

# awesome-single-cell

## awesome-single-cell




List of software packages (and the people developing these methods) for single-cell data analysis, including RNA-seq, ATAC-seq, etc. [Contributions welcome...](#)

### Citation

DOI [10.5281/zenodo.1117762](https://doi.org/10.5281/zenodo.1117762)

### Software packages

#### RNA-seq

- [alevin-fry](#) - [Rust] -  Rapid, accurate and memory-frugal preprocessing of single-cell and single-nucleus RNA-seq data.
- [anchor](#) - [Python] -  Find bimodal, unimodal, and multimodal features in your data
- [ascend](#) - [R] - ascend is an R package comprised of fast, streamlined analysis functions optimized to address the statistical challenges of single cell RNA-seq. The package incorporates novel and established methods to provide a flexible framework to perform filtering, quality control, normalization, dimension reduction, clustering, differential expression and a wide-range of plotting.
- [BayesPrism](#) - [R] - Bayesian cell Proportion Reconstruction Inferred using Statistical Marginalization (BayesPrism): A Fully Bayesian Inference of Tumor Microenvironment composition and gene expression.
- [bigSca](#) - [matlab] - An analytical framework for big-scale single cell data.
- [bonvoyage](#) - [Python] -  Transform percentage-based units into a 2d space to evaluate changes in distribution with both magnitude and direction.
- [bustools](#) - [C++] - A suite of tools for manipulating BUS files for single cell RNA-Seq pre-processing. bustools can be used to error correct barcodes, collapse UMIs, produce gene count or transcript compatibility count matrices, and is useful for many other tasks.
- [ccRemover](#) - [R] - Removes the Cell-Cycle Effect from Single-Cell RNA-Sequencing Data. [Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data.](#)
- [celda](#) - [R] - A suite of Bayesian hierarchical models and supporting functions to perform clustering of cells and genes for count data generated by scRNA-seq. [Celda: a Bayesian model to perform co-clustering of genes into modules and cells into subpopulations using single-cell RNA-seq data.](#) The package also includes [DecontX](#).

- [CellCNN](#) - [Python] - Representation Learning for detection of phenotype-associated cell subsets
- [CellRanger](#) - [Linux Binary] - Cell Ranger is a set of analysis pipelines that process Chromium single-cell RNA-seq output to align reads, generate gene-cell matrices and perform clustering and gene expression analysis. *Software requires registration with 10xgenomics.*
- [cellTree](#) - [R] - Cell population analysis and visualization from single cell RNA-seq data using a Latent Dirichlet Allocation model.
- [clusterExperiment](#) - [R] - Functions for running and comparing many different clusterings of single-cell sequencing data. Meant to work with SCONE and slingshot.
- [Clustergrammer](#) - [Python, JavaScript] - Interactive web-based heatmap for visualizing and analyzing high dimensional biological data, including single-cell RNA-seq. Clustergrammer can be used within a Jupyter notebook as an interactive widget that can be shared using GitHub and NBviewer, see [example notebook](#).
- [Clustergrammer2](#) - [Python, JavaScript] - Interactive WebGL web-based heatmap for visualizing and analyzing single-cell high-dimensional and location-based biological data. Clustergrammer can be used within a Jupyter notebook as an interactive widget that can be shared using GitHub and NBviewer, see [case studies](#).
- [CountClust](#) - [R] - Functions for fitting Grade-of-Membership models, also known as "Topic models", to RNA-seq counts. These models generalize clustering methods to allow that each cell may belong to more than one cluster/topic.
- [countsimQC](#) - [R] - Compare characteristics of one or more synthetic (e.g., RNA-seq) count matrices to a real count matrix, possibly the one based on which the synthetic data sets were generated.
- [cyclum](#) - [python] - Cyclum is a novel AutoEncoder approach that characterizes circular trajectories in the high-dimensional gene expression space. Applying Cyclum to removing cell-cycle effects leads to substantially improved delineations of cell subpopulations, which is useful for establishing various cell atlases and studying tumor heterogeneity. [bioRxiv](#)
- [CytoGuide](#) - [C++,D3] - [CyteGuide: Visual Guidance for Hierarchical Single-Cell Analysis](#)
- [DecontX](#) - [R] - DecontX is a Bayesian method to automatically estimate and remove read contamination in individual cells from scRNA-seq experiments even without learning any information from empty cell barcodes (identified by cell calling for droplet-based methods). [Decontamination of ambient RNA in single-cell RNA-seq with DecontX](#). Included in package [celda](#).
- [DESCEND](#) - [R] - DESCEND deconvolves the true gene expression distribution across cells for UMI scRNA-seq counts. It provides estimates of several distribution based statistics (five distribution measurements and the coefficients of covariates (such as batches or cell size)).

- DeLorean - [R] - Bayesian pseudotime estimation algorithm that uses Gaussian processes to model gene expression profiles and provides a full posterior for the pseudotimes.
- dittoSeg - [R] - Bioconductor package offering user friendly visualization tools for single-cell and Bulk RNA Sequencing. Color blindness friendly by default; novice coder friendly; highly customizable and powerful enough to build publication-ready figures; universal in that it works directly with Seurat, SingleCellExperiment, and SummarizedExperiment objects and has import capabilities for edgeR DGElists.
- dropkick - [Python] - Automated cell filtering for single-cell RNA sequencing data.
- dynamo - [Python] - Inclusive model of expression dynamics with scSLAM-seq and multiomics, vector field reconstruction and potential landscape mapping.
- embeddr - [R] - Embeddr creates a reduced dimensional representation of the gene space using a high-variance gene correlation graph and laplacian eigenmaps. It then fits a smooth pseudotime trajectory using principal curves.
- Falco - [AWS cloud] - Falco: A quick and flexible single-cell RNA-seq processing framework on the cloud.
- FastProject - [Python] - Signature analysis on low-dimensional projections of single-cell expression data.
- flotilla - [Python] - Reproducible machine learning analysis of gene expression and alternative splicing data
- GPfates - [Python] - Model transcriptional cell fates as mixtures of Gaussian Processes
- GSEAPy - [Python] - GSEAPy: Gene Set Enrichment Analysis in Python. GSEAPy is a Python/Rust implementation for GSEA and wrapper for Enrichr. GSEAPy can be used for RNA-seq, ChIP-seq, Microarray data. It can be used for convenient GO enrichment and to produce publication quality figures in python.
- HocusPocus - [R] - Basic PCA-based workflow for analysis and plotting of single cell RNA-seq data.
- HTSeq - [Python] - A Python library to facilitate programmatic analysis of data from high-throughput sequencing (HTS) experiments. A popular component of HTSeq is htseq-count , a script to quantify gene expression in bulk and single-cell RNA-Seq and similar experiments.
- IA-SVA - [R] - Iteratively Adjusted Surrogate Variable Analysis (IA-SVA) is a statistical framework to uncover hidden sources of variation even when these sources are correlated with the biological variable of interest. IA-SVA provides a flexible methodology to i) identify a hidden factor for unwanted heterogeneity while adjusting for all known factors; ii) test the significance of the putative hidden factor for explaining the variation in the data; and iii), if significant, use the estimated factor as an additional known factor in the next iteration to uncover further hidden factors.

- ICGS - [Python] - Iterative Clustering and Guide-gene Selection (Olsson et al. Nature 2016). Identify discrete, transitional and mixed-lineage states from diverse single-cell transcriptomics platforms. Integrated FASTQ pseudoalignment /quantification (Kallisto), differential expression, cell-type prediction and optional cell cycle exclusion analyses. Specialized methods for processing BAM and 10X Genomics spares matrix files. Associated single-cell splicing PSI methods (MultiPath-PSI). Apart of the AltAnalyze toolkit along with accompanying visualization methods (e.g., heatmap, t-SNE, SashimiPlots, network graphs). Easy-to-use graphical user and commandline interfaces.
- ivis - [Python or R] - Structure-preserving dimensionality reduction in single-cell datasets.
- kallisto - [C++] - kallisto is a program for quantifying abundances of transcripts or genes from bulk or single-cell RNA-Seq data, or more generally of target sequences using high-throughput sequencing reads. It is based on pseudoalignment for rapidly determining the compatibility of reads with targets, without the need for alignment.
- kb-python - [Python] - kb-python is a python package for processing single-cell RNA-sequencing. It wraps the kallisto | bustools single-cell RNA-seq command line tools in order to unify multiple processing workflows.
- knn-smoothing - [python or R or matlab] - The algorithm is based on the observation that across protocols, the technical noise exhibited by UMI-filtered scRNA-Seq data closely follows Poisson statistics. Smoothing is performed by first identifying the nearest neighbors of each cell in a step-wise fashion, based on variance-stabilized and partially smoothed expression profiles, and then aggregating their transcript counts.
- mfa - [R] - Probabilistic modeling of bifurcations in single-cell gene expression data using a Bayesian mixture of factor analyzers
- M3Drop - [R] - Michaelis-Menten Modelling of Dropouts for scRNASeq.
- MetaCell - [R, C++] - Analysis of single cell RNA-seq data by computing partitions of a cell similarity graph into small homogeneous groups of cells called metacells.
- MIMOSCA - [python] - A repository for the design and analysis of pooled single cell RNA-seq perturbation experiments (Perturb-seq).
- Monocle - [R] - Differential expression and time-series analysis for single-cell RNA-Seq.
- Muscat - [R] - muscat (Multi-sample multi-group scRNA-seq analysis tools ) provides various methods for Differential State (DS) analyses in multi-sample, multi-group, multi-(cell-)subpopulation scRNA-seq data.

- netSmooth - [R] - netSmooth is a network-diffusion based method that uses priors for the covariance structure of gene expression profiles on scRNA-seq experiments in order to smooth expression values. We demonstrate that netSmooth improves clustering results of scRNA-seq experiments from distinct cell populations, time-course experiments, and cancer genomics.
- NetworkInference - [Julia] - Fast implementation of single-cell network inference algorithms: Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures
- nimfa - [Python] - Nimfa is a Python scripting library which includes a number of published matrix factorization algorithms, initialization methods, quality and performance measures and facilitates the combination of these to produce new strategies. The library represents a unified and efficient interface to matrix factorization algorithms and methods.
- novoSpaRc - [Python] - Predict locations of single cells in space by solely using single-cell RNA sequencing data. An existing reference database of marker genes is not required, but significantly enhances performance if available. [bioRxiv](#).
- OEFinder - [R] - Identify ordering effect genes in single cell RNA-seq data. OEFinder shiny implemmentation depends on packages shiny, shinyFiles, gdata, and EBSeq.
- OncoNEM - [R] - OncoNEM is a probabilistic method for inferring intra-tumor evolutionarylineage trees from somatic single nucleotide variants of single cells. OncoNEM identifies homogeneous cellularsubpopulations and infers their genotypes as well as a tree describing their evolutionary relationships.
- outrigger - [Python] - Outrigger is a program to calculate alternative splicing scores of RNA-Seq data based on junction reads and a *de novo*, custom annotation created with a graph database, especially made for single-cell analyses.
- pcaReduce - [R] - hierarchical clustering of single cell transcriptional profiles.
- PyGMNormalize - [Python] - Python implementation of edgeR normalization method for count matrices.
- RAPIDS-singlecell - [Python] - A GPU-accelerated tool leveraging RAPIDS for scRNA analysis. Seamless scverse compatibility for efficient single-cell data processing and analysis. Replcates features from Scanpy, while also incorporating select functionalities from Squidpy and Decoupler.
- rMATS - [Python] - RNA-Seq Multivariate Analysis of Transcript Splicing.
- robustSingleCell - [R] - robustSingleCell is a pipeline designed to identify robust cell subpopulations using scRNAseq data and compare population compositions across tissues and experimental models via similarity analysis as described in Magen et al. (2019) [bioRxiv](#).
- SAVER - [R] - SAVER (Single-cell Analysis Via Expression Recovery) implements a regularized regression prediction and empirical Bayes method to recover the true gene expression profile in noisy and sparse single-cell RNA-seq data.

