Computational and Structural Biotechnology Journal 23 (2024) 2740–2745

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Software/web server article

LiverSCA: A comprehensive and user-friendly cell atlas in human hepatocellular carcinoma

Renwen Long [a,b,1], Suoangbaji [a,b,1], Irene Oi-Lin Ng [a,b], Daniel Wai-Hung Ho [a,b,*]

[a] State Key Laboratory of Liver Research, The University of Hong Kong, Hong Kong
[b] Department of Pathology, School of Clinical Medicine, The University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Liver cancer
ScRNA-seq
Cell atlas

ABSTRACT

We developed a cell atlas named LiverSCA on human liver cancer single-cell RNA sequencing data. It has a user-friendly web interface and comprehensive functionalities aiming to help researchers to make easy access to cellular and molecular landscapes of the tumor microenvironment in liver cancer. LiverSCA includes a complete analytical pipeline that allow mechanistic exploration on a wide variety of functionalities, such as cell clustering, cell annotation, identification of differentially expressed genes, functional enrichment analysis, analysis of cellular crosstalk, and pseudo-time trajectory analysis. Notably, our intuitive web interface allows users, particularly wet-lab researchers, to easily explore and undertake data discovery, without the need to handle any of the raw data.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer, and its incidence is increasing worldwide [1]. It is a deadly disease [2–4] and the major risk factors for HCC are infection with hepatitis B or hepatitis C viruses and excessive alcohol consumption [5]. However, in Western countries, non-alcoholic steatohepatitis associated with metabolic syndrome or diabetes mellitus are becoming a more frequent risk factor [6–8]. HCC tumor consists of an intricate tumor microenvironment (TME) involving the endeavor of malignant cells, immune cells and stromal cells to elicit hepatocarcinogenesis consequence [6,9]. A better understanding of the TME is crucial for identifying immune-related determinants of HCC progression and developing novel HCC immunotherapeutics.

Single-cell RNA sequencing (scRNA-seq) studies in HCC provided important insights into the heterogeneity, dynamics, and potential roles of TME in HCC progression and response to immunotherapies [10–12]. Importantly, sophisticated computational tools have been developed to fit various analytical purposes of scRNA-seq, such as CellRanger [13], Scanpy [14] and Seurat [15]. Nevertheless, computational/programming knowledge is usually the main hurdle for biologists to perform data discovery by utilizing scRNA-seq data or even get access to them. They heavily rely on computational tools that provide graphical user interface [16,17]. Therefore, bioinformaticians collaborate closely with biologists to fully comprehend the data discovery and interpretation of findings regarding scRNA-seq data.

There is an urgent need to provide a truly comprehensive and user-friendly cell atlas in HCC that can be easily accessible and utilized to perform mechanistic exploration. Here, we developed a cell atlas on human HCC called LiverSCA (**Liver S**ingle-**C**ell **A**tlas). It is a user-friendly web-based platform for visualizing and analyzing scRNA-seq data of HCC. Its intuitive user interface allows users, particularly wet-lab researchers, to conveniently explore the data, without the need to handle any of the raw data and coding.

2. Materials and methods

2.1. Implementation

This application, LiverSCA, is a code-free, all-in-one web interface with a core written in the Python3 programming language. All the charts are generated by light-weight JavaScript library and in-house scripts. MySQL database engine stores meta information. Detailed instructions and user tutorial are available at https://patholiver.hku.hk/liverp/.

* Correspondence to: Lab 704, 7th Floor, Laboratory Block, HKU Medical Center, 21 Sassoon Road, Pokfulam, Hong Kong.
E-mail address: dwhho@hku.hk (D.W.-H. Ho).
[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2024.06.031
Received 12 March 2024; Received in revised form 21 June 2024; Accepted 23 June 2024
Available online 24 June 2024

2.2. Sample cohort and data processing

We included scRNA-seq data from tissue samples of 35 HCC patients and 5 healthy livers sourced from both public and in-house sample cohorts [18–22] (Table 1). Specifically, for the HCC samples, only tumor tissue samples were utilized, and they were categorized based on reported etiological risk factors of patients (HBV, HCV, or nonviral).

