



OPEN ACCESS

EDITED BY

Tallulah Andrews,
Western University, Canada

REVIEWED BY

Nikolaus Fortelny,
University of Salzburg, Austria
Mai Chan Lau,
Bioinformatics Institute (A*STAR), Singapore
Shengquan Chen,
Nankai University, China

*CORRESPONDENCE

Marcin Cieslik,
✉ mcieslik@med.umich.edu
Arul M. Chinnaiyan,
✉ arul@med.umich.edu

RECEIVED 14 April 2024

ACCEPTED 11 June 2024

PUBLISHED 08 July 2024

CITATION

Gondal MN, Shah SUR, Chinnaiyan AM and
Cieslik M (2024), A systematic overview of
single-cell transcriptomics databases, their use
cases, and limitations.

Front. Bioinform. 4:1417428.

doi: 10.3389/fbinf.2024.1417428

COPYRIGHT

© 2024 Gondal, Shah, Chinnaiyan and Cieslik.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A systematic overview of single-cell transcriptomics databases, their use cases, and limitations

Mahnoor N. Gondal^{1,2}, Saad Ur Rehman Shah³,
Arul M. Chinnaiyan^{1,2,4,5,6,7*} and Marcin Cieslik^{1,2,4,7*}

¹Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States, ²Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, United States, ³Gies College of Business, University of Illinois Business College, Champaign, MI, United States, ⁴Department of Pathology, University of Michigan, Ann Arbor, MI, United States, ⁵Department of Urology, University of Michigan, Ann Arbor, MI, United States, ⁶Howard Hughes Medical Institute, Ann Arbor, MI, United States, ⁷University of Michigan Rogel Cancer Center, Ann Arbor, MI, United States

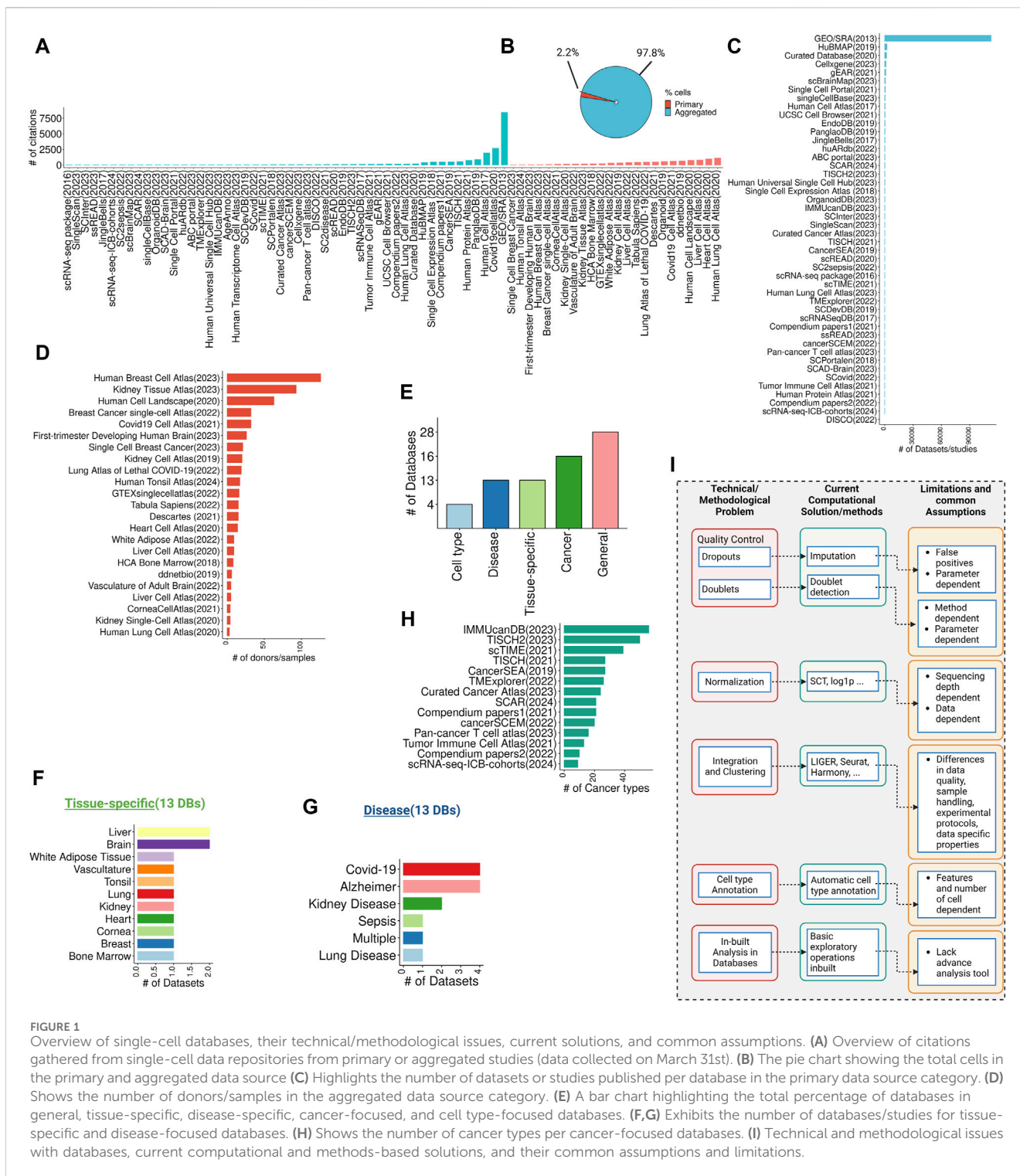
Rapid advancements in high-throughput single-cell RNA-seq (scRNA-seq) technologies and experimental protocols have led to the generation of vast amounts of transcriptomic data that populates several online databases and repositories. Here, we systematically examined large-scale scRNA-seq databases, categorizing them based on their scope and purpose such as general, tissue-specific databases, disease-specific databases, cancer-focused databases, and cell type-focused databases. Next, we discuss the technical and methodological challenges associated with curating large-scale scRNA-seq databases, along with current computational solutions. We argue that understanding scRNA-seq databases, including their limitations and assumptions, is crucial for effectively utilizing this data to make robust discoveries and identify novel biological insights. Such platforms can help bridge the gap between computational and wet lab scientists through user-friendly web-based interfaces needed for democratizing access to single-cell data. These platforms would facilitate interdisciplinary research, enabling researchers from various disciplines to collaborate effectively. This review underscores the importance of leveraging computational approaches to unravel the complexities of single-cell data and offers a promising direction for future research in the field.

