TITLE: singlecellVR: interactive visualization of single-cell data in virtual reality

AUTHORS: David F. Stein, Huidong Chen, Michael E. Vinyard, & Luca Pinello

AUTHOR INFORMATION

Contact information:

David F. Stein: david.stein@icahn.mssm.edu
Huidong Chen: huidong.chen@mgh.harvard.edu
Michael E. Vinyard: vinyard@g.harvard.edu
Luca Pinello: lpinello@mgh.harvard.edu

Affiliations:

1. **Graduate School of Biomedical Sciences, Icahn School Of Medicine at Mount Sinai, New York, NY, 10029, USA**
   David F. Stein

2. **Khoury College of Computer Science, Northeastern University, Boston, MA 02115, USA**
   David F. Stein

3. **Molecular Pathology Unit, Massachusetts General Hospital Research Institute, Charlestown, MA, 02129, USA**
   Huidong Chen, Michael E. Vinyard & Luca Pinello

4. **Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, 02129, USA**
   Huidong Chen, Michael E. Vinyard & Luca Pinello

5. **Department of Pathology, Harvard Medical School, Boston, MA, 02115, USA**
   Huidong Chen, Michael E. Vinyard, & Luca Pinello

6. **Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA**
   Huidong Chen, Michael E. Vinyard, & Luca Pinello

7. **Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02142, USA**
   Michael E. Vinyard

Author notes:

David F. Stein, Huidong Chen, and Michael E. Vinyard contributed equally to this work.

Corresponding author:

Correspondence to Luca Pinello.

ABSTRACT

Single-cell assays have transformed our ability to model heterogeneity within cell populations and tissues. Virtual reality has recently emerged as a powerful technology to dynamically explore complex data. However, expensive hardware or advanced data preprocessing skills are required to adapt such technology to single-cell data. To address current shortcomings, we built *singlecellVR*, a user-friendly website for visualizing single-cell data, designed for cheap and easily available virtual reality hardware (e.g., *Google Cardboard,* ~$8). We provide a companion package, *scvr* to streamline data conversion from the most widely-adopted single-cell analysis tools and a database of pre-analyzed datasets to which users can contribute.

KEY WORDS

Single-cell, scRNA-seq, virtual reality, VR, scATAC-seq, data visualization, clustering, trajectory inference

BACKGROUND

Characterization of cell type, while once dominated by pathological description, has over the past decade shifted towards a more quantitative and molecular approach. As such, molecular measurements in single cells have emerged as a centerpiece in the current paradigm of mechanistic biological investigation [1]. While advancements have been made such that all aspects of the central dogma of molecular biology are now accessed in single cells [2], single-cell RNA sequencing (scRNA-seq), a technique that samples the total mRNA of each cell, and single-cell Assay for Transposase Accessible Chromatin using sequencing (scATAC-seq), a technique that assesses genome-wide chromatin accessibility, are the most well-established and widely-used of these methods [2, 3]. Additionally, assays to profile DNA methylation[5] or protein levels are now maturing and becoming more widely-accessible [5, 6].

In scRNA-seq, the normalized count of each mRNA transcript acts as a dimension or feature by which cells may be characterized. Transcripts for >20,000 coded features are recorded and counted; the most informative features can distinguish one cell from another. In scATAC-seq, the feature space is even larger; cells are characterized by the genomic coordinates of the accessible regions or the features (e.g. transcription factor motifs, *k*-mer frequencies, etc.) derived from these regions. Initially performed in dozens to hundreds of cells, these experiments are now performed on the order of millions of cells. With high dimensionality as a result of thousands of features being considered for each cell and large (in cell number) experiments, analysis methods for this data has been required to advance concurrently with the development of these technologies [8, 9].

Among others, PCA, tSNE, and UMAP are dimensional reduction methods that have become common choices for enabling the visualization of high-dimensional single-cell datasets. Dimensionally reduced datasets are plotted such that cells, which are most similar cluster together and those that are more transcriptionally (and hopefully,

phenotypically) distant are likewise clustered apart. In addition to the visualization of cells, some trajectory inference methods can also learn a latent topology structure to reconstruct the putative time-ordering by which cells may progress along a dynamic biological process [10]. As single-cell technologies have advanced, techniques to cluster and organize cells transcriptionally (or based on other assays) have advanced alongside them to make key insights. However, representation of these dimensionally-reduced visualizations in press is limited to two or three dimensions. Even in three-dimensional plots from published studies, one cannot dynamically adjust or rotate the visualization to better understand the data from another angle. In addition, cells are typically annotated by features e.g. time points, cell type or clusters to investigate stratification along an axis of some biological process. To change the annotations presented in publication, one must often reprocess the raw data, which is time- and skill-intensive, highlighting the need for more dynamical visualization tools. Despite these limited, static representations, single-cell omic datasets are often information-rich and, in many cases, important biological heterogeneity is unable to be visualized if it is outside the scope of a given publication, without spending considerable cost and time to reanalyze the dataset from scratch. This limitation could be overcome with the ability to rapidly explore a preprocessed dataset of choice and toggle multiple cell annotations. Per the limitations mentioned above, many biologists who would like to query the data and generate hypotheses are unable to do so.

While other virtual reality (VR) visualization methods for single cell transcriptomic data have been recently proposed, they require either expensive hardware or very specific data inputs. Further, they have only demonstrated utility in the transcriptomic domain. Thus, a tool is required to enable researchers, especially those who are not able to efficiently reprocess the raw data, to explore the richness of published datasets (or their own unpublished data) through a simple, easy and affordable VR platform. Importantly, this platform must be flexible enough to accept all types of omics data from established processing tools currently employed by the single cell community.

At the time of this writing, two non-peer-reviewed methods that employ VR technology to overcome the limitations of two- and three-dimensional visualizations of scRNA-seq data have recently been developed. *CellexalVR* enables the visualization of standard scRNA-seq data though requires the user to preprocess their data through scripting, which requires intermediate to advanced programming skills [11]. Unfortunately, this tool requires expensive and dedicated VR hardware to operate. An alternative choice to *CellexalVR* is *starmap* [12], which enables the visualization of scRNA-seq data through inexpensive cardboard visor hardware, however lacks the advanced portability of outputs from commonly-used scRNA-seq analysis tools and cell annotation is limited to clustering results. Of note, there are currently no peer-reviewed tools available for the visualization of single-cell data in VR illustrating the novelty in this area of research. Additionally, there are no tools available that that support assays outside of scRNA-seq. To overcome their limitations as well as build on their qualities and initial progress, we here present *singlecellVR*, an interactive web application, which implements an innovative visualization for single-cell data built on VR technology. *singlecellVR* supports clustering, trajectory inference and abstract graph analysis for transcriptomic as well as epigenomic single cell data. *singlecellVR* is a browser-contained, free, and open-access tool. Importantly, we

have developed a one-command conversion tool called *scvr* to make the results of commonly-used single-cell analysis tools compatible with *singlecellVR*.