We obtained the expression matrices from the NCBI GEO database. Prior to integration, we conducted uniform quality control procedures, excluding cells with a low library size (<800), a high mitochondrial genome percentage (>10 %), and genes with low detection (UMI count < 200) (Fig.S1a).

The Harmony method [23] was utilized to eliminate undesirable technical variations among cohorts (i.e. batch-effect correction) and we combined the datasets of the same risk factor into a single Seurat object. Subsequently, the FindNeighbours, FindClusters, and RunUMAP/TSNE functions of Seurat (version 4) were employed to cluster and visualize the cells within the dataset. Cell-type markers detailed in the original papers were referenced during the process of cell type annotation.

The inferCNV R package (inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV) was used to estimate the copy number variation (CNV) status of potentially malignant cells. The CNV score for each cell was then re-standardized, and values were limited to 0 to 1. The CNV score of potential malignant cells of different risk factors was compared to normal hepatocytes (Fig.S1c).

3. Results

3.1. Overview of LiverSCA cell atlas on HCC

The LiverSCA includes 40 subjects (35 HCC patients of various etiological risk factors and 5 healthy controls) and has a total cell count of over 140,000 cells.

At the main interface, users can start by choosing a dataset (HBV, HCV, non-viral or all HCC cases). They can then opt to view the cellular landscape by cell type or cluster. Next, users can switch to different functional modules for subsequent analyses. LiverSCA's functionalities are organized into modules that provide analytical solutions for summarizing cell count/proportions, visualizing gene expression, detecting differential gene expression, conducting functional enrichment analysis, analyzing cell-cell communication, reconstructing cellular trajectories, and checking data quality (Fig. 1). In addition to LiverSCA's data visualization capabilities, all the plots in our cell atlas can be easily downloaded.

The datasets used in LiverSCA are from public and in-house sources. We annotated the major cell types in each dataset by using the markers reported in the original studies (Fig. 2a-b and supplementary table 1). There are malignant cells, normal hepatocytes, endothelial cells, fibroblasts, dendritic cells, B cells, plasma cells, macrophages, T cells, regulatory T cells, and natural killer cells. The availability of multiple etiological datasets allowed us to compare the cell compositions among them (Fig. 2c and Fig. S1 b).

3.2. Functional modules of LiverSCA

3.2.1. Dataset selection and cellular landscape

Users are required to select a dataset according to phenotype. At present, We included scRNA-seq data from tissue samples of 35 HCC patients and 5 healthy livers sourced from both public and in-house sample cohorts (Table 1). Users can examine a specific sample or a user-defined list of samples through "Sample ID" selection. After dataset selection, users can visualize the cellular landscape by tSNE or UMAP dimension reduction method and the cellular distribution can be stratified according to cell types or clusters. When users choose to display the cellular landscape by cell type, they can view the distribution of a specific cell type or various groups of cell types by selection at the "Celltype list". Similarly, LiverSCA can also display the cellular distribution according to clusters.

3.2.2. Cell proportion

In the cell proportion module, we offer two visualization methods (Fig. S2). The first is a bar plot that displays the cell population of each cell type/cluster. The second is a box plot that illustrates the distribution of gene counts per cell for each cell type/cluster. It should be noted that the results displayed by cluster or cell type are in line with the annotation methods (either cell type or cluster).

