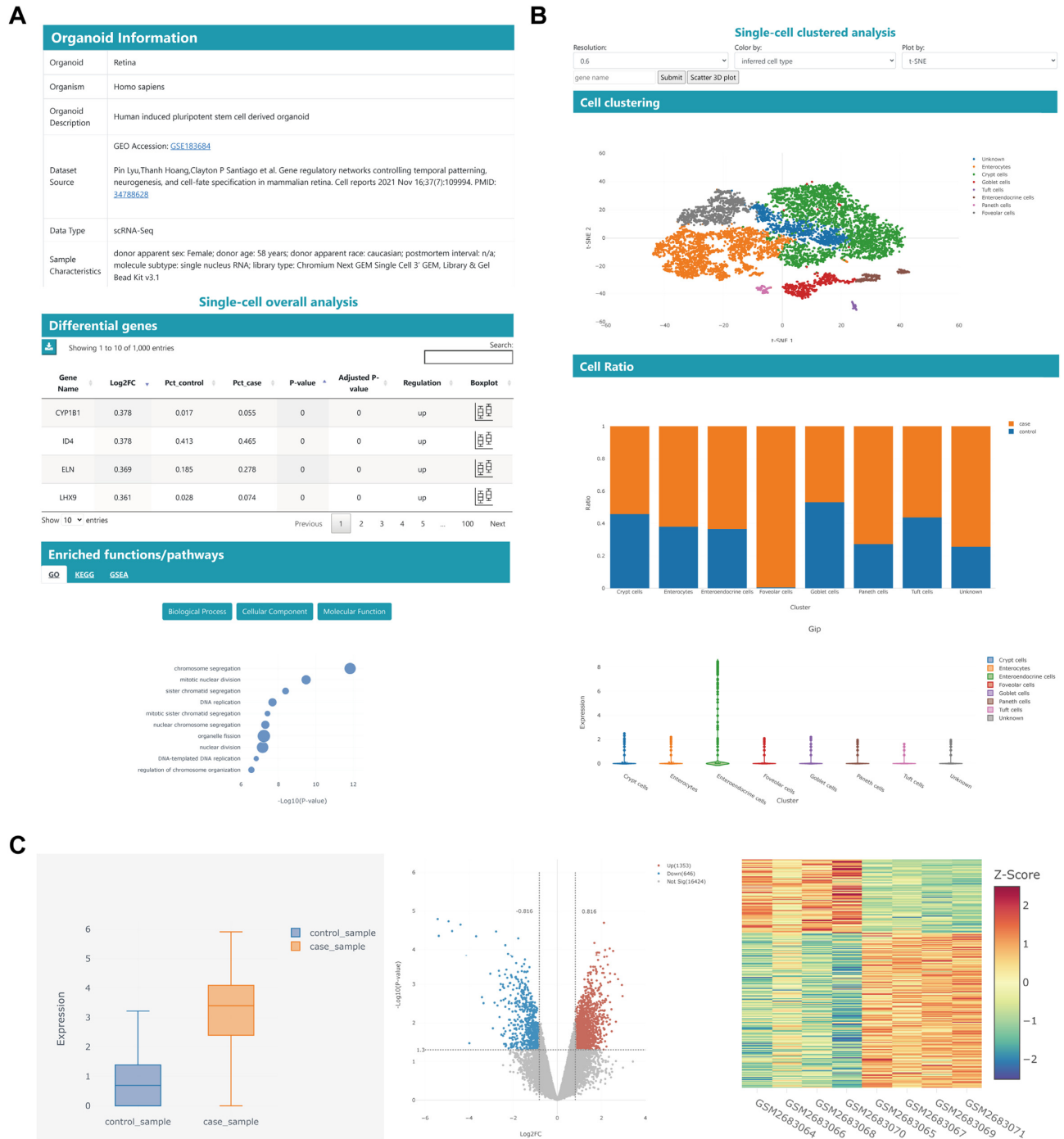
### Detailed information

The detailed information page contains the available organoid culture information for both samples and datasets (Figure 3A). Sample characteristics such as sample capture time and patient demographic information that the sample is derived from are provided. Furthermore, we annotated datasets with the original GSE number as well as publication information and PMID link when available. For bulk RNA datasets, the overall analysis results, including DEGs, enriched functions/pathways, and differential cell type markers, are provided. For single-cell RNA-seq datasets, in addition to the overall analysis, the single-cell

cluster analysis will generate a cell clustering plot with cell type annotation when available, cluster marker genes and DEGs in clusters (Figure 3B). Other interactive plots, including box plots, volcano plots and heatmaps, are also provided for data visualization in the detailed information page (Figure 3C).