RESULTS

singlecellVR user experience and overview

singlecellVR is an easy-to-use web platform and database that can be operated from inexpensive, cardboard visor hardware (~$8, available online from popular vendors including Google and Amazon). The webpage, available at http://www.singlecellvr.com enables users to explore several pre-loaded datasets or upload their own datasets for VR visualization. Visualization can be done either on a personal computer or smartphone. To facilitate the transition between the personal computer browser view and the phone-enabled VR visor (VR mode), we have implemented an easy way to transition between these two visualizations as described in the next sections. In VR mode an interactive visualization is presented to the user, allowing them to manipulate and visualize single-cell data and different annotations through the cardboard visor. Additionally, *singlecellVR* features the ability to receive as inputs, the standard output files of commonly-used tools for standard single-cell analysis: *Seurat*, *Scanpy (along with EpiScanpy)*, *STREAM*, and *PAGA*. A companion package enables the conversion of these standard outputs to VR-compatible objects in a single command.

Scanpy [13] and *Seurat* [14] are the most widely-used tools for performing routine single-cell analyses including steps for preprocessing, clustering, trajectory inference, differential expression, and ultimately, visualization. Both of these packages have enabled even novice computational biologists to process single-cell data. Importantly, functions to convert data objects that result from these packages have enabled easy data sharing. Here, we have enabled compatibility with both *Scanpy (AnnData)* and *Seurat* objects. In addition to these two mainstream workflows, we have also implemented compatibility with two trajectory inference tools, *PAGA* [9] and *STREAM* [10]. *PAGA* (partition-based graph abstraction) is a method that preserves topological features of scRNA-seq data in a dimensionally-reduced space through a graph-like map of a manifold fit to the data – this allows for variable resolution data structure inference [15]. *STREAM* (*single cell trajectories reconstruction, exploration and mapping)* developed by our group is a software package for the analysis of complex branching trajectories in single-cell omics data, supporting both scRNA-seq an scATAC-seq data [16].

As shown in **Figure 1**, to use *singlecellVR*, the user needs to select a precomputed dataset or convert their data from these pipelines (i.e. *Scanpy*, *EpiScanpy, Seurat*, *PAGA*, or *STREAM*). This can be easily accomplished by using *scvr*, a simple one-line command we provide that performs data conversion and produces a VR-compatible file for direct visualization with *singlecellVR*. To highlight the generalizability of *scvr* across data types, we have processed and visualized scATAC-seq data in addition to several scRNA-seq datasets. Taken together we believe, *singlecellVR* addresses the key limitations of previously-developed tools mentioned above and we compare them in detail below [4, 5].

VR Database and scvr preprocessing tool. *singlecellVR* provides a growing database of several datasets processed for VR visualization. This was possible thanks to the streamlined *scvr* utility. Importantly, this tool has been made available as a pip package and can be installed and run in just two commands, total (one for installation and one for running the conversion). To showcase this functionality, we have preprocessed with *scvr*, a collection of eight published datasets, which includes both scRNA-seq as well as scATAC-seq that are compatible with our VR engine and made them available in the user interface for immediate visualization. The file produced by the *scvr* companion package is formatted as a .json file and contains the 3-D coordinates of cells in a specified space (e.g. UMAP, LLE, etc.), cell annotations (e.g. FACS-sorting labels, clustering solutions, pseudotime, etc.), and feature quantification (gene expression levels, transcription factor deviation, etc.). It also contains the graph structure (the coordinates of nodes and edges) for trajectory inference methods. Excitingly, given the small footprint of these files, we are offering users the ability to submit their processed data to our repository (see **Supplementary Note 1**) to make our tool a general resource for the field. In this way, we hope to even further extend the ability of biologists to visualize once static datasets and easily generate new hypotheses through manipulation of a large number of rich datasets. Therefore we envision that our website will function as a repository for VR visualization data and information. In addition, we have made a Github repository to document how to generate and view datasets in VR with several step by step tutorials.

A simple, cloud-based web tool for VR visualization. To build *singlecellVR* we have adopted the web technology, *Dash* by *Plotly* and a recently-developed javascript framework for VR/AR, *A-FRAME*. This allowed us to create a tool that is portable and does not require any installation. The input to our visualization engine is a simple .json file. As discussed above, conversion from the standard output of any single-cell analysis tool to this format would normally pose a significant methodological roadblock to most users, especially non-computational biologists. To bridge this gap, we use the aforementioned companion package, *scvr* to parse and convert the outputs of the common single-cell data workflows, *Scanpy*, *EpiScanpy*, *Seurat*, *PAGA*, and *STREAM* and create an *A-FRAME*-compatible .json file (Supplementary Note 2). These workflows produce .h5ad, .loom, and .pkl files; a hypothetical expansion of this tool may be found in the ability to convert outputs from other tools that use these file formats as their output. We have provided a tutorial in an accompanying Github repository and filmed a short video tutorial found on the homepage of *singlecellVR* to assist users in preparing their *singlecellVR* data visualizations.

We predict that in most cases, users will prefer to upload their data through a computer. The website, http://www.singlecellvr.com can be reached through any web browser. Browser compatibility was tested against Google Chrome, Apple Safari, and Mozilla Firefox. All browsers demonstrated stable functionality with this VR tool on the array of web browsers for both Android and Apple smartphones.

Once the user has uploaded their data to the VR tool, they have the option to view and explore the data in 3-D directly in their web browser or to quickly jettison the data to their mobile device for visualization in a VR headset (**Figure 2**). A key challenge associated

with developing a method for visualization of single-cell data is transporting data that is typically processed on desktop setting to the smartphone-based VR visualization. To overcome this challenge and enable a seamless transition to a smartphone for VR view, our software dynamically generate a QR code that enables users to open the VR view on their phone of the website-uploaded data. This phone-based approach is particularly useful as most users are not processing single-cell data analysis from a phone nor would they keep the files on a mobile device.

