Human-Spa: An Online Platform Based on Spatial Transcriptome Data for Diseases of Human Systems

Yunyun Su
School of Computer Science and Technology
Hainan University
Haikou, China
yunyunsu@hainanu.edu.cn

Shiyu Yan
School of Life Sciences
Hainan University
Haikou, China
20223001644@hainanu.edu.cn

Chen Cao
School of Biomedical Engineering and Informatics
Nanjing Medical University
Nanjing,China
chen.cao@ucalgary.ca

Feifei Cui
School of Computer Science and Technology
Hainan University
Haikou, China
feifeicui@hainanu.edu.cn

Quan Zou*
Institute of Fundamental and Frontier Sciences
University of Electronic Science and Technology of China
Chengdu, China
zouquan@nclab.net

Zilong Zhang*
School of Computer Science and Technology
Hainan University
Haikou, China
zhangzilong@hainanu.edu.cn

2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385302

***Abstract*—Spatial transcriptomics has become a major method for high-throughput analysis of gene expression at the current level of cells, which can directly study gene expression changes in disease cells, reveal the occurrence and development mechanism of diseases, identify potential therapeutic targets related to diseases, and provide new clues for disease diagnosis and treatment. Although there have been major breakthroughs in the analysis and acquisition of transcriptome data, there are still many challenges in how to effectively manage, share and utilize these valuable data resources in the study of human systemic diseases. To solve this problem, we are committed to building a spatial transcriptome database website related to human systemic diseases, aiming to provide reliable datasets related to multiple diseases, and provide certain data information and analysis results to accelerate the progress of human systemic diseases research. This paper will introduce the construction process of the database website in detail, and discuss its application prospects in disease research, with the aim of promoting further development and innovation in the field of human health. Here, Human-spa mainly includes 12 Human systems, 38 disease types, 55 datasets, and Human-Spa provides a very friendly web interface for visualization and dataset parsing. In conclusion, the construction of Human-Spa will provide a powerful tool and resource platform for Human disease research. Human-Spa is available for free at http://www.human-spa.cn/**

***Keywords* — spatial transcriptome, diseases of human system, database website, data analysis**

I. Introduction

Human health and disease are determined by the synergies of different cell types and tissues. Based on the 12 major systems, the classification of human diseases can better understand and study human diseases. For example, we are familiar with Respiratory System, such as Lung Cancer [1], Pulmonary Fibrosis [2], Lung Adenocarcinoma cells [3] and other diseases; Urinary System, such as Adrenal Cancer [4], Adrenal Hyperplasia [5], Renal Clear Cell Carcinoma [6] and other diseases, each system is closely linked to maintain the normal function of the body. These classifications contribute to a better understanding of the characteristics of individual systems and the manifestations of diseases, and provide guidance for the study of human diseases. However, it should be noted that each disease may involve multiple systems, so it is necessary to consider the interaction of multiple systems comprehensively when studying and diagnosing diseases.

Spatial transcriptomics [7] is a technique and methodology used to investigate the spatial distribution and dynamics of gene expression with in cells. The rapid development of spatial transcriptomics technology allows us to better understand the type, function and spatial distribution of cells. Spatial transcriptome sequencing technology has a wide range of applications in biomedical research. In the beginning, Linnarsson et al. [8] the first proposed and developed single-cell RNA sequencing technology [8, 9], which laid the foundation for single-cell analysis. Subsequently, Ke et al. proposed Slide-seq technology [10] to realize spatial analysis of gene expression. So far, common sequencing technologies include: spatially-located RNA sequencing [11], Slide-seq [12], SeqFISH [13] and MERFISH [14], etc. It can help to study organ development, tumor microenvironment, immune cell distribution, and gene expression in the nervous system, it can also be used in disease diagnosis and drug development. Therefore, it is very convenient to build a related database website for the collation and application of transcriptome cell datasets.

