

# Getting to know CELESTA

2024-05-04

```
library(CELESTA)
library(Rmixmod)

## Loading required package: Rcpp
## Rmixmod v. 2.1.10 / URI: www.mixmod.org
library(spdep)

## Loading required package: spData
## To access larger datasets in this package, install the spDataLarge
## package with: `install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')`
## Loading required package: sf
## Linking to GEOS 3.10.2, GDAL 3.4.1, PROJ 8.2.1; sf_use_s2() is TRUE
library(ggplot2)
library(reshape2)
library(zeallot)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
print(getwd())

## [1] "/home/pampuch/studia/dls/DL_project/CELESTA"
writeLines(capture.output(sessionInfo()), "sessionInfo.txt")
```

## Function running celesta

Function based on CELESTA README on github (<https://github.com/plevritis-lab/CELESTA/blob/main/README.md>)

```
run_celesta <- function(marker_info_file, imaging_data_file, proj_title=NULL) {
  prior_marker_info <- read.csv(marker_info_file)
  imaging_data <- read.csv(imaging_data_file)
  colnames(prior_marker_info)[1] <- ""
  if(is.null(proj_title)) proj_title <- tail(strsplit(strsplit(imaging_data_file, "[.]")[[1]][1], "[/]").
  CelestaObj <- CreateCelestaObject(project_title = proj_title, prior_marker_info, imaging_data)
  CelestaObj <- FilterCells(CelestaObj, high_marker_threshold=0.9, low_marker_threshold=0.4)

  CelestaObj <- AssignCells(CelestaObj, max_iteration=10, cell_change_threshold=0.01,
                           high_expression_threshold_anchor=high_marker_threshold_anchor,
                           low_expression_threshold_anchor=low_marker_threshold_anchor,
                           high_expression_threshold_index=high_marker_threshold_iteration,
                           low_expression_threshold_index=low_marker_threshold_iteration)
}
```

## CELESTA on data provided on CELESTA github

1. prior marker info: [https://github.com/plevritis-lab/CELESTA/blob/main/data/prior\\_marker\\_info.csv](https://github.com/plevritis-lab/CELESTA/blob/main/data/prior_marker_info.csv)
2. imaging data: [https://github.com/plevritis-lab/CELESTA/blob/main/data/imaging\\_data.csv](https://github.com/plevritis-lab/CELESTA/blob/main/data/imaging_data.csv)

```
img_path <- "data/celesta_org_data/imaging_data.csv"
marker_info_path <- "data/celesta_org_data/prior_marker_info.csv"

run_celesta(marker_info_path, img_path, proj_title = "celesta_on celesta_data")
```

```
## Marker: CD31...vasculature.Cyc_19_ch_3
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: CD34...vasculature.Cyc_20_ch_3
## Marker: Cytokeratin...epithelia.Cyc_10_ch_2
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: aSMA...smooth.muscle.Cyc_11_ch_2
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: Podoplanin...lymphatics.Cyc_19_ch_4
## Marker: CD45...hematopoietic.cells.Cyc_4_ch_2
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: CD15...granulocytes.Cyc_14_ch_2
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: CD3...T.cells.Cyc_16_ch_4
## Marker: CD20...B.cells.Cyc_8_ch_3
## [1] "Warning: The marker expression potentially has too many zeros for\n      fitting. GMM fitting
## Marker: CD11c...DCs.Cyc_12_ch_3
## Marker: CD163...macrophages.Cyc_17_ch_3
## Marker: CD68...macrophages.Cyc_18_ch_4
## Marker: CD38...multifunctional.Cyc_20_ch_4
## Marker: CD56...NK.cells.Cyc_10_ch_4
## Marker: CD8...cytotoxic.T.cells.Cyc_3_ch_2
## Marker: CD4...T.helper.cells.Cyc_6_ch_3
## Marker: CD45RO...memory.cells.Cyc_18_ch_3
## Marker: FOXP3...regulatory.T.cells.Cyc_2_ch_3
## [1] "Getting the nearest neighbors"
## [1] "Identifying neighboring cells within a defined circle bandwidth"
## [1] "Identify distances for all the cells within the circle bandwidth"
## [1] "Total cells to be assigned in the current round: 2432"
## [1] "Start calculating the scoring function."
## [1] "Too few cells identified for: lymphatics"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## vasculature           1    128 0.05258833
## tumor cells           2     28 0.01150370
## aSMA+ stroma           3    423 0.17378800
## lymphatics            4      0 0.00000000
## immune cells           5    691 0.28389482
## [1] "Get distance from nearest assigned cells."
## [1] "Updating cell probability done."
## [1] "Too few cells identified for: lymphatics"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## vasculature           1    134 0.05505341
## tumor cells           2     49 0.02013147
```

```

## aSMA+ stroma          3    428 0.17584224
## lymphatics           4      0 0.00000000
## immune cells         5   1376 0.56532457
## [1] "Proportion of cells changed assignment in the last iteration: 0.294819078947368"
## [1] "Get distance from nearest assigned cells."
## [1] "Start calculating the scoring function."
## [1] "Updating cell probability done."
## [1] "Too few cells identified for: lymphatics"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## vasculature          1    134 0.05505341
## tumor cells          2     49 0.02013147
## aSMA+ stroma         3    428 0.17584224
## lymphatics           4      0 0.00000000
## immune cells         5   1376 0.56532457
## [1] "Proportion of cells changed assignment in the last iteration: 0"
## [1] "Total cells to be assigned in the current round: 1376"
## [1] "Start calculating the scoring function."
## [1] "Too few cells identified for: CD11c+ DCs"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: NK cells"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## CD3+ T cells          6    549 0.225554643
## CD15+ granulocytes    7    139 0.057107642
## B cells               8      3 0.001232539
## CD11c+ DCs           9      0 0.000000000
## CD68+CD163+ macrophages 10   148 0.060805259
## plasma cells         11     52 0.021364010
## NK cells             12      0 0.000000000
## [1] "Get distance from nearest assigned cells."
## [1] "Updating cell probability done."
## [1] "Too few cells identified for: CD11c+ DCs"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: NK cells"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## CD3+ T cells          6    553 0.227198028
## CD15+ granulocytes    7    140 0.057518488
## B cells               8      3 0.001232539
## CD11c+ DCs           9      0 0.000000000
## CD68+CD163+ macrophages 10   159 0.065324569
## plasma cells         11     53 0.021774856
## NK cells             12      0 0.000000000
## [1] "Proportion of cells changed assignment in the last iteration: 0.0123546511627907"
## [1] "Get distance from nearest assigned cells."
## [1] "Start calculating the scoring function."
## [1] "Updating cell probability done."
## [1] "Too few cells identified for: CD11c+ DCs"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: NK cells"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count proportion
## CD3+ T cells          6    553 0.227198028