KEYWORDS

single-cell RNA-seq, single-cell databases, single-cell atlases, single-cell data analysis, web-based platforms, cell heterogeneity, single-cell data integration, computational methods

1 Introduction

The first commercially available single-cell platform emerged in 2014 (Wu et al., 2018). Over the past decade, single-cell sequencing technologies have rapidly advanced, becoming faster and more cost-effective. Today, there are over 10 different commercially available platforms for high-throughput single-cell data collection (Valihrach et al., 2018; Mereu et al., 2020). This advancement has fueled remarkable growth in the field of single-cell RNA sequencing (scRNA-seq) research,



with nearly 2000 studies published to date (Svensson et al., 2020), populating numerous databases and repositories (Regev et al., 2017; Abugessaisa et al., 2018; Franzén et al., 2019; Wang et al., 2019; Papatheodorou et al., 2020). These studies have provided valuable insights into various biological processes, including development (Han et al., 2020), disease initiation and progression (Strzelecka et al., 2018), immune response (Chen et al., 2023), and identification of rare cell types (Lee et al., 2019;

Salcher et al., 2022). Alongside the generation of large-scale single-cell data, we also observe a sharp rise in scRNA-seq analysis tools, expected to reach 3,000 by the end of 2025 (Zappia and Theis, 2021; Davis, 2019).

Previous reviews and benchmarking analyses have extensively covered various aspects of scRNA-seq analysis such as quality control (Lähmemann et al., 2020), normalization (Hafemeister and Satija, 2019; Sina Boeshaghi et al., 2022), integration (Tran et al., 2020;

Luecken et al., 2022), and cell type annotation (Abdelal et al., 2019). However, the complexity of large-scale data necessitates a comprehensive evaluation of available scRNA-seq databases and repositories. This evaluation is crucial to understand concepts like integration in the context of large-scale databasing. Understanding the scope and limitations of these databases is crucial for storing, analyzing, and interpreting single-cell data directly from these repositories. In this review, we systematically address the limitations and common assumptions of existing scRNA-seq databases. We discuss the utility of these databases to meet the specific needs of researchers studying different biological systems and processes.

2 Landscape of single-cell transcriptomics databases

The rapid expansion of single-cell RNA-seq (scRNA-seq) studies has led to the development of numerous databases and repositories for storing, retrieving, and interpreting single-cell data (Regev et al., 2017; Abugessaisa et al., 2018; Franzén et al., 2019; Wang et al., 2019; Papatheodorou et al., 2020). These databases provide a resource for single-cell transcriptomic data that can be used to build computational models to investigate various biological processes. The data in scRNA-seq databases or atlases can come from either a “primary” source, that exclusively hosts data generated by the study itself and is not shared or aggregated with data from other studies, or an “aggregated” source, where data was collected and curated from multiple studies (Figures 1A–D). Single-cell databases can be further categorized into general (non-specific or broad category of databases), tissue-specific databases, disease-specific databases, cancer-focused databases, and cell type-focused databases (Figure 1E). We have curated a comprehensive list of scRNA-seq databases, accessible here (<https://github.com/mctp/single-cell-Databases/tree/main>, Supplementary Table S1). This list includes database names, years of establishment, PubMed IDs, citation counts (as of March 31st), URLs, web interface availability, number of datasets/studies, cell counts, primary vs. aggregated distinction, specific groups, tissue types, data types, file types, normalization methods (if mentioned), data locations, and types of web interfaces.

2.1 General single-cell databases

The establishment of the Human Cell Atlas (HCA) (Regev et al., 2017) in 2017 marked a significant effort to collect and integrate large-scale single-cell data into a comprehensive reference atlas for all human cells. HCA’s open-access resource forms one of the largest public databases for integrated single-cell data from large-scale sequencing projects comprising over 437 projects and 58.5 million cells across 18 tissues. Single-cell atlases offer high-resolution views of cellular composition in organs, leading to groundbreaking discoveries of rare cell types, developmental processes, and cell states associated with various disease processes (Argelaguet et al., 2021; Rood et al., 2022; Badia-I-Mompel et al., 2023).

Other general databases include Single Cell Portal (SCP) (Tarhan et al., 2023) and CZ CellxGene Discover (CZI Single-Cell Biology Program et al., 2023) which are more flexible and are ideal for retrieving

single-cell data focusing on a particular dataset of interest, focusing on unique features and variations within the datasets. The SCP and CZ CellxGene, developed by the Chan Zuckerberg Initiative (CZI) and Broad Institute, respectively, provide web-based interfaces for data exploration and analysis. CZ CellxGene hosts more than 1,284 datasets while SCP constitutes 654 datasets. They offer interactive visualization tools for exploring gene expression patterns, cell clusters, and cell type annotations in scRNA-seq data. Both platforms support the sharing of scRNA-seq datasets, allowing researchers to collaborate and access public datasets. Using the CZ CellxGene platform, users can also download the raw count data as an RDS file containing a Seurat object or an h5ad file with an AnnData object to perform their analysis. Similarly, SCP data can be downloaded as individual metadata files, raw count expressions, and normalized expression data from the website directly, however, the availability of raw or normalized data is subjective to each study in SCP. Another interesting example of a comprehensive database is Tabula Sapiens (Sapiens Consortium* et al., 2022) which houses primary data from 15 individuals across 24 tissues. This database enables the evaluation of gene expression in normal or baseline cell states, providing a valuable resource for developing gene regulation networks and trajectories (Gondal et al., 2024a). It offers a unique opportunity to study cell type-specific expression changes. The data is easily accessible through web platforms and can also be explored using tools like CZ CellxGene.

2.2 Tissue-specific single-cell databases

While cross-tissue general databases have their advantages, such as the ability to compare gene expression across different tissues and identify commonalities or differences. However, tissue-specific databases offer a more focused and detailed view of a particular tissue’s biology, making them valuable resources for researchers studying specific tissues. These databases can also provide in-depth insights into specialized functions, cell types, and tissue-specific immunity (Elmentaite et al., 2022). Towards this goal, one of the HCA’s sub-projects aims to develop tissue-specific reference atlases that serve as consensus representations of specific organs across multiple projects (Figure 1F). These atlases provide a standardized reference for specific tissues for comparing different datasets, facilitating cross-study comparisons and meta-analyses. An example of a tissue-specific single-cell database established by the HCA is the Human Lung Cell Atlas (HLCA) (Sikkema et al., 2023) which integrates 49 lung datasets encompassing 2.4 million cells from 486 individuals. The HLCA database development involved four main steps: data curation, integration method selection, cell type annotation, and data usage. Despite the benefits associated with large-scale tissue-specific single-cell databases, it is important to note that integrating data from diverse datasets, labs, and technologies presents challenges due to differences in data quality, sample handling, and experimental protocols. Moreover, ensuring consistency and standardization across datasets is crucial for meaningful comparisons, but achieving this in practice can be complex, particularly with the wide variety of cell types and states. For the HLCA project, the team primarily relied on harmonized manual annotation, integration benchmarking (Luecken et al., 2022), and insights from experts in the field, which can be subjective and may lack reproducibility.

