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 (lpinello@mgh.harvard.edu)

ABSTRACT

Single-cell assays have transformed our ability to model heterogeneity within cell populations and tissues. Virtual Reality (VR) 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 present *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]. Technological advancements have been made such that all aspects of the central dogma of molecular biology can be measured at the single-cell level[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. Thousands of transcripts are recorded and the most informative can be used to 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 (*pseudotime*) 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 using 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. While current representation are often limited and static, single-cell omic datasets are information-rich and, in many cases, important biological heterogeneity cannot be easily investigated and visualized outside the scope of the original publication, without spending considerable cost and time to reanalyze the datasets from scratch.

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 and emerging technologies and 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 also requires expensive and dedicated VR hardware to operate. An alternative to *CellexalVR* is *starmap [12]*, which enables the visualization of scRNA-seq data through inexpensive cardboard visor hardware, but lacks the advanced portability of outputs from commonly-used scRNA-seq analysis tools and limits cell annotation 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 and proteomic 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 using an array of 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 (and AnnData, a data structure from the creators of Scanpy and Episcanpy)* and *Seurat* objects. In addition to the support of these two mainstream workflows, we have also implemented compatibility with two popular 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 and single-cell proteomic data in addition to several scRNA-seq datasets. Taken together we believe, *singlecellVR* addresses the key limitations of previously-developed tools mentioned above, and a formal comparison is presented 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 Python pip package to streamline its installation and can convert a processed dataset for VR visualization with a simple command. To showcase this functionality, we have preprocessed with *scvr*, a collection of 15 published datasets, which includes both scRNA-seq as well as scATAC-seq and sc proteomic data and made them available for immediate VR 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) obtained from supported trajectory inference methods. Excitingly,