```

```

## CD15+ granulocytes          7   140 0.057518488
## B cells                     8     3 0.001232539
## CD11c+ DCs                  9     0 0.000000000
## CD68+CD163+ macrophages    10   159 0.065324569
## plasma cells                11    53 0.021774856
## NK cells                    12     0 0.000000000
## [1] "Proportion of cells changed assignment in the last iteration: 0"
## [1] "Total cells to be assigned in the current round: 553"
## [1] "Start calculating the scoring function."
##           cell_type_number count proportion
## CD8+ T cells          13    72 0.02958094
## CD4+ T cells          14   234 0.09613804
## [1] "Get distance from nearest assigned cells."
## [1] "Updating cell probability done."
##           cell_type_number count proportion
## CD8+ T cells          13   121 0.04971241
## CD4+ T cells          14   394 0.16187346
## [1] "Proportion of cells changed assignment in the last iteration: 0.377938517179023"
## [1] "Get distance from nearest assigned cells."
## [1] "Start calculating the scoring function."
## [1] "Updating cell probability done."
##           cell_type_number count proportion
## CD8+ T cells          13   121 0.04971241
## CD4+ T cells          14   395 0.16228431
## [1] "Proportion of cells changed assignment in the last iteration: 0.00180831826401447"
## [1] "Total cells to be assigned in the current round: 395"
## [1] "Start calculating the scoring function."
##           cell_type_number count proportion
## CD4+ T cells CD45R0+      15   136 0.05587510
## Tregs                  16    65 0.02670501
## [1] "Get distance from nearest assigned cells."
## [1] "Updating cell probability done."
##           cell_type_number count proportion
## CD4+ T cells CD45R0+      15   273 0.11216105
## Tregs                  16    65 0.02670501
## [1] "Proportion of cells changed assignment in the last iteration: 0.346835443037975"
## [1] "Get distance from nearest assigned cells."
## [1] "Start calculating the scoring function."
## [1] "Updating cell probability done."
##           cell_type_number count proportion
## CD4+ T cells CD45R0+      15   286 0.11750205
## Tregs                  16    65 0.02670501
## [1] "Proportion of cells changed assignment in the last iteration: 0.0329113924050633"
## [1] "Get distance from nearest assigned cells."
## [1] "Start calculating the scoring function."
## [1] "Updating cell probability done."
##           cell_type_number count proportion
## CD4+ T cells CD45R0+      15   286 0.11750205
## Tregs                  16    65 0.02670501
## [1] "Proportion of cells changed assignment in the last iteration: 0"

