Pseudotime

Ramiro

April 8, 2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Load library

```
library(monocle3,verbose = FALSE)
library(dplyr,verbose = FALSE)
```

Read the data

```
## Loads a sparse matrix RawCountsPseudotime
load("RawCountsPseudotime.rda")
dim(RawCountsPseudotime)
```

```
## [1] 20271  3127
```

Load cluster data

```
## Loads a factor variabele ClusterPseudotime containing cluster identities
load("ClusterPseudotime.rda")
## Change the name of variable to remind
DataClusters <- ClusterPseudotime

table(DataClusters)
```

```
## DataClusters
##        0.0        0.1        1.0        1.1       10.0       10.1       10.2
##        238        213        410         30          0          0         34
##   11.0_1_2       11.3       11.4       12.0       12.1       13.0       14.0
##          0          0          0          0          0          0          0
##       16.0       16.1       17.0       17.1       19.0       19.1        2.0
##        160         42        169         36          0          0        383
##       20.0       22.0       22.1     22.2_3       22.4       26.0       26.1
##          0          0          0          0          0          0          0
```

```
##       26.2     27.0_3       27.1       27.2      27.NA 28.0_2_3_4       28.1
##          0          0          0          0          0          0          0
##       29.0       29.1        3.0        3.1       30.0       31.0       33.0
##          0          0        253        119          0          0         90
##       34.0       35.0       35.1       36.0       37.0       37.1       38.0
##          0          0          0          0          0          0          0
##       39.0        4.0        4.1       40.0       41.0       42.0       42.1
##          0        206        142          0          0         40         27
##       43.0       44.0       45.0       46.0       48.0       49.0        5.0
##          0          0          0          0          0          0          0
##        5.1       50.0       51.0       51.1       52.0       53.0       54.0
##          0          0          0          0          0          0          0
##       55.0       56.0       57.0       59.0        6.0       60.0       61.0
##          0         43          0          0          0          0          0
##       62.0       63.0        7.0        8.0        8.1        8.2        8.3
##          0          0          0          0         77         27         15
##     10.0.0   10.0.1_4   10.0.2_3     10.0.5     10.1.0     10.1.1      8.0.0
##         45         59         57          9         33         21        103
##      8.0.1
##         46
```

```
length(DataClusters)
```

```
## [1] 3127
```

Convert DataClusters to a matrix format for input to Monocle

```
DataCluster.ID <- matrix(as.numeric(levels(DataClusters))[DataClusters], ncol = 1)
```

```
## Warning in matrix(as.numeric(levels(DataClusters))[DataClusters], ncol = 1): NAs
## introduced by coercion
```

```
rownames(DataCluster.ID) <- names(DataClusters)
colnames(DataCluster.ID) <- "Cluster.IDs"
DataCluster.ID[1:10,]
```

```
## cele-001-008.GATCAGTCAT cele-001-027.ACTCCGCCAA cele-001-042.TTCCTAGACC
##                       0                       0                       0
## cele-001-046.TTCTACGCCA cele-001-047.TTCGCTGCCT cele-001-047.ATGGAAGCAT
##                       0                       0                       0
## cele-001-064.AAGCTGACCT cele-001-065.GCCATCAACT cele-001-068.ACGGCAACCA
##                       0                       0                       0
## cele-001-071.GTCATTGCGC
##                       0
```

Generate matrix of gene short names for Monocle

```
geneNames <- matrix(rownames(RawCountsPseudotime), ncol = 1)
rownames(geneNames) <- rownames(RawCountsPseudotime)
colnames(geneNames) <- "gene_short_name"
head(geneNames)
```

```
##       gene_short_name
```

```
## aap-1 "aap-1"
## aat-1 "aat-1"
## aat-2 "aat-2"
## aat-3 "aat-3"
## aat-4 "aat-4"
## aat-5 "aat-5"
```

Initiate Monocle object

```
cds <- new_cell_data_set(expression_data = RawCountsPseudotime,
                         cell_metadata = DataCluster.ID,
                         gene_metadata = geneNames)


cds <- cds[,names(ClusterPseudotime[ClusterPseudotime %in% c("3.0","3.1")])]


colData(cds)
```