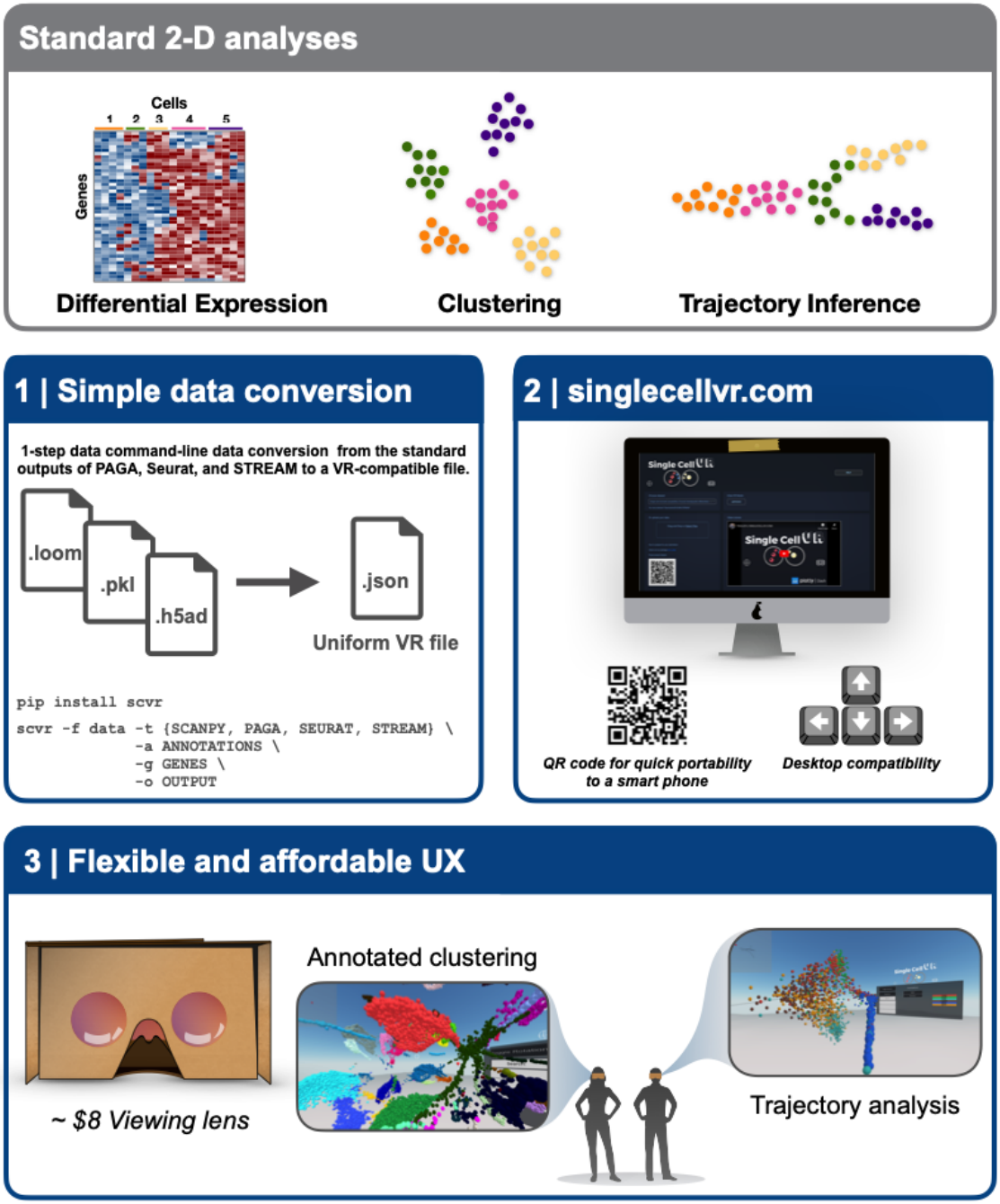

Figure 1 | An overview of the *singlecellVR* user experience. Top, grey: The outputs of a standard 2-dimensional scRNA-seq analysis. Middle and bottom, blue: a step-by-step overview of the *singlecellVR* workflow: **1**. Schematic of *flexible* data conversion. One command to install (via the Python pip package manager) and one command to convert the data to be VR-compatible. **2**. Webpage for uploading and exploring VR data. **3**. VR mode visualization using a cheap smartphone enabled headset.

given the small footprint of these files, we are offering users the ability to submit their processed data to our repository (see **Supplementary Figure 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 of single cell biological annotations. 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 recent web technologies, *Dash* by *Plotly* and *A-FRAME*, a recently-developed JavaScript framework for VR/AR. 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.

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** and **Supplementary 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. In fact we predict that in most cases, users will prefer to upload their data through a computer in which they may have run their analyses. To overcome this challenge and enable a seamless transition to a smartphone for VR view, our software dynamically generates 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 other technical or biological 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.