- SAKE - [R] - Single-cell RNA-Seq Analysis and Clustering Evaluation.
- SCALE - [R] - SCALE is a statistical framework for Single Cell ALlelic Expression analysis. SCALE estimates kinetic parameters that characterize the transcriptional bursting process at the allelic level, while accounting for technical bias.
- Scanpy - [Python] - Scanpy provides computationally efficient tools that scale up to very large data sets and enables simple integration of advanced machine learning algorithms.
- scbean - [Python] - Scbean integrates a range of models for single-cell data analysis, including dimensionality reduction, removing batch effects, and transferring well-annotated cell type labels from scRNA-seq to scATAC-seq and spatial resolved transcriptomics, and joint-analysis of paired multimodal single-cell data.
- SCCAF - [Python] Single Cell Clustering Assessment Framework (SCCAF) is a method for automated identification of putative cell types from single cell data by iteratively applying clustering and a machine learning approach. Putative cell type discovery from single-cell gene expression data
- SCell - [matlab] - SCell is an integrated software tool for quality filtering, normalization, feature selection, iterative dimensionality reduction, clustering and the estimation of gene-expression gradients from large ensembles of single-cell RNA-seq datasets. SCell is open source, and implemented with an intuitive graphical interface.
- scGEAToolbox - [matlab] - a Matlab toolbox for single-cell RNA-seq data analyses.
- schist - [Python] - schist is a scanpy-compatible python library which implements Nested Stochastic Block Models to identify cell groups in single cell experiments.
- Scillus - [R] - Scillus is an R wrapper package for enhanced processing and visualization of Seurat-based scRNA-seq data.
- SCINA - [R] - A semi-supervised category identification and assignment tool.
- SCP - [R] - SCP(Single Cell Pipeline) is an R package that provides a comprehensive set of tools for single cell data processing and downstream analysis.
- scVI - [python] - scVI is a ready-to-use scalable framework for the probabilistic representation and analysis of gene expression in single cells (batch correction, visualization, clustering, and differential expression). Deep generative modeling for single-cell transcriptomics
- scLM - [R] - Automatic detection of consensus gene clusters across multiple single-cell datasets
- scLVM - [R] - scLVM is a modelling framework for single-cell RNA-seq data that can be used to dissect the observed heterogeneity into different sources, thereby allowing for the correction of confounding sources of variation. scLVM was primarily designed to account for cell-cycle induced variations in single-cell RNA-seq data where cell cycle is the primary source of variability.

- scTDA - [Python] - scTDA is an object oriented python library for topological data analysis of high-throughput single-cell RNA-seq data. It includes tools for the preprocessing, analysis, and exploration of single-cell RNA-seq data based on topological representations.
- SCODE - [R/Julia]- an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation
- SCORE - [R] - Enhancing single-cell cellular state inference by incorporating molecular network features.
- SCOUP - [C++] - Uses probabilistic model based on the Ornstein-Uhlenbeck process to analyze single-cell expression data during differentiation.
- scrn - [R] - This package implements a variety of low-level analyses of single-cell RNA-seq data. Methods are provided for normalization of cell-specific biases, pool-based norms to estimate size factors, assignment of cell cycle phase, and detection of highly variable and significantly correlated genes.
- SCRL - [C++] - Network embedding-based representation learning for single cell RNA-seq data
- scruff - [R] - An R package for preprocessing single cell RNA-seq (scRNA-seq) FASTQ reads generated by CEL-Seq and CEL-Seq2 protocols. It demultiplexes reads according to a predetermined list of cell barcodes, maps reads to reference genome using Rsubread, and reports filtered UMI (Unique Molecular Identifier) count matrix ready for downstream analysis. scruff: an R/Bioconductor package for preprocessing single-cell RNA-sequencing data.
- scSVA - [R] - An R package for interactive two- and three-dimensional visualization and exploration of massive single-cell omics data ( $2 \cdot 10^9$  cells). scSVA supports interactive analytics in a cloud with containerized tools. It contains optimized implementation of diffusion maps and multi-threaded 3D force-directed layout (ForceAtlas2).
- scTCRseq - [python] - Map T-cell receptor (TCR) repertoires from single cell RNAseq.
- Seurat - [R] - It contains easy-to-use implementations of commonly used analytical techniques, including the identification of highly variable genes, dimensionality reduction (PCA, ICA, t-SNE), standard unsupervised clustering algorithms (density clustering, hierarchical clustering, k-means), and the discovery of differentially expressed genes and markers.
- SIMLR - [R, matlab] - SIMLR (Single-cell Interpretation via Multi-kernel LeaRning) learns an appropriate distance metric from the data for dimension reduction, clustering and visualization. SIMLR is capable of separating known subpopulations more accurately in single-cell data sets than do existing dimension reduction methods.

- sincell - [R] - Existing computational approaches for the assessment of cell-state hierarchies from single-cell data might be formalized under a general workflow composed of i) a metric to assess cell-to-cell similarities (combined or not with a dimensionality reduction step), and ii) a graph-building algorithm (optionally making use of a cells-clustering step). Sincell R package implements a methodological toolbox allowing flexible workflows under such framework.
- sincera - [R] - R-based pipeline for single-cell analysis including clustering and visualization.
- SingleSplice - [R, perl, C++] - A tool for detecting biological variation in alternative splicing within a population of single cells. See [Welch et al. 2016](#).
- singlet - [Python] - Single cell RNA-Seq analysis with phenotypes.
- soupX - [R] - An R package for the estimation and removal of cell free mRNA contamination in droplet based single cell RNA-seq data. The problem this package attempts to solve is that all droplet based single cell RNA-seq experiments also capture ambient mRNAs present in the input solution along with cell specific mRNAs of interest.
- SPRING - [matlab, javascript, python] - SPRING is a collection of pre-processing scripts and a web browser-based tool for visualizing and interacting with high dimensional data. SPRING was developed for single cell RNA-Seq data but can be applied more generally.
- scTOP - [Python] - Single-cell type order parameters. Physics-inspired method of processing single-cell RNA-seq and identifying cell fate, motivated by the epigenetic landscape.
- trendsceek - [R] - [Identification of spatial expression trends in single-cell gene expression data](#)
- VISION - [] - A tool for annotating the sources of variation in single cell RNA-seq data in an automated, unbiased and scalable manner. It produces an interactive, low latency and feature rich web-based report that can be easily shared amongst researchers.
- zUMIs - [R, perl, shell] - [zUMIs: A fast and flexible pipeline to process RNA-seq data with UMIs](#).
- STAR - [C/C++] - Splice-aware aligner for RNA-seq data, capable of mapping reads to a reference genome with high accuracy and speed.

## Quality control

- Cellity - [R] - Classification of low quality cells in scRNA-seq data using R
- gene\_network\_evaluation - [Python] - A flexible framework to evaluate the plausibility of gene programs inferred from single-cell genomic data. The assessment is broken down into themes such as goodness of fit (ability to explain the data), co-regulation, mechanistic interactions etc. Under each theme, multiple evaluation tasks are conceptualised and implemented using appropriate statistical tests.



- [scDiagnostics](#) - [R] - Package specifically designed to evaluate the fidelity of annotation transfer in scRNA-seq data. scDiagnostics provides a comprehensive set of diagnostic tools that assess the compatibility between query and reference datasets, helping to identify and mitigate risks of erroneous annotations.
- [SCONE](#) - [R] - SCONE (Single-Cell Overview of Normalized Expression), a package for single-cell RNA-seq data quality control (QC) and normalization. This data-driven framework uses summaries of expression data to assess the efficacy of normalization workflows.
- [SinQC](#) - [R] - A Method and Tool to Control Single-cell RNA-seq Data Quality.
- [scater](#) - [R] - Scater places an emphasis on tools for quality control, visualisation and pre-processing of data before further downstream analysis, filling a useful niche between raw RNA-sequencing count or transcripts-per-million data and more focused downstream modelling tools such as monocle, scLVM, SCDE, edgeR, limma and so on.

## Gene regulatory network identification

- [scPRINT](#) - [python] - scPRINT is pretrained on 50M cells to predict robust gene networks from single cell RNAseq data. [scPRINT: pre-training on 50 million cells allows robust gene network predictions](#)
- [Dictys](#) - [Python] - Dictys reconstructs and analyzes context specific and dynamic Gene Regulatory Networks from scRNA-seq and scATAC-seq datasets. [Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multiomics](#)
- [Normalizr](#) - [Python, Shell] - Normalizr infers Gene Regulatory Networks from Perturb-seq and other single-cell CRISPR screens. Its normalization and statistical association testing framework also unifies single-cell differential expression and co-expression. [Single-cell normalization and association testing unifying CRISPR screen and gene co-expression analyses with Normalizr](#).
- [SCENIC](#) - [R] - SCENIC is an R package to infer Gene Regulatory Networks and cell types from single-cell RNA-seq data. [SCENIC: single-cell regulatory network inference and clustering](#)
- [SCENIC+](#) - [python] - SCENIC+ is a python package to build gene regulatory networks using combined or separate scRNA-seq and scATAC-seq data. [SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks](#)
- [SINCERITIES](#) - [R/Matlab] - [Inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles](#)

## Immune receptor profiling

- [APackOfTheClones](#) - [R] - [APackOfTheClones: Visualization of clonal expansion with circle packing](#)
- [DALI](#) - [R] - Diversity Analysis Interface (DALI) is a tool that enables TCR and BCR analysis in the Seurat ecosystem. The functionality of the tool is also exposed via an interactive Shiny application.

- [Scirpy](#) - [Python] - [A Scanpy extension for analyzing single-cell T-cell receptor \(TCR\) sequencing data.](#)
- [scRepertoire](#) - [R] - [scRepertoire: An R-based toolkit for single-cell immune receptor analysis](#)
- [TraCeR](#) - [python] - Reconstruction of T-Cell receptor sequences from single-cell RNA-seq data.
- [TRAPeS](#) - [python, C++] - TRAPeS (TCR Reconstruction Algorithm for Paired-End Single-cell), a software for reconstruction of T cell receptors (TCR) using short, paired-end single-cell RNA-sequencing.

## Marker and differential gene expression identification

- [GPseudoClust](#) - [Python] - Software that clusters genes for pseudotemporally ordered data and quantifies the uncertainty in cluster allocations arising from the uncertainty in the pseudotime ordering.
- [GiniClust](#) - [Python/R] - GiniClust is a clustering method implemented in Python and R for detecting rare cell-types from large-scale single-cell gene expression data. GiniClust can be applied to datasets originating from different platforms, such as multiplex qPCR data, traditional single-cell RNAseq or newly emerging UMI-based single-cell RNAseq, e.g. inDrops and Drop-seq.
- [DECENT](#) - [R] - The unique features of scRNA-seq data have led to the development of novel methods for differential expression (DE) analysis. However, few of the existing DE methods for scRNA-seq data estimate the number of molecules pre-dropout and therefore do not explicitly distinguish technical and biological zeroes. We develop DECENT, a DE method for scRNA-seq data that adjusts for the imperfect capture efficiency by estimating the number of molecules pre-dropout.
- [MetaMarkers](#) - [R] - MetaMarkers proposes a simple methodology to pool marker information across dataset while keeping dataset independents to identify robust marker signatures from single-cell data. [How many markers are needed to robustly determine a cell's type?](#)
- [Phenotype Cover](#) - [Python] - Provides two algorithms for marker selection (G-PC, CEM-PC) introduced in [Multiset multicover methods for discriminative marker selection](#). Most marker selection methods focus on differential expression (DE) analysis. Although such methods work well for data with a few non-overlapping marker sets, they are not appropriate for large atlas-size datasets where several cell types and tissues are considered. To address this, we define the phenotype cover (PC) problem for marker selection and present algorithms that can improve the discriminative power of marker sets.
- [scDD](#) - [R] - scDD (Single-Cell Differential Distributions) is a framework to identify genes with different expression patterns between biological groups of interest. In addition to traditional differential expression, it can detect differences that are more complex and subtle than a mean shift.