3.2.3. Marker genes

The functions of the Marker Genes module vary depending on the cellular landscape presented by different cell types or clusters. When users select "cell type" for display, this module is capable of analyzing differently expressed genes (DEGs) for different cell types (Fig. 3). On the other hand, if users choose "cluster", it can reveal the DEGs of each cluster. If users choose to display by "cell type," this module allows for conducting differential expression analysis between one etiology and another one or the healthy dataset i.e. HBV vs HC or HBV vs healthy. Apart from listing out the DEGs in table format, we also offer volcano plot visualization of DEGs. Users can perform over-representation analysis of DEGs using gene pathway definitions by GO and KEGG for the top 10 signatures or all up/down-regulated DEGs. Additionally, gene set enrichment analysis (GSEA) using different gene set definitions (GO, KEGG, Reactome and MSigDB) are also available in this module.

Table 1
Detailed information on datasets used in LiverSCA.

Tissue Info	Etiology	Accession Number	Sample Info	Number of patients	Sample ID in LiverSCA	Cell Count	Total number & count
Healthy Liver tissue	Healthy	GSE115469 [18]	Liver grafts of five healthy neurologically deceased donors	5	Healthy_N1	8438	Total number of patients: 40
HCC tumor tissue	HBV	GSE112271 [19]	Sequenced four tumor regions	1	HBV_N1	4909	
		GSE149614 [20]	Each patient collected one primary tumor sample	5	HBV_N2	16386	Total cell count: 147,142
		GSE156625# [21]	Tumor tissue from multiple sectors	9	HBV_N3	8446	
		PRJNA932937 [22]	Each patient collected one primary tumor sample	9	HBV_N4	51174	
	HCV	GSE149614	Each patient collected one primary tumor sample	2	HCV_N1	7378	
	Non-viral	GSE112271	Sequenced three tumor regions	1	non_viral_N1	12973	
		GSE149614	Each patient collected one primary tumor sample	3	non_viral_N2	7720	
		GSE156625#	Tumor tissue from multiple sectors	5	non_viral_N3	29718	

Performed CD45 + and CD45- cells sorting then 1:1 mixing them to sequence. So, this dataset was not included in the cell type composition comparison.

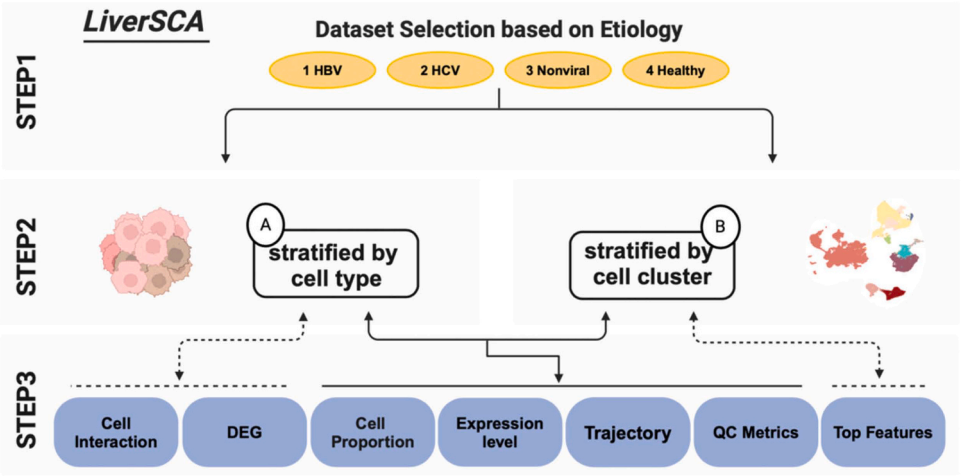

Fig. 1. Overview of LiverSCA cell atlas. Dash lines indicate the only available options for the corresponding stratification method in Step 2.

3.2.4. *FeaturePlot and VlnPlot modules*

It is important to mention that the output in these two modules will align with the choices made in the previous steps. We provide two visualization approaches (FeaturePlot and VlnPlot) to examine the expression level of one gene or a set of genes, respectively. For the choice of genes to be displayed, we offer two options. The first one is using canonical markers, which we used to annotate the cell types. Another one is a user-defined marker panel. Users can input the official gene name of any genes that are interested to view its expression level. And the main difference between these two modules is that in the latter, users can check a set of genes at once, and the results will be presented using both violin plot and dot plot.

