Figures Draft

Xinyi Lin

3/26/2021

## Figure A

confusion matrix

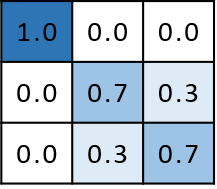
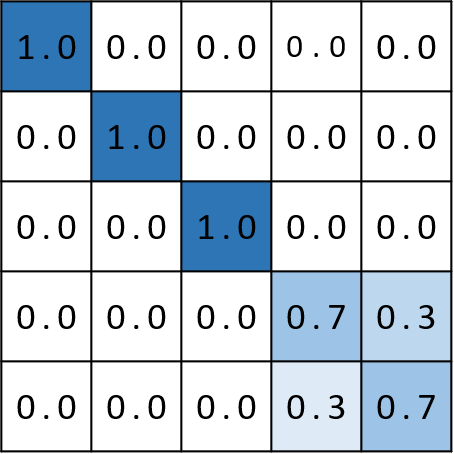
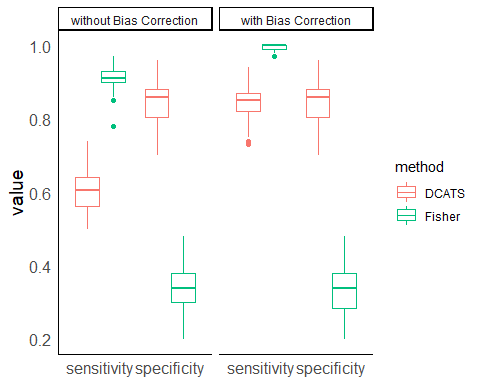
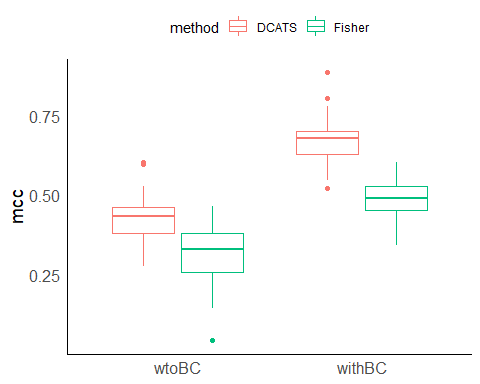


