

Simulation_plus3

Xinyi Lin

10/20/2020

Try to use parameters from HCC and mnnCT7 to do simulation.

In this file, I try:

1. Different batch size;
2. Different de.prob(differential expression probability)
3. Different cluster numbers

0.1, 0.2 keep the same, 0.1->0.05, 0.2->0.1, 0.2->0.25, 0.2->0.3

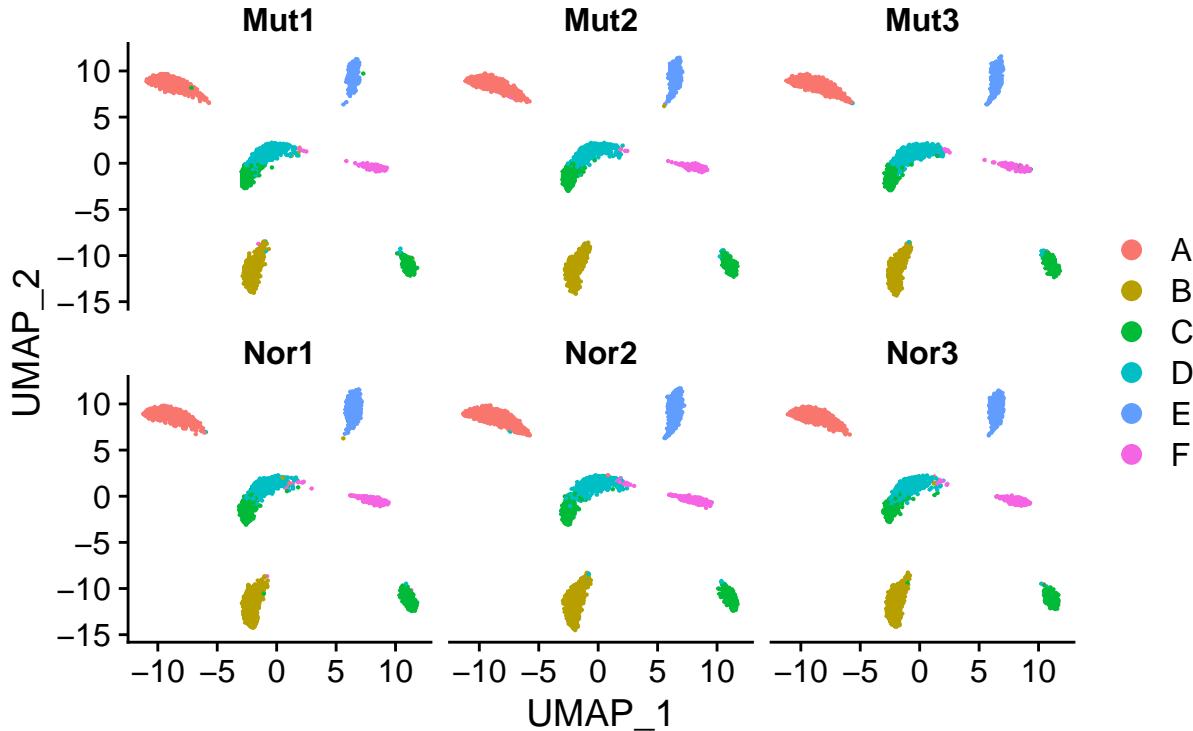
```
## sequential:  
## - args: function (expr, envir = parent.frame(), substitute = TRUE, lazy = FALSE, seed = NULL, global  
## - tweaked: FALSE  
## - call: NULL  
  
## multiprocess:  
## - args: function (expr, envir = parent.frame(), substitute = TRUE, lazy = FALSE, seed = NULL, global  
## - tweaked: TRUE  
## - call: plan("multiprocess", workers = 4)
```

Different batch size

Batch sizes are 5000, 10000, 15000, 20000.

```
##  
##      1   2   3   4   5   6  
##  A  0   0   0   0 3321   0  
##  B  7 2926   0   1   1   1  
##  C 1390   1   1   0   1 1336  
##  D 154   32  16   3  13 2393  
##  E  5   0   1 2243   0   1  
##  F  7   0 1143   1   2   0
```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.04185605
## 2 sepckle 0.01895189
## 3 dcats 0.58837914
## 4 diffcyt 0.03885007
##          A         B         C         D         E         F
## A 9.991218e-01 4.347294e-05 5.354929e-06 0.0006973049 0.000000e+00 1.320696e-04
## B 4.830640e-05 9.977614e-01 2.038288e-04 0.0018741656 8.408026e-05 2.822411e-05
## C 6.398791e-06 2.191918e-04 9.809678e-01 0.0182841017 1.409250e-04 3.816025e-04
## D 9.064472e-04 2.192512e-03 1.989066e-02 0.9741552203 7.984080e-04 2.056755e-03
## E 0.000000e+00 1.052786e-04 1.640876e-04 0.0008545491 9.987859e-01 9.017861e-05
## F 3.411725e-04 6.561540e-05 8.249700e-04 0.0040872755 1.674337e-04 9.945135e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1       A 4.346908e-04 1.004401e-03 2.647787e-09 6.384775e-02 6.680553e-05
## 2       B 7.965884e-01 7.627998e-01 7.110710e-01 9.060183e-01 8.155083e-01
## 3       C 3.454783e-01 1.384694e-01 3.749499e-03 3.221764e-01 2.348147e-01
## 4       D 7.457805e-03 1.522647e-05 5.478371e-38 8.002501e-05 3.145799e-06
## 5       E 1.163087e-16 3.330970e-07 9.289989e-49 4.487522e-06 3.243142e-22
## 6       F 4.680715e-07 7.015307e-07 6.295391e-35 1.572986e-04 1.200505e-12
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 2.691362e-08    1.39e-04        P        P        P        N
## 2 7.411629e-01    7.87e-01        N        N        N        N
## 3 6.314680e-03    1.38e-01        N        N        P        N
## 4 3.326962e-35    3.64e-11        P        P        P        P
## 5 1.648716e-46    1.50e-26        P        P        P        P
## 6 1.518161e-35    3.18e-25        P        P        P        P

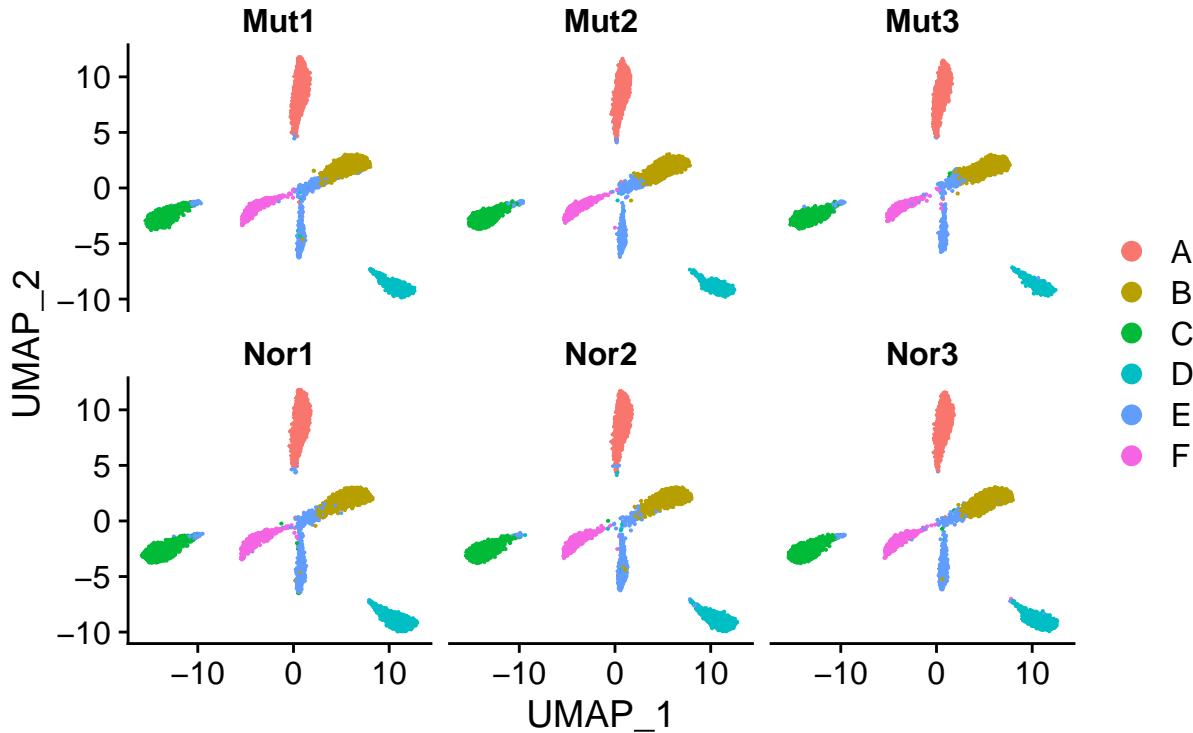
```

```

##   dcatsK_res fisher_res truth
## 1      P       P     P
## 2      N       N     N
## 3      N       P     N
## 4      P       P     P
## 5      P       P     P
## 6      P       P     P
##
##      1    2    3    4    5    6
## A    2    4    2    0 6629    1
## B    0    1   51    0    0 6339
## C    0 5642    4    0    0    1
## D    1    2   13 4423    0    1
## E   22   308 2202   42    58 1178
## F 3039    3    7    3    0    22

```

Cluster Results of Different Clusters in Different Samp



```

##   methods      time
## 1  fisher 0.05390501
## 2 sepckle 0.01096797
## 3  dcats 0.50463390
## 4 diffcyt 0.02792621
##
##          A         B         C         D         E         F
## A 9.977289e-01 1.690840e-05 2.949805e-05 5.809087e-05 0.002142226 2.433247e-05
## B 1.785804e-05 9.875029e-01 3.935005e-05 0.000000e+00 0.012376534 6.340072e-05
## C 3.493683e-05 4.412699e-05 9.962008e-01 0.000000e+00 0.003708520 1.159568e-05
## D 8.550504e-05 0.000000e+00 0.000000e+00 9.962387e-01 0.003455157 2.206697e-04