- SCDE - [R] - Differential expression using error models and overdispersion-based identification of important gene sets.
- SCMarker - [R] - SCMarker is a method performing ab initio marker gene set selection from scRNA-seq data to achieve improved clustering/cell-typing results. SCMarker: ab initio marker selection for single cell transcriptome profiling.
- SEPA - [R] - SEPA provides convenient functions for users to assign genes into different gene expression patterns such as constant, monotone increasing and increasing then decreasing. SEPA then performs GO enrichment analysis to analysis the functional roles of genes with same or similar patterns.
- switchde - [R] - Differential expression analysis across pseudotime. Identify genes that exhibit switch-like up or down regulation along single-cell trajectories along with where in the trajectory the regulation occurs.

## Cell clustering

- BackSPIN - [Python] - Biclustering algorithm developed taking into account intrinsic features of single-cell RNA-seq experiments.
- dropClust - [R/Python] - Efficient clustering of ultra-large scRNA-seq data.
- SC3 - [R] - SC3 is a tool for the unsupervised clustering of cells from single cell RNA-Seq experiments.
- TooManyCells - [Haskell, CLI program] - Suite of graph-based tools for efficient, global, and unbiased identification and visualization of cell clades.

## Dimension reduction

- torchdr - [python] - Dimensionality reduction toolbox using PyTorch, featuring various algorithms such as TSNE, UMAP, and more. Supports GPU acceleration to maximize computational efficiency.
- destiny - [R] - Diffusion maps are spectral method for non-linear dimension reduction introduced by Coifman et al.(2005). Diffusion maps are based on a distance metric (diffusion distance) which is conceptually relevant to how differentiating cells follow noisy diffusion-like dynamics, moving from a pluripotent state towards more differentiated states.
- PHATE - Potential of Heat-diffusion for Affinity-based Transition Embedding - [Python, R, matlab] - PHATE is a tool for visualizing high dimensional single-cell data with natural progressions or trajectories. PHATE uses a novel conceptual framework for learning and visualizing the manifold inherent to biological systems in which smooth transitions mark the progressions of cells from one state to another.
- [picasso] (<https://github.com/pachterlab/picasso>) - [python] - Map the points of an input matrix to user-defined, n-dimensional shape coordinates, while minimizing reconstruction error using an autoencoder neural network structure.
- scvis - [python] - Interpretable dimensionality reduction of single cell transcriptome data with deep generative models
- SWNE - [R] - Visualizing single-cell RNA-seq datasets with Similarity Weighted Nonnegative Embedding.(SWNE).

- ZIFA - [Python] - Zero-inflated dimensionality reduction algorithm for single-cell data.
- scPRINT - [python] - scPRINT is pretrained on 50M cells and generates multiple cell embeddings from single cell RNAseq profiles. scPRINT: pre-training on 50 million cells allows robust gene network predictions
- scDEED - [R] optimizing hyperparameters of UMAP/t-SNE, assigning each embedding a "reliability score" by permutation , manuscript open access: Statistical method scDEED for detecting dubious 2D single-cell embeddings and optimizing t-SNE and UMAP hyperparameters

## Archetypal analysis

- scAAnet - [Python] - scAAnet performs non-linear archetypal analysis through autoencoder networks to identify shared gene expression programs (GEPs) among heterogenous cell populations and infer relative activity of each GEP across cells.

## Count modelling and normalization

- BASiCS - [R] - Bayesian Analysis of single-cell RNA-seq data. Estimates cell-specific normalization constants. Technical variability is quantified based on spike-in genes. The total variability of the expression counts is decomposed into technical and biological components. BASiCS can also identify genes with differential expression/over-dispersion between two or more groups of cells.
- BEARscc - [R] - BEARscc makes use of ERCC spike-in measurements to model technical variance as a function of gene expression and technical dropout effects on lowly expressed genes.
- BPSC - [R] - Beta-Poisson model for single-cell RNA-seq data analyses
- dsb - [R or Python] - a method for normalizing and denoising protein data from antibody derived tags (ADT). Compatible with CITE-seq, ASAP-seq, TEA-seq, ICICLE-seq, MissionBio etc. Removes ambient and cell to cell technical noise from ADTs see vignettes on CRAN. Manuscript open access: Normalizing and denoising protein expression data from droplet-seq single cell profiling. Nature Communications (2022).
- MAST - [R] - Model-based Analysis of Single-cell Transcriptomics (MAST) fits a two-part, generalized linear models that are specially adapted for bimodal and/or zero-inflated single cell gene expression data
- SCnorm - [R] - A quantile regression based approach for robust normalization of single cell RNA-seq data.
- zinbwaveZinger - [R] - We introduce a weighting strategy, based on a zero-inflated negative binomial model, that identifies excess zero counts and generates gene- and cell-specific weights to unlock bulk RNA-seq DE pipelines for zero-inflated data, boosting performance for scRNA-seq. <https://doi.org/10.1186/s13059-018-1406-4>

## Batch-effect removal

- BatchEffectRemoval - [Python] - Removal of Batch Effects using Distribution-Matching Residual Networks
- ResPAN - [Python] - ResPAN is a light structured **Residual** autoencoder and mutual nearest neighbor **P**airing guided **A**dversarial **N**etwork for scRNA-seq batch correction.
- scPLS - [C++, R] - A normalization method to remove unwanted variation using both control and target genes. It takes advantage of the fact that genes in a scRNAseq study often can be naturally classified into two sets: a control set of genes that are free of effects of the predictor variables and a target set of genes that are of primary interest. By modeling the two sets of genes jointly using the partial least squares regression, scPLS is capable of making full use of the data to improve the inference of confounding effects.  
<https://www.nature.com/articles/s41598-017-13665-w>
- TASC - [C++, python] - To account for cell-to-cell technical differences, we propose a statistical framework, TASC (Toolkit for Analysis of Single Cell RNA-seq), an empirical Bayes approach to reliably model the cell-specific dropout rates and amplification bias by use of external RNA spike-ins. TASC incorporates the technical parameters, which reflect cell-to-cell batch effects, into a hierarchical mixture model to estimate the biological variance of a gene and detect differentially expressed genes. More importantly, TASC is able to adjust for covariates to further eliminate confounding that may originate from cell size and cell cycle differences.
- UNCURL - [Python] - Unsupervised and semi-supervised sampling effect removal for single-cell RNA-seq data.

## Cell projection and unimodal integration

- scmap - [R] - scmap is a method for projecting cells from a scRNA-seq experiment on to the cell-types identified in a different experiment.
- Monet - [python] - A package for analyzing and integrating scRNA-Seq data using PCA-based latent spaces.

## Simulation

- dropsim - [R] - Simulating droplet based scRNA-seq data.
- powsimR - [R] - Power analysis is essential to optimize the design of RNA-seq experiments and to assess and compare the power to detect differentially expressed genes. PowsimR is a flexible tool to simulate and evaluate differential expression from bulk and especially single-cell RNA-seq data making it suitable for a priori and posterior power analyses.
- splatter - [R] - Splatter is a package for the simulation of single-cell RNA sequencing count data. It provides a simple interface for creating complex simulations that are reproducible and well-documented.
- symsim - [R] - SymSim (Synthetic model of multiple variability factors for Simulation) is an R package for simulation of single cell RNA-Seq data.

## Pseudotime and trajectory inference

- CALISTA - [R] - CALISTA provides a user-friendly toolbox for the analysis of single cell expression data. CALISTA accomplishes three major tasks: 1) Identification of cell clusters in a cell population based on single-cell gene expression data, 2) Reconstruction of lineage progression and produce transition genes, and 3) Pseudotemporal ordering of cells along any given developmental paths in the lineage progression.
- CoSpar - [python] - CoSpar is a toolkit for dynamic inference by integrating state and lineage information. It gains superior robustness and accuracy by exploiting both the local coherence and sparsity of differentiation transitions, i.e., neighboring initial states share similar yet sparse fate outcomes. When only state information is available, CoSpar also improves upon existing dynamic inference methods by imposing sparsity and coherence.
- DensityPath - [.] - DensityPath: a level-set algorithm to visualize and reconstruct cell developmental trajectories for large-scale single-cell RNAseq data
- dynverse - [R] - A comparison of single-cell trajectory inference methods: towards more accurate and robust tools
- ECLAIR - [python] - ECLAIR stands for Ensemble Clustering for Lineage Analysis, Inference and Robustness. Robust and scalable inference of cell lineages from gene expression data.
- K-Branches - [R] - The main idea behind the K-Branches method is to identify regions of interest (branching regions and tips) in differentiation trajectories of single cells. So far, K-Branches is intended to be used on the diffusion map representation of the data, so the user should either provide the data in diffusion map space or use the destiny package perform diffusion map dimensionality reduction.
- MERLoT - [R/python] - Reconstructing complex lineage trees from scRNA-seq data using MERLoT.
- ouija - [R] - A descriptive marker gene approach to single-cell pseudotime inference
- ouijaflow - [python] - A descriptive marker gene approach to single-cell pseudotime inference
- Palantir - [Python] - Characterization of cell fate probabilities in single-cell data with Palantir
- PhenoPath - [R] - Single-cell pseudotime with heterogeneous genetic and environmental backgrounds, including Bayesian significance testing of interactions.
- pseudodynamics - [MATLAB] - Inferring population dynamics from single-cell RNA-sequencing time series data

- psupertime - [R] - psupertime is an R package which uses single cell RNAseq data, where the cells have labels following a known sequence (e.g. a time series), to identify a small number of genes which place cells in that known order. It can be used for discovery of relevant genes, for exploration of unlabelled data, and assessment of one dataset with respect to the labels known for another dataset. - preprint
- SCDIFF - [Python, JavaScript] - SCDIFF is a single-cell trajectory inference method with interactive visualizations powered by D3.js. SCDIFF utilized the TF regulatory information to mitigate the impact of enormous single-cell RNA-seq noise (such as drop-out). With the TF regulatory information, SCDIFF is also able to predict the TFs (and their activation time), which drive the cells to different cell fates. Such predictive power has been experimentally validated.
- SCIMITAR - [Python] - Single Cell Inference of Morphing Trajectories and their Associated Regulation module (SCIMITAR) is a method for inferring biological properties from a pseudotemporal ordering. It can also be used to obtain progression-associated genes that vary along the trajectory, and genes that change their correlation structure over the trajectory; progression co-associated genes.
- SCORPIUS - [R] - An accurate and easy tool for performing linear trajectory inference on single cells using single-cell RNA sequencing data. In addition, SCORPIUS provides functions for discovering the most important genes with respect to the reconstructed trajectory, as well as nice visualisation tools. Cannoodt et al. (2016) doi:10.1101/079509.
- SCUBA - [matlab/R] - SCUBA stands for "Single-cell Clustering Using Bifurcation Analysis." SCUBA is a novel computational method for extracting lineage relationships from single-cell gene expression data, and modeling the dynamic changes associated with cell differentiation.
- scVelo - [Python] - scVelo is a scalable toolkit for RNA velocity analysis in single cells. It generalizes the concept of RNA velocity by relaxing previously made assumptions with a dynamical model. It allows to identify putative driver genes, infer a latent time, estimate reaction rates of transcription, splicing and degradation, and detect competing kinetics.
- SLICER - [R] - Selective Locally linear Inference of Cellular Expression Relationships (SLICER) algorithm for inferring cell trajectories.
- slingshot - [R] - Functions for identifying and characterizing continuous developmental trajectories in single-cell sequencing data.
- SPADE - [R] - Visualization and cellular hierarchy inference of single-cell data using SPADE.

- TASIC - [matlab] - TASIC is a new method for determining temporal trajectories, branching and cell assignments in single cell time series experiments. Unlike prior approaches TASIC uses on a probabilistic graphical model to integrate expression and time information making it more robust to noise and stochastic variations.
- TopSLAM - [python] - Extracting and using probabilistic Waddington's landscape recreation from single cell gene expression measurements
- TSCAN - [R] - Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis.
- VELOCYTO - [Python, R] - Estimating RNA velocity in single cell RNA sequencing datasets.

