sctenifold

Cai Lab

Jun 17, 2022

METHODS

1 scTenifoldNet 1

2 scTenifoldKnk 7

3 scTenifoldXct 9

4 scTenifoldDev 11

5 License Agreement 13

CHAPTER

ONE

SCTENIFOLDNET

scTenifoldNet: A Machine Learning Workflow for Constructing and Comparing Transcriptome-wide Gene Regulatory Networks from Single-Cell Data

See Patterns paper for more information.

1.1 scTenifoldNet in R

1.1.1 Installation

In an *R* shell, type:

```
install.packages("scTenifoldNet")
```

1.1.2 Basic Use

```
library(scTenifoldNet)
library(Matrix)
X <- read.csv('X1.csv', header = FALSE)
colnames(X) <- paste0('X1_', seq_len(ncol(X)))
X <- as.matrix(X)
Y <- read.csv('X2.csv', header = FALSE)
colnames(Y) <- paste0('X2_', seq_len(ncol(Y)))
Y <- as.matrix(Y)
rownames(X) <- rownames(Y) <- readLines('genelist.csv')
set.seed(1)
DR <- scTenifoldNet(X = X, Y = Y)
save(DR, file = 'netOut.RData')
write.csv(DR$diffRegulation, row.names = FALSE, file = 'netResult.csv')
```

1.2 scTenifoldNet in MATLAB

1.2.1 Installation

Run the following code in *MATLAB*:

```
tic
disp('Installing scTenifoldNet...')
unzip('https://github.com/cailab-tamu/scTenifoldNet/archive/master.zip');
addpath('./scTenifoldNet-master/MATLAB');
toc
if exist('sctenifoldnet_m.m','file')
    disp('scTenifoldNet installed!')
end
```

1.2.2 Basic Use

Run ‘SCTENIFOLDNET` with an example data file clean_data_1Ctl_2FgF2.mat in *MATLAB*:

```
load clean_data_1Ctl_2FgF2.mat
sce=sce.selectgenes(1,0.15);
sce=sce.qcfilter;

%%
X0=sce.X(:,sce.c_batch_id==1);
X1=sce.X(:,sce.c_batch_id==2);
T=sctenifoldnet(X0,X1,sce.g,'savegrn',true);
writetable(T,'resT.txt');
Tr=e_fgsearun(T);
writetable(Tr,'resTr.txt');
tgenes=T.genelist(T.pAdjusted<0.1);
e_fgseanet(Tr);

load(ls('A0_*.mat'))
[y,i]=isemember(tgenes,genelist);
assert(all(y))
a0=A0(i,i);

load(ls('A1_*.mat'))
[y,i]=isemember(tgenes,genelist);
assert(all(y))
a1=A1(i,i);

g1=digraph(a1,tgenes,'omitselfloops');
g2=digraph(a2,tgenes,'omitselfloops');
gui.i_doublegraphs(g1,g2);
```

1.3 scTenifoldNet in Julia

1.3.1 Installation

Run the following code in *Julia*:

```
using Pkg
Pkg.add(PackageSpec(url="git://github.com/jamesjcai/ScTenifold.jl.git"))
Pkg.test("ScTenifold")
# or
# ] add https://github.com/jamesjcai/ScTenifoldNet.jl
```

1.3.2 Basic Use

Here is a simple example using randomly generated data.

```
using ScTenifold
using DelimitedFiles
# cd(dirname(@__FILE__))

X1=rand(100,1000);
X2=copy(X1);
X2[4,:].=0.0;
@time d,fc,p,adjp=ScTenifold.sctenifoldnet(X1,X2,donorm=false)

open("output_small.txt", "w") do io
    writedlm(io, [d fc p adjp])
end
```

1.3.3 Exported Functions

Code	Function
pcnet	Computes a gene regulatory network based on principal component regression
tensordecomp	Performs CANDECOMP/PARAFAC (CP) Tensor Decomposition
manialn	Performs non-linear manifold alignment of two gene regulatory networks
drgenes	Evaluates gene differential regulation based on manifold alignment distances
tenrnet	Subsamples cells, constructs single-cell gene regulatory networks (scGRNs) using principal component regression (pcnet), and denoises scGRNs using tensor decomposition (tensordecomp).