Supported tools and analysis. As previously mentioned, *Scanpy* and *Seurat* are two commonly-used tools for performing cell clustering as well as differential expression analysis. Here we demonstrate the utility of *singlecellVR* to visualize the common outputs of these tools, showcasing both the clustering solutions as well as differentially expressed genes or features that are visualized easily through the VR interface (**Figure 3**). A key advantage of our tool is the ability to supply multiple annotations to cells to visualize various attributes of the measured data, for example based on the biological query of interest or experimental design. This may include stratification by cluster identity, time points, tissues, or FACS-based labels. In **Figure 3**, we demonstrate the ability to select visualizations by various cluster identifications, which are user-customizable. With the advent of cross-experiment integration methods that can integrate not only multiple scRNA-seq experiments but experiments across modalities of single-cell data collection, this flexible labelling strategy should enable the user in the future to visualize even the most novel and complicated experiments in rich detail.

In addition to flexibility for visualizing complex experimental setups, *singlecellVR* is able to process large experiments. To demonstrate this utility, we processed (using *Scanpy* and *scvr*) and visualized on *singlecellVR*, the scRNA-seq data from the Chan-Zuckerberg Biohub *Tabula Muris* project, a dataset consisting of 44,949 cells and 20 tissues from seven mice [17]. In **Figure 3A**, clustering analyses of this dataset are projected into VR, colored by mouse tissue (**left**) and Louvain cluster identity (**right**).

Single-cell measurements are also particularly useful for capturing cross-section snapshots of a biological process. With dense cell sampling, one can often observe transient cell states that exist between two, more stable states. However, without an intrinsic understanding of the process being studied, it may be difficult to order these cells along a time axis of a biological process. To enable ordering cells by transcriptional state, pseudotemporal ordering has become a goal of the single-cell field using trajectory inference machine learning algorithms. Trajectory inference, like clustering, describes a high-dimensional biological process and being limited to a two/three-dimensional representation with a limited selection of visualized genes or pathways in a static paper is not ideal. Thus, we intend for our tool to leverage the richness of these datasets and make their general usefulness to the field more widespread. We therefore wanted to extend our VR visualization to the results of common trajectory inference tools (**Figure 3B**). *singlecellVR* supports two trajectory inference tools: *PAGA* is a partition-based graph abstraction trajectory inference method, while *STREAM* is a method to tease apart developmental trajectories and visualize the relative densities of cell populations along a developmental timeline.

To showcase the ability of *singlecellVR* to visualize trajectory inference results, we reprocessed a popular myeloid and erythroid differentiation dataset from *Paul, et al.*, 2015 [18], performing trajectory inference using *PAGA*. In the depiction of the *PAGA*-generated trajectory we are able to observe the nodes (gray) indicating each relative cell populations. Importantly, we are also able to explore the trajectory created by *PAGA* (indicated by the black lines between nodes) – being able to explore this trajectory is a key benefit of visualizing *PAGA*-analyzed datasets in VR as *PAGA* is designed specifically to preserve relative cell topology in constructing the trajectory along a pseudotime axis. Here we show the *PAGA* visualization colored by both clusters/nodes (**Figure 3B, top-left**) as well as by the relative gene expression of *klf1* (**Figure 3B, top-right**). Below the *PAGA* visualization in **Figure 3B**, we show *STREAM* trajectory plots of data that shows the developmental trajectories that occur in mouse blood [19]. These plots are colored by both cell identity (**Figure 3B, bottom-left**) as well as differential expression of *Gata1* (**Figure 3B, bottom-right**). In STREAM, a set of smooth curves, termed *principal graph*, are fit into a single-cell population. Each curve represents a developmental branch and the label (e.g.S0_S1) attached to it indicates the branch identity; within *singlecellVR*, we are able to easily explore this axis and observe qualitatively, the distribution of cells along each trajectory in the UMAP space. The branches of these trajectories are represented by the curves that cut through the cells.

Finally, we demonstrate the ability of *singlecellVR* (and *scvr*) to process and visualize epigenomic data. First, we used the *EpiScanpy* workflow to cluster the 10,000 cell PBMC (healthy donor) scATAC-seq dataset from 10x Genomics (**Figure 3C, left**). We then employed *singlecellVR* to visualize a *STREAM*-analyzed trajectory inference result from a scATAC-seq dataset, which captures mouse hematopoietic development (**Figure 3C, right**) [20]. Taken together, visualization of both the *EpiScanpy* clustering results of the 10x Genomics dataset and the *STREAM* trajectory inference analysis, *singlecellVR* proves to be a robust, generalizable tool for multiple modalities of single-cell analysis.

Comparison of singlecellVR to existing methods. As mentioned above, there are two previous reports of VR tools created to visualize single-cell data: *CellexalVR* [11] and *starmap* [12]. Notably, each of these tools offer only solutions for visualizing scRNA-seq clustering results in VR, leaving single-cell epigenomic analysis and trajectory inference unsolved. *CellexalVR* proposes a versatile, user-friendly visualization that can make use of standard scRNA-seq workflow outputs. A major drawback to *CellexalVR* is that it requires specific and expensive hardware (HTC Vive or HTC Vive Pro; > US$500-800). They also recommend a machine with an *Intel i7 processor, NVIDIA GTX1080, 16 GB ram, 1 TB solid state HD*. These are computational equipment that most biologists will not have at their disposal within their lab, likely limiting use of this tool to more computationally-focused labs. Additionally, software to pre-process the data in preparation for VR visualization is required and therefore requires of the user, an ability to perform scripting. In contrast, *singlecellVR* is made such that one can view their data quickly and easily with just a smartphone and cheap cardboard visor (~$8). With no advanced hardware required, *singlecellVR* is more accessible to a non-computational biologist or novice to the field, enabling ease of use.

In addition to building on the contributions of *CellexalVR*, *singlecellVR* is designed to address aspects of single-cell biology that have not been explored by *Starmap*, the only other available tool that we are aware of at the time of this writing. *Starmap* takes as input comma-separated values containing information of the three-dimensional coordinates of cells in the visualization as well as annotations (i.e., cluster ID), and up to 12 features per cell. This file must be prepared entirely by the user without assistance from the *Starmap* platform, limiting the audience of this tool to experienced computational biologists.

Notably, while both *CellexalVR and Starmap* provide a high-quality visualization of cell clustering, visualization of inferred cell trajectory is not incorporated into either tool. *singlecellVR* aims to take the best aspects of both of these tools (flexibility and ease of access) and expand them with important features to abate the current challenges each tool faces, individually. In doing so, we enable users to visualize their own precomputed data directly from the output of commonly-used single-cell RNA-seq analysis tools. Currently supported are *Scanpy*, *EpiScanpy, Seurat*, *PAGA*, and *STREAM*. *singlecellVR* is the only tool of the three discussed that features a QR code to quickly transport the VR data visualization to another device. Finally, *singlecellVR* is the only technology of those reported that has demonstrated utility in visualizing scATAC-seq data.