## Cell type identification and classification

- ceLLama - [R/Python] - ceLLama is a streamlined automation pipeline for cell type annotations using local large-language models (LLMs).
- cellassign - [R] - Automated, probabilistic assignment of scRNA-seq to known types. cellassign automatically assigns single-cell RNA-seq data to known cell types across thousands of cells accounting for patient and batch specific effects. Information about a priori known markers for cell types is provided as input to the model. cellassign then probabilistically assigns each cell to a cell type, removing subjective biases from typical unsupervised clustering workflows. [bioRxiv](#)
- CHETAH - [R] - CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. CHETAH (CHaracterization of cEll Types Aided by Hierarchical clustering) is an accurate cell type identification algorithm that is rapid and selective, including the possibility of intermediate or unassigned categories. Evidence for assignment is based on a classification tree of previously available scRNA-seq reference data and includes a confidence score based on the variance in gene expression per cell type. For cell types represented in the reference data, CHETAH's accuracy is as good as existing methods. Its specificity is superior when cells of an unknown type are encountered, such as malignant cells in tumor samples which it pinpoints as intermediate or unassigned. [bioRxiv](#)



- CIPR - [R] - (Cluster Identity PRedictor-pronounced cy-per). A Shiny web applet (and R-package) that helps annotating the cluster identities in single-cell RNA-sequencing (scRNA-seq) experiments. The algorithm compares gene expression signature of experimental clusters with known reference datasets. In addition to 7 reference datasets implemented in CIPR (2 from mouse and 5 from human), users can upload custom high-throughput reference data for specialized studies. The CIPR pipeline can be further tailored to different analytical contexts by excluding irrelevant reference subsets and low-variance reference genes from the analysis. The manuscript describing CIPR and comparing its performance against other similar software was published in BMC Bioinformatics. CIPR's fast and computationally efficient calculations and graphical outputs will facilitate scRNA-seq analysis where the user wants to try different clustering parameters iteratively and examine the cluster identities. Source code for the Shiny and R-package implementations are available on GitHub.
- easybio - [R] - easybio is an R package for cell type annotation using the CellMarker2.0 database. bioRxiv
- Garnett - [R] - Garnett is a software package that facilitates automated cell type classification from single-cell expression data. Garnett works by taking single-cell data, along with a cell type definition (marker) file, and training a regression-based classifier. Once a classifier is trained for a tissue/sample type, it can be applied to classify future datasets from similar tissues. In addition to describing training and classifying functions, this website aims to be a repository of previously trained classifiers. Supervised Classification Enables Rapid Annotation of Cell Atlases
- scANVI - [python] - single-cell ANnotation using Variational Inference (scANVI) is a semi-supervised variant of scVI designed to leverage any available cell state annotations — for instance when only one data set in a cohort is annotated, or when only a few cells in a single data set can be labeled using marker genes. Harmonization and Annotation of Single-cell Transcriptomics data with Deep Generative Models
- SignacX - [R] - Signac classifies the cellular phenotype for each individual cell in scRNA-seq data using neural networks trained with sorted bulk gene expression data from the Human Primary Cell Atlas. Signac can: map cells from one data set to another, classify non-human single cell data, identify novel cell types, and classify single cell data across many tissues, diseases and technologies. Cell type classification and discovery across diseases, technologies and tissues reveals conserved gene signatures and enables standardized single-cell readouts

- singleCellNet - [R] - A near-universal step in the analysis of single cell RNA-Seq data is to hypothesize the identity of each cell. Often, this is achieved by finding cells that express combinations of marker genes that had previously been implicated as being cell-type specific, an approach that is not quantitative and does not explicitly take advantage of other single cell RNA-Seq studies. SingleCellNet, which addresses these issues and enables the classification of query single cell RNA-Seq data in comparison to reference single cell RNA-Seq data. [bioRxiv](#)
- SingleR - [R] - SingleR leverages reference transcriptomic datasets of pure cell types to infer the cell of origin of each of the single cells independently. [Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage](#). *Nature Immunology* (2019).
- scCATCH - [R] - A single cell cluster-based annotation package from cluster marker genes identification to cluster annotation based on evidence-based score by matching the identified potential marker genes with known cell markers in tissue-specific cell taxonomy reference database (CellMatch) [Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data](#). *iScience* (2020).
- DeepSort - [python] - A reference-free cell-type annotation tool for single-cell RNA-seq data using deep learning with a weighted graph neural network, which is learned based on the most comprehensive single-cell transcriptomics atlases involving 764,741 cells across 88 tissues of human and mouse. [bioRxiv](#)
- ImmClassifier - [R,python,Docker] - A cell type annotation algorithm that employs a knowledge-based approach to annotating cells based on their underlying ontology and multitudes of previously-published data. By encoding immune cell hierarchy in a neural network, ImmClassifier is able to identify fine-grained cell types with high accuracy. By running in Docker the tool is platform-agnostic. [bioRxiv](#)
- Celltypist - [Python] - Celltypist is an automated cell type annotation tool for scRNA-seq datasets on the basis of logistic regression classifiers optimized by the stochastic gradient descent algorithm. Celltypist provides several different models for predictions, with a current focus on immune sub-populations, in order to assist in the accurate classification of different cell types and subtypes.
- scPRINT - [python] - scPRINT is pretrained on 50M cells to predict multiple cell labels de novo, from any single cell RNAseq profile. [scPRINT: pre-training on 50 million cells allows robust gene network predictions](#)

## Doublet Identification

- AMULET - [shell, Python, R] - A count based method for detecting multiplets from single nucleus ATAC-seq (snATAC-seq) data. [Genome Biology](#)
- demuxlet - [shell] - [Multiplexed droplet single-cell RNA-sequencing using natural genetic variation](#)
- DoubletFinder - [R] - Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. [BioRxiv](#)

- [DoubletDecon](#) - [R] - Cell-State Aware Removal of Single-Cell RNA-Seq Doublets. [BioRxiv](DoubletDecon: Cell-State Aware Removal of Single-Cell RNA-Seq Doublets)
- [DoubletDetection](#) - [R, Python] - A Python3 package to detect doublets (technical errors) in single-cell RNA-seq count matrices. An [R implementation](#) is in development.
- [Scrublet](#) - [Python] - Computational identification of cell doublets in single-cell transcriptomic data. [BioRxiv](#)
- [solo](#) - [Python] - Doublet detection via semi-supervised deep learning.

## Cell subsampling

- [geosketch](#) - [Python] - Method to subsample massive scRNA-seq datasets while preserving rare cell states. Resulting “sketch” accelerates clustering, visualization, and integration analyses. [Paper](#)

## Feature (Gene) imputation

- [G2S3](#) - [R] - [G2S3: a Sparse and Smooth Signal of Gene Graph-based imputation method for scRNA-seq data](#)
- [MAGIC](#) - [R, Python, MATLAB] - Markov Affinity-based Graph Imputation of Cells (MAGIC). A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. On [BioRxiv](#) and [published in Cell](#).
- [NetDECODE](#) - [R] - We develop an algorithm, called DECODE, to assess the extent of joint presence/absence of genes across different cells. We show that this network captures biologically-meaningful pathways, cell-type specific modules, and connectivity patterns characteristic of complex networks. We develop a model that uses this network to discriminate biological vs. technical zeros, by exploiting each gene's local neighborhood. For non-biological zeros, we build a predictive model to impute the missing value using their most informative neighbors.
- [scImpute](#) - [R] - [scImpute: Accurate And Robust Imputation For Single Cell RNA-Seq Data](#)
- [VIPER](#) - [C++, R] - A fast and accurate tool to impute zero values in single-cell RNA sequencing studies to facilitate accurate transcriptome quantification at the single-cell level. VIPER is based on nonnegative sparse regression models and is capable of progressively inferring a sparse set of local neighborhood cells that are most predictive of the expression levels of the cell of interest for imputation. A key feature of VIPER is its ability to preserve gene expression variability across cells after imputation.  
<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1575-1>
- [scPRINT](#) - [python] - scPRINT is pretrained on 50M cells to denoise and perform zero imputation of any single cell RNAseq profile. [scPRINT: pre-training on 50 million cells allows robust gene network predictions](#)

## Copy number analysis

- [aneufinder](#) - [R] - Bioconductor module for copy-number detection in single-cell whole genome sequencing (scWGS) and strand-seq data using a Hidden Markov Model or binary bisection method.
- [CopyKAT](#) - [R] - Inference of genomic copy number and subclonal structure from scRNA-seq data. Outperforms *inferCNV*. [Paper](#)
- [Ginkgo](#) - [R, C] - Ginkgo is a web application for single-cell copy-number variation analysis.
- [HoneyBADGER](#) - [R] - HoneyBADGER identifies and infers the presence of CNV and LOH events in single cells and reconstructs subclonal architecture using allele and expression information from single-cell RNA-sequencing data.
- [inferCNV](#) - [R] - Part of the TrinityCTAT (Trinity Cancer Transcriptome Analysis Toolkit). Provides tools for copy-number inference from single-cell RNA-seq data.
- [inferCNVpy](#) - [Python] - A Python/Scanpy re-implementation of *inferCNV* . Significantly faster than the R version.
- [MEDALT](#) - [R, Python] - This package performs lineage tracing using copy number profile from single cell sequencing technology. It will infer: 1. An rooted directed minimal spanning tree (RDMST) to represent aneuploidy evolution of tumor cells. 2. The focal and broad copy number alterations associated with lineage expansion.
- [Numbat](#) - [R] - Numbat is a haplotype-aware CNV caller from single-cell and spatial transcriptomics data. It integrates signals from gene expression, allelic ratio, and population-derived haplotype information to accurately infer allele-specific CNVs in single cells and reconstruct their lineage relationship. [Paper](#)
- [SCEVAN](#) - [R] - Easy-to-use package that starting from the raw count matrix of scRNA data automatically classifies the cells present in the biopsy by segregating non-malignant cells of tumor microenvironment from the malignant cells, outperforms *copyKAT*. It also infers the copy number profile of malignant cells, identifies subclonal structures and automatically analyses the specific and shared alterations of each subpopulation. [Preprint](#)
- [SCICoNE](#) - [C++, Python] - Single-cell copy number calling and event history reconstruction. SCICoNE reconstructs the history of copy number events in the tumour and uses these evolutionary relationships to identify the copy number profiles of the individual cells.

## Variant calling

- [cb\\_sniffer](#) - [python] - Mutation barcode caller, calls mutant and ref barcodes from 10x single cell data.
- [cerebra](#) - [python] - Cerebra is a tool for high-throughput summarizing of vcf entries following traditional variant calling for a sequencing experiment. Helps to extract relevant mutation information from among tens of thousands of vcf lines.
- [monovar](#) - [python] - Monovar is a single nucleotide variant (SNV) detection and genotyping algorithm for single-cell DNA sequencing data. It takes a list of bam files as input and outputs a vcf file containing the detected SNVs.

- octopus - [C++] - Bayesian haplotype-based mutation calling with single-cell calling model. Identifies clonal and subclonal mutations using phylogeny inference and accounts for allelic dropout.
- SCIPhi - [python] - Single-cell mutation identification via phylogenetic inference (SCIPhi) is a new approach to mutation detection in individual tumor cells by leveraging the evolutionary relationship among cells.
- SComatic SComatic is a tool that provides functionalities to detect somatic single-nucleotide mutations in high-throughput single-cell genomics and transcriptomics data sets, such as single-cell RNA-seq and single-cell ATAC-seq. Paper
- SSrGE - [python] - SSrGE is an approach to identify SNVs correlated with Gene Expression using multiple regularized linear regressions. It contains its own pipeline to infer SNVs from scRNA-seq reads and is able to identify and sort genes and SNVs for a given cell subgroup. Nature Communication (2017) Using single nucleotide variations in single-cell RNA-seq to identify subpopulations and genotype-phenotype linkage.

