*Published online 9 June 2023* *NAR Genomics and Bioinformatics, 2023, Vol. 5, No. 2* **1**
*https://doi.org/10.1093/nargab/lqad050*

# GRACE: a comprehensive web-based platform for integrative single-cell transcriptome analysis

**Hao Yu[1,2,3,4], Yuqing Wang[1,2], Xi Zhang[1,2,4,*] and Zheng Wang [1,2,3,4,*]**

[1]Medical Center of Hematology, Second Affiliated Hospital, Army Medical University, Chongqing 400037, China, [2]State Key Laboratory of Trauma, Burn and Combined Injury, Army Medical University, Chongqing 400037, China, [3]Bio-Med Informatics Research Center & Clinical Research Center, The Second Affiliated Hospital, Army Medical University, Chongqing 400037, China and [4]Jinfeng Laboratory, Chongqing 401329, China

Received December 16, 2022; Revised May 03, 2023; Editorial Decision May 08, 2023; Accepted May 18, 2023

## ABSTRACT

**Large-scale single-cell RNA sequencing (scRNA-seq) has emerged as a robust method for dissecting cellular heterogeneity at single-cell resolution. However, to meet the increasingly high computational demands of non-programming experts, a user-friendly, scalable, and accessible online platform for analyzing scRNA-seq data is urgently needed. Here, we have developed a web-based platform GRACE (GRaphical Analyzing Cell Explorer) (http://grace.flowhub.com.cn or http://grace.jflab.ac.cn:28080) that enables online massive single-cell transcriptome analysis, improving interactivity and reproducibility using high-quality visualization frameworks. GRACE provides easy access to interactive visualization, customized parameters, and publication-quality graphs. Furthermore, it comprehensively integrates preprocessing, clustering, developmental trajectory inference, cell-cell communication, cell-type annotation, subcluster analysis, and pathway enrichment. In addition to the website platform, we also provide a Docker version that can be easily deployed on private servers. The source code for GRACE is freely available at (https://github.com/th00516/GRACE). Documentation and video tutorials are accessible from website homepage (http://grace.flowhub.com.cn). GRACE can analyze massive scRNA-seq data more flexibly and be accessible to the scientific community. This platform fulfills the major gap that exists between experimental (wet lab) and bioinformatic (dry lab) research.**

## INTRODUCTION

With the development of single-cell technology, massive high-throughput transcriptome analysis has provided opportunities for groundbreaking discoveries (1,2). Many recently developed pipelines or packages, such as Seurat (3), Monocle (4) and Scanpy (5), integrate multiple functional modules to support bioinformatics programming tasks. However, users still require programming skills (at least R or Python) and a command-line interface to perform customized analysis. Additionally, wet labs may lack the appropriate talents to analyze massive single-cell data and operate powerful high-performance computing clusters. Thus, a major gap exists between experimental and bioinformatics research.

Several GUI-based platforms, such as Galaxy (6) (https://humancellatlas.usegalaxy.eu/), GranatumX (7), cellxgene (8), UCSC Cell Browser (9), ASAP (10), ICARUS (11) and SCP (https://singlecell.broadinstitute.org/single_cell) have been developed to analyze scRNA-seq data online. However, these platforms have not fully incorporated current comprehensive scRNA-seq analysis tools. A lack of batch-effect-removal function limits their scalability and utility in analyzing datasets from multiple samples (12). Without functional modularity, it is not easy to update their toolkits for these platforms. Thus, a robust platform is still required to provide a concise and stable environment for the integrated analysis of multiple datasets.

To address these challenges, we introduce GRACE, a web interface-based platform that supports online analysis, annotation, and visualization of scRNA-seq data. It comprehensively integrates the current standard pipelines in scRNA-seq analysis, empowering experimental researchers without programming skills to analyze and visualize datasets of single-cell transcriptome. GRACE is a GUI and Shiny-based service that offers a user-friendly, scalable and accessible online platform for analyzing scRNA-seq data, fulfilling the gap between experimental and bioinformatics research.

[*]To whom correspondence should be addressed. Tel: +86 023 68763198; Fax: +86 023 68763198; Email: biowz@mail.ustc.edu.cn
Correspondence may also be addressed to Xi Zhang. Email: zhangxxi@sina.com

© The Author(s) 2023. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

## MATERIALS AND METHODS

### Framework & design

GRACE is a code-free, open-source, all-in-one web platform designed to enable non-programming experts to analyze their single-cell transcriptome datasets. It employs Shiny (https://shiny.rstudio.com) to build GUI and Plotly's graphing library (https://plotly.com/r) for data visualization. Each function module follows a similar template to improve development efficiency. The structure of template comprises four main parts: 'Act' (activity) for specific action buttons, 'Ctl' (control) for special operations. 'Plt' (plot) and 'Tab' (table) are used for generating visualized graphs and tables, respectively. GRACE uses Anndata (https://anndata.readthedocs.io) as an internal data format to enhance data processing efficiency. Within internal data structure, 'X' or 'var' is used to store count matrix or each cell's metadata, respectively. GRACE utilizes SQLite3 database engine (https://sqlite.org/index.html) to store meta-information about each dataset, including sample IDs, species, groups, creation times, and file pathways. GRACE implements nextflow (13) package to facilitate online data uploading and internal data preprocessing. Furthermore, GRACE integrates best-practice pipelines and packages to achieve canonical analysis in a flexible and scalable way, including data preprocessing, dimension reduction, cell clustering & annotation, gene-expression analysis, trajectory analysis, cell-cell communication and single-cell gene ontology analysis (Figure 1).