### Result availability

Interactive plots are rendered with the Plotly js library in HTML5. Users can customize plot views by zooming and panning to areas of interest, as well as showing or hiding certain plot aspects. All processed data and related meta-data can be downloaded through the data table search results. All interactive plots can be downloaded in PNG and SVG format. Single-cell feature plots can be viewed and



**Figure 3.** Detailed information page of a single-cell RNA-seq dataset example. (A) Users can view detailed information, including organoid information, DEGs and enriched GOs/pathways/cell types. (B) For single-cell cluster analysis results, users can visualize cell clustering, the expression ratio for each cluster, and marker genes and DEGs within clusters. (C) Other interactive plots, including box plots, volcano plots and heatmaps, are also provided for data visualization. Due to space limitations, only selected parts of this page are shown here.

downloaded in the detailed information page of the corresponding single-cell dataset.

### User feedback and data submission

Usability tests are necessary for smooth database operation. Users can help us improve OrganoidDB by filling out a short questionnaire on the Feedback&Submit

page. The questionnaire collects information on the following: user intention, user satisfaction, platform usability, platform usefulness and platform ease of use (35,36). Users can give a short description of how they intend to use our platform and give a rating of whether they are successful in completing their tasks. In addition, users can also describe in detail their suggestions for OrganoidDB.

Users can also inform OrganoidDB of particular datasets of interest through the Feedback&Submit page. Users can submit their name, e-mail, and institution, as well as data source information such as GEO accession, link to the data source along with metadata describing the submission. OrganoidDB will review the submission and automatically inform the submitter when their submission is included in the database.

## CONCLUSIONS AND FUTURE DIRECTIONS

OrganoidDB is a user-friendly interactive database that provides comprehensively curated information and analysis results based on transcriptome data related to organoids from the GEO and ArrayExpress databases. Our database includes microarray, RNA-seq and single-cell RNA-seq data for multiple organ types in human and mouse, including intestinal organoids, brain organoids, and lung organoids. OrganoidDB makes it easy for organoid researchers to explore organoid-related transcriptomic data without the need for computational programming and to conduct their own organoid research or validate their results. Taking advantage of our in-house-built hybrid data storage architecture, OrganoidDB can quickly respond to user input and help users retrieve transcriptome profiles of interest. For example, researchers have found that aquaporin 5 (*AQP5*) is more highly expressed in human lung organoids than lung tissues in the datasets ‘Odd000570’, ‘Odd000657’ and ‘Odd001044’ through the *organoid comparison* module (Supplementary Figure S1A). It has been reported that *AQP5* is highly expressed in lung organoids as a marker of alveolar epithelial cells (37). In addition, users can also apply the *organoid development* module to find that the expression levels of some homeobox (*Hox*) genes and LIM homeobox 1 (*LHX1*) are highly correlated with the developmental time of kidney organoids in the dataset ‘Odd001105’ (Supplementary Figure S1B) (38). Additionally, through ‘protocol comparison’ of the organoid comparison module, users can find that the expression of hepatocyte markers such as *ALB* and *CYP3A4* (39) is generally upregulated in liver organoids in spinner flasks compared to static cultures in the dataset ‘Odd000172’ (Supplementary Figure S1C). Moreover, the GSEA results revealed the enrichment of ‘bile secretion’ (Supplementary Figure S1D), which is one of the major functions of the liver. In addition, cell marker enrichment analysis also indicated that hepatocyte cells might be enriched under spinner conditions (Supplementary Figure S1E). Thus, we can speculate that liver organoids cultured using spinner conditions more closely resemble liver tissue under physiological conditions, which was also confirmed by the previous researchers (40). In summary, OrganoidDB can meet the transcriptome analysis needs of organoid researchers and help improve the understanding of organoids as well as organoid culture protocols.