```

## CELESTA on a subset of our data

```
img_path_1 <- "data/exprs_data/IMMUcan_Batch20220908_S-220729-00002_002_roi_2.csv"
marker_info_path_1 <- "data/marker_info_data/markers_table_same_round.csv"
```

```
run_celesta(marker_info_path_1, img_path_1)
```

```
## Marker: CD68
## Marker: CD163
## Marker: CD14
## Marker: CD11c
## Marker: CD206
## Marker: CD33
## Marker: CD40
## Marker: SMA
## Marker: CD27
## Marker: CD45RA
## Marker: CD20
## Marker: CD3
## Marker: CD7
## Marker: GrzB
## Marker: TCF7
## Marker: CD45RO
## Marker: FOXP3
## Marker: ICOS
## Marker: CD8a
## Marker: CD4
## Marker: Ecad
## Marker: CD303
## Marker: HistoneH3
## Marker: DNA1
## Marker: DNA2
## Marker: CD38
## Marker: HLADR
## Marker: MPO
## Marker: CD16
## Marker: CD15
## Marker: Ido1
## Marker: LAG3
## Marker: PD1
## Marker: PDGFRb
## Marker: PDL1
## Marker: Ki67
## Marker: cleavedPARP
## Marker: B2M
## Marker: CarbonicAnhydrase
## Marker: VISTA
## [1] "Getting the nearest neighbors"
## [1] "Identifying neighboring cells within a defined circle bandwidth"
## [1] "Identify distances for all the cells within the circle bandwidth"
## [1] "Total cells to be assigned in the current round: 2020"
## [1] "Start calculating the scoring function."
## [1] "Too few cells identified for: B"
## [1] "Please consider relaxing the threshold."
```

```

## [1] "Too few cells identified for: DC"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: HLADR"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: NK"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: pDC"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Tumor"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Treg"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: CD4"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Mural"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: MacCD163"
## [1] "Please consider relaxing the threshold."
##           cell_type_number count   proportion
## B                1         0 0.0000000000
## Neutrophil        2         1 0.0004950495
## DC                3         0 0.0000000000
## HLADR             4         0 0.0000000000
## NK                5         0 0.0000000000
## plasma           6        32 0.0158415842
## pDC               7         0 0.0000000000
## Tumor            8         0 0.0000000000
## Treg             9         0 0.0000000000
## CD8              10         1 0.0004950495
## CD4              11         0 0.0000000000
## Mural            12         0 0.0000000000
## MacCD163         13         0 0.0000000000
## [1] "Get distance from nearest assigned cells."
## [1] "Updating cell probability done."
## [1] "Too few cells identified for: B"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: DC"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: HLADR"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: NK"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: pDC"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Tumor"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Treg"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: CD4"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: Mural"
## [1] "Please consider relaxing the threshold."
## [1] "Too few cells identified for: MacCD163"
## [1] "Please consider relaxing the threshold."

```

```
##          cell_type_number count  proportion
## B                      1    0 0.000000000
## Neutrophil             2    3 0.001485149
## DC                     3    0 0.000000000
## HLADR                   4    0 0.000000000
## NK                      5    0 0.000000000
## plasma                  6   32 0.015841584
## pDC                     7    0 0.000000000
## Tumor                   8    0 0.000000000
## Treg                    9    0 0.000000000
## CD8                     10   3 0.001485149
## CD4                     11    0 0.000000000
## Mural                   12    0 0.000000000
## MacCD163                13    0 0.000000000
## [1] "Proportion of cells changed assignment in the last iteration: 0.00198019801980198"
```