Figure A-1

#



## Saving 5 x 4 in image



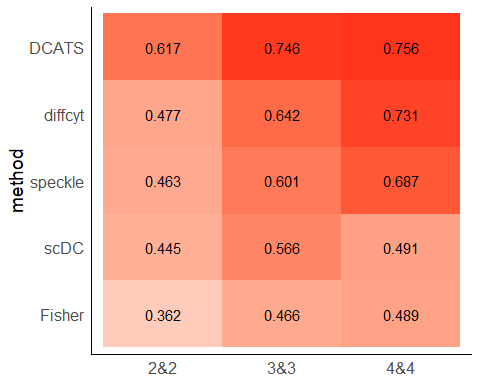
## Saving 5 x 4 in image

### plot

## Figure B

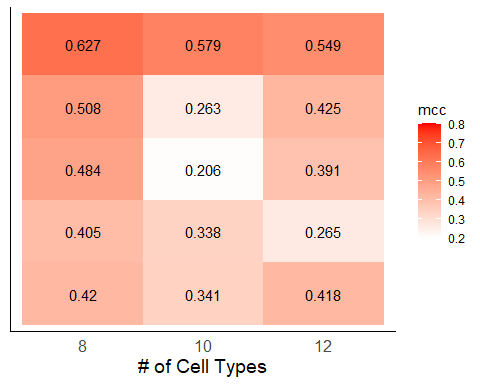
### Different number of replicates

Try to use heatmap

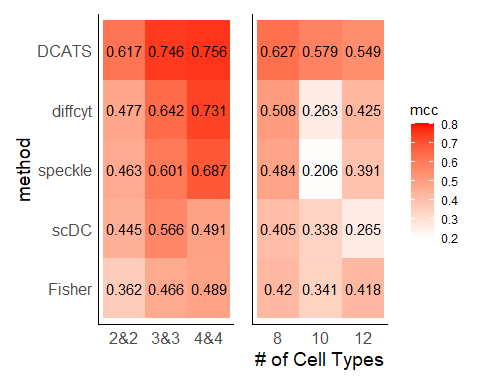
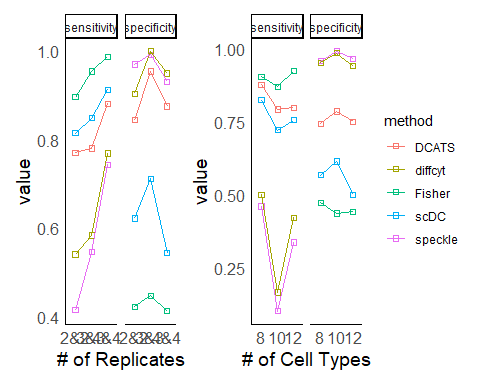
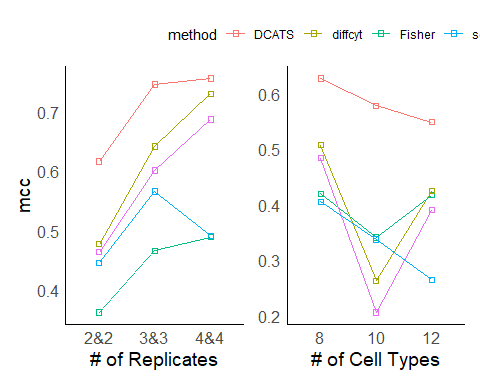


|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| method | mcc | auc | sensitivity | specificity | F1 | replicates |
| betabin\_wMSVM | 0.631 | 0.869 | 0.852 | 0.778 | 0.821 | 2&2 |
| betabin\_wMK | 0.620 | 0.872 | 0.874 | 0.741 | 0.819 | 2&2 |
| betabin\_wMU | 0.617 | 0.862 | 0.770 | 0.844 | 0.800 | 2&2 |
| betabin\_noBC | 0.491 | 0.802 | 0.696 | 0.793 | 0.732 | 2&2 |
| diffcyt | 0.477 | 0.849 | 0.541 | 0.904 | 0.661 | 2&2 |
| speckle | 0.463 | 0.838 | 0.415 | 0.970 | 0.574 | 2&2 |
| scDC | 0.445 | 0.758 | 0.815 | 0.622 | 0.743 | 2&2 |
| fisher | 0.362 | 0.854 | 0.896 | 0.422 | 0.725 | 2&2 |
| betabin\_wMU | 0.746 | 0.917 | 0.780 | 0.955 | 0.855 | 3&3 |
| betabin\_wMSVM | 0.705 | 0.929 | 0.924 | 0.773 | 0.859 | 3&3 |
| betabin\_wMK | 0.685 | 0.928 | 0.924 | 0.750 | 0.850 | 3&3 |
| diffcyt | 0.642 | 0.886 | 0.583 | 1.000 | 0.737 | 3&3 |
| speckle | 0.601 | 0.873 | 0.545 | 0.992 | 0.702 | 3&3 |
| betabin\_noBC | 0.566 | 0.882 | 0.712 | 0.848 | 0.764 | 3&3 |
| scDC | 0.566 | 0.849 | 0.848 | 0.712 | 0.794 | 3&3 |
| fisher | 0.466 | 0.916 | 0.955 | 0.447 | 0.761 | 3&3 |
| betabin\_wMU | 0.756 | 0.943 | 0.881 | 0.875 | 0.879 | 4&4 |
| diffcyt | 0.731 | 0.930 | 0.769 | 0.950 | 0.845 | 4&4 |
| betabin\_wMSVM | 0.716 | 0.939 | 0.956 | 0.744 | 0.864 | 4&4 |
| betabin\_wMK | 0.715 | 0.942 | 0.969 | 0.725 | 0.864 | 4&4 |
| speckle | 0.687 | 0.913 | 0.744 | 0.931 | 0.821 | 4&4 |
| betabin\_noBC | 0.650 | 0.899 | 0.812 | 0.838 | 0.823 | 4&4 |
| scDC | 0.491 | 0.862 | 0.912 | 0.544 | 0.770 | 4&4 |
| fisher | 0.489 | 0.947 | 0.988 | 0.412 | 0.767 | 4&4 |