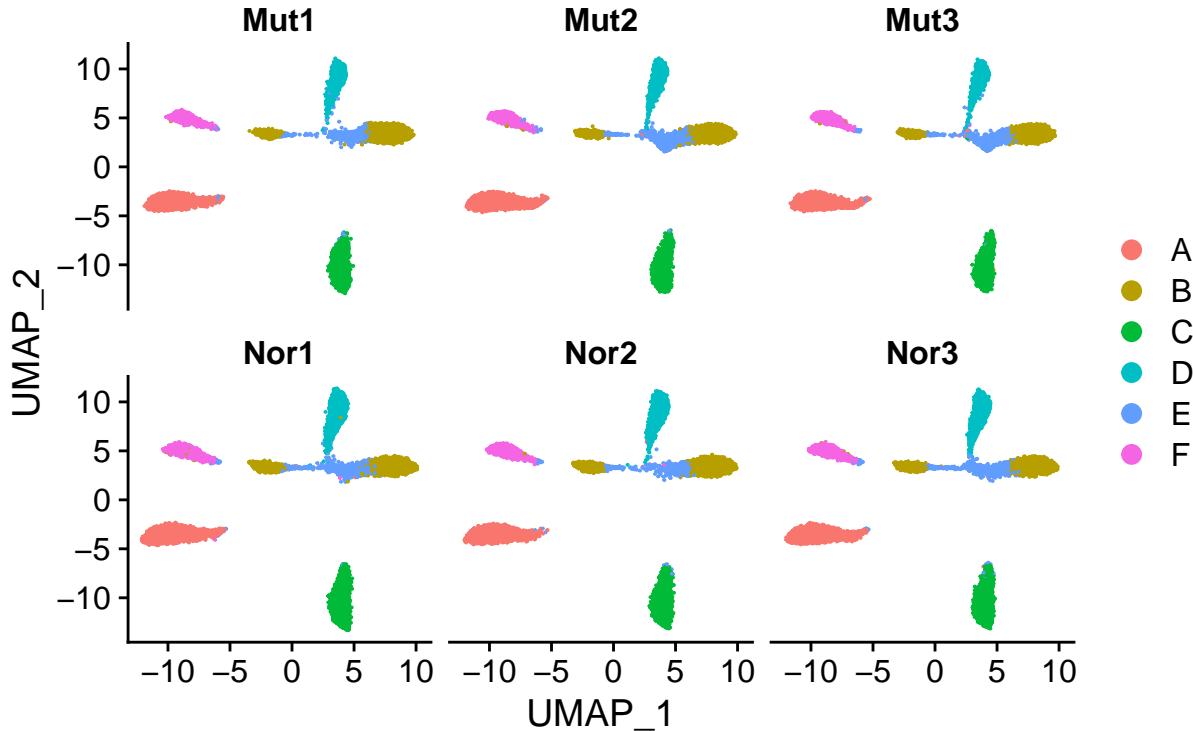
```

```

## E 3.566692e-03 1.951049e-02 5.213276e-03 3.908268e-03 0.964830985 2.970286e-03
## F 4.873164e-05 1.202231e-04 1.960788e-05 3.002505e-04 0.003572915 9.959383e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1      A 4.110717e-06 2.052018e-05 5.497407e-22 9.908739e-03 4.227433e-07
## 2      B 9.682902e-05 7.345687e-05 8.311940e-53 5.151962e-05 5.116960e-08
## 3      C 7.840063e-01 7.057095e-01 5.448079e-01 8.523953e-01 7.570875e-01
## 4      D 2.612613e-21 9.524866e-07 1.813683e-109 4.680693e-09 3.162609e-26
## 5      E 2.310453e-01 7.362790e-02 1.089132e-09 8.096435e-02 7.642005e-02
## 6      F 9.597719e-01 9.663989e-01 9.696270e-01 9.848964e-01 9.809223e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 1.771422e-19 1.05e-15 P P P P
## 2 3.386383e-47 3.20e-16 P P P P
## 3 5.759864e-01 7.73e-01 N N N N
## 4 2.329701e-105 1.03e-81 P P P P
## 5 4.104813e-09 2.33e-02 N N P N
## 6 9.852269e-01 9.73e-01 N N N N
##   dcatsK_res fisher_res truth
## 1      P P P
## 2      P P P
## 3      N N N
## 4      P P P
## 5      N P P
## 6      N N N
##
##          1   2   3   4   5   6
## A    2   2   8   1 10054  13
## B   34   1 2709   1   1 7288
## C    0 8953   4   0   0  14
## D    7   2   5 6649   0  20
## E   178  18 618  44  26 3945
## F  4333   0  46   2   3  19

```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.07380700
## 2 sepckle 0.01495481
## 3 dcats 0.51660299
## 4 diffcyt 0.02797198
##          A         B         C         D         E         F
## A 9.980055e-01 5.852391e-05 7.892189e-05 9.872232e-05 0.0016202449 1.380675e-04
## B 5.903616e-05 9.856962e-01 2.906601e-05 6.919380e-06 0.0128537602 1.355012e-03
## C 8.711422e-05 3.180477e-05 9.988276e-01 7.236566e-05 0.0009792971 1.809086e-06
## D 1.429380e-04 9.931496e-06 9.492339e-05 9.978807e-01 0.0018348552 3.668595e-05
## E 3.353257e-03 2.637130e-02 1.836154e-03 2.622742e-03 0.9606924799 5.124069e-03
## F 3.008368e-04 2.926832e-03 3.571142e-06 5.520862e-05 0.0053947122 9.913188e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1       A 4.732714e-07 1.045174e-04 2.234406e-31 2.972510e-03 2.502613e-09
## 2       B 3.188799e-01 1.694109e-01 2.896244e-08 1.535727e-01 7.742592e-02
## 3       B 3.188799e-01 1.694109e-01 2.896244e-08 1.535727e-01 7.742592e-02
## 4       C 5.903707e-01 3.803248e-01 2.829437e-01 8.080153e-01 5.347362e-01
## 5       D 8.274091e-30 3.482313e-06 7.491553e-174 5.823925e-12 1.014117e-51
## 6       F 7.817520e-01 6.951331e-01 6.228168e-01 9.187360e-01 8.230893e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 1.081503e-27    7.15e-12        P        P        P        P
## 2 2.221081e-07    1.42e-01        N        N        P        N
## 3 2.221081e-07    1.42e-01        N        N        P        N
## 4 3.166162e-01    5.94e-01        N        N        N        N
## 5 7.747212e-168   2.66e-56        P        P        P        P
## 6 6.429423e-01    7.56e-01        N        N        N        N

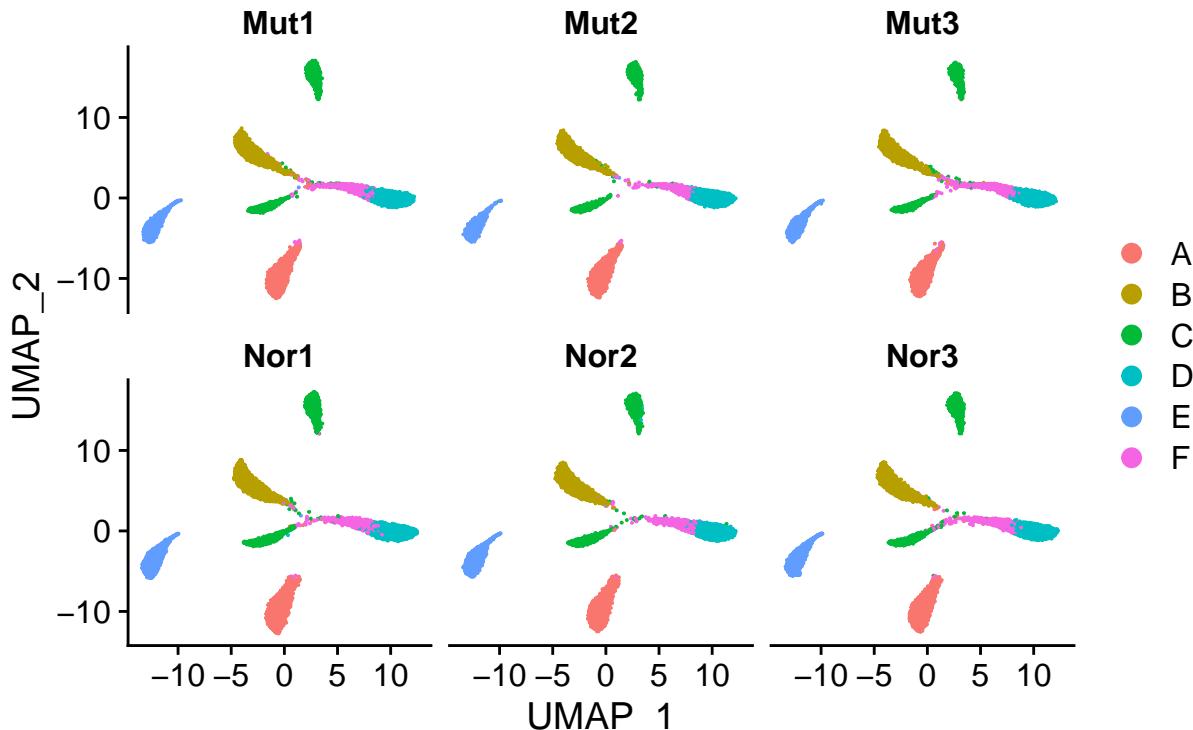
```

```

##   dcatsK_res fisher_res truth
## 1      P      P      P
## 2      N      P      P
## 3      N      P      P
## 4      N      N      N
## 5      P      P      P
## 6      N      N      N
##
##          1    2    3    4    5    6
## A     2    0    3    1 13293    5
## B    12 11864    6    2    7   16
## C   6014    5 4441    8    8   61
## D     2    0    3    0    0 10499
## E     0    0    1 8991    0    1
## F    26   17   77    2  131 4502

```

Cluster Results of Different Clusters in Different Samp



```

##   methods      time
## 1  fisher 0.08779597
## 2 sepckle 0.01795292
## 3  dcats 0.51063490
## 4 diffcyt 0.02789211
##
##           A         B         C         D         E         F
## A 9.987052e-01 1.081053e-04 0.0001955313 1.735983e-06 7.612385e-06 0.0009818124
## B 1.184317e-04 9.985651e-01 0.0003126131 3.009375e-06 4.670874e-05 0.0009541510
## C 2.3877763e-04 3.484663e-04 0.9963381178 2.025014e-04 1.607707e-04 0.0027113674
## D 2.223490e-06 3.518396e-06 0.0002123944 9.882467e-01 0.000000e+00 0.0115351979

```

```

## E 1.086454e-05 6.085095e-05 0.0001878982 0.000000e+00 9.996070e-01 0.0001333533
## F 2.655910e-03 2.356030e-03 0.0060061816 2.436241e-02 2.527542e-04 0.9643667167
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1      A 3.739565e-08 1.634285e-05 5.744595e-45 7.521988e-04 4.761322e-12
## 2      B 6.983992e-01 1.676407e-01 3.731802e-01 8.370491e-01 6.466992e-01
## 3      C 5.252712e-02 2.779588e-05 3.901479e-60 1.438340e-04 2.027729e-12
## 4      C 5.252712e-02 2.779588e-05 3.901479e-60 1.438340e-04 2.027729e-12
## 5      D 8.402902e-15 1.059657e-02 2.360937e-95 3.512535e-07 4.071013e-14
## 6      E 4.322300e-75 1.634285e-05 7.953288e-226 4.191137e-14 4.202046e-93
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 1.042967e-39 2.05e-14 P P P P
## 2 4.063116e-01 6.02e-01 N N N N
## 3 6.169017e-55 8.78e-16 P N P P
## 4 6.169017e-55 8.78e-16 P N P P
## 5 1.425410e-87 1.00e-04 P P P P
## 6 1.876120e-217 6.52e-39 P P P P
##   dcatsK_res fisher_res truth
## 1      P P P
## 2      N N N
## 3      P P N
## 4      P P P
## 5      P P P
## 6      P P P

## The following `from` values were not present in `x`: 6

```

Different de.prob

The de.prob of cluster B, C are 0.05, 0.06, 0.07, 0.08, 0.09.

