Although the analyses of scRNA-seq data are complex and specific procedures may vary for different purposes, many mature tools have been developed to integrate two or more functions that greatly simplify the independent operations. **Table 1** listed several popular tools for the scRNA-seq analysis.

1. **SC3, an R package for clustering**

SC3 was proposed by Kiselev et al. 1 in 2017. It is an interactive R package that uses a parallelization approach to avoid the need for user-specified parameters. SC3 was verified experimentally on 12 single-cell RNA-seq data sets. SC3 constrains parameter values via a pipeline, and was found to be superior to five other tested methods in terms of accuracy and stability. Because SC3 has a long run time, Kiselev et al. proposed randomly selecting subsets and constructing clusters based on the random matrix theory. They found that the estimated value was consistent with the number of original clusters suggested by them.

1. **Single-cell regulatory network inference and clustering (SCENIC)**

SCENIC 2 was proposed by Aibar et al. in 2017 who used it to identify stable cell states in tumor and brain single-cell RNA-seq data based on the activity of the gene regulatory networks in each cell. The authors proposed two complementary methods to handle the large dimensions of single-cell data: 1) small sample extraction to infer the gene regulatory network, and 2) gradient enhancement instead of random forest to achieve a more efficient solution. They demonstrated that single-cell data were suitable for gene regulation and that genomic regulatory codes can be used to guide the identification of transcription factors and cell states.

1. **Single-cell interpretation via multi-kernel learning (SIMLR)**

In 2017, Wang et al. 3 proposed SIMLR, a kernel-based similarity learning method, for dimensionality reduction of single-cell RNA-seq data. SIMLR also can be applied to large scale data sets. They conducted single kernel comparisons on four data sets without weight terms, and showed that adding weight terms significantly enhanced SIMLR performance

1. **Shared nearest neighbor (SNN)-Cliq)**

Xu et al. 4 developed SNN-Cliq in 2005 for grouping cells of the same type. Single-cell RNA-seq data usually have tens of thousands of dimensions, and only a few of the thousands of genes are significantly expressed in different types of cells, which makes the clustering problem difficult. SNN-Cliq combined with a SNN similarity metric can automatically determine the number of clusters, especially in high-dimensional single-cell data, which is a great advantage.

1. **Nonnegative Matrix Factorization (NMF)**

In 2016, Shao et al. 5 proposed NMF to identify subgroups in single-cell RNA-seq data sets. Identifying cell types from single-cell data is an unsupervised problem. Although PCA is used widely, single-cell data are generally too noisy. The first few principal components extracted from PCA can explain only a small part of the differences, and cell subgroups are not easy to distinguish through the projection of the first several dimensions. The NMF approach is different from PCA because its feature superposition constraint is non-negative. NMF was designed specially to detect single parts, which helps to detect the natural groupings of individual cells and functional cell subsets.

1. **BackSPIN**

Zeisel et al. 6 developed BackSPIN in 2015 and tested it on the adult nervous system, which is highly complex and has many cell types that are challenging to identify. Single-cell RNA-seq data were used to classify mammalian cortical cells. BackSPIN detected different types of cells based on molecule clustering and showed that transcription factors formed a complex hierarchical regulatory code, revealing the diversity of brain cell types and their transcriptomes.

1. **BiSNN-Walk**

Shi and Huang 7 proposed BiSNN-Walk, an iterative bi-clustering method based on SNN-Cliq 4. BiSNN-Walk differs from SNN-Cliq in that it returns a sorted list as a reliable indicator of a cluster. In addition, BiSNN-Walk uses a metric method based on entropy to select the starting point of clustering, and its clustering ability was tested on three single-cell RNA-seq data sets.

1. **SINCERA**

Guo et al. 8,9 proposed SINCERA, a pipeline for single-cell RNA-seq profiling analysis. SINCERA can identify cell types, gene signatures, and can determine key nodes. Analysis of mouse lung cells using the SINCERA pipeline distinguished the main cell types of fetal lung. Guo et al. subsequently introduced logistic regression models that predict gene sequences, providing a valuable tool for analyzing single-cell RNA-seq data.