### Best-practice pipelines with high performance

Compared to other similar software, GRACE offers high-performance pipelines that facilitate a seamless transition from upstream to downstream data analysis (Supplementary Table 1). At each step of scRNA-seq analysis, GRACE comprehensively integrated the most canonical software packages or modules currently available without any modification, apart from single-cell pathway enrichment. The specific software packages or modules used are detailed in Supplementary Table 2. GRACE comprehensively integrates analysis methods and packages with unmodified code to achieve the robustness and reproducibility. For example, GRACE integrates SCTransform (14) and DoubletFinder (15) for data standardization and doublet removal. To improve the flexibility of sample integration, there are five optional methods available for batch effect removal, including RPCA in Seurat v4 (3), FastMNN (16), Harmony (17), scVI (18) and svANVI (18). All these softwares are encapsulated into nextflow pipeline for streamlined and efficient data processing. In 'Feature Selection' module, GRACE provides three options: (1) selecting the top N highly variable genes (HVGs), which is implemented by calling Seurat's FindVariableFeatures function; (2) selecting genes based on modeling using the scikit-learn package, which employs linearSVC and ExtraTreesClassifier algorithms; and (3) directly uploading a customized gene list as selected features for downstream analysis. Furthermore, GRACE also provides the 'Feature filtering' module to filter out irrelevant or noisy genes, such as cell-cycle, ribosomal genes, or customized gene set uploaded by users. To infer 'cell-cell communication' analysis, GRACE incorporates CellPhoneDB (19), CellChat (20) and Cellcall (21) to caculate the strength of intercellular interaction and differential pairs of ligand-receptors.

To increase the efficiency of subclustering, GRACE provides an all-in-one pipeline for subclustering analyses, including reconstruction of lineage trajectories, intra-type cell communication and single-cell pathway enrichment. This section integrates several packages, such as Monocle2 (4), SPRING (22), Seurat v4 (3), CellPhoneDB (19), CellChat (20), and Cellcall (21). In 'Developmental trajectory' section, GRACE has integrated Monocle2 (4) and SPRING (22) to enable reverse-graph-embedding or force-directed-graph based trajectory inference. Notably, GRACE develops 'single-cell pathway enrichment' module to investigate the pathways of biological processes and the expression of specific genes across different states along the developmental trajectory, which make users interpret their data comprehensively and depict cellular heterogeneity at the single-cell resolution. GRACE uses all the HVGs in the first N of PCs (N can be set by the users) as input. Gene-set enrichment was first performed to identify classes of input genes and generate the annotated terms. The background value is computed by the number of genes in one cell against that of population total genes. At the single-cell level, the foreground value is computed by the number of genes in each pathway against that out of population total genes. F-score represents the ratio of foreground versus background value. Adjusted F-score was to reduce false positives and minimize false negatives. ClusterProfiler package (23) was incorporated to work on the main steps of gene-set enrichment analysis.

GRACE offers two automated methods for ultra-fast cell-type annotation, in addition to manual annotation using prior knowledge. The first method, which is referred to as 'Reference-based Annotation, utilizes the SingleR package (24) to annotate cells based on transcriptome datasets of pure cell types. It employs the 'Human Primary Cell Atlas Data" and 'Mouse RNAseq Data' references which are built-in data in the Celldex package (24). The second method uses the Weight Nearest Network (WNN) algorithm (3) to map query scRNA-seq data against a well-annotated single-cell transcriptome atlas, which is called 'Atlas-based Annotation'. In this module, GRACE offers four pre-generated atlases that can be accessed through the link (https://azimuth.hubmapconsortium.org). Additionally, it permits users to customize the annotated reference database, in addition to using the built-in references of the SingleR package. Data visualization is achieved using the Plotly package (https://plotly.com/r).

