

Automatic cell type annotation

☰ 相關類別

Dr. Po-Jung Huang, Chang Gung University, Taiwan. e-mail: pjhuang@gap.cgu.edu.tw



<https://reurl.cc/94yIMO>

1. Overview of cell annotation workflow
2. Strategies for automatic cell annotation
 - 2-1 Marker-based annotation approach
 - 2-2 Reference-based annotation approach
 - 2-3 Strengths & pitfalls of marker-based and reference-based annotation approaches
3. Reference-based annotation of scRNA-Seq (SingleR)
 - 3-1 Cell type reference datasets (celldex package)
 - 3-2 Schematic of SingleR
 - 3-3 SingleR browser
4. Marker-based annotation of scRNA-Seq
 - 4-1 ScType
 - 4-2 scCATCH
5. ShinySC
 - 5-1 Upload
 - 5-1-1 **Input format:**
 - 5-1-2 Supported organisms:
 - 5-2 Quality control
 - 5-3 Feature selection
 - 5-3-1 Identification of highly variable features & Dimension reduction (PCA)
 - 5-3-2 Determination of optimal clustering resolution
 - 5-3-2 Define optimal resolution
 - 5-4. Clustering
 - 5-4-1 Non-linear dimension reduction
 - 5-4-2 Heatmap & Vizdim
 - 5-5 Find markers
 - 5-6. Automated cell-type annotation
 - 5-6-1 ScType
 - 5-6-2 SingleR
 - 5-6-3

1. Overview of cell annotation workflow

Single-cell transcriptomics can profile thousands of cells in a single experiment and identify novel cell types, states and dynamics in a wide variety of tissues and organisms. Standard experimental protocols and analysis workflows have been developed to create single-cell transcriptomic maps from tissues. This tutorial focuses on how to interpret these data to identify cell types, states and other biologically relevant patterns with the objective of creating an annotated map of cells. We recommend a **three-step workflow** including **automatic cell annotation** (wherever possible), **manual cell annotation** and **verification**. Frequently encountered challenges are discussed, as well as strategies to address them. Guiding principles and specific recommendations for software tools and resources that can be used for each step are covered, and an R notebook is included to help run the recommended workflow. **Basic familiarity with computer software is assumed, and basic knowledge of programming (e.g., in the R language) is recommended.**

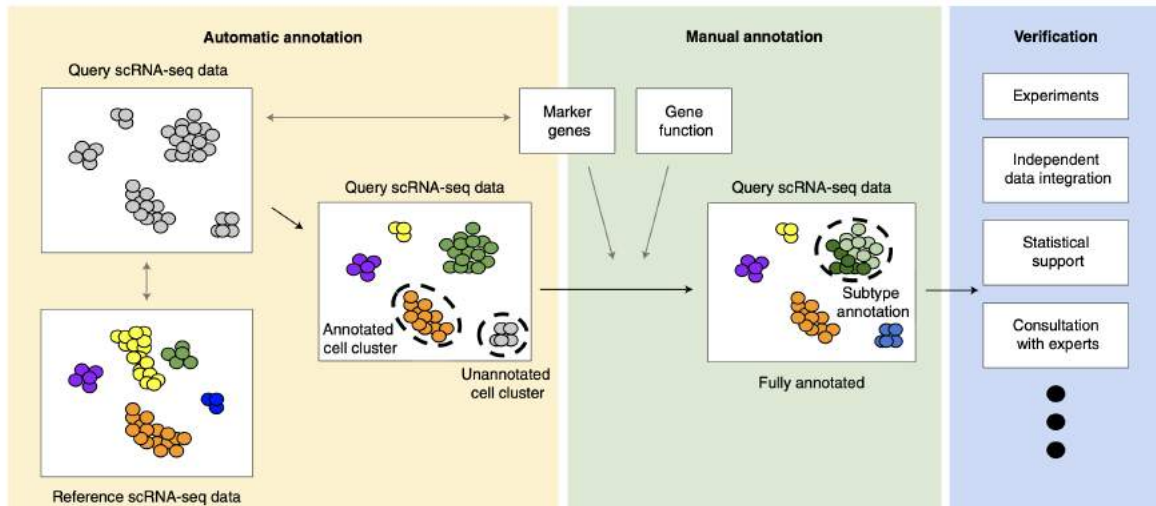


Fig. 2 | Cell annotation workflow. The recommended cell annotation process is composed of three major steps: automatic cell annotation, manual cell annotation and verification. The scRNA-seq data typically enter the workflow as a clustered gene-by-cell matrix, which is visualized using a dimensionality reduction method. An automatic cell annotation method is used to annotate cells either by comparison of the data with annotated reference data (e.g., a single-cell atlas) or using known marker genes indicative of a specific cell type. Manual annotation confirms or provides further detail for annotated cells or clusters or identifies the cell type of unlabeled clusters. Cell type can be manually inferred using a combination of marker genes, pathway analysis and differentially expressed genes with known functional information. Cell annotations are often verified using independent sources, such as new validation experiments, or comparison to complementary data, such as spatial transcriptomics data.

NATURE PROTOCOLS|VOL 16|JUNE 2021|2749–2764|

▼ Comparison of the caveats and recommendations for different approaches to cell annotation

Table 1 Comparison of the caveats and recommendations for different approaches to cell annotation			
Stage of analysis	Aspect of analysis	Potential caveats	Recommendation
Automatic cell annotation	All automatic methods	Fast, but not effective for poorly characterized cells	Use manual annotation for poorly characterized cells
	Annotating clusters	May miss important differences between cells	Use automatic annotation of clusters to get a general idea of cell type and then refine labels manually. In addition, use multiple cluster-based methods and compare results
	Annotating individual cells	Ideal, but requires high reads per cell	Experiments with low reads per cell require cluster-based annotation
	Marker-based annotation methods	Marker genes not easily accessible for all cell types; may result in conflicting or absent cell labels	Requires expert knowledge to curate more extensive marker lists
	Reference-based annotation methods	Perform poorly with incomplete or poorly matched reference data, which may result in conflicting or absent cell labels Often requires batch correction, which may reduce the accuracy of results Mistakes in reference data get carried over to results	Use well-matched reference data or marker-based methods if such data are unavailable Analyze the reference data for strong biological signals. Use a good experimental protocol that will prevail over batch effects Analyze reference data for potential errors before using
	Comparing results from different automatic annotation methods	Results may not agree with each other	Compare confidence scores of respective labels and consider label agreement (majority rule); resolve conflicts using manual annotation Consider the possibility of cell subtypes, new cell types or gradients and cell states
Expert manual cell annotation	All manual methods	Slow, labor-intensive Subjective	Whenever possible, begin with automatic annotation to determine general cell labels Work with an expert; consider multiple cell-type conclusions
	Marker-based annotation	Cell types not distinguishable by a single marker Known markers not distinguishing cell types Conflicting marker gene sets between sources	Use multiple markers for each cell type Curate larger lists of markers from the literature, additional experiments or experts Select a marker gene set that best represents the biological signal being looked for in the data (e.g., if looking for cell subtypes, use more extensive gene sets than what is used for general cell-type annotation)

2. Strategies for automatic cell annotation

2-1 Marker-based annotation approach

- Characteristic expression of **known marker genes**
 - Known relationships between **marker genes** and **cell types** from **databases**
- ▼ **MSigDB** (<https://www.gsea-msigdb.org/gsea/msigdb>)

MSigDB
Molecular Signatures Database

Molecular Signatures Database

Overview

The Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use with GSEA software, divided into Human and Mouse collections. From this web site, you can

- ▶ **Examine** a gene set and its annotations. See, for example, the HALLMARK_APOPTOSIS human gene set page.
- ▶ **Browse** gene sets by name or collection.
- ▶ **Search** for gene sets by keyword.
- ▶ **Investigate** gene sets:
 - ▶ **Compute overlaps** between your gene set and gene sets in MSigDB.
 - ▶ **Categorize** members of a gene set by gene families.
 - ▶ **View the expression profile** of a gene set in a provided public expression compendia.
 - ▶ Investigate the gene set in the online **biological network repository NDEx**
- ▶ **Download** gene sets.

License Terms

GSEA and MSigDB are available for use under these [license terms](#).

Please register to download the GSEA software and the MSigDB gene sets, and to use our web tools. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Current Version

Human MSigDB v2023.1.Hs updated March 2023. [Release notes](#).

Mouse MSigDB v2023.1.Mm updated March 2023. [Release notes](#).

Citing the MSigDB

To cite your use of the Molecular Signatures Database (MSigDB), a joint project of UC San Diego and Broad Institute, please reference Subramanian, Tamayo, et al. (2005, PNAS) and one or more of the following as appropriate: Liberzon, et al. (2011, Bioinformatics), Liberzon, et al. (2015, Cell Systems), and also the source for the gene set as listed on the gene set page.

Funding

GSEA and MSigDB are currently funded by a grant from NCI's Informatics Technology for Cancer Research (ITCR)

Human Collections

- H** **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.
- C5** **ontology gene sets** consist of genes annotated by the same ontology term.
- C1** **positional gene sets** corresponding to human chromosome cytogenetic bands.
- C6** **oncogenic signature gene sets** defined directly from microarray gene expression data from cancer gene perturbations.
- C2** **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.
- C7** **immunologic signature gene sets** represent cell states and perturbations within the immune system.
- C3** **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.
- C8** **cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of human tissue.
- C4** **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

Mouse Collections

- MH** **mouse-ortholog hallmark gene sets** are versions of gene sets in the MSigDB Hallmarks collection mapped to their mouse orthologs.
- M3** **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.
- M1** **positional gene sets** corresponding to mouse chromosome cytogenetic bands.
- M5** **ontology gene sets** consist of genes annotated by the same ontology term.
- M2** **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.
- M8** **cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of mouse tissue.

Other Gene Set Resources

- ▶ **Signatures of post-translational modification (PTM) sites** from the Proteomics group at the Broad Institute
- ▶ **Miscellaneous gene sets** from community contributors.

If you would like to suggest or contribute new gene sets, please contact us at genesets@broadinstitute.org.

- ▼ **PanglaoDB** (<https://panglaodb.se>)

PanglaoDB is a database for the scientific community interested in exploration of single cell RNA sequencing experiments from mouse and human. We collect and integrate data from multiple studies and present them through a unified framework.

Usage examples

- Run a gene search for [SOX2](#), [PECAM1](#) or [ACE2](#)
- Browse the full list of [samples](#)
- Explore the list of cell type markers for [Schwann cells](#)
- Browse cell types of the mouse [retina](#)
- Look at the expression of [CRX](#) in photoreceptor cells
- Find cell clusters where both [PECAM1](#) and [VCAM1](#) are expressed using a [boolean search](#) with the 'and' operator
- Find [quiescent neural stem cells](#) using AND+NOT

How to cite

Oscar Franzén, Li-Ming Gan, Johan L M Björkegren, *PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data*, *Database*, Volume 2019, 2019, baz046, doi:10.1093/database/baz046

What is single cell RNA sequencing?

Adapted from the [Wikipedia](#) article on the topic: *Single cell RNA sequencing examines the transcriptomes from individual cells with optimized next generation sequencing technologies, providing a higher resolution of gene expression and a better understanding of the function of an individual cell in the context of its microenvironment.*

Database statistics		
	<i>Mus musculus</i>	<i>Homo sapiens</i>
Samples	1063	306
Tissues	184	74
Cells	4,459,768	1,126,580
Clusters	8,651	1,748

Dataset of the day

Take a closer look at the cellular composition of [Calvaria](#), using a dataset which consists of 369 cells. Clustering of this dataset resulted in 1 cell clusters, containing among others, [Chondrocytes](#).

News

- 21-05-2020** Ongoing work to move to new hosting.
- 30-01-2020** A corrupted MySQL table caused dysfunction in the search function, the problem has now been fixed.
- 28-11-2019** We are looking for sponsors to host PanglaoDB. We have modest requirements (VPS with Ubuntu, etc). Please get in touch with us if you can provide help (contact@panglaodb.se).
- 01-07-2019** Updated the 2d view for data sets (now colors by cell type and not by cluster and colors are consistent across data sets). For example, see [this data set](#).
- 16-05-2019** Added more markers for [Tanycytes](#).
- 07-05-2019** Added markers for [Chromaffin cells](#).
- 01-05-2019** Markers for an additional cell types added: meet [the sebocyte](#).
- 30-04-2019** Added sensitivity and specificity to the [marker](#) list (shown separately for mouse and human).
[Show older news](#)

▼ [CellMarker \(http://117.50.127.228\)](http://117.50.127.228)

Welcome to CellMarker 2.0

Cell identification can be performed based on marker information stored in the CellMarker 2.0, which contains basic cells of organism formation (smooth muscle cells, epithelial cells...) and tumor cells (cancer stem cells, epithelial cancer cells...) that cause the organism to become cancerous and immune cells (T cells, B cells...) that fight off pathogens, etc.

Welcome to CellMarker 2.0

Click cells or tissues to quick search

Human Mouse

Quick Search

Cell type, Cell marker, Tissue Type OR Tissue

Human information

Species	Human
Tissue type	429
Cell type	1715
Cancer type	278
Cell marker sets	4334
Cell marker	15737

Mouse information

Species	Mouse
Tissue type	399
Cell type	1434
Cancer type	94
Cell marker sets	2185

▼ Literatures

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed®

Advanced

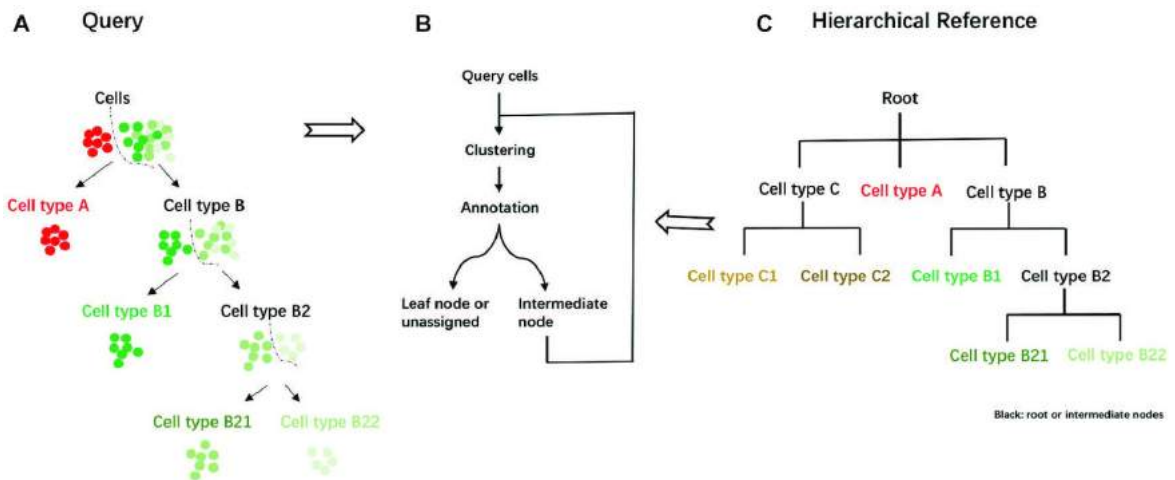
Search

PubMed® comprises more than 35 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.