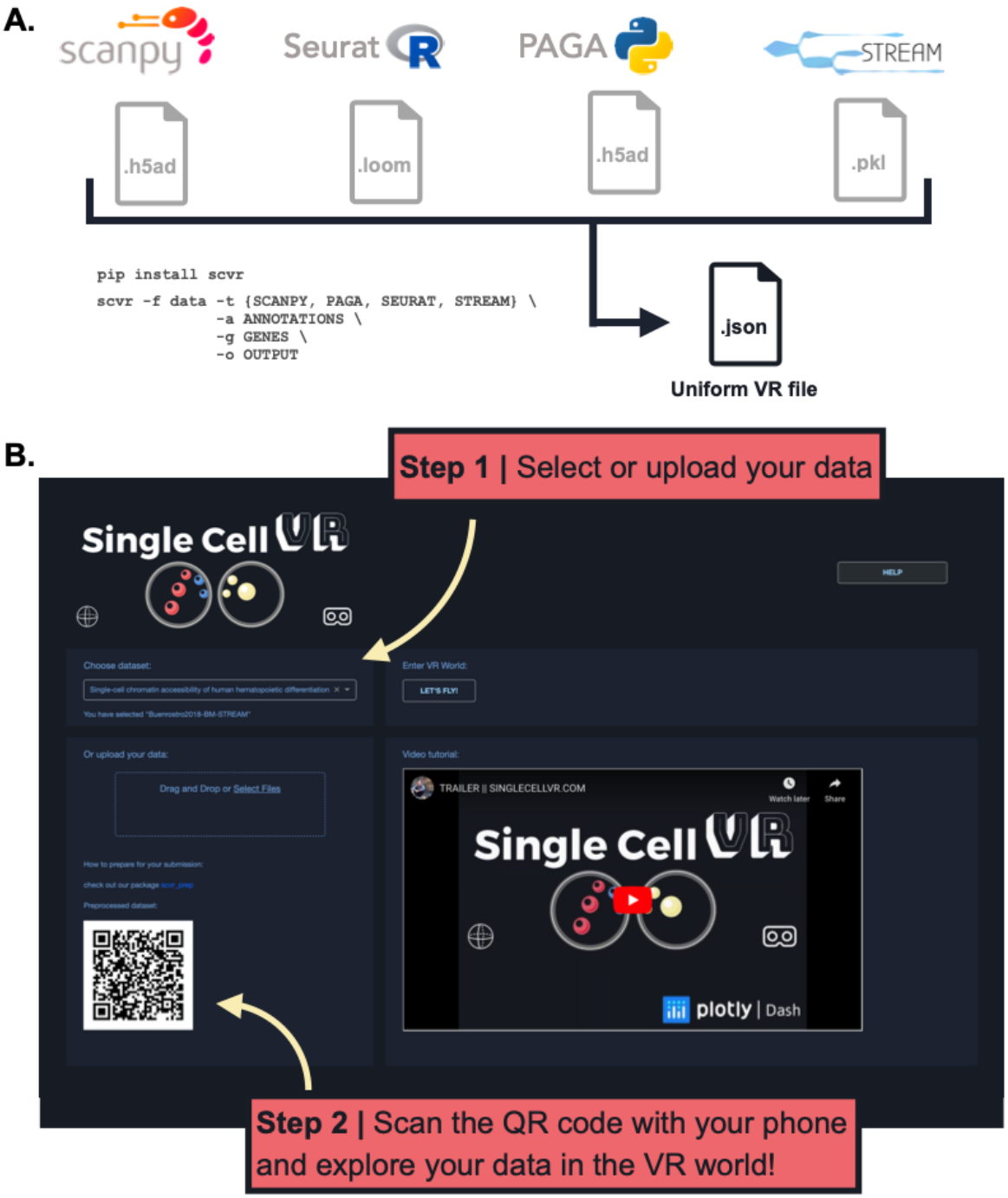

Figure 2 | Architecture of *singlecellVR*. **A**. Users can quickly pre-process their data from any of the sources listed at the top in a single command. **B**. Users can select from pre-processed data or upload their own data (step 1) and scan the dynamically generated QR code with their phone to begin the VR visualization (step 2).

In addition to flexibility for visualizing complex experimental setups, *singlecellVR* is able to visualize large experiments. To demonstrate this utility, we first 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 **Supplementary Figure 3A**, clustering analyses of this dataset are projected into VR, colored by mouse tissue (**left**) and Louvain cluster identity (**right**). With a quick rendering time (<1 second) for the *Tabula Muris* dataset, we next explored the realm of visualization for a modern, large atlas-scale dataset (>1M cells). Using

Scanpy and *scvr*, we successfully processed and visualized on our website, cells from the Allen Brain Institute that capture cortical and hippocampal development inside the mouse brain (**Figure 3A**) [18]. This dataset consists of 1,093,785 cells and is among the largest scRNA-seq datasets created, to date.

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 (or epigenomic) states, pseudotemporal ordering, based on trajectory inference and machine learning algorithms has become a useful technique for the single-cell field. Trajectory inference, like clustering, describes a high-dimensional biological process and being limited to a two/three-dimensional static visualization on paper, with a limited selection of genes or annotations 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*, a partition-based graph abstraction trajectory inference method and *STREAM,* a method based on principal graphs [19] that recovers a tree like structure to summarize developmental trajectories and to visualize the relative densities of cell populations along each branch.