### Dataset demonstration

Classical 'PBMC 3k' dataset (https://cf.10xgenomics.com/samples/cell/pbmc3k/) was converted into hdf5 format via DropletUtils (v1.14.2) (25) and fed into GRACE as one-group dataset (25,26). The other public dataset 'lymphoma' scRNA-seq (27) represents two-group dataset, including 'HC01' (healthy) and 'MA08' (lymphoma). This dataset was saved as CSV formatted file. Our GitHub repository also contains all test data

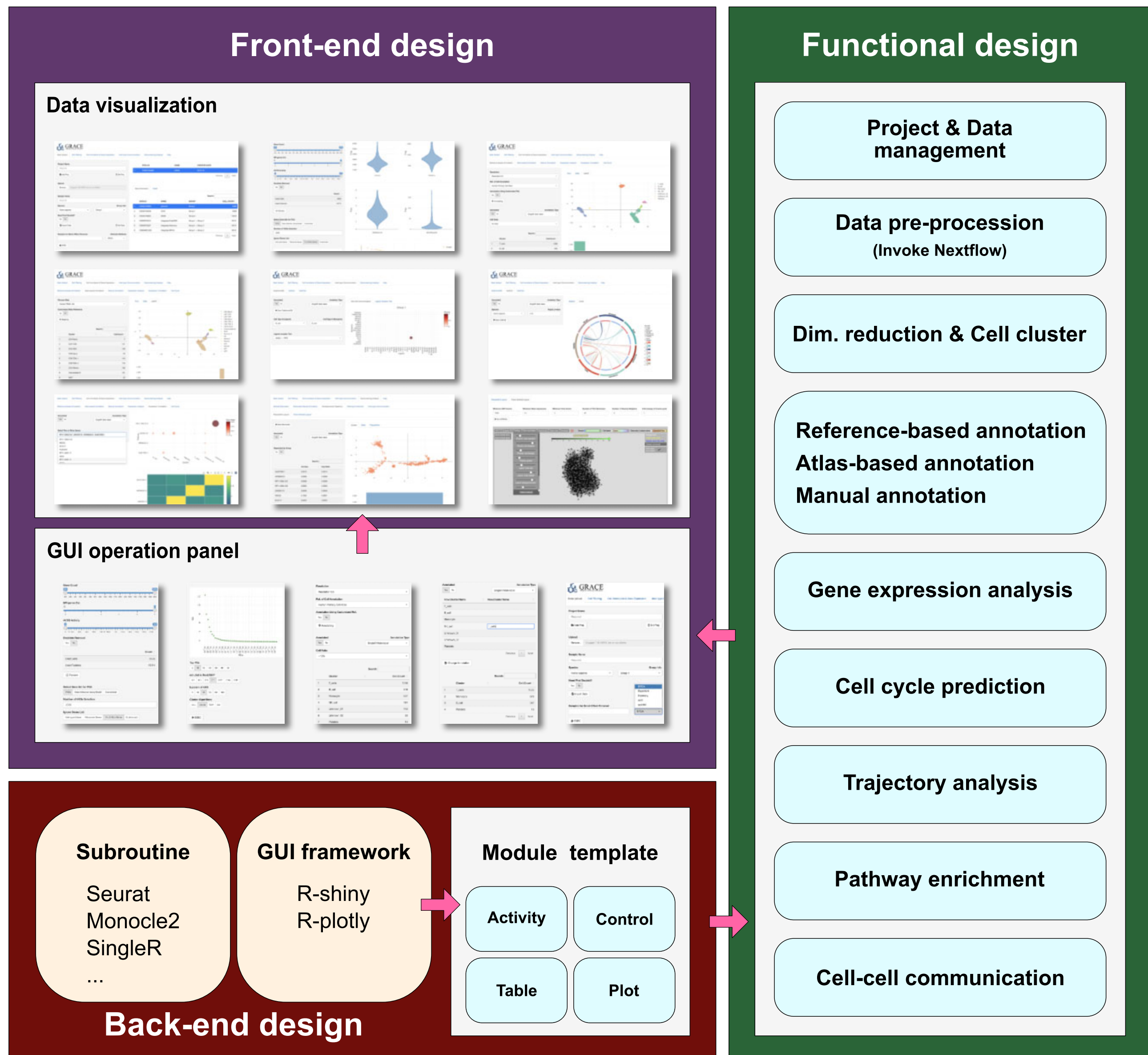


**Figure 1.** Summary of the workflow of GRACE**.** The analysis workflow of GRACE is divided into three parts: back-end development (framework and embedded software), front-end visualization and functional design. The back-end development utilizes the R-shiny and R-plotly packages to build the GUI framework. A unified module template is first built to implement each functional page, and then R packages such as Seurat and Monocle2 are incorporated into each page. The GUI framework is responsible for building the front-end graphical operation interface and data visualization. In terms of functional design, GRACE aims to provide comprehensive functions, including data preprocessing, cell clustering, and trajectory analysis.

(https://github.com/th00516/GRACE/tree/main/testdata). We mapped this query dataset onto an online annotated atlas 'PMBC 10k' (https://support.10xgenomics.com/single-cell-gene-expression/datasets/).

### Availability

All source code and its docker image can be free downloaded (https://github.com/th00516/GRACE)**.** The instructions and online video tutorials are available on the website homepage. (http://grace.flowhub.com.cn), The atlas reference is available on the Dropbox sharing link (https://www.dropbox.com/sh/zvsziyryjunqglm/AABY4sPTvGQPyxI5PlWzqBWXa?dl=0).