AUCell ¹⁰³	Marker based	R	Single cells	Area under the curve to estimate marker gene set enrichment	Yes	Because of low detection rates at the level of single cells, it requires many markers for every cell type
SCINA ³⁴	Marker based	R	Single cells	Expectation maximization, Gaussian mixture model	(Optional)	Simultaneously clusters and annotates cells; robust to the inclusion of incorrect marker genes
GSEA/GSVA ^{36,104}	Marker based	R/Java	Clusters of cells	Enrichment test	Yes	Marker gene lists must be reformatted in GMT format. Markers must all be differentially expressed in the same direction in the cluster

NATURE PROTOCOLS|VOL 16|JUNE 2021|2749–2764|

▼ **scMRMA: single cell multiresolution marker-based annotation** *Nucleic Acids Research*, Volume 50, Issue 2, 25 January 2022, Page e7,



2-2 Reference-based annotation approach

- transfer labels from reference cell or cluster (well annotated scRNA-seq data)
- Reference single-cell data are obtained from
 - ▼ Gene Expression Omnibus (GEO <https://www.ncbi.nlm.nih.gov/geo/>)

NCBI Resources How To Sign In to NCBI

GEO Home Documentation Query & Browse Email GEO

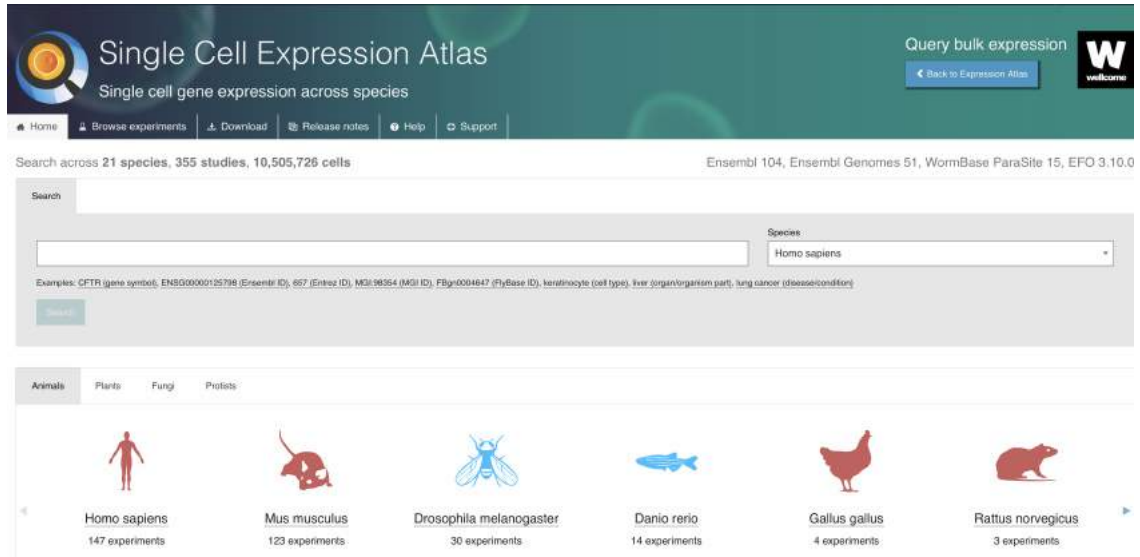
Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession

Getting Started	Tools	Browse Content
Overview FAQ About GEO DataSets About GEO Profiles About GEO2R Analysis How to Construct a Query How to Download Data	Search for Studies at GEO DataSets Search for Gene Expression at GEO Profiles Search GEO Documentation Analyze a Study with GEO2R Studies with Genome Data Viewer Tracks Programmatic Access FTP Site ENCODE Data Listings and Tracks	Repository Browser DataSets: 4348 Series: 204017 Platforms: 25181 Samples: 6522606
Information for Submitters	Submission Guidelines Update Guidelines	MIAME Standards Citing and Linking to GEO Guidelines for Reviewers GEO Publications

▼ Single Cell Expression Atlas (<https://www.ebi.ac.uk/gxa/sc/home>)



▼ Cell atlas projects

1. Human Cell Atlas (HCA) - The HCA aims to create comprehensive reference maps of all human cells, including their molecular profiles, for a better understanding of human health and disease.
Website: <https://www.humancellatlas.org/>
2. Mouse Cell Atlas (MCA) - The MCA project focuses on generating a comprehensive cell atlas of the mouse, providing a valuable resource for understanding mouse development, physiology, and disease models.
Website: <https://www.mousecellatlas.org/>
3. Human Protein Atlas (HPA) - The HPA aims to map the location of all human proteins in cells, tissues, and organs using various omics technologies, enabling researchers to explore protein expression and localization patterns.
Website: <https://www.proteinatlas.org/>
4. Tabula Muris - The Tabula Muris project focuses on creating a single-cell transcriptomic atlas of different organs and tissues in the mouse, providing insights into cellular diversity and function.
Website: <https://tabula-muris.ds.czbiohub.org/>
5. Fly Cell Atlas - The Fly Cell Atlas project aims to comprehensively map and characterize cell types in the fruit fly *Drosophila melanogaster*, enabling insights into fly development and physiology.
Website: <https://www.flycellatlas.org/>

singleCell Net ⁴²	Reference based	R	Single cells	Relative-expression gene pairs + random forest	Yes, but rarely does so even when it should ³³	10-100× slower than other methods; high accuracy
scmap-cluster ⁴¹	Reference based	R	Single cells	Consistent correlations	Yes	Fastest method available; balances false-positives and false-negatives; includes web interface for use with a large pre-built reference or custom reference set
scmap-cell ⁴¹	Reference based	R	Single cells	Approximate nearest neighbors	Yes	Assigns individual cells to nearest neighbor cells in reference; allows mapping of cell trajectories; fast and scalable
singleR ⁴³	Reference based	R	Single cells	Hierarchical clustering and Spearman correlations	No	Includes a large marker reference; does not scale to data sets of ≥10,000 cells; includes web interface with marker database
Scikit-learn ¹⁰²	Reference based	Python	Multiple possible	k-nearest neighbors, support vector machine, random forest, nearest mean classifier and linear discriminant analysis	(Optional)	Expertise required for correct design and appropriate training of classifier while avoiding overtraining

NATURE PROTOCOLS|VOL 16|JUNE 2021|2749–2764|

2-3 Strengths & pitfalls of marker-based and reference-based annotation approaches

Annotation Method	Strengths	Pitfalls

Marker-based	Fast and efficient for <u>well-characterized cell types</u>	May <u>not be effective</u> for <u>poorly characterized</u> or <u>rare cell types</u>
	Can be used to identify specific cell types based on <u>known markers</u>	May result in conflicting or absent cell labels if marker genes are not easily accessible for all cell types
		Requires expert knowledge to curate more extensive marker lists
Reference-based	Can be used to identify <u>novel cell types</u> or <u>subtypes</u> based on similarity to reference data	Requires <u>high-quality</u> and relevant <u>annotated reference data</u>
	Can be <u>more accurate</u> than marker-based methods for <u>well-matched</u> reference data	May perform poorly with <u>incomplete</u> or <u>poorly matched reference data</u> , which may result in <u>conflicting or absent cell labels</u>
	Can be used to integrate data from multiple sources	Often requires batch correction, which may reduce the accuracy of results


▼ ChatPDF

[s41596-021-00534-0.pdf](#)

ChatPDF - Chat with any PDF!

ChatPDF is the fast and easy way to chat with any PDF, free and without sign-in. Talk to books, research papers, manuals, essays, legal contracts, whatever you have! The intelligence revolution is here, ChatGPT was just the beginning!

<https://www.chatpdf.com/>



3. Reference-based annotation of scRNA-Seq (SingleR)

3-1 Cell type reference datasets (celldex package)

(2019) "Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage."

Nat. Immunol., **20**, 163-172. doi:10.1038/s41590-018-0276-y.

▼ Human (5)

BlueprintEncodeData

Blueprint (Martens and Stunnenberg 2013) and Encode (The ENCODE Project Consortium 2012)

DatabaseImmuneCellExpressionData

The Database for Immune Cell Expression/(eQTLs/Epigenomics)(Schmiedel et al. 2018)

HumanPrimaryCellAtlasData

The Human Primary Cell Atlas (Mabbott et al. 2013)

MonacolImmuneData

Monaco Immune Cell Data - GSE107011 (Monaco et al. 2019)

NovershternHematopoieticData

Novershtern Hematopoietic Cell Data - GSE24759

▼ Mouse (2)

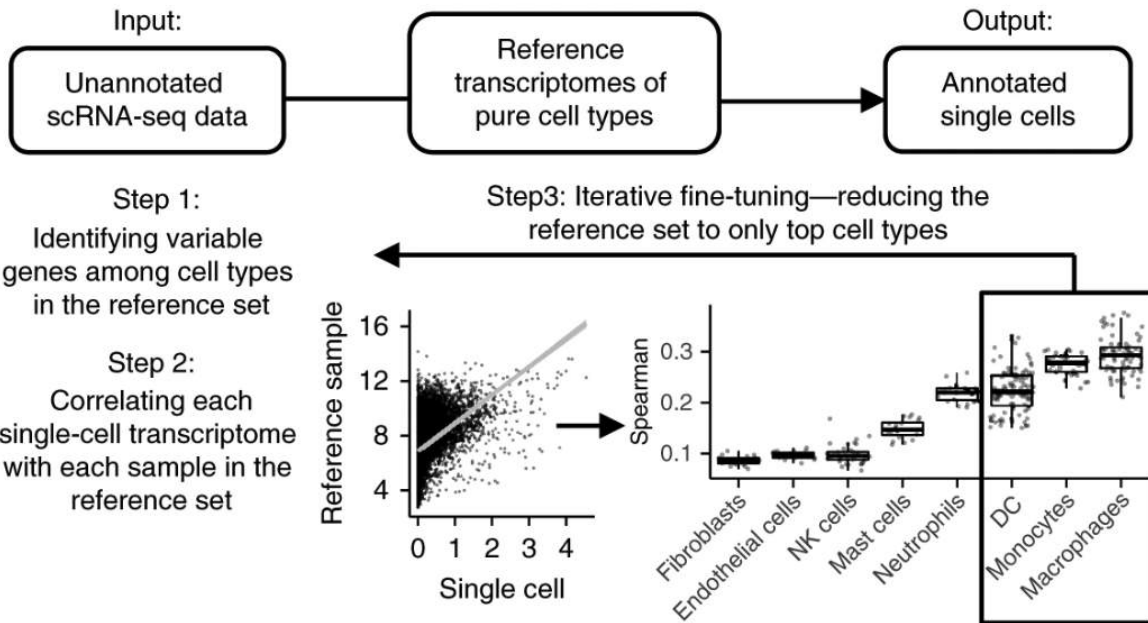
ImmGenData

the murine ImmGen (Heng et al. 2008)

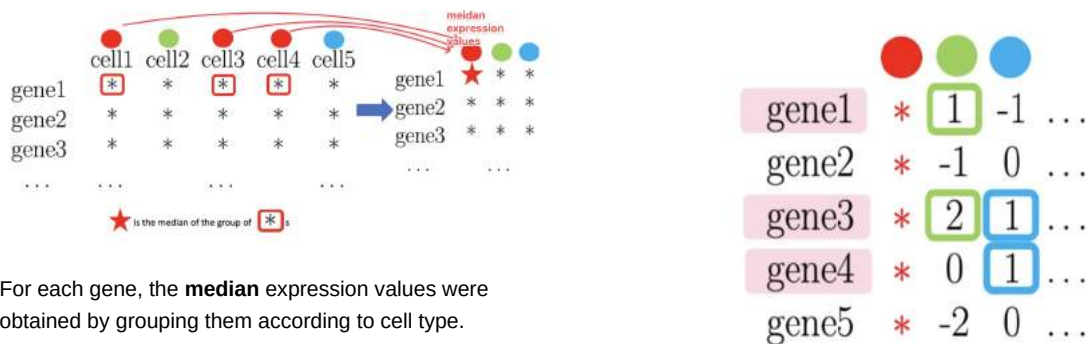
MouseRNAseqData

a collection of mouse data sets downloaded from GEO (Benayoun et al. 2019)

3-2 Schematic of SingleR



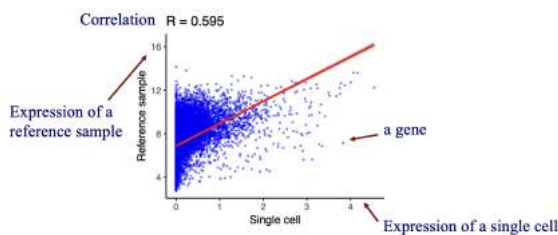
▼ Step1: Identifying variable genes

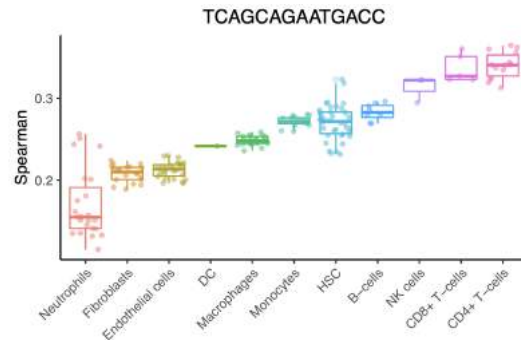


For each gene, the **median** expression values were obtained by grouping them according to cell type.

- Differential expression between each other cell type and the '**red**' cell type was calculated and all genes with positive differential expression values were selected
- The top *N* genes that showed the most difference in expression were chosen for the "red" cell type as variable genes.

▼ Step2: Correlation analysis

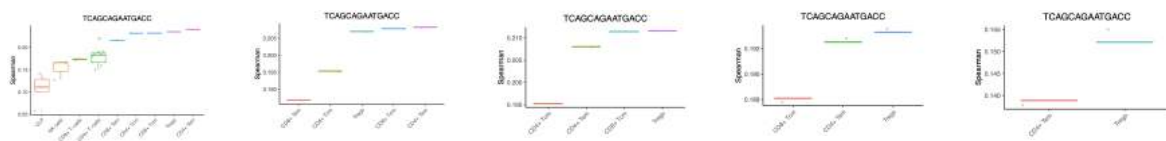




▼ Step3: Iterative fine-tuning (variable genes)

In this step *SingleR* reruns the correlation analysis, but only for the top cell types from step 2.

The analysis is performed only on variable genes between these cell types.



3-3 SingleR browser

singler - Default

<https://comphealth.ucsf.edu/app/singler>

▼ *SingleR* browser

SingleR Single-cell Recognition of cell types

Data sets Analysis Cluster Proportions Differential analysis

Important note: the *SingleR* browser has been upgraded. It now uses a new *SingleR* object. Please refer to <https://github.com/dviraran/SingleR> for more information.

To analyze single-cell data sets choose an item from the table and/or upload *SingleR* objects.

Import Data (SingleR object)

Browse... No file selected

If you would like to add a dataset to this list please contact dviraran@ucsf.edu.

