Supplementary data

Application Notes

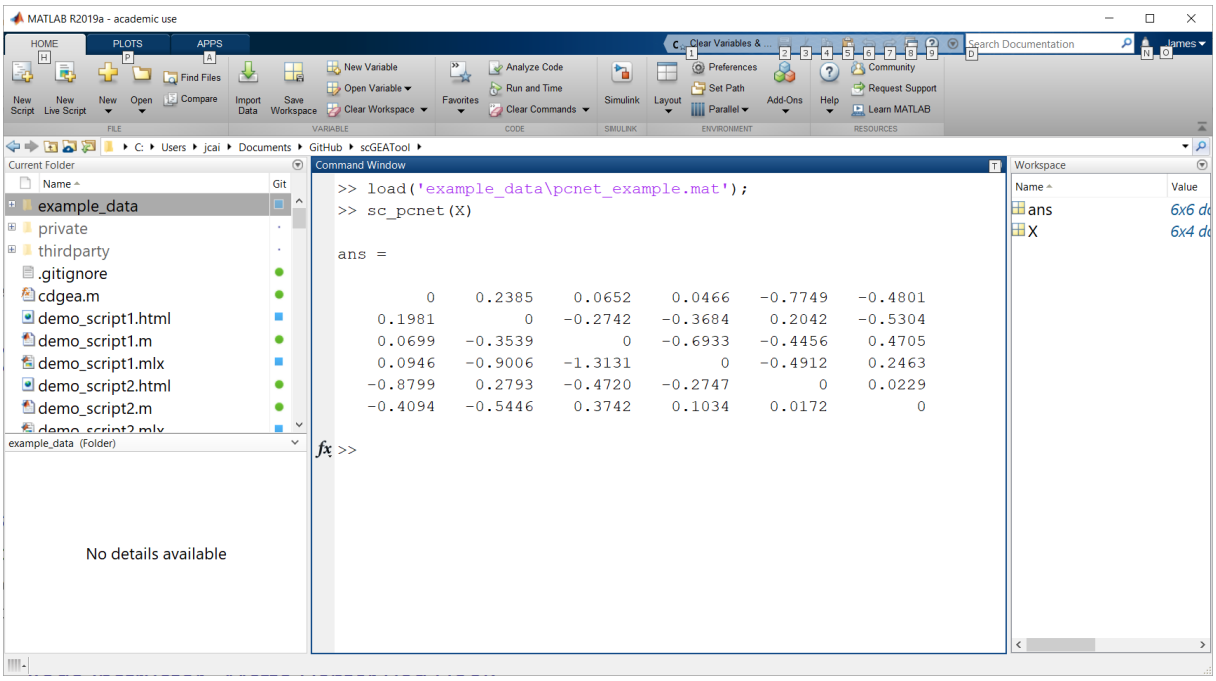
scGEAToolbox: a Matlab toolbox for single-cell RNA sequencing data analysis

James J. Cai[1,2]
[1]Department of Veterinary Integrative Biosciences, [2]Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843-4458, USA.

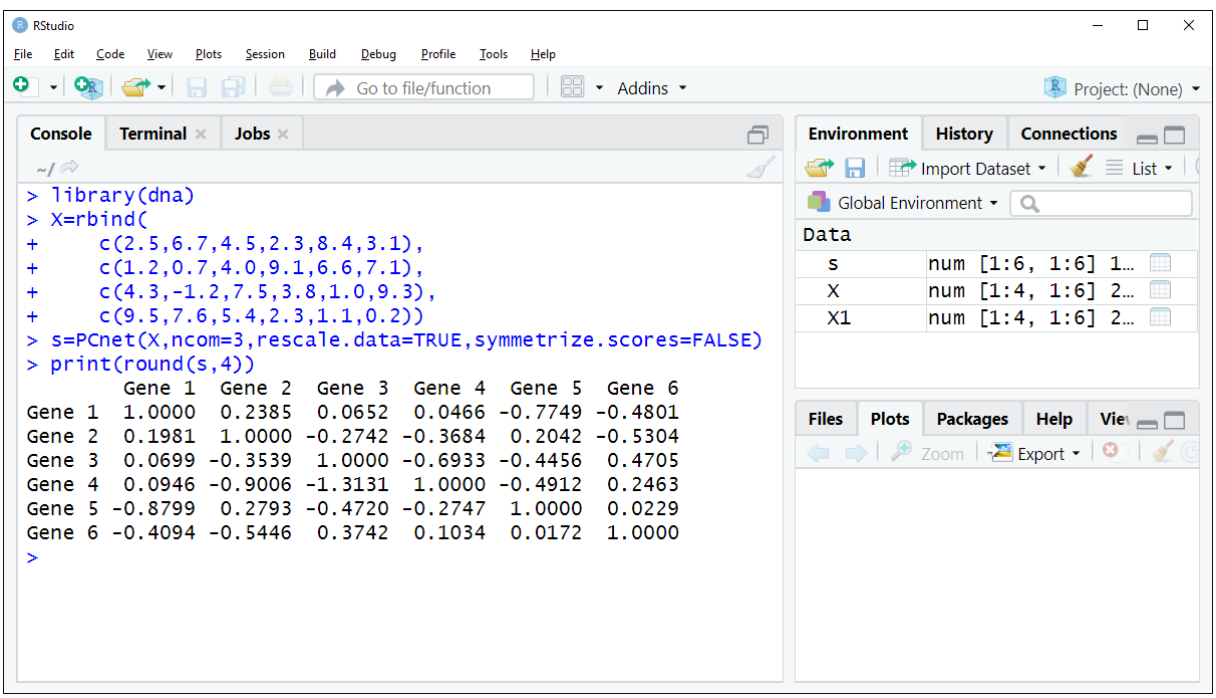
I. Numerical comparisons between selected scGEAToolbox Matlab functions and R functions

Comparison between scGEAToolbox PCNet function and dna/R PCNet function

Matlab code



R code



Comparison between scGEAToolbox function SC_SC3 and R/Bioconductor SC3

R code

```
library(SingleCellExperiment)
library(SC3)
library(scater)
sce <- SingleCellExperiment(
    assays = list(
        counts = as.matrix(yan),
        logcounts = log2(as.matrix(yan) + 1)
    ),
    colData = ann
)
# define feature names in feature_symbol column
rowData(sce)$feature_symbol <- rownames(sce)
# remove features with duplicated names
sce <- sce[!duplicated(rowData(sce)$feature_symbol), ]
sce <- sc3(sce, ks = 6)
sc3_export_results_xls(sce)
```

Matlab code

```
[X,genelist]=sc_readtsvfile('example_data/yan.csv');
t=readtable('example_data\yan_celltype.txt');
celltypelist=string(t.cell_type1);

c1=sc_sc3(X,6);
% Result of SC3/R pacakge
load example_data/sc3_results.txt
c0=sc3_results;
```

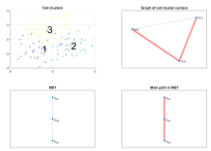
Compare clustering results using NMI

```
%% Compare clustering results
fun_cmp_clusters(c0,c1,"type","nmi")
The NMI value is 0.961969

ans =

    0.9620
```

Comparison between scGEAToolbox SC_TSCAN function and R/Bioconductor TSCAN

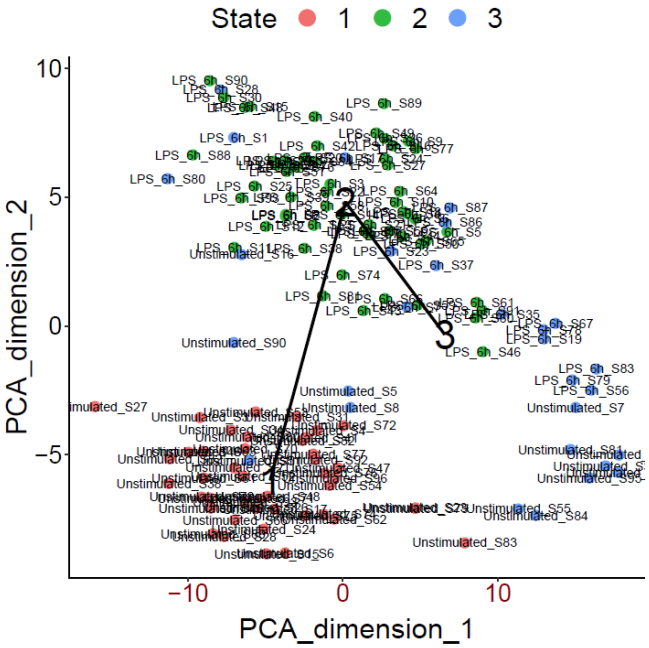
Matlab code

```
load example_data\tscan_lpsdata.mat
t=sc_tscan(X,'plotit',true);
```



R code

```
library(TSCAN)
data(lpsdata)
procdata <- preprocess(lpsdata)
lpsmclust <- exprmclust(procdata)
plotmclust(lpsmclust)
```


State
1
2
3
PCA_dimension_2
PCA_dimension_1
10
5
0
−5
−10
0
10

II. scGEAToolbox Demonstration Scripts

Demonstration of Filter, Normalization and Batch Correction of Data in scGEAToolbox

Read scRNA-seq data, X and Y

```
cdgea; % set working directory
[X,genelistx]=sc_readfile('example_data/GSM3204304_P_P_Expr.csv');
```

```
Reading example_data/GSM3204304_P_P_Expr.csv ...... done.
```

```
[Y,genelisty]=sc_readfile('example_data/GSM3204305_P_N_Expr.csv');
```

```
Reading example_data/GSM3204305_P_N_Expr.csv ...... done.
```

Select genes with at least 3 cells having more than 5 reads per cell.

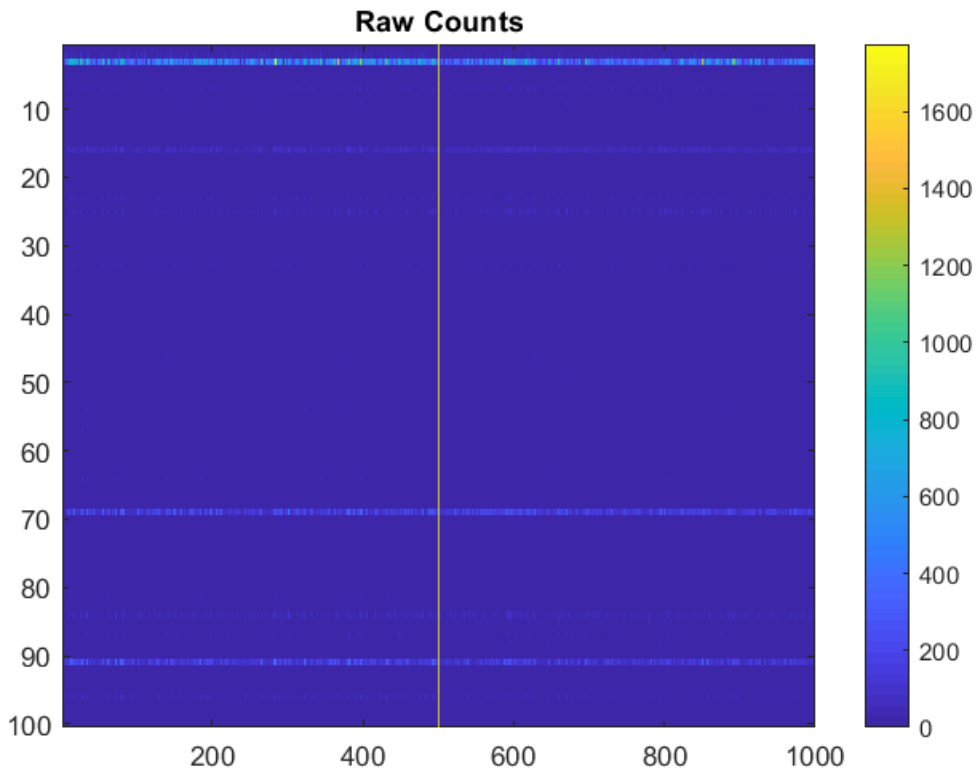
```
[X,genelistx]=sc_selectg(X,genelistx,5,3);
[Y,genelisty]=sc_selectg(Y,genelisty,5,3);
```

Obtain gene intersection of X and Y

```
[genelist,i,j]=intersect(genelistx,genelisty,'stable');
X=X(i,:);
Y=Y(j,:);
% libsizex=sum(X);
% libsizey=sum(Y);
% X=X(:,libsizex>quantile(libsizex,0.3)&libsizex<quantile(libsizex,0.95));
% Y=Y(:,libsizey>quantile(libsizey,0.3)&libsizey<quantile(libsizey,0.95));
clearvars -except X Y genelist
```

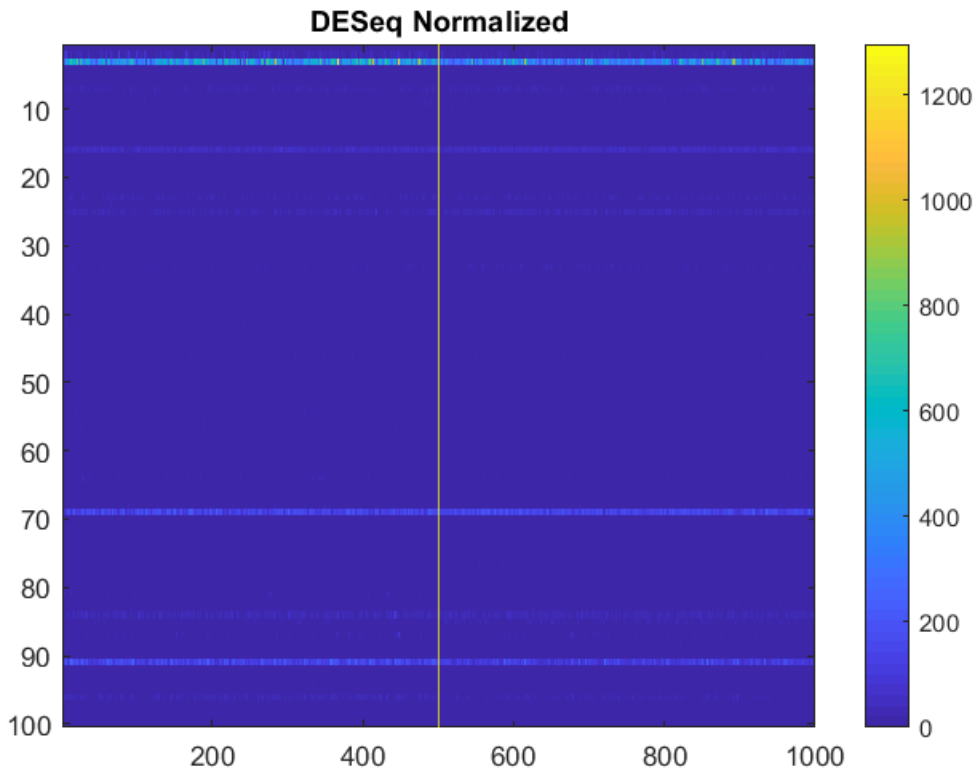
Show raw data

```
figure; imagesc([X(1:100,1:500) Y(1:100,1:500)]); title('Raw Counts'); colorbar;
xline(500,'y-');
```



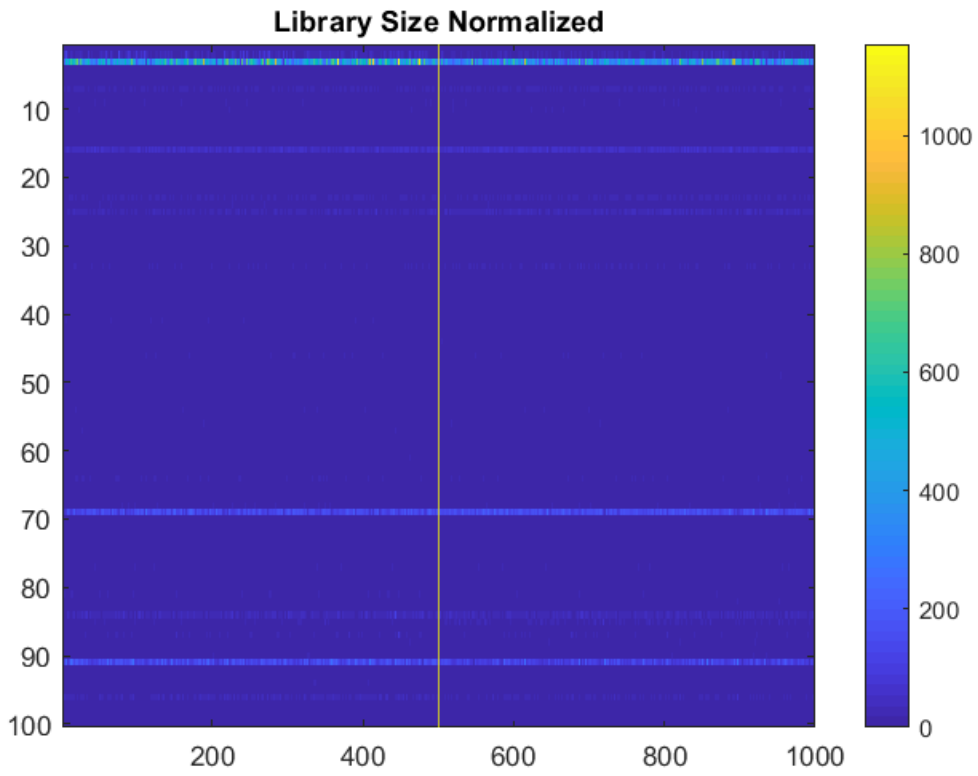
Show DESeq normalized data

```
[Xs]=sc_norm(X,'type','deseq');
[Ys]=sc_norm(Y,'type','deseq');
figure; imagesc([Xs(1:100,1:500) Ys(1:100,1:500)]); title('DESeq Normalized');
colorbar; xline(500,'y-');
```