DISCUSSION

The amount of publicly available scRNA-seq data has exploded in recent years. With new assays to capture chromatin accessibility, DNA methylation and protein levels in single cells, we predict a second wave of dataset generation. Each of these datasets are extremely high-dimensional and thus, rich with latent information about a given biological sample. Ideally, biologists would be able to explore this treasure-trove of data from any angle and make hypotheses regarding their own interests, perhaps rapidly testing *in silico* hypotheses at little to no time cost. Often however, experimental biologists lack the advanced computational skills and/or time required to reprocess and reanalyze raw data from published experiments to gain an understanding of the data from their desired angle of interest. Additionally, biologists who wish to thoroughly explore data prior to publication may rely on a computational specialist who is less connected to the biological problem of interest, introducing a disconnect in hypothesis-driven experimental turnover.

While once primarily reserved for entertainment, VR has found utility in both industrial and academic applications. In this manuscript we present *singlecellVR* a VR-based visualization platform for single cell data and discuss its differences with existing methods. Importantly, we provide a simple mechanism to prepare results from commonly-used single-cell analysis tools for VR visualization with a single command to considerably increase accessibility (see **methods**) With this added utility, we seek to empower non-computational biologists to explore their data and employ rapid hypothesis testing that could not be made from the traditional static representations typical of communication in a scientific report on paper or a computer screen.

We anticipate that VR will become increasingly useful as a research and education tool and that the construction of software libraries will aid such advancements. Our scalable and flexible VR visualization framework is not limited to scRNA-seq and it can be also easily adapted to other single-cell assays. *EpiScanpy* [21], *Seurat* [22], and *STREAM* [16] are able to process epigenomic data, most commonly scATAC data. Of note, as we have demonstrated here in **Figure 3C**, *singlecellVR* is able to visualize such epigenomic data. With the recent advances in spatially-resolved transcriptomics, a new sort of three-dimensional VR will also become especially useful [23]. Finally, technologies that derive the RNA velocity of single cells could hypothetically be visualized in VR; in future work, this framework could be extended to such data [24]. As software to analyze single cells reach their maturity, one could imagine the incorporation of such visualizations into more clinically translatable settings, such as medical devices.

CONCLUSION

singlecellVR is a platform that enables any researcher to easily visualize single-cell data in VR. Our platform is user-friendly, doesn't require advanced technical skills and dedicated hardware. Importantly, we have curated and preprocessed several recent single-cell datasets from key studies, providing the scientific community with an important resource from which they may readily explore and extract biological insight.

Methods

All datasets were processed using *Scanpy* (version 1.5.1), *EpiScanpy* (version 0.1.8, *Seurat* (version 3.1.5), *PAGA* (part of *Scanpy*, version 1.5.1), and *STREAM* (version 1.0), following their documentations. Jupyter notebooks to reproduce data processing are available at https://github.com/pinellolab/singlecellvr. Analyses were performed on a a 2019 MacBook Pro (2.4 GHz Intel Core i9, 16 GB RAM). To build *singlecellVR*, we used *A-FRAME* (version 1.0.0), *Dash* (version 1.13.3*)* by *Plotly*. The preprocessing package, *scvr* generates a series of .json files containing the spatial coordinates representative of cell embeddings in the UMAP or spectral embedding space and information including labels and features (e.g., gene expression, TF motif deviation, etc). These .json files are zipped upon output from *scvr* into a single file that can be easily uploaded to *singlecellVR* for visualization. The documentation for *scvr* is available here: https://github.com/pinellolab/singlecellvr. Video tutorials for learning about and running visualization experiments with *singlecellVR* (and using *scvr* to prepare the data) are available on YouTube, here: https://www.youtube.com/playlist?list=PLXqLNtGqlbeMaAuiBStnBzUNE6a-ULYx8.

Data Availability

The dataset shown in **Figure 3A** is from the Chan Zuckerberg Tabula Muris project and was downloaded here: https://figshare.com/projects/Tabula_Muris_Transcriptomic_characterization_of_20_organs_and_tissues_from_Mus_musculus_at_single_cell_resolution/27733. The dataset shown in **Figure 3B** (**top**) is from *Paul, et al.* 2015 [18] and was downloaded from

GSE72859. The dataset shown in **Figure 3B** (**bottom**) is from *Nestorowa, et al.* 2016 [19]. and was downloaded from GSE81682. The scATAC dataset shown in the clustering result of **Figure 3C** (**left**) is the 10x PBMC (healthy donor) generated by 10x Genomics and was downloaded here: https://support.10xgenomics.com/single-cell-atac/datasets/1.2.0/atac_pbmc_10k_v1. The scATAC dataset shown in the trajectory inference result in **Figure 3C** (**right**) is from *Buenrostro, et al.* 2018 [20] was downloaded from GSE96769.

The source code and the supporting data for this study are available online on GitHub at https://github.com/pinellolab/singlecellvr. The preprocessing package, scvr is included within that repository and further instruction for its use as well as the source code can be found at https://pypi.org/project/scvr/. The data in this manuscript can be reproduced using the Jupyter notebooks available at https://github.com/pinellolab/singlecellvr.

Ethics approval and consent to participate

Ethics approval was not needed for the study.

Competing interests

The authors declare that they have no competing interests.

Funding

This project has been made possible in part by grant number 2018-182734 to L.P. from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation. L.P. is also partially supported by the National Human Genome Research Institute (NHGRI) Career Development Award (R00HG008399) and Genomic Innovator Award (R35HG010717).

Contributions

DFS, HC, MEV, and LP conceived of this project and designed the experiment, which was begun at the 2019 *HackSeq*, at the University of British Columbia where input from collaborators mentioned in the acknowledgements was received. DFS, HC, and MEV processed the data hosted in the database. DFS led the development of the virtual reality framework. HC led the development of *Dash*-based website and the preprocessing module, scvr. MEV led the preparation of the manuscript. All authors performed user-testing of the software. LP supervised the development of this work and provided guidance. All authors wrote and approved the final manuscript.

Acknowledgements

We would like to acknowledge the organizers and participants of *HackSeq19 (University of British Columbia)*, where this project began. We would like to especially acknowledge those *HackSeq19* participants that contributed to the development of this project: Michelle Crown, Alexander Dungate, David Lin , Terry Lin, and Sepand Dyanatkar Motaghed. Additionally, we would like to thank Ashley Browne, Rebecca Combs, Salma Ibrahim, and Kate McCurley for their contributions to the construction of the virtual reality database as part of the Winsor Science Internship Program.