Show 10 entries

Search:

	Set.Name	Organism	Citation	Technology	SingleR.Refs	N.cells	Title	PMID
1	GSE74923	Mouse	Kimmerling et al 2016	C1	Immgen, Mouse-RNAseq	189	A microfluidic platform enabling single cell RNA-seq of multigenerational lineages	26732280
2	GSE78779	Mouse	Hashimshony et al. 2016	CEL-Seq2	Immgen, Mouse-RNAseq	188	CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq	27121950
3	GSE48968	Mouse	Shalek et al. 2014	Drop-seq	Immgen, Mouse-RNAseq	2311	Single-cell RNA-seq reveals dynamic paracrine control of cellular variation	24919153
4	10x (Zheng) - 2000cells	Human	Zheng et al. 2017	10X	HPCA, Blueprint_Encode	4099	Massively parallel digital transcriptional profiling of single cells.	28091601

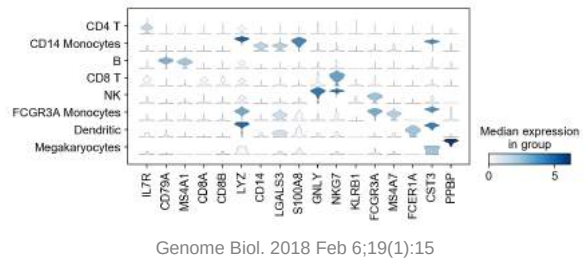
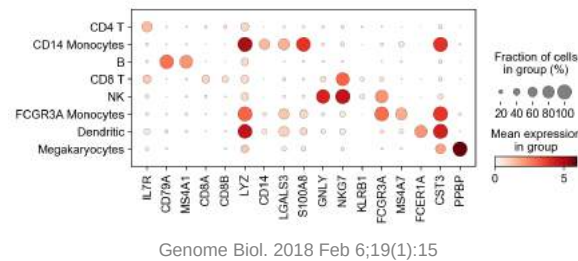
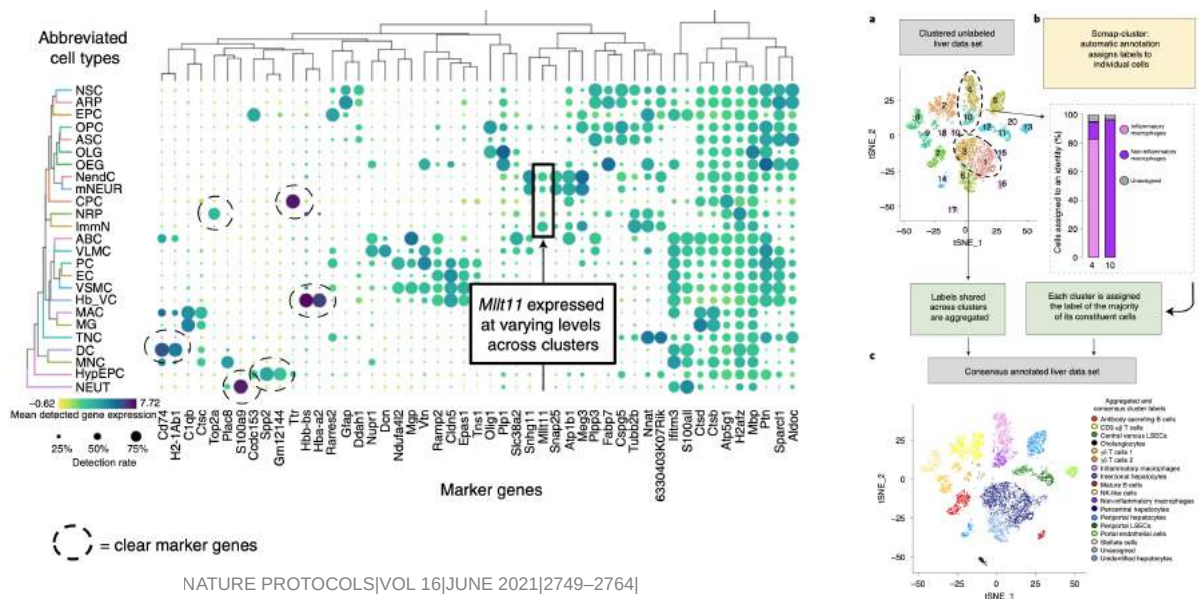
```
library(SingleR)
```

```
# Simplest use is running the wrapper function that creates both a SingleR and Seurat object:
```

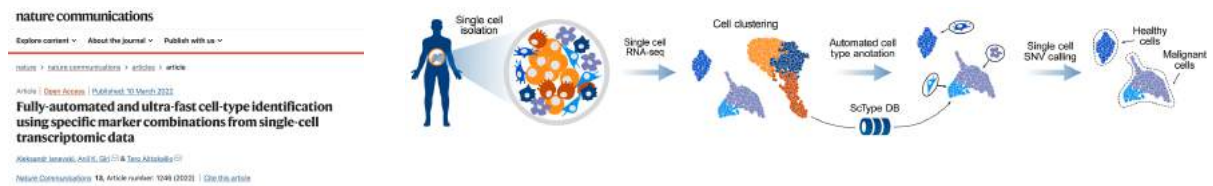
```
# counts file maybe a tab delimited text file, 10X directory or a matrix. annot is a tab delimited
# text file or a data.frame with the original identities. normalize.gene.length should be true if
# the data comes from a full-length platform. min.genes, min.cells, npca and regress.out are passed
# to Seurat to create a Seurat object:
singler = CreateSinglerSeuratObject(counts.file, annot, project.name,
  min.genes = 500, technology, species = "Human" (or "Mouse"), citation,
  normalize.gene.length = F, min.cells = 2, npca = 10
  regress.out = "nUMI", reduce.seurat.object = T)

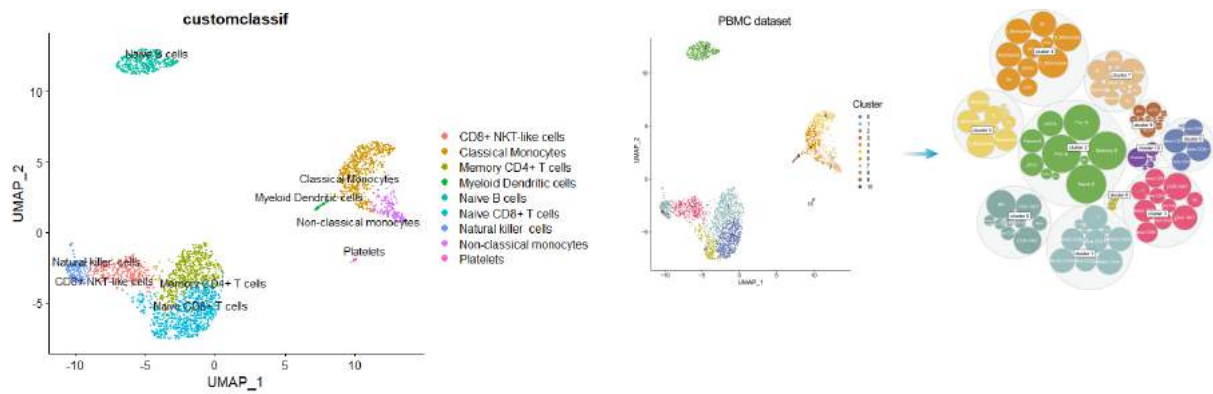
# The object can then be saved and uploaded to the SingleR web-app for further analysis and visualization or using functions avail
save(singler, file=paste0(project.name, '.RData')
```

4. Marker-based annotation of scRNA-Seq



4-1 ScType





[s41467-022-28803-w.pdf](#)

▼ ScType database

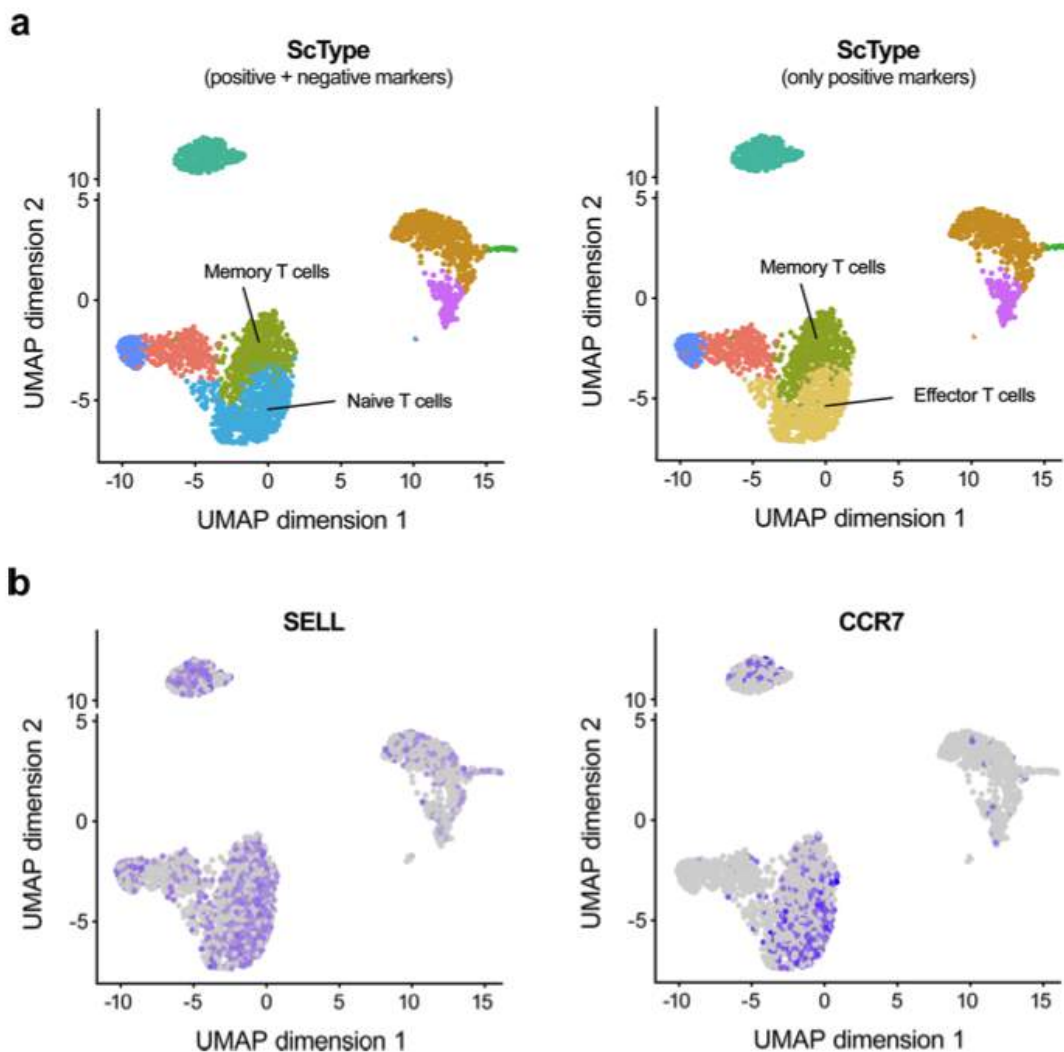
- CellMarker database
 - 13,605 cell markers for **467 cell types** in 158 human tissues/sub-tissues
 - 9,148 cell makers for **389 cell types** in 81 mouse tissues/sub-tissues
- PanglaoDB
 - 6,631 gene markers mapping to **155 cell type**
- Literature search
 - 37 negative markers
- User custom sets of positive and negative markers
 - domain knowledge
 - emerging studies

The widely applicable method is deployed both as **an interactive web-tool** (<https://sctype.app>), and as **an open-source R-package** (<https://github.com/lanevskiAleksandr/sc-type>).

sc-type
<https://sctype.app/database.php>

ScType utilizes both **positive** and **negative markers** for the cell type annotation

Naïve and memory T cells express **CCR7** and **SELL genes** for lymph node migration, while **effector T cells** do not.



Nat Commun. 2022 Mar 10;13(1):1246.

4-2 scCATCH

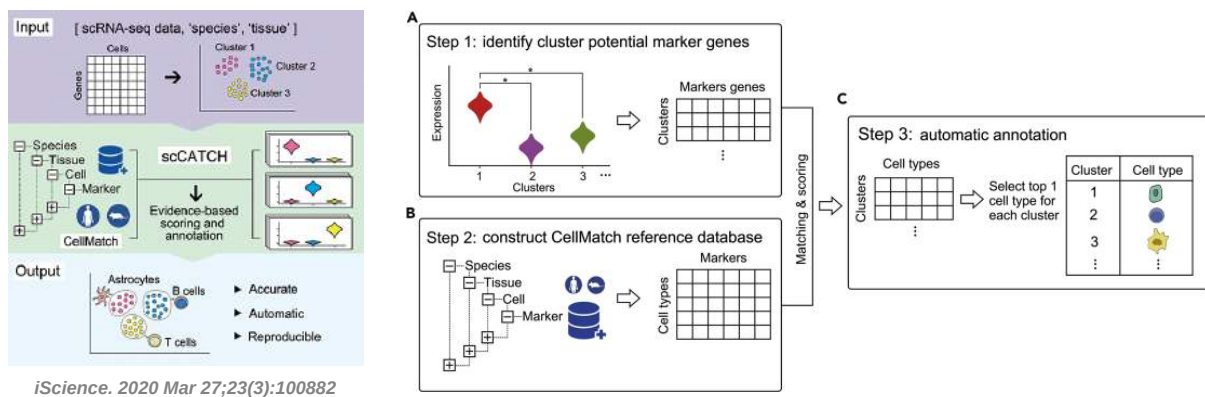
Accurate cell type identification is crucial for single-cell RNA sequencing studies, but current manual annotation methods can be time-consuming and subjective. The single cell Cluster-based Annotation Toolkit for Cellular Heterogeneity (scCATCH) offers a solution by automating the identification of cluster marker genes and annotation based on evidence-based scores and a tissue-specific cell taxonomy reference database (CellMatch).

CellMatch includes a panel of **353 cell types** and **related 686 subtypes** associated with **184 tissue types**, and **2,096 references of human and mouse**.