2.3 Disease-focused single-cell databases

These general databases are not designed to systematically gather data on gene expression specificity in different diseases. Given the diverse and heterogeneous nature of human diseases, which manifest unique gene expression profiles, there is a critical need for databases focused on disease-specific exploration (Figure 1G).

SC2disease (Zhao et al., 2021) is a manually curated scRNA-seq database that addresses this need, cataloging cell type-specific genes associated with 25 diverse diseases, including Huntington's disease, multiple sclerosis, and Alzheimer's disease. While SC2disease represents a pioneering effort in disease-specific gene expression profiling, there is also a growing need for more specific databases dedicated to the disease of interest. To address this, databases like SC2sepsis (Li et al., 2022), ssREAD (Wang C. et al., 2023), and SCovid (Qi et al., 2022) have emerged, focusing on individual diseases such as sepsis, Alzheimer's, and COVID-19, respectively. These databases aim to provide a more granular and disease-specific view of gene expression patterns, enhancing our understanding of disease mechanisms and potential therapeutic targets.

2.4 Cancer-focused single-cell databases

Cancer is a complex disease characterized by its highly heterogeneous and multifactorial nature (Gondal and Chaudhary, 2021). Traditional approaches to studying cancer, such as bulk RNA sequencing constitute a mixture of the cellular composition in tumors and often fail to accurately capture cancer cell-specific gene expression (Ding et al., 2020; Huang et al., 2023). scRNA-seq technologies offer unprecedented insights into tumor heterogeneity, evolution, and responses to therapy (Wang et al., 2020; Zeng et al., 2022; Gondal et al., 2024b). As a result, numerous databases hosting cancer-focused scRNA-seq data have emerged (Figure 1H).

One such example is CancerSEA (Yuan et al., 2019) which was launched in 2019 as a resource utilizing single-cell data from cancer datasets to decode the functional states of cancer cells, these states included stemness, invasion, metastasis, proliferation, epithelial-to-mesenchymal transition (EMT), angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, hypoxia, inflammation, and quiescence. In a salient study, Dohmen et al. (Dohmen et al., 2022) utilized CancerSEA's functional states to validate gene sets derived from their machine-learning model, while Zhao et al. (Zhao et al., 2023) demonstrated the necessity of NF-KB for initiating oncogenesis using CancerSEA's functional states. Several other studies (Lan et al., 2020; Deng et al., 2022; Tang et al., 2022; Wang L. et al., 2023) have leveraged CancerSEA to correlate their gene or gene set findings with cancer single-cell data, showcasing the utility of this resource.

However, CancerSEA has limitations, including hosting only 93,475 malignant cells and an inability to study interactions between stromal or immune cells and cancer cells. It also lacks a user-friendly web interface to support data exploration and visualization. In an attempt to overcome these challenges, TISCH was originally developed in 2021 (Sun et al., 2021), with version two released in 2023 (Han et al., 2023). TISCH2 curates cancer datasets with both malignant and non-malignant cell types, currently hosting 190 datasets, encompassing 50 cancer types, and spanning 6 million cells. To illustrate the utility

of TISCH in computational models, Xu et al. (2023) employed TISCH's data to analyze the correlation between FOXM1 and immune cells. Similarly, Zhang et al. (2023) employed TISCH to evaluate m7G regulators expression in osteosarcoma scRNA-seq data. As such, numerous studies have utilized TISCH to evaluate the expression of genes of interest across cancer datasets (Zhao et al., 2022; Liu et al., 2023a; Benedetti et al., 2023; Liu et al., 2023b; Liu Y. et al., 2023; Zhang et al., 2023). Although TISCH provides a valuable resource to the cancer research community, it is important to be aware of the assumptions and limitations of TISCH data. While many studies aim to understand gene expression in malignant cells, TISCH also contains treatment data from immune checkpoint blockade (ICB), chemotherapy, and targeted therapy. Therefore, it is important to ensure that the results are not confounded, as gene expression varies after treatments and can yield diverse results (Liu et al., 2024). Additionally, TISCH includes data from multiple stages of cancer, such as primary tumors or metastatic sites. Therefore, users need to carefully extract only relevant information when employing TISCH data. TISCH employs an automatic cell-type annotation method, which may lead to a lack of consensus with the original dataset's manual annotation. Importantly, all downloaded datasets in TISCH are in fixed expression matrices (Zeng et al., 2022), and users cannot download the raw count data. Therefore, any attempts to further integrate or normalize the data might result in technical variation rather than biological results. These limitations could potentially introduce bias into analyses or hinder comparability across different databases.

2.5 Cell-type-focused single-cell databases

To better understand the intricacies of cell biology, dedicated resources focused on cell-type profiling of single cells have emerged. JingleBells (Ner-Gaon et al., 2017), introduced in 2017, represented an advancement in this direction by providing a comprehensive immune cell resource. JingleBells facilitates the study of immune cell involvement in various diseases, including cancer, and infectious diseases, providing valuable insights into disease mechanisms and potential therapeutic targets. However, JingleBells lacks an interactive web interface and only allows for BAM file download which means analyzing and interpreting single-cell data from JingleBells requires specialized computational tools and expertise, limiting accessibility to researchers with specific skills. In comparison, the human Antigen Receptor database (huARdb) (Wu et al., 2022), published originally in 2022, is a comprehensive human single-cell immune profiling database, housing 444,794 high-confidence T or B cells (hcT/B cells) with complete TCR/BCR sequences and transcriptomes sourced from 215 datasets. To enhance user experience, the authors have created a user-friendly web interface that offers interactive visualization modules, enabling biologists to analyze transcriptome and TCR/BCR features at the single-cell level with ease. Fan et al. (2023a) utilized huARdb by analyzing ulcerative colitis (UC) patients' immune cells derived from huARdb. Similarly, they also employed huARdb to investigate the healthy and UC composition of peripheral blood immune cells and colonic cells (Fan et al., 2023b). Additional cell-type-focused single-cell databases include EndoDB (Khan et al., 2019) which hosts endothelial cells transcriptomics data from 360 datasets and ABC portal (Gao et al.,

2023), a database for blood cells across 198 datasets, allowing for a blood cell-type-specific exploration.

3 Challenges associated with the utilization of large-scale single-cell databases and their examples from current literature

While the scRNA-seq field is progressing towards the improvement and development of large-scale single-cell databases, their application in research comes with certain caveats and despite their vastness, they must be used judiciously (Figure 11). Some of the key considerations and limitations include:

3.1 Data quality

Ensuring data quality in scRNA-seq is critical for accurate interpretation and analysis. A fundamental assumption of droplet-based scRNA-seq is that each droplet, where molecular tagging and reverse transcription occur, contains messenger RNA (mRNA) from a single cell. However, in practice, this assumption is often violated, leading to potential distortions in the interpretation of scRNA-seq data. Common examples include doublets which are real cells however they contain multiple cells and dropouts where the expression of one gene was not detected in one cell. This becomes a major issue because large-scale databases such as the Human Cell Atlas (HCA) rely heavily on the accuracy and cellular specificity of transcriptional readouts generated by scRNA-seq.

