User guide of CIRCLET

Abstract

CIRCLET is a powerful and robust circular trajectory reconstruction tool without specifying a starting cell for resolving cell cycle phases of single cells by considering multi-scale features of chromosomal architectures. CIRCLET reveals its best superiority based on the combination of a feature set about global information and two feature sets about local interactional information in terms of designed evaluation indexes and verification strategies from a collection of cell cycle Hi-C maps. For more information about CIRCLET, explore our publication.

Instruction

- CIRCLET consists six key steps **(Figure 1)**:

(1) Extracting feature: multi-scale feature sets are extracted from single-cell Hi-C maps.

(2) Reducing feature dimensions: the dimension of these feature sets are further reduced to a low n-dimensional space via diffusion maps (e.g., 2 dimension as an example).

(3) Constructing a KNN graph: CIRCLET constructs a k-nearest-neighbor graph in the n-dimensional embedded space and selects a set of cells called “waypoints”, one of which is randomly selected as the starting cell s.

(4) Computing initial ordering: an initial ordering of cells is obtained by the shortest path distance from s (e.g. distance $D_{s,t}$ marked by red solid line from s to t cell).

(5) Detecting the orientation and refining the ordering: CIRCLET computes a perspective matrix $\mathbf{P}$, which records the shorting path distance of each cell to the starting cell s from the viewpoint of waypoints (e.g. the distance of cell t to s from the viewpoint of w_1 is $P_{w,t} = D_{s,w_1} + D_{w_1,t}$). These waypoints’ perspective is firstly used to identify the clockwise (CW) or counterclockwise (CCW) semicircle of cells from s.

(6) Obtaining final trajectory: CIRCLET iteratively executes the step (5) until convergence, eventually obtaining a high-resolution cell-cycle trajectory.

- CIRCLET can help you perform two main types of circular trajectory reconstruction:

1) CICRLET helps to reconstruct a cell-cycle trajectory from single cell Hi-C maps, and use the designed indexes to evaluate and visualize the results.
2) The part of CIRCLET also helps to reconstruct a cell-cycle trajectory from single cell RNA-seq dataset, and use the designed indexes to evaluate and visualize the results.