▼ CellMatch database

	species	tissue	cancer	condition	subtype1	subtype2	s
1	Human	Kidney	Normal	Normal cell	NA	NA	^
2	Human	Liver	Normal	Normal cell	NA	NA	^
3	Human	Endometrium	Normal	Normal cell	NA	NA	^
4	Human	Germ	Normal	Normal cell	Primordial	NA	^
5	Human	Corneal epithelium	Normal	Normal cell	NA	NA	^
6	Human	Placenta	Normal	Normal cell	NA	NA	^
7	Human	Periosteum	Normal	Normal cell	Periosteum-Derived	NA	^

	species	tissue	cancer	condition	subtype1	subtype2	s
8	Human	Periosteum	Normal	Normal cell	Periosteum-Derived	NA	^
9	Human	Periosteum	Normal	Normal cell	Periosteum-Derived	NA	^
10	Human	Periosteum	Normal	Normal cell	Periosteum-Derived	NA	^
11	Human	Amniotic membrane	Normal	Normal cell	Amnion	NA	^
12	Human	Amniotic membrane	Normal	Normal cell	Amnion	NA	^
13	Human	Primitive streak	Normal	Normal cell	Primitive	NA	^
14	Human	Primitive streak	Normal	Normal cell	Primitive	NA	^
15	Human	Adipose tissue	Normal	Normal cell	NA	NA	^
16	Human	Scalp	Normal	Normal cell	Bulge	NA	^
17	Human	Heart	Normal	Normal cell	NA	NA	^
18	Human	Liver	Normal	Normal cell	NA	NA	^
19	Human	Liver	Normal	Normal cell	NA	NA	^



```
obj <- findmarkergene(object = obj,
  species = "Human",
  marker = cellmatch,
  tissue = c("Blood", "Peripheral blood", "Plasma", "Serum", "Umbilical cord blood"))
```

```
obj <- findmarkergene(object = obj,
  species = "Human",
  marker = cellmatch,
  tissue = c("Blood", "Peripheral blood", "Serum", "Colon", "Colorectum", "Intestine"),
  cancer = c("Colon Cancer", "Colorectal Cancer"))
```

Allow users to select different combination of tissues or cancers for annotation

```
# Example
cellmatch_new <- cellmatch[cellmatch$species == "Mouse" & cellmatch$tissue %in% c("Kidney", "Liver", "Lung", "Brain"), ]
obj <- findmarkergene(object = obj, if_use_custom_marker = TRUE, marker = cellmatch_new)
obj <- findcelltype(obj)

# Example
cellmatch_new <- cellmatch[cellmatch$species == "Mouse" & cellmatch$cancer %in% c("Lung Cancer", "Lymph node", "Renal Cell Carcinoma"), ]
obj <- findmarkergene(object = obj, if_use_custom_marker = TRUE, marker = cellmatch_new)
obj <- findcelltype(obj)

# Example
cellmatch_new <- cellmatch[cellmatch$species == "Mouse", ]
cellmatch_new <- cellmatch[cellmatch$cancer %in% c("Lung Cancer", "Lymph node", "Renal Cell Carcinoma", "Prostate Cancer") | cellmatch$tissue %in% c("Blood", "Peripheral blood", "Plasma", "Serum", "Umbilical cord blood"), ]
obj <- findmarkergene(object = obj, if_use_custom_marker = TRUE, marker = cellmatch_new)
obj <- findcelltype(obj)
```

5. ShinySC

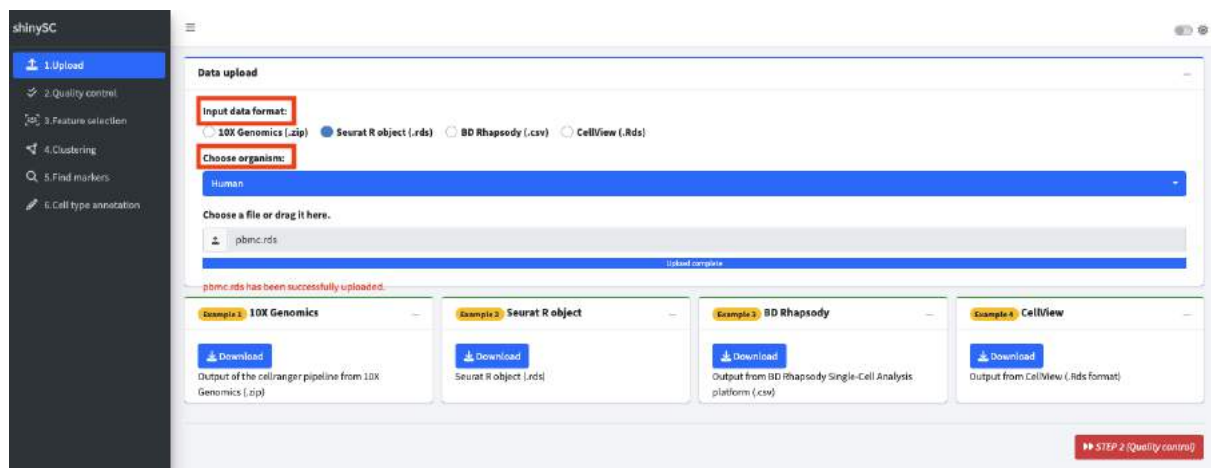
5-1 Upload

5-1-1 Input format:

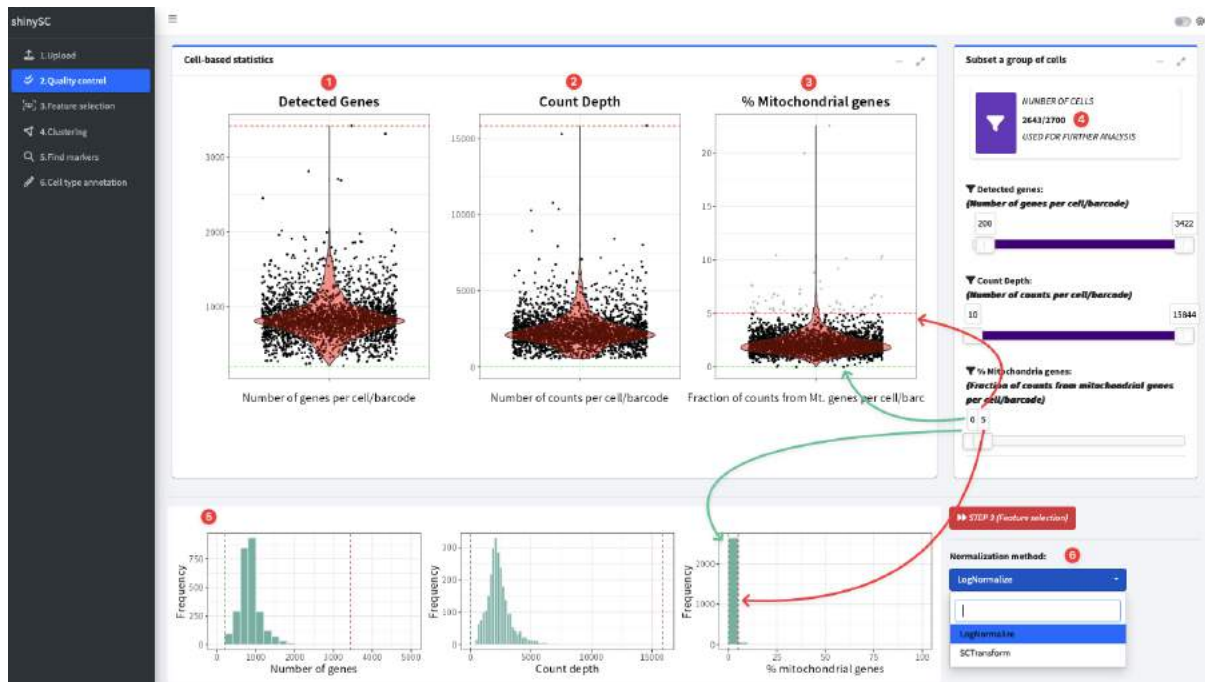
- 10X Genomics (.zip)
- Seurat R object (.rds)
- BD Rhapsody (.csv)
- CellView (.Rds)

5-1-2 Supported organisms:

- Human
- Mouse

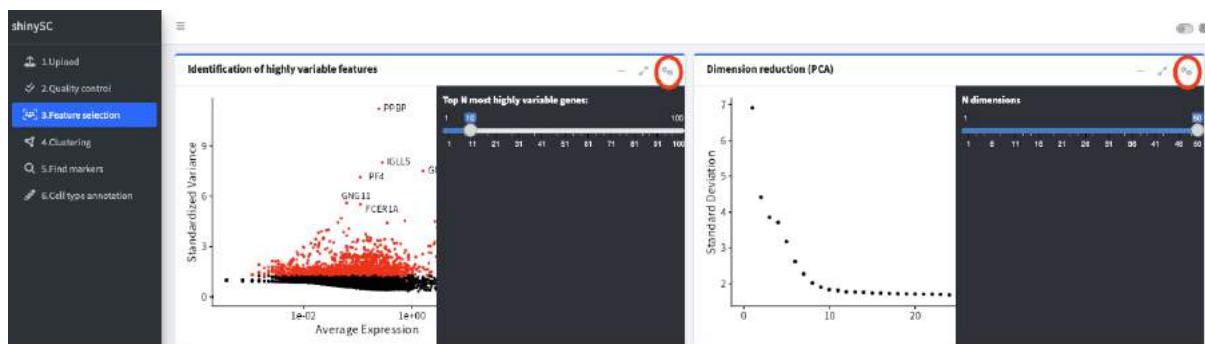
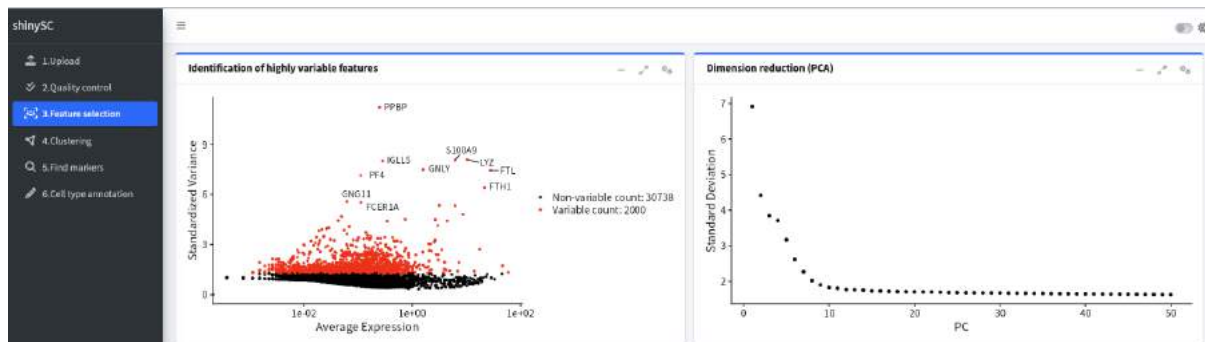


5-2 Quality control



5-3 Feature selection

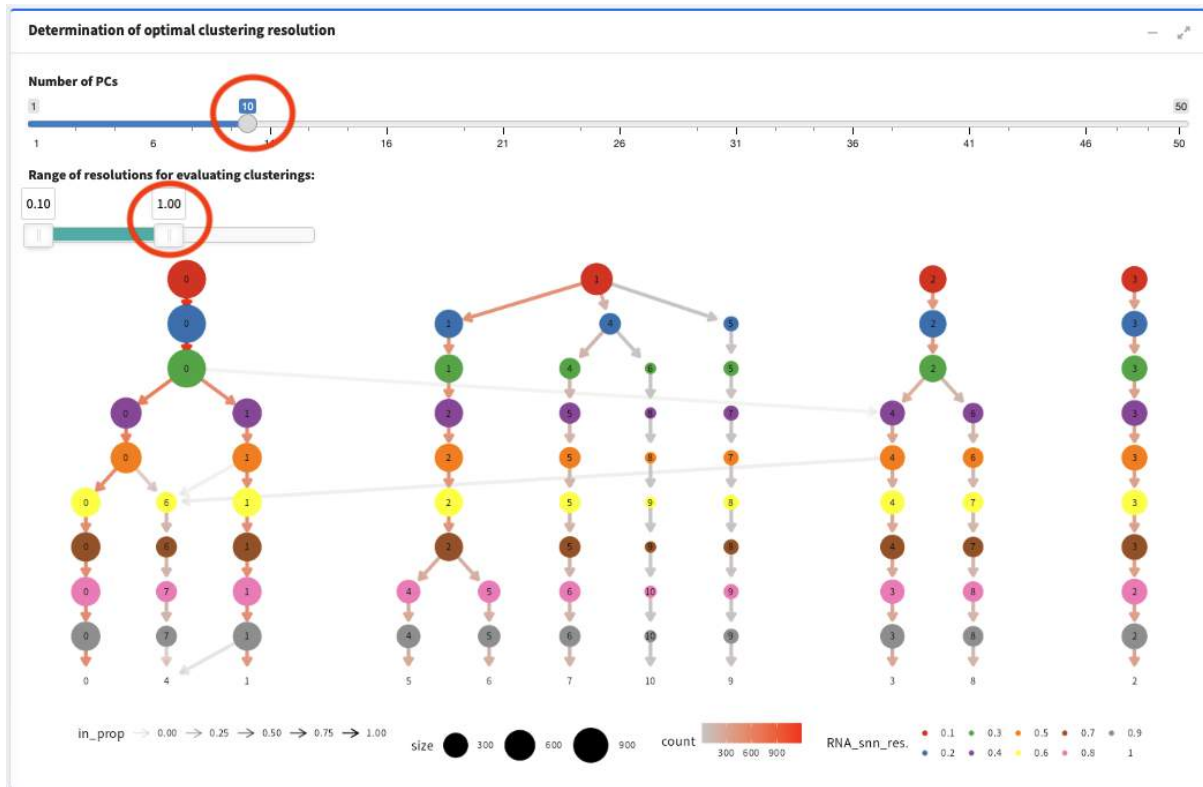
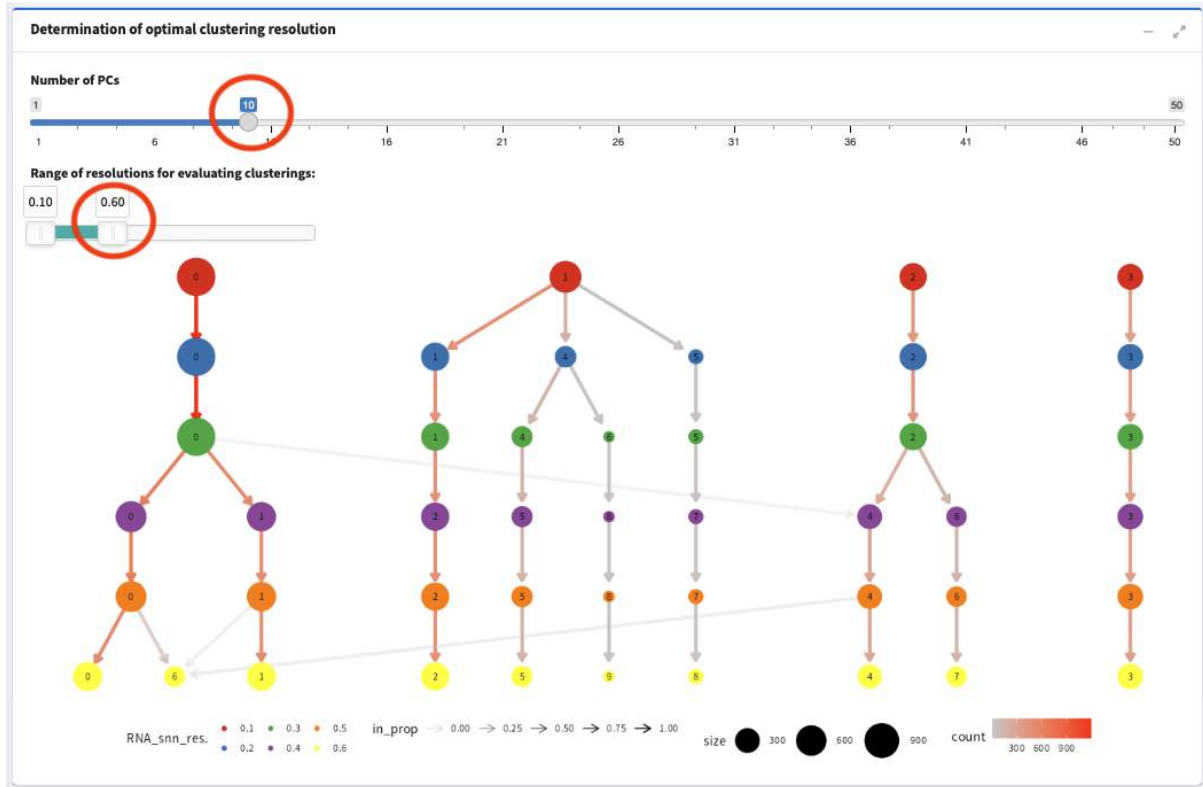
5-3-1 Identification of highly variable features & Dimension reduction (PCA)



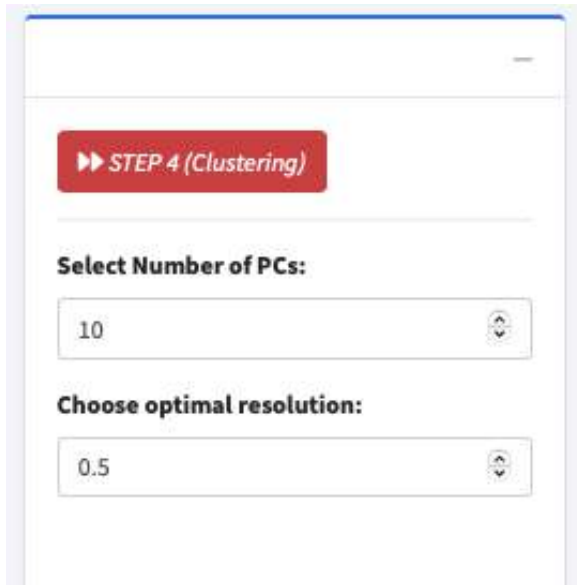
5-3-2 Determination of optimal clustering resolution

Tools and techniques for single-cell RNA sequencing data

<https://lazappi.github.io/phd-thesis/4-clust-trees.html>



5-3-2 Define optimal resolution

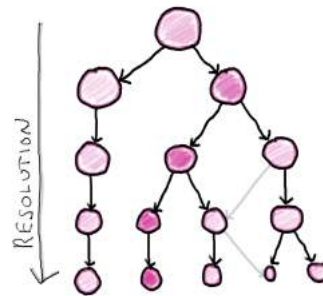
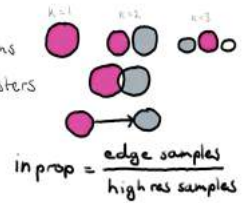


4. CLUSTERING TREES

How many clusters to use?