## Epigenomics

- ArchR - [R] - ArchR is a full-featured R package for processing and analyzing single-cell ATAC-seq data. ArchR provides the most extensive suite of scATAC-seq analysis tools of any software available. ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis.
- ATACdemultiplex - [Go] - Suites of low-level multi-threaded utilities to efficiently manipulate large single-cell ATAC-Seq data (BAM, BED/fragments, Fastq files). Very efficient for creating sparse matrices, subset fragment/BED files, annotate peaks, estimate FDR corrected fisher features, create bigwig files and compute TSS enrichments (global and at the single-cell level).
- AtacWorks - [python] - AtacWorks is a deep learning tool to denoise and identify accessible chromatin regions from low-coverage, low cell count, or low-quality ATAC-seq data. AtacWorks can denoise signal and identify peaks from rare cellular subtypes in a mixed population. Biorxiv
- BIRD - [C++/R] - BIRD is a tool for predicting chromatin accessibility and inferring regulatory element activities in single cells using scRNA-seq. Global prediction of chromatin accessibility using small-cell-number and single-cell RNA-seq
- ChromA - [C++/Fortran] - Chromatin Accessibility Annotation Tool.
- ChromVAR - [R] - Determine variations in chromatin accessibility across sets of annotations or peaks. Designed primarily for single-cell or sparse chromatin accessibility data, e.g. from scATAC-seq or sparse bulk ATAC or DNase-seq experiments. BioRxiv
- cisTopic - [R/python] - A probabilistic framework used to simultaneously discover coaccessible enhancers and stable cell states from sparse single-cell epigenomics data. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data

- cicero - [R] - Predicts enhancer-gene pairs by co-accessibility. Also adapts monocle for single-cell ATAC-seq (clustering, trajectories, differential accessibility).
- DeepCpg - [python] - DeepCpG is a deep neural network for predicting the methylation state of CpG dinucleotides in multiple cells. It allows to accurately impute incomplete DNA methylation profiles, to discover predictive sequence motifs, and to quantify the effect of sequence mutations.
- EpiScanpy - [python] - EpiScanpy is the epigenomic extension of scRNA-seq analysis tool Scanpy. It analyses single-cell open chromatin (scATAC-seq) and single-cell DNA methylation (for example scBS-seq) data. EpiScanpy: integrated single-cell epigenomic analysis
- Enhlink - [Go/Binary] - Enhlink is a fast, easy to install, scalable, and robust computational approach that can infer linkages from high-dimensional, sparse, mono- or multi-omic single-cell datasets. Enhlink can be extended to infer distal, covariates, and clusters linkages. Compared to alternative methods such as Cicero, Archr, or Signac, Enhlink is more flexible, accurate and robust, and performs much faster. Enhlink can easily process data generated by the Cell Ranger pipelines, or any sparse matrices saved in the appropriate format. Enhlink infers distal and context-specific enhancer-promoter linkages
- Melissa - [R] - Melissa (MEthylation Lation Inference for Single cell Analysis), a Bayesian hierarchical method to quantify spatially-varying methylation profiles across genomic regions from single-cell bisulfite sequencing data (scBS-seq). Melissa clusters individual cells based on local methylation patterns, enabling the discovery of epigenetic differences and similarities among individual cells. The clustering also acts as an effective regularisation method for imputation of methylation on unassayed CpG sites, enabling transfer of information between individual cells. BioRxiv
- SCALE - [python] - SCALE is a deeplearning tool combining GMM with VAE for single-cell ATAC-seq analysis (visualization, clustering, imputation, batch effect removal, downstream analysis for celltype-specific TFs). SCALE method for single-cell ATAC-seq analysis via latent feature extraction
- SCATE - [R] - SCATE reconstructs activities of individual cis-regulatory elements (CREs) from single-cell ATAC-seq data by adaptively integrating information from co-activated CREs, similar cells, and publicly available regulome data. Single-cell ATAC-seq signal extraction and enhancement with SCATE
- scbs - [python] - A command line tool for the analysis of Single-Cell Bisulfite-Sequencing data. scbs makes it easy to obtain a cell×region methylation matrix (≈count matrix) from raw single-cell methylation files and enables efficient storage, quality control and visualization. Furthermore, scbs allows you to scan the entire genome for variably or differentially methylated regions (VMRs or DMRs), and implements a novel approach for more accurate quantification of methylation in genomic intervals. bioRxiv

- scE2G - [python] - Family of models to predict enhancer-gene regulation based on single-cell data. These models use features derived from either single-cell ATAC-seq or multiomicRNA and ATAC-seq data in supervised classifiers trained on a new harmonized CRISPR perturbation dataset including over 13,000 examples.
- SCRAT - [R] - SCRAT provides essential tools for users to read in single-cell regulome data (ChIP-seq, ATAC-seq, DNase-seq) and summarize into different types of features. It also allows users to visualize the features, cluster samples and identify key features. Single-cell regulome data analysis by SCRAT
- Signac - [R] - Signac is an extension of Seurat for the analysis, interpretation, and exploration of single-cell chromatin datasets.

## Multi-assay data integration

- bindSC - [R] - bindSC (Bi-dimensional INtegration of multi-omics Data from Single Cell sequencing technologies) is an R package for single cell multi-omic integration analysis, developed and maintained by Ken chen's lab in MDACC. bindSC is developed to address the challenge of single-cell multi-omic data integration that consists of unpaired cells measured with unmatched features across modalities. Bi-order multimodal integration of single-cell data
- CellWalkR - [R] - An R Package for integrating single-cell and bulk data to resolve regulatory elements.
- CITE-seq-Count - [python] Cite-seq-Count is a python package that deals with Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) and cell hashing data. CITE-seq is a multimodal single cell phenotyping method that allows for immunophenotyping of cells with a potentially limitless number of markers and unbiased transcriptome analysis using existing single-cell sequencing approaches.
- clonealign - [R] - Integrate single-cell RNA and single-cell DNA-seq measured in separate cells from the same tumour to infer cancer-clone-specific gene expression profiles clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers, Genome Biology 2019
- Cobolt - [python] - Cobolt is a novel method that not only allows for analyzing the data from joint-modality platforms, but provides a coherent framework for the integration of multiple datasets measured on different modalities. Cobolt: integrative analysis of multimodal single-cell sequencing data
- FigR - [R] - FigR (Functional inference of gene Regulation) is a computational framework for supporting the integration of single cell chromatin accessibility and gene expression data to infer transcriptional regulators of target genes. Functional inference of gene regulation using single-cell multi-omics
- gimVI - [python] - gimVI is a joint generative model for imputation of missing genes in spatial transcriptomics assay from unpaired scRNA-seq data. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements

- GLUE - [python] - GLUE (Graph-Linked Unified Embedding) is a deep learning method for unpaired single-cell multi-omics data integration and regulatory inference ([Paper](#)).
- LIGER - [R] - LIGER (Linked Inference of Genomic Experimental Relationships) is a package for integrating and analyzing multiple single-cell datasets. [Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity](#).
- MATCHER - [python] - [MATCHER: An algorithm for integrating single cell transcriptomic and epigenomic data using manifold alignment](#). MATCHER takes multiple types of single cell measurements performed on distinct single cells and infers single cell multi-omic profiles.
- MultiVI - [python] - MultiVI is a probabilistic framework that leverages deep neural networks to jointly analyze scRNA, scATAC and multiomic (scRNA + scATAC) data. [MultiVI: deep generative model for the integration of multi-modal data](#)
- MOFA - [python, R] - Multi-Omics Factor Analysis, a framework for unsupervised integration of multi-omics data sets. MOFA is a method for disentangling the different sources of heterogeneity in bulk and single-cell multi-omics data sets. It identifies the latent factors that drive unique and shared variability in the different assays. The factors can be used for visualisation, pseudotime reconstruction, imputation, among other functionalities. [Paper](#)
- SCALEX - [python] - SCALEX provides a VAE framework for integration of heterogeneous single-cell data by disentangling batch-invariant components from batch-related variations and projecting the batch-invariant components into a generalized, low-dimensional cell-embedding space. [Construction of continuously expandable single-cell atlases through integration of heterogeneous datasets in a generalized cell-embedding space](#)
- scarf - [python] - 🧑‍🔬 Toolkit for highly memory efficient analysis of single-cell RNA-Seq, scATAC-Seq and CITE-Seq data. Analyze atlas scale datasets with millions of cells on laptop. [Preprint](#)
- scDART - [python] - scDART is a deep learning framework that integrates scRNA-seq and scATAC-seq data and learns cross-modalities relationships simultaneously. [scDART: integrating unmatched scRNA-seq and scATAC-seq data and learning cross-modality relationship simultaneously](#).
- SISUA - [python] - In this study, we propose models based on the Bayesian generative approach, where protein quantification available as CITE-seq counts from the same cells are used to constrain the learning process, thus forming a semi-supervised model. The generative model is based on the deep variational autoencoder (VAE) neural network architecture. [bioRxiv](#)



- TotalVI - [python] - Total Variational Inference (totalVI) is a coupled generative model and inference procedure for CITE-seq data. TotalVI deals with modelisation of the background noise of protein measurements, harmonization of multiple CITE-seq experiments and imputation of missing proteins. [A Joint Model of RNA Expression and Surface Protein Abundance in Single Cells](#)

### Rare cell detection

- FiRE - [python, R, C++] - Finder of rare entities (FiRE) helps identify rare cell types in voluminous single-cell datasets. Design of FiRE is inspired by the observation that rareness estimation of a particular data point is the flip side of measuring the density around it. In principle, FiRE uses the Sketching technique, a variant of locality sensitive hashing, to assign rareness score to every cell. [Paper](#)

### Cellular interactions/communication

- CellPhoneDB - [python] - Publicly available repository of curated receptors, ligands and their interactions in human (subunit architecture is included for both ligands and receptors, representing heteromeric complexes accurately). [Paper](#)
- NicheNet - [R] - To study intercellular communication from a computational perspective. It uses human or mouse gene expression data of interacting cells as input and combines this with a prior model that integrates existing knowledge on ligand-to-target signaling paths. This allows to predict ligand-receptor interactions that might drive gene expression changes in cells of interest. [Paper](#)
- COMUNET - [python] - It streamlines the interpretation of the results from cell-cell communication analyses by using multiplex networks to represent and cluster all potential communication pathways between cell types. [Paper](#)
- CellChat - [R] - It predicts major signaling inputs and outputs for cells and how those cells and signals coordinate for functions using network analysis and pattern recognition approaches. Through manifold learning and quantitative contrasts, CellChat classifies signaling pathways and delineates conserved and context-specific pathways across different datasets. [Paper](#)
- LIANA - [R, python] - LIANA enables the use of any combination of ligand-receptor methods and resources, and their consensus. [Paper](#)

### Single cell large model

- geneformer - [Python] a single-cell large model training on 30 million human single-cell transcriptomics, supporting batch integration, gene dosage sensitivity predictions, chromatin dynamics prediction, network dynamics prediction, etc. manuscript open access: [Transfer learning enables predictions in network biology](#)
- scGPT - [Python] a single-cell large model training on 33 million human single-cell transcriptomics, supporting single-cell annotation, batch integration, perturbation prediction manuscript open access: [scGPT: toward building a foundation model for single-cell multi-omics using generative AI](#)

- scFoundation - [Python] a single-cell large model training on 50 million human single-cell transcriptomics with 100 million parameters, supporting single-cell clustering, drug response prediction, perturbation prediction, single-cell annotation, gene module inference, etc. manuscript open access: [Large-scale foundation model on single-cell transcriptomics](#)
- CellPLM - [Python] the first single-Cell Pre-trained Language Model that encodes cell-cell relations and it consistently outperforms existing pre-trained and non-pre-trained models in diverse downstream tasks, with 100x higher inference speed compared to existing pre-trained models, training on 9 million scRNA-seq cells and 2 million SRT cells. manuscript open access: [CellPLM: Pre-training of Cell Language Model Beyond Single Cells](#)