Author information

Author notes

1. David F. Stein, Huidong Chen, and Michael E. Vinyard contributed equally to this work.

Affiliations

1. **Graduate School of Biomedical Sciences, Icahn School Of Medicine at Mount Sinai, New York, NY, 10029, USA**
   David F. Stein

2. **Khoury College of Computer Science, Northeastern University, Boston, MA 02115, USA**
   David F. Stein

3. **Molecular Pathology Unit, Massachusetts General Hospital Research Institute, Charlestown, MA, 02129, USA**
   Huidong Chen, Michael E. Vinyard & Luca Pinello

4. **Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, 02129, USA**
   Huidong Chen, Michael E. Vinyard & Luca Pinello

5. **Department of Pathology, Harvard Medical School, Boston, MA, 02115, USA**
   Huidong Chen, Michael E. Vinyard, & Luca Pinello

6. **Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA**
   Huidong Chen, Michael E. Vinyard, & Luca Pinello

7. **Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02142, USA**
   Michael E. Vinyard

Corresponding author

Correspondence to Luca Pinello.

References

[1] C. Trapnell, "Defining cell types and states with single-cell genomics," *Genome Research*. 2015, doi: 10.1101/gr.190595.115.

[2] T. Stuart and R. Satija, "Integrative single-cell analysis," *Nature Reviews Genetics*. 2019, doi: 10.1038/s41576-019-0093-7.

[3] D. Lähnemann *et al.*, "Eleven grand challenges in single-cell data science," *Genome Biology*. 2020, doi: 10.1186/s13059-020-1926-6.

[4] J. D. Buenrostro *et al.*, "Single-cell chromatin accessibility reveals principles of regulatory variation," *Nature*, 2015, doi: 10.1038/nature14590.

[5] C. Luo, P. Hajkova, and J. R. Ecker, "Dynamic DNA methylation: In the right place at the right time," *Science*. 2018, doi: 10.1126/science.aat6806.

[6] H. Specht, E. Emmott, T. Koller, and N. Slavov, "High-throughput single-cell proteomics quantifies the emergence of macrophage heterogeneity," *bioRxiv*, 2019, doi: 10.1101/665307.

[7] M. Labib and S. O. Kelley, "Single-cell analysis targeting the proteome," *Nature Reviews Chemistry*. 2020, doi: 10.1038/s41570-020-0162-7.

[8] L. Tian *et al.*, "Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments," *Nat. Methods*, 2019, doi: 10.1038/s41592-019-0425-8.

[9] H. Chen *et al.*, "Assessment of computational methods for the analysis of single-cell ATAC-seq data," *Genome Biol.*, 2019, doi: 10.1186/s13059-019-1854-5.

[10] W. Saelens, R. Cannoodt, H. Todorov, and Y. Saeys, "A comparison of single-cell trajectory inference methods: towards more accurate and robust tools," *bioRxiv*, 2018, doi: 10.1101/276907.

[11] O. Legetth, J. Rodhe, J. Pålsson, M. Wallergård, S. Lang, and S. Soneji, "CellexalVR: A virtual reality platform for the visualisation and analysis of single-cell gene expression data," *bioRxiv*, 2019, doi: 10.1101/329102.

[12] A. Yang, Y. Yao, J. Li, and J. W. Ho, "starmap: Immersive visualisation of single cell data using smartphone-enabled virtual reality," *bioRxiv*, 2018, doi: 10.1101/324855.

[13] F. A. Wolf, P. Angerer, and F. J. Theis, "SCANPY: Large-scale single-cell gene expression data analysis," *Genome Biol.*, 2018, doi: 10.1186/s13059-017-1382-0.

[14] T. Stuart *et al.*, "Comprehensive Integration of Single-Cell Data," *Cell*, 2019, doi: 10.1016/j.cell.2019.05.031.

[15] F. A. Wolf *et al.*, "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells," *Genome Biol.*, 2019, doi: 10.1186/s13059-019-1663-x.

[16] H. Chen *et al.*, "Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM," *Nat. Commun.*, 2019, doi: 10.1038/s41467-019-09670-4.

[17] N. Schaum *et al.*, "Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris," *Nature*, 2018, doi: 10.1038/s41586-018-0590-4.

[18] F. Paul *et al.*, "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors," *Cell*, 2015, doi: 10.1016/j.cell.2015.11.013.

[19] S. Nestorowa *et al.*, "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation," *Blood*, 2016, doi: 10.1182/blood-2016-05-716480.

[20] J. D. Buenrostro *et al.*, “Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation,” *Cell*, 2018, doi: 10.1016/j.cell.2018.03.074.
[21] A. Danese, M. L. Richter, D. S. Fischer, F. J. Theis, and M. Colomé-Tatché, “EpiScanpy: integrated single-cell epigenomic analysis,” *bioRxiv*, 2019, doi: 10.1101/648097.
[22] T. Stuart *et al.*, “Seurat v3,” *Cell*, 2019, doi: 10.1016/j.cell.2019.05.031.
[23] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, “Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity,” *Cell*, 2019, doi: 10.1016/j.cell.2019.05.006.
[24] G. La Manno *et al.*, “RNA velocity of single cells,” *Nature*. 2018, doi: 10.1038/s41586-018-0414-6.