1.3.4 Loading ScTenifoldNet

Once installed, **ScTenifoldNet.jl** can be loaded typing:

```
using ScTenifoldNet
```

1.3.5 Simulating of a dataset

Here we simulate a dataset of 2000 cells (columns) and 100 genes (rows) following the negative binomial distribution with high sparsity (~67%).

```
d=NegativeBinomial(20,0.98)
X=rand(d,100,2000)
```

1.3.6 Generating a perturbed network

We generate a perturbed network modifying the expression of genes 10, 2, and 3 and replacing them with the expression of genes 50, 11, and 5.

```
Y=copy(X)
Y[10,:]=Y[50,:]
Y[2,:]=Y[11,:]
Y[3,:]=Y[5,:]

X=X[:,vec(sum(X,dims=1).>30)]
Y=Y[:,vec(sum(Y,dims=1).>30)]
```

1.3.7 Generating networks

Here we run **ScTenifoldNet** under the H0 (there is no change in the regulation of the gene) using the same matrix as input and under the HA (there is a change in the regulation of the genes) using the control and the perturbed network.

```
Z0=ScTenifoldNet.tenrnet(X, donorm=true)
Z1=ScTenifoldNet.tenrnet(Y, donorm=true)
```

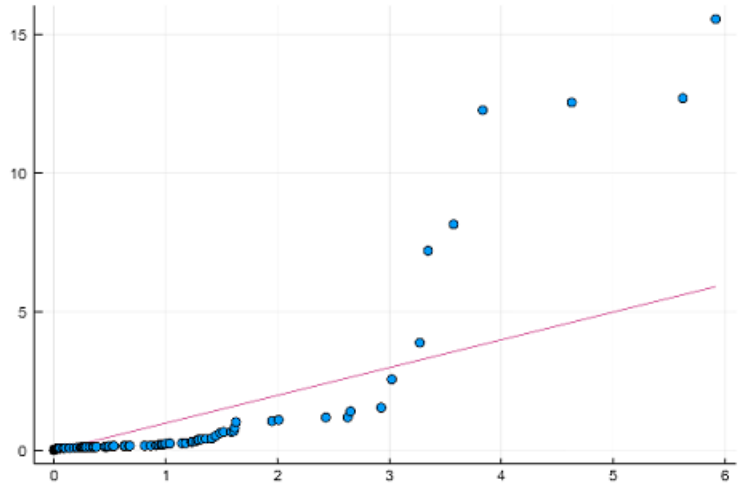
1.3.8 Differential regulation based on manifold alignment distances

As is shown below, under the H0, none of the genes shown a significative difference in regulatory profiles using an FDR cut-off of 0.1, but under the HA, the 6 genes involved in the perturbation (50, 11, 2, 10, 5, and 3) are identified as perturbed.

```
d,aln0,aln1=ScTenifoldNet.manialn(Z0,Z1)
fc,p,adjp=ScTenifoldNet.drgenes(d)
```

1.3.9 Plotting the results

Results can be easily displayed using quantile-quantile plots.




```
using StatsPlots, Distributions
x=rand(Chisq(1), length(fc))
qqplot(x, fc)
```

CHAPTER

TWO

SCTENIFOLDKNK

scTenifoldKnk: an efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation

See Patterns paper for more information.

The characterization of the perturbation profiles caused by a gene knockout allows identifying genes under its direct regulation. Due to the economical and biological limitations to perform the experimental systematic knockout of all genes in a cell-type-specific manner, the development of computational tools to predict the effect of gene knockouts is needed. Following that purpose, we introduced scTenifoldKnk, a machine learning workflow performing virtual knockout experiments on single-cell gene regulatory networks. scTenifoldKnk only requires wild-type single-cell RNA-seq data as input, and returns a weighted list of genes, that ranked by the regulatory effect predicted for the knockout gene over all the other genes expressed in the cell.

2.1 scTenifoldKnk in MATLAB

2.1.1 Quick installation

Run the following code in *MATLAB*:

```
tic
disp('Installing scTenifoldKnk...')
unzip('https://github.com/cailab-tamu/scTenifoldKnk/archive/master.zip');
addpath('./scTenifoldKnk-master/MATLAB');
toc
if exist('sctenifoldknk.m','file')
    disp('scTenifoldKnk installed!')
end
```