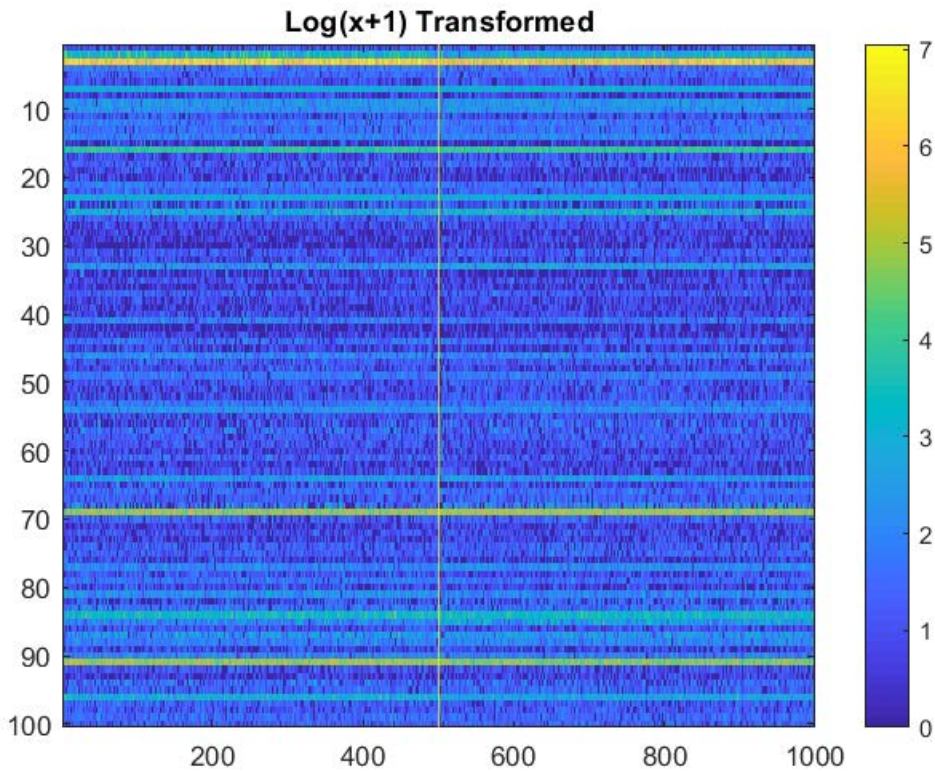
Show library-size normalized data

```
[X]=sc_norm(X,'type','libsize');
[Y]=sc_norm(Y,'type','libsize');
figure; imagesc([X(1:100,1:500) Y(1:100,1:500)]); title('Library Size Normalized');
colorbar; xline(500,'y-');
```



Log(x+1) transformed normalized data

```
X=log(X+1);
Y=log(Y+1);
figure; imagesc([X(1:100,1:500) Y(1:100,1:500)]); title('Log(x+1) Transformed');
colorbar; xline(500,'y-');
```



Show data subject to MAGIC imputation

```
Xo=run_magic(X);
```

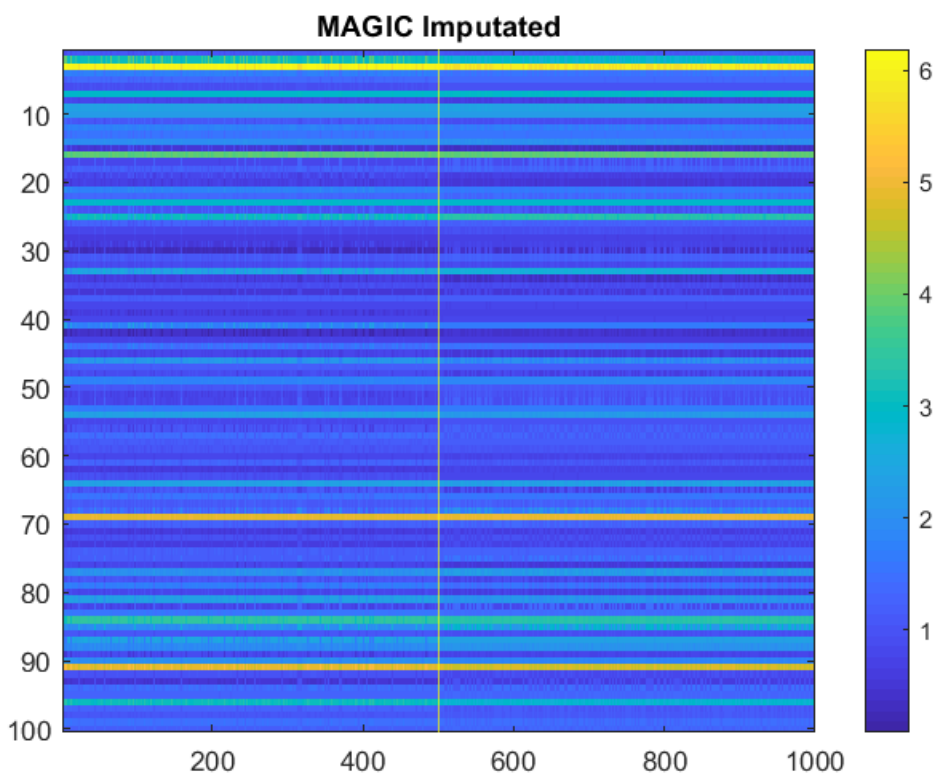
```
doing PCA
computing kernel
Computing alpha decay kernel:
Number of samples = 835
First iteration: k = 300
Number of samples below the threshold from 1st iter: 802
Using radius based search for the rest
   Symmetrize affinities
   Done computing kernel
imputing using optimal t
t = 1
t = 2
t = 3
t = 4
t = 5
t = 6
t = 7
t = 8
t = 9
```

```
Yo=run_magic(Y);
```

```
doing PCA
computing kernel
Computing alpha decay kernel:
Number of samples = 644
First iteration: k = 300
Number of samples below the threshold from 1st iter: 631
```

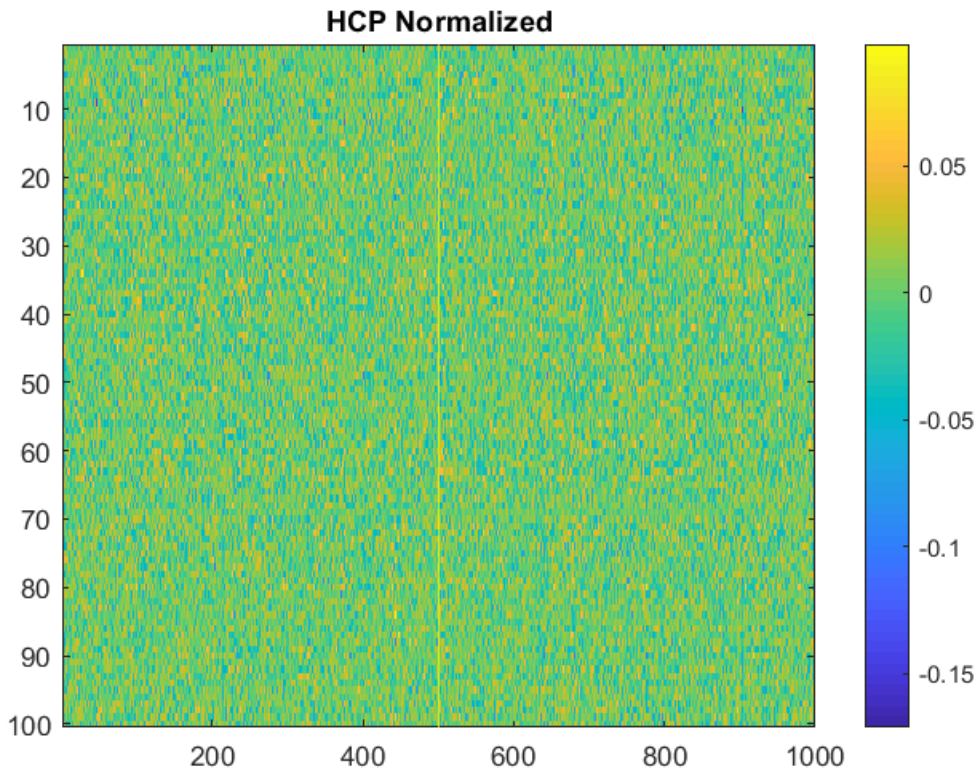
```
Using radius based search for the rest
   Symmetrize affinities
   Done computing kernel
imputing using optimal t
t = 1
t = 2
t = 3
t = 4
t = 5
t = 6
t = 7
t = 8
t = 9
```

```
figure; imagesc([Xo(1:100,1:500) Yo(1:100,1:500)]); title('MAGIC Imputated');
colorbar; xline(500,'y-');
```



Show HCP normalized data

```
[Xm,Ym]=run_hcp(X,Y);
figure; imagesc([Xm(1:100,1:500) Ym(1:100,1:500)]); title('HCP Normalized');
colorbar; xline(500,'y-');
```

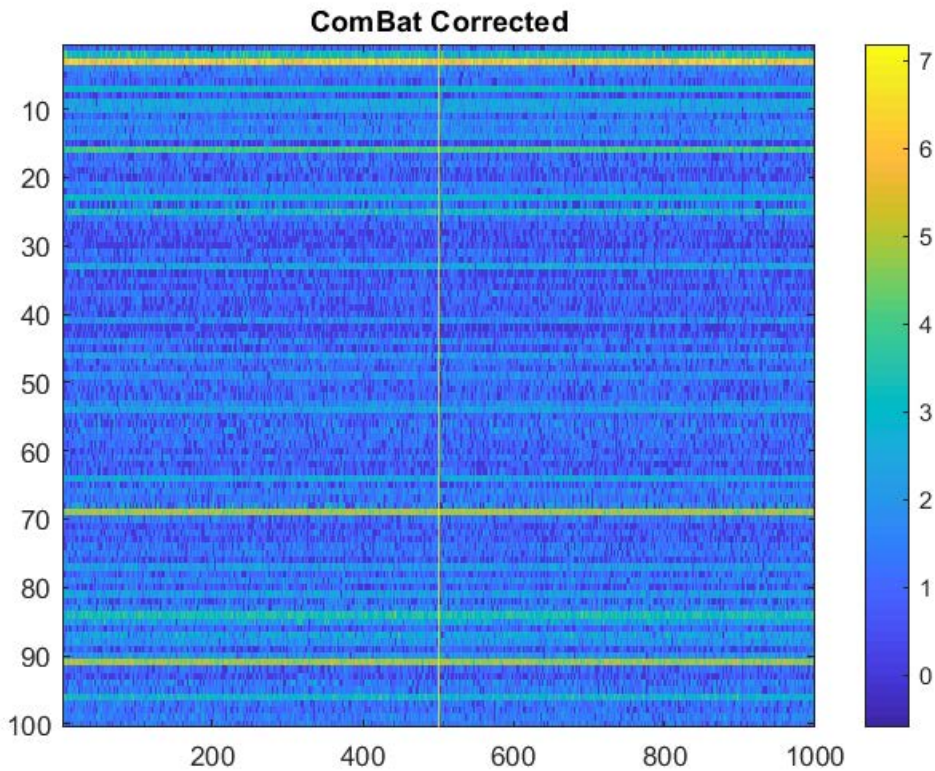


Show data with ComBat batch correction

```
[Xn,Yn]=run_combat2(X,Y);
```

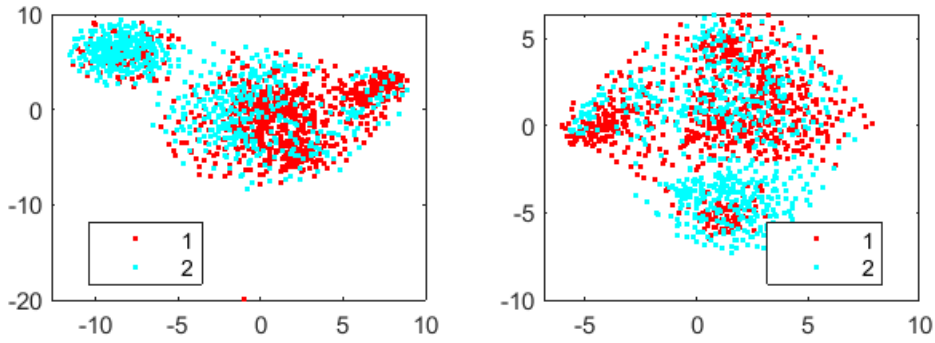
```
[combat] Found 2 batches
[combat] Adjusting for 0 covariate(s) of covariate level(s)
[combat] Standardizing Data across features
[combat] Fitting L/S model and finding priors
[combat] Finding parametric adjustments
[combat] Adjusting the Data
```

```
figure; imagesc([Xn(1:100,1:500) Yn(1:100,1:500)]); title('ComBat Corrected');
colorbar; xline(500,'y-');
```



Visulize cells before and after ComBat batch correction

```
batchidx=[1*ones(size(X,2),1); 2*ones(size(Y,2),1)];

figure;
subplot(2,2,1)
[s]=sc_tsne([X Y]);
gscatter(s(:,1),s(:,2),batchidx,'','',5);
subplot(2,2,2)
[s]=sc_tsne([Xn Yn]);
gscatter(s(:,1),s(:,2),batchidx,'','',5);
```



The End

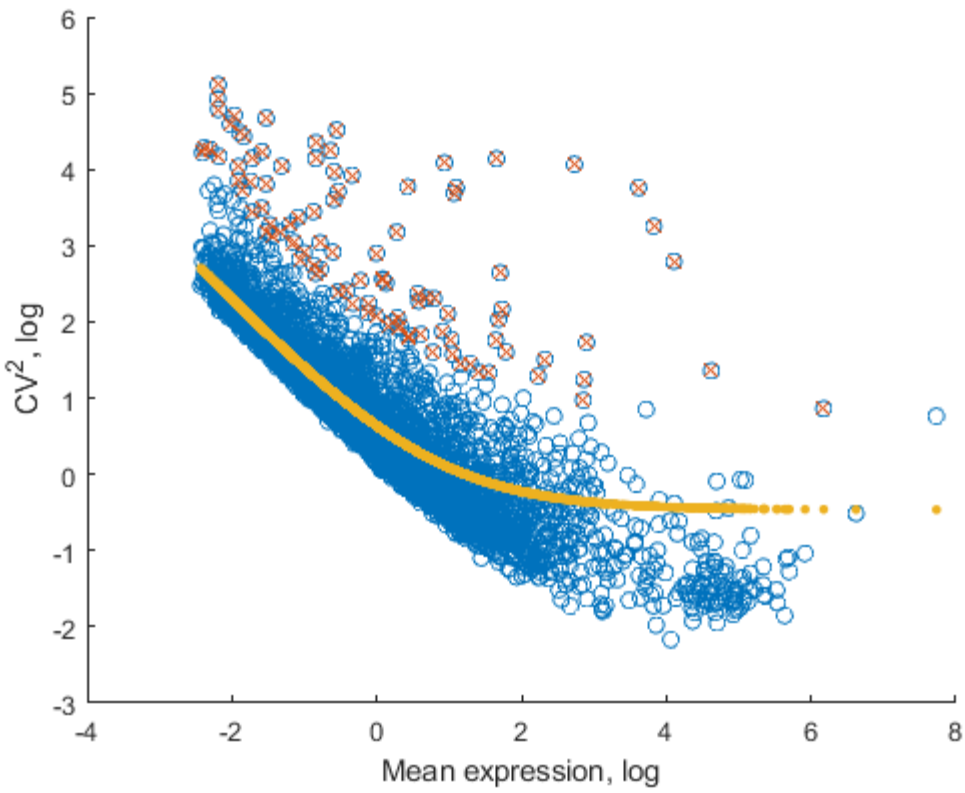
Demonstration of Feature Selection Functions in scGEApp