### Different number of cell types



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| method | mcc | auc | sensitivity | specificity | F1 | clustersN |
| betabin\_wMSVM | 0.593 | 0.878 | 0.800 | 0.793 | 0.797 | 10 |
| betabin\_wMK | 0.579 | 0.877 | 0.793 | 0.786 | 0.790 | 10 |
| betabin\_wMU | 0.522 | 0.800 | 0.607 | 0.893 | 0.708 | 10 |
| fisher | 0.341 | 0.792 | 0.871 | 0.436 | 0.716 | 10 |
| scDC | 0.338 | 0.753 | 0.721 | 0.614 | 0.685 | 10 |
| betabin\_noBC | 0.329 | 0.707 | 0.507 | 0.807 | 0.597 | 10 |
| diffcyt | 0.263 | 0.774 | 0.164 | 0.986 | 0.279 | 10 |
| speckle | 0.206 | 0.770 | 0.100 | 0.993 | 0.181 | 10 |
| betabin\_wMK | 0.549 | 0.864 | 0.798 | 0.750 | 0.779 | 12 |
| betabin\_wMSVM | 0.549 | 0.861 | 0.798 | 0.750 | 0.779 | 12 |
| betabin\_wMU | 0.546 | 0.840 | 0.689 | 0.850 | 0.749 | 12 |
| betabin\_noBC | 0.508 | 0.827 | 0.706 | 0.800 | 0.740 | 12 |
| diffcyt | 0.425 | 0.809 | 0.420 | 0.942 | 0.568 | 12 |
| fisher | 0.418 | 0.834 | 0.924 | 0.442 | 0.743 | 12 |
| speckle | 0.391 | 0.822 | 0.336 | 0.967 | 0.491 | 12 |
| scDC | 0.265 | 0.712 | 0.756 | 0.500 | 0.669 | 12 |
| betabin\_wMK | 0.627 | 0.885 | 0.878 | 0.743 | 0.823 | 8 |
| betabin\_wMSVM | 0.613 | 0.886 | 0.865 | 0.743 | 0.815 | 8 |
| betabin\_wMU | 0.610 | 0.867 | 0.764 | 0.845 | 0.796 | 8 |
| diffcyt | 0.508 | 0.855 | 0.500 | 0.953 | 0.646 | 8 |
| betabin\_noBC | 0.506 | 0.815 | 0.676 | 0.824 | 0.730 | 8 |
| speckle | 0.484 | 0.844 | 0.459 | 0.959 | 0.613 | 8 |
| fisher | 0.420 | 0.845 | 0.905 | 0.473 | 0.744 | 8 |
| scDC | 0.405 | 0.768 | 0.824 | 0.568 | 0.731 | 8 |