## Other applications

- BASIC - [python] - BASIC is a semi-de novo assembly method to determine the full-length sequence of the BCR in single B cells from scRNA-seq data.
- BPCells - [R] - A package for high performance single cell analysis of RNA-seq and ATAC-seq datasets. BPCells provides disk-backed streaming calculations that reduce memory requirements for normalization and PCA of large scRNA-seq datasets by nearly 100-fold compared to in-memory workflows such as Scanpy.
- CytoSpill - [R] - The goal of CytoSpill is to estimate and compensate spillover noises in CyTOF data, without requiring any training data.
- dropEst - [C++, R] - High-performance pipeline for initial analysis of droplet-based single-cell RNA-seq data (Drop-seq, inDrop, 10x and some others). Allows to estimate gene count matrix as well as diagnostic stats from fastq files with raw reads. Implements corrections for different noise sources.
- dropSeqPipe - [python, R, snakemake] - An automatic data handling pipeline for drop-seq/scrb-seq data. It runs from raw fastq.gz data until the final count matrix with QC plots along the way.
- ffq - [python] - Fetch run and metadata information for single-cell genomics datasets.
- gget - [Python] - `gget` is a free, open-source command-line tool and Python package that enables efficient querying of genomic databases. `gget` consists of a collection of separate but interoperable modules, each designed to facilitate one type of database querying in a single line of code.
- immunarch - [R] - R Package for Fast and Painless Exploration of Single-cell and Bulk T-cell/Antibody Immune Repertoires
- MetaNeighbor - [R] - [MetaNeighbor: a method to rapidly assess cell type identity using both functional and random gene sets](#)
- sasc - [C] - sasc stands for Simulated Annealing Single-Cell, an algorithm for performing phylogenetic analysis of single-cell cancer samples. Manuscript [here](#).
- scDataviz - [R] - scDataviz: single cell dataviz and downstream analyses, with a primary focus on flow and mass cytometry.

- [SCIFIL](#) - [Matlab] - SCIFIL: Single Cell Inference of Fitness Landscape is a computational method for in vivo inference of clonal selection and estimate of fitness landscapes of heterogeneous cancer cell populations from single cell sequencing data.
- [SCope](#) - [python] - SCoPe is a fast visualization tool for large-scale and high dimensional scRNA-seq datasets. Publication [here](#).
- [scTE](#) - [python] - Quantifying transposable element expression from single-cell sequencing data.
- [SCIFER](#) - [shell, python] - Approach for analysis of LINE-1 mRNA expression in single cells at a single locus resolution.
- [SiFit](#) - [Java] - [SiFit: A Method for Inferring Tumor Trees from Single-Cell Sequencing Data under Finite-site Models](#)
- [Sinto](#) - [python] - A toolkit for working with aligned single-cell reads. Includes functions to split BAM files by cell barcode, add cell barcodes as read tags, move cell barcode information to read groups, and create a scATAC-seq fragment file from a BAM file.
- [sircel](#) - [python] - sircel (pronounced "circle") separates reads in a fastq file based on barcode sequences that occur at known positions of reads. This is an essential first step in analyzing single-cell genomics data from experiments such as Drop-Seq. Barcode sequences often contain deletion and/or mismatch errors that arise during barcode synthesis and sequencing, and we have designed our barcode recovery approach with these issues in mind. In addition to identifying barcodes in an unbiased manner, sircel also quantifies their abundances. [doi](#)
- [Snakemake single-cell-rna-seq workflow](#) - [python, R, snakemake] - An automated pipeline for single cell RNA-seq analysis.
- [VisCello](#) - [R] - VisCello for C.elegans embryogenesis.
- [Wishbone](#) - [python] - [Wishbone is an algorithm to identify bifurcating developmental trajectories from single cell data. Wishbone can applied to both single cell RNA-seq and mass cytometry datasets.](#)

## **Spatial transcriptomics**

- [BayesSpace](#) - [R] A Bayesian statistical model for clustering and resolution enhancement of spatial gene expression experiments. manuscript open access: [Spatial transcriptomics at subspot resolution with BayesSpace](#)
- [CellTrek](#) - [R] CellTrek is a computational method to achieve single-cell spatial mapping through coembedding, random forest and metric learning approaches. manuscript open access: [Spatial charting of single-cell transcriptomes in tissues](#)
- [cell2location](#) - [Python] A Bayesian model that perform spatial deconvolution in SRT data and create cellular maps of diverse tissues based on negative binomial distribution. manuscript open access: [Cell2location maps fine-grained cell types in spatial transcriptomics](#)

- DestVI - [Python] A spatial deconvolution method designed with single-cell Latent Variable Model, scLVM(variational auto-encoder architecture). manuscript open access: [DestVI identifies continuums of cell types in spatial transcriptomics data](#)
- DSTG - [Python] A spatial deconvolution method designed with graph-based convolutional networks. manuscript open access: [DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence](#)
- Giotto Suite - [R] Flexible, composable, and technology-agnostic data framework in R, which allows the representation and integration of virtually any type of spatial omics data at any spatial resolution. manuscript open access: [Giotto Suite: a multi-scale and technology-agnostic spatial multi-omics analysis ecosystème](#)
- Merfishtools - [Python] - MERFISHtools implement a Bayesian framework for accurately predicting gene or transcript expression from MERFISH data.
- NMFreg - [Python] - The method is proposed in Slide-seq paper and reconstructs expression of each Slide-seq bead as a weighted combination of metagene factors, each corresponding to the expression signature of an individual cell type, defined from scRNA-seq.
- PASTE - [Python] A spatial alignment tool for aligning homogeneous spatial transcriptomic slices based on optimal transport and euclidean distance. manuscript open access: [Alignment and integration of spatial transcriptomics data](#)
- PASTE2 - [Python] A spatial alignment tool for aligning homogeneous spatial transcriptomic slices based on the partial extension of the Fused Gromov-Wasserstein optimal transport. manuscript open access: [PASTE2: Partial Alignment of Multi-slice Spatially Resolved Transcriptomics Data](#)
- RCTD - [R] A statistical model to deconvolute cell types of spatial spots based on scRNA-seq reference by poisson distribution and maximum likelihood estimation. manuscript open access: [Robust decomposition of cell type mixtures in spatial transcriptomics](#)
- SLAT - [Python] SLAT is to align both homogeneous and heterogeneous (the first work) single cell spatial omics data by employing a graph alignment framework consists of LGCN and adversarial discriminator. manuscript open access: [Spatial-linked alignment tool \(SLAT\) for aligning heterogenous slices](#)
- SpaGCN - [Python] SpaGCN is a graph convolutional network to integrate gene expression and histology to identify spatial domains and spatially variable genes. manuscript open access: [SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network](#)
- SpaTalk - [R] - SpaTalk is a cell-cell communication inference method for either single-cell or spot-based spatially resolved transcriptomic data, e.g., STARmap, MERFISH, seqFISH, Slide-seq, 10X Visium.

- SpatialDe - [Python] - SpatialDE is a statistical test to identify genes with spatial patterns of expression variation from multiplexed imaging or spatial RNA-sequencing data.
- SpatialDWLS - [R] A method to identify the cell types at each location with Giotto and determine the cell type composition using dampened weighted least squares. manuscript open access: SpatialDWLS: accurate deconvolution of spatial transcriptomic data
- SpatialPrompt - [Python] SpatialPrompt is a spot deconvolution and domain identification method for spatially resolved transcriptomics datasets. Main advantage of this tool is, it is highly scalable for large datasets.
- spatialGE - [R] An analysis suite allowing users to study spatial transcriptomics data from multiple platforms (e.g., Visium, CosMx). The package includes methods for pre-processing, clustering/domain detection, spatially variable genes, and functional analysis via the detection of gene expression gradients and/or gene set enrichment spatial patterns.
- Splotch - [Python] Splotch is a hierarchical generative probabilistic model for analyzing Spatial Transcriptomics data.
- SPOTlight - [R] SPOTlight enables the deconvolution of SRT data from a single-cell reference by a non-negative matrix factorization regression(NMFreg) model. manuscript open access: SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes
- squidpy - [Python] - Squidpy is a Python package for the analysis and visualization of spatial molecular data. It provides tools to process, analyze, and visualize spatial transcriptomics data, including spatially resolved transcriptomics and spatial proteomics. Squidpy: a scalable framework for spatial omics data analysis.
- Starspace - [Python] - Defines a schema for gene or protein expression data containing spatially localized information. Converts data from a variety of assay types, including Spatial Transcriptomics, CODEX, In-situ Sequencing, MERFISH, osmFISH, and starMAP. Demonstrates how to visualize and interact with these data using common analysis packages, and convert the formats into loom and anndata objects, for downstream analysis in R and Python.
- STAGATE - [Python] STAGATE is designed for spatial clustering and denoising expressions of spatial resolved transcriptomics (ST) data by learning low-dimensional latent embeddings with both spatial information and gene expressions via a graph attention auto-encoder(GATE). manuscript open access: Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder

- [STAligner](#) - [Python] STAligner is designed for alignment and integration of spatially resolved transcriptomics data, it employs a graph attention auto-encoder neural network(GATE) to extract spatially aware embedding and introduces the triplet loss to update the spot embedding to reduce the distance from the anchor to positive spot. manuscript open access: [STAligner enables the integration and alignment of multiple spatial transcriptomics datasets](#)
- [Tangram](#) - [Python] Tangram is used to map single-cell (or single-nucleus) gene expression data onto spatial gene expression data designed with optimizing a specially designed mapping objective loss. manuscript open access: [Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram](#)

## Tutorials and workflows

- [Analysis of single cell RNA-seq data](#) - [R and Python] - The [course](#) is taught through the University of Cambridge Bioinformatics training unit, but the material found on these pages is meant to be used for anyone interested in learning about computational analysis of scRNA-seq data.
- [Aaron Lun's Single Cell workflow on Bioconductor](#) - [R] - This article describes a computational workflow for basic analysis of scRNA-seq data using software packages from the open-source Bioconductor project.
- [ATAC-Seq Pipeline](#) - [Shell and R] - [Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors](#).
- [Bioconductor2016 Single-cell-RNA-sequencing workshop by Sandrine Dudoit lab](#) - [R] - SCONE, clusterExperiment, and slingshot tutorial.
- [BiomedCentral Single Cell Omics collectin](#) - collection of papers describing techniques for single-cell analysis and protocols.
- [Clustering 3K PBMCs with Scanpy in Galaxy](#) - Galaxy Training Material.
- **CRUK CI** [Introduction to single-cell RNA-seq data analysis](#) | [website](#)
- [CSHL Single Cell Analysis - Bioinformatics](#) course materials - Uses Shalek 2013 and Macaulay 2016 datasets to teach machine learning to biologists
- [CytoTOF Workflow](#) - [R] - An R-based pipeline for differential analyses of high dimensional mass cytometry data, primarily based on Bioconductor packages. [Link](#) to the paper describing a high-level introduction to the core concepts and code.
- **Dana-Farber Cancer Institute** [Trajectory inference across conditions: differential expression and differential progression](#) | [website](#)
- **Dan Beiting** [DIY Transcriptomics](#) | [website](#) - A hybrid course covering best practices for bulk and single cell RNA-seq data analysis, with a primary focus on empowering students to be independent in the use of lightweight and open-source software and the R/bioconductor environment.
- [EBI Single cell RNA-seq tutorial](#) | [website](#)
- [ELIXIR EXCELERATE Single RNA-seq data analysis with R](#) | [website](#)