3.1.1 Examples from current literature and benchmarking studies

3.1.1.1 Dropouts

To overcome the issue of dropouts, numerous single-cell imputation methods have been developed. However, imputation affects downstream results and some of these methods may introduce false correlations. For example, Breda et al. (2021)'s comparison of MAGIC (van Dijk et al., 2018) results with Sanity (SAmpling-Noise-corrected Inference of Transcription activitY), elicited that MAGIC introduced strong positive correlations where no or low correlation was expected. A comparative study by Zhang and Zhang (2020) highlighted that the number of cells and method parameters also affected imputation results and some methods preferred similar cells while imputing. Therefore, imputation results can be variable, and downstream analysis will be affected by imputation.

3.1.1.2 Doublets

Doublets are major confounders in scRNA-seq data analysis. However, there are computational methods that exist to detect doublets in single-cell data. A benchmarking study (Xi and Li, 2021) compared nine doublet detection methods, revealing that there is still room for improvement in detection accuracy. Generally, these methods performed better on datasets with higher doublet rates, larger sequencing depths, more cell types, or greater heterogeneity between cell types. However, the removal of doublets by these methods led to improvements in various downstream analyses. It enhanced the identification of

Differentially Expressed (DE) genes, reduced the presence of spurious cell clusters, and improved the inference of cell trajectories. However, the extent of improvement varied across different methods, highlighting the need for further refinement and development in this area.

3.2 Normalization

Normalization is another critical aspect of scRNA-seq data analysis and can be a complex problem when dealing with multiple datasets. Specifically, variability in experimental protocols and data processing methods can pose challenges in data normalization, affecting the comparability of results across datasets in a database. Differences in normalization approaches can lead to discrepancies in gene expression profiles, making it difficult to draw meaningful conclusions from the downstream analyses.

3.2.1 Examples from current literature and benchmarking studies

There are several methods to perform single-cell data normalization such as SCT transformation, and log1p normalization (Hafemeister and Satija, 2019). The choice of the method, however, is dependent on various features of the data including sequencing depth as both lowly and highly abundance genes are confounded by sequencing depth (Hafemeister and Satija, 2019). Sina Boeshaghi et al. (2022) demonstrated that the assumptions implied in the choice of normalization methods will affect downstream analysis in determining whether the variation is technical or biological. In a salient example, TISCH2 (Han et al., 2023) database hosts single-cell gene expression matrices for each dataset. In our analysis of TISCH2 data, this matrix is already normalized and integrated, users incorporating this data in their research need to be aware of this normalization to make accurate assessments of data and not re-normalize or merge it directly with other datasets which might result in substandard results. Therefore, in our opinion, when using datasets directly from single-cell databases it is necessary to be aware of the pre-processing steps and how they affect downstream results to ensure accurate analysis and interpretation.

3.3 Integration and batch effects removal

The integration and batch effect removal of scRNA-seq data from diverse datasets, labs, and technologies can be complex (Stuart et al., 2019). Variations in data formats, processing pipelines, and batch effects can affect the robustness and reliability of integrated analyses, potentially masking true biological signals. Methods for integrating heterogeneous datasets are continually evolving, with efforts focused on minimizing batch effects and preserving biological variability. There are more than 50 integration methods published to date (Zappia et al., 2018; Luecken et al., 2022).

3.3.1 Examples from current literature and benchmarking studies

Large databases host numerous datasets from multiple studies, however, it is also important to be aware of the properties associated with each study during integration. For

example, [Salcher et al. \(2022\)](#) established a large non-small cell lung cancer (NSCLC) atlas comprising 29 datasets spanning 1,283,972 single cells from 556 samples. Although this effort resulted in the in-depth characterization of a neutrophil subpopulation, however, according to our re-analysis of this data, among the 29 datasets, [Maynard et al. \(2020\)](#)'s NSCLC samples were also incorporated which were not treatment-naive. This can be a potential confounder in downstream analysis. Therefore, it is the user's responsibility to be aware of this data-specific property and to use atlases and databases with care to derive robust biological insights. Similarly, several attempts have been made to benchmark integration methods for single-cell data ([Tran et al., 2020](#); [Luecken et al., 2022](#)). While [Tran et al. \(2020\)](#) showed that LIGER ([Liu et al., 2020](#)), Seurat 3 ([Satija et al., 2015](#)), and Harmony ([Korsunsky et al., 2019](#)) performed the best among 11 other methods, [Luecken et al. \(2022\)](#) revealed that LIGER and Seurat v3 favor the removal of batch effect over the conservation of biological variation. This highlights the importance of considering the dataset and the specific research question when selecting an integration method. Selecting the right method is crucial as it directly impacts the biological insights that can be generated from the integrated data.

3.4 Cell-type annotation

Accurate annotation of cell types in scRNA-seq databases is crucial for interpreting results accurately. Harmonizing cell type annotations across different datasets is essential for facilitating cross-study comparisons and meta-analysis. While automatic cell-type annotation methods are convenient, they may lack consensus with manual annotations from original datasets. This can introduce ambiguity in cell-type assignments and lead to misinterpretation of results.

3.4.1 Examples from current literature and benchmarking studies

In our recent re-analysis of Tabula Sapiens data ([Sapiens et al., 2022](#)), we observed that 10% of the heart cells were mislabelled as hepatocytes in the study's original metadata. This is biologically incorrect since hepatocytes cannot be in the heart, these are liver epithelial cells ([Gong et al., 2022](#)). One potential reason for this mislabelling can be that Tabula Sapiens data was annotated using an automatic cell-type annotation tool, another reason could be sample mishandling. Therefore, diligent manual intervention for cell type annotation needs to be practiced to ensure accurate and robust results. Additionally, [Abdelaal et al. \(2019\)](#) carried out a performance comparison analysis between 22 automatic cell-type identification methods in single-cell data. Although the authors did not state a preference they noted that the results can vary depending on input features and the number of cells which means that they cannot be solely relied on, there will be some manual intervention for accurate cell type annotation.

3.5 "Zero-code" single-cell analysis platforms

Single-cell data plays a crucial role in validating and enhancing the accuracy of wet lab results and hypothesis-driven publications

([Bao et al., 2023](#)). To facilitate easy access and analysis of this data, many databases provide built-in tools that allow researchers without computational expertise to explore existing datasets and assess their hypotheses using basic operations like exploring gene expressions, isolating cell subsets for individual analysis, and identifying clusters within the data. However, for more complex analyses that require significant computational resources, these tools are often not available directly on the database platforms.