A tree of clusters!

1. Cluster at multiple resolutions
2. Calculate overlap between clusters
3. Build a graph
4. Weight edges
5. Visualise

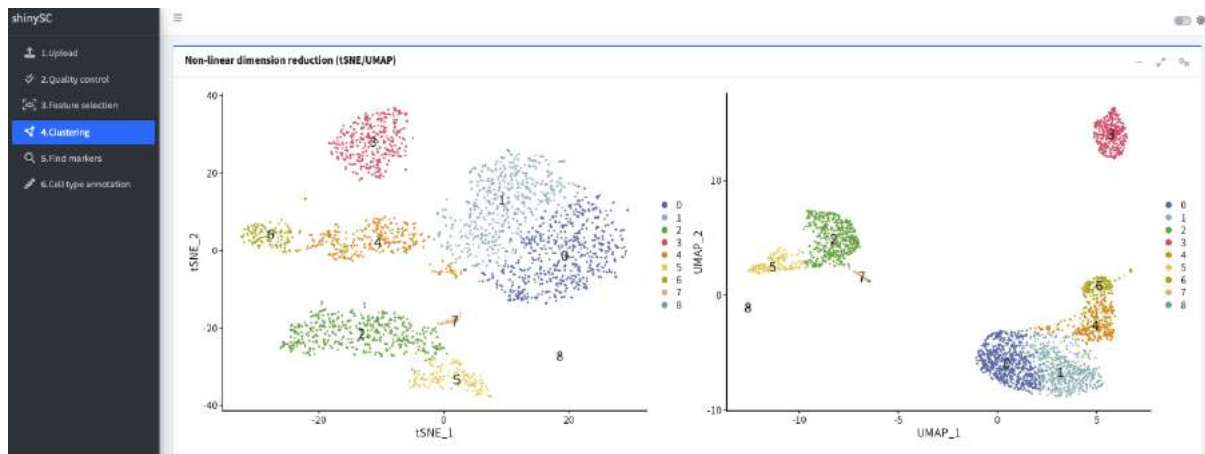


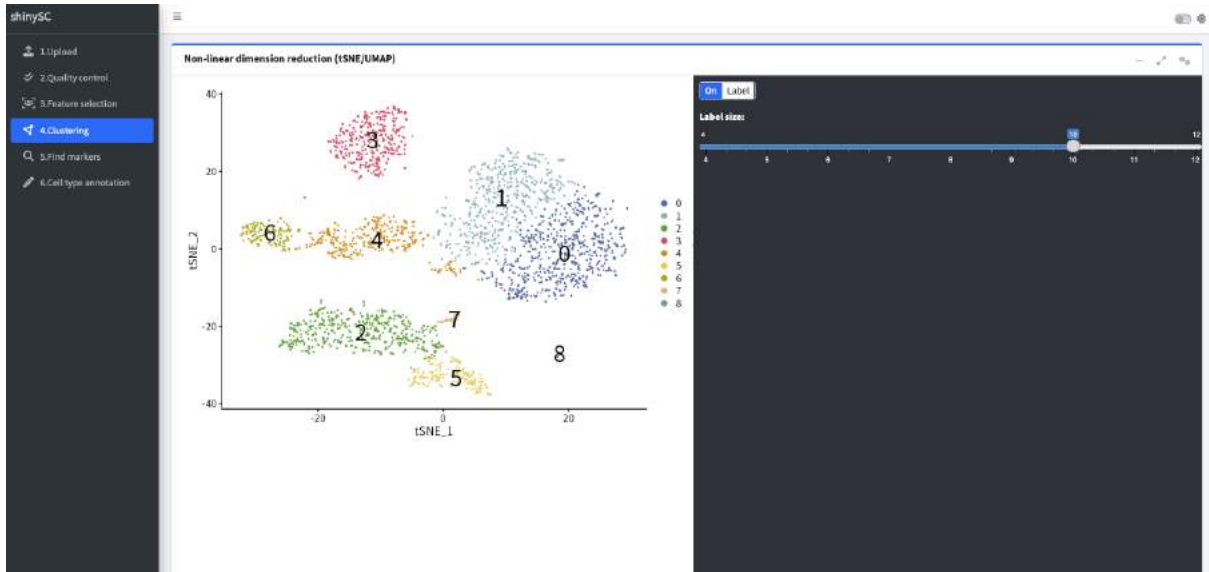
Structure can show stability

Show information across resolutions

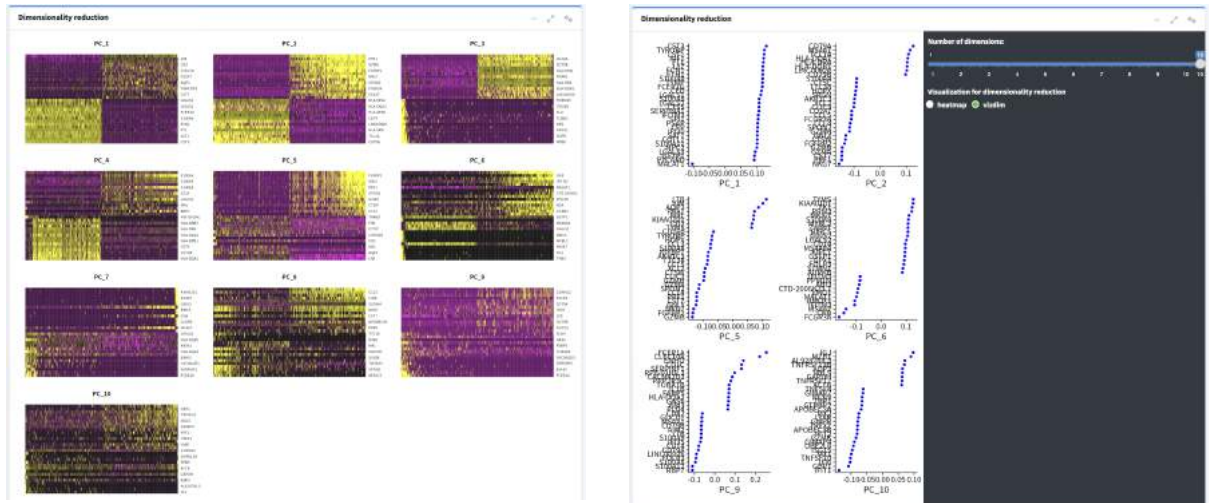
5-4. Clustering

5-4-1 Non-linear dimension reduction





5-4-2 Heatmap & Vizdim



5-5 Find markers

▼ Differential expression test

Wilcoxon Rank Sum test
▼

Wilcoxon Rank Sum test
▼

Student's t-test
▼

▼ Cluster

Cluster:

0

0

1

2

3

4

5

6

7

8

▼ Gene description

Biomart: Human (ok), Mouse (待確認Gene name格式)

New Count Results URL XML Perl Help

Export all results to File TSV Unique results only Go

Email notification to

View 200 rows as HTML Unique results only

Gene name	Gene description	GeneCards ID
MT-1F	mitochondrially encoded tRNA-Phe (UUU/C) [Source:HGNC Symbol;Acc:HGNC:7461]	7461
MT-RNR1	mitochondrially encoded 12S rRNA [Source:HGNC Symbol;Acc:HGNC:7470]	7470
MT-TV	mitochondrially encoded tRNA-Val (GUN) [Source:HGNC Symbol;Acc:HGNC:7500]	7500
MT-RNR2	mitochondrially encoded 16S rRNA [Source:HGNC Symbol;Acc:HGNC:7471]	7471
MT-TL1	mitochondrially encoded tRNA-Leu (UUA/G) 1 [Source:HGNC Symbol;Acc:HGNC:7490]	7490
MT-ND1	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1 [Source:HGNC Symbol;Acc:HGNC:7455]	7455
MT-TI	mitochondrially encoded tRNA-Ile (AAU/C) [Source:HGNC Symbol;Acc:HGNC:7488]	7488
MT-TQ	mitochondrially encoded tRNA-Gln (CAA/G) [Source:HGNC Symbol;Acc:HGNC:7495]	7495
MT-TM	mitochondrially encoded tRNA-Met (AUA/G) [Source:HGNC Symbol;Acc:HGNC:7492]	7492
MT-ND2	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2 [Source:HGNC Symbol;Acc:HGNC:7456]	7456
MT-TW	mitochondrially encoded tRNA-Trp (UGA/G) [Source:HGNC Symbol;Acc:HGNC:7501]	7501
MT-TA	mitochondrially encoded tRNA-Ala (GCN) [Source:HGNC Symbol;Acc:HGNC:7475]	7475
MT-TN	mitochondrially encoded tRNA-Aan (AAU/C) [Source:HGNC Symbol;Acc:HGNC:7493]	7493
MT-TC	mitochondrially encoded tRNA-Cys (UGU/C) [Source:HGNC Symbol;Acc:HGNC:7477]	7477
MT-TY	mitochondrially encoded tRNA-Tyr (UAU/C) [Source:HGNC Symbol;Acc:HGNC:7502]	7502
MT-CO1	mitochondrially encoded cytochrome c oxidase I [Source:HGNC Symbol;Acc:HGNC:7419]	7419
MT-TS1	mitochondrially encoded tRNA-Ser (UCN) 1 [Source:HGNC Symbol;Acc:HGNC:7497]	7497

▼ GeneCards

(GeneCardsID → GeneCards web site 連結待建立)

VDAC1P4 Gene - Voltage Dependent Anion Channel 1 Pseudogene 4

Pseudogene (Updated: Mar 21, 2023 ; GC01P180434 ; GIFIS: 10)

Jump to section	Aliases Paralogs	Disorders Pathways	Domains Products	Drugs Proteins	Expression Publications	Function Sources	Genomics Summaries	Localization Transcripts	Orthologs Variants
Research Products	Antibodies Primers	Assays	Proteins	Inhib. RNA	CRISPR	miRNA	Drugs	Cell Lines	Clones

Proteins Primary Antibodies
ELISAs Antibody Arrays
Activity Assays

Online Vector Design Platform
Virus Packaging (AAV/Lentil)
CDR3 Library Construction

Proteins Antibodies Assays
Genes shRNA Primers
CRISPR Lentiviral Particles

CRISPR Knockout Kit sgRNA
Engineered Cells
Edited iPSCs

Aliases for VDAC1P4 Gene

Aliases for VDAC1P4 Gene

GeneCards Symbol: **VDAC1P4** ² ⁵

Voltage Dependent Anion Channel 1 Pseudogene 4 ² ³ ⁵

Voltage-Dependent Anion Channel 4 Pseudogene ² ³

VDAC4P ³ ⁵

VDAC4 ³ ⁵

Voltage-Dependent Anion Channel 4 ²

External ids for VDAC1P4 Gene

HGNC: 12675 NCBI Entrez Gene: 7418 Ensembl: ENSG00000235060 OMIM®: 610030

Previous HGNC Symbols for VDAC1P4 Gene

VDAC4, VDAC4P

Previous GeneCards identifiers for VDAC1P4 Gene

GC01U902394, GC01P180404, GC01P151634

Search aliases for VDAC1P4 gene in PubMed and other databases

GeneAlaCart
GENECARDS BATCH QUERIES

> 190 Integrated Biomedical Sources

API, JSON, CSV, EXCEL

shinySC

Differential expression test: **1** Cluster: **2** avg_logFC >= p_val_adj <

Wilcoxon Rank Sum test 0

Cluster-specific gene marker(s)

Copy CSV Excel PDF Print Search:

gene	cluster	avg_log2FC	pct.1	pct.2	p_val_adj	p_val	description	GeneCardsID
CCR7	0	1.389	0.468	0.111	1.2805e-02700000e-07	4.29207278967883e-92	C-C motif chemokine receptor 7 [Source:HGNC Symbol;Acc:HGNC:1608]	1608
LDLRAP1	0	1.176	0.266	0.084	1.0525590972489e-30	3.21508922269195e-35	low density lipoprotein receptor adaptor protein 1 [Source:HGNC Symbol;Acc:HGNC:18640]	18640
LEF1	0	1.092	0.358	0.105	1.74637802299883e-49	5.33440657034282e-54	lymphoid enhancer binding factor 1 [Source:HGNC Symbol;Acc:HGNC:6551]	6551
PRKQC-AS1	0	1.082	0.356	0.109	2.26908277300734e-46	6.93103663329264e-51	PRKQC antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44689]	44689
MAL	0	1.020	0.29	0.085	7.83275525166572e-37	2.39255765522198e-41	mal, T cell differentiation protein [Source:HGNC Symbol;Acc:HGNC:6817]	6817
LDHB	0	1.020	0.924	0.598	2.38977511626939e-100	7.29969795427147e-105	lactate dehydrogenase B [Source:HGNC Symbol;Acc:HGNC:6541]	6541
PIK3IP1	0	1.005	0.467	0.183	2.14836457136785e-45	6.56229632649476e-50	phosphoinositide-3-kinase interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:24942]	24942
NOSIP	0	0.975	0.657	0.358	2.21001609005487e-47	6.75061424049959e-52	nitric oxide synthase interacting protein [Source:HGNC Symbol;Acc:HGNC:17946]	17946
CD3D	0	0.902	0.85	0.417	5.17941746277998e-66	1.5817575379558e-70	CD3 delta subunit of T-cell receptor complex [Source:HGNC Symbol;Acc:HGNC:1673]	1673
CD3E	0	0.844	0.736	0.408	1.61002580496134e-45	4.91791131089665e-50	CD3 epsilon subunit of T-cell receptor complex [Source:HGNC Symbol;Acc:HGNC:1674]	1674