3.2.5. *Cell-cell module*

To perform cell-cell communication analysis, we utilize the algorithm of CellphoneDB [24] and CellChat [25] at our cell atlas (Fig. 3). This module provides an overview heatmap of cellular crosstalk between different cell types. Besides, it can also display the summary of interaction pairs between two specific cell types (users need to select the cell types for partner A and partner B in the module). The option "Number" means the number of top interacting ligand-receptor pairs to be used in making the plot. For example, choosing "10″ in the "Number" option indicates the top 10 pairs with the highest interaction scores. Besides, users can also depict the dot plot using a specific ligand-receptor pairs that specified by the "Interaction Proteins(A|B)" option. For users who are interested in examining the immune checkpoint pairs, they can simply select the "Immune-Checkpoint" option. This option not only contains all the functions mentioned in the general ligand-receptor pairs but specifically focuses on the immune checkpoint pairs between antigen-presented cells (APC) and T cells. In addition, we provide a dedicated database containing some well-known immune checkpoint pairs that have been reported in literature.

3.2.6. *Trajectory module*

In this module, we provide the functionality of pseudo-time inference analysis (Fig. 3). Users have the option to choose a specific cell type and examine the reconstruction of its cell state transition process organized by clusters or subtypes. The results are presented via UMAP plots, with one focusing on stratifying cells according to clusters/subtypes whereas another one pinpointing the status of pseudo-time. Users can take advantage of the output of this module to determine the cellular progression of cells and deduce their transitional trajectory.

3.2.7. *QC Metrics module*

In this module, users can verify the quality of the scRNA-seq data. Here, we show the data information from three parameters: the number of unique genes detected in each cell (nFeature_RNA), the total number of molecules detected in a cell (nCount_RNA) and the percentage of reads that map to the mitochondrial genome (percentage_MT). The results are shown by the violin plot as well as the bar plot. Users can use "Group by" selection to decide the display of results by sample. If users prefer to view all samples in a single plot, "horizontal stack" option should be selected.

3.3. *Comparison with other web-based tools*

There are not many publicly available cell atlases regarding liver tissue and that of liver cancer tissue is particularly scanty. Therefore, we conducted a functional comparison and directly compared LiverSCA with several other liver/liver cancer-related cell atlases [26,27] (Table 2). In general, most of the other tools (LiverSCA, GepLiver, Liver Cell Atlas and Liver Single Cell Atlas) have gene expression visualization function. In fact, Liver Cell Atlas (https://www.livercellatlas.org) and Liver Single Cell Atlas (http://liveratlas-vilarinholab.med.yale.edu/) only provide this simple function. Moreover, GepLiver (http://www.gepliver.org/) has further functionalities on DEG detection and survival analysis. On the other hand, CancerLiver (https://webs.iiitd.edu.in/raghava/cancerliver/) is more focused on functional enrichment and it includes GSEA on both GO and KEGG. In summary, LiverSCA includes all the aforementioned functionalities (except survival analysis due to the current small sample size but we have plan to include more cohorts and implement this function in our future update). More importantly, LiverSCA also has other unique and important functions that distinguish itself from the other existing tools. In the HCC tumor microenvironment, there are prominent communications among different cell populations that elicit pivotal influence in supporting hepatocarcinogenesis and tumor development. Throughout the process, the cellular transition is critical to the oncogenic mechanism and they also likely contribute immunosuppressive effect in leading to immune escape that spares the HCC tumor from immune surveillance. Pinpointing the above perspectives, LiverSCA includes specific functional modules to allow exploration on cell-cell communication and cellular trajectory analyses. Apart from the functional comparison that distinguishing the specific emphasis of individual tools, we also included the computation time of making analysis plots in the comparison (Table S2). Waiting time is an important indicator of user satisfaction and it allows objective and fair interpretation of usage experience of different tools. LiverSCA were able to complete different tasks within a reasonable timeframe, which was also at least comparable to or even faster than the other tools in general (Table S2). Taken together, to our best knowledge, we believe LiverSCA is currently one of the most comprehensive liver-related cell atlases and we will continue to work