To address this challenge, several web-based platforms have been created to enable online analysis of scRNA-seq data. Among them, the Automated Single-cell Analysis Pipeline (ASAP) ([Gardeux et al., 2017](#)) was published in 2017 and provided basic processing analysis of scRNA-seq data post-alignment from filtering to cell type annotation and functional gene set enrichment. Later, this web-based pipeline improved in 2020 ([David et al., 2020](#)) with more scalable options. However, ASAP did not include advanced scRNA-seq analysis tools such as regulon activity assessment. To address this issue, ICARUS_v2 ([Jiang et al., 2023a](#)) was launched which also incorporates the Drug-Gene Interaction database to facilitate drug discovery. This platform was improved and ICARUS_v3 ([Jiang et al., 2023b](#)) was published for zero-code single-cell analysis. ICARUS_v3 employs a geometric cell sketching method to subsample representative cells from the dataset to store in memory. This enables advanced scRNA-seq analysis through a user-friendly web interface. ICARUS_v3 can seamlessly integrate with output files from databases like Single Cell Portal (SCP) ([Tarhan et al., 2023](#)) and CZ CellxGene Discover (CZI Single-Cell Biology Program [et al., 2023](#)), eliminating the need for coding expertise. Users can leverage this platform to conduct a wide range of analyses, including differential expression analysis, gene regulatory network construction, trajectory analysis, and cell-cell communication inference. While such tools facilitate online analysis of scRNA-seq data, offering user-friendly interfaces and automated workflows. However, these platforms come with limitations and assumptions. They often assume users have high-quality, pre-processed data and a stable internet connection, which may not always be the case. The platforms also impose constraints on data size and complexity due to server limitations, potentially limiting the depth of analysis for large or intricate datasets. Additionally, the algorithms and default parameters embedded in these platforms may not be optimal for all types of scRNA-seq data, leading to less tailored analyses compared to custom pipelines. Despite these limitations, such platforms provide valuable accessibility and convenience for many researchers.

4 Platforms for hosting and visualizing large-scale single-cell data

As the volume of single-cell data continues to grow, scalability becomes a significant concern. Developing methods and infrastructure that can handle the increasing complexity and size of single-cell datasets is crucial for future research. Towards this aim, for easy, fast, and customizable exploration of single-cell data for public use, numerous user-driven platforms have emerged ([Rue-Albrecht et al., 2018](#); [Feng et al., 2019](#)).

One such platform, the Interactive SummarizedExperiment Explorer (iSEE) ([Rue-Albrecht et al., 2018](#)), launched in 2018,

enables users to host their SummarizedExperiment data. Researchers such as Graf et al. (2019) and Newton et al. (2022) have employed iSEE to visualize their single-cell data, demonstrating its utility in data exploration. Similarly, the Single Cell Explorer (Feng et al., 2019) allows users to input loom and Seurat objects, making the data more accessible.

ShinyCell (Ouyang et al., 2021) is another example of a platform offering web-based interfaces for exploring and analyzing data. These interfaces can be customized for maximum usability and can be uploaded to online platforms to broaden access to published data. ShinyCell supports various common single-cell data formats, including SingleCellExperiment, h5ad, loom, and Seurat objects, as inputs. In a salient example, Ma et al. (2023) used ShinyCell to host their pan-cancer single-cell data, showcasing its versatility and effectiveness in data dissemination. Likewise, Curras-Alonso et al. (2023) developed their web application using ShinyCell, highlighting its widespread adoption in the research community. By providing easy-to-use tools for data analysis, these platforms help democratize access to single-cell data and facilitate collaboration between researchers from different disciplines.

5 Discussion

The rapid expansion of single-cell RNA-seq (scRNA-seq) studies has ushered in a plethora of databases and repositories dedicated to storing, retrieving, and interpreting single-cell data. These databases provide a wealth of single-cell transcriptomic data that can be used to build computational models to understand various biological processes. However, challenges such as data quality, normalization, integration, and annotation can affect the reliability and comparability of results across different datasets and studies.

While the existing databases are valuable for basic scRNA-seq analysis, they cannot often perform advanced analyses such as regulon activity assessment, pseudobulking, and differential gene expression analysis. Users still need to possess programming skills and be familiar with using a command-line interface to conduct customized analysis. Furthermore, many wet labs may not have the necessary resources to manage high-performance computing clusters. To address this gap and enable wet-lab researchers to conduct advanced scRNA-seq analysis, platforms like ICARUS_v3 (Jiang et al., 2023b) offer web-based analysis tools. These platforms provide an accessible way for researchers to explore and analyze single-cell data, bridging the gap between wet lab experimentation and bioinformatics analysis.

Taken together, in this mini-review, we address the utility and applicability of large-scale scRNA-seq databases. We address some of the challenges and common assumptions that need to be considered when using these databases for hypothesis-driven studies, highlighting platforms for hosting customized scRNA-seq data for community usage. While challenges remain, the development of user-friendly platforms is narrowing the gap between wet-lab experimentation and bioinformatics analysis, ultimately advancing our understanding of cellular processes at a single-cell level.

Author contributions

MG: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Visualization, Writing—original

draft, Writing—review and editing. SS: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing. AC: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing. MC: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study was supported by the National Cancer Institute (NCI) Outstanding Investigator Award R35CA231996 (AC), NCI Prostate SPOR grant P50CA186786 (AC), and NCI Michigan-VUMC Biomarker Characterization Center grant U2CCA271854 (AC). AC is also a Howard Hughes Medical Institute Investigator, A. Alfred Taubman Scholar, and American Cancer Society Professor. This manuscript was also supported in part by funding from the Innovation in Cancer Informatics (ICI398672) and the V Foundation for Cancer Research (T2019-006) to MC.

Acknowledgments

We would like to acknowledge the use of ChatGPT, an AI language model developed by OpenAI, for assisting with language editing and suggestions during the preparation of this manuscript.