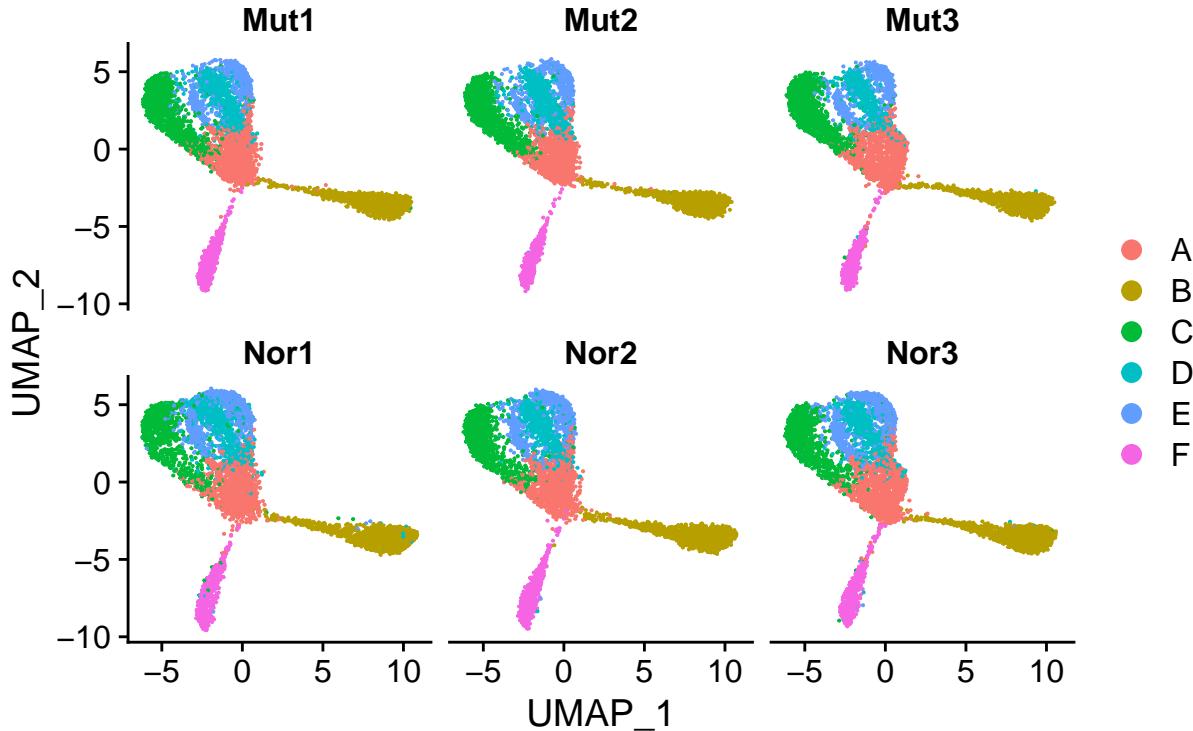
When de.prob equals to 0.04, only the fisher's exact test give the "correct" result.

```

##
##      1   2   3   4   5   6
##  A  24   59  803 1790 2443 1755
##  B   1 5898    0    8    6    1
##  C   3    2   83  190  102 5523
##  D   0    0  141  298 3805   91
##  E   0    1 1247 2179  323 167
##  F 3036    0    5    3    8    5

```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.04687405
## 2 sepckle 0.01097298
## 3 dcats 0.62449288
## 4 diffcyt 0.02792597
##          A         B         C         D         E         F
## A 0.966524836 2.104556e-03 0.0107420311 1.232316e-02 6.762506e-03 1.542911e-03
## B 0.002340986 9.975926e-01 0.0000229815 3.652967e-06 2.434477e-05 1.547400e-05
## C 0.012561866 2.416060e-05 0.9708858213 6.642102e-03 9.810901e-03 7.514940e-05
## D 0.019681366 5.244940e-06 0.0090713295 9.492400e-01 2.195974e-02 4.233484e-05
## E 0.011948460 3.866977e-05 0.0148233184 2.429396e-02 9.487843e-01 1.112533e-04
## F 0.003166482 2.854962e-05 0.0001318845 5.440024e-05 1.292245e-04 9.964895e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1        B 0.97022840 0.9344479655 9.306126e-01 9.889538e-01 9.685793e-01
## 2        C 0.01026009 0.0003058568 2.162315e-38 4.581661e-04 9.982388e-05
## 3        D 0.98334426 0.9344479655 6.339307e-01 8.886633e-01 8.191959e-01
## 4        E 0.17834519 0.0003058568 1.161482e-66 2.690104e-06 1.083348e-06
## 5        E 0.17834519 0.0003058568 1.161482e-66 2.690104e-06 1.083348e-06
## 6        F 0.95567780 0.9344479655 9.543423e-01 9.837464e-01 9.785621e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 9.454043e-01    9.62e-01      N       N       N       N
## 2 1.225888e-34   7.71e-15      P       P       P       P
## 3 6.588426e-01   9.62e-01      N       N       N       N
## 4 6.215174e-64   5.76e-22      P       N       P       P
## 5 6.215174e-64   5.76e-22      P       N       P       P
## 6 9.703812e-01   9.62e-01      N       N       N       N