Spatial transcriptome data has also exerted great influence in the information age. scRNA-seq [15] (single-cell RNA sequencing), as a high-throughput molecular biology technique utilized for sequencing and analyzing the transcriptome of individual cells [16, 17]. It has a wide range of applications in life science research, including precise classification and definition of cell subsets, identification of cell state transitions, analysis of cellular development and differentiation pathways, as well as investigation into disease pathogenesis [18, 19]. At present, there are two methods [20, 21] for integrating scRNA-seq and spatial transcriptome data, mainly are Deconvolution [22] and Mapping [23]. Deconvolution aims to isolate discrete cell subpopulations from a mixture of mRNA transcripts at each capture point based on single-cell data; Mapping has two aspects: locating a specified scRNA-seq to each cell on the HPRI [24] map profile and locating each scRNA-seq cell to a specific niche or region of the tissue. Under the leadership of deep learning, machine learning, and artificial intelligence, very effective

spatial transcriptome data can be obtained, making significant contributions to scientific research.

With the continuous development of transcriptomic and spatial transcriptomic imaging technologies, an increasing number of spatial transcriptomic database websites are beginning to involve the integration and analysis of spatial transcriptomic data. For example, in 2018, Stuart et al. developed a website called "Tabula Muris Spatial" [25] that integrates spatial transcriptomic data from 18 different tissues in mice and provides online querying and analysis tools. In 2020, Z. Fan et al. released the first single-cell spatial transcriptomic database and an online visualization platform called "SpatialDB" [26]. They classified the dataset using 8 different sequencing technologies, providing a resource database for studying the spatial cellular structure of tissues and potentially offering new insights into the cellular microenvironment in diseases. More recently, Z. Fan et al. developed a new spatial transcriptomic database website called "SPASCER" [27], aimed at aiding the understanding of tissue heterogeneity, microenvironments in specific regions, and intercellular interactions across tissue structures at multiple levels.

Combining the existing spatial transcriptome data sites with the lack of idling databases for specific human system diseases, we developed Human-Spa, a manually collated resource of spatial transcriptome datasets for researchers to efficiently investigate and reuse these published data. In the current version, Human-Spa is divided into 12 categories based on Human systems, for the vacancy of idling database sites for diseases of human systems. Including Cardiovascular System (4), Endocrine System (3), Digestive System (2), Hematopoietic System (2), Surgical System (1), Epidermal System (5), Immune System (4), Urinary System (4), Respiratory System (8), Reproductive System (16), Musculoskeletal System (3), Nervous System (3), including a total of 38 disease types, an online site that provides partial spatial visualization of 55 datasets and multiple common spatial transcriptome techniques. Human-Spa is available for free at ***http://www.human-spa.cn/*** and no login required.

II. WEBSITE OVERVIEW

The Human Systems Disease Spatial Transcriptome Database (Human-Spa) is an online platform for storing and providing spatial transcriptome data related to various systemic diseases of the human body. Human-Spa is committed to collecting, collating, and sharing transcriptome data from multiple tissues and disease states so that scientists, researchers, and physicians can conduct research in understanding disease mechanisms and developing new treatments (Fig.1).

In order to enhance the navigability of our database website, we delve into the specific gene expression patterns and cellular subpopulation differentiations across various cell types in different tissues and organs. We need to carefully consider several factors, including data sources, data processing and storage, database construction, and data visualization analysis. Let's take a closer look at each of them.

A. Data collection

These datasets involved in this website are mainly obtained from GEO database [28], PubMed literature database [29] and other platforms, The above are all open databases, which contain the literature of published research data. According to the different data sources, the paper information and datasets involved in the specific data are collected, including keywords and systematic review. After the collected datasets are classified according to disease type, the dataset containing expression matrix and normalized processing are preferentially retained, so as to facilitate the subsequent data integration and improve the efficiency of data visualization analysis. We obtained 38 disease types and 55 datasets from the search results (Fig.1 and TABLE I). For each datasets, we read the original paper, extract the original data, and provide uniform details of the data publication information, experimental design, and data description on the website.

