

Simulation in Different Scenario

Xinyi Lin

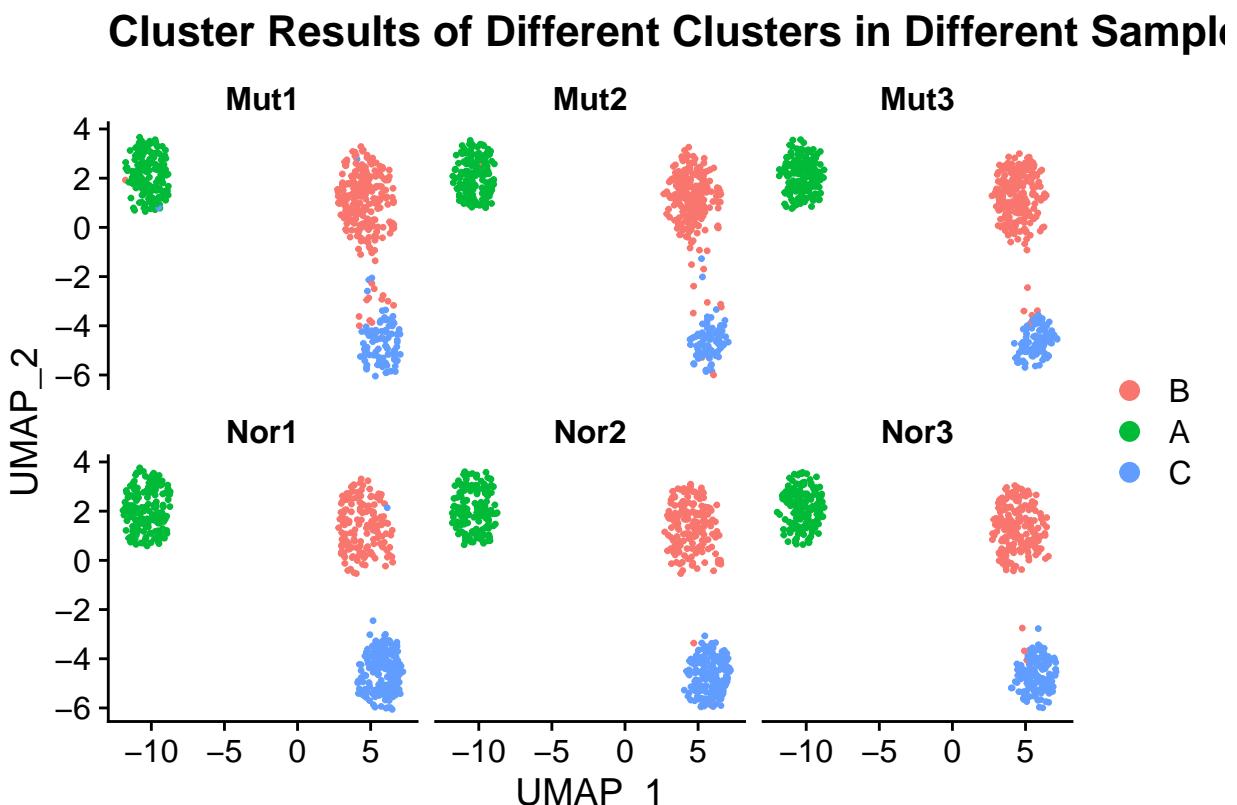
9/17/2020

Problems need to solve:

1. DCATS: label problems
2. scDC: functions don't work

Different batch size

Batch sizes are 500, 1000, 1500, 2000.



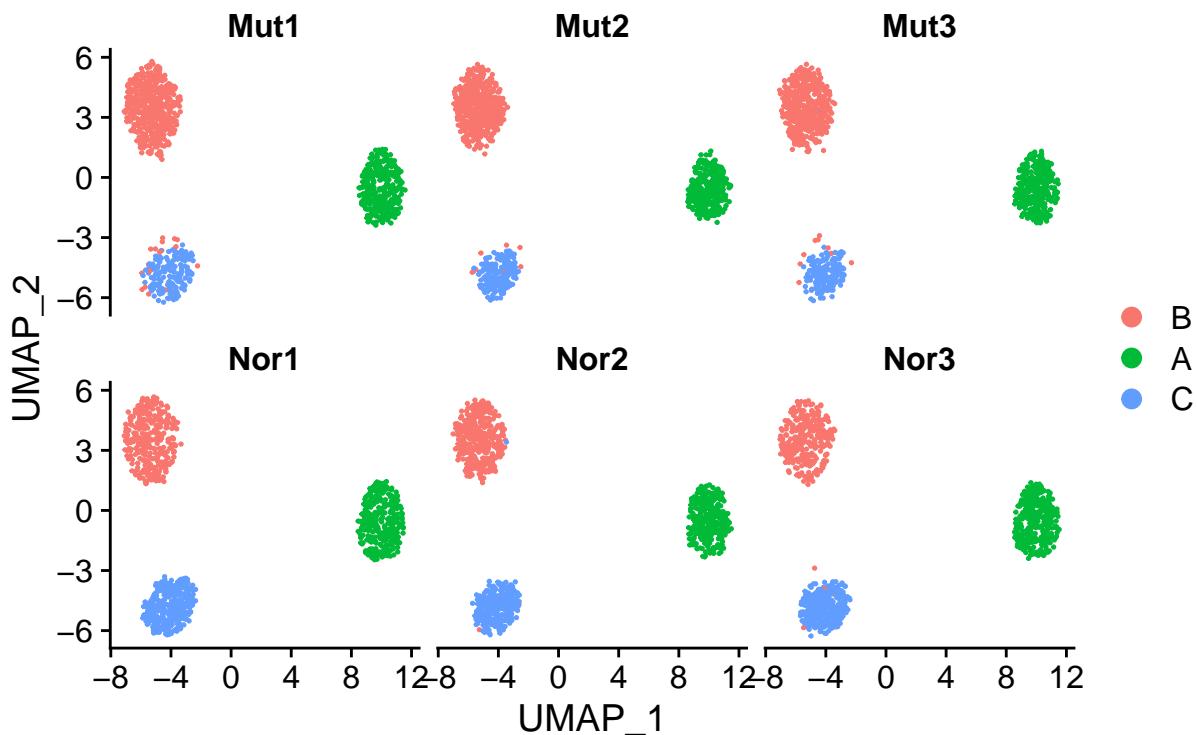
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 6.929452e-01  5.659978e-01 4.135431e-01 4.361157e-01
```