1. **SEURAT**

Satija et al. 10 proposed SEURAT, a toolbox for cell spatial localization. SEURAT combines single-cell RNA-seq data with in situ RNA patterns to predict cell spatial localization. The toolbox was applied to infer the spatial location of a complete transcriptome and correctly located unusual subpopulations. The reliability of SEURAT was verified using the RNA-seq data of 851 single cells from *Danio rerio* embryos.

1. **Monocle**

To study cell differentiation, the expression profiles of individual cells are required. Monocle was developed by Trapnell et al. 11 as an unsupervised algorithm for analyzing single cell gene expression data to reveal the expression sequence of key regulatory factors and the interactions associate with differentiation. The authors used the Monocle algorithm to study mouse myoblasts and found eight transcription factors that had not been considered previously. Single-cell RNA-Seq data collected at different time points can help to reveal key events in differentiation.

1. **Single Cell Representation Learning (SCRL)**

SCRL 12 is a non-linear dimensionality reduction method based on machine learning and clustering that was developed by Li et al. in 2017. To process drop-out events of single-cell RNA data, SCRL uses biological knowledge such as high-throughput RNA sequencing and adopts a network embedding method to express more abundant and low-dimensional expression of single-cell RNA-seq data.

**Table** Summary of popular analytical tools for scRNA-seq. *Note*: the package link is bound to each tool name and the publication link is bound to its year.