Showing 1 to 10 of 166 entries Previous 2 3 4 5 ... 17 Next

💡 Identification of Cluster-specific gene markers

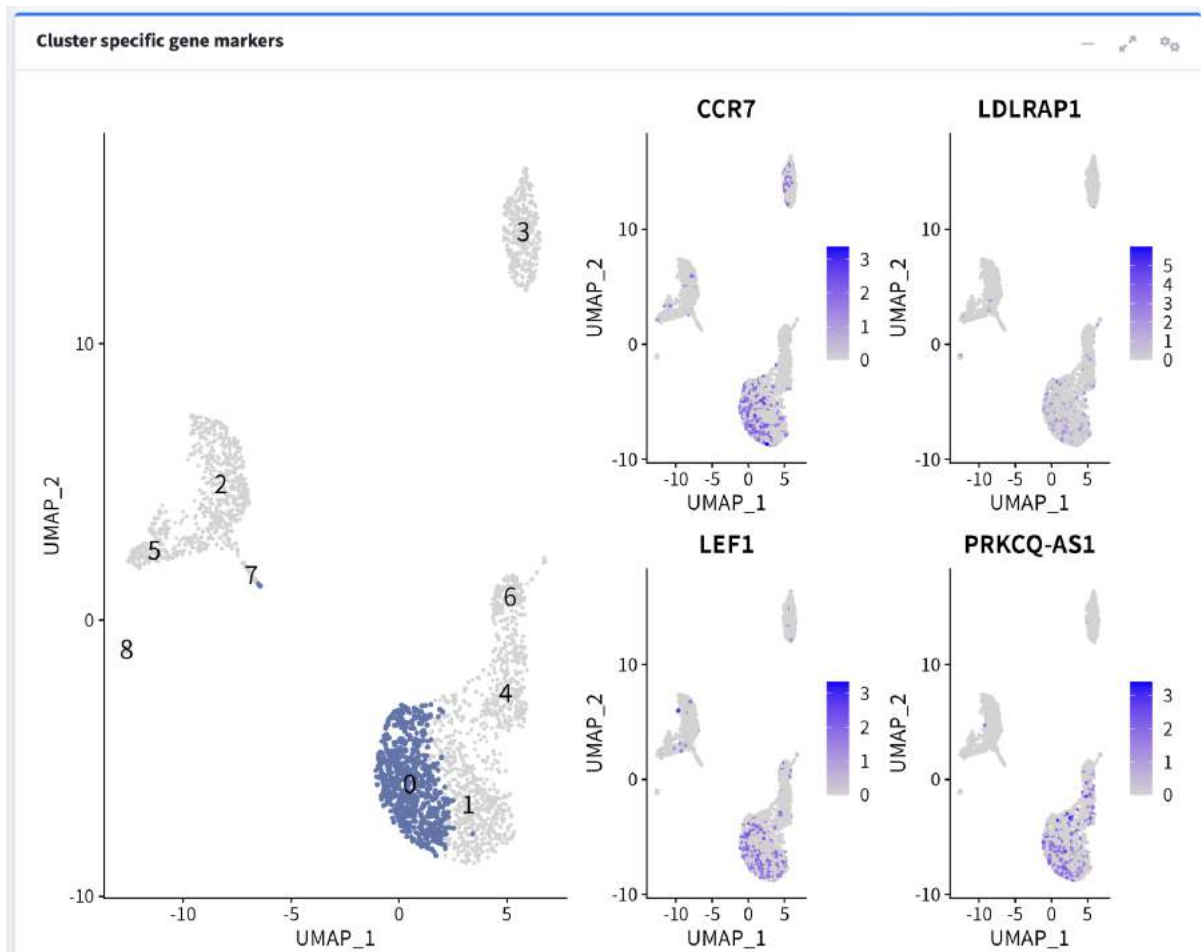
Cluster-specific gene marker(s)

Copy CSV Excel PDF Print Search:

gene	cluster	avg_log2FC	pct.1	pct.2	p_val_adj	p_val	description
CCR7	0	1.389	0.468	0.111	1.402128799888e-41	4.29207278967683e-92	C-C motif chemokine receptor 7 [Source:HGNC Symbol;Acc:HGNC:1608]
LDLRAP1	0	1.176	0.266	0.084	1.851333087388e-30	3.21508922269195e-35	low density lipoprotein receptor adaptor protein 1 [Source:HGNC Symbol;Acc:HGNC:18640]
LEF1	0	1.092	0.358	0.105	1.7481746222988e-40	5.33440657034282e-54	lymphoid enhancer binding factor 1 [Source:HGNC Symbol;Acc:HGNC:6551]
PRKCQ-AS1	0	1.082	0.356	0.109	1.28888377320373e-40	6.93103663329264e-51	PRKCQ antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44689]
MAL	0	1.020	0.29	0.085	7.83275525166572e-37	2.39255765522198e-41	mal, T cell differentiation protein [Source:HGNC Symbol;Acc:HGNC:6817]
LDHB	0	1.020	0.924	0.598	2.38977511626939e-100	7.29969795427147e-105	lactate dehydrogenase B [Source:HGNC Symbol;Acc:HGNC:6541]
PIK3IP1	0	1.005	0.467	0.183	2.14836457136785e-45	6.56229632649476e-50	phosphoinositide-3-kinase interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:24942]
NOSIP	0	0.975	0.657	0.358	2.21001609005487e-47	6.75061424049995e-52	nitric oxide synthase interacting protein [Source:HGNC Symbol;Acc:HGNC:17946]
CD3D	0	0.902	0.85	0.417	5.17841746277598e-66	1.58177575379558e-70	CD3 delta subunit of T-cell receptor complex [Source:HGNC Symbol;Acc:HGNC:1673]
CD3E	0	0.844	0.736	0.408	1.61002580496134e-45	4.91791131089665e-50	CD3 epsilon subunit of T-cell receptor complex [Source:HGNC Symbol;Acc:HGNC:1674]

Showing 1 to 10 of 166 entries

Previous 1 2 3 4 5 ... 17 Next



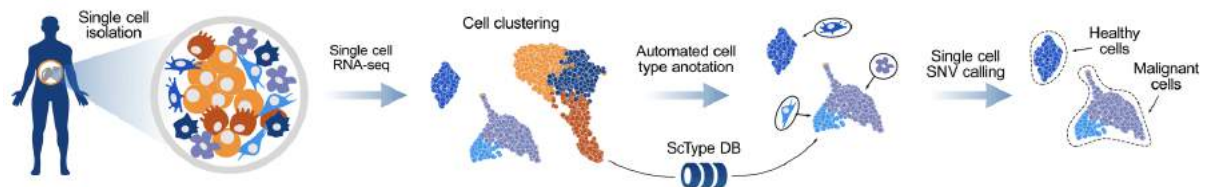
5-6. Automated cell-type annotation

5-6-1 ScType

Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data

Aleksandr Ianevski, Anil K. Giri & Tero Aittokallio

Nature Communications 13, Article number: 1246 (2022) | [Cite this article](#)



[s41467-022-28803-w.pdf](#)

The widely applicable method is deployed both as **an interactive web-tool** (<https://sctype.app>), and as **an open-source R-package**.

<https://github.com/ianeviskiAleksandr/sc-type>

The ScType tool is currently undergoing an update and is expected to be back up and running within the next few days!

Please use our R-based implementation in the meanwhile:
<https://github.com/ianeviskiAleksandr/sc-type>

shinySC

1. Upload
2. Quality control
3. Feature selection
4. Clustering
5. Find markers
6. Cell type annotation

Cell-type identification methods

ScType SingleR scCATCH

The tissue type your data belongs to:

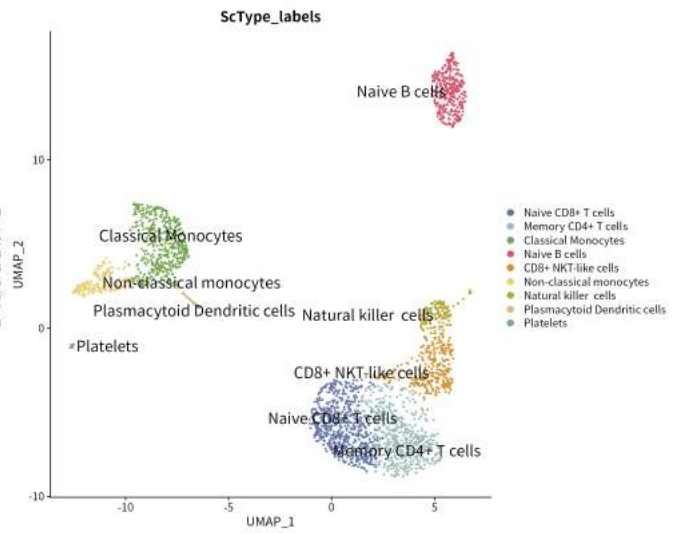
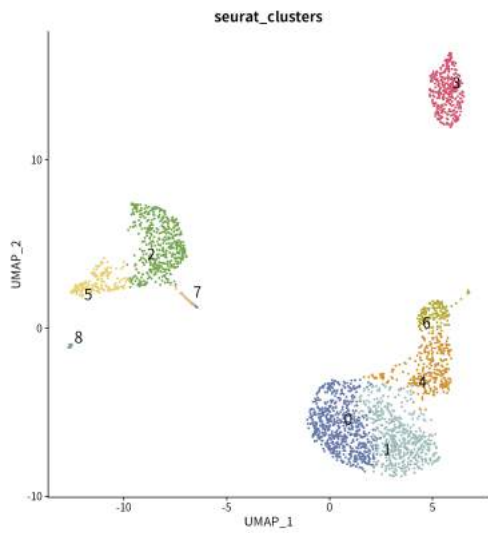
Immune system

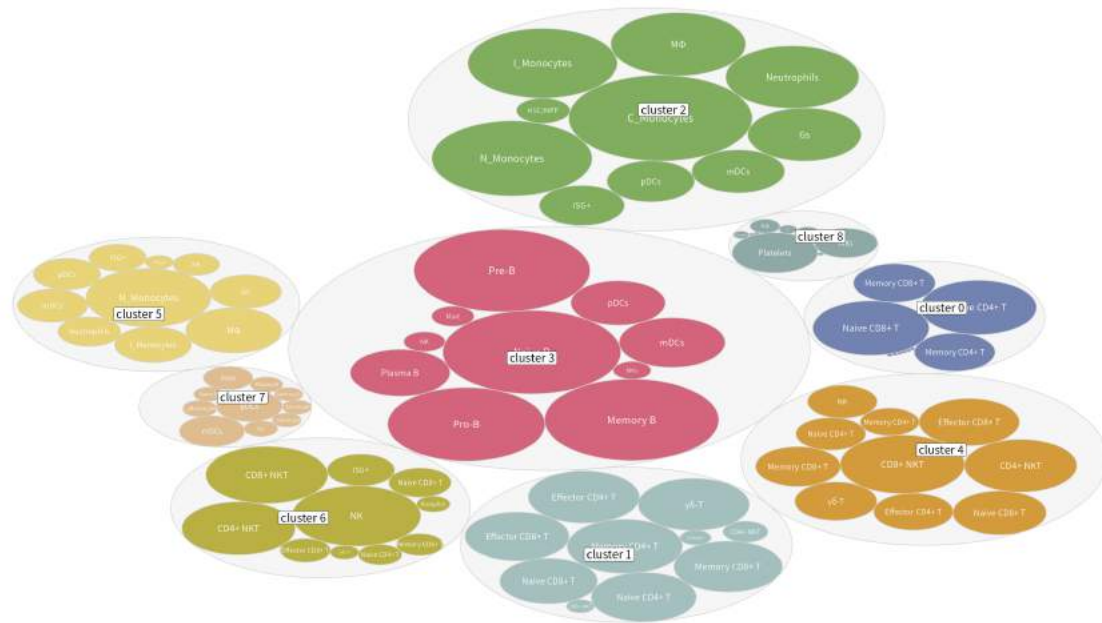
Run

Immune system

- Pancreas
- Liver
- Eye
- Kidney
- Brain
- Lung
- Adrenal
- Heart
- Intestine
- Muscle
- Placenta
- Spleen
- Stomach
- Thymus

15





5-6-2 SingleR

Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage

[Dvir Aran](#), [Agnieszka P. Looney](#), [Leqian Liu](#), [Esther Wu](#), [Valerie Fong](#), [Austin Hsu](#), [Suzanna Chak](#), [Ram P. Naikawadi](#), [Paul J. Wolters](#), [Adam R. Abate](#), [Atul J. Butte](#) & [Mallar Bhattacharya](#) ✉

Nature Immunology **20**, 163–172 (2019) | [Cite this article](#)

[s41590-018-0276-y.pdf](#)

SingleR

Performs unbiased cell type recognition from single-cell RNA sequencing data, by leveraging reference transcriptomic datasets of pure cell types to infer the cell of origin of each single cell independently.

<https://bioconductor.org/packages/release/bioc/html/SingleR.html>



singler - Default

<https://comphealth.ucsf.edu/app/singler>

▼ **singler web application**

SingleR Single-cell Recognition of cell types

Data sets Analysis Cluster Proportions Differential analysis

Important note: the SingleR browser has been upgraded. It now uses a new SingleR object. Please refer to <https://github.com/dviraran/SingleR> for more information.

To analyze single-cell data sets choose an item from the table and/or upload SingleR objects.

Import Data (SingleR object)

Browse... No file selected

If you would like to add a dataset to this list please contact dvir.aran@ucsf.edu.