Conflict of interest

AC is a co-founder of and serves as a Scientific Advisory Board member for LynxDx, Esanik Therapeutics, Medsyn, and Flamingo Therapeutics. AC is a scientific advisor or consultant for EdenRoc, Aurigene Oncology, Ascentage Pharma, Proteovant, Belharra, Rapta Therapeutics, and Tempus.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2024.1417428/full#supplementary-material>

References

- Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M. J. T., et al. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.* 20 (1), 194. doi:10.1186/s13059-019-1795-z
- Abugessaisa, I., Noguchi, S., Böttcher, M., Hasegawa, A., Kouno, T., Kato, S., et al. (2018). SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.* 46 (D1), D781–D787. doi:10.1093/nar/gkx949
- Argelaguet, R., Cuomo, A. S. E., Stegle, O., and Marioni, J. C. (2021). Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.* 39 (10), 1202–1215. doi:10.1038/s41587-021-00895-7
- Badia-I-Mompel, P., Wessels, L., Müller-Dott, S., Trimbou, R., Ramirez Flores, R. O., Argelaguet, R., et al. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* 24 (11), 739–754. doi:10.1038/s41576-023-00618-5
- Bao, Y., Qiao, Y., Choi, J. E., Zhang, Y., Mannan, R., Cheng, C., et al. (2023). Targeting the lipid kinase PIKfyve upregulates surface expression of MHC class I to augment cancer immunotherapy. *Proc. Natl. Acad. Sci. U. S. A.* 120 (49), e2314416120. doi:10.1073/pnas.2314416120
- Benedetti, E., Liu, E. M., Tang, C., Kuo, F., Buyukozkan, M., Park, T., et al. (2023). A multimodal atlas of tumour metabolism reveals the architecture of gene-metabolite covariation. *Nat. Metab.* 5 (6), 1029–1044. doi:10.1038/s42255-023-00817-8
- Breda, J., Zavolan, M., and van Nimwegen, E. (2021). Bayesian inference of gene expression states from single-cell RNA-seq data. *Nat. Biotechnol.* 39 (8), 1008–1016. doi:10.1038/s41587-021-00875-x
- Chen, D., Luo, Y., and Cheng, G. (2023). Single cell and immunity: better understanding immune cell heterogeneities with single cell sequencing. *Clin. Transl. Med.* 13 (1), e1159. doi:10.1002/ctm2.1159
- Curras-Alonso, S., Soulier, J., Defard, T., Weber, C., Heinrich, S., Laporte, H., et al. (2023). An interactive murine single-cell atlas of the lung responses to radiation injury. *Nat. Commun.* 14 (1), 2445. doi:10.1038/s41467-023-38134-z
- Czi Single-Cell Biology Program Abdulla, S., Aebermann, B., Assis, P., Badajoz, S., Bell, S. M., Bezzi, E., et al. (2023). CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. bioRxiv. Available at: <https://www.biorxiv.org/content/10.1101/2023.10.30.563174v1>.
- David, F. P. A., Litovchenko, M., Deplancke, B., and Gardeux, V. (2020). ASAP 2020 update: an open, scalable and interactive web-based portal for (single-cell) omics analyses. *Nucleic Acids Res.* 48 (W1), W403–W414. doi:10.1093/nar/gkaa412
- Davis, S. (2019). Awesome-single-cell: community-curated list of software packages and data resources for single-cell, including RNA-seq, ATAC-seq, etc. Available at: <https://github.com/seandavi/awesome-single-cell>.
- Deng, L., Jiang, A., Zeng, H., Peng, X., and Song, L. (2022). Comprehensive analyses of PDHA1 that serves as a predictive biomarker for immunotherapy response in cancer. *Front. Pharmacol.* 13, 947372. doi:10.3389/fphar.2022.947372
- Ding, S., Chen, X., and Shen, K. (2020). Single-cell RNA sequencing in breast cancer: understanding tumour heterogeneity and paving roads to individualized therapy. *Cancer Commun.* 40 (8), 329–344. doi:10.1002/cac2.12078
- Dohmen, J., Baranovskii, A., Ronen, J., Uyar, B., Franke, V., and Akalin, A. (2022). Identifying tumor cells at the single-cell level using machine learning. *Genome Biol.* 23 (1), 123. doi:10.1186/s13059-022-02683-1
- Elmentaite, R., Dominguez Conde, C., Yang, L., and Teichmann, S. A. (2022). Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* 23 (7), 395–410. doi:10.1038/s41576-022-00449-w
- Fan, Q., Dai, W., Li, M., Wang, T., Li, X., Deng, Z., et al. (2023b). Inhibition of α 2,6-sialyltransferase relieves symptoms of ulcerative colitis by regulating Th17 cells polarization. *Int. Immunopharmacol.* 125 (Pt A), 111130. doi:10.1016/j.intimp.2023.111130
- Fan, Q., Li, M., Zhao, W., Zhang, K., Li, M., and Li, W. (2023a). Hyper α 2,6-Sialylation promotes CD4+ T-cell activation and induces the occurrence of ulcerative colitis. *Adv. Sci.* 10 (26), e2302607. doi:10.1002/advs.202302607
- Feng, D., Whitehurst, C. E., Shan, D., Hill, J. D., and Yue, Y. G. (2019). Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNA-seq data. *BMC Genomics* 20 (1), 676. doi:10.1186/s12864-019-6053-y
- Franzén, O., Gan, L. M., and Björkregren, J. L. M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, baz046. doi:10.1093/database/baz046
- Gao, X., Hong, F., Hu, Z., Zhang, Z., Lei, Y., Li, X., et al. (2023). ABC portal: a single-cell database and web server for blood cells. *Nucleic Acids Res.* 51 (D1), D792–D804. doi:10.1093/nar/gkac646