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 [20], performing trajectory inference using *PAGA*. *PAGA* is designed specifically to preserve relative cell topology in constructing the trajectory along a pseudotime axis. In the depiction of the *PAGA*-generated trajectory, nodes (gray) correspond to cell groups, and edges (black lines between nodes) connecting the groups quantify their connectivity and confidence (thickness) (**Figure 3B, top**). To showcase the VR output of *STREAM* we reprocessed a popular mouse blood dataset from *Nestorowa et al*., 2016 [21]. In STREAM, a set of smooth curves, termed *principal graph*, are fitted to the data and each curve represents a developmental branch. Within *singlecellVR*, we are able to easily explore these trajectories in VR and observe qualitatively, the distribution of cells along each branch in the UMAP space (**Figure 3B**, **bottom**). The branches of these trajectories are represented by the curves that cut through the cells.

singlecellVR and *scvr* also support processing and visualizing epigenomic data. To showcase this functionality, we first used the *EpiScanpy* workflow to cluster a scATAC-seq dataset from 10x Genomics containing 10,000 cell PBMC (healthy donor) (**Figure 3C, left**). Next we reprocessed with STREAM a scATAC-seq dataset profiling human hematopoiesis from *Buenrostro et al.,2019* [22] (**Figure 3C, right**). In addition, we extend *singlecellVR* to single-cell quantitative proteomics data, To this end we reprocessed data from *SCoPE2*, a recent assay to quantitate proteins in single cells using mass spectrometry [6]. We performed trajectory inference using *STREAM* on one *SCoPE2* dataset profiling the transition from monocytes to macrophages in the absence of polarizing cytokines. Our analysis revealed a bifurcated branch structure as cells progress towards macrophage phenotypes (**Supplemental Figure 3B**). Importantly, such bifurcation is not readily visualized in previous reports in two dimensions. Finally, we took advantage of the recent advances in the multi-omics field, using *Seurat* to integrate and co-embed PBMC cells profiled by scRNA-seq and scATAC-seq by *10x Genomics* (**Figure 3D**). Taken together, *singlecellVR* proves to be a robust, generalizable tool for multiple modalities of single-cell analysis.