The marker info had lineage level corresponding to considering all cell types in the same round (probably does not make sense):

```
read.csv(marker_info_path_1)
```

```
##          X Lineage_level CD68 CD163 CD14 CD11c CD206 CD33 CD40 SMA CD27
## 1         B           1_0_1    0    0    0    1    0    0    1    0  NA
## 2  Neutrophil       1_0_2    0    0    0    1    0    1    0    0    0
## 3         DC           1_0_3    0    0    0    1    NA    0    1    0    0
## 4        HLADR       1_0_4    1    0    0    NA    0    0    0    0  NA
## 5         NK           1_0_5    0    0    0    0    0    0    0    0    0
## 6        plasma     1_0_6    0    0    0    0    0    0    0    0    1
## 7         pDC        1_0_7    0    0    0    0    0    0    1    0    0
## 8         Tumor      1_0_8    0    0    0    0    0    NA    0    0    0
## 9         Treg       1_0_9    0    0    0    0    0    0    0    0  NA
## 10        CD8        1_0_10   0    0    0    0    0    0    0    0  NA
## 11        CD4        1_0_11   0    0    0    0    0    0    0    0  NA
## 12        Mural      1_0_12   0    0    0    0    0    0    0    1    0
## 13  MacCD163        1_0_13   1    1    1    1    1    NA    1    0    0
##      CD45RA CD20 CD3 CD7 GrzB TCF7 CD45RO FOXP3 ICOS CD8a CD4 Ecad CD303
## 1         1    1    0    0    0    NA    0    0    0    0    0    0    0
## 2         0    0    0    0    0    0    0    0    0    0    0    0    0
## 3         0    0    0    0    0    0    0    0    NA    0    0    0    0
## 4         0    0    0    0    0    0    0    0    0    0    0    0    0
## 5         0    0    0    1    NA    0    0    0    0    0    0    0    0
## 6         0    0    0    0    0    NA    0    0    0    0    0    0    0
## 7         1    0    0    0    NA    NA    0    0    NA    0    0    0    1
## 8         0    0    0    0    0    0    0    0    0    0    0    1    0
## 9         NA    0    1    1    0    NA    NA    1    NA    0    1    0    0
## 10        NA    0    1    1    NA    NA    NA    0    0    1    0    0    0
## 11        NA    0    1    1    0    NA    NA    0    NA    0    1    0    0
## 12         0    0    0    0    0    0    0    0    0    0    0    0    0
## 13         0    0    0    0    0    0    0    0    0    0    NA    0    0
##      HistoneH3 DNA1 DNA2 CD38 HLADR MPO CD16 CD15 Ido1 LAG3 PD1 PDGFRb PDL1 Ki67
## 1             1    1    1    NA    1    0    0    0    NA    NA    NA    NA    NA
## 2             1    1    1    0    0    1    0    1    NA    0    NA    NA    NA
## 3             1    1    1    NA    1    0    0    0    NA    NA    NA    NA    NA
## 4             1    1    1    0    1    0    0    0    NA    0    NA    NA    NA
## 5             1    1    1    NA    0    0    NA    0    NA    NA    NA    NA    NA
```

```
## 6      1      1      1      1      0      0      0      0      NA      NA      NA      NA      0      0
## 7      1      1      1      NA      1      0      0      0      NA      NA      NA      NA      NA      NA
## 8      1      1      1      0      0      0      0      NA      NA      0      0      NA      NA      NA
## 9      1      1      1      0      NA      0      0      0      NA      NA      NA      0      NA      NA
## 10     1      1      1      0      NA      0      0      0      NA      NA      NA      NA      NA      NA
## 11     1      1      1      0      NA      0      0      0      NA      NA      NA      NA      NA      NA
## 12     1      1      1      0      NA      0      0      0      NA      0      0      NA      NA      NA
## 13     1      1      1      0      1      0      NA      0      NA      0      0      NA      NA      NA
##      cleavedPARP B2M CarbonicAnhydrase VISTA
## 1      NA      NA      NA      NA
## 2      NA      NA      NA      NA
## 3      NA      NA      NA      NA
## 4      NA      NA      NA      NA
## 5      NA      NA      NA      NA
## 6      NA      NA      NA      NA
## 7      NA      NA      NA      NA
## 8      NA      NA      NA      NA
## 9      NA      NA      NA      NA
## 10     NA      NA      NA      NA
## 11     NA      NA      NA      NA
## 12     NA      NA      NA      NA
## 13     NA      NA      NA      NA
```