```

## 2      B 6.072355e-08 1.575240e-04 6.889945e-15 6.488365e-15
## 3      C 6.889945e-15 4.182365e-05 1.171748e-21 3.144598e-22
##
## $time_df
##   methods      time
## 1  fisher 0.007979155
## 2 sepckle 0.015957832
## 3 dcats 0.246373177

```

Cluster Results of Different Clusters in Different Samples

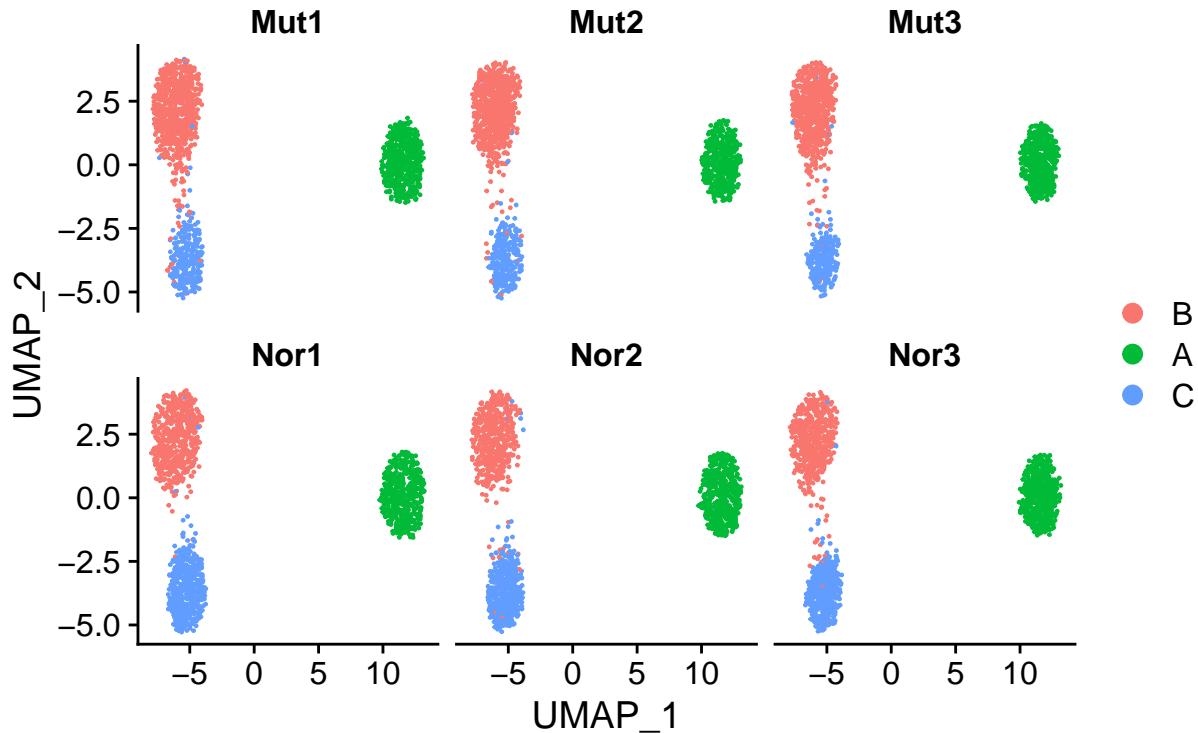


```

## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1      A 7.742517e-01 7.417927e-01 6.594604e-01 6.795236e-01
## 2      B 2.115282e-18 2.939822e-07 1.027788e-41 2.905624e-42
## 3      C 1.027788e-41 2.022382e-07 8.260421e-49 9.059806e-51
##
## $time_df
##   methods      time
## 1  fisher 0.011950016
## 2 sepckle 0.004987001
## 3 dcats 0.253323078

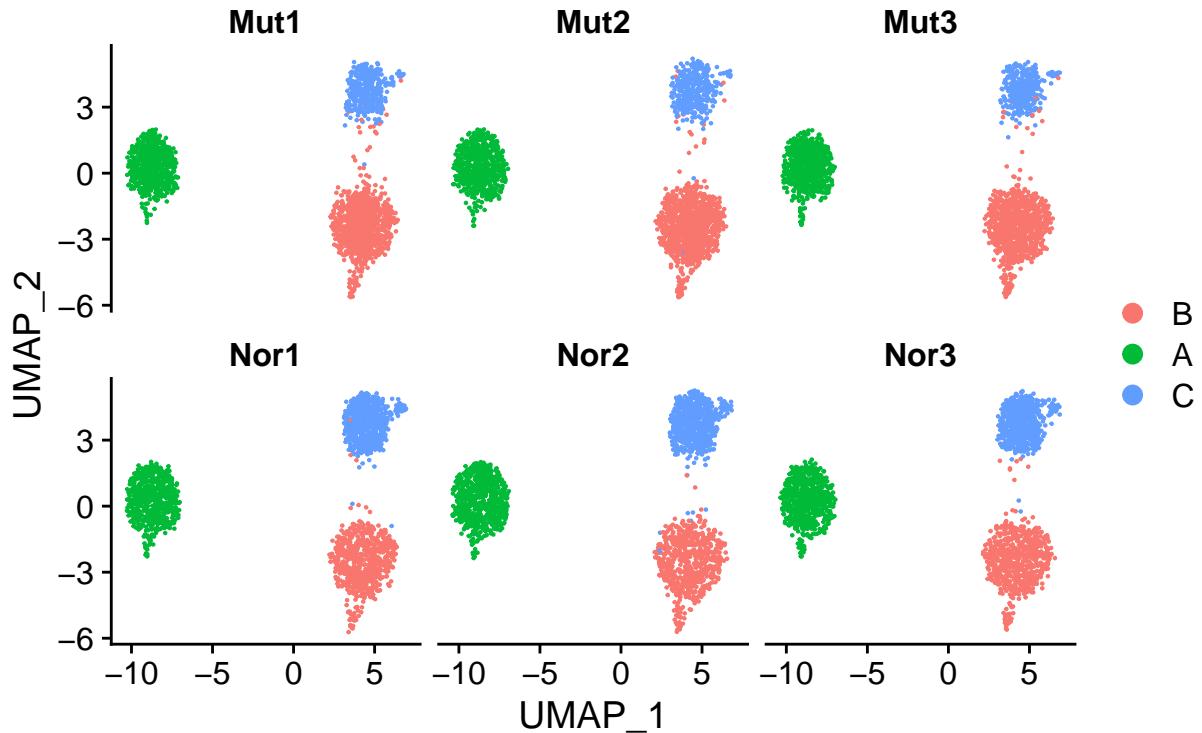
```

Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 7.457880e-01  5.578942e-01 5.456519e-01 5.606287e-01
## 2       B 4.352388e-08  1.224444e-08 2.210894e-44 8.788459e-45
## 3       C 2.210894e-44  8.998110e-09 1.732844e-51 6.719757e-53
##
## $time_df
##   methods      time
## 1  fisher 0.015955925
## 2 sepckle 0.004952192
## 3 dcats 0.259330988
```