2.2 Case Study: Identifying biological processes regulators using genome-wide perturbation profiles

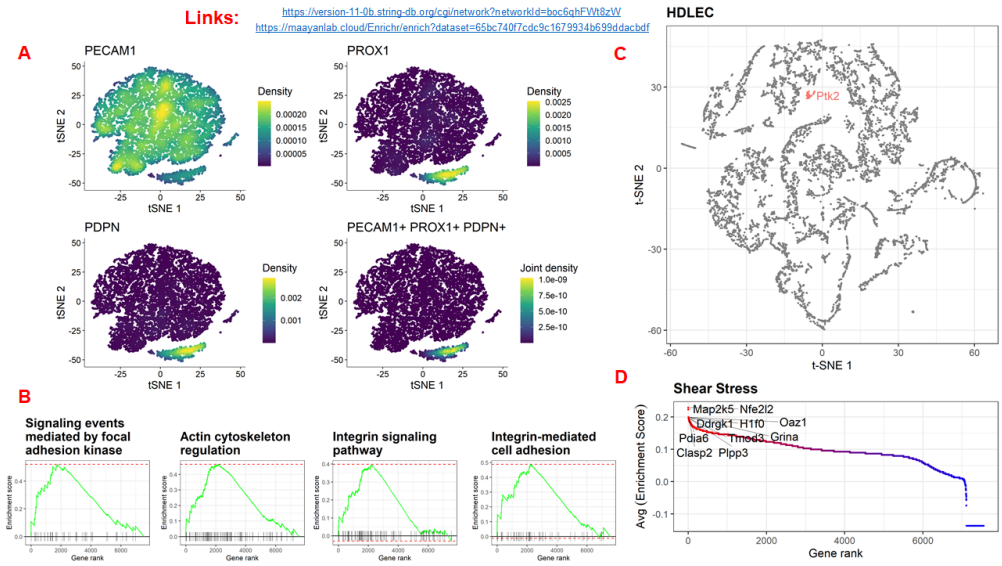
2.2.1 Shear Stress Response in Human Dermal Lymphatic Endothelial Cells

Using this weighted list together with the gene sets provided by the gene ontology (GO), it is also possible to predict unknown regulators of the already characterized biological processes. As an example, we performed the systematic knockout of 7,548 genes expressed in 885 Mice Dermal Lymphatic Endothelial Cells (MDLEC). We collected by the integration of 23 public datasets containing PECAM1+, PROX1+, and PDPN+ endothelial cells from the PanglaoDB database

(SRS2749416, SRS2532206, SRS4004491, SRS3348010, SRS3044263, SRS3044258, SRS3044262, SRS2874285, SRS2874279, SRS2874276, SRS2874271, SRS3600293, SRS3600294, SRS3600295, SRS3600296, SRS3600297, SRS3600298, SRS3600299, SRS3600300, SRS3600301, SRS3020563, SRS4388158, and SRS4388159)–that is, we harmonized all publicly available scRNA-seq datasets for endothelial cells.

We used the predicted perturbation profile of each gene to identify novel regulators associated with the response to shear stress using the gene sets reported by the gene ontology for this process (GO:0034616, GO:0071498, GO:0071499). We did this applying single-sample gene set enrichment analysis (ssGSEA included in the GSVA R package) on the predicted perturbation profiles for all genes and then ranked them based on the decreasing average enrichment score for the selected three gene sets. Using this approach, we found *Nfe2l2* (PMID: 25563726, 28877882), *Map2k5* (PMID: 26416763, 21166929), *Ddrgk1*/*Ufbp1* (PMID: 29461087), *Grina*, *H1f0* (PMID: 23802622), *Oaz1*, *Clasp2*, *Pdia6*, *Plpp3* (30429326, 26034042), *Tmod3* as the top 10 predicted candidate genes regulating the response to shear stress in MDLEC cells.

https://version-11-0b.string-db.org/cgi/network?networkId=boc6qhFWt8zW https://maayanlab.cloud/Enrichr/enrich?dataset=65bc740f7cdc9c1679934b699ddacbdf



CHAPTER

THREE

SCTENIFOLDXCT

scTenifoldXct: a semi-supervised method for predicting cell-cell interactions and mapping cellular communication graphs via manifold alignment of intra- and inter-cellular gene regulatory networks

The latest single-cell RNA sequencing (scRNA-seq) technology allows transcriptomic information to be gathered from thousands of cells in a single assay, providing unprecedented cellular heterogeneity insight. Data from scRNA-seq enables the detection of cell-cell interactions in a tissue sample. We present scTenifoldXct, a semi-supervised computational tool for detecting ligand-receptor (L-R)-mediated cell-cell interactions and mapping cellular communication graphs. Our method is based on manifold alignment, using L-R pairs as inter-data correspondences to embed ligand and receptor genes expressed in interacting cells into a unified latent space. Deep neural networks are employed to minimize the distance between corresponding genes while preserving the structure of gene regulatory networks. We apply scTenifoldXct to real data sets for testing. We show that our method detects cell-cell interactions with high sensitivity and reveals weak but biologically relevant interactions that tend to be overlooked by other methods. We demonstrate how scTenifoldXct can be used to compare different samples, such as healthy vs. diseased, to identify differential interactions, thereby revealing changes in the communication status of cells.

Visit GitHub site for more information.

CHAPTER

FOUR

SCTENIFOLDDEV

CHAPTER

FIVE

LICENSE AGREEMENT