Show 10 entries

Search:

Set.Name	Organism	Citation	Technology	SingleR.Refs	N.cells	Title	PMID
1 GSE74923	Mouse	Kimmerling et al. 2016	C1	Immgen, Mouse-RNAseq	189	A microfluidic platform enabling single cell RNA-seq of multigenerational lineages	26732280
2 GSE78779	Mouse	Hashimshony et al. 2016	CEL-Seq2	Immgen, Mouse-RNAseq	188	CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq	27121950
3 GSE48968	Mouse	Shalek et al. 2014	Drop-seq	Immgen, Mouse-RNAseq	2311	Single-cell RNA-seq reveals dynamic paracrine control of cellular variation	24919153
4 10x (Zheng) - 2000cells	Human	Zheng et al. 2017	10X	HPCA, Blueprint_Encode	4099	Massively parallel digital transcriptional profiling of single cells.	28091801
5 GSE111664	Mouse	Aran et al.	Drop-seq	Immgen, GSE94135+GSE49932, Mouse-RNAseq	8405	Single-cell RNA-seq reveals profibrotic macrophages in lung fibrosis	NA
6 PBMC (healthy)	Human	10X datasets	10X	HPCA, Blueprint_Encode	8391	Massively parallel digital transcriptional profiling of single cells.	28091601
7 TabulaMuris-FACS	Mouse	Tabula Muris	Smart-Seq2	Immgen, Mouse-RNAseq	44949	Single-cell transcriptomics of 20 mouse organs creates aTabula Muris	NA
8 GSE52529	Human	Trapnell et al. 2014	C1	HPCA, Blueprint_Encode	379	Pseudo-temporal ordering of individual cells reveals regulators of differentiation	24658544
9 GSE54006	Mouse	Jaitin et al. 2014	MARS-Seq	Immgen, Mouse-RNAseq	1864	Massively parallel single-cell RNA-Seq for dissecting cell type and cell state compositions	24531970
10 GSE57872	Human	Patel et al. 2014	Smart-Seq	HPCA, Blueprint_Encode	96	Single cell RNA-seq of primary human glioblastomas	24825914

Showing 1 to 10 of 120 entries

Previous 1 2 3 4 5 ... 12 Next



Reference Index for cell type

```
biocmanager::install("celldex")
```

- 建立local databases

```
ref = celldex::DatabaseImmuneCellExpressionData()
saveRDS(ref, "DatabaseImmuneCellExpressionData.rds")
```

1. Upload

2. Quality control

3. Feature selection

4. Clustering

5. Find markers

6. Cell type annotation

Cell-type identification methods

ScType SingleR scCATCH

Reference dataset:

(Human) NovershternHematopoieticData

(Human) NovershternHematopoieticData: immune Murine datasets for sorted hematopoietic cell populations (Novershtern et al. 2015)

(Human) DatabaseImmuneCellExpressionData: Immune Bulk RNA-seq datasets from the DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) project

(Human) MonacImmuneData: Immune Bulk RNA-seq samples of sorted immune cell populations-a0 (Monaco et al. 2019)

(Human) HumanPrimaryCellAtlasData: General Publicly available microarray datasets derived from human primary cells (Mabbott et al. 2013)

(Human) BlueprintEncodeData: General Bulk RNA-seq data for pure stroma and immune cells generated by Blueprint-a0 and ENCODE projects.

☰

Cell-type identification methods

ScType
 SingleR
 scCATCH

Reference dataset:

(Mouse) MouseRNAseqData

(Mouse) MouseRNAseqData General: Bulk RNA-seq data sets from the gene expression omnibus (Benayoun et al. 2019)

(Mouse) ImmGenData Immune: Microarray profiles of pure mouse immune cells from Immunological Genome Project (ImmGen)

shinySC ☰

1. Upload
2. Quality control
3. Feature selection
4. Clustering
5. Find markers
6. Cell type annotation

Cell-type identification methods

ScType
 SingleR
 scCATCH

Reference dataset:

(Human) Novershtern/HematopoieticData

▶▶ Run

Cell type annotation

SingleR_labels

UMAP_2

UMAP_1

- B cells
- Basophils
- CD4+ T cells
- CD8+ T cells
- Dendritic cells
- Granulocytes
- HSCs
- Megakaryocytes
- Monocytes
- NK cells
- NK T cells

Cell-type identification methods

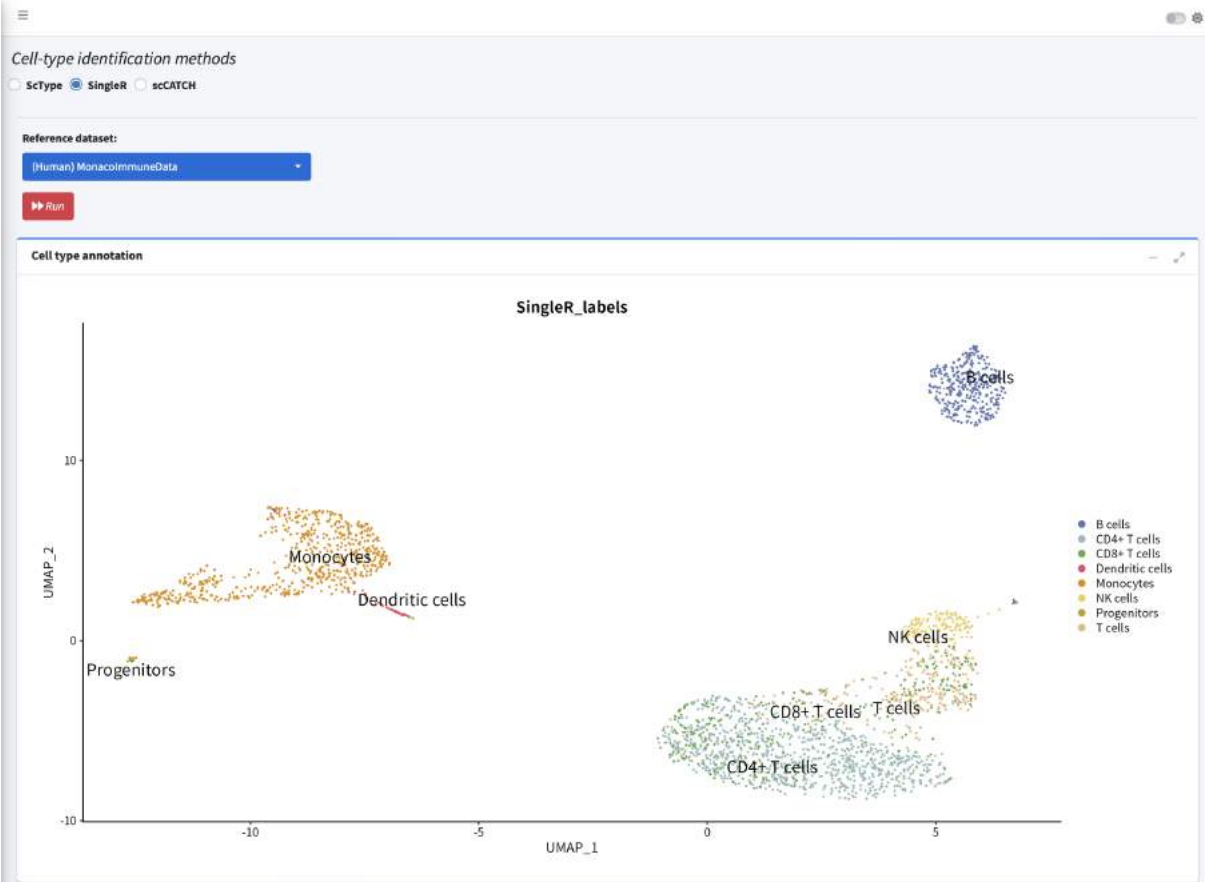
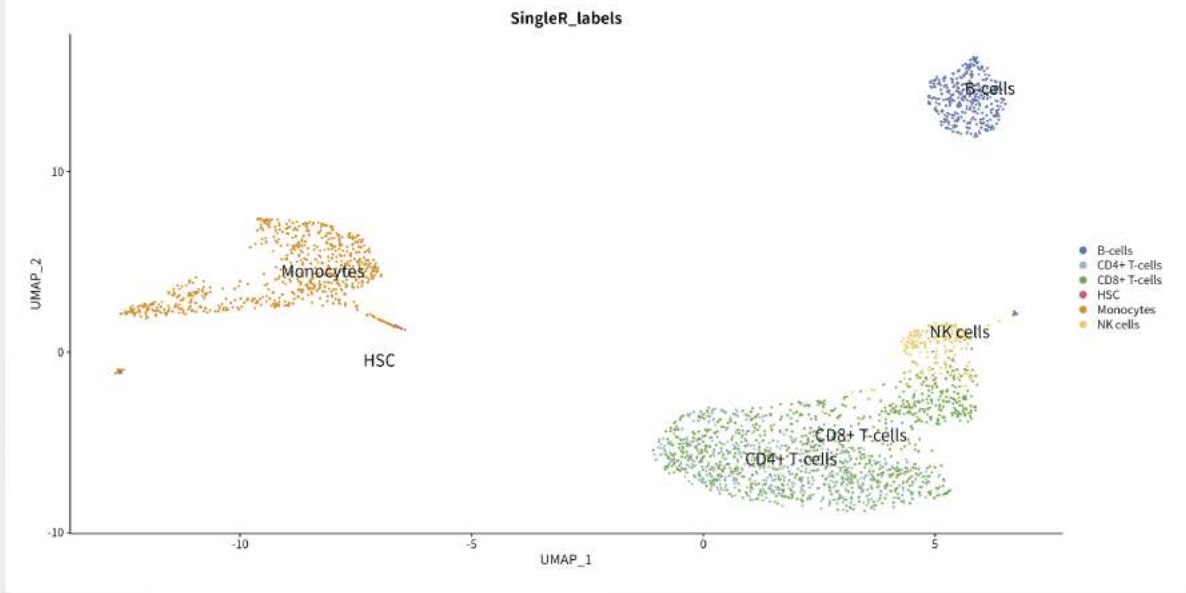
ScType SingleR scCATCH

Reference dataset:

(Human) BlueprintEncodeData

Run

Cell type annotation





5-6-3

scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data

Xin Shao ¹, Jie Liao ¹, Xiaoyan Lu ¹, Rui Xue ¹, Ni Ai ¹, Xiaohui Fan ²

Affiliations + expand

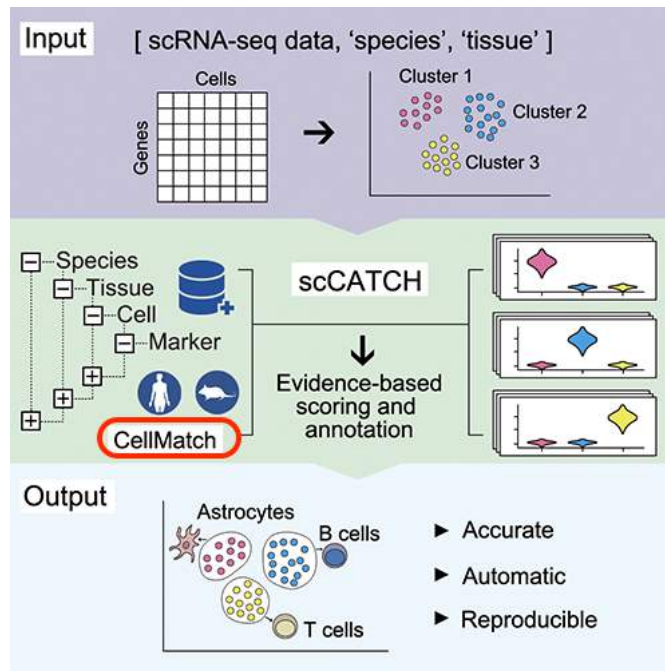
PMID: 32062421 PMCID: PMC7031312 DOI: 10.1016/j.isci.2020.100882

[main.pdf](#)

<https://github.com/ZJUFanLab/scCATCH>

Knowledge in **CellMatch reference database** was derived from various resources, such as **CellMarker** (Zhang et al., 2019b), **MCA** (Han et al., 2018), **CancerSEA** (Yuan et al., 2019), and the **CD Marker Handbook**.

a panel of 353 cell types and related 686 subtypes associated with **184 tissue types**, **20,792 cell-specific marker genes**, and **2,097 references of humans and mice** were introduced into scCATCH as the reference database.



- Cascade filters for CellMatch reference database

species	cancer	tissue	celltype	gene	pmid	resource	condition	subtype1	subtype2	subtype3
153	Human	Normal	Blood	T Cell	CD3D	26453327	Experiment	Normal cell	Angiogenic	
154	Human	Normal	Blood	T Cell	CD3E	26453327	Experiment	Normal cell	Angiogenic	
155	Human	Normal	Blood	T Cell	CD3G	26453327	Experiment	Normal cell	Angiogenic	
156	Human	Normal	Blood	T Cell	PECAM1	26453327	Experiment	Normal cell	Angiogenic	
157	Human	Normal	Blood	T Cell	CXCR4	26453327	Experiment	Normal cell	Angiogenic	
789	Human	Normal	Blood	Dendritic Cell	AXL	26429389	Single-cell sequencing	Normal cell	AXL+	
181	Human	Normal	Blood	Monocyte	CD14	19201052	Experiment	Normal cell	CD14+	
184	Human	Normal	Blood	Monocyte	CD14	19201052	Experiment	Normal cell	CD14+	
99	Human	Normal	Blood	Progenitor Cell	PROM1	24055327	Experiment	Normal cell	Circulating	
100	Human	Normal	Blood	Progenitor Cell	CD34	24055327	Experiment	Normal cell	Circulating	

Normal/Cancer:

Normal

- Kaposi's Sarcoma
- Larynx Cancer
- Leukemia
- Lipoma
- Liver Cancer
- Lung Adenocarcinoma
- Lung Cancer
- Lung Squamous Cell Carcinoma
- Lymphoma
- Malignant Insulinoma
- Malignant Mesothelioma
- Malignant Peripheral Nerve Sheath Tumor
- Medulloblastoma
- Melanoma
- Mucoepidermoid Carcinoma
- Multiple Myeloma
- Myeloma
- Natural Killer Cell Lymphoma
- Nephroblastoma
- Non-Small Cell Lung Cancer

Normal

- Esophageal Cancer
- Oligodendroglioma
- Oral Cancer
- Oral Squamous Cell Carcinoma
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Pancreatic Ductal Adenocarcinomas
- Papillary Thyroid Carcinoma
- Prostate Cancer
- Renal Cell Carcinoma
- Renal Clear Cell Carcinoma
- Retinoblastoma
- Salivary Gland Tumor
- Sarcoma
- Small Cell Lung Cancer
- Testicular Germ Cell Tumor
- Thyroid Cancer
- Tongue Cancer
- Uterine Leiomyoma

Tissue:

Blood

Select All Deselect All

Tissue

- Abdominal adipose tissue
- Adipose tissue
- Adrenal gland
- Adventitia
- Airway epithelium
- Alveolus
- Amniotic fluid
- Amniotic membrane
- Antecubital vein
- Anterior cruciate ligament
- Artery
- Ascites
- Bladder
- Blood
- Blood vessel
- Bone
- Bone marrow
- Brain
- Breast
- Bronchoalveolar system
- Brown adipose tissue
- Cartilage
- Chorionic villus
- Colon
- Colorectum
- Cornea
- Corneal endothelium
- Corneal epithelium
- Corpus luteum
- Deciduous tooth
- Dental pulp
- Dermis
- Dorsolateral prefrontal cortex
- Embryo
- Embryoid body
- Embryonic brain

Cell type:

B Cell, Basophil, Circulating Fetal Cell, Decidual Cell, Dendritic Cell, Endothelial Ce

Select All Deselect All

celltype

- B Cell
- Basophil
- Circulating Fetal Cell
- Decidual Cell
- Dendritic Cell
- Endothelial Cell
- Eosinophil
- Epithelial Cell
- Erythroblast
- Granulocyte
- Hematopoietic Cell
- Killer Cell
- Leukocyte
- Lymphocyte
- Lymphoid Cell
- Macrophage
- Megakaryocyte
- Monocyte
- Mystloid Cell
- Neutrophil
- Plasmblast
- Platelet
- Progenitor Cell
- Progenitor-like-Angiogenesis-promoting Cell
- Red Blood Cell (Erythrocyte)
- Stem Cell
- T Cell
- Thymic Emigrant Cell
- White Blood Cell

- The differences between Seurat and scCATCH

Cell-type identification methods

ScType SingleR scCATCH

Select reference datasets for cell type annotation

Normal/Cancer:

Normal

Tissue:

Blood

Cell type:

B Cell, Basophil, Circulating Fetal Cell, Decidual Cell, Dendritic Cell, Endothelial Cr

Select method for marker genes identification

scCATCH (Slow) Seurat (Fast)