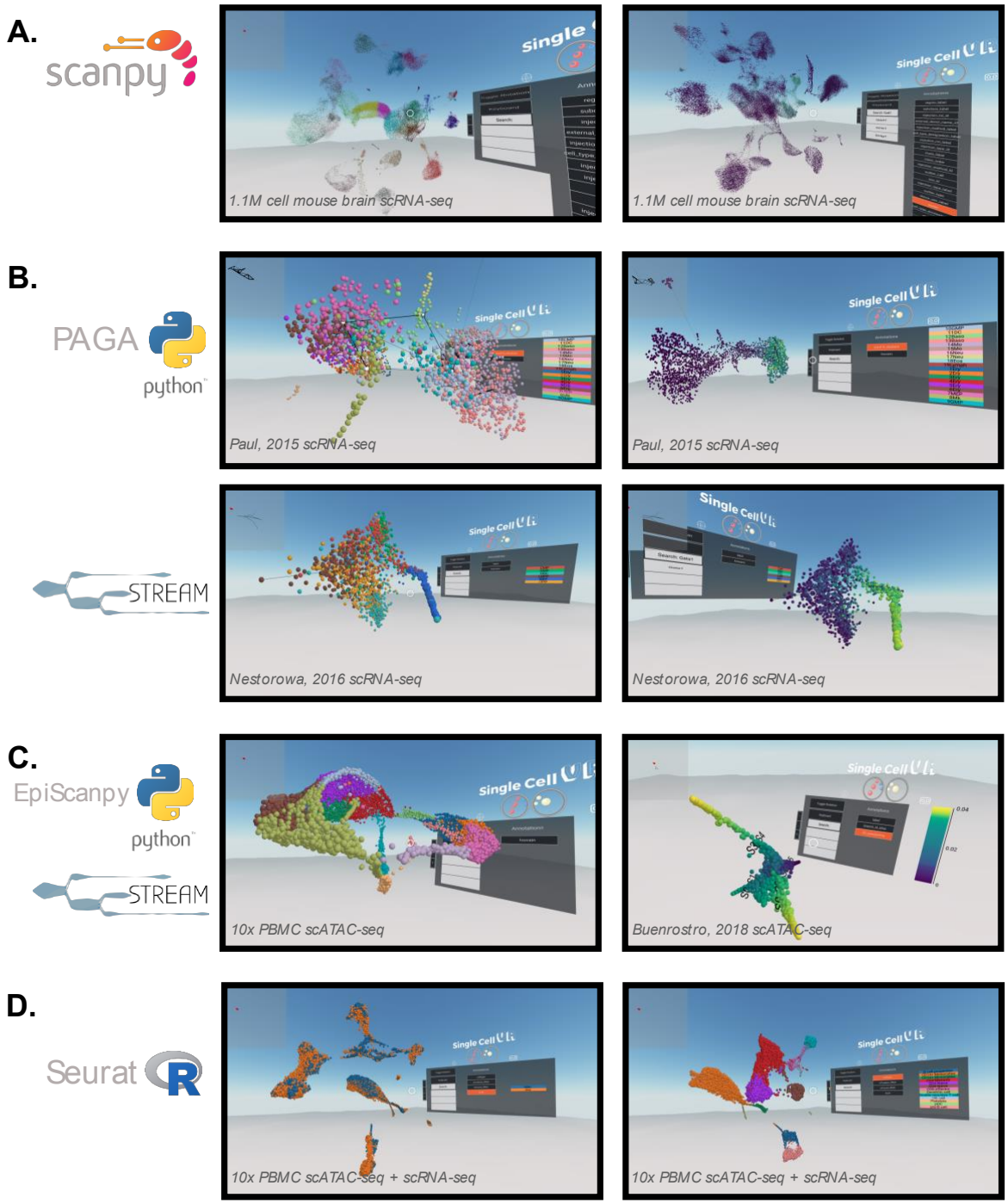

Figure 3 | VR visualization of single-cell processed datasets profiled by different technologies and analyzed by different computational tools of . **A**. Scanpy offers solutions for clustering single-cell data. Shown is a UMAP of the Allen Brain Atlas mouse brain scRNA-seq dataset from *Yao, et al.,* 2020 [18] and processed by Scanpy. Leiden clustering solution (**left**) and expression of *Gad1* (**right**). **B**. Trajectory inference applications. *PAGA* offers a partition-based graph abstraction to uncover potential trajectories (edges) between group of cells (nodes) (**top-left**) relative gene expression (e.g., *Klf1,* ***top-right***), amongst other annotations. The *PAGA*-analyzed dataset shown here is from *Paul, et al.*, 2015 [20]. *STREAM* offers the visualization of developmental trajectories, which can be visualized by cell identity (**bottom-left**) or by relative gene expression (e.g., *Gata1,* ***bottom-right***), amongst other annotations. The STREAM-analyzed dataset shown here is from *Nestorowa, et al.*, 2016 [21]. **C**. Epigenomic applications. *EpiScanpy* enables the clustering and visualization of scATAC-seq data (**left**). PBMC (healthy donor) 10,000 cells dataset analyzed by *EpiScanpy* and with colors corresponding to clustering solutions (Louvain clustering). *STREAM* was used to perform trajectory inference on the *Buenrostro, et al.,* 2018 scATAC-seq dataset [22] (**right**). **D**. Seurat offers solutions for clustering single-cell data as well as integrating datasets across experiments. Shown is a Seurat-integrated scRNA-seq and scATAC-seq PBMC dataset from 10x Genomics, colored by technology (**left**) and cell type (**right**).