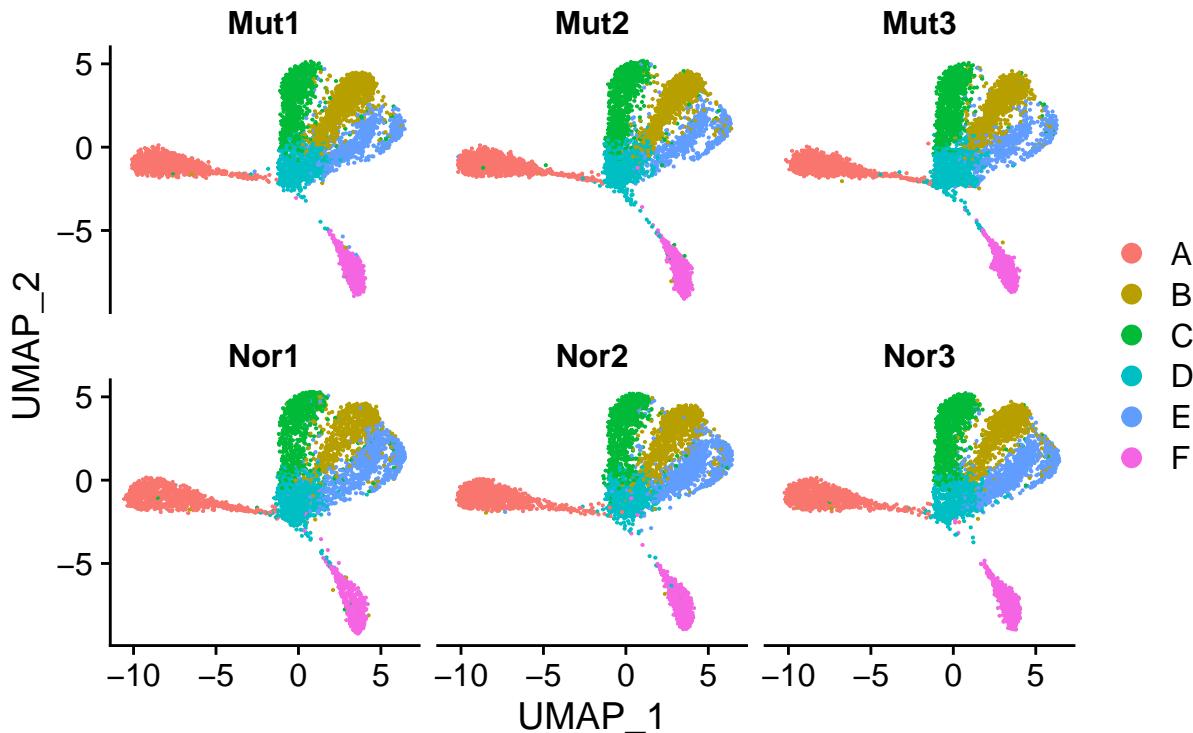
```

```

##   dcatsK_res fisher_res truth
## 1      N       N     N
## 2      P       P     P
## 3      N       N     P
## 4      P       P     P
## 5      P       P     P
## 6      N       N     N
##
##      1    2    3    4    5    6
## A 0 5904 2 9 6 11
## B 1 1 202 244 165 5095
## C 5 1 94 161 4913 115
## D 149 53 491 1049 1389 2060
## E 0 1 1489 3000 210 249
## F 2909 0 1 5 4 12

```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.05182195
## 2 sepckle 0.01001310
## 3 dcats 0.48772597
## 4 diffcyt 0.02393579
##
##          A         B         C         D         E         F
## A 9.970672e-01 0.0001278907 0.0001151160 0.002603402 7.060062e-05 1.577062e-05
## B 1.406849e-04 0.9617795973 0.0078193506 0.011337752 1.853716e-02 3.854508e-04
## C 1.338389e-04 0.0082643529 0.9674433746 0.010755061 1.325395e-02 1.494216e-04
## D 3.073682e-03 0.0121684735 0.0109215397 0.959020599 1.199869e-02 2.817019e-03

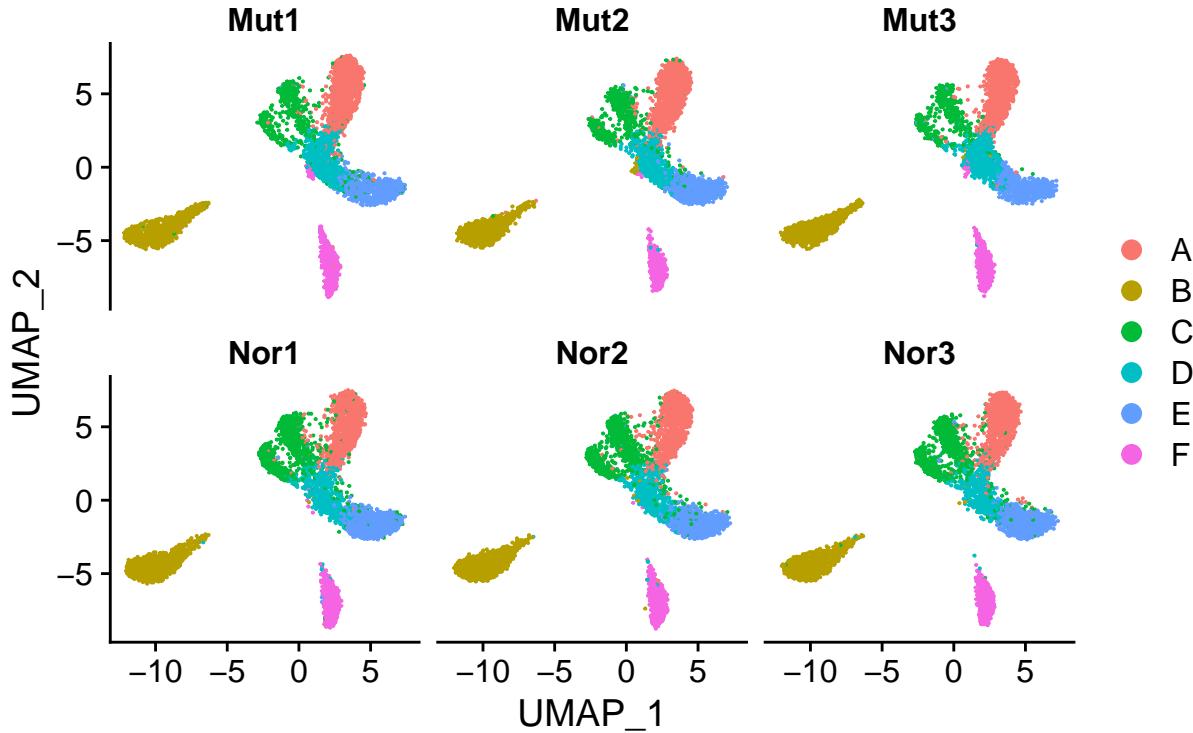
```

```

## E 8.802034e-05 0.0210091860 0.0142125861 0.012670404 9.517578e-01 2.619766e-04
## F 3.052372e-05 0.0006781865 0.0002487455 0.004618073 4.067024e-04 9.940178e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1      A  0.9663620  0.834541937 8.847446e-01 9.518397e-01 9.508904e-01
## 2      B  0.1178334  0.001297265 5.590656e-19 1.854061e-02 1.049034e-02
## 3      C  0.2695605  0.045375162 5.383782e-07 1.928102e-01 1.639702e-01
## 4      E  0.1843249  0.001297265 4.211659e-85 2.858224e-07 3.407413e-07
## 5      E  0.1843249  0.001297265 4.211659e-85 2.858224e-07 3.407413e-07
## 6      F  0.8990268  0.617115740 4.718562e-01 8.439701e-01 7.098347e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 9.020314e-01    9.16e-01      N      N      N      N
## 2 3.598157e-17    8.59e-07      P      N      P      P
## 3 2.001959e-06    2.74e-02      P      N      P      N
## 4 8.795744e-80    8.28e-11      P      N      P      P
## 5 8.795744e-80    8.28e-11      P      N      P      P
## 6 4.953584e-01    7.50e-01      N      N      N      N
##   dcatsK_res fisher_res truth
## 1      N      N      N
## 2      P      P      P
## 3      N      P      P
## 4      P      P      P
## 5      P      P      P
## 6      N      N      N
##
##          1   2   3   4   5   6
## A  1   6  77 255  28 6154
## B  0 5918   8   3   0   5
## C  2   5 1692 3120 132 117
## D  57  29 443  921 2070 1242
## E  1   1  56  161 4455  23
## F 3003   1   3   8   2   1

```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.07001400
## 2 sepckle 0.01790214
## 3 dcats 0.51764798
## 4 diffcyt 0.02493715
##          A         B         C         D         E         F
## A 9.793847e-01 2.107541e-04 9.817495e-03 0.009332560 1.223043e-03 3.142660e-05
## B 2.231724e-04 9.978759e-01 1.130691e-04 0.001740735 4.518732e-06 4.258577e-05
## C 1.261551e-02 1.372093e-04 9.644797e-01 0.011109139 1.162821e-02 3.024952e-05
## D 1.259277e-02 2.218138e-03 1.166532e-02 0.958600263 1.253783e-02 2.385675e-03
## E 1.682783e-03 5.871359e-06 1.245074e-02 0.012784636 9.730718e-01 4.129307e-06
## F 6.245719e-05 7.992543e-05 4.678423e-05 0.003513790 5.964523e-06 9.962911e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1       A 0.01552135 0.048021795 3.463396e-15 2.772036e-02 1.473143e-02
## 2       B 0.97806857 0.666903541 8.392020e-01 9.767998e-01 9.349693e-01
## 3       C 0.05934641 0.001525332 2.832867e-179 1.074217e-14 4.853664e-16
## 4       C 0.05934641 0.001525332 2.832867e-179 1.074217e-14 4.853664e-16
## 5       E 0.46102908 0.251896186 3.881744e-09 9.187144e-02 6.677076e-02
## 6       F 0.88519887 0.666903541 7.587660e-01 9.333350e-01 9.041742e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 1.635641e-13 1.75e-02 P P P P
## 2 8.588972e-01 9.09e-01 N N N N
## 3 9.212672e-174 8.09e-12 P N P P
## 4 9.212672e-174 8.09e-12 P N P P
## 5 1.845198e-08 2.36e-01 N N P N
## 6 7.793437e-01 9.09e-01 N N N N

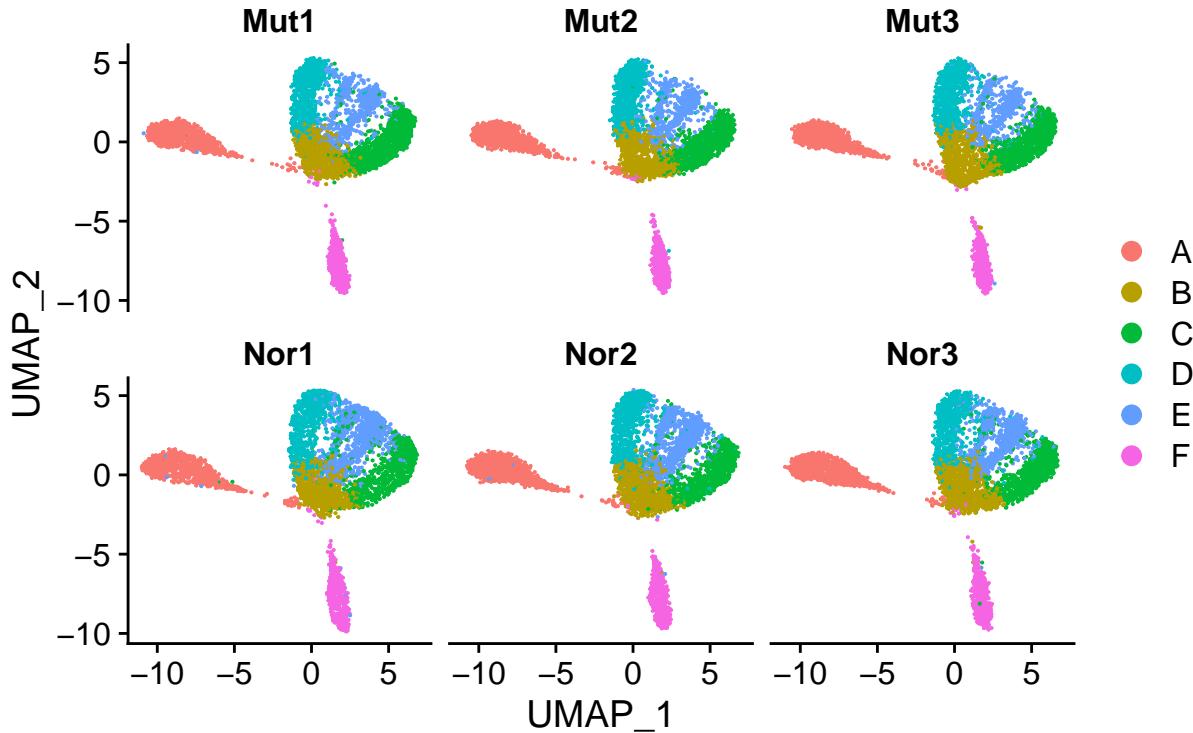
```

```

##   dcatsK_res fisher_res truth
## 1      P      P      P
## 2      N      N      N
## 3      P      P      P
## 4      P      P      P
## 5      N      P      P
## 6      N      N      N
##
##      1    2    3    4    5    6
## A 0 5910 0 0 6 4
## B 22 42 606 1174 1777 1952
## C 0 1 161 167 37 5146
## D 0 0 181 124 4618 81
## E 1 7 1323 3000 242 351
## F 3041 0 8 3 7 8

```

Cluster Results of Different Clusters in Different Samp



```

##   methods      time
## 1 fisher 0.05089402
## 2 sepckle 0.01097679
## 3 dcats 0.52060604
## 4 diffcyt 0.02492714
##
##           A          B          C          D          E          F
## A 9.979666e-01 0.001876832 9.483069e-06 3.120805e-05 8.845008e-05 2.741241e-05
## B 2.073173e-03 0.958874568 1.336571e-02 1.403240e-02 9.476719e-03 2.177427e-03
## C 1.063040e-05 0.013563837 9.656425e-01 3.163771e-03 1.757883e-02 4.044976e-05
## D 3.816195e-05 0.015534089 3.451187e-03 9.652969e-01 1.565073e-02 2.897305e-05