- Download package:
  Download page – Get the latest version of CIRCLET

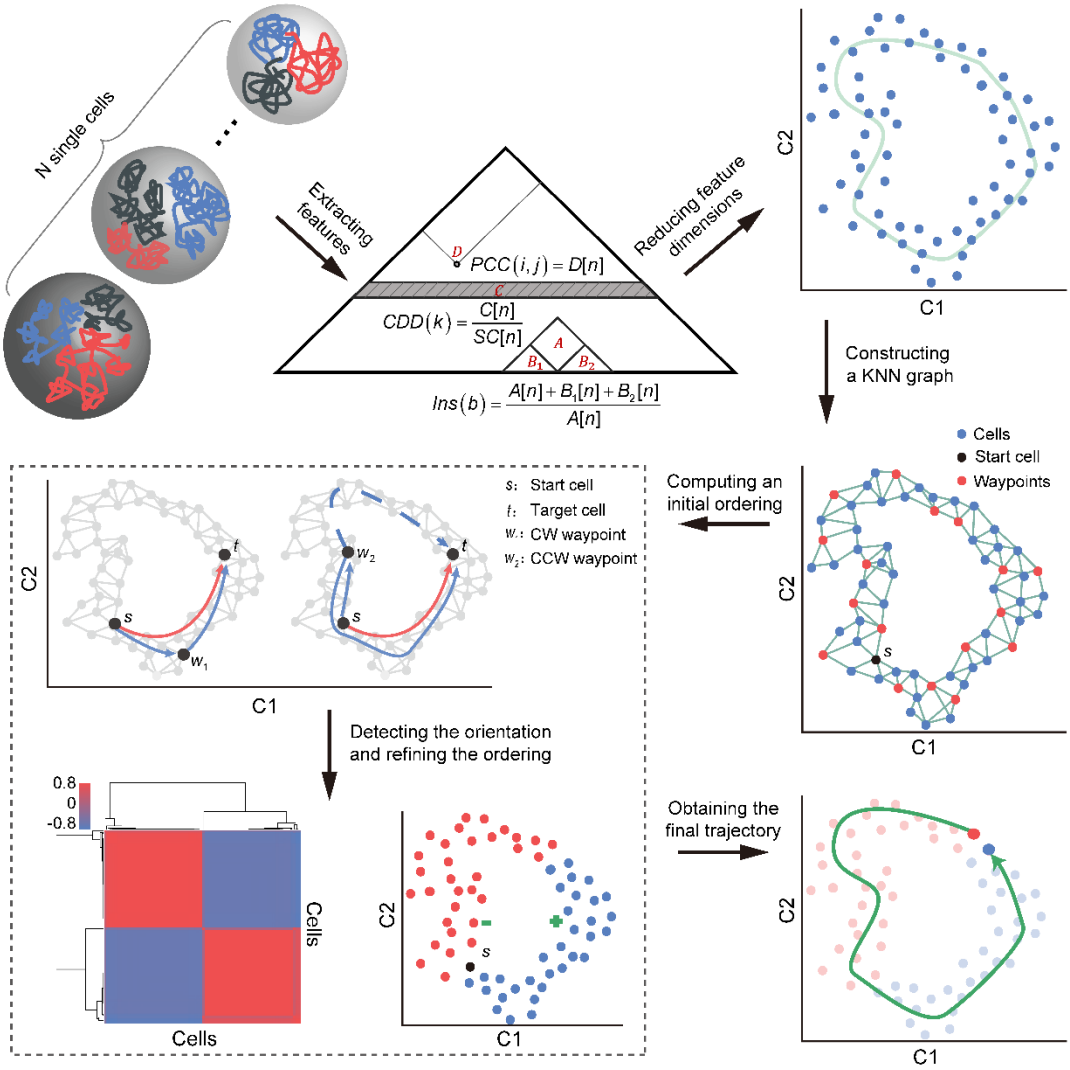

Figure 1. Illustration of CIRCLET.

Installing CICRLET:

1) It is recommended to install CICRLET in python 3.6 or above environment.
2) Installing requiring packages, including numpy, pandas, seaborn using pip or conda.
3) CICRLET build upon a package called Wishbone. Before installing CICRLET, first, follow these instructions to install.
4) Extracting the downloaded installation package to the specified directory and run the command: python setup.py install.

Getting started:

- Circular trajectory reconstruction of Hi-C maps:

Import CIRCLET

0) Next, we evaluate the trajectory in Nagano et al. [1] and get initial information of Hi-C dataset. Now you can execute the commands below.

 # The fig result is shown in Figure 2A.
Nature_evaluation, index, data_type, UBI, passed_qc_sc_DF_RO =
evaluate_Nagano_study(Nagano_dir)

1) Extracting feature: We have extracted multi-scale different multi-scale features of chromosomal architectures from single Hi-C maps, including Contact probability distribution versus genomic distance (CDD), Pairs' contact coverage (PCC), Insulation score of each bin (Ins), and Multiple composite metrics (MCM). In the step, the users need to read the required feature sets to program. We suggest to use the combination of three feature sets (MCM, PCC and CDD) to reconstruct the trajectory according to the analysis in our publication. Now you can execute the commands below to get combination of multiple features set.
   # Get input matrix for CIRCLET
   HiC_dir='./src/CIRCLET/DATA/Hi-Cmaps'
   scdata,filename=Get_SC_HiCmap_Features(HiC_dir,index)

2) Reducing feature dimensions: now, you can execute the commands below to reduce dimension of original space.
   # tsne result with FACS label is shown in the in Figure 2B.
   data=Reduce_dimension_ HiCmap(scdata,data_type)

3) Getting the circular trajectory from Hi-C maps: now, you can execute the commands to command trajectory reconstruction.
   # trajectory reconstruction based on KNN
   Result_file=Getting_trajectory_HiCmap(HiC_dir,scdata,data,filename)