Celesta output:

```
celesta_out_1 <- paste(tail(strsplit(strsplit(img_path_1, "[.]")[[1]][1], "[/"])[[1]], n=1), "final_cell")
print(celesta_out_1)
```

```
## [1] "IMMUcan_Batch20220908_S-220729-00002_002_roi_2_final_cell_type_assignment.csv"
celesta_out_1 <- read.csv(celesta_out_1)
head(celesta_out_1)
```

```
##      Round.1 Cell.type.number Final.cell.type      X      Y      CD68
## 1 Unknown      0      Unknown 375.72727 0.4545455 1.913765
## 2 Unknown      0      Unknown 436.81579 1.1842105 2.171190
## 3 Unknown      0      Unknown 273.63158 1.4210526 2.043997
## 4 Unknown      0      Unknown 310.06000 1.8000000 1.621160
## 5 Unknown      0      Unknown 321.73913 1.7173913 2.454538
## 6 Unknown      0      Unknown 67.71875 1.7812500 1.666919
##      CD163      CD14      CD11c      CD206      CD33      CD40      SMA      CD27
## 1 0.2596372 2.873650 1.231679 0.8633644 2.941737 1.91863429 2.6320329 2.891771
## 2 0.3680121 2.643814 1.769263 0.6342356 2.516529 0.34836876 1.0035133 1.572870
## 3 0.3716634 2.199514 1.064684 0.8642605 1.716621 0.02645026 0.0000000 1.472907
## 4 0.1509251 2.843119 1.547418 0.7051552 2.327627 0.38148938 0.5601294 1.357708
## 5 0.5583428 2.685838 1.992943 0.5332607 2.537487 0.68353399 0.4934767 1.101725
## 6 0.2402320 2.278385 1.590739 0.4693274 1.271210 0.24118241 0.5936743 2.075888
##      CD45RA      CD20      CD3      CD7      GrzB      TCF7      CD45RO      FOXP3
## 1 1.1099978 0.7276653 3.4646132 2.9367051 1.886247 0.833374 3.521541 3.1531404
## 2 1.6180227 1.1461222 1.8097367 1.2383724 2.155987 1.551714 3.384255 2.0378247
## 3 1.5856688 1.8548379 1.2291504 0.7636131 1.257550 1.347632 3.462684 1.3905660
## 4 1.4476364 0.9827928 1.1008167 1.1509480 2.003261 1.540989 3.363497 1.3865457
## 5 1.6107307 1.1938577 1.0271608 0.6437306 1.917804 1.312343 2.908856 1.7946530
## 6 0.8307044 0.3667732 0.9189267 0.3939230 1.089349 1.020104 3.110873 0.6393201
##      ICOS      CD8a      CD4      Ecad      CD303      HistoneH3      DNA1      DNA2
## 1 3.0454243 1.2010700 2.9307795 2.011174 1.132185 4.662441 4.694400 5.183200
```



```
## 2 1.4688487 1.2394218 1.7879023 4.404330 1.651497 4.858639 5.465577 5.960342
## 3 0.7929060 0.5338481 0.7709894 3.517929 1.287079 4.879359 5.731619 6.289294
## 4 1.3136431 1.3509379 1.7379908 4.631852 1.574281 4.402132 5.546083 6.120431
## 5 1.5142858 1.5024045 1.6904741 4.157759 1.697365 4.416976 4.803598 5.307182
## 6 0.5785684 0.4271875 1.1026120 1.770104 1.149303 4.568509 5.101809 5.660453
##      CD38      HLADR      MP0      CD16      CD15      Ido1      LAG3      PD1
## 1 1.5800046 4.460672 0.2694547 1.548095 0.2815353 0.8853757 1.1922724 0.8721140
## 2 1.2254055 2.953500 0.4821959 1.812707 2.1547884 1.5919059 0.9511183 1.0540736
## 3 0.4876684 2.068147 0.4194118 1.791760 0.5714060 1.0672473 0.8773543 0.8836842
## 4 1.4323210 2.588182 0.5360185 1.138565 2.0096698 1.3650783 0.6439501 1.0193657
## 5 1.2463916 2.900253 0.4895882 1.949723 1.2985587 1.4380467 0.8919202 0.6460546
## 6 2.2293396 2.635072 0.4363070 1.883864 0.6641002 0.7765715 0.4348450 0.4675359
##      PDGFRb      PDL1      Ki67      cleavedPARP      B2M      CarbonicAnhydrase
## 1 2.1553708 0.9162011 2.9559447 0.1369803 3.609471 1.252276
## 2 0.8801272 0.6356995 0.7003455 0.5229714 3.050585 1.938937
## 3 0.9833068 0.5541773 0.7209461 0.3095616 2.611706 2.164138
## 4 0.7334459 0.7210392 0.6176802 0.7292117 3.186465 2.130194
## 5 0.8965111 0.6663735 1.5192377 0.3180405 3.193999 1.872949
## 6 1.5464852 0.3501807 0.2725620 0.1902399 2.747736 1.112212
##      VISTA
## 1 2.2257712
## 2 0.8968787
## 3 0.6305858
## 4 0.6693036
## 5 0.7665576
## 6 1.0388566
```

```
print("Assigned cells stats")
```

```
## [1] "Assigned cells stats"
```

```
table(celesta_out_1$Final.cell.type)
```

```
##
##      CD8      Neutrophil      plasma      Unknown
##      3          3          32         1982
```