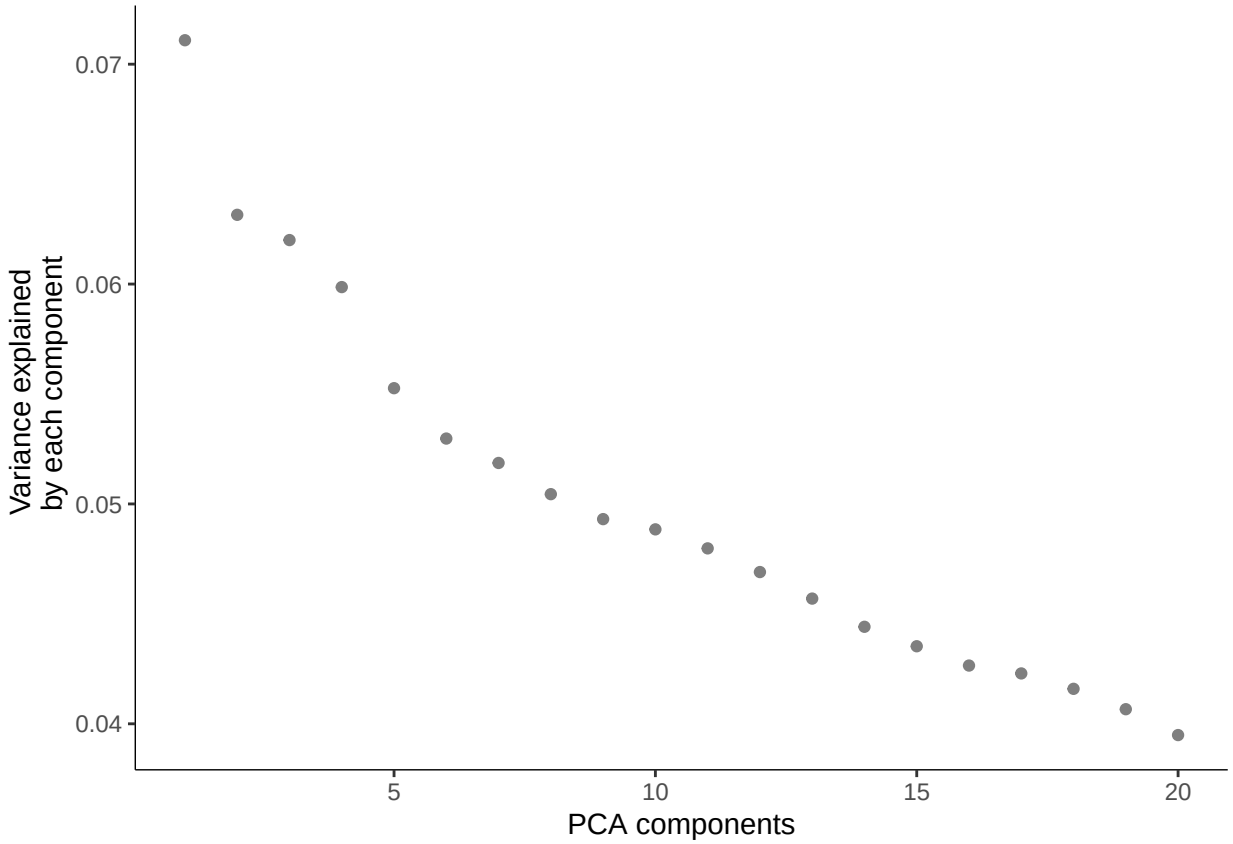
```
## DataFrame with 372 rows and 2 columns
##                          Cluster.IDs       Size_Factor
##                            <numeric>         <numeric>
## cele-001-002.ACGACCAATA          3 0.512985911598468
## cele-001-003.GGATTCTATC          3 0.625482822036728
## cele-001-003.TCCAGAAGGT          3 0.764978990980171
## cele-001-013.GATATGGTCT          3 0.787478373067823
## cele-001-030.CGTCCGTCCT          3 0.503986158763407
## ...                            ...               ...
## cele-010-072.CCTGACGTTC        3.1  1.50745859987269
## cele-010-081.AAGTCTTCCG        3.1  1.69645340940897
## cele-010-088.TGCCAGATGG        3.1  2.12394166907436
## cele-010-090.CAGGCGCCAT        3.1  2.78542350245133
## cele-010-092.GAAGTCCGTC        3.1  1.16996786855791
```

```
## Column Cluster.IDs contains the original DataCluster IDs
colData(cds)$Cluster.IDs <- factor(colData(cds)$Cluster.IDs)
colData(cds)
```

```
## DataFrame with 372 rows and 2 columns
##                          Cluster.IDs       Size_Factor
##                             <factor>         <numeric>
## cele-001-002.ACGACCAATA          3 0.512985911598468
## cele-001-003.GGATTCTATC          3 0.625482822036728
## cele-001-003.TCCAGAAGGT          3 0.764978990980171
## cele-001-013.GATATGGTCT          3 0.787478373067823
## cele-001-030.CGTCCGTCCT          3 0.503986158763407
## ...                            ...               ...
## cele-010-072.CCTGACGTTC        3.1  1.50745859987269
## cele-010-081.AAGTCTTCCG        3.1  1.69645340940897
## cele-010-088.TGCCAGATGG        3.1  2.12394166907436
## cele-010-090.CAGGCGCCAT        3.1  2.78542350245133
## cele-010-092.GAAGTCCGTC        3.1  1.16996786855791
```