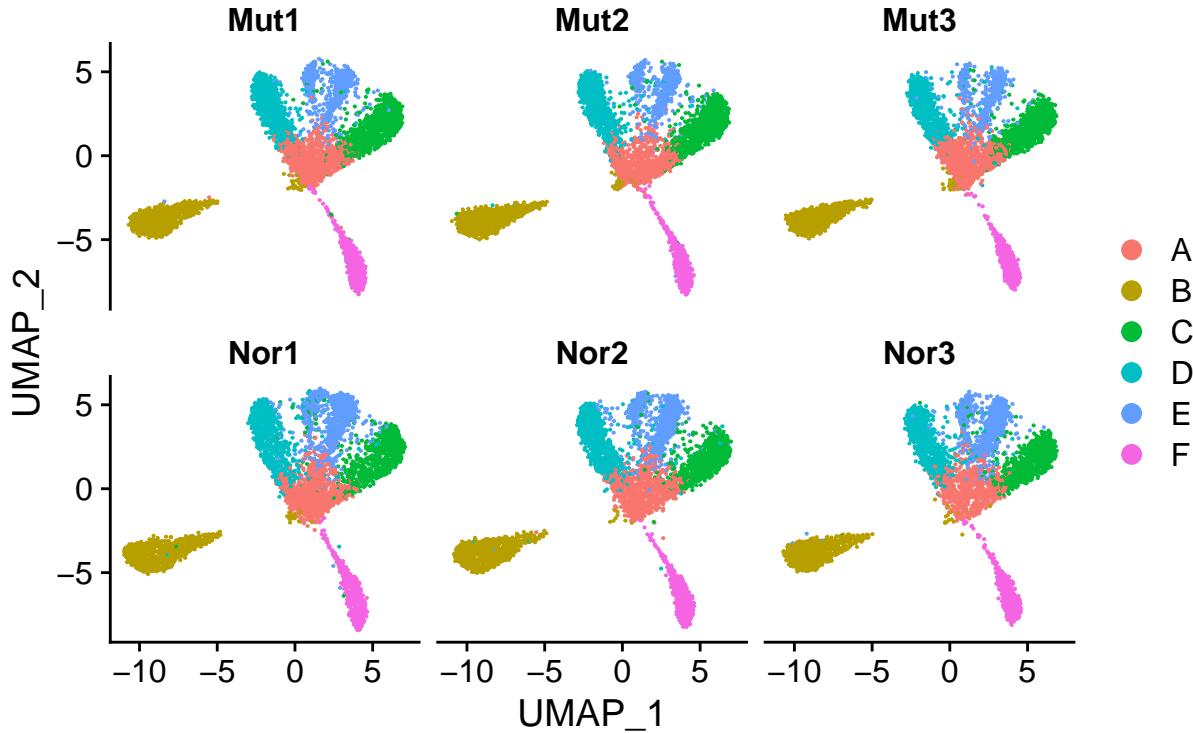
```

```

## E 1.107007e-04 0.010737423 1.962645e-02 1.601854e-02 9.533975e-01 1.093708e-04
## F 5.014930e-05 0.003606212 6.601354e-05 4.334590e-05 1.598700e-04 9.960744e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1      A 0.9723052 9.032333e-01 9.537270e-01 9.959828e-01 9.746732e-01
## 2      C 0.1012964 5.125631e-06 3.771135e-32 1.989809e-03 1.035046e-03
## 3      D 0.3422716 5.559765e-04 5.386375e-11 8.820782e-02 6.722945e-02
## 4      E 0.1080625 1.269437e-04 1.760859e-67 2.814684e-06 1.009855e-06
## 5      E 0.1080625 1.269437e-04 1.760859e-67 2.814684e-06 1.009855e-06
## 6      F 0.8960234 7.401990e-01 7.749865e-01 9.399270e-01 8.789363e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 9.672410e-01 9.60e-01 N N N N
## 2 3.040281e-29 2.94e-24 P N P P
## 3 4.125965e-10 1.60e-08 P N P N
## 4 6.561447e-63 2.51e-25 P N P P
## 5 6.561447e-63 2.51e-25 P N P P
## 6 7.952414e-01 9.50e-01 N N N N
##   dcatsK_res fisher_res truth
## 1      N N N
## 2      P P P
## 3      N P P
## 4      P P P
## 5      P P P
## 6      N N N
##
##      1   2   3   4   5   6
## A 27 70 548 1192 2007 2327
## B 1 5882 2 1 3 3
## C 2 2 167 132 34 4988
## D 2 5 150 117 4547 81
## E 0 1 1407 3022 94 139
## F 3032 0 5 4 2 4

```

Cluster Results of Different Clusters in Different Samples



```

##   methods      time
## 1 fisher 0.05285811
## 2 sepckle 0.01396298
## 3 dcats 0.62623787
## 4 diffcyt 0.02789211
##          A         B         C         D         E         F
## A 0.962432532 2.421248e-03 1.236631e-02 0.0115886958 9.556818e-03 1.634399e-03
## B 0.002485244 9.972562e-01 5.734189e-05 0.0001178015 5.624218e-05 2.716535e-05
## C 0.014581382 6.587201e-05 9.703398e-01 0.0033357665 1.165846e-02 1.870544e-05
## D 0.014697028 1.455512e-04 3.587831e-03 0.9706823202 1.073346e-02 1.538072e-04
## E 0.012732302 7.300052e-05 1.317273e-02 0.0112755692 9.626881e-01 5.830742e-05
## F 0.003052706 4.943249e-05 2.963031e-05 0.0002265211 8.174423e-05 9.965600e-01
##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1       B 0.88962863 0.42748103 6.008508e-01 9.010412e-01 8.124934e-01
## 2       C 0.05891228 0.02177380 2.479476e-13 4.802097e-02 2.029009e-02
## 3       D 0.55830112 0.72135359 1.694113e-01 7.007192e-01 6.681386e-01
## 4       E 0.12216055 0.00196979 1.096566e-116 4.396108e-11 3.802278e-13
## 5       E 0.12216055 0.00196979 1.096566e-116 4.396108e-11 3.802278e-13
## 6       F 0.99741583 0.78425884 8.634268e-01 9.746008e-01 9.356872e-01
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 6.311201e-01 8.51e-01 N N N N
## 2 3.591310e-12 2.08e-03 P N P P
## 3 1.947292e-01 8.51e-01 N N N N
## 4 8.494085e-112 2.03e-12 P N P P
## 5 8.494085e-112 2.03e-12 P N P P
## 6 8.817499e-01 9.00e-01 N N N N

```

```

##   dcatsK_res fisher_res truth
## 1      N       N     N
## 2      P       P     P
## 3      N       N     P
## 4      P       P     P
## 5      P       P     P
## 6      N       N     N

```

Different clusters number

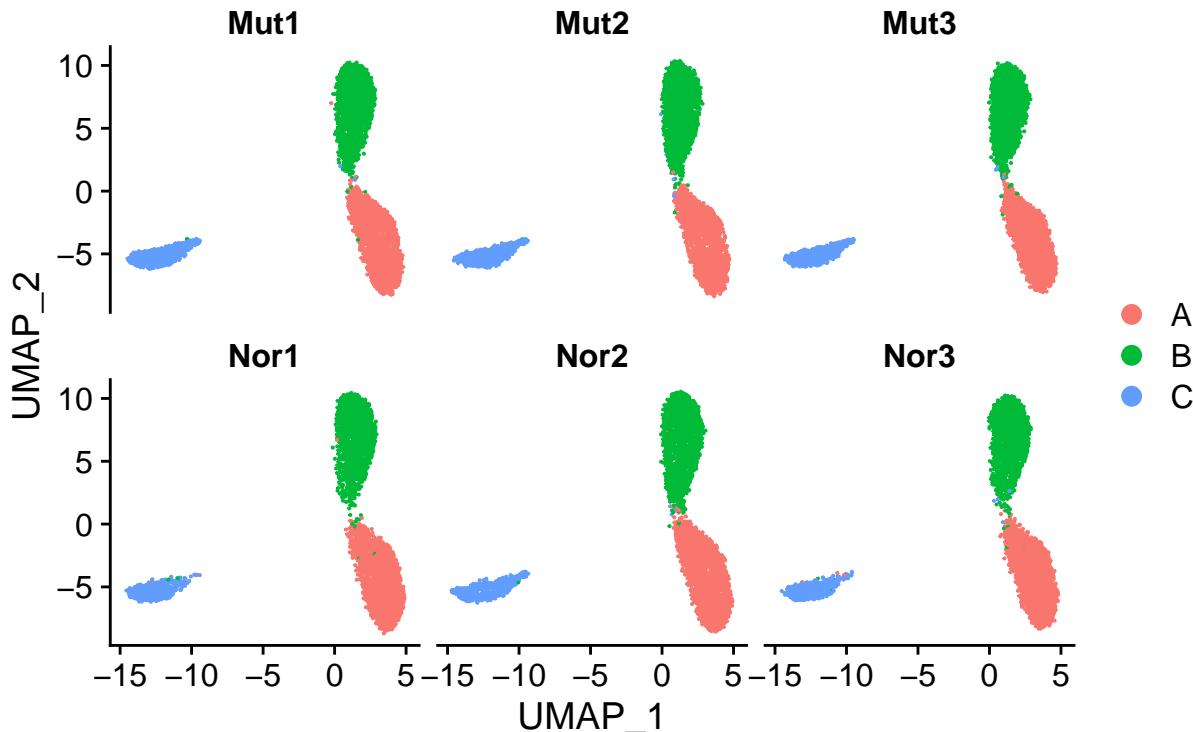
In the following example, there are no small clusters. Cluster numbers are 3, 6, 12.

```

##
##      1    2    3
## A 48 12 13345
## B 11959 16 35
## C 18 4561 6

```

Cluster Results of Different Clusters in Different Samples



```

## `summarise()` regrouping output by 'condition' (override with `$.groups` argument)

## group variable has 2 levels, t-tests will be performed

## FlowSOM clustering completed in 0.8 seconds

```

```

##   methods      time
## 1 fisher 0.04188800
## 2 sepckle 0.01097107
## 3 dcats 0.25435305
## 4 diffcyt 0.02595901

##          A           B           C
## A 0.997672077 0.001924262 0.0004036615
## B 0.002137693 0.997167847 0.0006944594
## C 0.001110053 0.001719064 0.9971708837