Cluster Results of Different Clusters in Different Samples



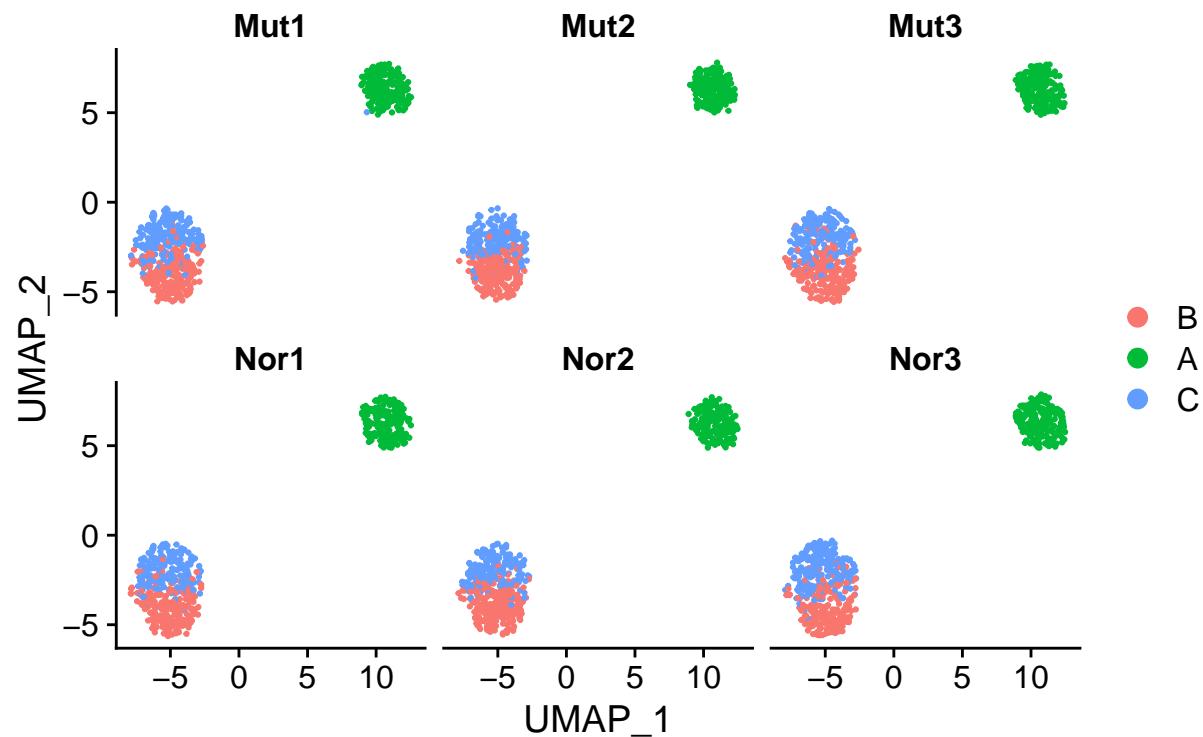
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 7.685354e-01  6.601008e-01 6.284930e-01 6.422941e-01
## 2       B 2.679213e-18  2.057987e-09 1.511594e-67 2.325315e-68
## 3       C 1.511594e-67  1.180343e-09 4.464267e-81 5.211656e-84
##
## $time_df
##   methods      time
## 1   fisher 0.019949198
## 2 sepckle 0.005979776
## 3 dcats 0.235368967
```

Different de.prob

The de.prob of cluster B, C are 0.02, 0.04, 0.06, 0.08.

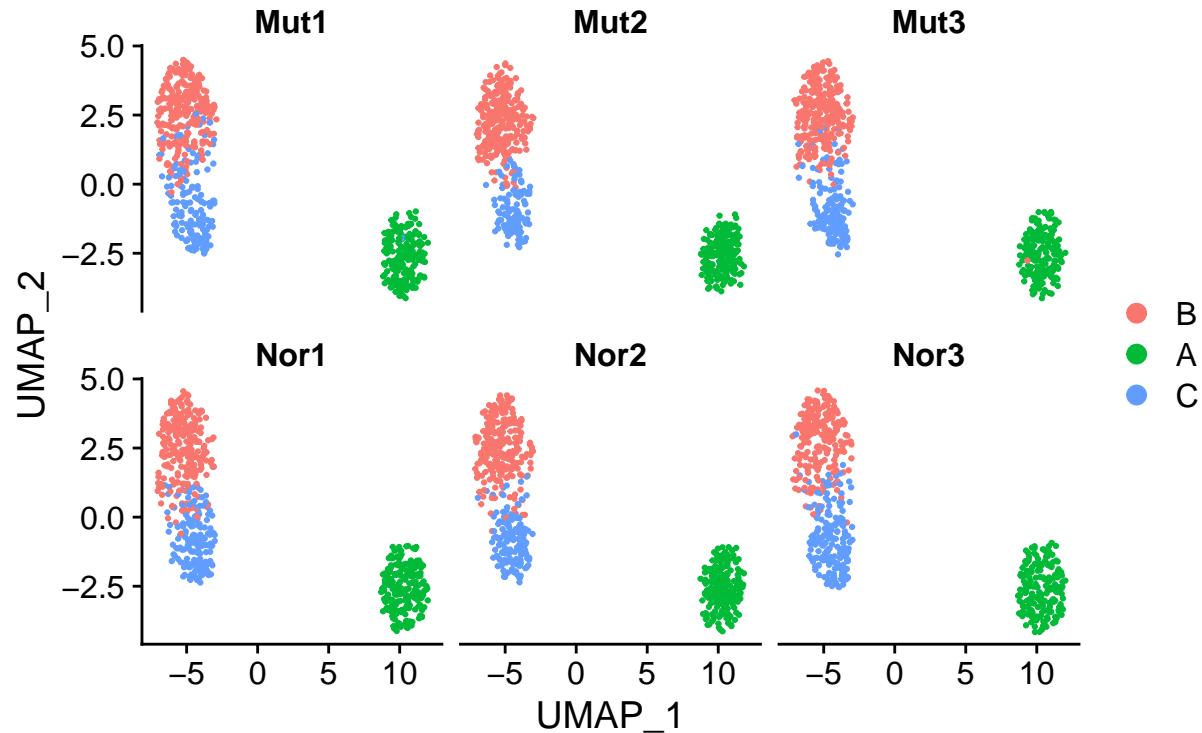
When de.prob equals to 0.04, only the fisher's exact test give the “correct” result.

Cluster Results of Different Clusters in Different Samples



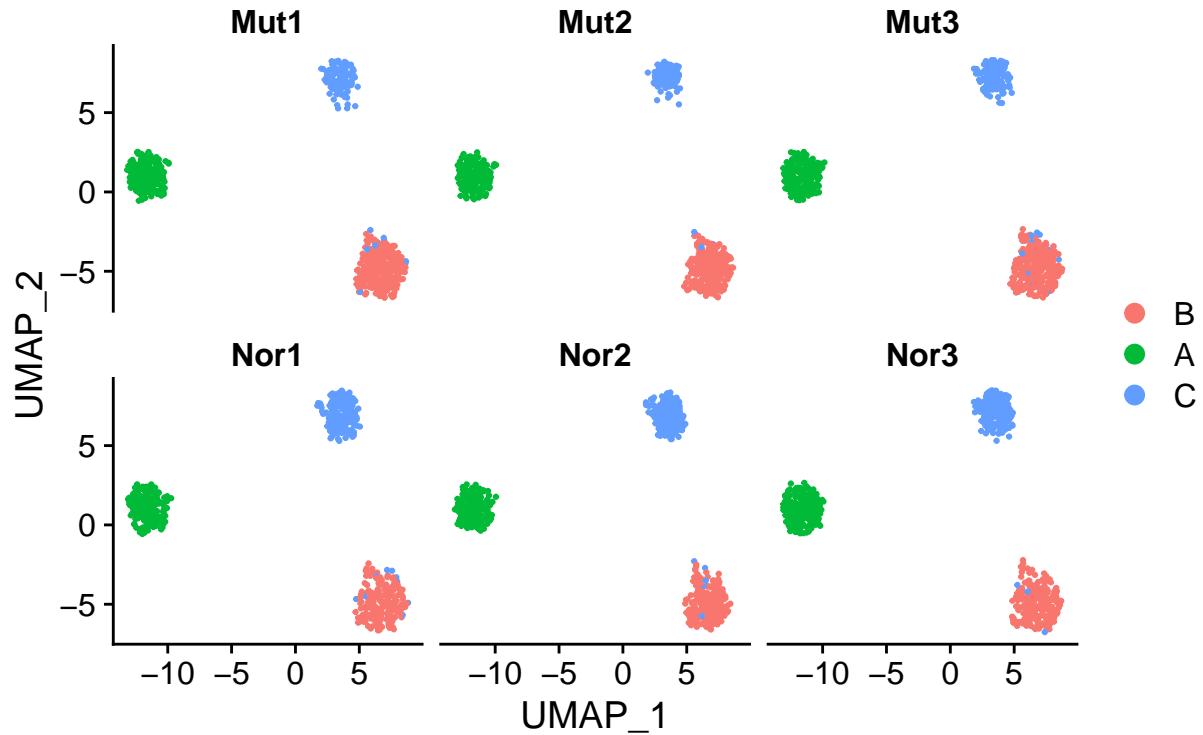
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A    0.9431431    0.8301337    0.4762864    0.4986057
## 2       B    0.9069592    0.8301337    0.5086470    0.5312071
## 3       C    0.5086470    0.9781097    0.9714788    1.0000000
##
## $time_df
##   methods      time
## 1  fisher 0.007977962
## 2 sepckle 0.005032063
## 3 dcats  0.244347095
```