|  |  |  |  |
| --- | --- | --- | --- |
| Tools | Year | Program | Tags |
| [SAMtools](https://github.com/samtools/samtools) | [2009](https://www.ncbi.nlm.nih.gov/pubmed/19505943) | C | (Not scRNA-seq specific) post-alignment processing |
| [SART](https://github.com/alexdobin/STAR) | [2013](https://www.ncbi.nlm.nih.gov/pubmed/23104886) | C | (Not scRNA-seq specific) alignment |
| [Monocle2](https://github.com/cole-trapnell-lab/monocle2-rge-paper) | [2014](https://www.nature.com/articles/nmeth.4402) | R | Clustering, differential expression, dimensionality reduction, visualization |
| [BackSPIN](https://github.com/linnarsson-lab/BackSPIN) | [2015](http://science.sciencemag.org/content/347/6226/1138) | Python | Gene filtering, biclustering, cell type prediction |
| [SINCERA](https://github.com/xu-lab/SINCERA) | [2015](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004575) | R | Quality control, normalization, gene filtering, clustering, differential expression, marker genes, cell type prediction |
| [MAST](https://github.com/RGLab/MAST) | [2015](https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0844-5) | R | Quality control, normalization, differential expression, network construction |
| [Kallisto](https://github.com/pachterlab/kallisto) | [2016](https://www.nature.com/articles/nbt.3519) | C | Quantification |
| [BPSC](https://github.com/nghiavtr/BPSC) | [2016](https://academic.oup.com/bioinformatics/article/32/14/2128/2288270) | R | beta-Poisson mixture model |
| [salmon](https://github.com/COMBINE-lab/salmon) | [2017](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5600148/) | C++ | UMI, quantification |
| [UMI-tools](https://github.com/CGATOxford/UMI-tools) | [2017](https://genome.cshlp.org/content/early/2017/01/18/gr.209601.116.abstract) | Python | UMI, quantification |
| [CellRanger](https://github.com/10XGenomics/cellranger) | [2017](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger) | Python/R | Alignment, UMI, quantification, quality control, clustering, differential expression, marker genes, dimensionality reduction, visualization |
| [SC3](https://github.com/hemberg-lab/sc3) | [2017](https://www.nature.com/articles/nmeth.4236) | R | Gene filtering, clustering, cell type prediction |
| [Scater](https://github.com/davismcc/scater) | [2017](https://academic.oup.com/bioinformatics/article/33/8/1179/2907823) | R | Quantification, quality control, normalization, dimensional reduction, visualization |
| [SCENIC](https://github.com/aertslab/SCENIC) | [2017](https://www.nature.com/articles/nmeth.4463) | R/Python | Clustering, network construction, regulon prediction, visualization |
| [Seurat](https://github.com/satijalab/seurat) | [2018](https://www.nature.com/articles/nbt.4096) | R | Normalization, gene filtering, clustering, differential expression, marker gene, dimensionality reduction, visualization |
| [SAVER](https://github.com/mohuangx/SAVER) | [2018](https://www.nature.com/articles/s41592-018-0033-z) | R | Imputation |
| [SCDE](https://github.com/hms-dbmi/scde) | [2014](https://www.nature.com/articles/nmeth.2967),[2016](https://www.nature.com/articles/nmeth.3734) | R | Differential expression, pathway analysis, visualization |
| [QUBIC2](https://github.com/maqin2001/qubic2) | *preprint* | C | Biclustering, network construction |
| [GeneQC](http://bmbl.sdstate.edu/GeneQC/home.html) | [2018](https://www.frontiersin.org/articles/10.3389/fgene.2018.00313/full) | Server | Alignment, mapping uncertainty, realignment, quantification |
| [IRIS-EDA](http://bmbl.sdstate.edu/IRIS/) | [2018](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006792) | Correlation analysis, clustering, differential expression, visualization, dimensionality reduction |
| [IRIS3](http://bmbl.sdstate.edu/iris3) | - | Integrated web server for gene regulation elucidation |
| [KEGG](https://www.genome.jp/kegg/) | [2000](https://www.ncbi.nlm.nih.gov/pubmed/10592173) | Database | Gene annotation |
| [EnrichR](http://amp.pharm.mssm.edu/Enrichr/) | [2013](https://www.ncbi.nlm.nih.gov/pubmed/23586463),[2016](https://www.ncbi.nlm.nih.gov/pubmed/27141961) | Enrichment analysis |
| [DAVID](https://david.ncifcrf.gov/) | [2009](https://www.ncbi.nlm.nih.gov/pubmed/19033363?dopt=Abstract) | Gene annotation |
| [Harmonizome](http://amp.pharm.mssm.edu/Harmonizome/) | [2016](http://database.oxfordjournals.org/content/2016/baw100.short) | Gene/protein function |
| [SwissRegulon](http://swissregulon.unibas.ch/sr/annotations) | [2007](https://www.ncbi.nlm.nih.gov/pubmed/17130146),[2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531101/) | Regulon database |
| [reactome](https://reactome.org/) | [2004](https://www.ncbi.nlm.nih.gov/pubmed/18428737) | Gene annotation, pathway construction |

Ref

1 Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483-486, doi:10.1038/nmeth.4236 (2017).

2 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

3 Wang, B., Zhu, J., Pierson, E., Ramazzotti, D. & Batzoglou, S. J. N. m. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. **14**, 414 (2017).

4 Xu, C. & Su, Z. J. B. Identification of cell types from single-cell transcriptomes using a novel clustering method. **31**, 1974-1980 (2015).

5 Shao, C. & Höfer, T. J. B. Robust classification of single-cell transcriptome data by nonnegative matrix factorization. **33**, 235-242 (2017).

6 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).

7 Shi, F. & Huang, H. J. J. o. C. B. Identifying cell subpopulations and their genetic drivers from single-cell RNA-seq data using a biclustering approach. **24**, 663-674 (2017).

8 Guo, M., Wang, H., Potter, S. S., Whitsett, J. A. & Xu, Y. SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput Biol* **11**, e1004575, doi:10.1371/journal.pcbi.1004575 (2015).

9 Guo, M. & Xu, Y. in *Transcriptome Data Analysis* 209-222 (Springer, 2018).

10 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

11 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. **32**, 381 (2014).

12 Li, X. *et al.* Network embedding-based representation learning for single cell RNA-seq data. **45**, e166-e166 (2017).