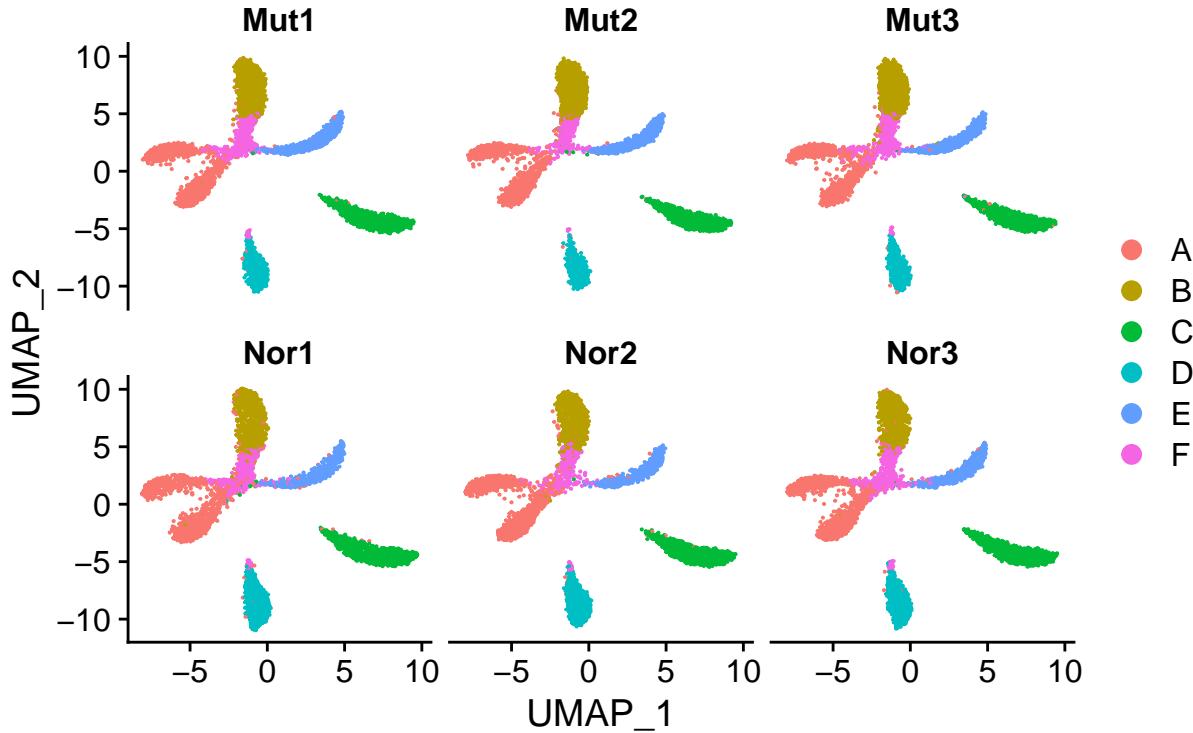
## The following `from` values were not present in `x`: 4, 5, 6

##   cluster dcats_pvalsT speckle_pvals  dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1      A 3.956054e-11 1.644249e-10 1.639826e-68 2.784758e-05 5.295492e-15
## 2      B 9.909737e-01 9.361834e-01 9.436358e-01 9.826189e-01 9.796955e-01
## 3      C 2.030104e-25 6.739673e-12 2.407199e-124 2.067835e-07 8.280758e-35
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 1.640628e-49      2.12e-32          P          P          P          P
## 2 9.593137e-01      9.60e-01          N          N          N          N
## 3 7.132562e-120     2.75e-103         P          P          P          P
##   dcatsK_res fisher_res truth
## 1          P          P          P
## 2          N          N          N
## 3          P          P          P

##
##          1    2    3    4    5    6
## A 2735    5   50    3    6 4422
## B  2     0 5971    0    1 106
## C  4 5881    1    0    0 12
## D  0     1    1    0 4313    9
## E  0     0    3 3737    2 15
## F 325   12 1516   26 185 656

```

Cluster Results of Different Clusters in Different Samples



```

## `summarise()` regrouping output by 'condition' (override with `groups` argument)

## group variable has 2 levels, t-tests will be performed

## FlowSOM clustering completed in 0.7 seconds

##   methods      time
## 1 fisher 0.05186081
## 2 sepckle 0.01097012
## 3 dcats 0.49271393
## 4 diffcyt 0.02489781

##          A         B         C         D         E         F
## A 0.9895304224 3.721324e-03 1.972355e-04 4.016931e-04 1.894285e-04 0.005959897
## B 0.0045396809 9.860537e-01 1.092652e-05 4.185775e-06 0.000000e+00 0.009391529
## C 0.0002374481 1.078295e-05 9.985897e-01 0.000000e+00 7.831397e-06 0.001154205
## D 0.0006544651 5.590365e-06 0.000000e+00 9.965252e-01 0.000000e+00 0.002814773
## E 0.0003376020 0.000000e+00 1.159352e-05 0.000000e+00 9.965409e-01 0.003109914
## F 0.0154764427 1.999129e-02 2.489613e-03 4.486252e-03 4.531283e-03 0.953025118

## The following `from` values were not present in `x`: 6

##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1       A 4.177698e-01 1.056767e-02 6.915345e-14 6.419232e-02 8.004242e-03

```

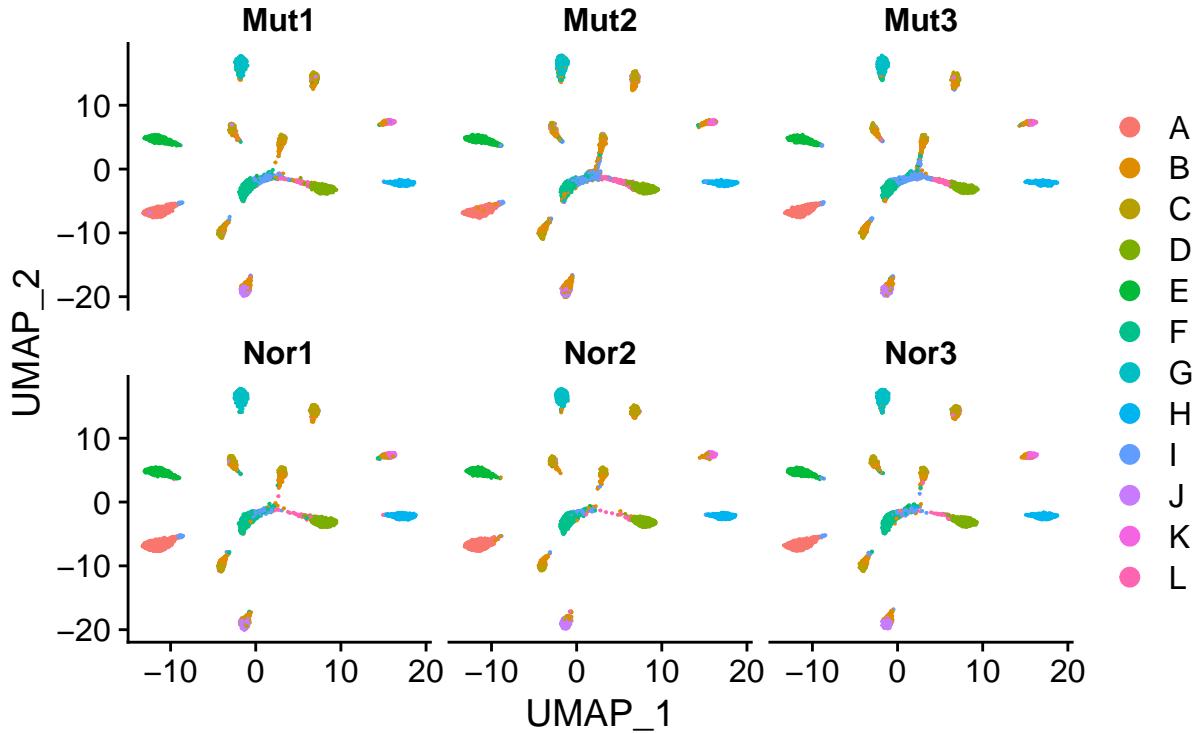
```

## 2      A 4.177698e-01  1.056767e-02  6.915345e-14  6.419232e-02  8.004242e-03
## 3      B 2.968052e-04  3.890542e-04  1.535730e-51  6.475499e-05  6.401807e-08
## 4      C 8.970005e-01  8.171343e-01  7.937153e-01  9.449535e-01  8.853123e-01
## 5      D 2.373009e-13  1.362793e-05  4.134732e-107 3.578108e-10  7.052909e-30
## 6      E 2.017931e-08  6.673056e-05  1.048835e-35  5.737274e-04  4.189064e-11
##   fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1 3.548384e-12      6.27e-03          P          N          P          N
## 2 3.548384e-12      6.27e-03          P          N          P          N
## 3 2.037301e-46      9.07e-09          P          P          P          P
## 4 8.156760e-01      8.74e-01          N          N          N          N
## 5 2.140089e-103    1.08e-33          P          P          P          P
## 6 7.006790e-34      1.29e-14          P          P          P          P
##   dcatsK_res fisher_res truth
## 1          P          P          N
## 2          P          P          P
## 3          P          P          P
## 4          N          N          N
## 5          P          P          P
## 6          P          P          P

##
##          1   2   3   4   5   6   7   8   9   10  11  12
## A  0   0   0   7   0 4204   6   8   0   0   11   0
## B  62   6 752 740   0   36 549 325  20   77 647 446
## C  5   0 619 607   0   0 781  85   1   0 169 1127
## D  3   0   9   0   0   0   1   8 3047   0   2   9
## E  1 2963   0   1   0   1   6   0   0   0   4   6
## F 131   2  21  51   0   9  32  30   5 2550   8  21
## G 2765   0   3   2   0   0   2  11   0  10   0   0
## H  0   0   2   1 2284   2   1   3   0   0   1   1
## I  44  68  70  58   6 167  38  36  50 1038  81  44
## J  0   0  20   4   0   5  56   3   3   0 1268 137
## K  4   0  23  26   0   0  25 725   1   5  17  35
## L  0   0   8   0   1   0   3   8 604   2  12   6

```

Cluster Results of Different Clusters in Different Samples



```

## `summarise()` regrouping output by 'condition' (override with `groups` argument)

## group variable has 2 levels, t-tests will be performed

##          A          B          C          D          E          F
## A 9.942028e-01 0.0031452694 6.742332e-05 0.000000e+00 3.759913e-06 2.099630e-04
## B 3.909429e-03 0.9280714710 2.607961e-02 3.029337e-03 1.054891e-03 1.164796e-02
## C 8.555579e-05 0.0266247073 9.421715e-01 1.974002e-04 7.029532e-05 0.000000e+00
## D 0.000000e+00 0.0033133533 2.114871e-04 9.871260e-01 0.000000e+00 3.958693e-05
## E 5.060673e-06 0.0011423059 7.456202e-05 0.000000e+00 9.960284e-01 8.626908e-05
## F 3.023873e-04 0.0134963076 0.000000e+00 4.193696e-05 9.230930e-05 9.621254e-01
## G 0.000000e+00 0.0041482774 3.544576e-04 0.000000e+00 0.000000e+00 3.125822e-03
## H 0.000000e+00 0.0002172208 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## I 4.480828e-03 0.0075344731 0.000000e+00 0.000000e+00 4.735808e-03 3.449745e-02
## J 1.417477e-03 0.0228328873 4.583093e-02 2.711766e-04 2.151443e-05 0.000000e+00
## K 0.000000e+00 0.0248989673 3.704138e-02 4.007333e-04 0.000000e+00 8.440051e-05
## L 0.000000e+00 0.0104371494 0.000000e+00 4.130576e-02 0.000000e+00 8.506028e-04
##          G          H          I          J          K          L
## A 0.000000e+00 0.0000000000 0.001865028 5.057648e-04 0.000000e+00 0.000000e+00
## B 3.481082e-03 0.0001577653 0.003897943 1.012626e-02 6.450177e-03 2.094077e-03
## C 3.036648e-04 0.0000000000 0.000000000 2.075060e-02 9.796282e-03 0.000000e+00
## D 0.000000e+00 0.0000000000 0.000000000 1.315408e-04 1.135444e-04 9.064452e-03
## E 0.000000e+00 0.0000000000 0.002653088 1.033221e-05 0.000000e+00 0.000000e+00
## F 3.039317e-03 0.0000000000 0.020679253 0.000000e+00 2.533382e-05 1.977438e-04
## G 9.917684e-01 0.0000000000 0.000000000 0.000000e+00 5.980887e-04 4.906133e-06

```