4) Evaluating the results: now you can execute the commands below to evaluate the reconstruction trajectory.
   # The fig result is shown in Figure 2C and evaluation results is shown in Table 1
   type_names=['G1','ES','MS','G2']
   evaluation=evaluate_result_RNAseq(passed_qc_sc_DF_RO,data_type,Result_file,UBI, type_names,software='CIRCLET')

- Circular trajectory reconstruction of RNA-seq datasets:

1) Extracting feature: This process only uses the part of CIRCLET, not including the step of Extracting feature. Thus, we collected a single-cell RNA-seq dataset consisting of 182 cells for G1, S and G2/M phases and use a set of 959 annotated genes of cell cycle for analysis with variation above the background level in [2]. Now you can execute the commands below to get combination of multiple features set.
   # inputing address of RNA-seq dataset
   Rnaseq_dir='./src/CIRCLET/DATA/MESC_RNA-seq/EMSC_RNA-seq.xlsx'
   filename='EMSC_RNA-seq'

 # choosing features of RNA-seq dataset based on annotated genes
#To perform this step, you must include 'GO_term_summary_CellCycle.xlsx' in # the folder of 'EMSC_RNA-seq.xlsx'.
scdata, data_types=Get_SC_RNAseq_Features(Rnaseq_dir, filename)

2) Reducing feature dimensions: now, you can execute the commands below to reduce dimension of original space.
 # tsne result with FACS label is shown in the in Figure 2D.
data=Reduce_dimension_RNAseq(scdata,data_type)

3) Getting the circular trajectory from RNA-seq dataset: now, you can execute the commands to command trajectory reconstruction.
Result_file=Getting_trajectory_RNAseq(Rnaseq_dir,scdata,data,filename='MESC_RNA-seq')
4) Evaluating the results: now you can execute the commands below to evaluate the reconstruction trajectory.
 # The fig result is shown in Figure 2E and evaluation results is shown in Table 2
type_names=['G1','S','G2/M']
 # return four evaluation indexes for the reconstructed cell-cycle trajectory by #CIRCLET based on single-cell RNA-seq dataset.
evaluation=evaluate_result_RNAseq(Result_file,data_types,type_names,software='CIRCLET')

- Circular trajectory reconstruction of RNA-seq datasets:

5) Extracting feature: This process only uses the part of CIRCLET, not including the step of Extracting feature. Thus, we collected a single-cell RNA-seq dataset consisting of 182 cells for G1, S and G2/M phases and use a set of 959 annotated genes of cell cycle for analysis with variation above the background level in [2]. Now you can execute the commands below to get combination of multiple features set.
 # inputing address of RNA-seq dataset
Rnaseq_dir='./src/CIRCLET/DATA/MESC_RNA-seq/EMSC_RNA-seq.xlsx'
filename='EMSC_RNA-seq'
 # choosing features of RNA-seq dataset based on annotated genes
#To perform this step, you must include 'GO_term_summary_CellCycle.xlsx' in # the folder of 'EMSC_RNA-seq.xlsx'.
scdata, data_types=Get_SC_RNAseq_Features(Rnaseq_dir, filename)

6) Reducing feature dimensions: now, you can execute the commands below to reduce dimension of original space.
 # tsne result with FACS label is shown in the in Figure 2D.
data=Reduce_dimension_RNAseq(scdata,data_type)

7) Getting the circular trajectory from RNA-seq dataset: now, you can execute the commands to command trajectory reconstruction.

Result_file=Getting_trajectory_RNAseq(Rnaseq_dir,scdata,data,filename='MESC_RNA-seq')

8) Evaluating the results: now you can execute the commands below to evaluate the reconstruction trajectory.