Cluster Results of Different Clusters in Different Samples



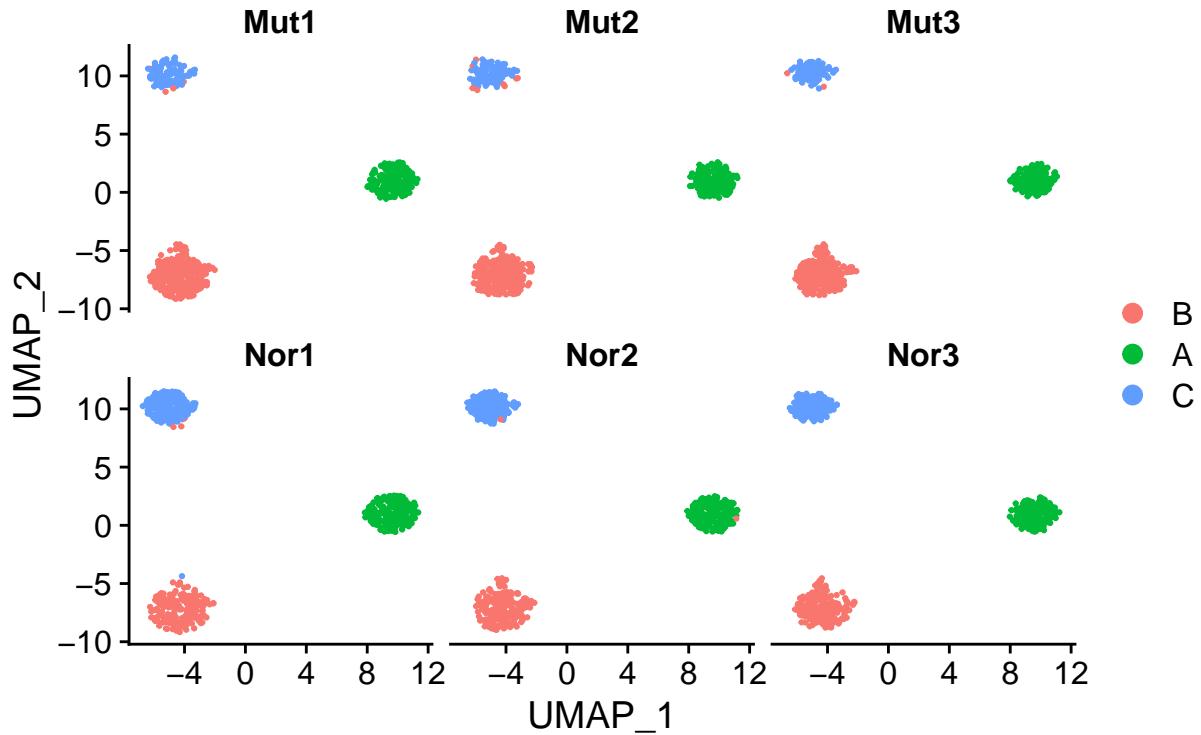
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A    0.94563519     0.4400677  0.476286380 0.4986056694
## 2       B    0.31698036     0.1261368  0.007686760 0.0084599195
## 3       C    0.00768676     0.1261368  0.000156066 0.0001759251
##
## $time_df
##   methods      time
## 1   fisher 0.008964062
## 2 sepckle 0.004000902
## 3 dcats 0.275225163
```

Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 8.202455e-01  5.041440e-01 4.762864e-01 4.986057e-01
## 2       B 5.978434e-06  8.724020e-06 4.142148e-15 4.010225e-15
## 3       C 4.142148e-15  2.009836e-06 1.624364e-21 5.921696e-22
##
## $time_df
##   methods      time
## 1   fisher 0.009021044
## 2 sepckle 0.004976988
## 3 dcats 0.252388000
```