- Gardeux, V., David, F. P. A., Shajkofci, A., Schwalie, P. C., and Deplancke, B. (2017). ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. *Bioinformatics* 33 (19), 3123–3125. doi:10.1093/bioinformatics/btx337
- Gondal, M. N., and Chaudhary, S. U. (2021). Navigating multi-scale cancer systems biology towards model-driven clinical Oncology and its applications in personalized therapeutics. *Front. Oncol.* 11, 712505. doi:10.3389/fonc.2021.712505
- Gondal, M. N., Cieslik, M., and Chinnaiyan, A. M. (2024b). Integrated cancer cell-specific single-cell RNA-seq datasets of immune checkpoint blockade-treated patients. *bioRxiv* 3, 576110. doi:10.1101/2024.01.17.576110
- Gondal, M. N., Mannan, R., Bao, Y., Hu, J., Cieslik, M., and Chinnaiyan, A. M. (2024a). Abstract 860: pan-tissue master regulator inference reveals mechanisms of MHC alterations in cancers. *Cancer Res.* 84 (6_Suppl. ment), 860. doi:10.1158/1538-7445.am2024-860
- Gong, J., Tu, W., Liu, J., and Tian, D. (2022). Hepatocytes: a key role in liver inflammation. *Front. Immunol.* 13, 1083780. doi:10.3389/fimmu.2022.1083780
- Graf, C., Wilgenbus, P., Pagel, S., Pott, J., Marini, F., Reyda, S., et al. (2019). Myeloid cell-synthesized coagulation factor X dampens antitumor immunity. *Sci. Immunol.* 4 (39), eaaw8405. doi:10.1126/sciimmunol.aaw8405
- Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20 (1), 296. doi:10.1186/s13059-019-1874-1
- Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., et al. (2020). Construction of a human cell landscape at single-cell level. *Nature* 581 (7808), 303–309. doi:10.1038/s41586-020-2157-4
- Han, Y., Wang, Y., Dong, X., Sun, D., Liu, Z., Yue, J., et al. (2023). TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res.* 51 (D1), D1425–D1431. doi:10.1093/nar/gkac959
- Huang, D., Ma, N., Li, X., Gou, Y., Duan, Y., Liu, B., et al. (2023). Advances in single-cell RNA sequencing and its applications in cancer research. *J. Hematol. Oncol.* 16 (1), 98. doi:10.1186/s13045-023-01494-6
- Jiang, A., Snell, R. G., and Lehnert, K. (2023b). ICARUS v3, a massively scalable web server for single cell RNA-seq analysis of millions of cells. bioRxiv. Available at: <https://www.biorxiv.org/content/biorxiv/early/2023/11/21/2023.11.20.567692>.
- Jiang, A., You, L., Snell, R. G., and Lehnert, K. (2023a). Delineation of complex gene expression patterns in single cell RNA-seq data with ICARUS v2.0. *Nar. Genom. Bioinform* 5 (2), lqad032. doi:10.1093/nargab/lqad032
- Khan, S., Taverna, F., Rohlenova, K., Treps, L., Geldhof, V., de Rooij, L., et al. (2019). EndoDB: a database of endothelial cell transcriptomics data. *Nucleic Acids Res.* 47 (D1), D736–D744. doi:10.1093/nar/gky997
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16 (12), 1289–1296. doi:10.1038/s41592-019-0619-0
- Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21 (1), 31. doi:10.1186/s13059-020-1926-6
- Lan, Z., Yao, X., Sun, K., Li, A., Liu, S., and Wang, X. (2020). The interaction between lncRNA SNHG6 and hnRNP A1 contributes to the growth of colorectal cancer by enhancing aerobic glycolysis through the regulation of alternative splicing of PKM. *Front. Oncol.* 10, 363. doi:10.3389/fonc.2020.00363
- Lee, H., Yu, H., and Welch, J. (2019). A beginner's guide to single-cell transcriptomics. *Biochem.* 41 (5), 34–38. doi:10.1042/bio04105034
- Li, Y., Tan, R., Chen, Y., Liu, Z., Chen, E., Pan, T., et al. (2022). SC2sepsis: sepsis single-cell whole gene expression database. *Database*, baac061. doi:10.1093/database/baac061
- Liu, J., Gao, C., Sodicoff, J., Kozareva, V., Macosko, E. Z., and Welch, J. D. (2020). Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat. Protoc.* 15 (11), 3632–3662. doi:10.1038/s41596-020-0391-8
- Liu, Y., Altreuter, J., Bodapati, S., Cristea, S., Wong, C. J., Wu, C. J., et al. (2024). Predicting patient outcomes after treatment with immune checkpoint blockade: a review of biomarkers derived from diverse data modalities. *Cell Genom* 4 (1), 100444. doi:10.1016/j.xgen.2023.100444
- Liu, Y., Yang, J., Wang, T., Luo, M., Chen, Y., Chen, C., et al. (2023c). Expanding PROTACTable genome universe of E3 ligases. *Nat. Commun.* 14 (1), 6509. doi:10.1038/s41467-023-42233-2
- Liu, Y., 'e, Lu, S., Sun, Y., Wang, F., Yu, S., Chen, X., et al. (2023a). Deciphering the role of QPCTL in glioma progression and cancer immunotherapy. *Front. Immunol.* 14, 1166377. doi:10.3389/fimmu.2023.1166377
- Liu, Y., Zhao, S., Chen, Y., Ma, W., Lu, S., He, L., et al. (2023b). Vimentin promotes glioma progression and maintains glioma cell resistance to oxidative phosphorylation inhibition. *Cell Oncol.* 46 (6), 1791–1806. doi:10.1007/s13402-023-00844-3
- Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* 19 (1), 41–50. doi:10.1038/s41592-021-01336-8
- Ma, C., Yang, C., Peng, A., Sun, T., Ji, X., Mi, J., et al. (2023). Pan-cancer spatially resolved single-cell analysis reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment. *Mol. Cancer* 22 (1), 170. doi:10.1186/s12943-023-01876-x