HVG analysis with single data X

```
cdgea; % set working directory
[X,genelist]=sc_readfile('example_data/GSM3044891_GeneExp.UMIs.10X1.txt');
```

```
Reading example_data/GSM3044891_GeneExp.UMIs.10X1.txt ...... done.
```

```
[X,genelist]=sc_selectg(X,genelist,3,1);
% Normalize data with DESeq method
Xn=sc_norm(X,'type','deseq');
[T]=sc_hvg(Xn,genelist,true,true);
```




```
% Highly variable genes (HVGenes), FDR<0.05
HVGenes=T.genes(T.fdr<0.05)
```

```
HVGenes = 1227×1 string array
    "BCL2A1"
    "CYP1B1"
    "TXNIP"
    "RSPH1"
    "RP11-856M7.6"
    "BRD3"
    "HIST2H2AA3"
    "JCHAIN"
    "LMNA"
```

```
    "MACROD2"
    "MIR3142HG"
    "PEG10"
    "CCL22"
    "PRKCDBP"
```

Spline-fit feature selection with single data X

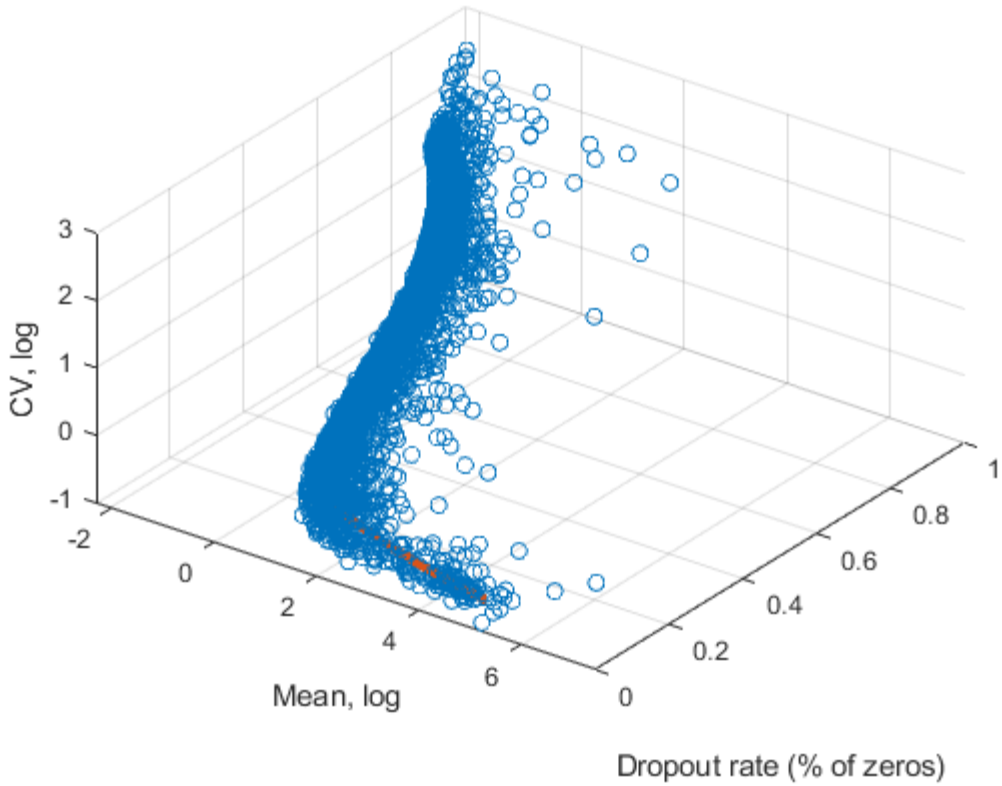
```
[X,genelist]=sc_readfile('example_data/GSM3044891_GeneExp.UMIs.10X1.txt');
```

```
Reading example_data/GSM3044891_GeneExp.UMIs.10X1.txt ...... done.
```

```
[X,genelist]=sc_selectg(X,genelist,3,1);

sortit=true;
[T1]=sc_splinefit(X,genelist,sortit);
% Top 50 featured genes with highest deviation (D) values
T1.genes(1:50)
```

```
ans = 50×1 string array
    "IGLC2"
    "IGHG1"
    "IGKC"
    "IGHG3"
    "CCL22"
    "IGHM"
    "IGHG4"
    "WFDC2"
    "IGKV1-12"
    "CCL4"
    "IGLC3"
    "CCL3L3"
    "FXYD2"
    "PRKCDBP"
```

```
dofit=true;
showdata=true;
% Show data points and the spline-fit curve
figure;
sc_scatter3(X,genelist,dofit,showdata);
view([36.39 46.25])
```



Analysis of differentially deviated (DD) genes using spline-fit feature selection with data X and Y

Read and pre-process two data sets, X and Y

```
[X,genelistx]=sc_readfile('example_data/GSM3204304_P_P_Expr_999cells.csv');
```

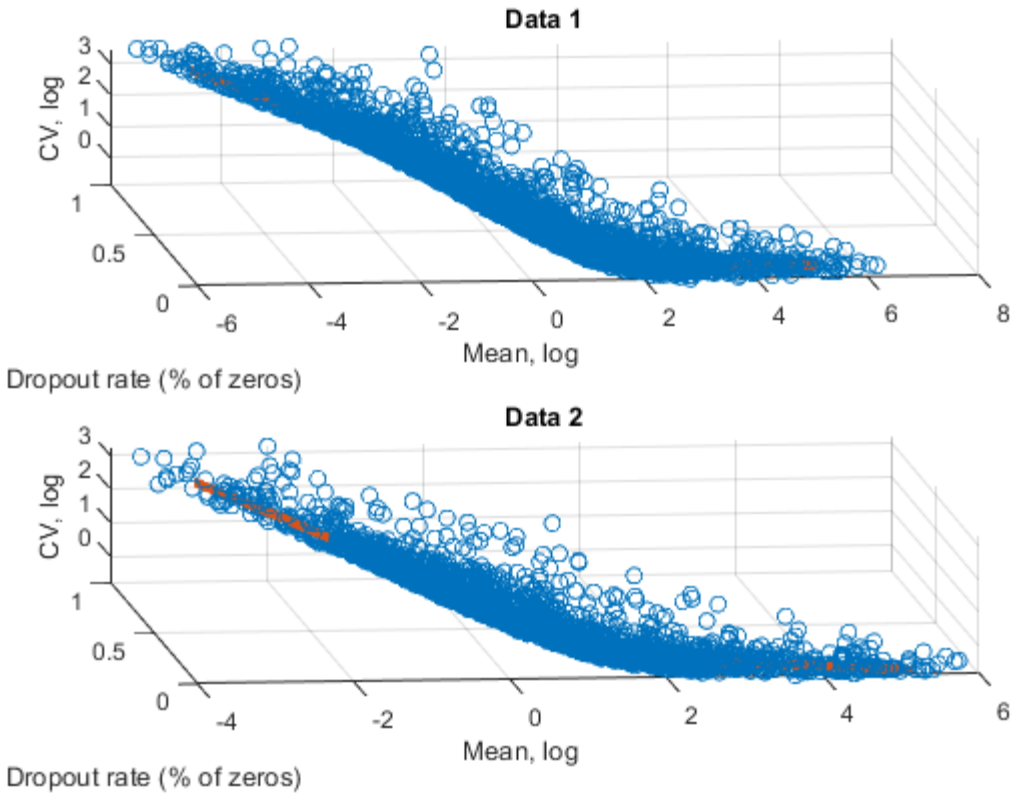
```
Reading example_data/GSM3204304_P_P_Expr_999cells.csv ...... done.
```

```
[Y,genelisty]=sc_readfile('example_data/GSM3204305_P_N_Expr_999cells.csv');
```

```
Reading example_data/GSM3204305_P_N_Expr_999cells.csv ...... done.
```

```
[X,genelistx]=sc_selectg(X,genelistx,3,1);
[Y,genelisty]=sc_selectg(Y,genelisty,3,1);

% Show 3D scatter plot and spline-fit curve for X
figure;
dofit=true;
showdata=true;
subplot(2,1,1)
sc_scatter3(X,genelistx,dofit,showdata);
title('Data 1')
view([-6.39 36.70])

% Show 3D scatter plot and spline-fit curve for Y
%figure;
subplot(2,1,2)
sc_scatter3(Y,genelisty,dofit,showdata);
title('Data 2')
```

```
% view([24.08 32.68])
view([-6.39 36.70])
```



Using function SC_SPLINEFIT2 to fit X and Y separately and obtain DD

value for each gene.

```
[T2]=sc_splinefit2(X,Y,genelistx,genelisty,true);
```

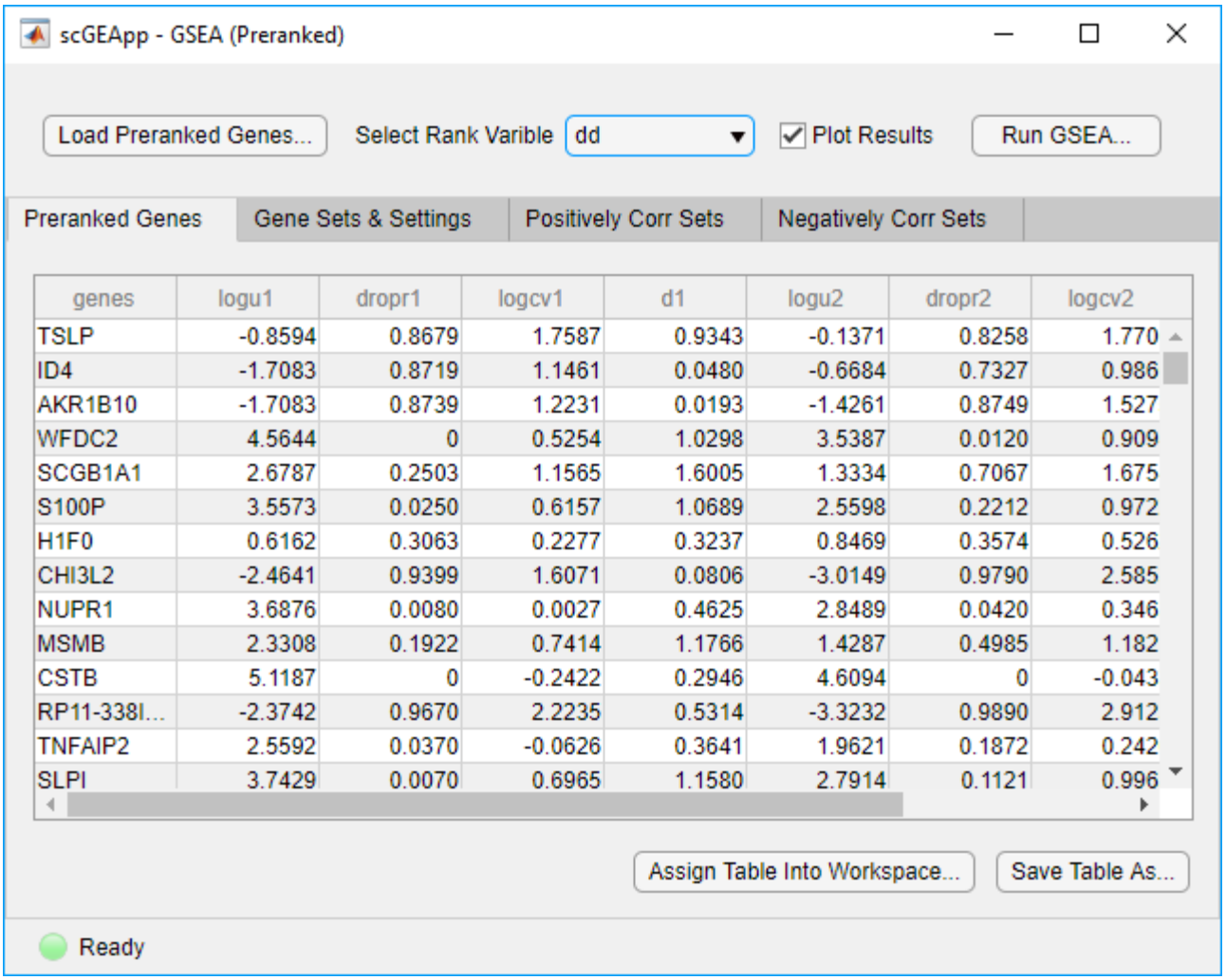
Top 50 genes with highest DD value.

```
T2.genes(1:50)
```

```
ans = 50×1 string array
    "TSLP"
    "ID4"
    "AKR1B10"
    "WFDC2"
    "SCGB1A1"
    "S100P"
    "H1F0"
    "CHI3L2"
    "NUPR1"
    "MSMB"
    "CSTB"
    "RP11-338I21.1"
    "TNFAIP2"
    "SLPI"
```

Run GSEAPreranked App with genes ranked with DD

```
addpath('thirdparty/GSEA');
a=gseapr;
a.load_data(T2);
```



Click gene set names in GSEA result table to show GSEA plots

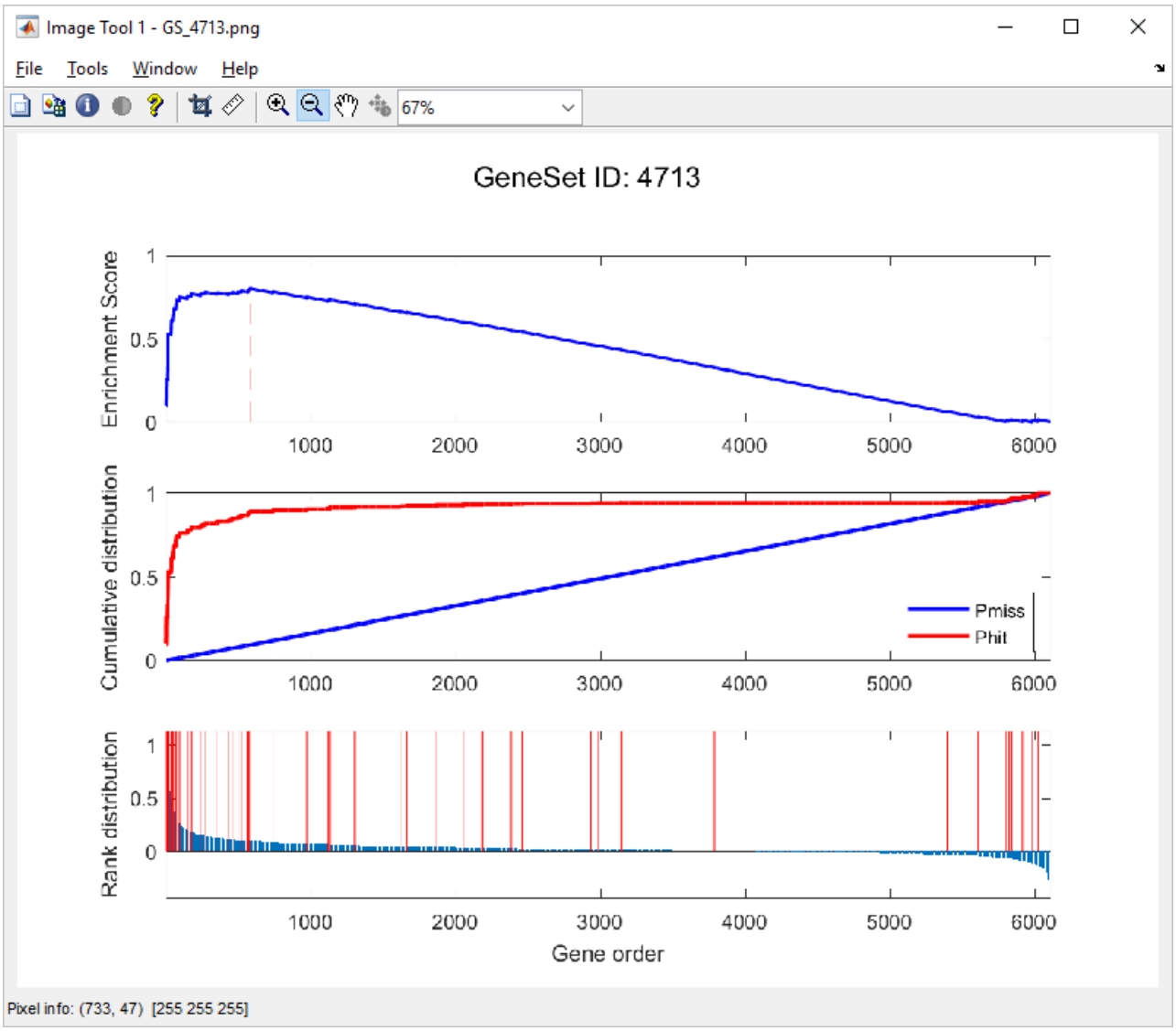


Image Tool 1 - GS_4713.png
File Tools Window Help
67%
GeneSet ID: 4713
Enrichment Score
Cumulative distribution
Rank distribution
Pmiss
Phit
Gene order
Pixel info: (733, 47) [255 255 255]