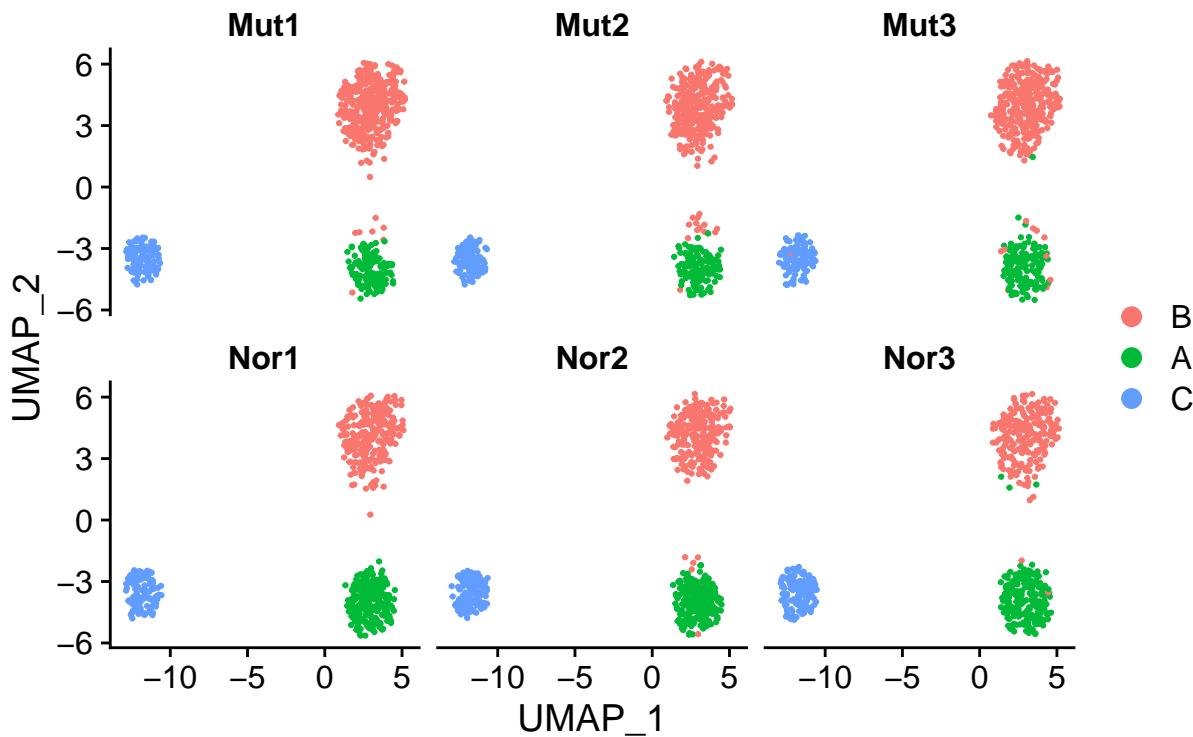
Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 6.758255e-01  5.158060e-01 4.762864e-01 4.986057e-01
## 2       B 3.188108e-12  3.284386e-06 1.808849e-20 1.383208e-20
## 3       C 1.808849e-20  6.352862e-07 6.356245e-29 6.146485e-30
##
## $time_df
##   methods      time
## 1   fisher 0.008975983
## 2 sepckle 0.004972935
## 3 dcats  0.238315105
```

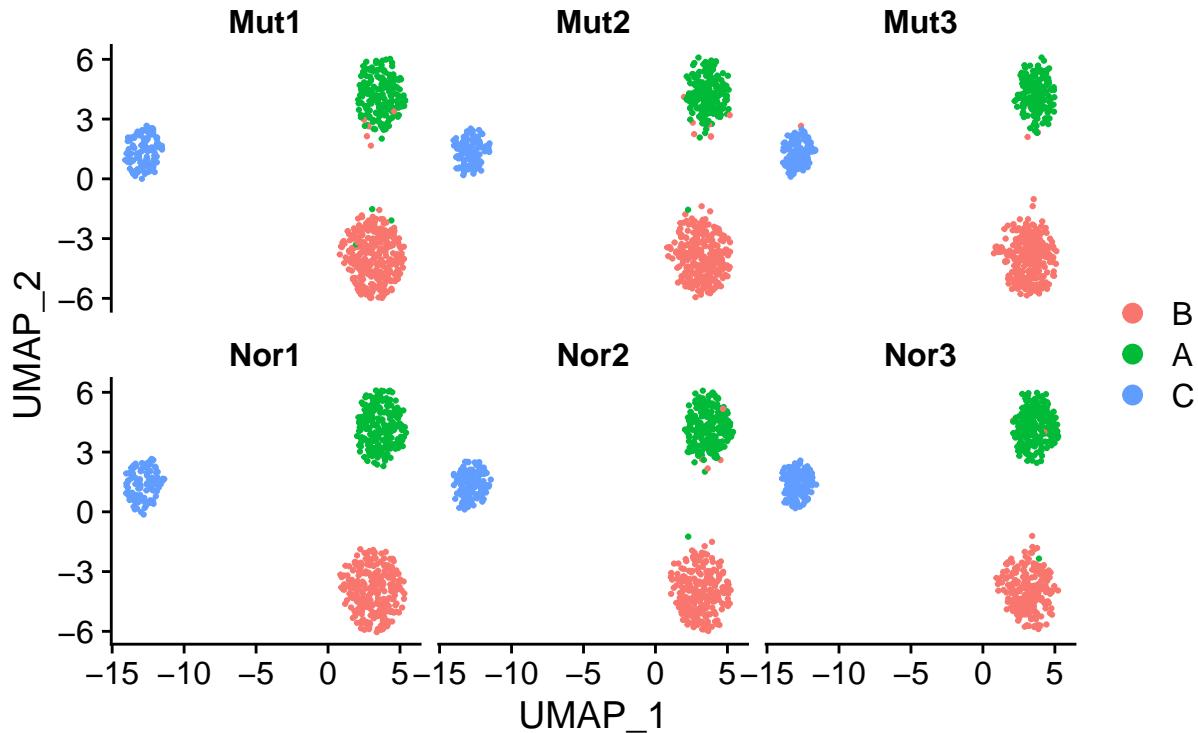
Different composition

Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 3.329999e-06 4.758358e-06 8.141138e-23 3.541373e-23
## 2       B 7.174435e-06 4.758358e-06 8.984597e-24 6.055524e-24
## 3       C 2.272624e-01 3.463323e-01 2.272624e-01 2.435776e-01
##
## $time_df
##   methods      time
## 1   fisher 0.009023190
## 2 sepckle 0.003991842
## 3   dcats 0.241353989
```

Cluster Results of Different Clusters in Different Samples

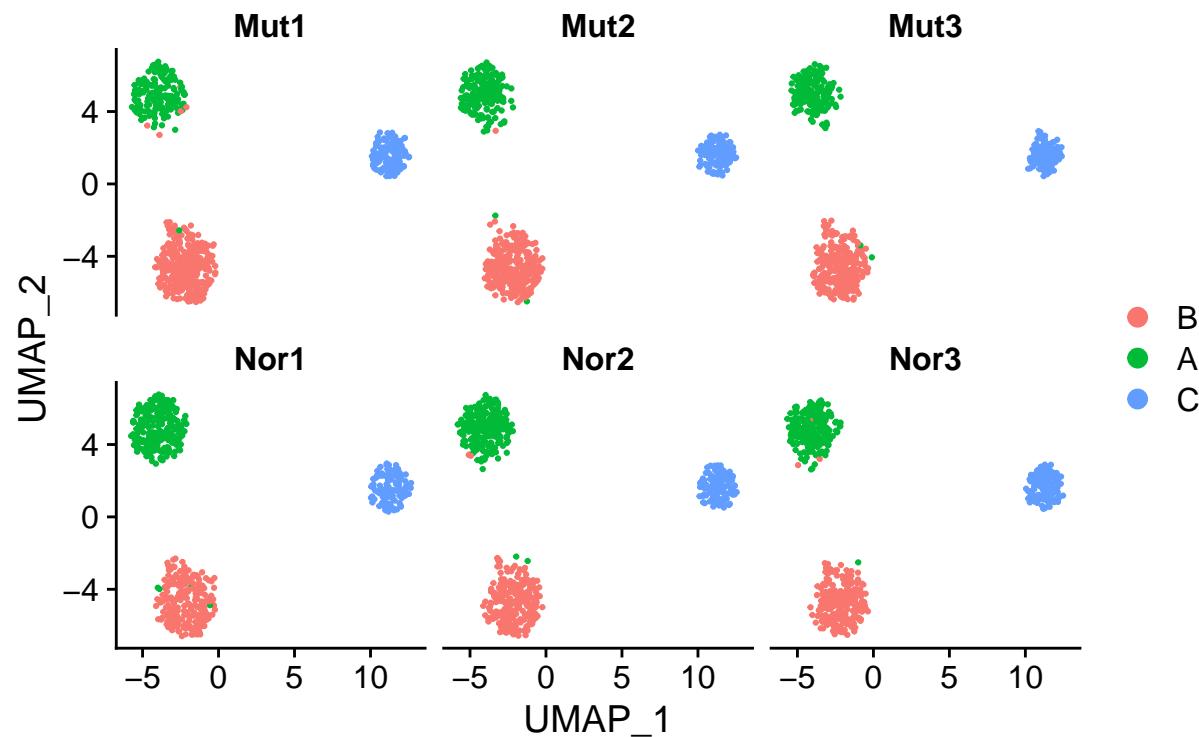


```

## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A  0.01543553  0.0005798027 4.379707e-07 4.920662e-07
## 2       B  0.01113654  0.0005798027 1.444596e-06 1.641263e-06
## 3       C  1.00000000  1.0000000000 1.000000e+00 1.000000e+00
##
## $time_df
##   methods      time
## 1  fisher 0.008979082
## 2 sepckle 0.004981995
## 3 dcats  0.250333071