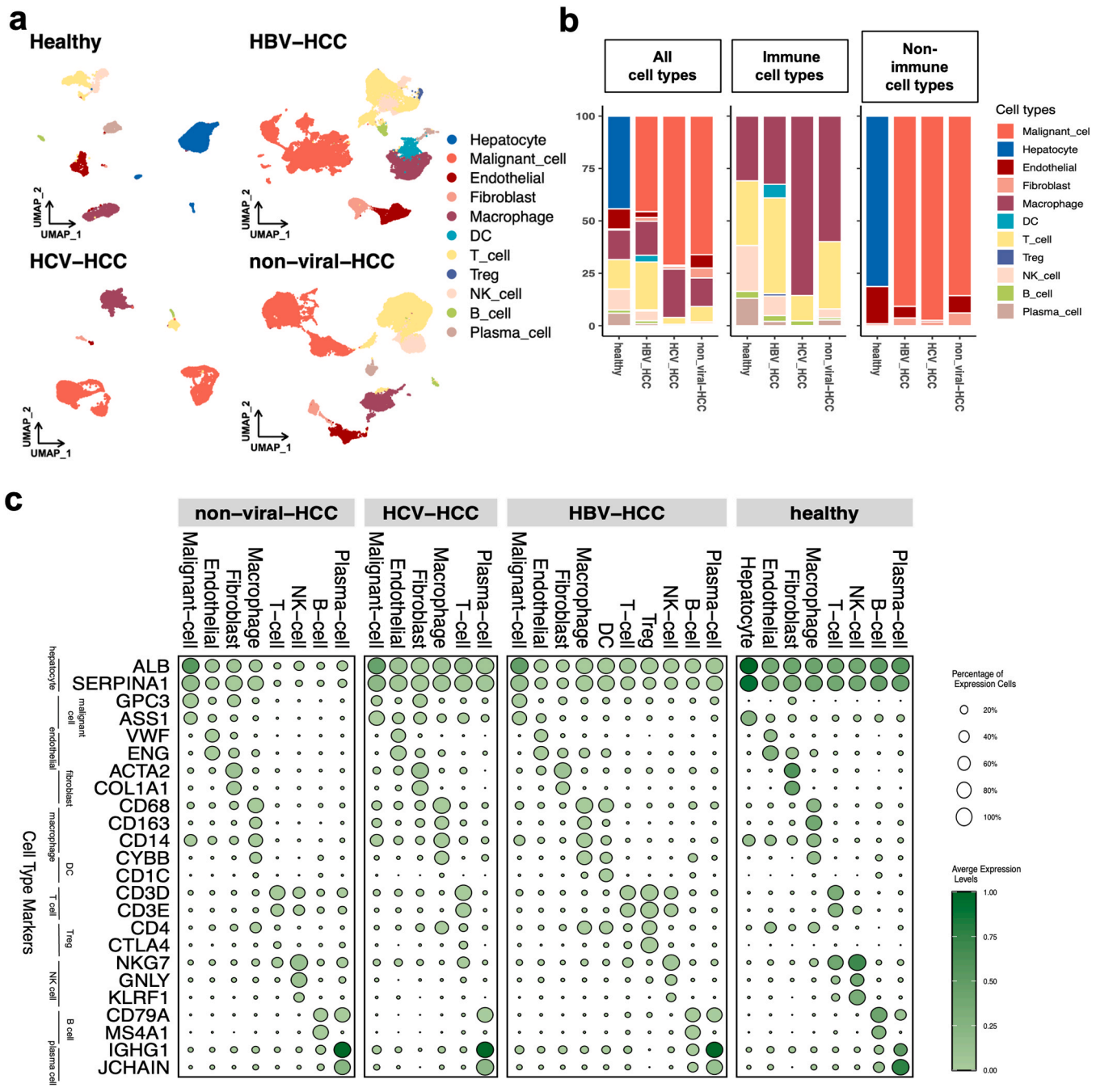

Fig. 2. Summary of datasets in LiverSCA. a. UMAP cell type distributions for four phenotypes, respectively. b. The expression levels of representative markers for each cell type among phenotypes. c. Cell type composition comparison among etiologies grouped by all cell types, immune cell types, and non-immune cell types, respectively.

towards future enhancement that pinpointing more sample cohorts, larger cell count, more comprehensive annotation of cell types and the possibility to identify rare cell subpopulations.

4. Conclusions

To sum up, our LiverSCA cell atlas on HCC website tool of scRNA-seq analysis pipeline offers a complete analysis pipeline that has comprehensive set of functions to investigate cellular and molecular landscapes in intricate biological systems of HCC. This may help identifying useful biomarkers for translational application of HCC surveillance [28] and shed light on pinpointing specific interactions or action of cells [29–32]. It is user-friendly and adaptable, enabling users to customize their analysis to fit their research needs. Our intuitive graphical interface is designed to assist researchers in discovering new mechanistic insights regarding the complex tumor microenvironment of HCC. We aim to make this process easier and more efficient for researchers with sophisticated computational/programming knowledge. We hope this can alleviate the obstacle for clinicians and cell biologists in undertaking bioinformatics discovery. Moreover, LiverSCA reduces the knowledge gap between different domains of research and accelerates knowledge transfer in the scientific community. Since scRNA-seq technology is still in rapid development, more functionalities, such as single spatial transcriptomics and multimodal omics, will also be considered for future updates. With our aims of continuous development and maintenance of LiverSCA, we believe that it can provide good and useful user