The End

Demonstration of Visualization Functions in scGEAToolbox

Load and pre-process three data sets, X, Y and Z

```
cdgea; % set working directory
[X,genelistx]=sc_readfile('example_data/GSM3204304_P_P_Expr.csv');
```

```
Reading example_data/GSM3204304_P_P_Expr.csv ...... done.
```

```
[Y,genelisty]=sc_readfile('example_data/GSM3204305_P_N_Expr.csv');
```

```
Reading example_data/GSM3204305_P_N_Expr.csv ...... done.
```

```
[Z,genelistz]=sc_readfile('example_data/GSM3044891_GeneExp.UMIs.10X1.txt');
```

```
Reading example_data/GSM3044891_GeneExp.UMIs.10X1.txt ...... done.
```

```
[X,genelistx]=sc_selectg(X,genelistx,3,1);
[Y,genelisty]=sc_selectg(Y,genelisty,3,1);
[Z,genelistz]=sc_selectg(Z,genelistz,3,1);
```

Intersection of common genes in X, Y and Z

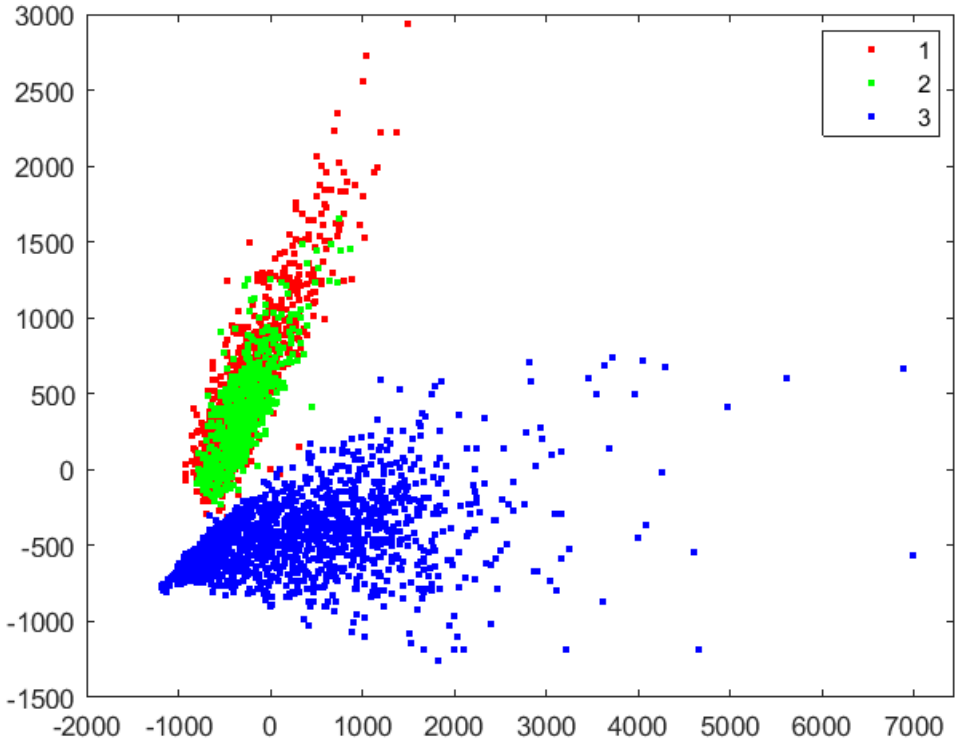
```
[genelist]=intersect(intersect(genelistx,genelisty,'stable'),genelistz,'stable');
% Remove genes encoded in the mitochondrial genome
i=startsWith(genelist,'MT-');
genelist(i)=[];
[~,i1]=ismember(genelist,genelistx);
[~,i2]=ismember(genelist,genelisty);
[~,i3]=ismember(genelist,genelistz);
X=X(i1,:); genelistx=genelist;
Y=Y(i2,:); genelisty=genelist;
Z=Z(i3,:); genelistz=genelist;
% [X]=sc_norm(X,'type','deseq');
% [Y]=sc_norm(Y,'type','deseq');
% [Z]=sc_norm(Z,'type','deseq');
```

Label cells

```
cellidx=[1*ones(size(X,2),1); 2*ones(size(Y,2),1); 3*ones(size(Z,2),1)];
```

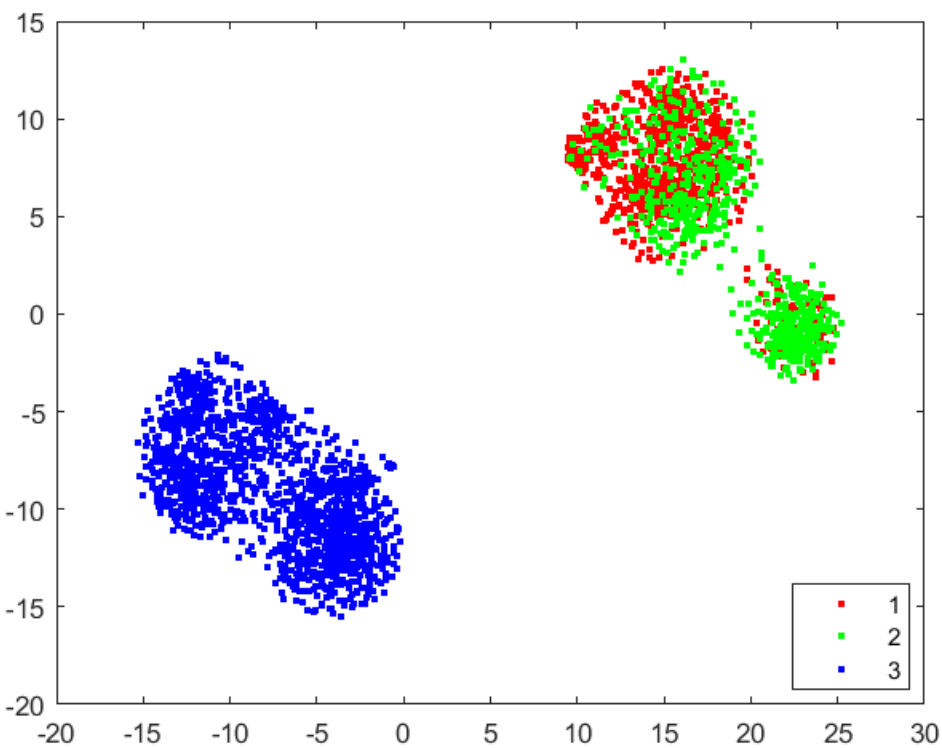
PCA

```
[~,s]=pca([X Y Z]');
gscatter(s(:,1),s(:,2),cellidx,'','',8);
```



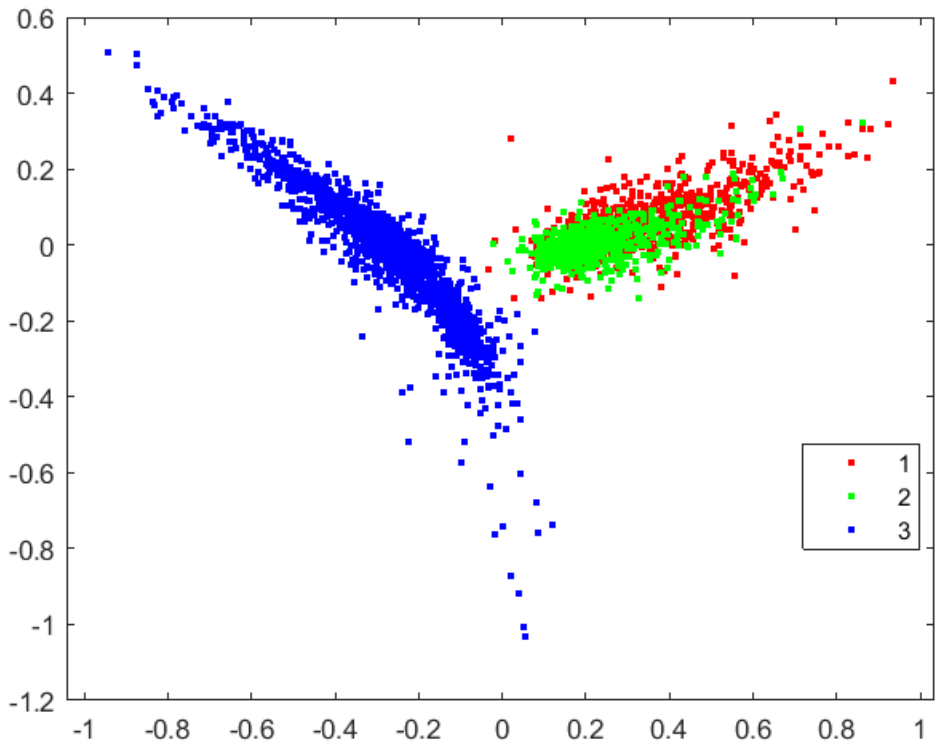
t-SNE

```
[s]=sc_tsne([X Y Z],2);
gscatter(s(:,1),s(:,2),cellidx,'','',8);
```



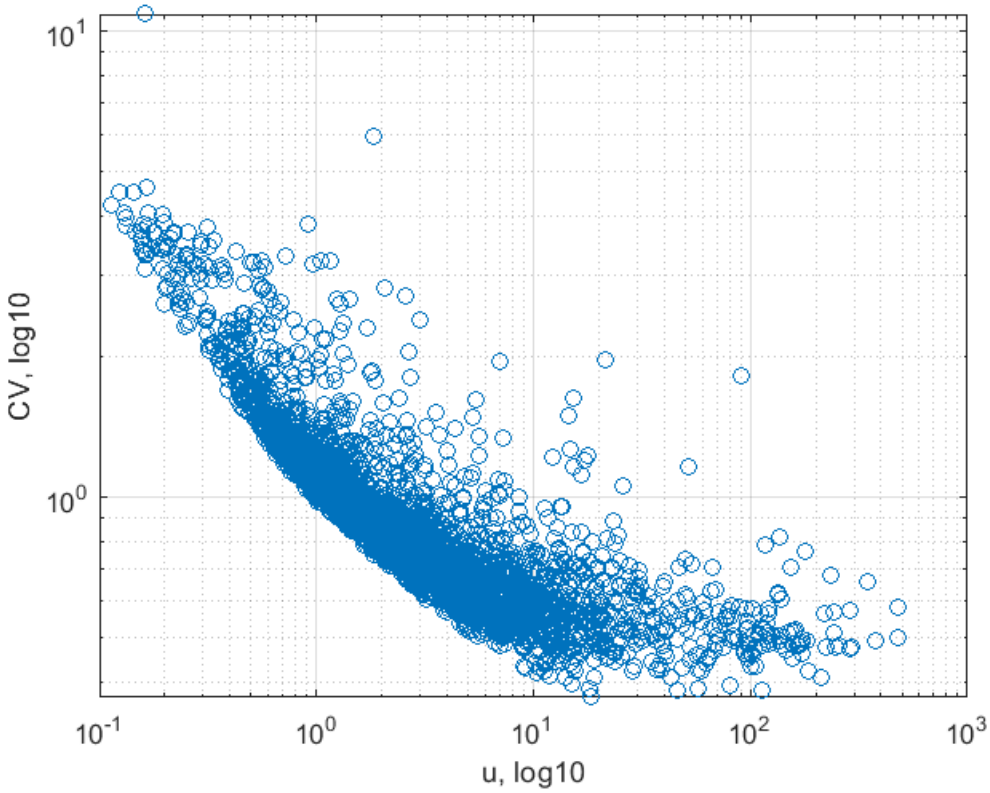
Diffusion Map

```
[s]=sc_diffuse([X Y Z]);
gscatter(s(:,1),s(:,2),cellidx,'','',8);
```

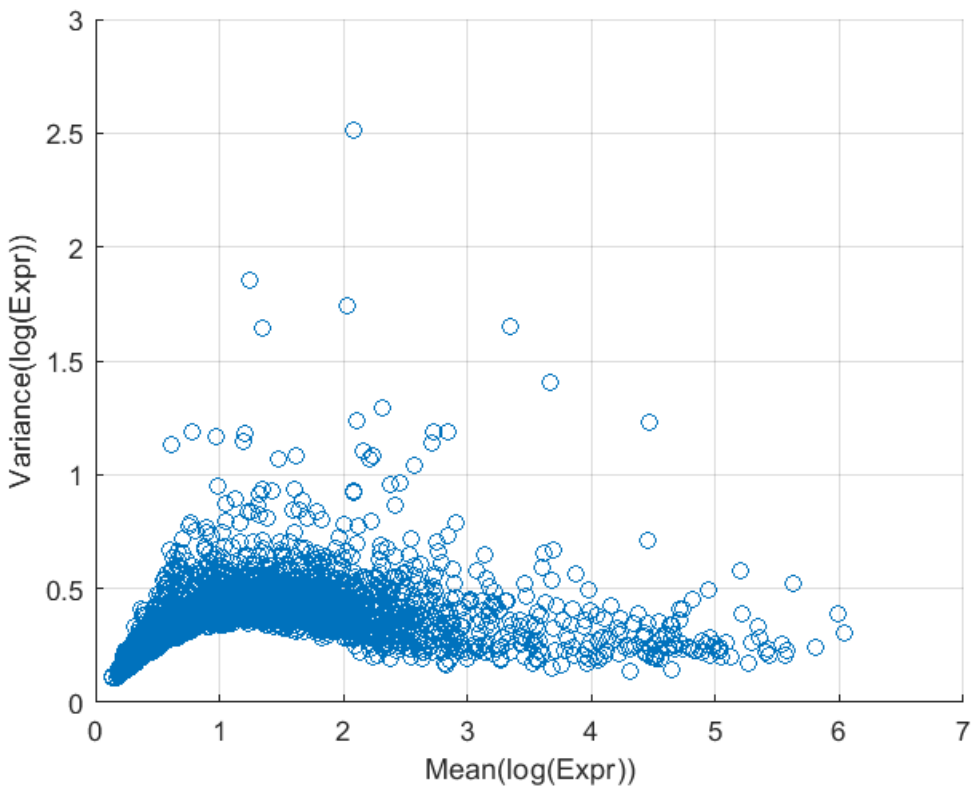


Scatter plots

```
figure;
sc_scatter(X,genelistx,'mean_cv');
```

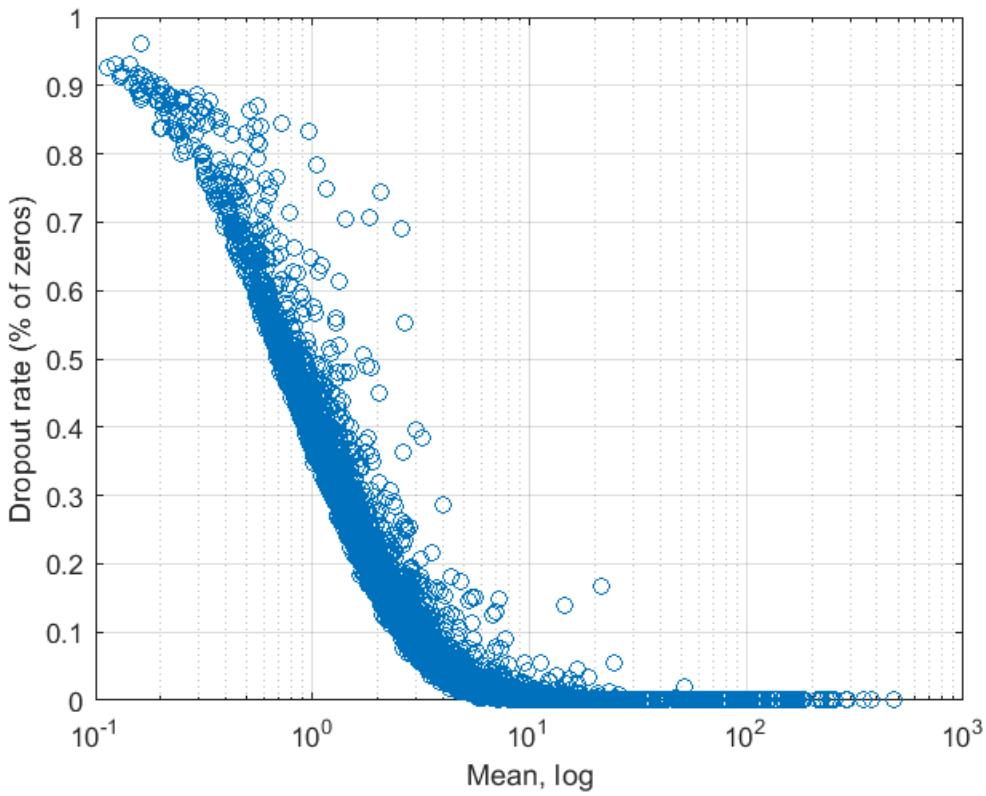



```
figure;
sc_scatter(X,genelistx,'meanlg_varlg');
```