- [Festival of Genomics California Single Cell Workshop](#) - [R] - Explores basic workflow from exploratory data analysis to normalization and downstream analyses using a dataset of 1679 cells from the Allen Brain Atlas.
- [Gilad Lab Single Cell Data Exploration](#) - R-based exploration of single cell sequence data. Lots of experimentation.
- [GPU accelerated single-cell analysis using RAPIDS](#) - NVIDIA tutorials on using RAPIDS (<https://www.rapids.ai/>) to accelerate single-cell analysis on GPUs.
- [\*\*Harvard Chan Bioinformatics Core\*\* Single-cell RNA-seq data analysis workshop | website](#)
- [Harvard STEM Cell Institute Single Cell Workshop 2015](#) - workshop on common computational analysis techniques for scRNA-seq data from differential expression to subpopulation identification and network analysis. [See course description for more information](#)
- [Hemberg Lab scRNA-seq course materials](#)
- [kallistobustools](#) - kallisto | bustools workflow for pre-processing single-cell RNA-seq data.
- [\*\*NBIS\*\* Single cell RNA-seq analysis workshop | website](#)
- [\*\*MGC/BioSB\*\* Course - Single Cell Analysis](#)
- [nf-core/scrnaseq](#) - nf-core/scrnaseq is a bioinformatics best-practice analysis pipeline for processing 10x Genomics single-cell RNA-seq data. The pipeline is built using Nextflow, a workflow tool to run tasks across multiple compute infrastructures in a very portable manner. It uses Docker/Singularity containers making installation trivial and results highly reproducible.
- [nf-core/scflow](#) - nf-core/scflow is a bioinformatics pipeline for scalable, reproducible, best-practice analyses of single-cell/nuclei RNA-sequencing data. The pipeline is built using Nextflow, a workflow tool to run tasks across multiple compute infrastructures in a very portable manner.
- [nf-core/scnanoseq](#) - nf-core/scnanoseq is a Nextflow analysis pipeline for processing 10X Genomics single-cell/nuclei RNA-seq data derived from the Oxford Nanopore long-read sequencer. The pipeline has been designed to be scalable to large datasets (including PromethION data), portable and reproducible.
- [Orchestrating Single-Cell Analysis with Bioconductor](#) - [R] - This blogdown book describes a comprehensive and reproducible workflow for the analysis of single-cell RNA-sequencing data.
- [Pre-processing of 10X Single-Cell RNA Datasets in Galaxy](#) - Galaxy Training Material.
- [Theis Lab Single Cell Tutorial](#) - The main part of this repository is a case study where the best-practices established in the manuscript are applied to a mouse intestinal epithelium regions dataset from Haber et al., Nature 551 (2018) available from the GEO under GSE92332.

- [Using Seurat \(v1.2\) for unsupervised clustering and biomarker discovery](#) - 301 single cells across diverse tissues from (Pollen et al., Nature Biotechnology, 2014). Original tutorial using Seurat 1.2
- [Using Seurat \(v1.2\) for spatial inference in single-cell data](#) - 851 single cells from zebrafish embryogenesis (Satija\*, Farrell\* et al., Nature Biotechnology, 2015). Original tutorial using Seurat 1.2
- [Seurat \(v3.0\) - Guided Clustering Tutorial](#) - new tutorial using Seurat 3.0
- [SIB Single-cell Transcriptomics | website](#)
- [SIB NBIS/SciLifeLab Advanced topics in Single Cell Omics | website](#)
- [SIB Advanced topics in single-cell transcriptomics](#)
- [WEHI Single cell RNA-seq analysis workshop | website](#)
- [Wellcome Sanger Institute Analysis of single cell RNA-seq data | website](#)

## Web portals, apps, and databases

### Web portals and databases

- [10X Genomics datasets](#) - 10x genomics public datasets, including 1.3M cell mouse brain dataset.
- [ASAP](#) - Automated Single-cell Analysis Pipeline (deposited in [BioRxiv](#) on December 22, 2016).
- [cellBrowser](#) - [Python, Javascript] Python pipeline and Javascript scatter plot library for single-cell datasets. [Demo](#)
- [CellView](#) - CellView is an R Shiny web application that allows knowledge-based and hypothesis-driven exploration of processed single cell transcriptomic data. [ref.](#)
- [Cell BLAST](#) - A Web portal powered by Cell\_BLAST (scRNA-seq querying tool) and ACA (scRNA-seq database).
- [CELLxGENE](#) - CELLxGENE is a suite of tools that help scientists to find, download, explore, analyze, annotate, and publish single cell datasets. It includes several powerful tools with various features to help you to engage with single cell data.
- [conquer](#) - A repository of consistently processed, analysis-ready single-cell RNA-seq data sets.
- [Curated Database of single-cell studies](#) - Available as a `tsv` download. Over 500 single cell transcriptomics studies have been published to date. Many of these have data available, but the links between data, study, and systems studied can be hard to identify through literature search. This manuscript describes a nearly exhaustive and manually curated database of single cell transcriptomics studies with descriptions of what kind of data and what biological systems have been studied. [bioRxiv.](#)
- [D<sup>3</sup>E](#) - Discrete Distributional Differential Expression (D<sup>3</sup>E) is a tool for identifying differentially-expressed genes, based on single-cell RNA-seq data.




- [dseqr](#) - Dseqr runs end-to-end multi-sample single-cell and bulk RNA-seq analyses using a user friendly web app built around best practices from the OSCA handbook. Features include pseudobulk differential expression analysis, automated cluster annotation, reference mapping with Azimuth, Gene Ontology analysis, and drug connectivity mapping. Projects can either be analysed online or locally using the [dseqr R package](#).
- [EBI Single Cell Expression Atlas](#) - The Single Cell Expression Atlas contains uniformly re-analysed single cell expression data across different species and provides interactive visualizations to explore that data.
- [Galaxy Single Cell Omics Workbench](#) - dedicated Galaxy server for analyzing single cell data.
- [IRIS3](#) - IRIS3 (integrated cell-type-specific regulon inference server from single-cell RNA-Seq) is an easy-to-use server empowered by over 20 functionalities to support comprehensive interpretations and graphical visualizations of identified cell-type-specific regulons.
- [JingleBells](#) - A repository of standardized single cell RNA-Seq datasets for analysis and visualization in IGV at the single cell level. Currently focused on immune cells (<http://www.jimmunol.org/content/198/9/3375.long>).
- [SCPortalen](#) - SCPortalen: human and mouse single-cell centric database. [ref](#)
- [scRNA.seq.datasets](#) - Collection of public scRNA-Seq datasets used by [Hemberg Lab](#)
- [scRNASeqDB](#) - A database aggregating human single-cell RNA-seq datasets. [ref](#)
- [Single Cell Portal](#) - The Single-Cell Portal was developed to facilitate open data and open science in Single-cell Genomics. The portal currently focuses on sharing scientific results interactively, and sharing associated datasets.
- [Single-Cell Tumor Immune Atlas project](#) - [R] - We generated a [single-cell tumor immune atlas](#), jointly analyzing >500,000 cells from 217 patients and 13 cancer types, providing the basis for a patient stratification based on immune cell compositions.
- [V-SVA](#) - An R Shiny application for detecting and annotating hidden sources of variation in single cell RNA-seq data.
- [WOT](#) - Waddington Optimal Transport (wot) uses time-course data to infer how the probability distribution of cells in gene-expression space evolves over time, by using the mathematical approach of optimal transport.

## **Interactive visualization and analysis**

- [Asc-Seurat](#) - [R, Docker] - Asc-Seurat is a web application based on [Shiny](#). Pronounced as "ask Seurat", it provides an easy-to-install and easy-to-use interface that allows the execution of all steps necessary for scRNA-seq analysis. It integrates many of the capabilities of the [Seurat](#) and [Dynverse](#) and also allows an instantaneous functional annotation of genes of interest using [BioMart](#).

- Cellar - [Python] - Cellar is an easy to use, interactive, and comprehensive software tool for the assignment of cell types in single-cell studies. It supports preprocessing, dimensionality reduction, clustering, differential expression & enrichment analysis, spatial transcriptomics, label transfer, semi-supervised clustering and more. A live web server running Cellar is available [here](#). [NatureComm](#).
- cellBrowser - [Python, Javascript] Python pipeline and Javascript scatter plot library for single-cell datasets. [Demo](#)
- CellView - CellView is an R Shiny web application that allows knowledge-based and hypothesis-driven exploration of processed single cell transcriptomic data. [ref](#).
- cellxgene VIP - cellxgene VIP is a web-based interactive tool built upon cellxgene but greatly extended its plotting and analytical capabilities by integrating state-of-the-art tools in this field. It allows users with no programming experience to rapidly explore scRNA-seq data and create high-resolution figures commonly seen in high-profile publications. Furthermore, it is the first tool to the author's knowledge to allow computational biologists to write their own code to communicate with the hosting server via a mini Jupyter notebook like interface. It opens up unlimited capabilities even beyond the rich set of plotting functions provided in the tool. [ref](#).
- Cerebro - [R] - Cerebro (cell report browser), a Shiny- and Electron-based standalone desktop application for macOS and Windows, which allows investigation and inspection of pre-processed single-cell transcriptomics data without requiring bioinformatic experience of the user. Through an interactive and intuitive graphical interface, users can i) explore similarities and heterogeneity between samples and cells clusters in 2D or 3D projections such as t-SNE or UMAP, ii) display the expression level of single genes or genes sets of interest, iii) browse tables of most expressed genes and marker genes for each sample and cluster. [bioRxiv](#)
- ChromSCape - [R] - Interactive & complete analysis of single-cell epigenomic landscapes with Shiny. Includes counting, QC, filtering, dimensionality reduction, clustering, visualisation, coverage, peak calling, differential & gene set analysis) - (scChIP-seq, scCUT&TAG, scATAC-seq, scChIC-seq...). ([Prompsy et al. 2020](#))
- CirroCumulus - CirroCumulus is an interactive visualization tool for large-scale single-cell genomics (e.g. sc/snRNA-seq, spatial) data.
- CReSCENT - [R, Javascript, Python] - CReSCENT: Cancer Single Cell Expression Toolkit ([Mohanraj et al. 2020](#)), is an intuitive and scalable web portal incorporating a containerized pipeline execution engine for standardized analysis of cancer scRNA-seq data and associated metadata. CReSCENT uses public data sets and preconfigured pipelines that are accessible to computational biology non-experts and are user-editable to allow optimization, comparison, and reanalysis for specific experiments. Users can also upload their own scRNA-seq data for analysis and results can be kept private or shared with other users.

- FASTGenomics - [Python, R] - FASTGenomics is an online platform to share single-cell RNA sequencing data and analyses using reproducible workflows. Gene expression data can be shared meeting European data protection standards (GDPR). FASTGenomics enables the user to upload their own data and generate customized and reproducible workflows for the exploration and analysis of gene expression data ([Scholz et al. 2018](#)). Follow us on [Twitter](#).
- Ginkgo - [R, C] - Ginkgo is a web application for single-cell copy-number variation analysis and visualization.
- Granatum - Granatum  is a graphical single-cell RNA-seq (scRNA-seq) analysis pipeline for genomics scientists. [Published in December 2017](#).
- histoCAT [MATLAB]- Histology Topography Cytometry Analysis Toolbox (histoCAT) is a package to visualize and analyse highly multiplexed image cytometry data. [histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data](#)
- InterCellar - [R] - an R/Shiny app for interactive analysis and exploration of cell-cell communication based on single-cell transcriptomics data. Starting from pre-computed ligand-receptor interactions, InterCellar provides filtering options, annotations and multiple visualizations to explore cell clusters, genes and functions. Moreover, based on functional annotation from Gene Ontology and pathway databases, InterCellar implements data-driven analyses to investigate cell-cell communication in one or multiple conditions. Every step of the analysis can be performed interactively, thus not requiring any programming skills. ([Interlandi et al. 2022](#)).
- iS-CellR - iS-CellR (Interactive platform for Single-cell RNAseq) is a web-based Shiny app that integrates the Seurat package with Shiny's reactive programming framework to provide comprehensive analysis and interactive visualization of single-cell RNAseq data. [Paper](#)
- iSEE - [R] - iSEE, interactive SummarizedExperiment Explorer. The iSEE package aims to provide an interactive user interface for exploring data in objects derived from the SummarizedExperiment class. Particular focus will be given to single-cell data in the SingleCellExperiment derived class. The interface is implemented with RStudio's Shiny, with a multi-panel setup for ease of navigation. Features include: dynamically linked charts, support for reproducibility by recording the exact code for every output, as well as guided tours to learn step-by-step the salient features of the user interface and of the data. A demo instance of the app is available at this address: <http://shiny.imbei.uni-mainz.de:3838/iSEE>.
- Millefy - [R] - An R package and a Docker image with JupyterLab for visualizing read coverage of scRNA-seq datasets in genomic contexts. By dynamically and automatically reorder single cells based on 'locus-specific' pseudotime, Millefy highlights cell-to-cell heterogeneity in read coverage of scRNA-seq data. [paper](#).