## Figure C

Figure C

Figure C

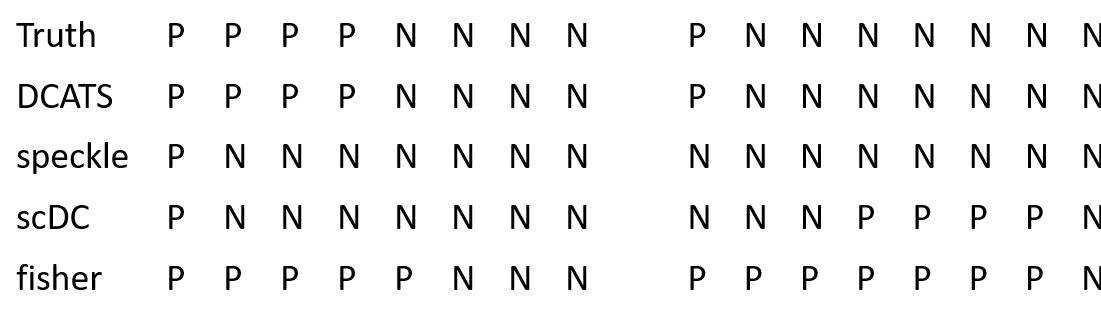


Figure C1

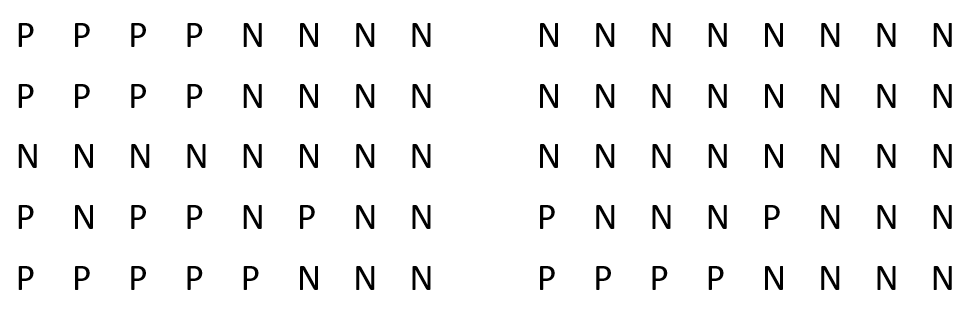


Figure C2

## real-world data 1 - Experiment 7

The ‘group’ column started with ‘B’ is the indicators of replicates

## batch barcode condition clusterRes x  
## 1 B1 AAACATACCACAAC Control Enterocyte.Progenitor Enterocyte.Progenitor  
## 2 B1 AAACGCACGAGGAC Control Stem Stem  
## 3 B1 AAACGCACTAGCCA Control Stem Stem  
## 4 B1 AAACGCACTGTCCC Control Stem Stem  
## 5 B1 AAACTTGACCACCT Control Enterocyte.Progenitor Enterocyte.Progenitor  
## 6 B1 AAAGATCTACCTTT Control Enterocyte.Progenitor Enterocyte.Progenitor

## batch barcode condition clusterRes   
## Length:9842 Length:9842 Length:9842 Length:9842   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
## x   
## Length:9842   
## Class :character   
## Mode :character

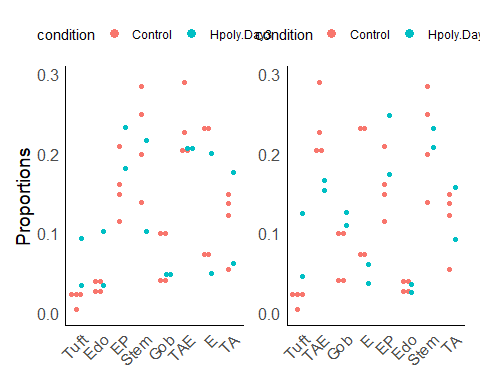
## `summarise()` has grouped output by 'batch'. You can override using the `.groups` argument.

## # A tibble: 10 x 3  
## # Groups: batch [10]  
## batch condition n  
## <chr> <chr> <int>  
## 1 B1 Control 840  
## 2 B10 Salmonella 950  
## 3 B2 Control 200  
## 4 B3 Control 1258  
## 5 B4 Control 942  
## 6 B5 Hpoly.Day3 1490  
## 7 B6 Hpoly.Day3 631  
## 8 B7 Hpoly.Day10 1169  
## 9 B8 Hpoly.Day10 1542  
## 10 B9 Salmonella 820

## `summarise()` has grouped output by 'batch', 'condition'. You can override using the `.groups` argument.

## `summarise()` has grouped output by 'condition'. You can override using the `.groups` argument.

## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.