B. Data processing

In the process of matrix expression and normalization of the collected spatial transcription datasets, data preprocessing is firstly carried out. In the process of data preprocessing, considering the influence of different sequencing technologies on the results of data analysis, the quality control of the collected data is taken into account, the data with poor quality is cleaned and deleted, and the data dimension is reduced by gene screening to eliminate the data batch effect. To eliminate differences that may occur in horizontal comparisons. The matrix representation and data normalization are further carried out. When the expression matrix was generated, gene readings and FPKM/RPKM [30] values were taken, and gene expression levels were arranged by gene row and sample column. A series of standardized gene samples were performed to eliminate differences in expression between them. The main purpose of the above processing is to obtain a dataset that meets the requirements of the analysis.

To facilitate data visualization, we store the processed datasets in a MySQL database and create a data table for it, connect via Navicat, and communicate with the front page via Express.

C. Database construction

Human-Spa is built on a CentOS Linux server. The Web server is built using Apache, the website is developed using PHP, the front-end web page is developed using Layui framework, and the Echarts framework is used to realize visualization. The visual data analysis presented in Human-Spa uses common data analysis packages such as Seurat, Scanpy, etc.

D. Data analysis and visualization

A graph of cluster analysis results for relevant disease datasets is available in Human-Spa. About UMAP [31] and t-SNE [32] clustering are popular nonlinear reduction and visualization techniques. It is mainly used for dimensionality reduction, but it can also be used for clustering tasks. UMAP cluster graph is a dataset generated by UMAP algorithm, and t-SNE cluster graph is a visual representation of the dataset generated by t-SNE algorithm to help us observe and understand the cluster structure of the dataset. A cluster map can identify cluster patterns, cluster distribution, and similarity relationships among data points in a dataset.

We will cluster and visualize datasets, use dimensionality reduction algorithms such as UMAP and t-SNE to reduce high-dimensional data to two or many dimensions., data dimensionality reduction, and map the data to a low-dimensional manifold to preserve the local structure of the