- [NASQAR](#) - Nucleic Acid SeQuence Analysis Resource, a web-based platform that provides an intuitive interface for popular tools (like DESeq2, Seurat, and others) to perform standard downstream analysis workflows for RNAseq data. The portal hosts a number of R Shiny apps.
- [PIVOT](#) - Platform for Interactive analysis and Visualization Of Transcriptomics data. [ref](#)
- [scClustViz](#) - An interactive R Shiny tool for visualizing single-cell RNAseq clustering results from common analysis pipelines (SingleCellExperiment or Seurat, currently). Its main goal is two-fold: A: to help select a biologically appropriate resolution or K from clustering results by assessing differential expression between the resulting clusters; and B: help annotate cell types and identify marker genes. [See the demo app here!](#) scClustViz can also be used to generate R data packages for sharing published data - see the website for details and a list of published datasets.
- [scRNAseqApp](#) - The scRNAseqApp is a Shiny app package designed for interactive visualization of single-cell data. It is an enhanced version derived from the [ShinyCell](#), repackaged to accommodate multiple datasets. The app enables users to visualize data containing various types of information simultaneously, facilitating comprehensive analysis. Additionally, it includes a user management system to regulate database accessibility for different users.
- [SeuratWizard](#) - a web-based (wizard style) interactive R Shiny application to perform guided single-cell RNA-seq data analysis and clustering. [demo](#)
- [SeuratV3Wizard](#) - a web-based (wizard style) interactive R Shiny application to perform guided single-cell RNA-seq data analysis and clustering based on Seurat v3. [demo](#)
- [ShinyArchRUiO](#) - [R] - Shiny based web app for visualization of single-cell ATAC-seq data using ArchR. [ShinyArchR.UiO: user-friendly, integrative and open-source tool for visualization of single-cell ATAC-seq data using ArchR](#)
- [ShinyCortex](#) - a resource that brings together data from recent scRNA-seq studies of the developing cortex for further analysis. ShinyCortex is based in R and displays recently published scRNA-seq data from the human and mouse cortex in a comprehensible, dynamic and accessible way, suitable for data exploration by biologists. [paper](#)
- [singleCellTK](#) - The singleCellTK is an R/Shiny package and GUI for analyzing and visualizing scRNA-Seq through a web interface. Analysis modules include data summary and filtering, dimensionality reduction and clustering, batch correction, differential expression analysis, pathway activity analysis, and power analysis. [Comprehensive generation, visualization, and reporting of quality control metrics for single-cell RNA sequencing data.](#)

- [spatialGE-web](#) - A web application providing point-and-click access to the methods in the [spatialGE R package](#) and other tools (including STdeconvolve, InSituType, SpaGCN). The web application requires no coding experience. User accounts can be created to safely keep samples and results organized within projects.
- [STREAM](#) - STREAM is an interactive computational pipeline for reconstructing complex cellular developmental trajectories from sc-qPCR, scRNA-seq or scATAC-seq data. [preprint](#).
- [Vitessce](#) - [JavaScript, Python, R] - A framework for integrative visualization of multi-modal single-cell data, supporting microscopy images, cell segmentations, dimensionality reduction scatterplots, gene expression heatmaps, and genome browser tracks. Vitessce is packaged as a [React component](#), [Jupyter widget](#), and [R htmlwidget](#).
- [V-SVA](#) - An R Shiny application for detecting and annotating hidden sources of variation in single cell RNA-seq data.

## Journal articles of general interest

### Paper collections

- [Mendeley Single Cell Sequencing Analysis](#)
- [BioMedCentral Single-Cell -omics collection](#)
- [Single-Cell Genomics in the Journal Science](#) - Special issue on Single-Cell Genomics
- [The emerging field of single-cell analysis](#) - Special issue on single cell analysis

### Big data approach overview

- [Single-cell Transcriptome Study as Big Data](#)

### Experimental design

- [Design and computational analysis of single-cell RNA-sequencing experiments](#)
- [How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives](#)
- [Sensei. How to determine the number of patients required to ascertain a cell-type abundance change estimated from scRNA-seq experiment.](#)

### Methods comparisons

- [Comparative analysis of single-cell RNA sequencing methods](#) - a comparison of wet lab protocols for scRNA sequencing.
- [Comparison of computational methods for imputing single-cell RNA-sequencing data](#) - We compared eight imputation methods, evaluated their power in recovering original real data, and performed broad analyses to explore their effects on clustering cell types, detecting differentially expressed genes, and reconstructing lineage trajectories in the context of both simulated and real data. Simulated datasets and case studies highlight that there are no one method performs the best in all the situations.

- [Comparison of methods to detect differentially expressed genes between single-cell populations](#) - comparison of five statistical methods to detect differentially expressed genes between two distinct single-cell populations.
- [Bias, Robustness And Scalability In Differential Expression Analysis Of Single-Cell RNA-Seq Data](#) - comparison of 36 statistical methods to detect differentially expressed genes between two annotated populations from the [conquer](#) database of consistently processed scRNA-seq datasets.
- [Single-Cell RNA-Sequencing: Assessment of Differential Expression Analysis Methods](#) - an assessment of main bulk and single-cell differential analysis methods used to analyze scRNA-seq data.
- [A comparison of single-cell trajectory inference methods](#) - Unsure which of the more than 70 **trajectory inference** methods to use for your single-cell dataset? We evaluated 45 methods based on four criteria: the accuracy of the trajectory, how scalable the method is, how stable its outputs are, and the usability of the tool. These are summarised in a "*funky heatmap*" (Figures 2 & 3). Check out [dynverse.org](#) for more information.
- [Evaluation of methods to assign cell type labels to cell clusters from single-cell RNA-sequencing data](#) - In this study, we benchmarked five methods (CIBERSORT, GSEA, GSVA, ORA and METANEIGHBOR) for the task of assigning cell type labels to cell clusters from scRNA-seq data. We used five scRNA-seq datasets: human liver, 11 Tabula Muris mouse tissues, two human peripheral blood mononuclear cell datasets, and mouse retinal neurons, for which reference cell type signatures were available. Our results show that, in general, all five methods perform well in the task as evaluated by receiver operating characteristic curve analysis (average area under the curve (AUC) = 0.91, sd = 0.06), whereas precision-recall analyses show a wide variation depending on the method and dataset (average AUC = 0.53, sd = 0.24). GSVA was the overall top performer and was more robust in cell type signature subsampling simulations, although different methods performed well using different datasets. METANEIGHBOR and GSVA were the fastest methods.
- [Evaluation of single-cell classifiers for single-cell RNA sequencing data sets](#) - In this article, nine tools have been systematically compared. The article provides a guideline for researchers to select and apply suitable single cell and cluster classification tools in their analysis workflows and sheds some lights on potential direction of future improvement on classification tools.
- [Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data](#) - a comparison of gene regulatory network inference methods using simulated and real single-cell RNA-seq datasets

## Similar lists and collections

- [CrazyHotTommy's RNA-seq analysis list](#) - Very broad list that includes some single cell RNA-seq packages and papers.

- [Museum of Spatial Transcriptomics](#) - A comprehensive catalog of spatial transcriptomics data sets and methods.
- [scRNA-tools.org](#) - Database of scRNA-seq analysis tools and their functions. Managed through this [Github repository](#).
- [agitter's Pseudotime estimation list](#) - An overview of algorithms for estimating pseudotime in single-cell RNA-seq data.

## People

Gender bias at conferences is a well known problem

(<http://www.sciencemag.org/careers/2015/07/countering-gender-bias-conferences>).

Creating a list of potential speakers can help mitigate this bias and a community of people developing and maintaining helps to further diversify this list beyond smaller networks.

## Female

- [Rhonda Bacher \(University of Wisconsin-Madison, USA\)](#).
- [Barbara Di Camillo \(Information Engineering Department, University of Padova, Italy\)](#).
- [Jinmiao Chen \(Singapore Immunology Network, A\\*STAR, Singapore\)](#).
- [Jean Fan \(Johns Hopkins University, USA\)](#).
- [Brooke Fridley \(Children's Mercy Hospital, USA\)](#).
- [Sandrine Dudoit \(UC Berkeley, USA\)](#).
- [Lana X. Garmire, \(University of Hawaii Cancer Center, USA\)](#).
- [Laleh Haghverdi \(EMBL, Germany\)](#).
- [Stephanie Hicks \(Johns Hopkins Bloomberg School of Public Health, USA\)](#).
- [Christina Kendzioriski \(University of Wisconsin-Madison, USA\)](#).
- [Keegan Korthauer \(Dana Farber Cancer Institute, USA\)](#).
- [Smita Krishnaswamy \(Yale University\)](#).
- [Ning Leng \(Morgridge Institute for Research, USA\)](#).
- [Elisabetta Mereu \(Centre for Genomic Regulation, Barcelona\)](#).
- [Samantha Morris \(Depts of Dev. Bio. and Genetics, Washington University, St. Louis\)](#).
- [Alicia Oshlack \(Murdoch Children's Research Institute, Australia\)](#).
- [Dana Pe'er \(Memorial Sloan Kettering Cancer Center, USA\)](#).
- [Emma Pierson \(Stanford University, USA\)](#).
- [Aviv Regev \(Broad Institute, USA\)](#).
- [Sarah Snelling \(University of Oxford, UK\)](#).
- [Charlotte Soneson \(Institute of Molecular Life Sciences, University of Zurich\)](#).
- [Sarah Teichmann \(Wellcome Trust Sanger Institute, UK\)](#).
- [Barbara Treutlein \(ETH Zurich, CH\)](#).
- [Catalina Vallejos \(The Alan Turing Institute & UCL, UK\)](#).
- [Sanja Vickovic \(New York Genome Center & Columbia University, USA\)](#).

## Male

- Stein Aerts (KU Leuven Center for Human Genetics, Belgium).
- Ahmet Coskun (Georgia Tech, USA).
- Bart DePlancke (EPFL, School of Life sciences, Institute of Bioengineering, Switzerland).
- Raphael Gottardo (Fred Hutchinson Cancer Research Center, USA).
- Chung Chau Hon (RIKEN Centre for Integrative Medical Sciences, Yokohama).
- Yanxiang Deng (University of Pennsylvania, USA).
- Martin Hemberg (Sanger Institute, UK).
- John Hickey (Duke University, USA).
- Holger Heyn (Centre for Genomic Regulation, Barcelona).
- Peter Kharchenko (Department of Biomedical Informatics, Harvard Medical School, USA).
- Sten Linnarson (Karolinska Institutet, Sweden).
- Aaron Lun (Cancer Research UK, UK).
- John Marioni (EBI, UK).
- Davis McCarthy (EBI, UK).
- John Reid (MRC Biostatistics Unit, Cambridge University, UK).
- Mark Robinson (Institute of Molecular Life Sciences, University of Zurich).
- Yvan Saeys (Vlaams Instituut voor Biotechnologie, Ghent, Belgium).
- Rickard Sandberg (Karolinska Institutet, SE).
- Neville Sanjana (New York Genome Center & NYU).
- Rahul Satija (New York Genome Center).
- Peter Sims (Columbia University, Department of Systems Biology).
- Oliver Stegle (EBI, UK).
- Fabian Theis (Institute of Computational Biology, Helmholtz Zentrum München).
- Cole Trapnell (University of Washington, Department of Genome Sciences).
- Itai Yanai (New York University, School of Medicine, Institute for Computational Medicine, USA).