```

Cluster Results of Different Clusters in Different Samples

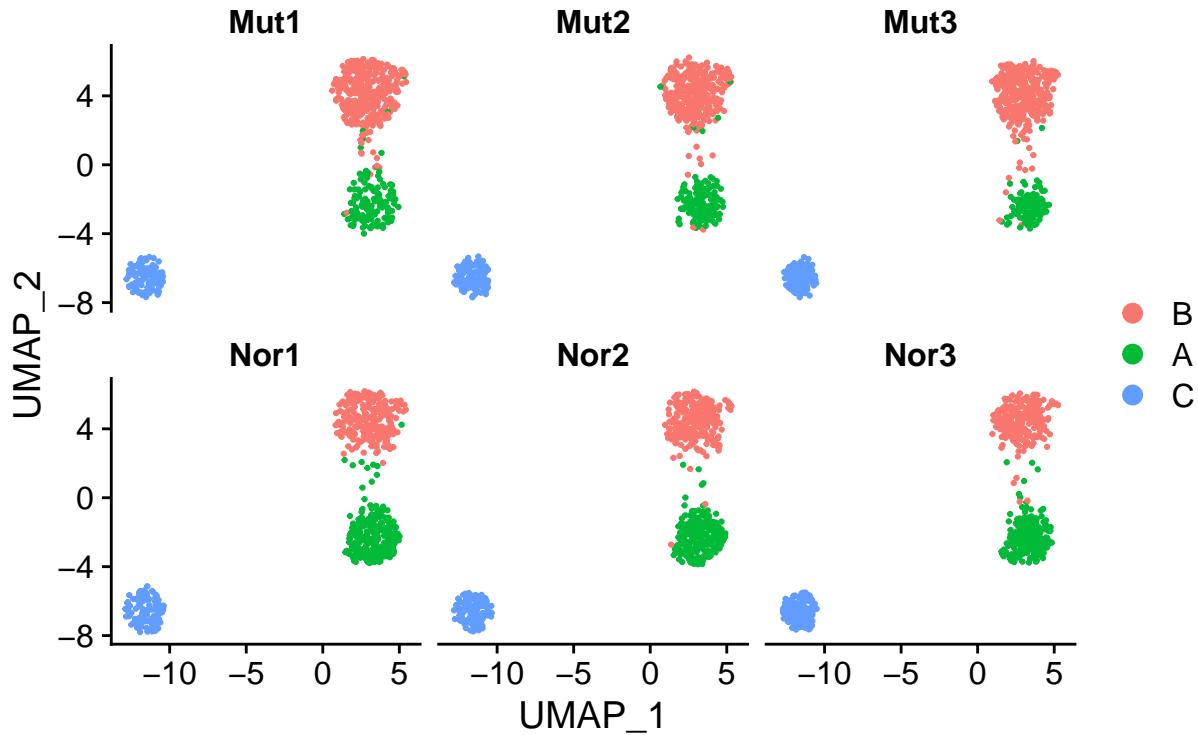


```

## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A  0.002236473  0.000165966 1.761302e-10 1.874273e-10
## 2       B  0.003602618  0.000165966 8.823016e-10 9.801246e-10
## 3       C  1.000000000  1.000000000 1.000000e+00 1.000000e+00
##
## $time_df
##   methods      time
## 1  fisher 0.007977962
## 2 sepckle 0.004937172
## 3 dcats  0.248368025

```

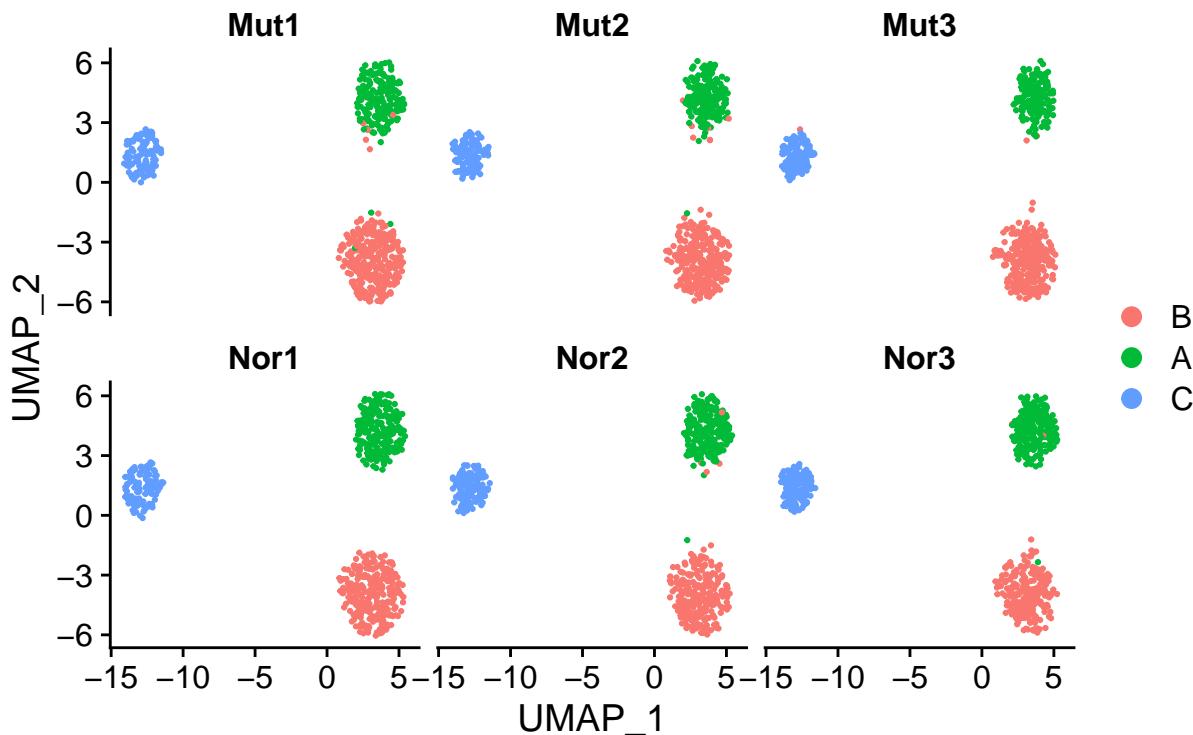
Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A 3.410972e-06  1.041730e-06 1.534467e-29 2.859680e-30