## Saving 5 x 4 in image  
## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.

## real-world data2

## cell tsne1 tsne2 ind condition cluster clusterRes  
## 1 AAACATACAATGCC-1 -4.277833 -19.294709 107 ctrl 5 CD4 T cells  
## 2 AAACATACATTTCC-1 -27.640373 14.966629 1016 ctrl 9 CD14+ Monocytes  
## 3 AAACATACCAGAAA-1 -27.493646 28.924885 1256 ctrl 9 CD14+ Monocytes  
## 4 AAACATACCAGCTA-1 -28.132584 24.925484 1256 ctrl 9 CD14+ Monocytes  
## 5 AAACATACCATGCA-1 -10.468194 -5.984389 1488 ctrl 3 CD4 T cells  
## 6 AAACATACCTCGCT-1 -24.367997 20.429285 1256 ctrl 9 CD14+ Monocytes  
## multiplets  
## 1 doublet  
## 2 singlet  
## 3 singlet  
## 4 doublet  
## 5 singlet  
## 6 singlet

## `summarise()` has grouped output by 'condition'. You can override using the `.groups` argument.

## # A tibble: 16 x 3  
## # Groups: condition [2]  
## condition clusterRes n  
## <chr> <chr> <int>  
## 1 ctrl B cells 1488  
## 2 ctrl CD14+ Monocytes 3365  
## 3 ctrl CD4 T cells 6005  
## 4 ctrl CD8 T cells 1409  
## 5 ctrl Dendritic cells 227  
## 6 ctrl FCGR3A+ Monocytes 906  
## 7 ctrl Megakaryocytes 166  
## 8 ctrl NK cells 1051  
## 9 stim B cells 1392  
## 10 stim CD14+ Monocytes 3082  
## 11 stim CD4 T cells 6028  
## 12 stim CD8 T cells 1225  
## 13 stim Dendritic cells 245  
## 14 stim FCGR3A+ Monocytes 1008  
## 15 stim Megakaryocytes 180  
## 16 stim NK cells 1279

## # A tibble: 0 x 2  
## # ... with 2 variables: cell <chr>, n <int>

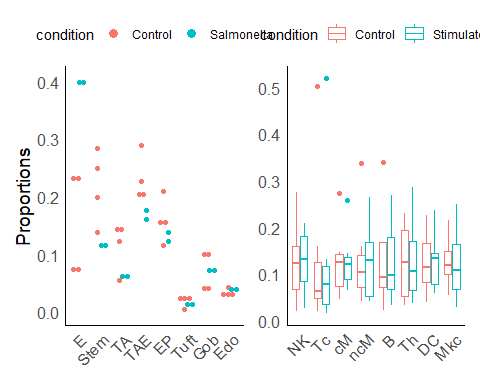
## BARCODE RD.TOTL RD.PASS RD.UNIQ N.SNP BEST SNG.1ST  
## 1 AAACATACAATGCC-1 5799 280 269 185 DBL-107-1244-0.500 107  
## 2 AAACATACATTTCC-1 5466 592 501 236 SNG-1016 1016  
## 3 AAACATACCAGAAA-1 4337 330 300 154 SNG-1256 1256  
## 4 AAACATACCAGCTA-1 7120 418 338 179 DBL-1256-1244-0.500 1256  
## 5 AAACATACCATGCA-1 2422 86 76 54 SNG-1488 1488  
## 6 AAACATACCTCGCT-1 5312 561 497 207 SNG-1256 1256  
## SNG.LLK1 SNG.2ND SNG.LLK2 SNG.LLK0 DBL.1ST DBL.2ND ALPHA LLK12  
## 1 -78.5388 1244 -92.3864 -75.9878 107 1244 0.5 -62.8745  
## 2 -61.8525 1256 -193.3940 -135.9317 1016 1256 0.5 -93.0213  
## 3 -50.6213 1015 -100.6965 -67.2587 101 1256 0.5 -64.5022  
## 4 -70.2734 1244 -110.3962 -82.1916 1256 1244 0.5 -67.3787  
## 5 -16.1294 1015 -31.7146 -27.0167 1015 1488 0.5 -18.5646  
## 6 -66.3586 1015 -180.3066 -115.0556 1256 101 0.5 -101.4645  
## LLK1 LLK2 LLK10 LLK20 LLK00 PRB.DBL PRB.SNG1  
## 1 -78.5388 -92.3864 -88.6649 -94.1498 -77.4107 1.00e+00 1  
## 2 -61.8525 -193.3940 -103.2284 -168.3927 -132.6010 4.15e-15 1  
## 3 -107.8190 -50.6213 -111.3319 -64.5022 -72.4244 1.95e-07 1  
## 4 -70.2734 -110.3962 -80.3745 -103.1706 -85.1143 7.21e-01 1  
## 5 -31.7146 -16.1294 -30.2045 -22.5295 -28.2980 1.28e-02 1  
## 6 -66.3586 -190.9367 -101.4645 -194.9529 -124.2197 8.42e-17 1