## RESULTS

### Case study 1: 'PBMC 3k' as one-group dataset

This case study demonstrates best-practice preprocessing and cell-type annotation methods. Data preprocessing, including removal of less-expressed genes, doublets, and low-quality cells, as well as dimensional reduction, was performed automatically after dataset uploading using the internal implementation of nextflow (13) (Figure S2A). For cell-type annotation, we initially used a reference-based method with the SingleR default reference after unsupervised clustering. However, one unknown cluster remained in addition to annotated cell types (Figure S2B). To resolve this, we employed the 'Atlas-based Annotation' module of

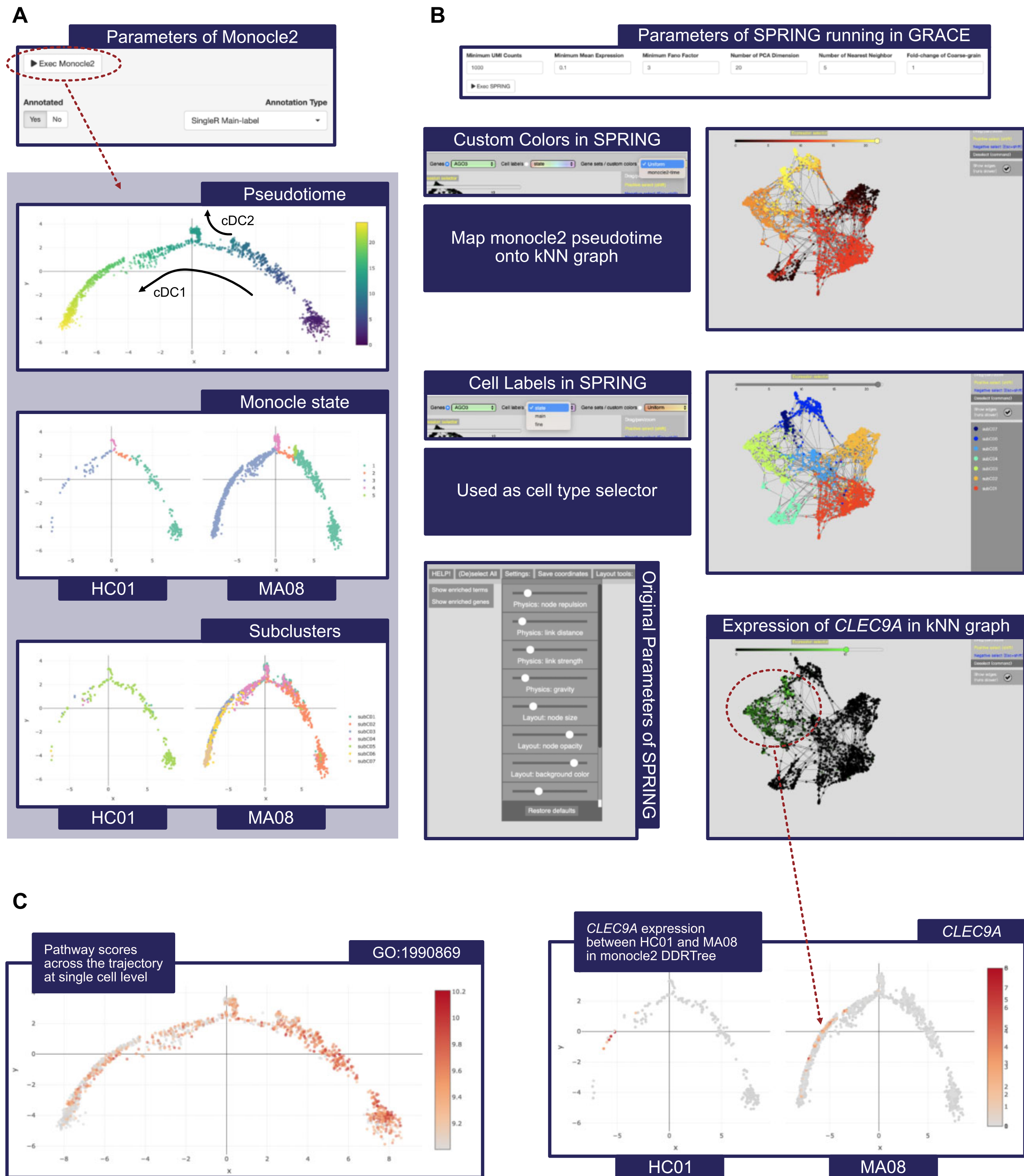


**Figure 2.** The developmental trajectory module in GRACE integrates both Monocle2 and SPRING to provide a comprehensive analysis of single-cell transcriptome data. (**A**) In the monocle2 section, the analysis was performed on the DC cluster of HC01 and MA08 integrated samples. The main results include pseudotime, monocle state, and the mapping of DC subclusters to DDRTree (Subclusters) results. The monocle state and subclusters support two-group comparisons, with HC01 representing group 1 and MA08 representing group 2. (**B**) In the SPRING section, there are six groups of parameters that can be set. GRACE integrated the original parameters of the SPRING Viewer, which can be used to adjust the Force-directed graph parameters. The SPRING Viewer supports displaying gene expression, such as the expression of CLEC9A. The original cell labels are used in GRACE to store cell type/cluster information, while the original custom colors are used in GRACE to store pseudotime information from monocle2. This enables integration and analysis of monocle2 and SPRING results. (**C**) The expression level of CLEC9A mapped onto DDRTree corresponds to the red circle in the kNN graph in (B), which represents subcluster 3 at a resolution of 0.2. This subcluster was annotated as DC:monocyte-derived.