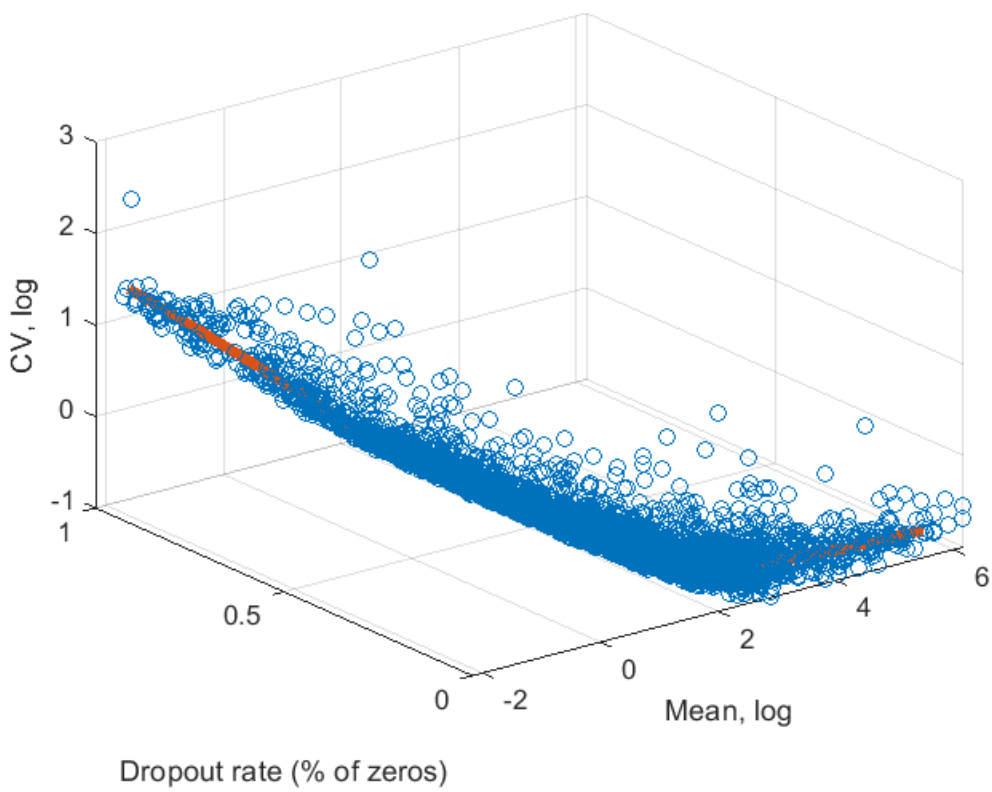

```
figure;
```

```
sc_scatter(X,genelistx,'mean_dropr');
```



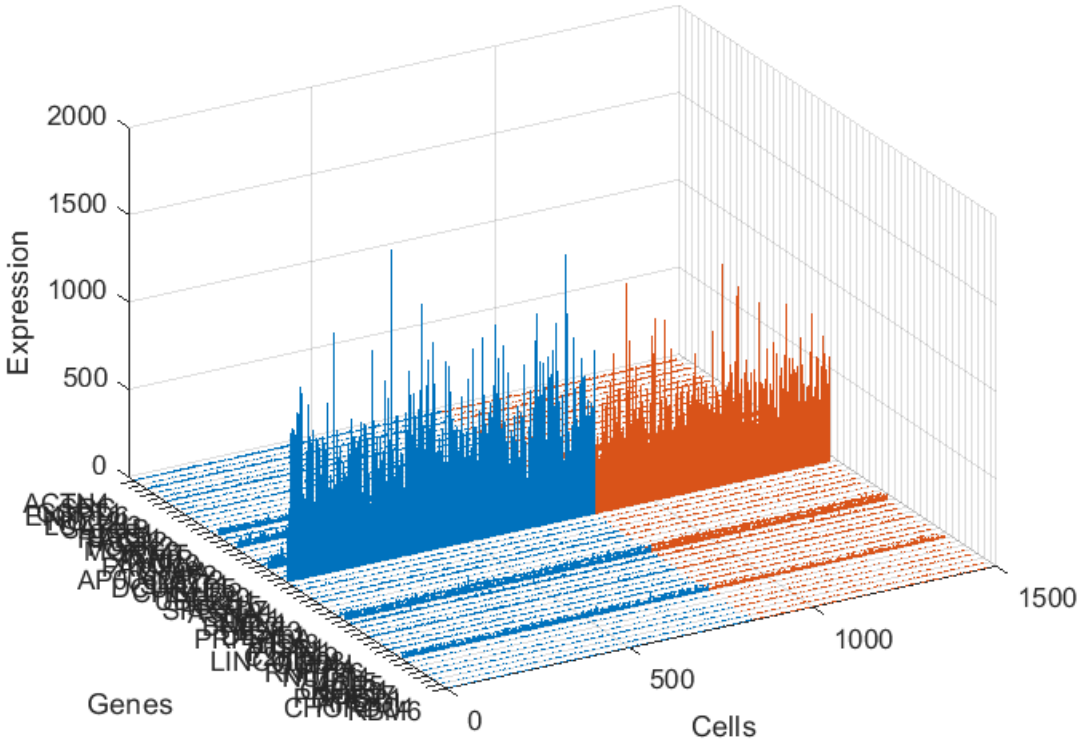
3D scatter plot with spline fit

```
figure;
sc_scatter3(X,genelistx,true,true);
```



Feature selection and show top 50 differentially deviated (DD) genes

```
T=sc_splinefit2(X,Y,genelistx,genelisty);
T=sortrows(T,size(T,2),'descend');
[~,idx1]=ismember(table2array(T(:,1)),genelistx);
[~,idx2]=ismember(table2array(T(:,1)),genelisty);
figure;
sc_stem3(X(idx1,:),Y(idx2,:),genelistx(idx1),50);
```



The End

Demonstration of Clustering Functions in scGEAToolbox

Load example data

```
cdgea; % set working directory
% load('example_data/example10xdata2.mat','X','genelist');
[X,genelistx]=sc_readfile('example_data/GSM3204304_P_P_Expr.csv');
```

```
Reading example_data/GSM3204304_P_P_Expr.csv ...... done.
```

```
[Y,genelisty]=sc_readfile('example_data/GSM3204305_P_N_Expr.csv');
```

```
Reading example_data/GSM3204305_P_N_Expr.csv ...... done.
```

```
[X,genelistx]=sc_selectg(X,genelistx,3,1);
[Y,genelisty]=sc_selectg(Y,genelisty,3,1);
```

Intersection of common genes in X, Y and Z

```
[genelist]=intersect(genelistx,genelisty,'stable');
% Remove genes encoded in the mitochondrial genome
i=startsWith(genelist,'MT-');
genelist(i)=[];
[~,i1]=ismember(genelist,genelistx);
[~,i2]=ismember(genelist,genelisty);
X=X(i1,:); genelistx=genelist;
Y=Y(i2,:); genelisty=genelist;
```

Label cells

```
cellidx=[1*ones(size(X,2),1); 2*ones(size(Y,2),1)];
```

Cluster cells using SIMLR

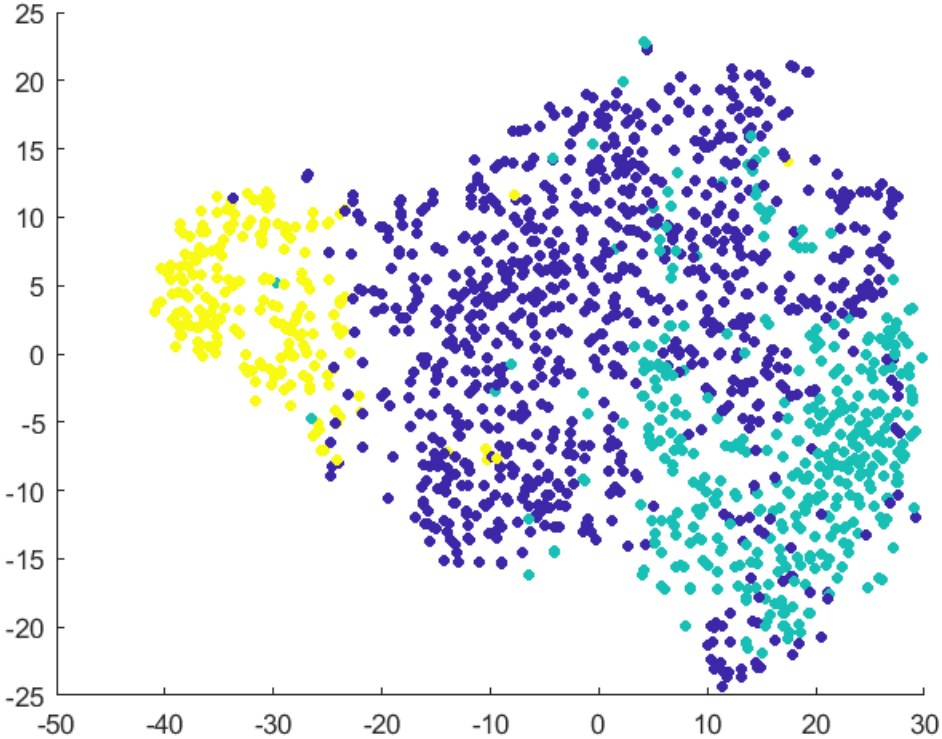
```
C=sc_cluster([X Y],'type','simlr');
```

```
To specify k, use RUN_SIMLR(X,k).
Iteration 10: error is 1.8059
Iteration 20: error is 1.7879
Iteration 30: error is 1.0038
Iteration 40: error is 0.80932
Iteration 50: error is 0.73918
Iteration 60: error is 0.69578
Iteration 70: error is 0.66487
Iteration 80: error is 0.64181
Iteration 90: error is 0.62523
Iteration 100: error is 0.6085
Iteration 110: error is 0.59795
Iteration 120: error is 0.58727
Iteration 130: error is 0.57768
Iteration 140: error is 0.57028
```

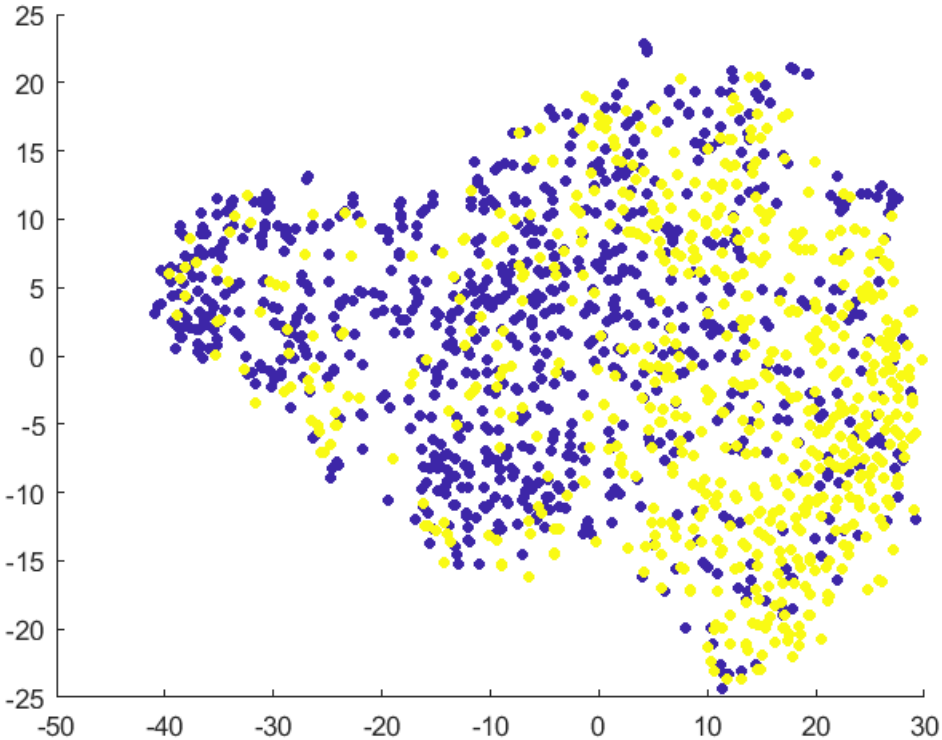
Plot the clustering result

```
s=sc_tsne([X Y],2,false,true,false);   % s=sc_tsne(X,ndim,plotit,donorm,dolog1p);
```

```
figure;
scatter(s(:,1),s(:,2),20,C,'filled')
```




```
figure;
scatter(s(:,1),s(:,2),20,cellidx,'filled')
```



Using SC3 example data yan.csv

```
[X,genelist]=sc_readtsvfile('example_data/yan.csv');
```

```
Reading example_data/yan.csv ...... done.
```

```
t=readtable('example_data\yan_celltype.txt');
celltypelist=string(t.cell_type1);
rng(235); showlegend=true;
% rng(111); showlegend=false;
% rng(113); showlegend=true;

s=sc_tsne(X,2);
c1=sc_sc3(X,6);
```

```
CLUSTER ENSEMBLES using CSPA
wgraph: writing graph0

C:\Users\jcai\Documents\GitHub\scGEATool\thirdparty\ClusterPack>pmetis
 graph0 6
**********************************************************************

  METIS 3.0   Copyright 1997, Regents of the University of Minnesota

Graph Information ---------------------------------------------------
  Name: graph0, #Vertices: 90, #Edges: 1111, #Parts: 6

Recursive Partitioning... -------------------------------------------
  6-way Edge-Cut: 16923384, Balance:  1.13
```

```
c2=run_simlr(X,6);
```

```
Iteration 10: error is 0.21153
Iteration 20: error is 1.1375
Iteration 30: error is 0.62867
Iteration 40: error is 0.25251
Iteration 50: error is 0.15781
Iteration 60: error is 0.26267
Iteration 70: error is 0.10414
Iteration 80: error is 0.074384
Iteration 90: error is 0.33669
Iteration 100: error is 0.35016
Iteration 110: error is 0.17267
Iteration 120: error is 0.27486
Iteration 130: error is 0.26362
Iteration 140: error is 0.10059
Iteration 150: error is 0.14544
```

```
c3=run_soptsc(X,'k',6);
```

```
Iter  Err
1, 6.867832
2, 5.876795
3, 5.042887
4, 4.362326
5, 3.807127
6, 3.351183
7, 2.973775
8, 2.658954
9, 0.711817
10, 0.112101
11, 0.103572
12, 0.098797
13, 0.093055
14, 0.087015
15, 0.081145
16, 0.075632
17, 0.070527
18, 0.062194
19, 0.050105
```

```
% Result of SC3/R pacakge
load example_data/sc3_results.txt
c0=sc3_results;
```

Compare clustering results between SC3/R vs SC3, SIMILR and SoptSC

```
Cal_NMI(c0,c1)
```

```
ans = 0.9205
```

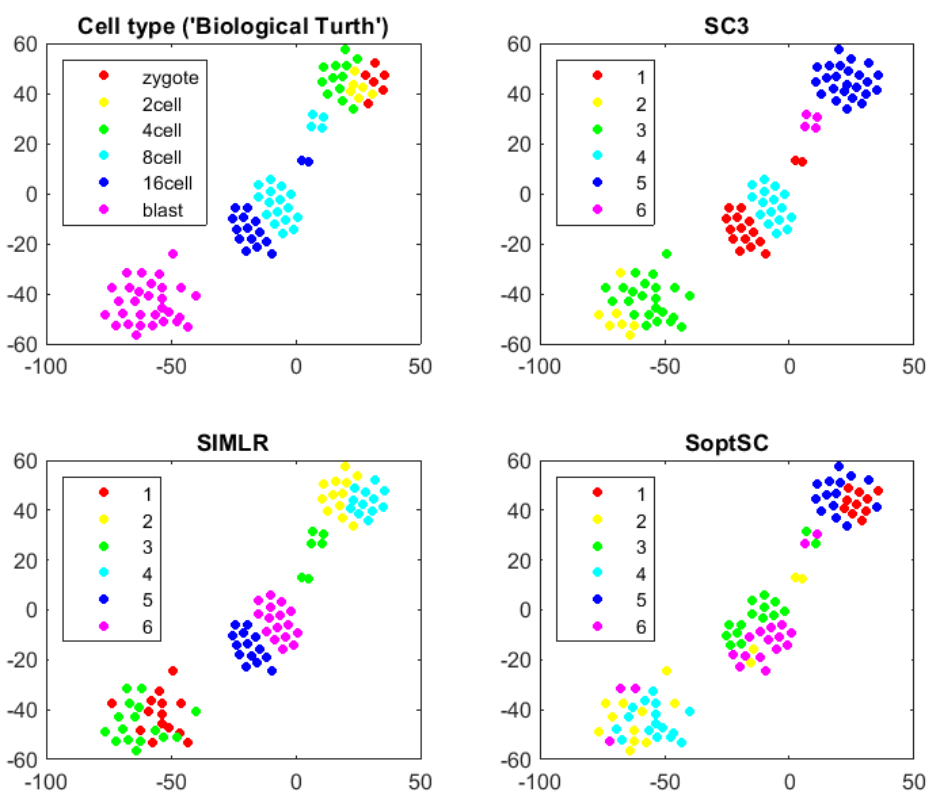
```
Cal_NMI(c0,c2)
```

```
ans = 0.8200
```

```
Cal_NMI(c0,c3)
```

```
ans = 0.5464
```

```
fh=figure;
subplot(2,2,1)
gscatter(s(:,1),s(:,2),celltypelist)
if showlegend, legend('Location','northwest'); else, legend off; end
title('Cell type (''Biological Turth'')')

subplot(2,2,2)
gscatter(s(:,1),s(:,2),c1)
if showlegend, legend('Location','northwest'); else, legend off; end
title('SC3')

subplot(2,2,3)
gscatter(s(:,1),s(:,2),c2)
if showlegend, legend('Location','northwest'); else, legend off; end
title('SIMLR')

subplot(2,2,4)
gscatter(s(:,1),s(:,2),c3)
if showlegend, legend('Location','northwest'); else, legend off; end
title('SoptSC')
fh.Position=[fh.Position(1) fh.Position(2)-100 fh.Position(3)+100
fh.Position(4)+100];
```



The End

Demonstration of Pseudotime Analysis and Gene Network Functions in scGEAToolbox

Load examle data set, X

```
cdgea; % set working directory
[X,genelist]=sc_readfile('example_data/GSM3044891_GeneExp.UMIs.10X1.txt');
```

```
Reading example_data/GSM3044891_GeneExp.UMIs.10X1.txt ...... done.
```

Select genes with at least 3 cells having more than 5 reads per cell.