## 2       B 1.502514e-05  1.283716e-06 1.029409e-25 6.121191e-26
## 3       C 1.000000e+00  1.000000e+00 1.000000e+00 1.000000e+00
##
## $time_df
##   methods      time
## 1  fisher 0.007944822
## 2 sepckle 0.005020142
## 3 dcats  0.254271984
```

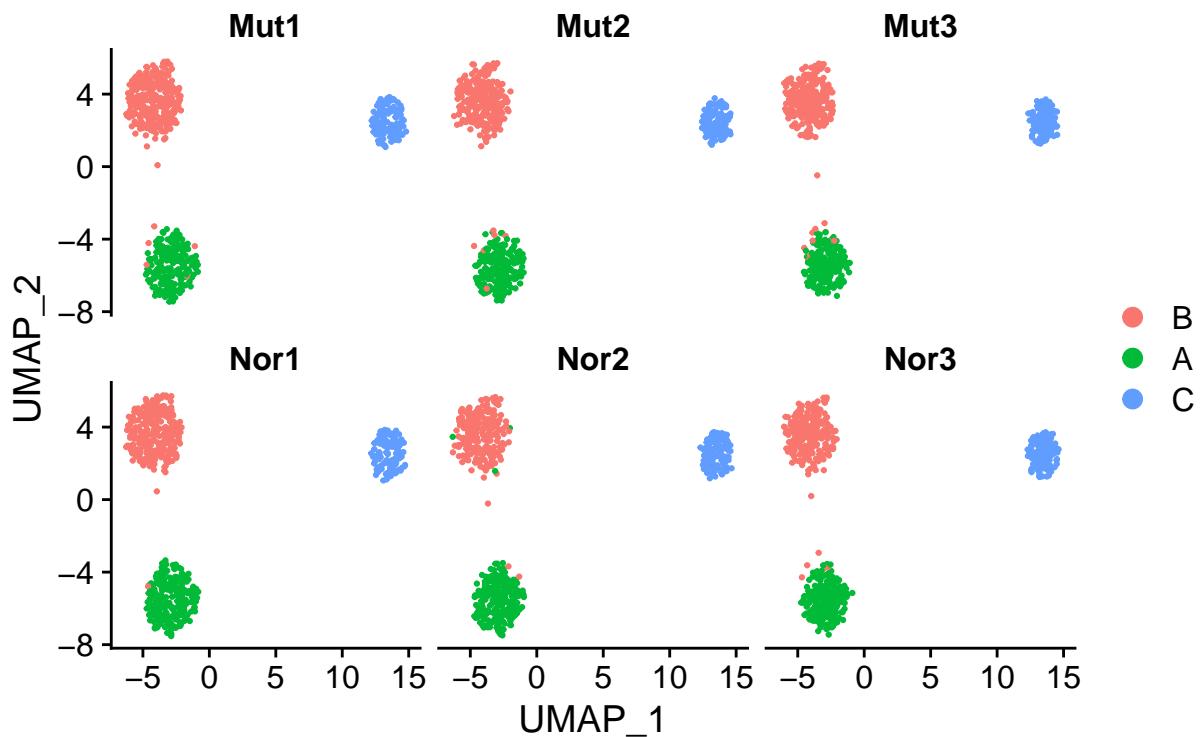
negative

Cluster Results of Different Clusters in Different Samples



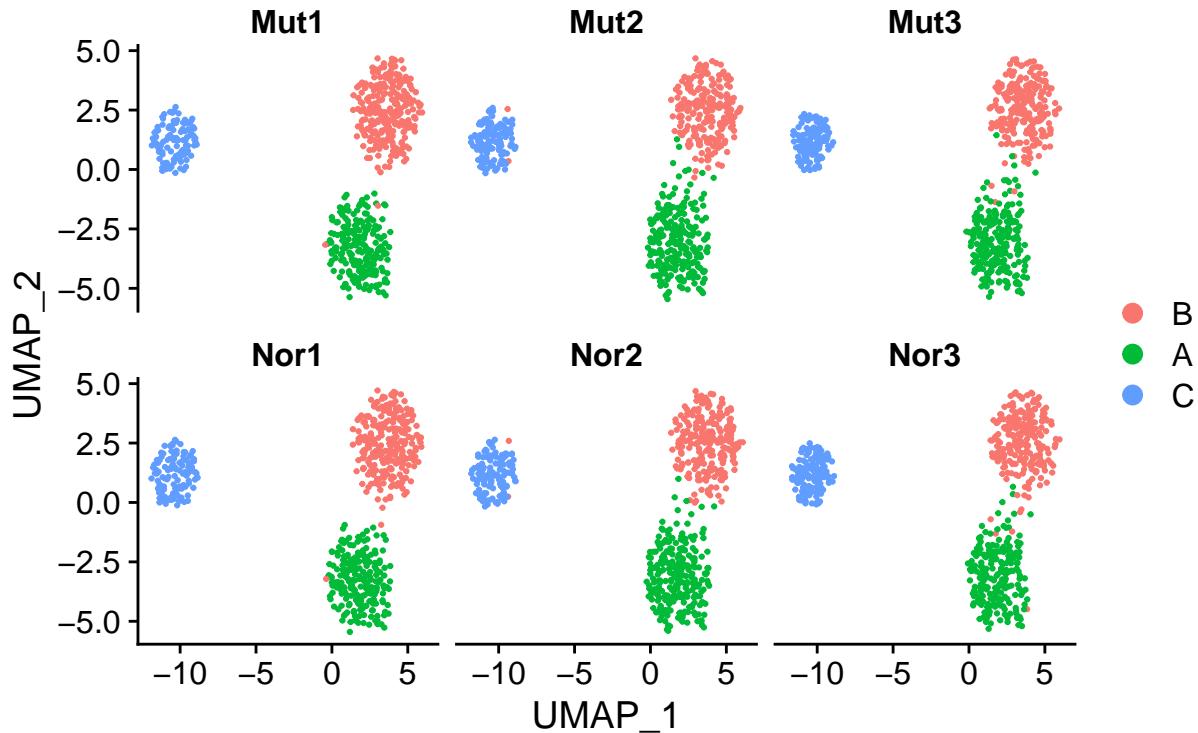
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A  0.01543553  0.0005798027 4.379707e-07 4.920662e-07
## 2       B  0.01113654  0.0005798027 1.444596e-06 1.641263e-06
## 3       C  1.00000000  1.0000000000 1.000000e+00 1.000000e+00
##
## $time_df
##   methods      time
## 1   fisher 0.007977962
## 2 sepckle 0.003996134
## 3   dcats 0.248369932
```