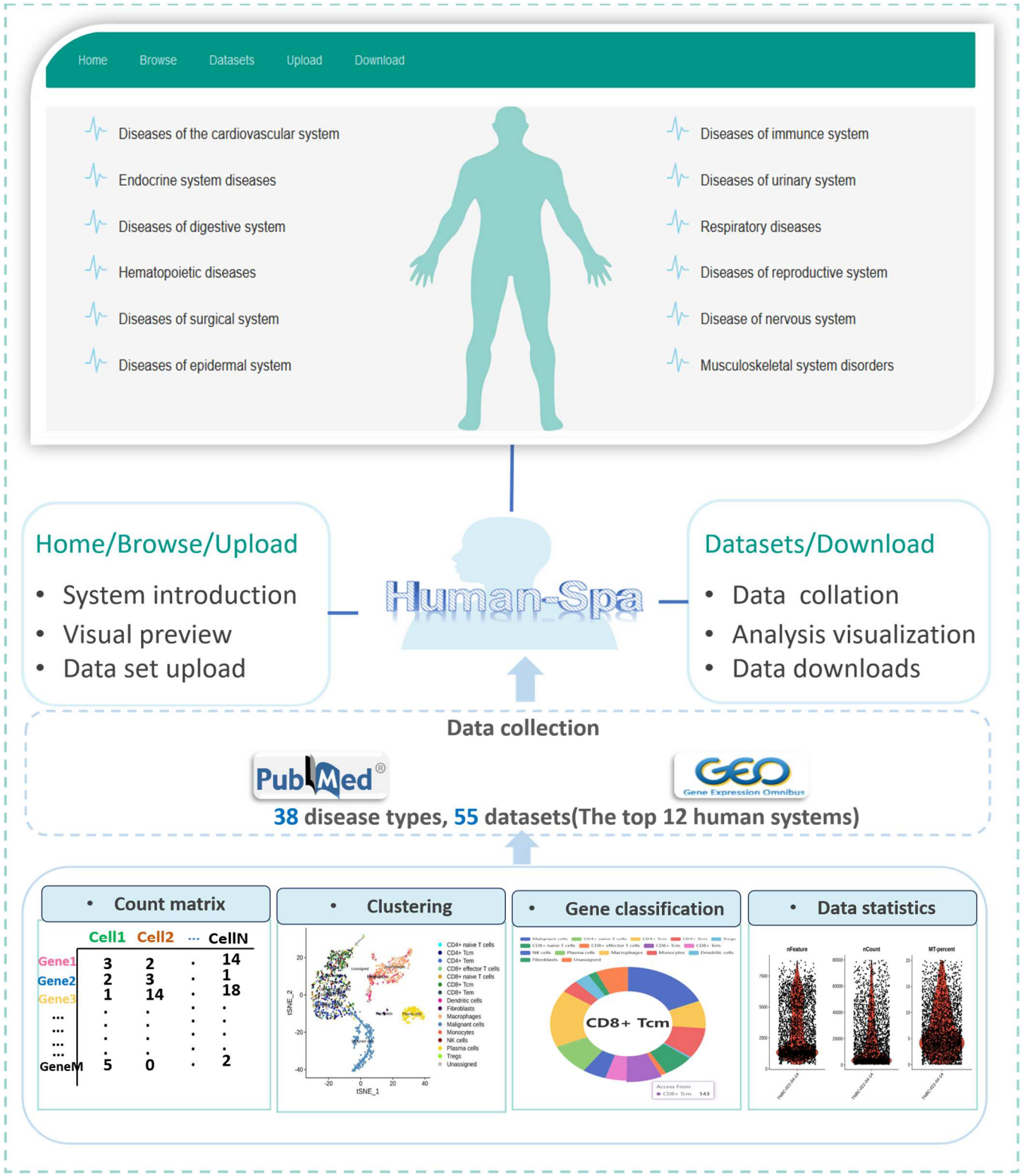


Fig.1. Overview of the Human-Spa database, the spatial transcriptome data of diseases corresponding to 12 major human systems were collected and organized. Human-Spa provides a web interface for the integration of such datasets, where users can preview, upload, download datasets, and view visual analysis annotations of datasets.

data. Clustering algorithms, such as K-means [33] and hierarchical clustering [34], are used to perform cluster analysis on the data after dimensionality reduction, and the clustering results are visualized using scatter plots or other graphical methods. In a cluster graph, the data points of different clusters often have different colors or markers in order to distinguish them.

E. Data preview, download and upload

In the process of building the website, the Echarts framework was used to carry out statistics on the corresponding disease types of each human system, and detailed explanations were made in the form of charts. In the process of building the website, the Echarts framework can be used to carry out statistics on the disease types corresponding to each human system and explain them in

TABLE I PROVIDES DETAILED STATISTICS ON THE COLLECTED DATASETS IN HUMAN-SPA

(*)OUTSIDE IS THE NUMBER OF DATASETS, INSIDE IS THE NUMBER OF DISEASE TYPES

Human System	Datasets No.*	Disease Type
Cardiovascular system	4（4）	Pulmonary hypertension, Venous hypertension, Dilated myocarditis, Carotid atherosclerosis
Endocrine system	3（3）	Type 1 diabetes, Type 2 diabetes, Adrenal hyperplasia
Digestive system	2（2）	Gastric cancer, Liver cancer
Hematopoietic system	3（3）	Chronic lymphocytic leukemia, Acute Promyelocytic Leukemia, Acute Lymphoblastic Leukemia
Surgical system	1（1）	Dermatitis
Epidermal system	5（3）	Melanoma, Squamous-cell Carcinoma, Herpes virus
Immune system	4（2）	Rheumatoid arthritis, Systemic lupus erythematosus
Urinary system	4（3）	Adrenal carcinoma, Clear cell carcinoma of kidney, Wilms tumor
Respiratory system	8（4）	Lung cancer, Pulmonary fibrosis，Alveolar adenocarcinoma, COVID-19
Reproductive system	15（6）	Triple negative breast cancer, Breast cancers，Cervical cancer, Ovarian cancer，Prostate cancer, HPV
Musculoskeletal system	3（3）	Human osteosarcoma, Lumbar disc herniation, Osteoarthritis
Nervous system	3（3）	Glioma, Intracranial Aneurysms, Alzheimer's disease