 # The fig result is shown in Figure 2E and evaluation results is shown in Table 2
type_names=['G1','S','G2/M']
 # return four evaluation indexes for the reconstructed cell-cycle trajectory by #CIRCLET based on single-cell RNA-seq dataset.
evaluation=evaluate_result_RNAseq(Result_file,data_types,type_names,software='CIRCLET')

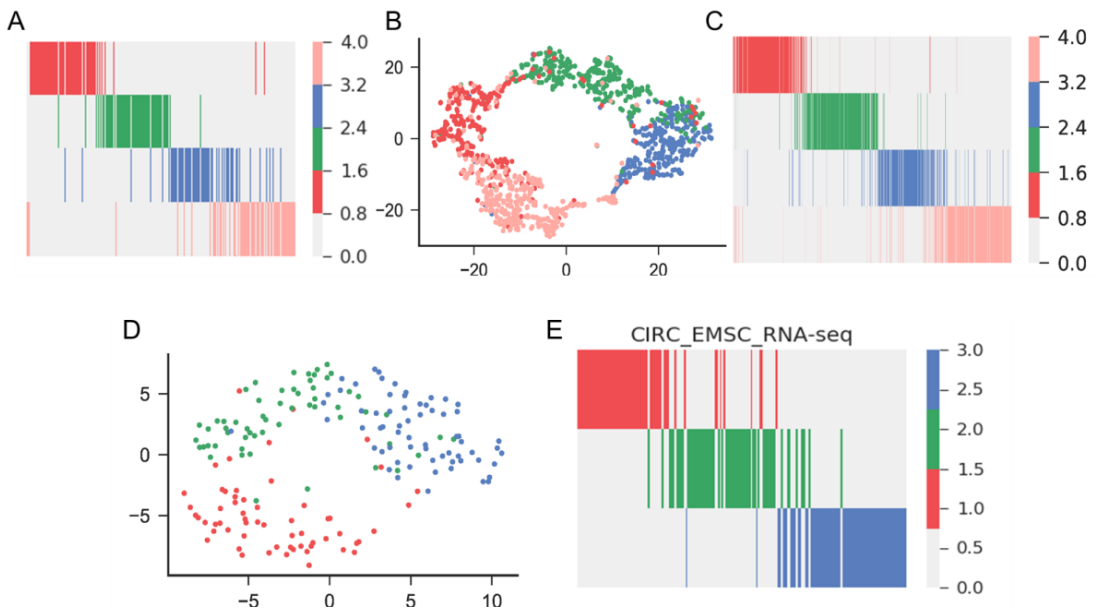

Figure 2. A. The reconstructed cell-cycle trajectories by Nagano et al.' method from four FACS-sorted cells (G1, ES, MS, LS/G2) based on single cell Hi-C datasets. B. tSNE maps by CIRCLET from four FACS-sorted cell phases (G1, ES, MS, LS/G2) based on single cell Hi-C datasets. C. The reconstructed cell-cycle trajectories by CIRCLET from four FACS-sorted cells (G1, ES, MS, LS/G2) based on single cell Hi-C datasets. D-E. tSNE map and reconstructed cell-cycle trajectory by CIRCLET from four FACS-sorted cell phases (G1, S, G2/M) based on 959 cell-cycle annotated genes from a single-cell RNA-seq dataset of 182 cells.

Table 1. Comparison of the five evaluation indexes for the reconstructed trajectory by CIRCLET based on single cell Hi-C datasets.

Evaluation Indexes	Scores
AUC:G1-ES	0.961669
AUC:ES-MS	0.935530
AUC:MS-G2	0.887522
AUC:G2-G1	0.975066
LCS	0.801200

Table 2. Comparison of the five evaluation indexes for the reconstructed trajectory by CIRCLET based on single cell RNA-seq datasets.

Evaluation Indexes	Scores
AUC:G1-S	0.911748
AUC:S-G2/M	0.963130
AUC:G2/M-G1	0.997132
LCS	0.814607

References:

1. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, Mendelson Cohen N, Wingett S, Fraser P, Tanay A: **Cell-cycle dynamics of chromosomal organization at single-cell resolution.** *Nature* 2017, **547:**61-67.
2. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells.** *Nature Biotechnology* 2015, **33:**155.