[2] D. Lähnemann *et al.*, “Eleven grand challenges in single-cell data science,” *Genome Biology*. 2020, doi: 10.1186/s13059-020-1926-6.
[3] J. D. Buenrostro *et al.*, “Single-cell chromatin accessibility reveals principles of regulatory variation,” *Nature*, 2015, doi: 10.1038/nature14590.
[4] C. Luo, P. Hajkova, and J. R. Ecker, “Dynamic DNA methylation: In the right place at the right time,” *Science*. 2018, doi: 10.1126/science.aat6806.
[5] H. Specht, E. Emmott, T. Koller, and N. Slavov, “High-throughput single-cell proteomics quantifies the emergence of macrophage heterogeneity,” *bioRxiv*, 2019, doi: 10.1101/665307.
[6] M. Labib and S. O. Kelley, “Single-cell analysis targeting the proteome,” *Nature Reviews Chemistry*. 2020, doi: 10.1038/s41570-020-0162-7.
[7] W. Saelens, R. Cannoodt, H. Todorov, and Y. Saeys, “A comparison of single-cell trajectory inference methods: towards more accurate and robust tools,” *bioRxiv*, 2018, doi: 10.1101/276907.
[8] O. Legetth, J. Rodhe, J. Pålsson, M. Wallergård, S. Lang, and S. Soneji, “CellexalVR: A virtual reality platform for the visualisation and analysis of single-cell gene expression data,” *bioRxiv*, 2019, doi: 10.1101/329102.
[9] A. Yang, Y. Yao, J. Li, and J. W. Ho, “starmap: Immersive visualisation of single cell data using smartphone-enabled virtual reality,” *bioRxiv*, 2018, doi: 10.1101/324855.
[10] A. Haque, J. Engel, S. A. Teichmann, and T. Lönnberg, “A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications,” *Genome Medicine*. 2017, doi: 10.1186/s13073-017-0467-4.
[11] F. A. Wolf, P. Angerer, and F. J. Theis, “SCANPY: Large-scale single-cell gene expression data analysis,” *Genome Biol.*, 2018, doi: 10.1186/s13059-017-1382-0.
[12] T. Stuart *et al.*, “Comprehensive Integration of Single-Cell Data,” *Cell*, 2019, doi: 10.1016/j.cell.2019.05.031.
[13] F. A. Wolf *et al.*, “PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells,” *Genome Biol.*, 2019, doi: 10.1186/s13059-019-1663-x.
[14] H. Chen *et al.*, “Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM,” *Nat. Commun.*, 2019, doi: 10.1038/s41467-019-09670-4.
[15] N. Schaum *et al.*, “Single-cell transcriptomics of 20 mouse organs creates a Tabula

Muris," *Nature*, 2018, doi: 10.1038/s41586-018-0590-4.

[16] F. Paul *et al.*, "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors," *Cell*, 2015, doi: 10.1016/j.cell.2015.11.013.

[17] J. D. Buenrostro *et al.*, "Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation," *Cell*, 2018, doi: 10.1016/j.cell.2018.03.074.

[18] A. Danese, M. L. Richter, D. S. Fischer, F. J. Theis, and M. Colomé-Tatché, "EpiScanpy: integrated single-cell epigenomic analysis," *bioRxiv*, 2019, doi: 10.1101/648097.

[19] T. Stuart *et al.*, "Seurat v3," *Cell*, 2019, doi: 10.1016/j.cell.2019.05.031.

[20] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, "Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity," *Cell*, 2019, doi: 10.1016/j.cell.2019.05.006.

[21] G. La Manno *et al.*, "RNA velocity of single cells," *Nature*. 2018, doi: 10.1038/s41586-018-0414-6.

[22] S. Nestorowa *et al.*, "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation," *Blood*, 2016, doi: 10.1182/blood-2016-05-716480.

Figure legends

Figure 1 | An overview of the *singlecellVR* user experience. Top, grey: The outputs of a standard 2-dimensional scRNA-seq analysis. Middle and bottom, blue: a step-by-step overview of the *singlecellVR* workflow: **1**. Schematic of *flexible* data conversion. One command to install (via pip) and one command to convert the data to be VR-compatible. **2**. Rendering of an interactive webpage for uploading and exploring VR data. **3**. A sample of what the viewer might see in a given visualization as well as a generic version of a common VR smart-phone-adaptive headset.

Figure 2 | Architecture of *singlecellVR*. **A**. Users can quickly pre-process their data from any of the sources listed at the top in a single command. **B**. Once users have selected their data, they can then upload their data (step 1) and scan the QR code with their phone to begin the VR visualization (step 2).

Figure 3 | Rendering the single-cell virtual reality visualization. **A**. Clustering applications. *Scanpy* and *Seurat* offer tools for clustering, which can be visualized using *singlecellVR*. Cells can be visualized and colored by various annotations (shown: mouse tissue type, **left**) or their cluster ID (**right**). The *Scanpy*-analyzed dataset shown here is from *Nestorowa, et al.*, 2016 [19]. **B**. Trajectory inference applications. *PAGA* offers a partition-based graph abstraction, which can be visualized by individual graph nodes (**top-left**) or relative gene expression (**top-right**). The *PAGA*-analyzed dataset shown here is from *Paul, et al.*, 2015 [18]. *STREAM* offers the visualization of developmental trajectories, which can be visualized by individual branch trajectory (**bottom-left**) or by relative gene expression (**bottom-right**). **C**. Epigenomic applications. *EpiScanpy* enables the clustering and visualization of scATAC-seq data (**left**). The *EpiScanpy*-analyzed dataset shown here is from the 10x PBMC (healthy donor) 10,000 cells dataset is colored by Louvain clustering solution. *STREAM* was used to perform trajectory inference on the *Buenrostro, et al.,* 2018 scATAC-seq dataset [20] (**right**).

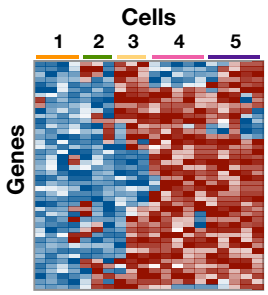
Supplementary Notes

1. **Contributing VR-processed data to the singlecellVR data repository**. Users that wish to contribute to the growing repository of VR datasets my do so by means of a pull request here: https://github.com/pinellolab/singlecellvr. We ask that users add the 'VR Dataset' flag (purple, already added to the sidebar) to their pull request. In addition, in the commit message of the object or comment section of the pull request, please describe your data and the methods by which it was processed, summarizing all labels and genes included.

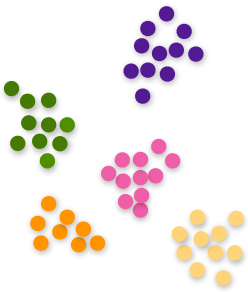
2. **Usage**. Users may navigate the virtual reality visualization via a combination of gaze controls and keyboard inputs. A circle, centered in the user's FOV, indicates the direction that a user will move through the virtual space and also acts as the appendage through which the user will interact with objects in the visualization. Particularly, a user may gaze at the buttons on the menu to select from various clustering annotations and gene expression colorations and to toggle the rotation of the entire visualization. Movement proceeds in the direction of the user's gaze and is controlled with the forward and back arrows on the keyboard. Alternatively, users of VR goggles with a button allowing interaction with a phone screen, may hold the button down to move forward in the direction toward which they are facing. Alphanumeric character input appears in the search field of the menu and may be used to search for available gene expression profiles which will appear in the result fields.The enter key clears the current search value and removes old results from the menu. Additionally, the space bar may be used to summon the menu into the user's FOV. When the user is finished with the menu, they may press the space bar while holding down the control key to return the menu to its starting position. Ocassionally, due to the wide range of data sources supported by singlecellVR, cells may obscure other features of a visualization or may be too small to easily explore. On such occasions, users may hold down the control key and press either the plus or minus key. These combinations will increase and decrease the size of the cells respectivey. Finally, in VR mode on a mobile device, holding the shift key will summon the HUD into view. Subsequently releasing shift will hide the HUD.