detail in the form of charts. First, a dataset of human systems and disease types is organized, and the data is stored in a JSON file, each system corresponds to an array, and the array contains different disease types under that system. Import the Echarts library, import the online CDN version through the <script> tag or download the library file to the local import, create the chart container, specify the container element as the parent element of the chart, configure the chart parameters.

In addition, users can also download the required data through the “Download” page. If users wish to share their data, they can send the necessary information to us via the “Upload” page. We will process this data and add it to the Human-Spa database.

F. Gene proportion

According to the expression pattern classification of genes in different tissues or physiological conditions in the data set, the number of species was integrated, including gene name, proportion and classification. Meanwhile, the total proportion of classification was calculated, pie chart example was created and parameters were configured (Fig. 2)

III. DISCUSSION

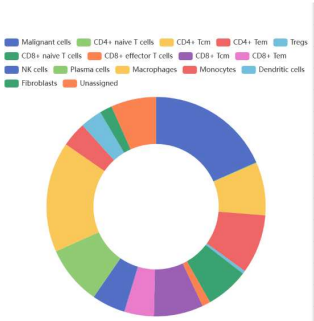
At present, We have completed the research on the website construction of the spatial transcription database of human system diseases, mainly using the front-end framework and database to achieve the construction of the website, and also made a lot of information collection and sorting in the website content, including some data analysis content. The main purpose of constructing this website is to realize the integration of human spatial transcription dataset. This paper mainly describes some important work made. At present, spatial transcriptome sequencing technology is one of the most popular analysis methods, which can provide spatial information of gene expression at the tissue and organ level.

At the same time, the scarcity of websites targeting spatial transcriptome databases provides a more favorable approach to data analysis in the medical field. Human-Spa integrates existing spatial transcription data and provides detailed descriptions of data sources, data processing, data analysis and visualization on the website. By storing and correlating large amounts of transcriptome data in a database, Human-Spa provides users with an important knowledge discovery platform. The implementation process of the existing dataset is described in detail, and users can conduct more detailed research according to the actual situation and the papers in which the data appear, and find the common characteristics, gene expression patterns and possible biological mechanisms between different diseases. This is very helpful in deepening the

understanding of disease occurrence and development, providing clues for new treatments and drug development.

IV. SHORTAGE AND PROSPECT

Despite the above advantages, Human-Spa still has some limitations. First of all, Human-Spa does not provide an online analysis method for visualization, but only provides the user with the visualized results of our implementation. There are a number of relatively mature online tools and platforms (SPACECAT, Seurat Spatial, etc.) available for online analysis of spatial transcriptome data, which provide various analytical methods and visualization capabilities to help researchers resolve spatial features of gene expression. At the same time, Human-Spa as a whole is a static interactive web interface, which means that it provides the ability to display and interact with existing data, but the data analysis visualization of Human-Spa is not yet mature. To address these deficiencies, we can make our website richer and more powerful with the help of online analysis tools and cutting-edge frameworks to quickly build a responsive and dynamic interface.



Subtype	Cell Count	Percentage
Malignant cells	361	18.380%
CD4+ naive T cells	2	0.100%
CD4+ Tcm	153	7.790%
CD4+ Tem	171	8.710%
Tregs	9	0.460%
CD8+ naive T cells	125	6.360%
CD8+ effector T cells	24	1.220%
CD8+ Tcm	143	7.280%
CD8+ Tem	86	4.380%
NK cells	97	4.940%
Plasma cells	171	8.710%
Macrophages	320	16.290%
Monocytes	72	3.670%
Dendritic cells	63	3.210%
Fibroblasts	36	1.830%
Unassigned	131	6.670%

Fig.2. This is a picture of the GEO database, dataset GSE148673. That is, 46,501 cells in 21 tumors of triple-negative breast cancer were data processed to retain 1966 cells with obvious characteristics corresponding to the classification and proportion of different gene types.