## BARCODE RD.TOTL RD.PASS RD.UNIQ N.SNP BEST SNG.1ST SNG.LLK1  
## 1 AAACATACCAAGCT-1 2491 239 228 141 SNG-101 101 -31.2822  
## 2 AAACATACCCCTAC-1 3747 282 274 173 SNG-1488 1488 -37.4261  
## 3 AAACATACCCGTAA-1 2844 144 133 108 SNG-1244 1244 -35.0043  
## 4 AAACATACCCTCGT-1 3575 199 189 115 SNG-1488 1488 -31.5489  
## 5 AAACATACGAGGTG-1 2649 170 158 99 SNG-1488 1488 -33.2782  
## 6 AAACATACGCGAAG-1 9874 880 822 345 SNG-101 101 -139.7810  
## SNG.2ND SNG.LLK2 SNG.LLK0 DBL.1ST DBL.2ND ALPHA LLK12 LLK1  
## 1 1015 -81.3393 -59.5871 101 1015 0.5 -47.8026 -31.2822  
## 2 1256 -104.7874 -74.9247 1256 1488 0.5 -53.8991 -104.7874  
## 3 1488 -61.5604 -51.6974 1488 1244 0.5 -36.9775 -61.5604  
## 4 1244 -79.8040 -57.9696 1256 1488 0.5 -42.8116 -83.5802  
## 5 1244 -66.4966 -48.9522 101 1488 0.5 -36.4330 -75.8228  
## 6 1256 -266.4763 -170.1924 101 1488 0.5 -154.9333 -139.7810  
## LLK2 LLK10 LLK20 LLK00 PRB.DBL PRB.SNG1  
## 1 -81.3393 -32.1591 -47.8026 -64.9647 1.07e-08 1  
## 2 -37.4261 -101.7666 -64.5707 -79.2115 1.00e-08 1  
## 3 -35.0043 -72.6479 -50.8297 -52.6835 2.03e-02 1  
## 4 -31.5489 -80.4661 -49.8506 -58.0733 2.38e-06 1  
## 5 -33.2782 -72.2344 -36.4330 -48.3017 8.06e-03 1  
## 6 -273.8806 -136.9040 -154.9333 -175.5614 3.82e-08 1

## cell batch  
## 1 AAACATACATTTCC-1 SNG-1016  
## 2 AAACATACCAGAAA-1 SNG-1256  
## 3 AAACATACCATGCA-1 SNG-1488  
## 4 AAACATACCTCGCT-1 SNG-1256  
## 5 AAACATACCTGGTA-1 SNG-1039  
## 6 AAACATACGATGAA-1 SNG-1488

## `summarise()` has grouped output by 'condition', 'clusterRes'. You can override using the `.groups` argument.

## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.



## Saving 5 x 4 in image  
## `stat\_bindot()` using `bins = 30`. Pick better value with `binwidth`.