Figure 1

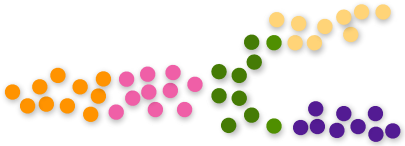
Standard 2-D analyses



Differential Expression



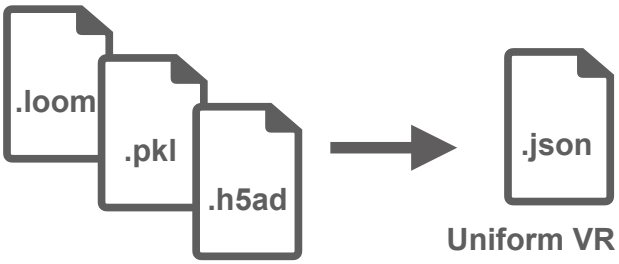
Clustering



Trajectory Inference

1 | Simple data conversion

1-step data command-line data conversion from the standard outputs of PAGA, Seurat, and STREAM to a VR-compatible file.



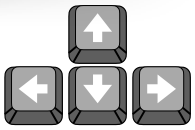
Uniform VR file

```
pip install scvr-prep

scvr_prep -f data.zip -t {PAGA, SEURAT, STREAM} \
                      -a ANNOTATIONS \
                      -g GENES \
                      -o OUTPUT
```

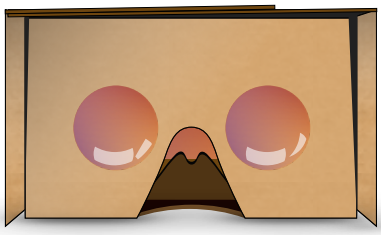
2 | singlecellvr.com

QR code for quick portability to a smart phone



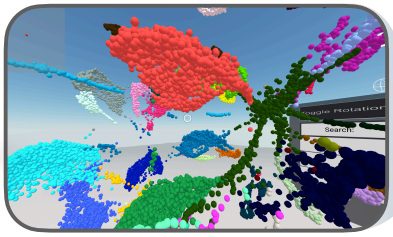
Desktop compatibility

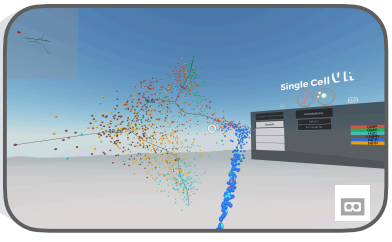
3 | Flexible and affordable UX



~ $15 Viewing lens

Annotated clustering

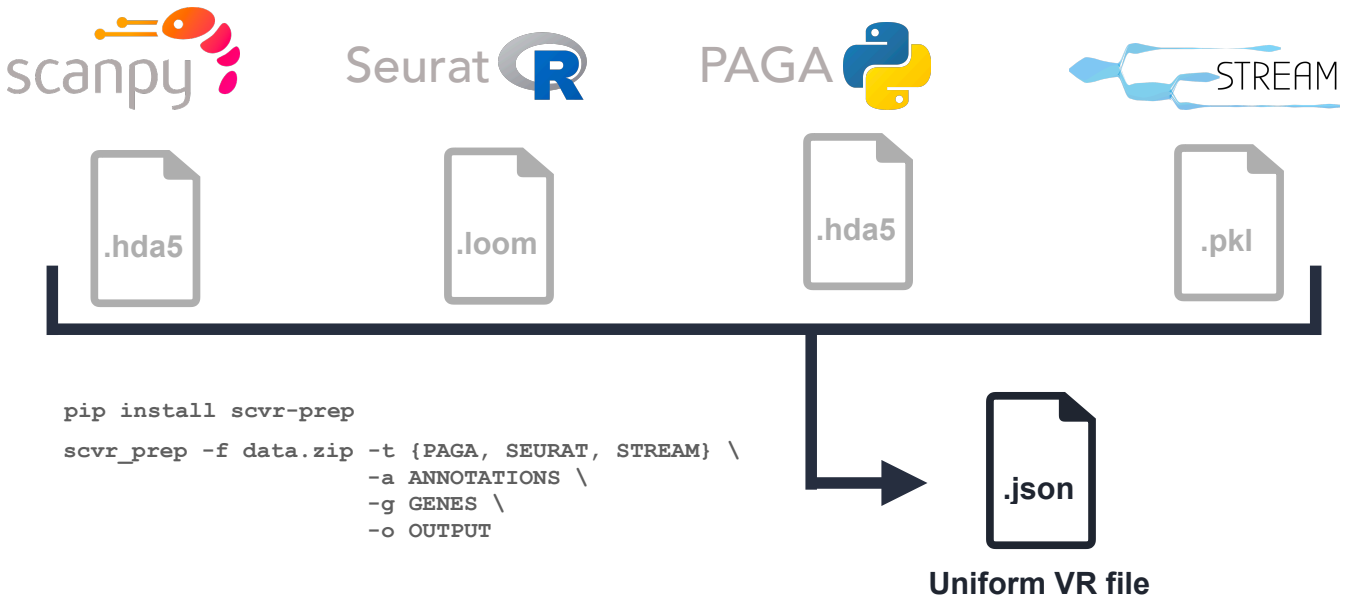




Trajectory analysis

Figure 2

A.



B.