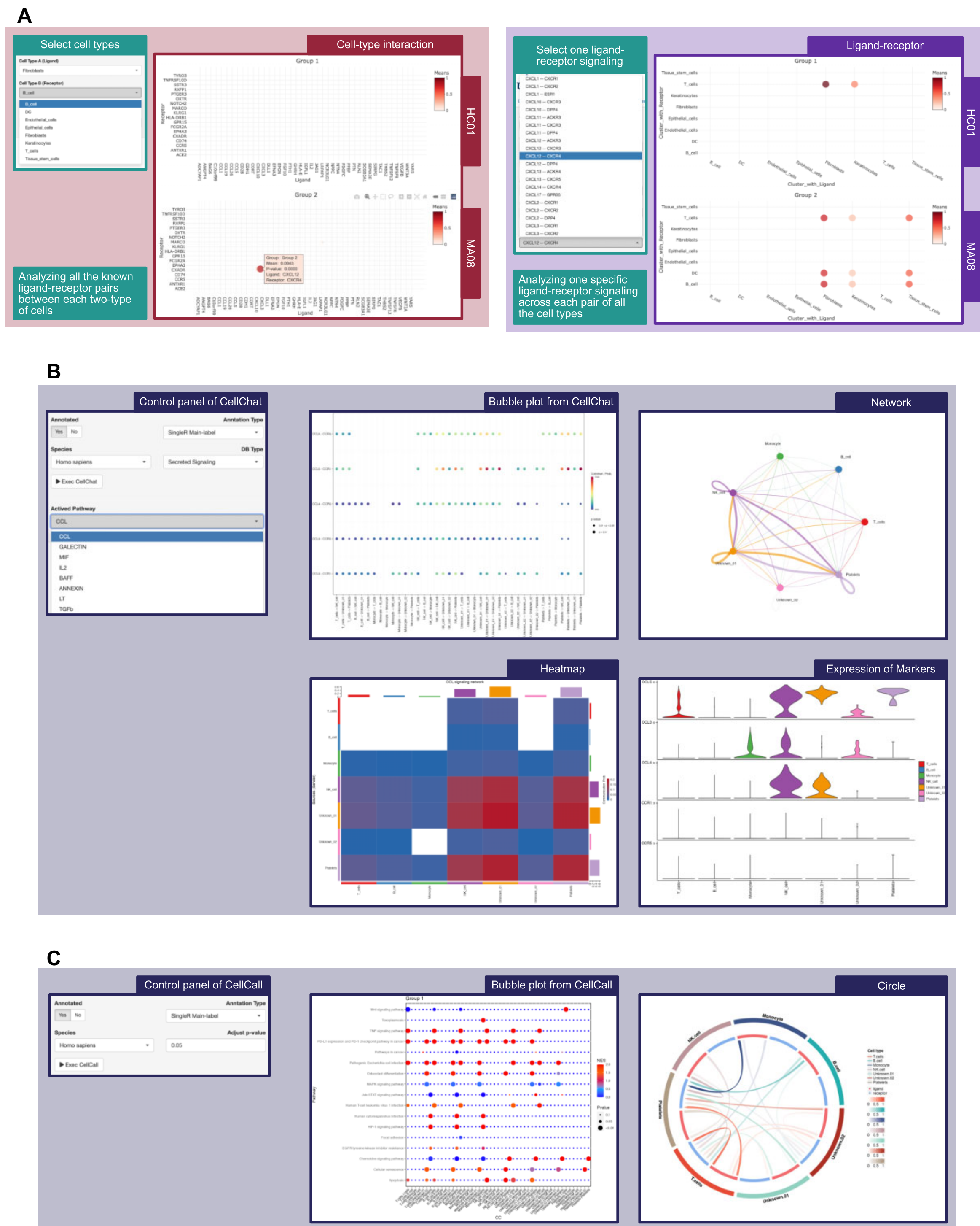


**Figure 3.** Demonstration of the Cell-to-cell communication module, including CellphoneDB, CellChat, and Cellcall. (**A**) Workflow of **CellphoneDB** analysis. The Control panel showed each cell type providing either 'Ligand' or 'Receptor', respectively, when two types of cells were defined for prediction. The dotplot showed that the 'healthy' group (HC01) had significantly different ligand-receptor profiles, compared to the 'lymphoma' group (MA08). Control panel representing how to analyze the specific 'ligand-receptor' pair, and the '*CXCL12–CXCR4*' pair as an example. The dotplot showed the relative interaction of each cell group based on a measure of '*CXCL12–CXCR4*' pair. *P*-value representing by circle size, and the color representing the average expression level of ligand-receptor pairs. (**B**) Workflow of **CellChat** analysis. Before running CellChat, users need to set the species information and database type in the control panel. The CellChat results can be displayed based on different activated pathways. Taking the pbmc3k dataset as an example, Grace provided the activation level of different ligand-receptor pairs between different cell types (bubble plot), the strength of interactions between different cell types (network and heatmap), and the differential expression of markers in the selected activated pathway between different cell types (expression of markers). (**C**) Workflow of **Cellcall** analysis. During running, Cellcall requires users to provide species information and FDR threshold (Control panel of CellCall). Using the pbmc3k dataset as an example, Grace provided the activation levels of different ligand-receptor pairs between different cell types (Bubble plot), and the strength of the interaction from ligand to receptor between different cell types (circle plot).

GRACE to map the query dataset onto an annotated atlas ('PBMC 10k'). This module successfully annotated the unknown clusters as effector memory CD8+ T cells, which was confirmed by marker genes (CD8A, CD8B, GZMB and NKG7) (28) (Figure S2C–E). The co-expression analysis module further allowed investigation of potential molecular relationships within the selected subset (Figure S2F).

### Case study 2: 'lymphoma' as two-group dataset

This case study demonstrates the functional modules used for analyzing a two-group dataset. Sample integration and batch effect correction were achieved using the RPCA algorithm (3) and nextflow package (13) (Figure S3A). After performing quality control, dimension reduction, and cell-type annotation (Figure S1), 'DC' cells were selected to examine subclustering function. The expression of marker genes and automatic cell-type annotation simultaneously annotated DC subsets (Figure S3A). The implemented Monocle2 package (4) was performed to further study the trajectory of