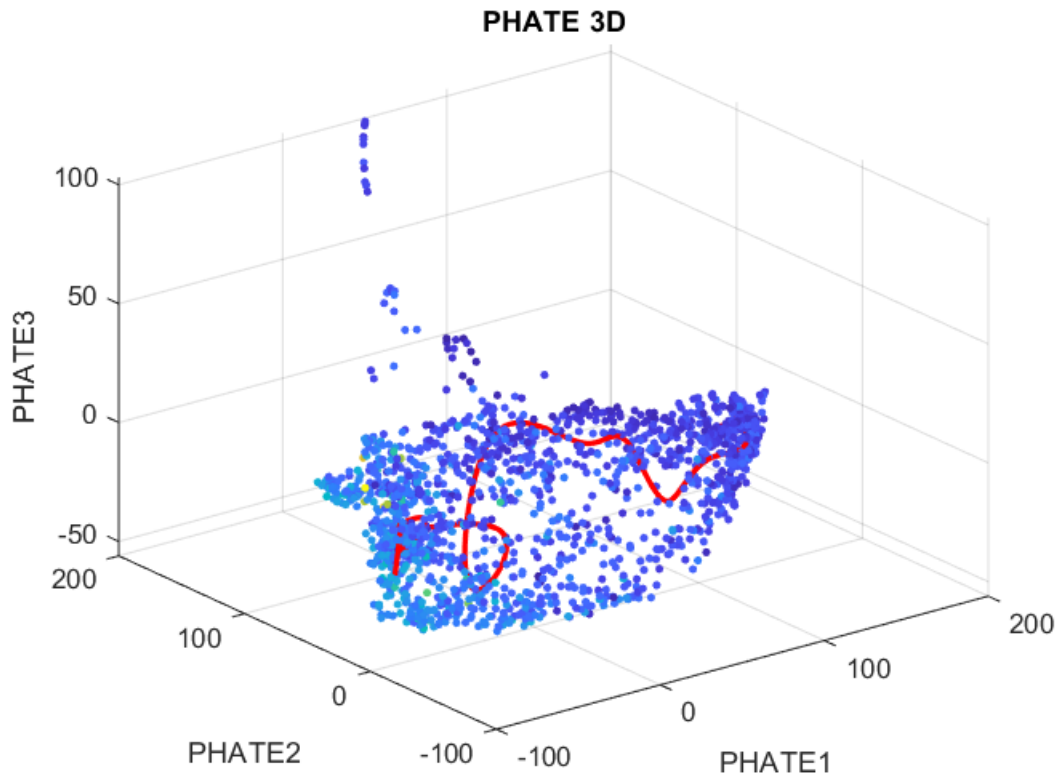
```
[X,genelist]=sc_selectg(X,genelist,5,3);
```

Trajectory analysis using the PHATE+splinefit method

s=run_phate(X,3,false,true); [t,xyz1]=i_pseudotime_by_splinefit(s,1); hold on plot3(xyz1(:,1),xyz1(:,2),xyz1(:,3),'-r','linewidth',2);

```
% Calculte pseudotime T
figure;
t=sc_trajectory(X,"type","splinefit","plotit",true);
```

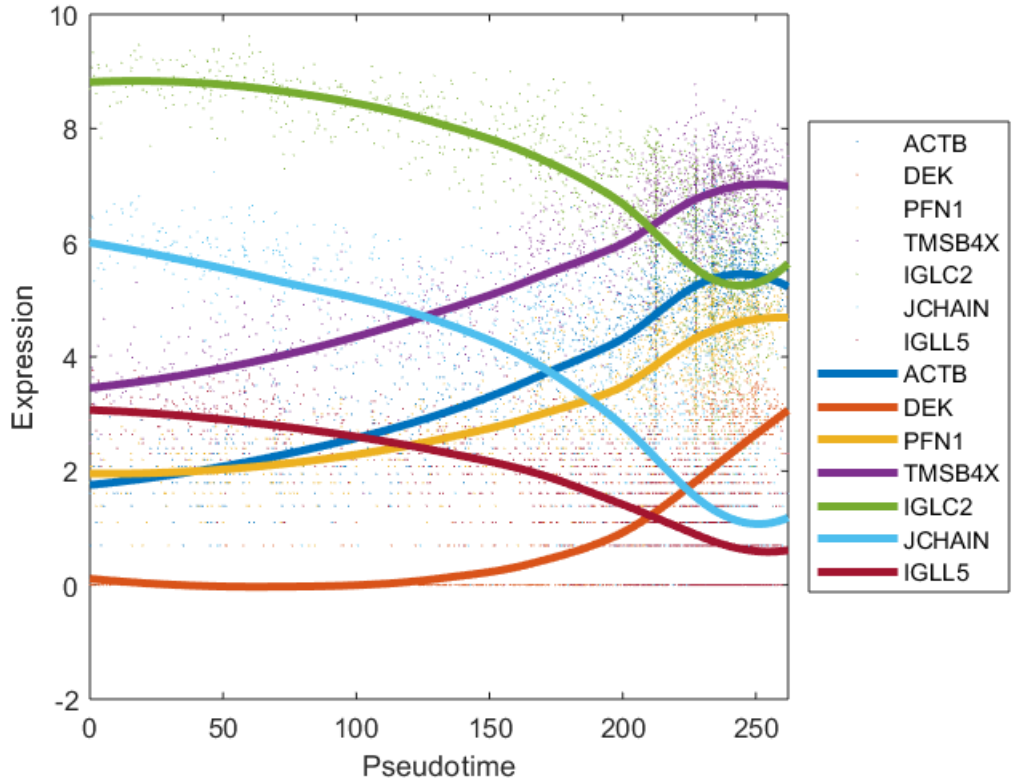
```
Doing PCA
PCA using random SVD
PCA took 0.44292 seconds
using alpha decaying kernel
Computing alpha decay kernel:
Number of samples = 1567
First iteration: k = 100
Number of samples below the threshold from 1st iter: 1566
Using radius based search for the rest
   Symmetrize affinities
   Done computing kernel
Computing kernel took 0.2895 seconds
Make kernel row stochastic
Running PHATE without landmarking
Diffusing operator
```




```
12
12
14
10
13
12
26
17
13
17
```

Plot gene expression profile of cells ordered according to their pseudotime T.

```
r=corr(t,X',type,spearman); % Calculate linear correlation between gene
expression profile and T
[~,idxp]= maxk(r,4);  % Select top 4 positively correlated genes
[~,idxn]= mink(r,3);  % Select top 3 negatively correlated genes
selectedg=genelist([idxp idxn]);

% Plot expression profile of the 5 selected genes
figure;
i_plot_pseudotimeseries(log(1+X),genelist,t,selectedg)
```



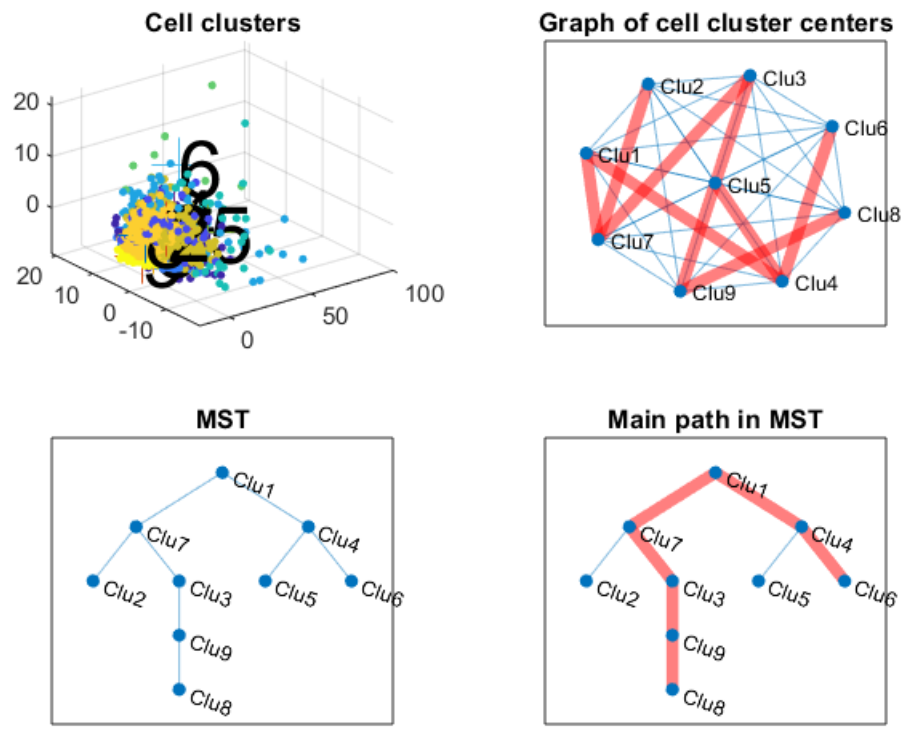

```
% % Nonlinear correlation
%
% r=zeros(size(X,1),1);
% for k=1:size(X,1)
%     k
%     r(k)=distcorr(t,double(X(k,:))');
% end
% [~,idxp]= maxk(r,3);
% [~,idxn]= mink(r,2);
% selectedg=genelist([idxp; idxn]);
% figure;
% i_plot_pseudotimeseries(log(1+X),genelist,t,selectedg)
```

Trajectory analysis using TSCAN

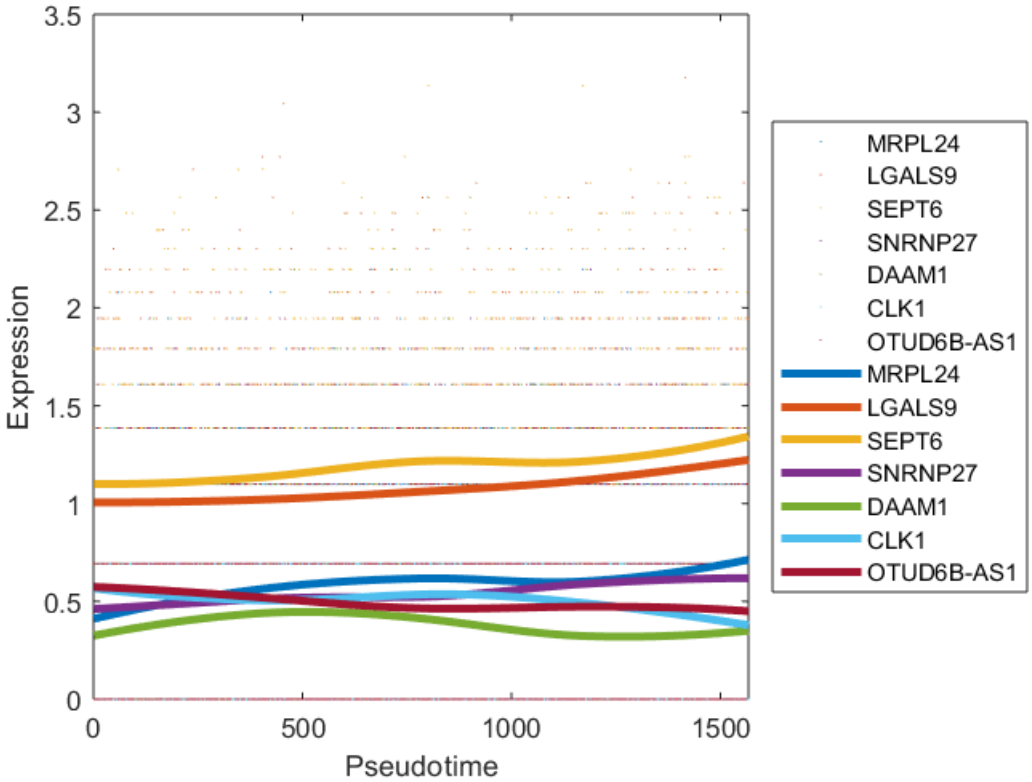
Calculte pseudotime T

```
figure;
t=sc_trajectory(X,"type","tscan","plotit",true);
```

```
Warning: Failed to converge in 100 iterations for gmdistribution with 9 components
```



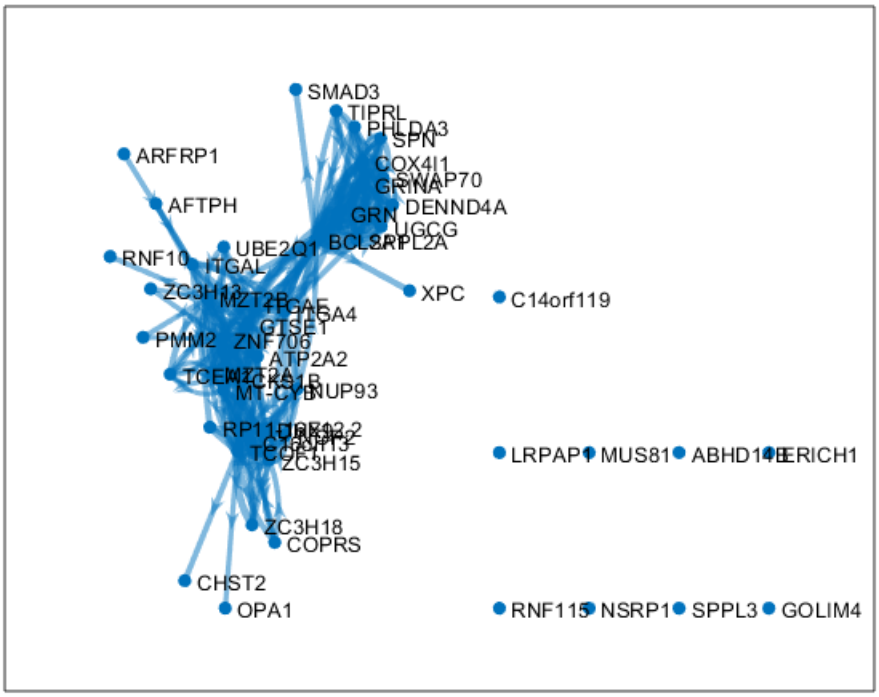

```
r=corr(t,X','type','spearman'); % Calculate linear correlation between gene
expression profile and T
[~,idxp]= maxk(r,4);  % Select top 4 positively correlated genes
[~,idxn]= mink(r,3);  % Select top 3 negatively correlated genes
selectedg=genelist([idxp idxn]);

% Plot expression profile of the 5 selected genes
figure;
i_plot_pseudotimeseries(log(1+X),genelist,t,selectedg)
```



Construct single-cell gene regulatory network (scGRN)