To the best of our knowledge, OrganoidDB is the first publicly available database that provides an organized resource for researchers to explore transcriptomic differences in organoids in terms of their tissue of origin, treatment protocol, growth protocol, development time, and organoid specificity. OrganoidDB is expected to provide a better un-

derstanding of organoids and help to improve organoid culture protocols to yield organoids that fully recapitulate the structure of the modelled organs. In the future, we will continue to maintain OrganoidDB and make regular updates—new data can be added to OrganoidDB with our standardized pipelines for collecting metadata from GEO and ArrayExpress and processing raw transcriptomic data. We plan to review new GEO and ArrayExpress entries every 6 months and update our database accordingly. Furthermore, from the Feedback&Submit page, users can upload new organoid datasets, and we will update OrganoidDB accordingly after reviewing the submission. User feedback will be reviewed weekly to continue to improve the usability of OrganoidDB and make timely updates. We will also expand the database in the following directions: (i) incorporate more types of high-throughput data, such as genomic and proteomic data, to make OrganoidDB a multiomics resource; (ii) collect and analyse organoid data for drug processing to provide data resources for target discovery and drug development; (iii) build an in-house pipeline to allow users to analyse their data using our standard procedures and to integrate the results with OrganoidDB datasets and (iv) develop a series of analytics tools and algorithms to evaluate the organoid models. We believe that OrganoidDB will benefit the organoid research community and expand the application of this field.

## DATA AVAILABILITY

OrganoidDB is freely available online at [http://www.inbirg.com/organoid\\_db/](http://www.inbirg.com/organoid_db/), and there is no login requirement.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The computing work in this paper was partly supported by the Supercomputing Center of Chongqing Medical University. The authors thank all members of Pan’s lab for beta-testing the website.

## FUNDING

Research Startup Funds of Chongqing Medical University; National Natural Science Foundation of China [82104063]; project of the top-notch talent cultivation program for the graduate students of Chongqing Medical University [BJRC202214]; University Innovation Research Group Project of Chongqing [CXQT21016]; High-Level Innovation Platform Project of Chongqing (No. 14). Funding for open access charge: the Research Startup Funds of Chongqing Medical University.

*Conflict of interest statement.* None declared.

## REFERENCES

- Hofer, M. and Lutolf, M.P. (2021) Engineering organoids. *Nat. Rev. Mater.*, **6**, 402–420.
- Lancaster, M.A. and Knoblich, J.A. (2014) Organogenesis in a dish: modeling development and disease using organoid technologies. *Science*, **345**, 1247125.