▶▶ RUN

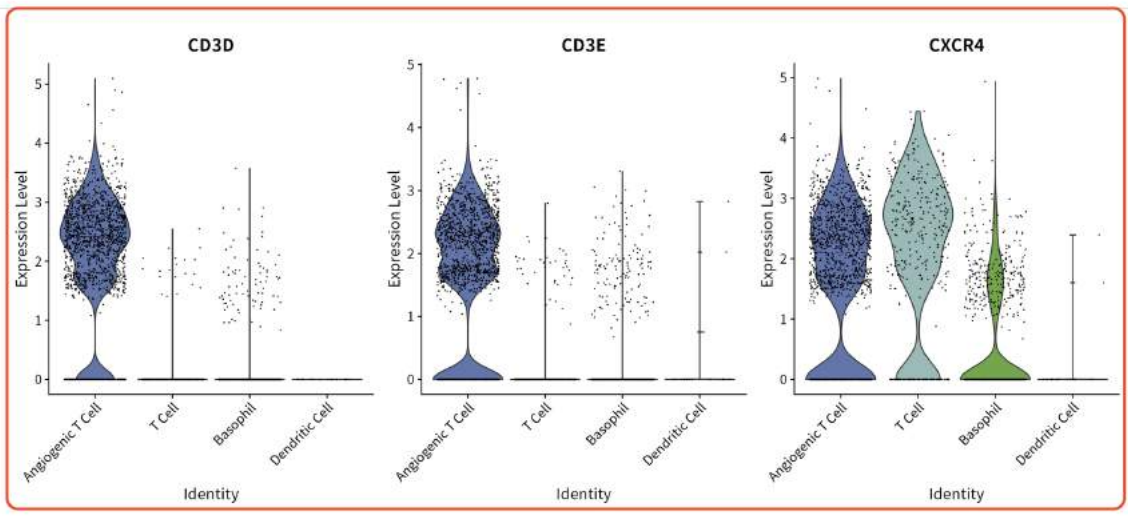
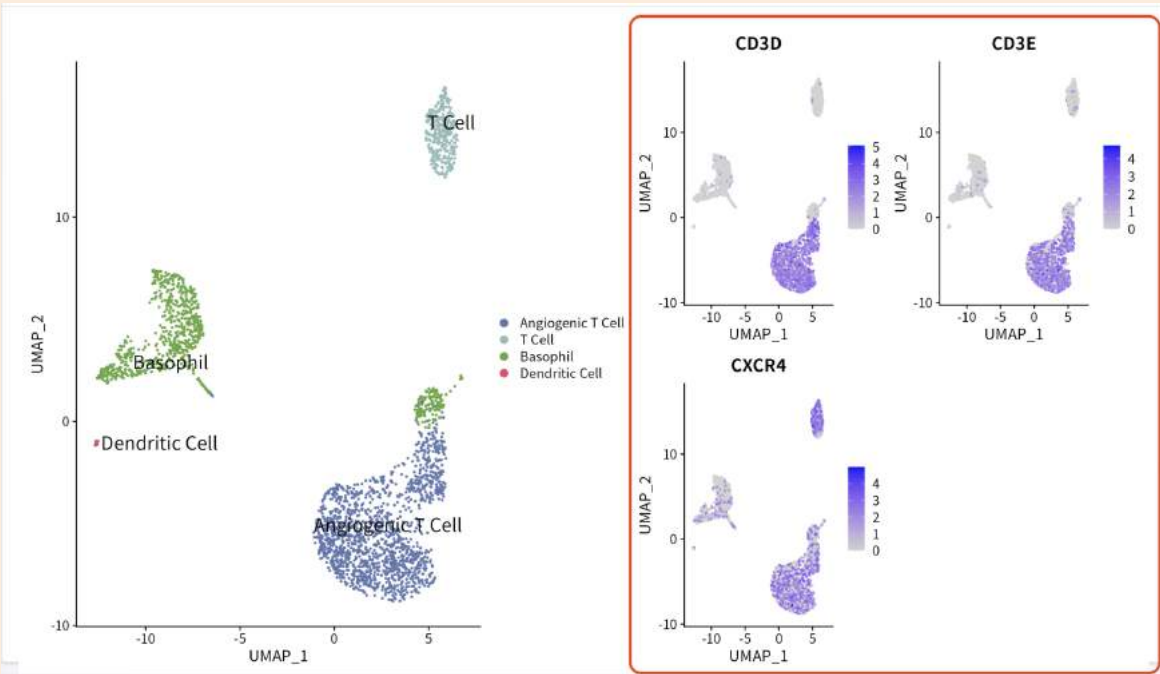
According to page 5 of the PDF, Seurat and scCATCH differ in their approach to finding marker genes for clusters. Seurat uses a one-against-all approach, potentially generating a set of pseudo cluster potential marker genes (highly expressed in at least two clusters). On the other hand, scCATCH carries out paired comparisons to identify differentially expressed genes in only one specific cluster to ensure accuracy in matching the CellMatch database. The cluster potential marker genes identified through scCATCH analysis usually were a subset of genes determined via Seurat. However, an increased number of cluster potential marker genes did not benefit cell annotation.

▼ Seurat approach

Show 5 entries Search:

cluster	cluster_marker	cell_type	celltype_score	celltype_related_marker	PMID
0	JUN, SELL, CXCR4, IL7R, LTB, GIMAP7, RGS10, CD3E, CD3D, EVL, JUNB, DNAJB1, PIK3IP1	Angiogenic T Cell	0.97	CD3D, CD3E, CXCR4	26453327
1	CD52, JUN, CD2, TXNIP, S100A10, CD48, BIN1, CXCR4, IL7R, DUSP1, HLA-E, LTB, GIMAP7, GIMAP4, CD99, PLP2, TSC2D3, ANXA1, IFITM1, PTPRCAP, CD3E, CD3D, SLC2A3, CD69, NAP1L1, ITM2B, EVL, CRIP1, JUNB	Angiogenic T Cell	0.92	CD3D, CD3E, CXCR4	26453327 28342911, 26184676, 21421855,
2	ISG15, IFI6, CTSS, S100A10, S100A11, S100A9, S100A8, S100A4, CD48, ID2, TKT, ANXA5, DUSP1, LST1, AIF1, HLA-DPA1, HLA-DPB1, SNX3, ACTB, SAT1, TIMP1, SSR4, LY6E, FCN1, PSAP, FTH1, NEAT1, CASP4, CASP1, CD63, LYZ, NPC2, MT2A, ARRB2, TYROBP, FTL, TYMP, ITGB2	Basophil	0.88	CD63	20558998, 19945674, 18817894, 14707480
3	CD52, SELL, CXCR4, PLAC8, LTB, HLA-DPA1, HLA-DPB1, SNX3, PTPRCAP, CD69	T Cell	0.75	CXCR4, SELL, CD69	26453327, 22697005, 28566371
4	CD2, ID2, DUSP2, IL7R, HLA-E, CD99, ANXA1, CTSW, PTPRCAP, CD3E, CD3D	Angiogenic T Cell	0.91	CD3D, CD3E	26453327

Showing 1 to 5 of 9 entries Previous Next

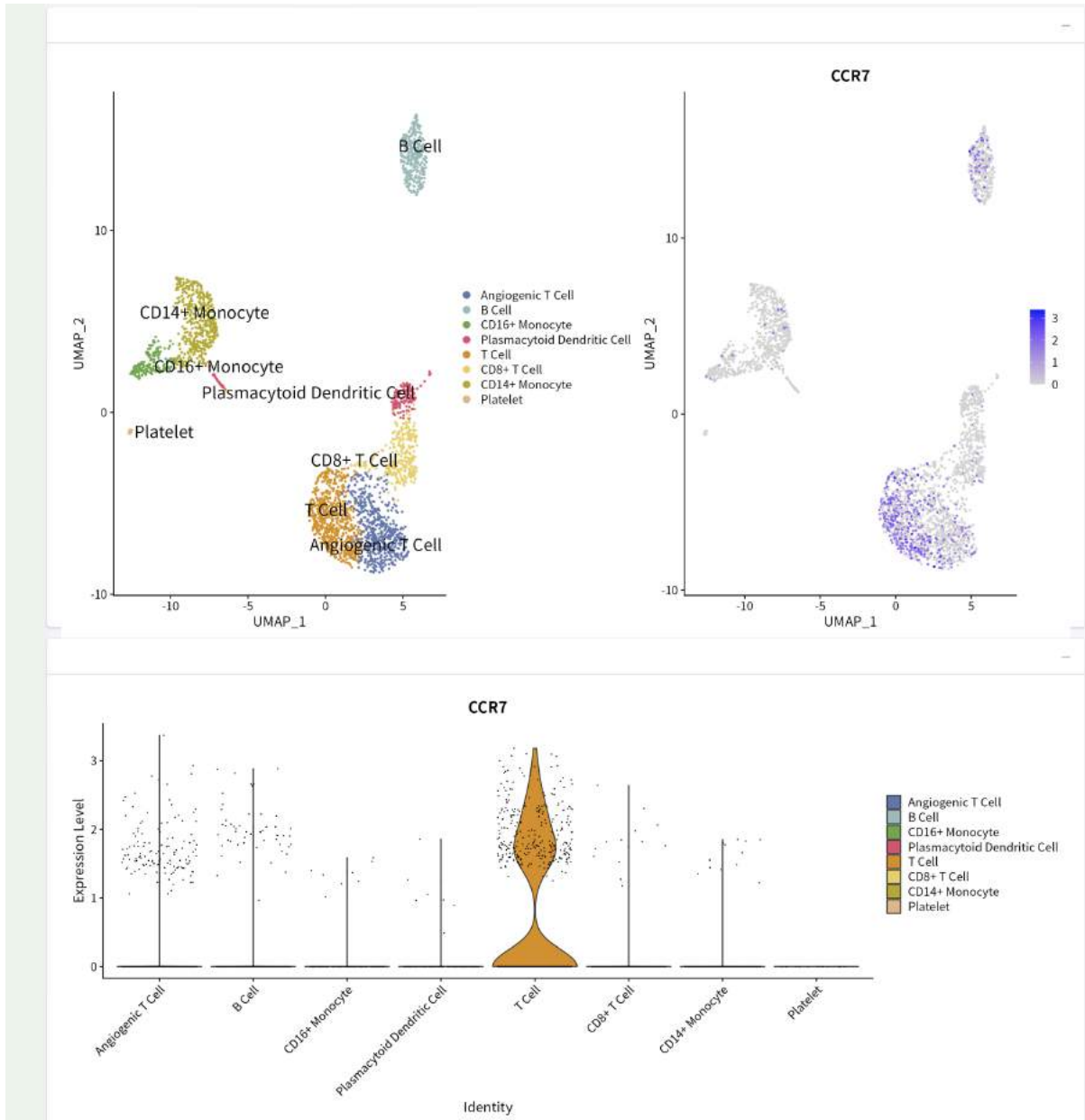


▼ scCATCH approach

Show 5 entries Search:

cluster	cluster_marker	cell_type	celltype_score	celltype_related_marker	PMID
0	CCR7	T Cell	0.5	CCR7	28929596
1	CD2, IL7R, LTB, GIMAP7, CD3E, CD3D	Angiogenic T Cell	0.91	CD3D, CD3E	26453327
2	S100A9, S100A12, S100A8, CD14, FCN1, ASGR1	CD14+ Monocyte	0.88	CD14	19001052
3	MS4A1, TCL1A, CD79B	B Cell	0.75	MS4A1	25254006, 6970772, CD Handbook
4	CD8A	CD8+ T Cell	0.8	CD8A	10754519, 10640724

Showing 1 to 5 of 9 entries Previous **1** 2 Next



 Tissue: 'Blood', 'Peripheral blood', 'Bone marrow'

Cell-type identification methods

ScType SingleR scCATCH

Select reference datasets for cell type annotation

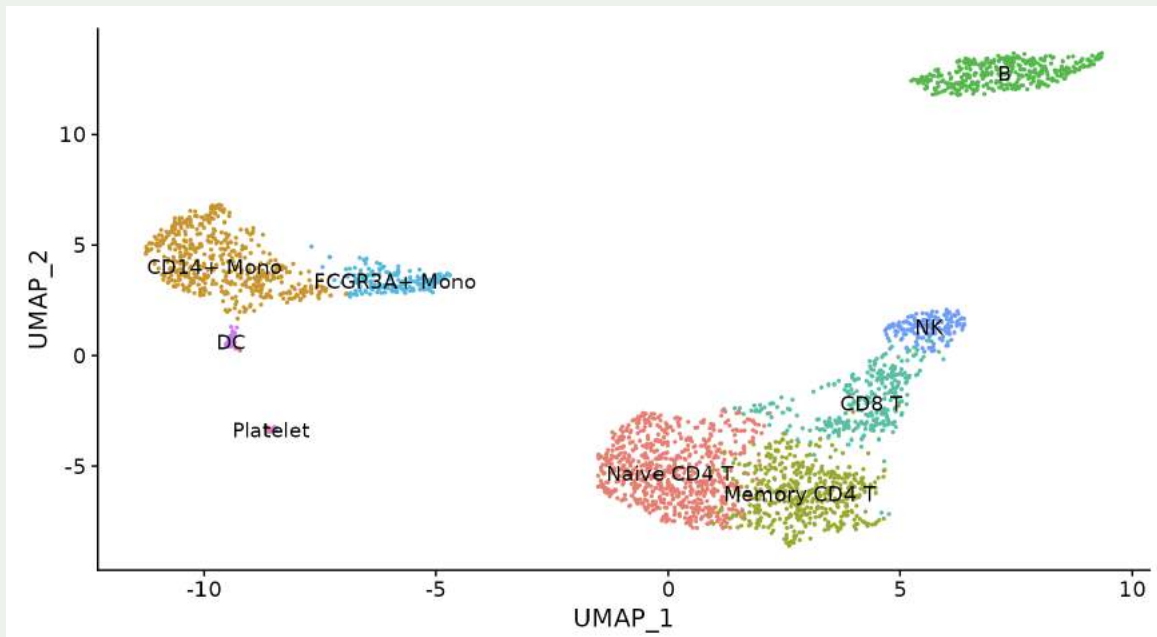
Normal/Cancer:

Normal

Tissue:

Blood, Bone marrow, Peripheral blood

▼ Seurat 提供範例



Show 5 entries Search:

cluster	cluster_marker	cell_type	celltype_score	celltype_related_marker	PMID
0	CCR7	Naive T Cell	0.85	CCR7	23044634, 29361178, 28622514
1	CD2, IL7R, LTB, GIMAP7, CD3E, CD3D, LDHB	Regulatory T Cell	0.96	CD3D, CD3E, IL7R	28777444, 26049548, 23769051
2	S100A9, S100A12, S100A8, CD14, FCN1, ASGR1	CD14+ Monocyte	0.94	CD14	19001052, 29361178
3	MS4A1, TCL1A, ISG20, CD79B, FCER2, CD79A, CD37	B Cell	0.92	MS4A1, CD79A, CD79B, CD37	25799053, 25254006, 6970772, 16181617, 12360049, 1592393, 2365993, 29230012, 29361178, CD Handbook
4	CD8A, GZMK	CD8+ T Cell	0.93	CD8A	28263960, 10754519, 10640724, 29230012, 29361178

Showing 1 to 5 of 9 entries Previous 1 2 Next