Using principal component regression (PCNet) method

```
X50=X(1:50,:);
genelist50=genelist(1:50);
A=sc_pcnet(X50);

% Plot constructed network
%
A=A.*(abs(A)>quantile(abs(A(:)),0.9));
G=digraph(A,genelist50);
LWidths=abs(5*G.Edges.Weight/max(G.Edges.Weight));
LWidths(LWidths==0)=1e-5;
figure;
plot(G,'LineWidth',LWidths);
```




```
p.MarkerSize = 7;
p.Marker = 's';
p.NodeColor = 'r';
```

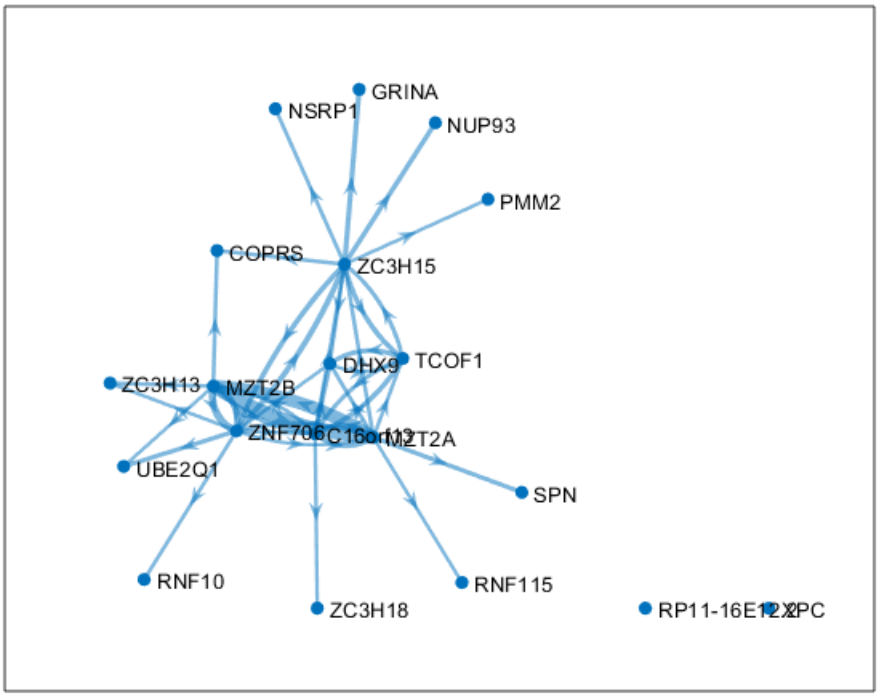
Using GENIE3 method

```
X20=X(1:20,:);
genelist20=genelist(1:20);
A=run_genie3(X20);
```

```
Tree method: RF
K: sqrt
Number of trees: 1000

Gene 1/20...
Table prediction_values = 1567000 x 1
Gene 2/20...
Table prediction_values = 1567000 x 1
Gene 3/20...
Table prediction_values = 1567000 x 1
Gene 4/20...
Table prediction_values = 1567000 x 1
Gene 5/20...
Table prediction_values = 1567000 x 1
Gene 6/20...
```

```
% Plot constructed network
%
A=A.*(abs(A)>quantile(abs(A(:)),0.9));
G=digraph(A,genelist20);
```

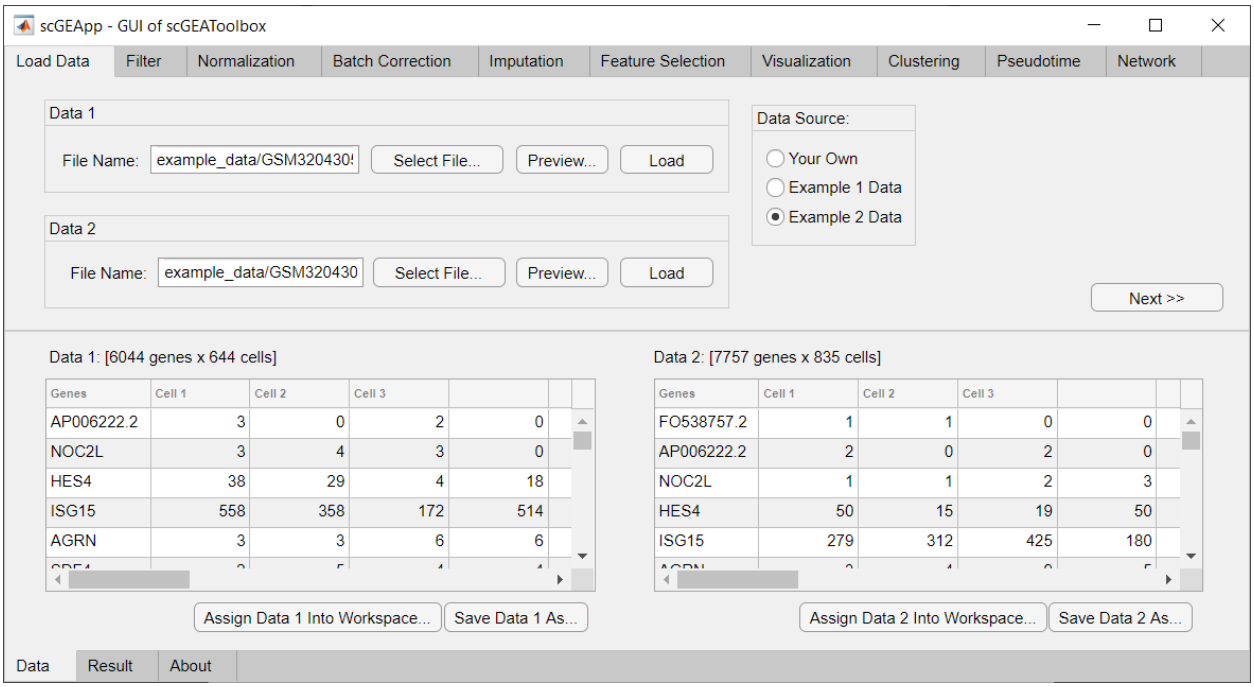
```
LWidths=abs(5*G.Edges.Weight/max(G.Edges.Weight));
LWidths(LWidths==0)=1e-5;
figure;
plot(G,'LineWidth',LWidths);
```

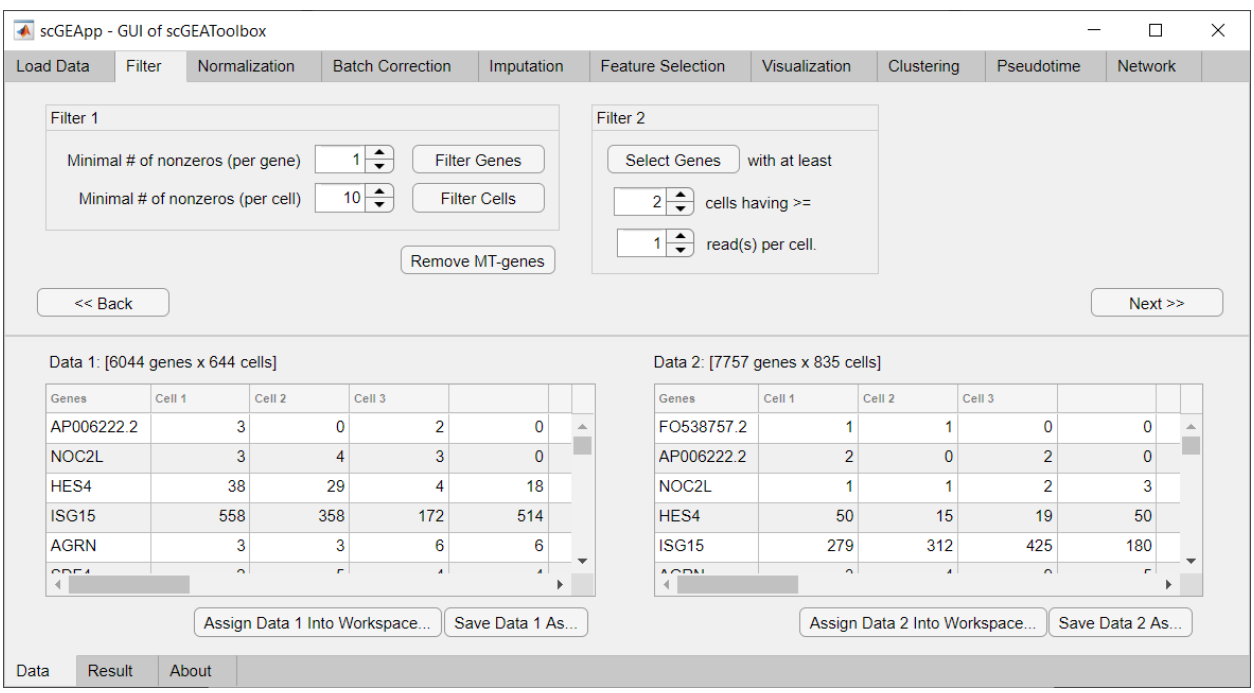


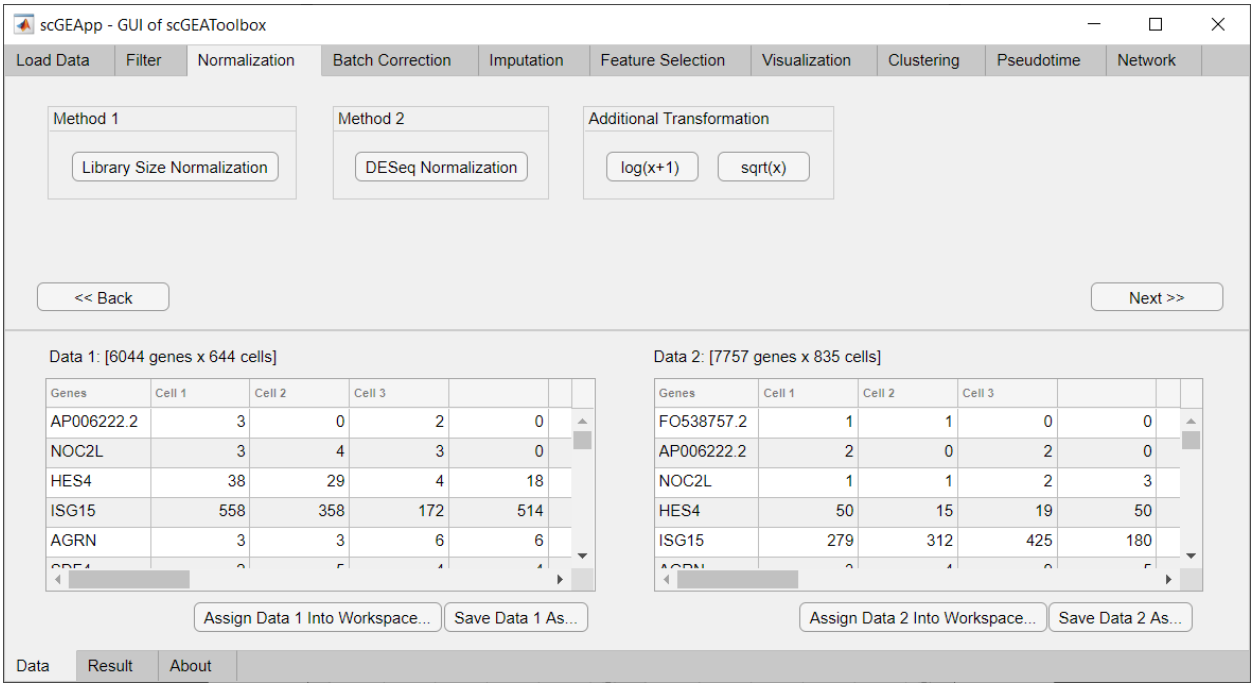

```
p.MarkerSize = 7;
p.Marker = 's';
p.NodeColor = 'r';
```

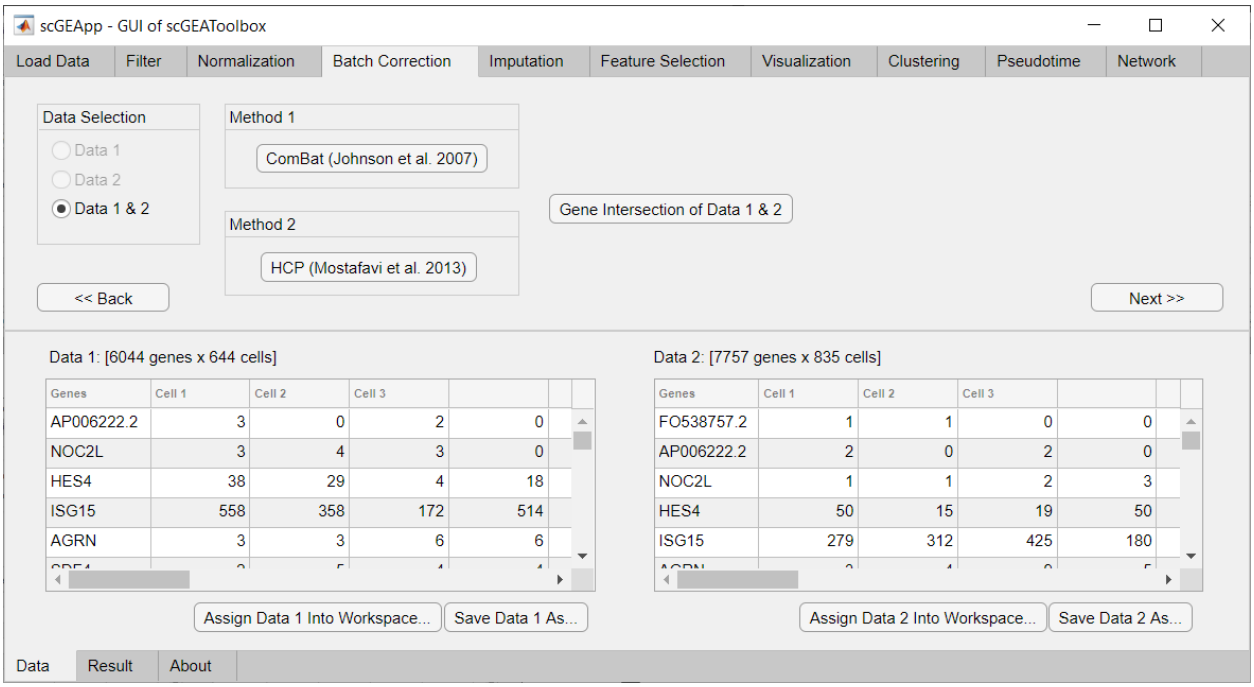
The End

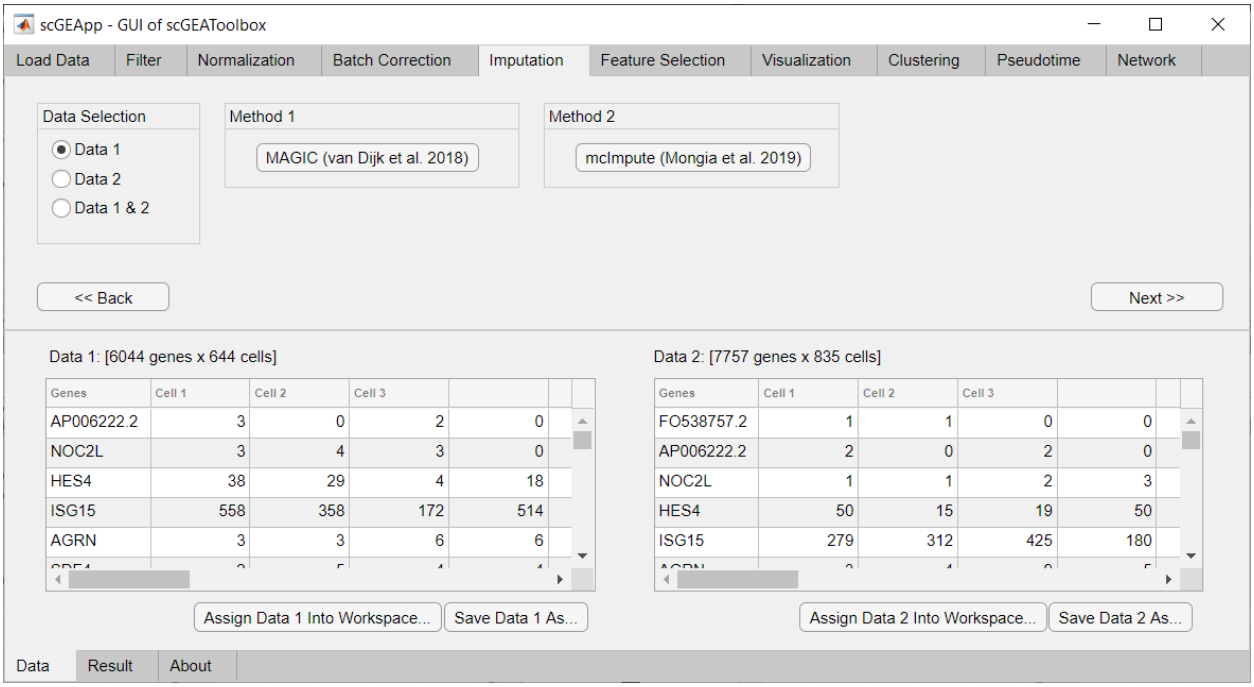
Supplementary Fig. S1. Screenshots of scGEApp.