DC subsets across different lineages. Two conventional dendritic cells (cDCs) including cDC1 lineage (expressing *CLEC9A* and *XCR1*) and cDC2 lineage (expressing *CD11b*/*ITGAM* and *CD172a*/*SIRPA*) could be identified (29) (Figure 2A, S3B). To better visualize these fate-defined subsets, the tree with two branches corresponding to two major lineages was generated independently (Figure 2A, S3B). As expected, the cDC1 lineage (including subC03, subC06 and subC07) can easily be distinguished from the cDC2 lineage (Figure S3B). GRACE has integrated SPRING, which reconstructed the kNN graph for 'DC' cells and mapped Monocle2's pseudotime onto the kNN graph (Figure 2B). The expression level of CLEC9A was also mapped onto both Monocle2-DDRTree and SPRING kNN graph simultaneously (Figure 2B, C). The red circle indicates subcluster 3 under resolution 0.2, which was annotated as DC:monocyte-derived (Figure 2B, C). Furthermore, the expression of selected genes could be viewed to study the potential roles in lineage differentiation (Figure 2C, Supplementary Figure S3B).

GRACE provides module to perform pathway enrichment at single-cell resolution. The fold-changes and p-values of each gene-set can be calculated for individual cells (Figure S4A). The diverse sets of biological processes among different subtypes or along developmental trajectory could be observed, respectively (Figure 2C, S3C). Simultaneous visualization of gene expression in selected pathway (For instance, the visualization of *CCR1*, *CCL2* and CXCL2 from GO:1990869 pathway using both UMAP and DDRtree plots) could facilitate an empirical investigation of the relationship between subsets and diverse molecular signatures (Figure S4B).

For cell-cell communication, GRACE provides two ways to study cell-cell communication. First, users can analyze all the known ligand-receptor pairs between each two-type of cells (Figure 3A). For example, *CXCL12–CXCR4*-axis had stronger interaction between fibroblast and B cells (Figure 3A), which has already been considered an attractive target for cancer therapies (30). Second, GRACE also supports the investigation of signaling transduction between each pair of cells with one specific ligand-receptor pair (Figure 3A). For instance, the enhanced interaction between B and fibroblasts existed in lymphoma tissue, compared to healthy tissue (Figure 3A).

## DISCUSSION

Single-cell transcriptome technology requires the comprehensive and sophisticated computational software for data integration, processing, analysis, mining and interpretation. In this study, we developed the web-based platform GRACE, based on a GUI computing server, to bridge the gap between experimental and bioinformatics research. It provides a user-friendly framework for investigating cellular heterogeneity, discovering cell subtypes, and revealing important biological processes.