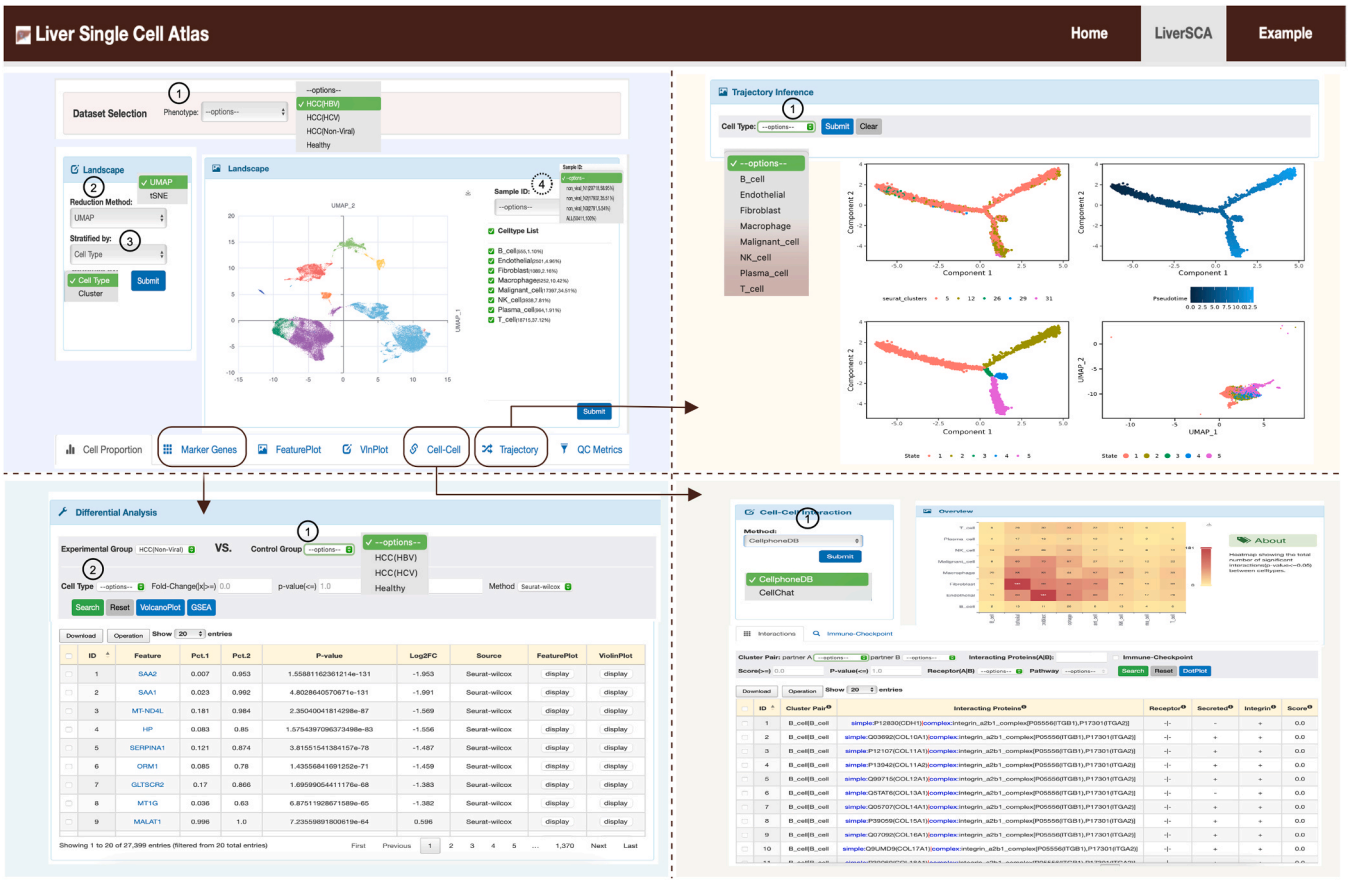

Fig. 3. The web interface of major functional modules in LiverSCA. The first step is to select the dataset and the stratification method. Next, users can switch to differently expressed gene, cell-cell communication and trajectory inference modules for the corresponding analyses.

Table 2
Comparison of functionality of cell atlases.

	Functional Modules						
	Gene Expression	DEG	ORA	GSEA	CCC	Trajectory	Survival
LiverSCA	√	√	√	√	√	√	×
GepLiver	√	√	×	×	×	×	√
CancerLiver	×	×	√	√	×	×	×
Liver Cell Atlas	√	×	×	×	×	×	×
Liver Single Cell Atlas	√	×	×	×	×	×	×

Note: DEG: differently expressed gene; ORA: over-representation analysis of DEG; GSEA: gene set enrichment analysis; CCC: cell-cell communication

experience.

Financial support

The study was supported by the Hong Kong Research Grants Council Theme-based Research Scheme (T12-704/16-R and T12-716/22-R), Innovation and Technology Commission grant to State Key Laboratory of Liver Research (ITC PD/17-9), Health and Medical Research Fund (10212956 & 07182546), RGC General Research Fund (17100021 & 17117019), National Natural Science Foundation of China (81872222), and University Development Fund of The University of Hong Kong. I.O. L. Ng is Loke Yew Professor in Pathology.

Conflict of Interest

The authors declare no conflicts of interest that pertain to this work.

Data Availability

LiverSCA is freely accessible using the link: https://patholiver.hku.hk/liverp/. The usage instruction and user tutorial are also available there.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.06.031.