Cluster Results of Different Clusters in Different Samples



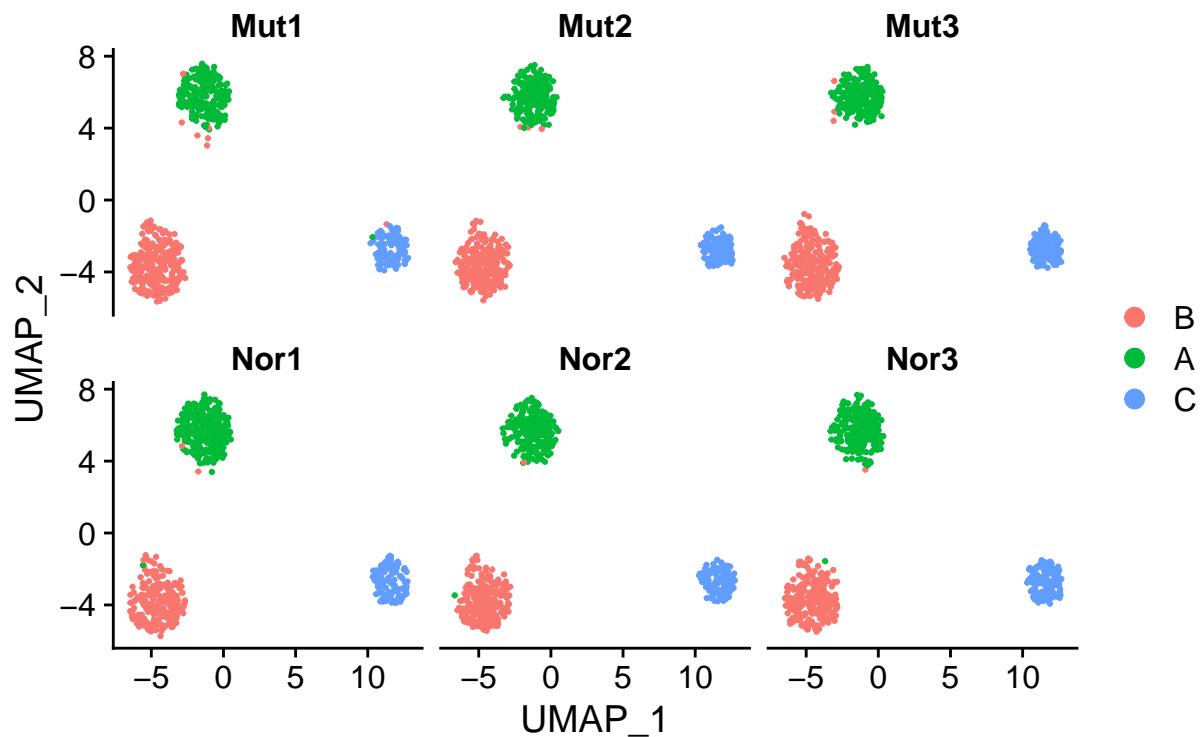
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A     0.2403083    0.07441124    0.01689223    0.01849425
## 2       B     0.2265072    0.07441124    0.02012900    0.02195635
## 3       C     1.0000000    1.00000000    1.00000000    1.00000000
##
## $time_df
##   methods      time
## 1   fisher 0.007978916
## 2 sepckle 0.004955053
## 3 dcats 0.255319118
```

Cluster Results of Different Clusters in Different Samples



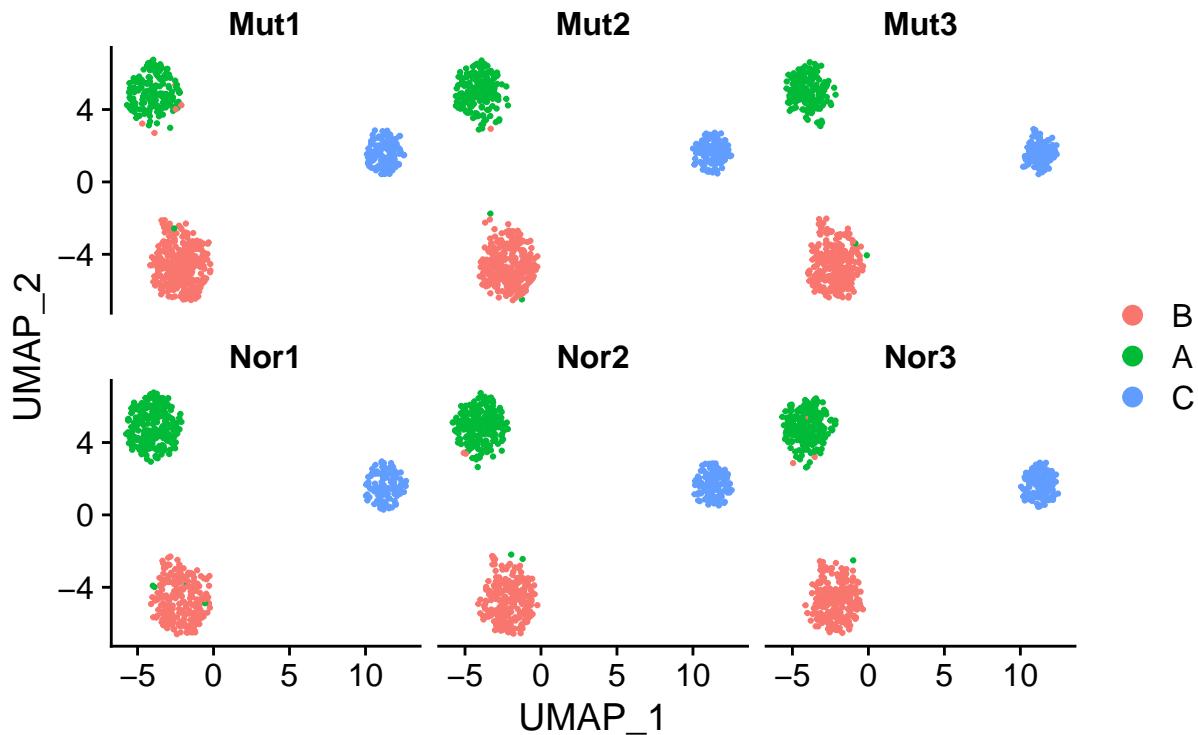
```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A    0.9683336           1    0.8914458    0.9184761
## 2       B    0.9684534           1    0.8916806    0.9186528
## 3       C    1.0000000           1    1.0000000    1.0000000
##
## $time_df
##   methods      time
## 1  fisher 0.006983042
## 2 sepckle 0.004970074
## 3 dcats  0.259294033
```

Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A  0.06127541    0.006763024 0.0001261974 0.0001422849
## 2       B  0.07204367    0.006763024 0.0001835532 0.0002066208
## 3       C  1.000000000  1.000000000 1.00000000000 1.00000000000
##
## $time_df
##   methods      time
## 1  fisher 0.009003878
## 2 sepckle 0.005949020
## 3 dcats  0.239344120
```

Cluster Results of Different Clusters in Different Samples



```
## $Res_df
##   cluster dcats_pvals1 speckle_pvals dcats_pvals2 fisher_pvals
## 1       A  0.002236473  0.000165966 1.761302e-10 1.874273e-10
## 2       B  0.003602618  0.000165966 8.823016e-10 9.801246e-10
## 3       C  1.000000000  1.000000000 1.000000e+00 1.000000e+00
##
## $time_df
##   methods      time
## 1  fisher 0.014991045
## 2 sepckle 0.003992796
## 3 dcats  0.230386019
```

Codes

```
knitr::opts_chunk$set(echo = FALSE)
library(splatter)
library(Seurat)
library(speckle)
library(DCATS)
library(ggplot2)
library(tidyverse)
source("functions.r")
set.seed(123)
probNor = c(1/3,1/3,1/3)
```

```

probMut = c(1/2,1/6,1/3)
de_prob = c(0.1,0.1,0.5)

batch_size_list = seq(500,2000,500)
setresolu_list = c(0.5, 0.3, 0.3, 0.3)
for (i in 1:4) {
  batch_size = batch_size_list[i]
  setresolu = setresolu_list[i]

  sim_list = simualtion(probNor, probMut, de_prob, batch_size)
  integratedSamples = runSeurat(sim_list, batch_size, setresolu)
  time = rep(NA,3)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTime(integratedSamples, sim_list)
  print(Res)
}

set.seed(123)
probNor = c(1/3,1/3,1/3)
probMut = c(1/2,1/6,1/3)
batch_size = 600

de_prob_list = seq(0.02, 0.08, 0.02)
setresolu_list = c(0.4, 0.4, 0.4, 0.3)
for (i in 1:4) {
  de_prob = c(de_prob_list[i], de_prob_list[i], 0.5)
  setresolu = setresolu_list[i]

  sim_list = simualtion(probNor, probMut, de_prob, batch_size)
  integratedSamples = runSeurat(sim_list, batch_size, setresolu)
  time = rep(NA,3)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTime(integratedSamples, sim_list)
  print(Res)
}

set.seed(123)
probNor = c(0.4, 0.4, 0.2)
batch_size = 600
de_prob = c(0.06,0.06,0.5)

probMut_list = c(0.2, 0.3, 0.5, 0.6)
setresolu_list = c(0.3, 0.3, 0.3, 0.3)
for (i in 1:4) {
  probMut = c(probMut_list[i], 0.8 - probMut_list[i], 0.2)
  setresolu = setresolu_list[i]

  sim_list = simualtion(probNor, probMut, de_prob, batch_size)
  integratedSamples = runSeurat(sim_list, batch_size, setresolu)
  time = rep(NA,3)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTime(integratedSamples, sim_list)
}

```

```

print(Res)
}
set.seed(123)
probNor = c(0.4, 0.4, 0.2)
batch_size = 600
de_prob = c(0.06,0.06,0.5)

probMut_list = seq(0.3,0.5, 0.05)
setresolu_list = c(0.3, 0.3, 0.3, 0.3, 0.3)
for (i in 1:5) {
  probMut = c(probMut_list[i], 0.8 - probMut_list[i], 0.2)
  setresolu = setresolu_list[i]

  sim_list = simualtion(probNor, probMut, de_prob, batch_size)
  integratedSamples = runSeurat(sim_list, batch_size, setresolu)
  time = rep(NA,3)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTime(integratedSamples, sim_list)
  print(Res)
}

```