3. Kim, J., Koo, B.K. and Knoblich, J.A. (2020) Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.*, **21**, 571–584.
4. Rossi, G., Manfrin, A. and Lutolf, M.P. (2018) Progress and potential in organoid research. *Nat. Rev. Genet.*, **19**, 671–687.
5. Lee, M.O., Lee, S.G., Jung, C.R., Son, Y.S., Ryu, J.W., Jung, K.B., Ahn, J.H., Oh, J.H., Lee, H.A., Lim, J.H. *et al.* (2021) Development of a quantitative prediction algorithm for target organ-specific similarity of human pluripotent stem cell-derived organoids and cells. *Nat. Commun.*, **12**, 4492.
6. Bock, C., Boutros, M., Camp, J.G., Clarke, L., Clevers, H., Knoblich, J.A., Liberali, P., Regev, A., Rios, A.C., Stegle, O. *et al.* (2021) The organoid cell atlas. *Nat. Biotechnol.*, **39**, 13–17.
7. Brazovskaja, A., Treutlein, B. and Camp, J.G. (2019) High-throughput single-cell transcriptomics on organoids. *Curr. Opin. Biotechnol.*, **55**, 167–171.
8. Clough, E. and Barrett, T. (2016) The gene expression omnibus database. *Methods Mol. Biol.*, **1418**, 93–110.
9. Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., Snow, C., Fonseca, N.A., Petryszak, R., Papatheodorou, I. *et al.* (2019) ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Res.*, **47**, D711–D715.
10. Sean, D. and Meltzer, P.S. (2007) GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics*, **23**, 1846–1847.
11. Harrison, P.W., Ahamed, A., Aslam, R., Alako, B.T.F., Burgin, J., Buso, N., Courtot, M., Fan, J., Gupta, D., Haseeb, M. *et al.* (2021) The european nucleotide archive in 2020. *Nucleic Acids Res.*, **49**, D82–D85.
12. GTEx Consortium (2015) The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
13. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P. *et al.* (2021) A single-cell type transcriptomics map of human tissues. *Sci. Adv.*, **7**, eab2169.
14. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y. and Shi, L. (2017) A comprehensive mouse transcriptomic bodymap across 17 tissues by RNA-seq. *Sci. Rep.*, **7**, 4200.
15. Schaum, N., Karkanas, J., Neff, N.F., May, A.P., Quake, S.R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M.B. *et al.* (2018) Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, **562**, 367–372.
16. Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
17. Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-Based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, **37**, 907–915.
18. Liao, Y., Smyth, G.K. and Shi, W. (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
19. Zhang, Y., Zou, D., Zhu, T., Xu, T., Chen, M., Niu, G., Zong, W., Pan, R., Jing, W., Sang, J. *et al.* (2022) Gene expression nebulas (GEN): a comprehensive data portal integrating transcriptomic profiles across multiple species at both bulk and single-cell levels. *Nucleic Acids Res.*, **50**, D1016–D1024.
20. Zhang, Z., Cui, F., Lin, C., Zhao, L., Wang, C. and Zou, Q. (2021) Critical downstream analysis steps for single-cell RNA sequencing data. *Brief. Bioinform.*, **22**, bbab105.
21. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
22. Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O. *et al.* (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.*, **17**, 77.
23. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
24. Choudhary, S. and Satija, R. (2022) Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.*, **23**, 27.
25. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R. and Raychaudhuri, S. (2019) Fast, sensitive, and accurate integration of single cell data with harmony. *Nat. Methods*, **16**, 1289–1296.
26. Ianevski, A., Giri, A.K. and Aittokallio, T. (2022) Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun.*, **13**, 1246.
27. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
28. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
29. Chen, Y., Lun, A.T.L. and Smyth, G.K. (2016) From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, **5**, 1438.
30. The Gene Ontology Consortium (2021) The gene ontology resource: enriching a Gold mine. *Nucleic Acids Res.*, **49**, D325–D334.
31. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. and Tanabe, M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.*, **49**, D545–D551.
32. Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M. *et al.* (2019) CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.*, **47**, D721–D728.
33. Franzén, O., Gan, L.M. and Björkegren, J.L.M. (2019) PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, **2019**, baz046.
34. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L. *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov.*, **2**, 100141.
35. Chin, J.P., Diehl, V.A. and Norman, K.L. (1988) Development of an instrument measuring user satisfaction of the human-computer interface. *Conf. Hum. Factors Comput. Syst. - Proc.*, **Part F130202**, 213–218.
36. DeLone, W.H. and McLean, E.R. (2003) The delone and mclean model of information systems success: a ten-year update. *J. Manag. Inf. Syst.*, **19**, 9–30.
37. Tran, E., Shi, T., Li, X., Chowdhury, A.Y., Jiang, D., Liu, Y., Wang, H., Yan, C., Wallace, W.D., Lu, R. *et al.* (2022) Development of human alveolar epithelial cell models to study distal lung biology and disease. *IScience*, **25**, 103780.
38. Khoshdel Rad, N., Aghdami, N. and Moghadasali, R. (2020) Cellular and molecular mechanisms of kidney development: from the embryo to the kidney organoid. *Front. Cell Dev. Biol.*, **8**, 183.
39. Guo, D.L., Wang, Z.G., Xiong, L.K., Pan, L.Y., Zhu, Q., Yuan, Y.F. and Liu, Z.S. (2017) Hepatogenic differentiation from human adipose-derived stem cells and application for mouse acute liver injury. *Artif. Cells Nanomed. Biotechnol.*, **45**, 224–232.
40. Schneeberger, K., Sánchez-Romero, N., Ye, S., van Steenbeek, F.G., Oosterhoff, L.A., Pla Palacin, I., Chen, C., van Wolferen, M.E., van Tienderen, G., Lieshout, R. *et al.* (2020) Large-Scale production of LGR5-positive bipotential human liver stem cells. *Hepatology*, **72**, 257–270.