References

[1] Llovet JM, et al. Immunotherapies for hepatocellular carcinoma. Nat Rev Clin Oncol 2022;19(3):151–72.
[2] Kung JWC, Ng KKC. Role of locoregional therapies in the management of patients with hepatocellular carcinoma. Hepatoma Res 2022;8:17.
[3] Fonseca LG, Carrilho F, J. Current landscape and future directions for systemic treatments of hepatocellular carcinoma. Hepatoma Res 2023;9:27.

[4] Wu Y-C, et al. Issue on combined locoregional and systemic treatment for hepatocellular carcinoma. Hepatoma Res 2023;9:6.
[5] Ho DW, et al. Molecular pathogenesis of hepatocellular carcinoma. Liver Cancer 2016;5(4):290–302.
[6] Llovet JM, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2021;7(1):6.
[7] Piciotti R, et al. Old-fashioned and newly discovered biomarkers: the future of NAFLD-related HCC screening and monitoring. Hepatoma Res 2022;8:37.
[8] Desjonqueres E, Gigante E. Hepatocellular carcinoma on the background of nonalcoholic fatty liver disease: epidemiological update. Hepatoma Res 2022;8:16.
[9] Chan LK, et al. Cellular heterogeneity and plasticity in liver cancer. Semin Cancer Biol 2022;82:134–49.
[10] Xiong X, et al. Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. Mol Cell 2019;75(3):644–60. e5.
[11] Pfister D, et al. NASH limits anti-tumour surveillance in immunotherapy-treated HCC. Nature 2021;592(7854):450–6.
[12] Yuen VW-H, et al. Using mouse liver cancer models based on somatic genome editing to predict immune checkpoint inhibitor responses. J Hepatol 2023;78(2):376–89.
[13] Zheng, G., et al., *Massively parallel digital transcriptional profiling of single cells. Nat Commun 8: 14049.* Data Set5. Putative transcription factors binding motifs identified for genes in trans-eQTL (expression quantitative trait loci) hotspots Data Set6. Putative master regulators in the trans-eQTL (expression quantitative trait loci) hotspots Figure S, 2017. 1.
[14] Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19(1):5.
[15] Stuart T, et al. Comprehensive integration of single-cell data. Cell 2019;177(7):1888–902. e21.
[16] Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6(269). p. pl1-pl1.
[17] Dennis G, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4(9):1–11.
[18] MacParland SA, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun 2018;9(1):4383.
[19] Losic B, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. Nat Commun 2020;11:*291*.
[20] Lu Y, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nat Commun 2022;13(1):4594.
[21] Sharma A, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell 2020;183(2):377–94. e21.
[22] Zhang Q, et al. Reciprocal interactions between malignant cells and macrophages enhance cancer stemness and M2 polarization in HBV-associated hepatocellular carcinoma. Theranostics 2024;14(2):892.
[23] Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16(12):1289–96.
[24] Efremova M, et al. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc 2020;15(4):1484–506.
[25] Jin S, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021;12(1):1088.
[26] Li Z, et al. GepLiver: an integrative liver expression atlas spanning developmental stages and liver disease phases. Sci Data 2023;10(1):376.
[27] Kaur H, et al. CancerLivER: a database of liver cancer gene expression resources and biomarkers. Database 2020;2020:baaa012.
[28] Ahn JC, et al. Hepatocellular carcinoma surveillance: current practice and future directions. Hepatoma Res 2022;8:10.
[29] Chen W-M, et al. Targeting the interactions between lymphocytes and liver cancer stem cells in combination with immunotherapy is a promising therapeutic strategy. Hepatoma Res 2023;9:2.
[30] Cadamuro M, et al. The tumor microenvironment in hepatocarcinoma: dissecting the functions of cancer-associated fibroblasts. Hepatoma Res 2023;9:47.
[31] Ho DW, et al. Single-cell RNA sequencing shows the immunosuppressive landscape and tumor heterogeneity of HBV-associated hepatocellular carcinoma. Nat Commun 2021;12(1):3684.
[32] Ho DW, et al. Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. Cancer Lett 2019;459:176–85.