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; > $500-800 and 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 single-cell epigenomics and single-cell proteomics data alongside a large database of precomputed single cell datasets.

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 assisted by *in silico* analysis 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 innovations and 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 and tools that already support epigenomic data and/or single-cell proteomic data (*EpiScanpy* [23], *Seurat* [24], and *STREAM* [16]). With the recent advances in spatially-resolved transcriptomics, we believe that a new sort of three-dimensional VR will also become especially useful [25]. 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 [26]. 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 scalable platform that enables any researcher to easily visualize single-cell data in VR. Our platform is user-friendly, requires no advanced technical skills or dedicated hardware. Importantly, we have curated and preprocessed several recent single-cell datasets from key studies across various modalities of data generation, 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 2019 MacBook Pro (2.4 GHz Intel Core i9, 16 GB RAM). To build *singlecellVR*, we used *A-FRAME* (version 1.0.0), *Dash* by *Plotly* (version 1.13.3*)*. The preprocessing package, *scvr* generates a series of .json files containing the spatial coordinates representative of cell embeddings in 3D embedding (e.g. PCA, UMAP, etc.) 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.

Data Availability

The dataset shown in **Figure 3A** is from *Yao, et al*. 2020 [18] and was downloaded and processed from https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x. The dataset shown in **Figure 3B** (**top**) is from *Paul, et al.* 2015 [20] and was downloaded from GSE72859. The dataset shown in **Figure 3B** (**bottom**) is from *Nestorowa, et al.* 2016 [21] 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 from 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 [22] and was downloaded from GSE96769. The scRNA-seq and scATAC-seq data shown in **Figure 3D** was obtained from the *Seurat v3* vignette available here: https://satijalab.org/seurat/v3.2/atacseq_integration_vignette.html. The dataset shown in **Supplementary Figure 3A** is from the Chan Zuckerberg *Tabula Muris* project [17] and was downloaded from: https://figshare.com/projects/Tabula_Muris_Transcriptomic_characterization_of_20_organs_and_tissues_from_Mus_musculus_at_single_cell_resolution/27733. The data shown in **Supplementary Figure 3B** was obtained from https://scope2.slavovlab.net/docs/data.

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 https://pypi.org/project/scvr/. 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.

All the analyses 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 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.

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,”

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

471 [18] Z. Yao, “A taxonomy of transcriptomic cell types across the isocortex and hippocampal
472 formation,” *bioRxiv2*, 2020, doi: https://doi.org/10.1101/2020.03.30.015214.

473 [19] L. Albergante *et al.*, “Robust and scalable learning of complex intrinsic dataset geometry via
474 ElPiGraph,” *Entropy*, 2020, doi: 10.3390/e22030296.

475 [20] F. Paul *et al.*, “Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors,”
476 *Cell*, 2015, doi: 10.1016/j.cell.2015.11.013.

477 [21] S. Nestorowa *et al.*, “A single-cell resolution map of mouse hematopoietic stem and progenitor
478 cell differentiation,” *Blood*, 2016, doi: 10.1182/blood-2016-05-716480.

479 [22] J. D. Buenrostro *et al.*, “Integrated Single-Cell Analysis Maps the Continuous Regulatory
480 Landscape of Human Hematopoietic Differentiation,” *Cell*, 2018, doi:
481 10.1016/j.cell.2018.03.074.

482 [23] A. Danese, M. L. Richter, D. S. Fischer, F. J. Theis, and M. Colomé-Tatché, “EpiScanpy:
483 integrated single-cell epigenomic analysis,” *bioRxiv*, 2019, doi: 10.1101/648097.

484 [24] T. Stuart *et al.*, “Seurat v3,” *Cell*, 2019, doi: 10.1016/j.cell.2019.05.031.

485 [25] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, “Single-
486 Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity,” *Cell*, 2019,
487 doi: 10.1016/j.cell.2019.05.006.

488 [26] G. La Manno *et al.*, “RNA velocity of single cells,” *Nature*. 2018, doi: 10.1038/s41586-018-0414-
489 6.