```

## H 0.000000e+00 0.9986528808 0.001091606 0.000000e+00 0.000000e+00 3.829205e-05
## I 0.000000e+00 0.0015324758 0.939622110 0.000000e+00 0.000000e+00 7.596856e-03
## J 0.000000e+00 0.0000000000 0.0000000000 9.251344e-01 4.491657e-03 0.000000e+00
## K 1.937411e-03 0.0000000000 0.0000000000 7.689613e-03 9.279475e-01 0.000000e+00
## L 2.051989e-05 0.0001386143 0.019588703 0.000000e+00 0.000000e+00 9.276586e-01

## FlowSOM clustering completed in 0.8 seconds

##   methods      time
## 1 fisher 0.06283307
## 2 sepckle 0.02988911
## 3 dcats 0.98340297
## 4 diffcyt 0.02892113

##          A         B         C         D         E         F
## A 9.942028e-01 0.0031452694 6.742332e-05 0.000000e+00 3.759913e-06 2.099630e-04
## B 3.909429e-03 0.9280714710 2.607961e-02 3.029337e-03 1.054891e-03 1.164796e-02
## C 8.555579e-05 0.0266247073 9.421715e-01 1.974002e-04 7.029532e-05 0.000000e+00
## D 0.000000e+00 0.0033133533 2.114871e-04 9.871260e-01 0.000000e+00 3.958693e-05
## E 5.060673e-06 0.0011423059 7.456202e-05 0.000000e+00 9.960284e-01 8.626908e-05
## F 3.023873e-04 0.0134963076 0.000000e+00 4.193696e-05 9.230930e-05 9.621254e-01
## G 0.000000e+00 0.0041482774 3.544576e-04 0.000000e+00 0.000000e+00 3.125822e-03
## H 0.000000e+00 0.0002172208 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## I 4.480828e-03 0.0075344731 0.000000e+00 0.000000e+00 4.735808e-03 3.449745e-02
## J 1.417477e-03 0.0228328873 4.583093e-02 2.711766e-04 2.151443e-05 0.000000e+00
## K 0.000000e+00 0.0248989673 3.704138e-02 4.007333e-04 0.000000e+00 8.440051e-05
## L 0.000000e+00 0.0104371494 0.000000e+00 4.130576e-02 0.000000e+00 8.506028e-04
##          G         H         I         J         K         L
## A 0.000000e+00 0.0000000000 0.001865028 5.057648e-04 0.000000e+00 0.000000e+00
## B 3.481082e-03 0.0001577653 0.003897943 1.012626e-02 6.450177e-03 2.094077e-03
## C 3.036648e-04 0.0000000000 0.0000000000 2.075060e-02 9.796282e-03 0.000000e+00
## D 0.000000e+00 0.0000000000 0.0000000000 1.315408e-04 1.135444e-04 9.064452e-03
## E 0.000000e+00 0.0000000000 0.002653088 1.033221e-05 0.000000e+00 0.000000e+00
## F 3.039317e-03 0.0000000000 0.020679253 0.000000e+00 2.533382e-05 1.977438e-04
## G 9.917684e-01 0.0000000000 0.0000000000 0.000000e+00 5.980887e-04 4.906133e-06
## H 0.000000e+00 0.9986528808 0.001091606 0.000000e+00 0.000000e+00 3.829205e-05
## I 0.000000e+00 0.0015324758 0.939622110 0.000000e+00 0.000000e+00 7.596856e-03
## J 0.000000e+00 0.0000000000 0.0000000000 9.251344e-01 4.491657e-03 0.000000e+00
## K 1.937411e-03 0.0000000000 0.0000000000 7.689613e-03 9.279475e-01 0.000000e+00
## L 2.051989e-05 0.0001386143 0.019588703 0.000000e+00 0.000000e+00 9.276586e-01

## The following `from` values were not present in `x`: 9, 12

##   cluster dcats_pvalsT speckle_pvals dcats_pvalsI dcats_pvalsU dcats_pvalsK
## 1          A 1.366602e-17 4.384371e-06 2.318417e-144 3.439411e-14 2.359555e-32
## 2          B 3.098999e-01 2.022811e-02 4.369705e-59 3.031216e-06 7.482894e-06
## 3          B 3.098999e-01 2.022811e-02 4.369705e-59 3.031216e-06 7.482894e-06
## 4          C 7.811424e-01 1.266076e-01 9.293011e-16 1.474389e-02 3.702031e-02
## 5          C 7.811424e-01 1.266076e-01 9.293011e-16 1.474389e-02 3.702031e-02
## 6          D 1.456067e-06 1.437613e-01 2.131318e-07 9.087476e-02 1.613793e-02
## 7          E 9.896959e-01 4.292095e-01 4.176920e-01 7.888033e-01 7.111766e-01
## 8          F 1.092523e-03 2.207706e-01 1.619117e-04 2.456426e-01 9.114757e-02
## 9          G 5.632034e-01 2.184163e-01 8.248474e-03 4.544517e-01 1.919576e-01

```

```

## 10      H 2.131207e-15  1.546210e-04  1.231673e-51 2.525137e-07 1.323557e-23
## 11      J 3.852530e-02  1.266076e-01  6.061071e-09 6.191537e-02 7.212741e-02
## 12      K 5.605161e-01  2.022811e-02  2.721754e-21 2.520262e-03 9.183845e-04
##    fisher_pvals diffcyt_pvals speckle_res dcatsT_res dcatsI_res dcatsU_res
## 1  1.615763e-142   5.49e-41          P          P          P          P
## 2  7.007956e-57   8.01e-04          P          N          P          P
## 3  7.007956e-57   8.01e-04          P          N          P          P
## 4  6.014620e-15   1.19e-01          N          N          P          P
## 5  6.014620e-15   1.19e-01          N          N          P          P
## 6  4.760586e-07   1.41e-01          N          P          P          N
## 7  4.411900e-01   6.62e-01          N          N          N          N
## 8  2.512444e-04   2.67e-01          N          P          P          N
## 9  1.051793e-02   3.14e-01          N          N          P          N
## 10 8.188840e-52   3.92e-15          P          P          P          P
## 11 9.257346e-09   7.53e-02          N          P          P          N
## 12 6.152285e-22   2.56e-04          P          N          P          P
##    dcatsK_res fisher_res truth
## 1          P          P          P
## 2          P          P          N
## 3          P          P          N
## 4          P          P          P
## 5          P          P          P
## 6          P          P          P
## 7          N          N          N
## 8          N          P          P
## 9          N          P          N
## 10         P          P          P
## 11         N          P          P
## 12         P          P          P

```

Different uniform matrix

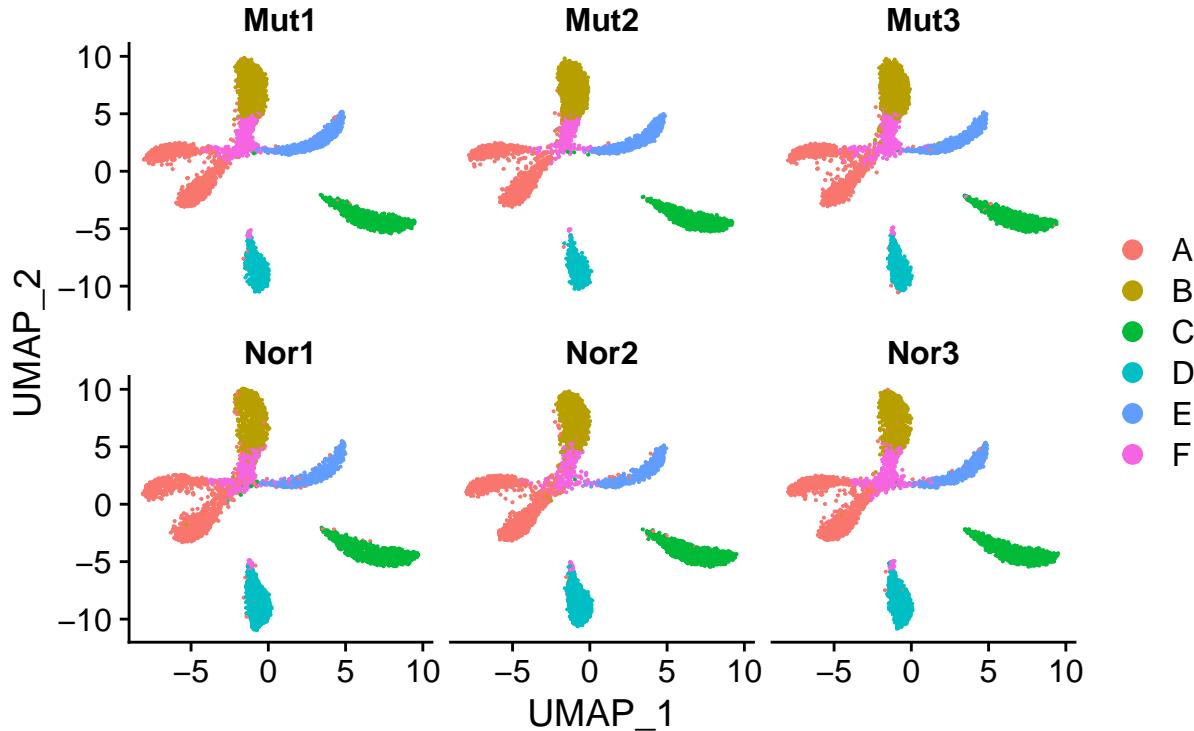
Test whether different uniform matrix will influence the results, U1 is closer to identity matrix

```

##
##      1   2   3   4   5   6
##  A 2735   5  50   3   6 4422
##  B   2   0 5971   0   1 106
##  C   4 5881   1   0   0  12
##  D   0   1   1   0 4313   9
##  E   0   0   3 3737   2  15
##  F 325  12 1516  26 185  656

```

Cluster Results of Different Clusters in Different Samples