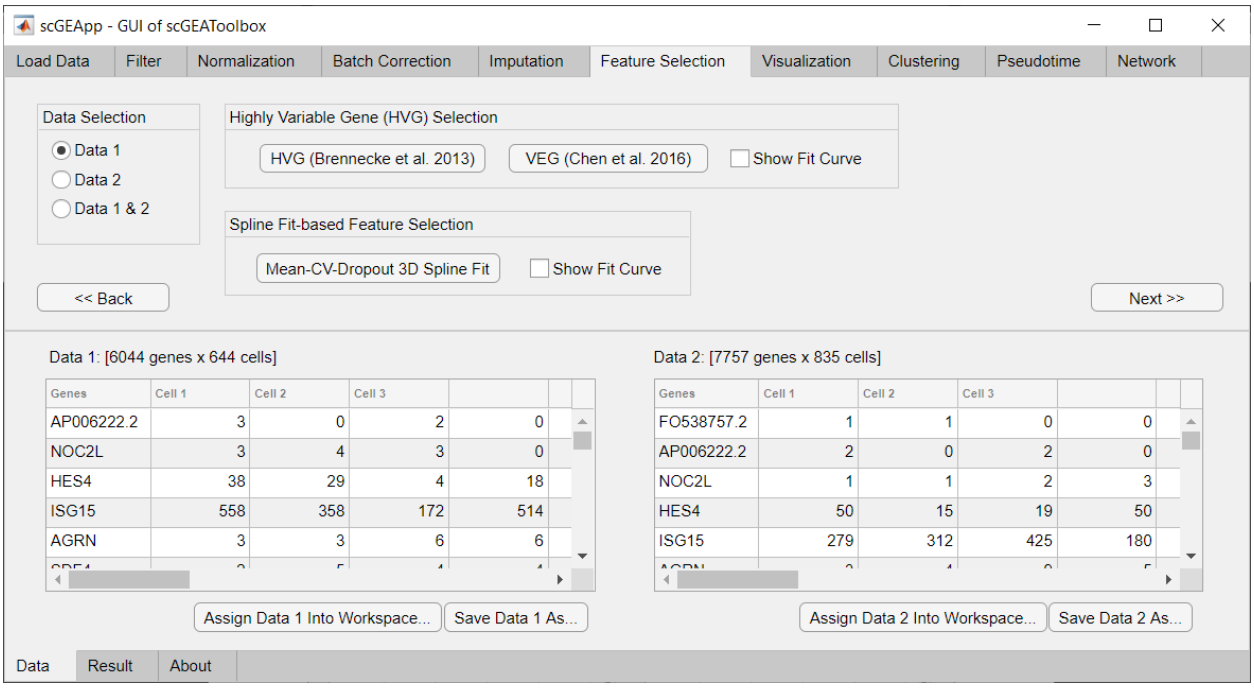


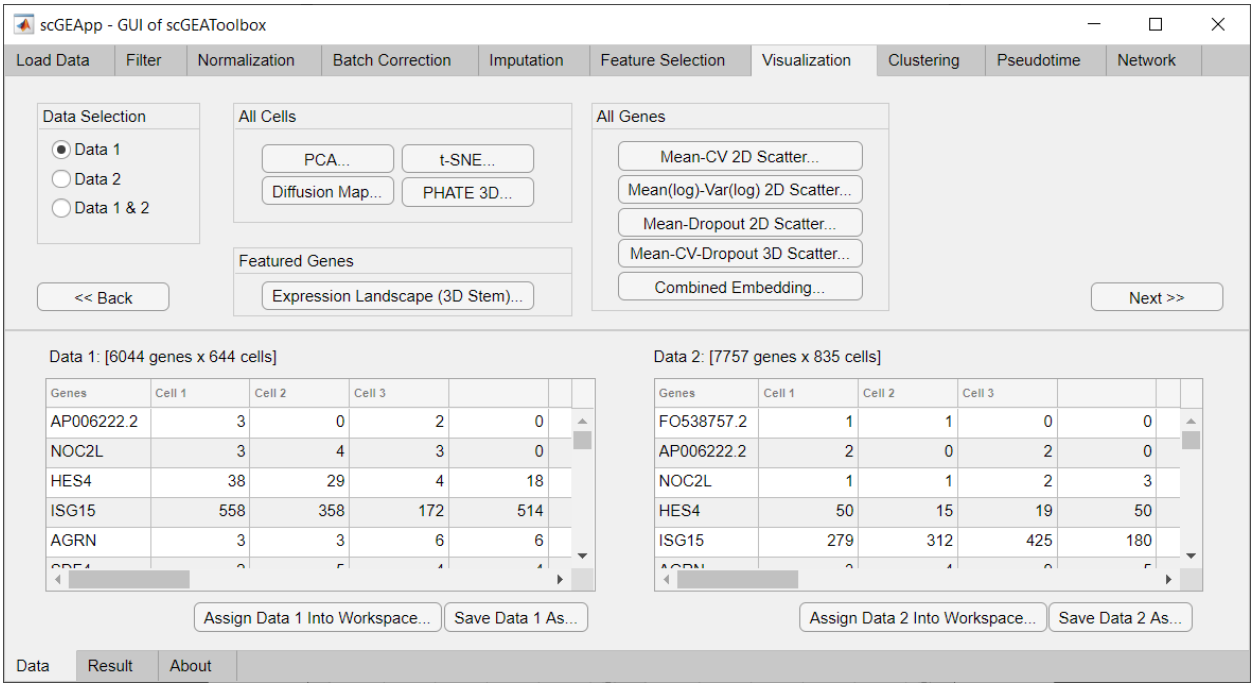


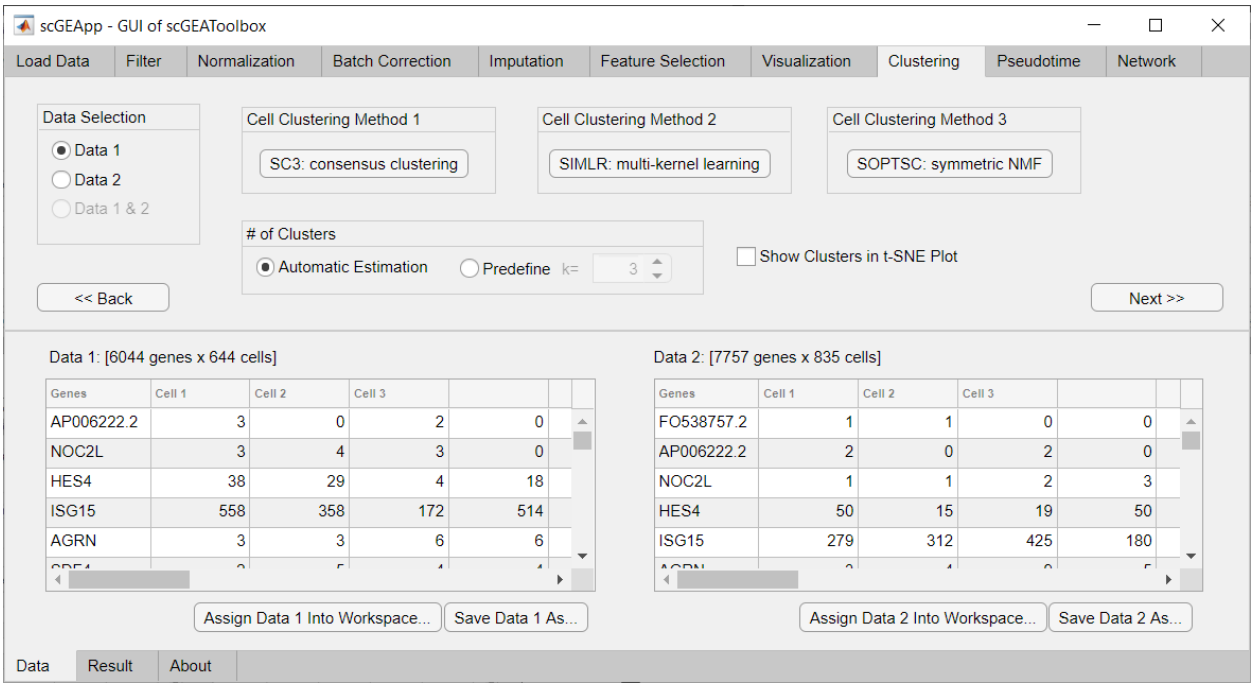

scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Method 1
Library Size Normalization
Method 2
DESeq Normalization
Additional Transformation
log(x+1)
sqrt(x)
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About

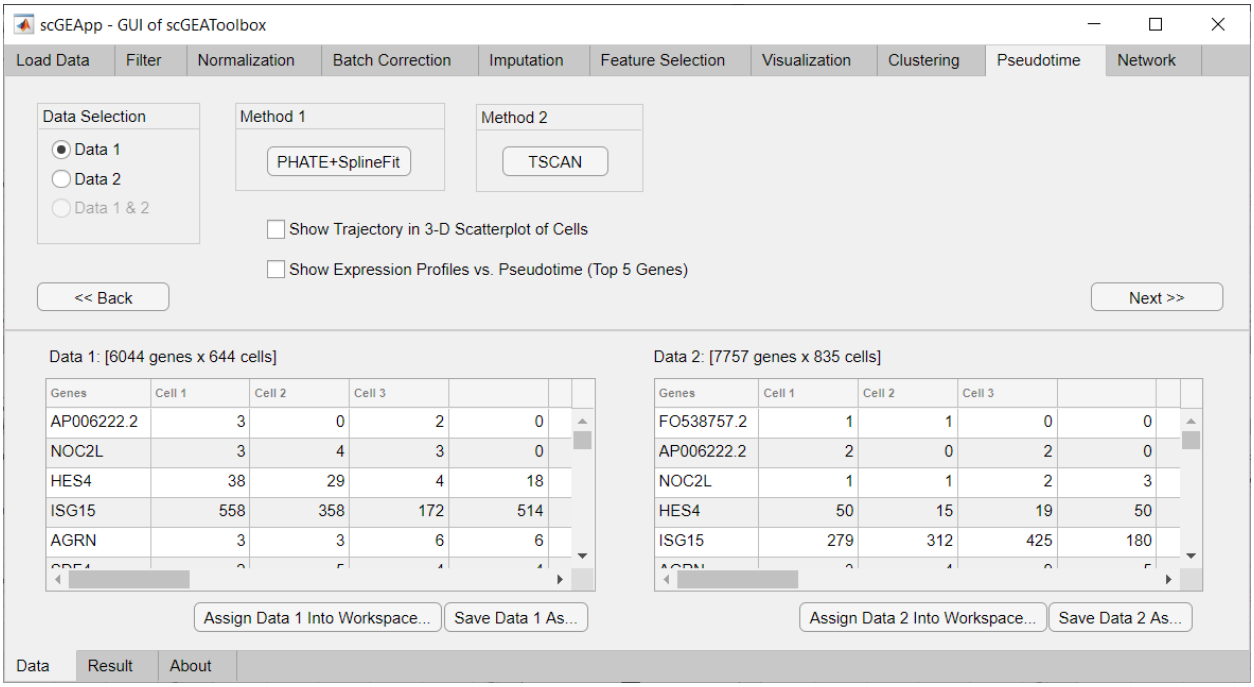

scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Method 1
ComBat (Johnson et al. 2007)
Method 2
HCP (Mostafavi et al. 2013)
Gene Intersection of Data 1 & 2
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About

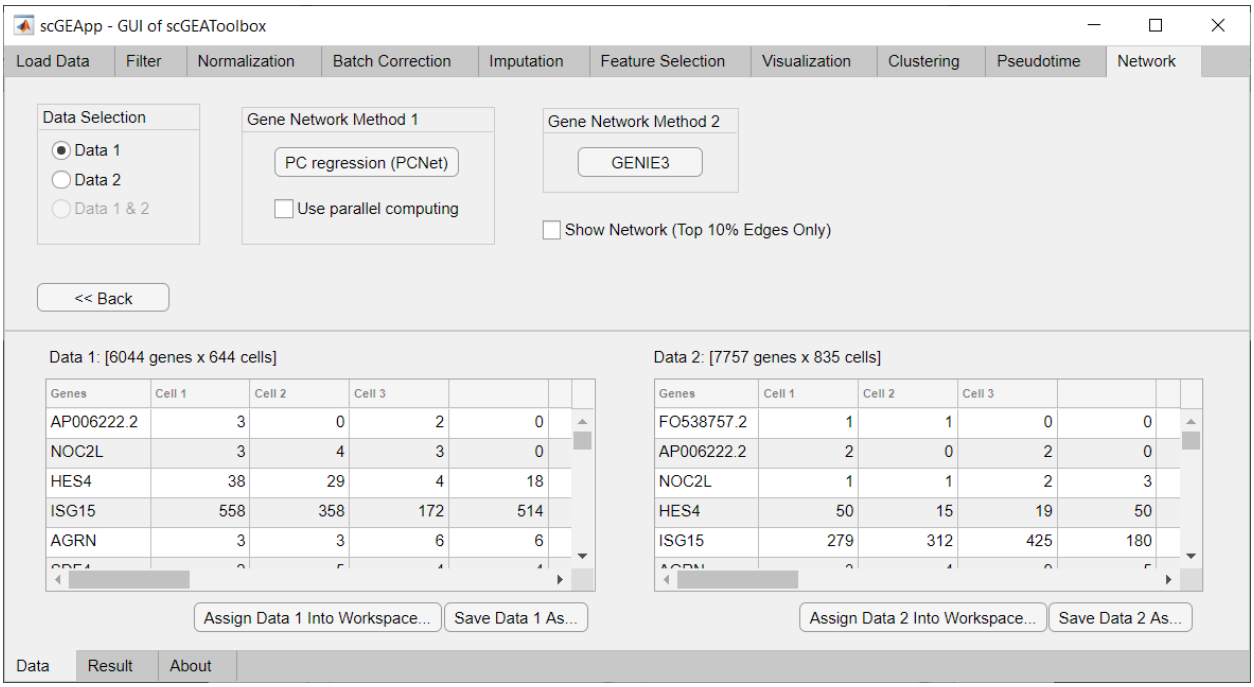

scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Method 1
MAGIC (van Dijk et al. 2018)
Method 2
mcImpute (Mongia et al. 2019)
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About


scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Highly Variable Gene (HVG) Selection
HVG (Brennecke et al. 2013)
VEG (Chen et al. 2016)
Show Fit Curve
Spline Fit-based Feature Selection
Mean-CV-Dropout 3D Spline Fit
Show Fit Curve
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About


scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
All Cells
PCA...
t-SNE...
Diffusion Map...
PHATE 3D...
Featured Genes
Expression Landscape (3D Stem)...
All Genes
Mean-CV 2D Scatter...
Mean(log)-Var(log) 2D Scatter...
Mean-Dropout 2D Scatter...
Mean-CV-Dropout 3D Scatter...
Combined Embedding...
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About


scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Cell Clustering Method 1
SC3: consensus clustering
Cell Clustering Method 2
SIMLR: multi-kernel learning
Cell Clustering Method 3
SOPTSC: symmetric NMF
# of Clusters
Automatic Estimation
Predefine
k=
3
Show Clusters in t-SNE Plot
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About


scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Method 1
PHATE+SplineFit
Method 2
TSCAN
Show Trajectory in 3-D Scatterplot of Cells
Show Expression Profiles vs. Pseudotime (Top 5 Genes)
<< Back
Next >>
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About


scGEApp - GUI of scGEAToolbox
Load Data
Filter
Normalization
Batch Correction
Imputation
Feature Selection
Visualization
Clustering
Pseudotime
Network
Data Selection
Data 1
Data 2
Data 1 & 2
Gene Network Method 1
PC regression (PCNet)
Use parallel computing
Gene Network Method 2
GENIE3
Show Network (Top 10% Edges Only)
<< Back
Data 1: [6044 genes x 644 cells]
Genes Cell 1 Cell 2 Cell 3
AP006222.2 3 0 2 0
NOC2L 3 4 3 0
HES4 38 29 4 18
ISG15 558 358 172 514
AGRN 3 3 6 6
Data 2: [7757 genes x 835 cells]
Genes Cell 1 Cell 2 Cell 3
FO538757.2 1 1 0 0
AP006222.2 2 0 2 0
NOC2L 1 1 2 3
HES4 50 15 19 50
ISG15 279 312 425 180
Assign Data 1 Into Workspace...
Save Data 1 As...
Assign Data 2 Into Workspace...
Save Data 2 As...
Data
Result
About