The advancement of the Human Cell Atlas project provides a new perspective for life science research. It is expected that more spatial transcriptomic techniques will be developed and spatial gene expression data will be accumulated rapidly. We will continue to collect new technologies and datasets to update Human-Spa. In addition, we will integrate more tools and data sources to analyze data. Collecting new technologies and datasets and integrating them into Human-Spa will enable the platform to offer more comprehensive and diverse data analysis and visualization capabilities. This is extremely valuable for researchers, allowing them to delve deeper into cell function and interactions. In addition, the integration of more tools and data sources is also a good direction. By combining different tools and data sources, it can provide more diversified analysis methods and richer data resources, further expanding the scope of application of Human-Spa. Continuously updating and integrating new technologies, datasets, tools, and data sources, and improving on the shortcomings of Human-Spa, Human-Spa will become a more powerful and comprehensive platform that offers more possibilities for researchers and pushes life science research to the next level.

V. WEBSITE INSTRUCTIONS

Human-Spa is a website for researchers and practitioners in the fields of spatial transcriptome, data mining and bioinformatics. It mainly includes the following five functional modules: home page, preview, dataset, upload and download. The home page provides a brief overview of the site's introduction, purpose, and dataset information for the 12 major systems of the human body. The preview page displays the distribution of the dataset in the form of bar charts and pie charts, allowing users to intuitively understand the composition and distribution of the data. The dataset page classifies by body system, showing details of each dataset and some of the data analysis results. Users can view data information about a particular dataset and obtain an analysis sample graph of the dataset in the Submit function on the dataset page. The upload page provides the details of the dataset submitted by the user, where the user can provide the dataset information, which we then summarize and add to our database based on this information. The download page provides the download function of all datasets, and users can get the datasets they need from here. Through the above functional modules, Human-Spa provides a convenient platform for researchers and practitioners to browse, upload and download datasets related to spatial transcriptome, data mining and biological information, which will help them better conduct research and promote the development of related fields.

FUNDING

The work was supported by the National Natural Science Foundation of China (No. 62102064, No. 62261018).

REFERENCES

[1] B. C. Bade and C. S. Dela Cruz, "Lung Cancer 2020: Epidemiology, Etiology, and Prevention," (in eng), Clin Chest Med, vol. 41, no. 1, pp. 1-24, Mar 2020, doi: 10.1016/j.ccm.2019.10.001.

[2] P. W. Noble, C. E. Barkauskas, and D. Jiang, "Pulmonary fibrosis: patterns and perpetrators," (in eng), J Clin Invest, vol. 122, no. 8, pp. 2756-62, Aug 2012, doi: 10.1172/JCI60323.

[3] N. Kim et al., "Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma," (in eng), Nat Commun, vol. 11, no. 1, p. 2285, May 08 2020, doi: 10.1038/s41467-020-16164-1.

[4] G. D. Hammer, M. Fassnacht, and E. Lalli, "Adrenal cancer: scientific advances," (in eng), Mol Cell Endocrinol, vol. 351, no. 1, p. 1, Mar 31 2012, doi: 10.1016/j.mce.2011.11.010.

[5] S. F. Witchel, "Congenital Adrenal Hyperplasia," (in eng), J Pediatr Adolesc Gynecol, vol. 30, no. 5, pp. 520-534, Oct 2017, doi: 10.1016/j.jpag.2017.04.001.

[6] H. I. Wettersten, O. A. Aboud, P. N. Lara, and R. H. Weiss, "Metabolic reprogramming in clear cell renal cell carcinoma," (in eng), Nat Rev Nephrol, vol. 13, no. 7, pp. 410-419, Jul 2017, doi: 10.1038/nrneph.2017.59.