```
## NULL
## NULL
## The following `from` values were not present in `x`: 6
##   cluster dcats_pvalsU1 dcats_pvalsU2 dcats_pvalsU3 truth
## 1       A 3.874062e-02 8.072225e-02 1.414918e-01     N
## 2       A 3.874062e-02 8.072225e-02 1.414918e-01     P
## 3       B 8.748132e-05 9.299262e-04 6.633241e-03     P
## 4       C 9.477409e-01 9.713016e-01 9.608215e-01     N
## 5       D 4.138199e-09 5.150710e-07 2.303158e-05     P
## 6       E 1.064904e-03 2.927763e-03 1.657375e-02     P
```

Codes

```
knitr::opts_chunk$set(echo = FALSE)
library(splatter)
library(Seurat)
library(speckle)
library(DCATS)
library(ggplot2)
library(tidyverse)
```

```

library(diffcyt)
source("functions.r")
library(future)
# check the current active plan
plan()
# change the current plan to access parallelization
plan("multiprocess", workers = 4)
plan()
options(future.globals.maxSize = 2000 * 1024^2)
set.seed(123)
probNor = c(0.1, 0.2, 0.1, 0.2, 0.2, 0.2)
probMut = c(0.1, 0.2, 0.05, 0.1, 0.25, 0.3)
de_prob = runif(6, 0.2, 0.5)

batch_size_list = seq(2500, 10000, length = 4)
setresolu_list = c(0.2, 0.2, 0.15, 0.2)
mnnCT7params = readRDS("./data/mnnCT7params.rds")
for (i in 1:4) {
  batch_size = batch_size_list[i]
  setresolu = setresolu_list[i]

  sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
  #integratedSamples = runSeurat(sim_list, batch_size, setresolu)

  name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", i)
  #saveRDS(integratedSamples, name)
  integratedSamples = readRDS(name)

  conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
  print(conf.mat)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTimeFSDD6(integratedSamples, sim_list, batch_size)
  print(Res$time_df)
  print(Res$knn_mat)

  cluster_map = apply(conf.mat, 2, which.max) %>%
    plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
    as_tibble() %>%
    mutate(truth = c("N", "N", "P", "P", "P", "P")) %>%
    dplyr::rename(cluster = value)

  Res$Res_df %>%
    mutate(speckle_res = ifelse(speckle_pvals < 0.05, "P", "N"),
      dcatsT_res = ifelse(dcats_pvalsT < 0.05, "P", "N"),
      dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
      dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
      dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
      fisher_res = ifelse(fisher_pvals < 0.05, "P", "N")) %>%
    merge(cluster_map, by = "cluster") %>%
    print()
}

}

```

```

#sendEmail("Simulation is done!")
cluster_map = apply(conf.mat, 2, which.max) %>%
  plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
  as_tibble() %>%
  mutate(truth = c("N", "N", "P", "P", "P", "P")) %>%
  dplyr::rename(cluster = value)
set.seed(123)
probNor = c(0.1, 0.2, 0.1, 0.2, 0.2, 0.2)
probMut = c(0.1, 0.2, 0.05, 0.1, 0.25, 0.3)
batch_size = 5000

de_prob_list = seq(0.05, 0.10, 0.01)
setresolu_list = c(0.25, 0.2, 0.2, 0.2, 0.2)
for (i in 1:5) {
  de_prob = c(0.5, 0.5, 0.1, 0.1, de_prob_list[i], de_prob_list[i])
  setresolu = setresolu_list[i]

  sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
  #integratedSamples = runSeurat(sim_list, batch_size, setresolu)

  name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", i+4)
  #saveRDS(integratedSamples, name)
  integratedSamples = readRDS(name)

  conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
  print(conf.mat)
  plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster")
  print(plot)
  Res = getPandTimeFSDD6(integratedSamples, sim_list, batch_size)
  print(Res$time_df)
  print(Res$knn_mat)

  cluster_map = apply(conf.mat, 2, which.max) %>%
    plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
    as_tibble() %>%
    mutate(truth = c("N", "N", "P", "P", "P", "P")) %>%
    dplyr::rename(cluster = value)

  Res$Res_df %>%
    mutate(speckle_res = ifelse(speckle_pvals < 0.05, "P", "N"),
      dcatsT_res = ifelse(dcats_pvalsT < 0.05, "P", "N"),
      dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
      dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
      dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
      fisher_res = ifelse(fisher_pvals < 0.05, "P", "N")) %>%
    merge(cluster_map, by = "cluster") %>%
    print()
}

#sendEmail("Simulation is done!")
## Three clusters
set.seed(123)
probNor = c(0.4, 0.1, 0.5)

```

```

probMut = c(0.4, 0.2, 0.4)
batch_size = 5000
de_prob = runif(3, 0.01, 0.4)
setresolu = 0.1

mnnCT7params = readRDS("./data/mnnCT7params.rds")
sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
#integratedSamples = runSeurat(sim_list, batch_size, setresolu)
name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", 10)
#saveRDS(integratedSamples, name)
integratedSamples = readRDS(name)

conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
print(conf.mat)
plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster 1")
print(plot)
Res = getPandTimeFSDD3(integratedSamples, sim_list, batch_size)
print(Res$time_df)
print(Res$knn_mat)

cluster_map = apply(conf.mat, 2, which.max) %>%
  plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
  as_tibble() %>%
  mutate(truth = c("N", "P", "P")) %>%
  dplyr::rename(cluster = value)
Res$Res_df %>%
  mutate(speckle_res = ifelse(speckle_pvals < 0.05, "P", "N"),
         dcatsT_res = ifelse(dcats_pvalsT < 0.05, "P", "N"),
         dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
         dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
         dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
         fisher_res = ifelse(fisher_pvals < 0.05, "P", "N")) %>%
  merge(cluster_map, by = "cluster") %>%
  print()

#sendEmail("Simulation is done!")
## Six clusters
set.seed(123)
probNor = c(0.1, 0.2, 0.2, 0.1, 0.2, 0.2)
probMut = c(0.1, 0.2, 0.3, 0.15, 0.1, 0.15)
batch_size = 5000
de_prob = runif(6, 0.01, 0.6)
setresolu = 0.2

mnnCT7params = readRDS("./data/mnnCT7params.rds")
sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
#integratedSamples = runSeurat(sim_list, batch_size, setresolu)
name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", 11)
#saveRDS(integratedSamples, name)
integratedSamples = readRDS(name)

conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
print(conf.mat)

```

```

plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster 1")
print(plot)
Res = getPandTimeFSDD6(integratedSamples, sim_list, batch_size)
print(Res$time_df)
print(Res$knn_mat)

cluster_map = apply(conf.mat, 2, which.max) %>%
  plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
  as_tibble() %>%
  mutate(truth = c("N", "N", "P", "P", "P", "P")) %>%
  dplyr::rename(cluster = value)
Res$Res_df %>%
  mutate(speckle_res = ifelse(speckle_pvals < 0.05, "P", "N"),
         dcatsT_res = ifelse(dcats_pvalsT < 0.05, "P", "N"),
         dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
         dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
         dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
         fisher_res = ifelse(fisher_pvals < 0.05, "P", "N")) %>%
  merge(cluster_map, by = "cluster") %>%
  print()

#sendEmail("Simulation is done!")
## Twelve clusters
set.seed(123)
probNor = c(0.1, 0.1, 0.05, 0.05, 0.1, 0.2, 0.05, 0.05, 0.1, 0.1, 0.05, 0.05)
probMut = c(0.1, 0.1, 0.05, 0.05, 0.05, 0.1, 0.05, 0.03, 0.15, 0.15, 0.1, 0.07)
batch_size = 5000
de_prob = runif(12, 0.3, 0.5)
setresolu = 0.58

mnnCT7params = readRDS("./data/mnnCT7params.rds")
sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
#integratedSamples = runSeurat(sim_list, batch_size, setresolu)
name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", 12)
#saveRDS(integratedSamples, name)
integratedSamples = readRDS(name)

conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
print(conf.mat)
plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster 1")
print(plot)
Res = getPandTimeFSDD12(integratedSamples, sim_list, batch_size)
print(Res$time_df)
print(Res$knn_mat)

cluster_map = apply(conf.mat, 2, which.max) %>%
  plyr::mapvalues(from = c(1:12), to = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L")) %>%
  as_tibble() %>%
  mutate(truth = c(rep("N",4), rep("P", 8))) %>%
  dplyr::rename(cluster = value)
Res$Res_df %>%
  mutate(speckle_res = ifelse(speckle_pvals < 0.05, "P", "N"),
         dcatsT_res = ifelse(dcats_pvalsT < 0.05, "P", "N"),
         dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
         dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
         dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
         fisher_res = ifelse(fisher_pvals < 0.05, "P", "N"))

```

```

dcatsI_res = ifelse(dcats_pvalsI < 0.05, "P", "N"),
dcatsU_res = ifelse(dcats_pvalsU < 0.05, "P", "N"),
dcatsK_res = ifelse(dcats_pvalsK < 0.05, "P", "N"),
fisher_res = ifelse(fisher_pvals < 0.05, "P", "N")) %>%
  merge(cluster_map, by = "cluster") %>%
  print()

#sendEmail("Simulation is done!")
set.seed(123)
probNor = c(0.1, 0.2, 0.2, 0.1, 0.2, 0.2)
probMut = c(0.1, 0.2, 0.3, 0.15, 0.1, 0.15)
batch_size = 5000
de_prob = runif(6, 0.01, 0.5)
setresolu = 0.2

mnnCT7params = readRDS("./data/mnnCT7params.rds")
sim_list = simualtionRWD(probNor, probMut, de_prob, batch_size, params = mnnCT7params)
#integratedSamples = runSeurat(sim_list, batch_size, setresolu)
name = sprintf("./data/simulation_plus3/integratedSamples%d.rds", 11)
#saveRDS(integratedSamples, name)
integratedSamples = readRDS(name)

conf.mat<-table(Idents(integratedSamples), sim_list$origLabels)
print(conf.mat)
plot = DimPlot(integratedSamples, ncol = 3, reduction = "umap", split.by = "batch") + ggtitle("Cluster Map")
print(plot)

Res = getPandTimeTestU(integratedSamples, sim_list)
print(Res$time_df)
print(Res$knn_mat)

cluster_map = apply(conf.mat, 2, which.max) %>%
  plyr::mapvalues(from = c(1:6), to = c("A", "B", "C", "D", "E", "F")) %>%
  as_tibble() %>%
  mutate(truth = c("N", "N", "P", "P", "P", "P")) %>%
  dplyr::rename(cluster = value)
Res %>%
  merge(cluster_map, by = "cluster") %>%
  print()

```