This platform improves on many novel functions in various ways. GRACE is the only platform that offers comprehensive analysis capabilities, including doublet removal, dimensionality reduction, batch effect removal, clustering, cell annotation, developmental trajectory, and cell-to-cell communication analysis, as compared to all the other platforms (Supplementary table 1). Only GRACE has data management and long-term data storage capabilities, which enable users to easily modify parameters or switch to different software or algorithms. In addition to incorporating more software packages at each step to provide a greater range of analysis methods, GRACE also allows users to customize parameter settings as much as possible. The fully integrated analysis offers greater versatility than simply assembling modules such as Galaxy (6), or GranatumX (7), which only allows users to perform a few analysis tasks at a time. Furthermore, GRACE is the only platform that provides subcluster analysis, which enables users to perform in-depth investigations, particularly after cell-type annotation (Supplementary table 1). The heterogeneity and functionality of cellular subpopulations are currently hot topics in cutting-edge research (31–33). Moreover, the all-in-one format of GRACE is advantageous when linking data analysis and visualization. Compared to UCSC Cell Browser (9), SCP (https://singlecell.broadinstitute.org/single_cell) and cellxgene (8), GRACE allows seamless linkage that enriches user experience., GRACE also supports multiple data formats frequently used in scRNA-seq analysis, including tab-separated value/text (tsv/txt), comma-separated value (csv) formats directly downloaded from GEO database, normal RDS format as input of Seurat package (3), and hierarchical data format (HDF5) format generated from 10x Cellranger package. Fifth, many best-practice pipelines were implemented into GRACE, contributing to scalable, reproducible and straightforward development in both online platform and local server. In conclusion, GRACE provides easy access to interactive visualization, customized parameters, and publication-ready graphs for scRNA-seq analysis.

## DATA AVAILABILITY

All source code and its docker image can be free downloaded (https://github.com/th00516/GRACE) and Zenodo

(https://doi.org/10.5281/zenodo.7923930). The instructions and online video tutorials are available on the website homepage. (http://grace.flowhub.com.cn). The atlas reference is available on the Dropbox sharing link (https://www.dropbox.com/sh/zvsziyryjunqglm/AABY4sPTvGQPyxI5PlWzqBWXa?dl=0) and Zenodo (https://doi.org/10.5281/zenodo.7923930).

## SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

## ACKNOWLEDGEMENTS

*Author contributions*: Z.W. and H.Y. designed the project. H.Y and Z.W. developed GRACE software. Z.W., X.Z. and H.Y. wrote the manuscript and organized all figures, Y.W. designed and developed the project website, and edited the tutorial videos. All authors discussed results and commented on the manuscript.

## FUNDING

National Key R&D Program of China [2022YFA1103300, 2022YFA1103303, 2022YFA1103304]; National Natural Science Foundation of China [82020108004]; Natural Science Foundation of Chongqing [cstc2019jcyj-msxmX0421]; Translational Research Grant of NCRCH [2020ZKZC02]; Youth Talent Development Program from Second Affiliated Hospital, Army Medical University [2022YQB014].
*Conflict of interest statement.* None declared.