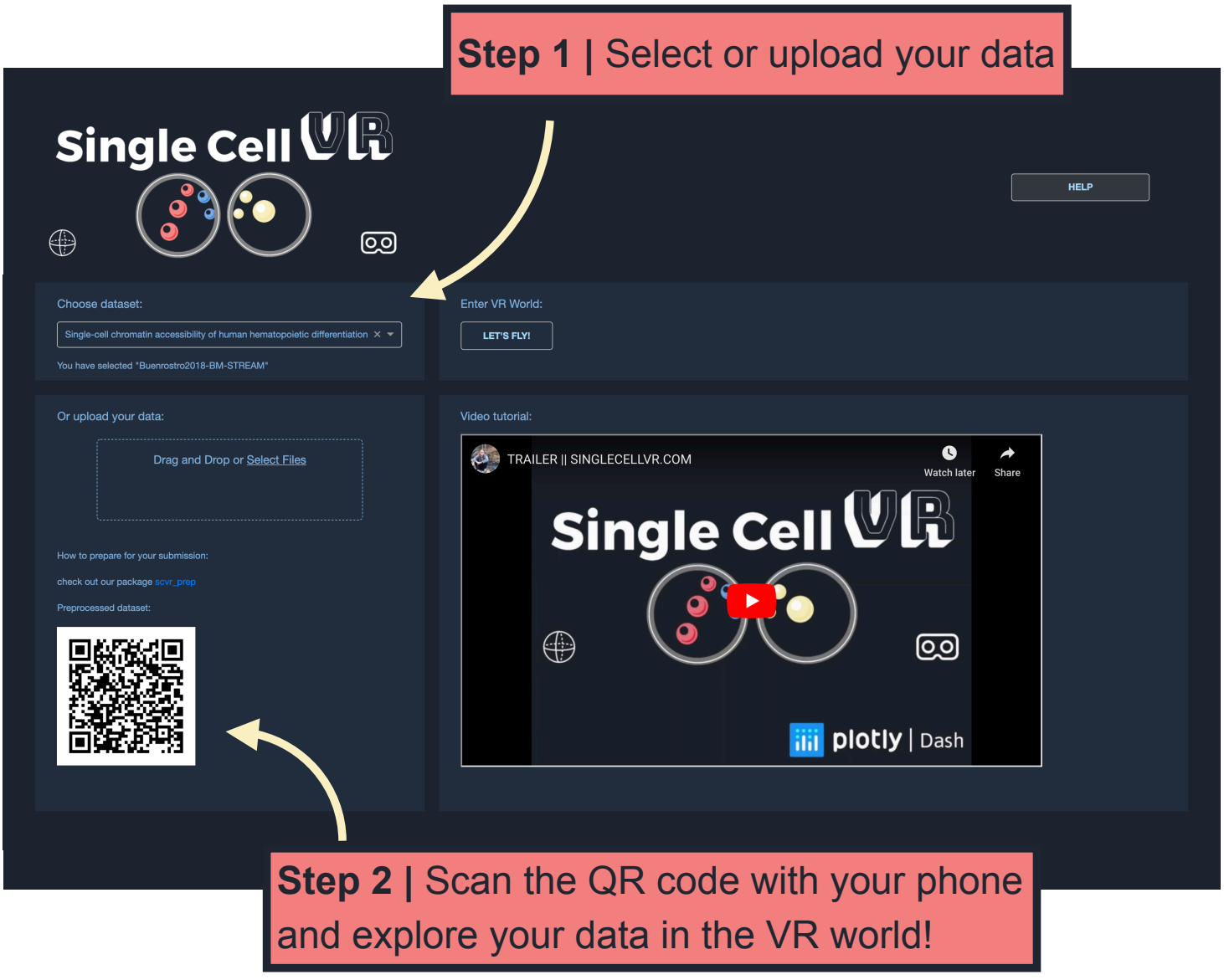
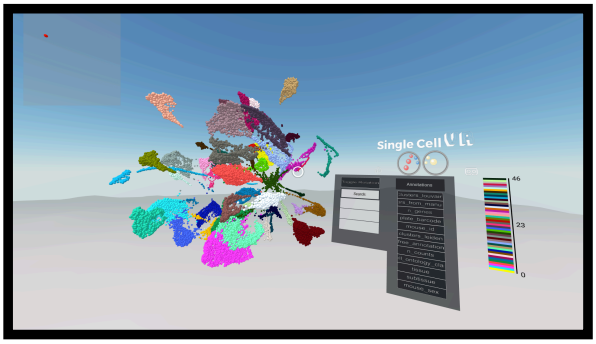


Figure 3

A.

scanpy

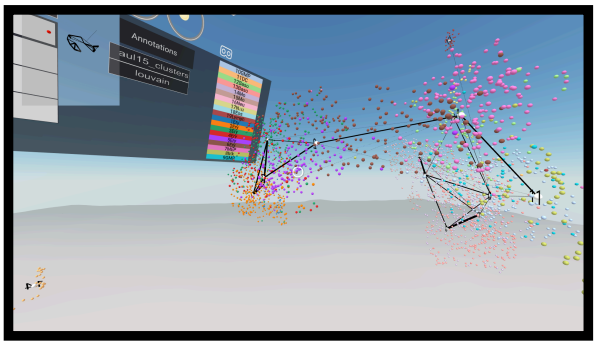
Seurat

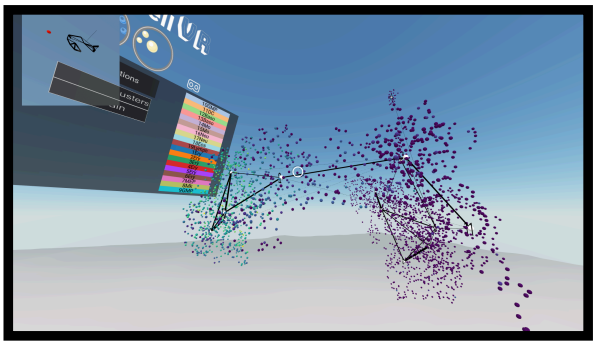


B.

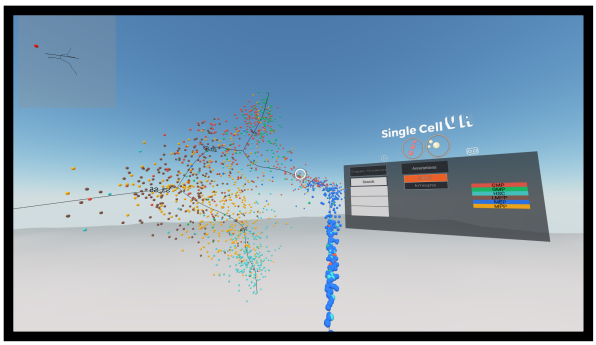
PAGA

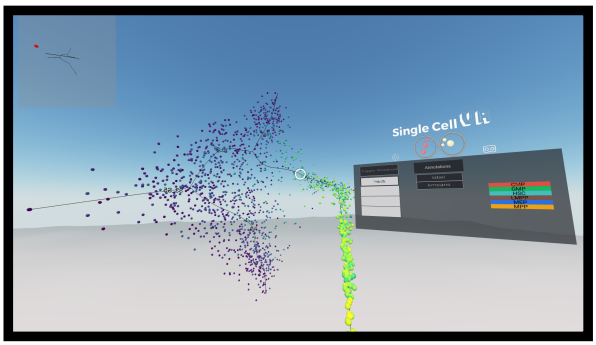
python





STREAM





C.

EpiScanpy

python

STREAM