[7] X. Shen, Y. Zhao, Z. Wang, and Q. Shi, "Recent advances in high-throughput single-cell transcriptomics and spatial transcriptomics," (in eng), Lab Chip, vol. 22, no. 24, pp. 4774-4791, Dec 06 2022, doi: 10.1039/d2lc00633b.

[8] S. Linnarsson, "Sequencing Single Cells Reveals Sequential Stem Cell States," (in eng), Cell Stem Cell, vol. 17, no. 3, pp. 251-2, Sep 03 2015, doi: 10.1016/j.stem.2015.08.016.

[9] S. Slovin et al., "Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview," (in eng), Methods Mol Biol, vol. 2284, pp. 343-365, 2021, doi: 10.1007/978-1-0716-1307-8_19.

[10] S. Vickovic et al., "High-definition spatial transcriptomics for in situ tissue profiling," (in eng), Nat Methods, vol. 16, no. 10, pp. 987-990, Oct 2019, doi: 10.1038/s41592-019-0548-y.

[11] E. Jia et al., "Spatial Transcriptome Profiling of Mouse Hippocampal Single Cell Microzone in Parkinson's Disease," (in eng), Int J Mol Sci, vol. 24, no. 3, Jan 17 2023, doi: 10.3390/ijms24031810.

[12] S. G. Rodriques et al., "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution," (in eng), Science, vol. 363, no. 6434, pp. 1463-1467, Mar 29 2019, doi: 10.1126/science.aaw1219.

[13] C. L. Eng et al., "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH," (in eng), Nature, vol. 568, no. 7751, pp. 235-239, Apr 2019, doi: 10.1038/s41586-019-1049-y.

[14] J. R. Moffitt and X. Zhuang, "RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH)," (in eng), Methods Enzymol, vol. 572, pp. 1-49, 2016, doi: 10.1016/bs.mie.2016.03.020.

[15] G. Chen, B. Ning, and T. Shi, "Single-Cell RNA-Seq Technologies and Related Computational Data Analysis," (in eng), Front Genet, vol. 10, p. 317, 2019, doi: 10.3389/fgene.2019.00317.

[16] Z. Zhang, F. Cui, C. Wang, L. Zhao, and Q. Zou, "Goals and approaches for each processing step for single-cell RNA sequencing data," (in eng), Brief Bioinform, vol. 22, no. 4, Jul 20 2021, doi: 10.1093/bib/bbaa314.

[17] Z. Zhang, F. Cui, C. Lin, L. Zhao, C. Wang, and Q. Zou, "Critical downstream analysis steps for single-cell RNA sequencing data," (in eng), Brief Bioinform, vol. 22, no. 5, Sep 2 2021, doi: 10.1093/bib/bbab105.

[18] Z. Zhang, F. Cui, M. Zhou, S. Wu, Q. Zou, and B. Gao, "Single-cell RNA Sequencing Analysis Identifies Key Genes in Brain Metastasis from Lung Adenocarcinoma," (in eng), Curr Gene Ther, vol. 21, no. 4, pp. 338-348, 2021, doi: 10.2174/1566523221666210319104752.

[19] Z. Zhang, F. Cui, C. Cao, Q. Wang, and Q. Zou, "Single-cell RNA analysis reveals the potential risk of organ-specific cell types vulnerable to SARS-CoV-2 infections," (in eng), Comput Biol Med, vol. 140, p. 105092, Nov 29 2021, doi: 10.1016/j.compbiomed.2021.105092.

[20] J. Chen et al., "Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data," (in eng), Nat Commun, vol. 13, no. 1, p. 6494, Oct 30 2022, doi: 10.1038/s41467-022-34277-7.

[21] Z. Zhang et al., "webSCST: an interactive web application for single-cell RNA-sequencing data and spatial transcriptomic data integration," (in eng), Bioinformatics, vol. 38, no. 13, pp. 3488-3489, Jun 27 2022, doi: 10.1093/bioinformatics/btac350.