## REFERENCES

1. Lei,Y., Tang,R., Xu,J., Wang,W., Zhang,B., Liu,J., Yu,X. and Shi,S. (2021) Applications of single-cell sequencing in cancer research: progress and perspectives. *J. Hematol. Oncol. J. Hematol. Oncol.*, **14**, 91.
2. Guo,R., Lü,M., Cao,F., Wu,G., Gao,F., Pang,H., Li,Y., Zhang,Y., Xing,H., Liang,C. *et al.* (2021) Single-cell map of diverse immune phenotypes in the acute myeloid leukemia microenvironment. *Biomark. Res.*, **9**, 15.
3. Hao,Y., Hao,S., Andersen-Nissen,E., Mauck,W.M., Zheng,S., Butler,A., Lee,M.J., Wilk,A.J., Darby,C., Zager,M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
4. Qiu,X., Mao,Q., Tang,Y., Wang,L., Chawla,R., Pliner,H.A. and Trapnell,C. (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*, **14**, 979–982.
5. Wolf,F.A., Angerer,P. and Theis,F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.*, **19**, 15.
6. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
7. Garmire,D.G., Zhu,X., Mantravadi,A., Huang,Q., Yunits,B., Liu,Y., Wolfgruber,T., Poirion,O., Zhao,T., Arisdakessian,C. *et al.* (2021) GranatumX: a community-engaging, modularized, and flexible webtool for single-cell data analysis. *Genomics Proteomics Bioinformatics*, **19**, 452–460.
8. Megill,C., Martin,B., Weaver,C., Bell,S., Prins,L., Badajoz,S., McCandless,B., Pisco,A.O., Kinsella,M., Griffin,F. *et al.* (2021) cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. bioRxiv doi: https://doi.org/10.1101/2021.04.05.438318, 06 April 2021, preprint: not peer reviewed.
9. Speir,M.L., Bhaduri,A., Markov,N.S., Moreno,P., Nowakowski,T.J., Papatheodorou,I., Pollen,A.A., Raney,B.J., Seninge,L., Kent,W.J. *et al.* (2021) UCSC Cell Browser: visualize your single-cell data. *Bioinformatics*, **37**, 4578–4580.
10. 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**, 3123–3125.
11. Jiang,A., Lehnert,K., You,L. and Snell,R.G. (2022) ICARUS, an interactive web server for single cell RNA-seq analysis. *Nucleic Acids Res.*, **50**, W427–W433.
12. Tran,H.T.N., Ang,K.S., Chevrier,M., Zhang,X., Lee,N.Y.S., Goh,M. and Chen,J. (2020) A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.*, **21**, 12.
13. Di Tommaso,P., Chatzou,M., Floden,E.W., Barja,P.P., Palumbo,E. and Notredame,C. (2017) Nextflow enables reproducible computational workflows. *Nat. Biotechnol.*, **35**, 316–319.
14. Hafemeister,C. and Satija,R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.*, **20**, 296.
15. McGinnis,C.S., Murrow,L.M. and Gartner,Z.J. (2019) DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.*, **8**, 329–337.
16. Haghverdi,L., Lun,A.T.L., Morgan,M.D. and Marioni,J.C. (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.*, **36**, 421–427.
17. Korsunsky,I., Millard,N., Fan,J., Slowikowski,K., Zhang,F., Wei,K., Baglaenko,Y., Brenner,M., Loh,P. and Raychaudhuri,S. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*, **16**, 1289–1296.
18. Gayoso,A., Lopez,R., Xing,G., Boyeau,P., Valiollah Pour Amiri,V., Hong,J., Wu,K., Jayasuriya,M., Mehlman,E., Langevin,M. *et al.* (2022) A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.*, **40**, 163–166.
19. Efremova,M., Vento-Tormo,M., Teichmann,S.A. and Vento-Tormo,R. (2020) CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.*, **15**, 1484–1506.
20. Jin,S., Guerrero-Juarez,C.F., Zhang,L., Chang,I., Ramos,R., Kuan,C.-H., Myung,P., Plikus,M.V. and Nie,Q. (2021) Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.*, **12**, 1088.
21. Zhang,Y., Liu,T., Hu,X., Wang,M., Wang,J., Zou,B., Tan,P., Cui,T., Dou,Y., Ning,L. *et al.* (2021) CellCall: integrating paired ligand–receptor and transcription factor activities for cell–cell communication. *Nucleic Acids Res.*, **49**, 8520–8534.
22. Weinreb,C., Wolock,S. and Klein,A.M. (2018) SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, **34**, 1246–1248.
23. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R Package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.*, **16**, 284–287.
24. Aran,D., Looney,A.P., Liu,L., Wu,E., Fong,V., Hsu,A., Chak,S., Naikawadi,R.P., Wolters,P.J., Abate,A.R. *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.*, **20**, 163–172.
25. participants in the 1st Human Cell Atlas Jamboree, Lun,A.T.L., Riesenfeld,S., Andrews,T., Dao,T.P., Gomes,T. and Marioni,J.C. (2019) EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.*, **20**, 63.
26. Griffiths,J.A., Richard,A.C., Bach,K., Lun,A.T.L. and Marioni,J.C. (2018) Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.*, **9**, 2667.
27. Rindler,K., Jonak,C., Alkon,N., Thaler,F.M., Kurz,H., Shaw,L.E., Stingl,G., Weninger,W., Halbritter,F., Bauer,W.M. *et al.* (2021) Single-cell RNA sequencing reveals markers of disease progression in primary cutaneous T-cell lymphoma. *Mol. Cancer*, **20**, 124.
28. Hornburg,M., Desbois,M., Lu,S., Guan,Y., Lo,A.A., Kaufman,S., Elrod,A., Lotstein,A., DesRochers,T.M., Munoz-Rodriguez,J.L. *et al.* (2021) Single-cell dissection of cellular components and interactions shaping the tumor immune phenotypes in ovarian cancer. *Cancer Cell*, **39**, 928–944.
29. Villani,A.-C., Satija,R., Reynolds,G., Sarkizova,S., Shekhar,K., Fletcher,J., Griesbeck,M., Butler,A., Zheng,S., Lazo,S. *et al.* (2017)

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, **356**, eaah4573.
30. AlMusawi,S., Ahmed,M. and Nateri,A.S. (2021) Understanding cell-cell communication and signaling in the colorectal cancer microenvironment. *Clin. Transl. Med.*, **11**, e308.
31. Kinker,G.S., Greenwald,A.C., Tal,R., Orlova,Z., Cuoco,M.S., McFarland,J.M., Warren,A., Rodman,C., Roth,J.A., Bender,S.A. *et al.* (2020) Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.*, **52**, 1208–1218.
32. Brown,C.C., Gudjonson,H., Pritykin,Y., Deep,D., Lavallée,V.-P., Mendoza,A., Fromme,R., Mazutis,L., Ariyan,C., Leslie,C. *et al.* (2019) Transcriptional basis of mouse and human dendritic cell heterogeneity. *Cell*, **179**, 846–863.
33. Papalexi,E. and Satija,R. (2018) Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.*, **18**, 35–45.

© The Author(s) 2023. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com