- Maynard, A., McCoach, C. E., Rotow, J. K., Harris, L., Haderk, F., Kerr, D. L., et al. (2020). Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* 182 (5), 1232–1251.e22. doi:10.1016/j.cell.2020.07.017
- Mereu, E., Lafzi, A., Moutinho, C., Ziegenhain, C., McCarthy, D. J., Álvarez-Varela, A., et al. (2020). Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* 38 (6), 747–755. doi:10.1038/s41587-020-0469-4
- Ner-Gaon, H., Melchior, A., Golan, N., Ben-Haim, Y., and Shay, T. (2017). JingleBells: a repository of immune-related single-cell RNA-sequencing datasets. *J. Immunol.* 198 (9), 3375–3379. doi:10.4049/jimmunol.1700272
- Newton, A. H., Williams, S. M., Major, A. T., and Smith, C. A. (2022). Cell lineage specification and signalling pathway use during development of the lateral plate mesoderm and forelimb mesenchyme. *Development* 149 (18), dev200702. doi:10.1242/dev.200702
- Ouyang, J. F., Kamaraj, U. S., Cao, E. Y., and Rackham, O. J. L. (2021). ShinyCell: simple and sharable visualization of single-cell gene expression data. *Bioinformatics* 37 (19), 3374–3376. doi:10.1093/bioinformatics/btab209
- Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M. P., George, N., Fexova, S., et al. (2020). Expression Atlas update: from tissues to single cells. *Nucleic Acids Res.* 48 (D1), D77–D83. doi:10.1093/nar/gkz947
- Qi, C., Wang, C., Zhao, L., Zhu, Z., Wang, P., Zhang, S., et al. (2022). SCovid: single-cell atlases for exposing molecular characteristics of COVID-19 across 10 human tissues. *Nucleic Acids Res.* 50 (D1), D867–D874. doi:10.1093/nar/gkab881
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., et al. (2017). The human cell atlas. *Elife* 6, e27041. doi:10.7554/eLife.27041
- Rood, J. E., Maartens, A., Hupalowska, A., Teichmann, S. A., and Regev, A. (2022). Impact of the human cell atlas on medicine. *Nat. Med.* 28 (12), 2486–2496. doi:10.1038/s41591-022-02104-7
- Rue-Albrecht, K., Marini, F., Soneson, C., and Lun, A. T. L. (2018). iSEE: interactive SummarizedExperiment explorer. *F1000Res* 7, 741. doi:10.12688/f1000research.14966.1
- Salcher, S., Sturm, G., Horvath, L., Untergasser, G., Kuempers, C., Fotakis, G., et al. (2022). High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* 40 (12), 1503–1520.e8. doi:10.1016/j.ccell.2022.10.008
- Sapiens, T., Jones, R. C., Karkanas, J., Krasnow, M. A., Pisco, A. O., Quake, S. R., et al. (2022). The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376 (6594), eabl4896. doi:10.1126/science.abl4896
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33 (5), 495–502. doi:10.1038/nbt.3192
- Sikkema, L., Ramírez-Suástegui, C., Strobl, D. C., Gillett, T. E., Zappia, L., Madisson, E., et al. (2023). An integrated cell atlas of the lung in health and disease. *Nat. Med.* 29 (6), 1563–1577. doi:10.1038/s41591-023-02327-2
- Sina Boeshaghi, A., Hallgrímsson, I. B., Gálvez-Merchán, Á., and Pachter, L. (2022). Depth normalization for single-cell genomics count data. *bioRxiv*. doi:10.1101/2022.05.06.490859v1
- Strzelecka, P. M., Ranzoni, A. M., and Cvejic, A. (2018). Dissecting human disease with single-cell omics: application in model systems and in the clinic. *Dis. Model Mech.* 11 (11), dmm036525. doi:10.1242/dmm.036525
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177 (7), 1888–1902.e21. doi:10.1016/j.cell.2019.05.031
- Sun, D., Wang, J., Han, Y., Dong, X., Ge, J., Zheng, R., et al. (2021). TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 49 (D1), D1420–D1430. doi:10.1093/nar/gkaa1020
- Svensson, V., da Veiga Beltrame, E., and Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. *Database*, baaa073. doi:10.1093/database/baaa073
- Tang, Y., Kwiatkowski, D. J., and Henske, E. P. (2022). Midkine expression by stem-like tumor cells drives persistence to mTOR inhibition and an immune-suppressive microenvironment. *Nat. Commun.* 13 (1), 5018. doi:10.1038/s41467-022-32673-7
- Tarhan, L., Bistline, J., Chang, J., Galloway, B., Hanna, E., and Weitz, E. (2023). Single Cell Portal: an interactive home for single-cell genomics data. *bioRxiv*, 2023.07.13.548886. doi:10.1101/2023.07.13.548886
- Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., et al. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 21 (1), 12. doi:10.1186/s13059-019-1850-9
- Valihrach, L., Androvic, P., and Kubista, M. (2018). Platforms for single-cell collection and analysis. *Int. J. Mol. Sci.* 19 (3), 807. doi:10.3390/ijms19030807
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174 (3), 716–729.e27. doi:10.1016/j.cell.2018.05.061
- Wang, C., McNutt, M., Ma, A., Fu, H., and Ma, Q. (2023a). ssREAD: a single-cell and spatial RNA-seq database for alzheimer's disease. *bioRxiv*. doi:10.1101/2023.09.08.556944
- Wang, L., Cao, Y., Guo, W., and Xu, J. (2023b). High expression of cuproptosis-related gene FDX1 in relation to good prognosis and immune cells infiltration in colon adenocarcinoma (COAD). *J. Cancer Res. Clin. Oncol.* 149 (1), 15–24. doi:10.1007/s00432-022-04382-7
- Wang, Y., Mashock, M., Tong, Z., Mu, X., Chen, H., Zhou, X., et al. (2020). Changing technologies of RNA sequencing and their applications in clinical Oncology. *Front. Oncol.* 10, 447. doi:10.3389/fonc.2020.00447
- Wang, Z., Feng, X., and Li, S. C. (2019). SCDevDB: a database for insights into single-cell gene expression profiles during human developmental processes. *Front. Genet.* 10, 903. doi:10.3389/fgene.2019.00903
- Wu, L., Xue, Z., Jin, S., Zhang, J., Guo, Y., Bai, Y., et al. (2022). huARdb: human Antigen Receptor database for interactive clonotype-transcriptome analysis at the single-cell level. *Nucleic Acids Res.* 50 (D1), D1244–D1254. doi:10.1093/nar/gkab857
- Wu, X., Yang, B., Udo-Inyang, I., Ji, S., Ozog, D., Zhou, L., et al. (2018). Research techniques made simple: single-cell RNA sequencing and its applications in dermatology. *J. Invest Dermatol.* 138 (5), 1004–1009. doi:10.1016/j.jid.2018.01.026
- Xi, N. M., and Li, J. J. (2021). Benchmarking computational doublet-detection methods for single-cell RNA sequencing data. *Cell Syst.* 12 (2), 176–194.e6. doi:10.1016/j.cels.2020.11.008
- Xu, Z., Pei, C., Cheng, H., Song, K., Yang, J., Li, Y., et al. (2023). Comprehensive analysis of FOXM1 immune infiltrates, m6a, glycolysis and ceRNA network in human hepatocellular carcinoma. *Front. Immunol.* 14, 1138524. doi:10.3389/fimmu.2023.1138524
- Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., et al. (2019). CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* 47 (D1), D900–D908. doi:10.1093/nar/gky939
- Zappia, L., Phipson, B., and Oshlack, A. (2018). Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* 14 (6), e1006245. doi:10.1371/journal.pcbi.1006245
- Zappia, L., and Theis, F. J. (2021). Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol.* 22 (1), 301. doi:10.1186/s13059-021-02519-4
- Zeng, J., Zhang, Y., Shang, Y., Mai, J., Shi, S., Lu, M., et al. (2022). CancerSCEM: a database of single-cell expression map across various human cancers. *Nucleic Acids Res.* 50 (D1), D1147–D1155. doi:10.1093/nar/gkab905
- Zhang, L., and Zhang, S. (2020). Comparison of computational methods for imputing single-cell RNA-sequencing data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17 (2), 376–389. doi:10.1109/TCBB.2018.2848633
- Zhang, Y., Gan, W., Ru, N., Xue, Z., Chen, W., Chen, Z., et al. (2023). Comprehensive multi-omics analysis reveals m7G-related signature for evaluating prognosis and immunotherapy efficacy in osteosarcoma. *J. Bone Oncol.* 40, 100481. doi:10.1016/j.jbo.2023.100481
- Zhao, M., Chauhan, P., Sherman, C. A., Singh, A., Kaileh, M., Mazan-Mamczarz, K., et al. (2023). NF-κB subunits direct kinetically distinct transcriptional cascades in antigen receptor-activated B cells. *Nat. Immunol.* 24 (9), 1552–1564. doi:10.1038/s41590-023-01561-7
- Zhao, S., Chi, H., Ji, W., He, Q., Lai, G., Peng, G., et al. (2022). A bioinformatics-based analysis of an anoikis-related gene signature predicts the prognosis of patients with low-grade gliomas. *Brain Sci.* 12 (10), 1349. doi:10.3390/brainsci12101349
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., et al. (2021). SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.* 49 (D1), D1413–D1419. doi:10.1093/nar/gkaa838