[22] N. Wang, X. Li, R. Wang, and Z. Ding, "Spatial transcriptomics and proteomics technologies for deconvoluting the tumor microenvironment," (in eng), Biotechnol J, vol. 16, no. 9, p. e2100041, Sep 2021, doi: 10.1002/biot.202100041.

[23] I. Harrow et al., "Ontology mapping for semantically enabled applications," (in eng), Drug Discov Today, vol. 24, no. 10, pp. 2068-2075, Oct 2019, doi: 10.1016/j.drudis.2019.05.020.

[24] T. C. Ricketts, K. Thompson, S. Neuwahl, and V. McGee, "HPRI data tracks. Developing an index of surgical underservice," (in eng), Bull Am Coll Surg, vol. 96, no. 7, pp. 45-7, 57, Jul 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22315900.

[25] T. M. Consortium, O. c. S. N. K. J. N. N. F. M. A. P. Q. S. R. q. s. e. f. W.-C. T. twc@ stanford. edu 4 5 6 g Darmanis Spyros spyros. darmanis@ czbiohub. org 2 h, L. c. B. J. B. O. C. M. B. C. S. G. F. J. R. C. M. A. P. L. P. A. Oliveira 2 Sit Rene V. 2 Stanley Geoffrey M. 3 Webber James T. 2 Zanini Fabio 3, and C. d. a. B. J. B. O. C. P. C. D. D. S. D. J. L. K. J. P. A. O. Stanley, "Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris," Nature, vol. 562, no. 7727, pp. 367-372, 2018.

[26] Z. Fan, R. Chen, and X. Chen, "SpatialDB: a database for spatially resolved transcriptomes," Nucleic acids research, vol. 48, no. D1, pp. D233-D237, 2020.

[27] Z. Fan et al., "SPASCER: spatial transcriptomics annotation at single-cell resolution," Nucleic Acids Research, vol. 51, no. D1, pp. D1138-D1149, 2023.

[28] T. Barrett et al., "NCBI GEO: archive for functional genomics data sets--update," (in eng), Nucleic Acids Res, vol. 41, no. Database issue, pp. D991-5, Jan 2013, doi: 10.1093/nar/gks1193.

[29] N. Fiorini, D. J. Lipman, and Z. Lu, "Towards PubMed 2.0," (in eng), Elife, vol. 6, Oct 30 2017, doi: 10.7554/eLife.28801.

[30] M. S. I. Dona, L. A. Prendergast, S. Mathivanan, S. Keerthikumar, and A. Salim, "Powerful differential expression analysis incorporating network topology for next-generation sequencing data," (in eng), Bioinformatics, vol. 33, no. 10, pp. 1505-1513, May 15 2017, doi: 10.1093/bioinformatics/btw833.

[31] A. Diaz-Papkovich, L. Anderson-Trocmé, and S. Gravel, "A review of UMAP in population genetics," (in eng), J Hum Genet, vol. 66, no. 1, pp. 85-91, Jan 2021, doi: 10.1038/s10038-020-00851-4.

[32] N. Pezzotti, B. P. F. Lelieveldt, L. Van Der Maaten, T. Hollt, E. Eisemann, and A. Vilanova, "Approximated and User Steerable tSNE for Progressive Visual Analytics," (in eng), IEEE Trans Vis Comput Graph, vol. 23, no. 7, pp. 1739-1752, Jul 2017, doi: 10.1109/TVCG.2016.2570755.

[33] B. Liu, T. Zhang, Y. Li, Z. Liu, and Z. Zhang, "Kernel Probabilistic K-Means Clustering," (in eng), Sensors (Basel), vol. 21, no. 5, Mar 08 2021, doi: 10.3390/s21051892.

[34] Z. Zou, K. Hua, and X. Zhang, "HGC: fast hierarchical clustering f or large-scale single-cell data," (in eng), Bioinformatics, vol. 37, no . 21, pp. 3964-3965, Nov 05 2021, doi: 10.1093/bioinformatics/btab420.