Step 1: Normalize and pre-process the data

```
cds <- preprocess_cds(cds, num_dim = 20)
plot_pc_variance_explained(cds)
```



Step 2: Reduce the dimensions using UMAP

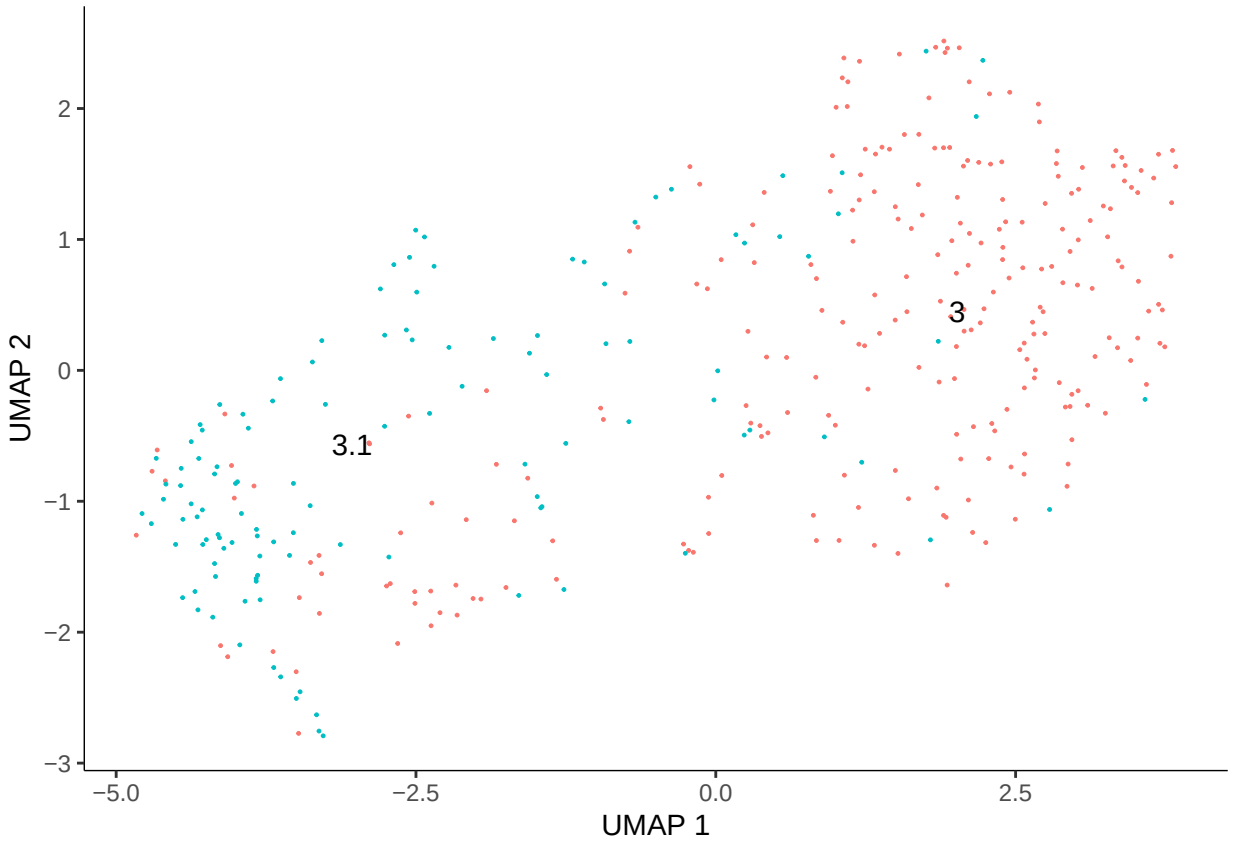
```
cds <- reduce_dimension(cds,umap.min_dist = 0.1,cores = 8)
```

```
## No preprocess_method specified, using preprocess_method = 'PCA'
```

```
## Note: reduce_dimension will produce slightly different output each time you run it unless you set 'u
```

```
plot_cells(cds, color_cells_by = "Cluster.IDs", group_label_size = 4, cell_size = 0.5)
```

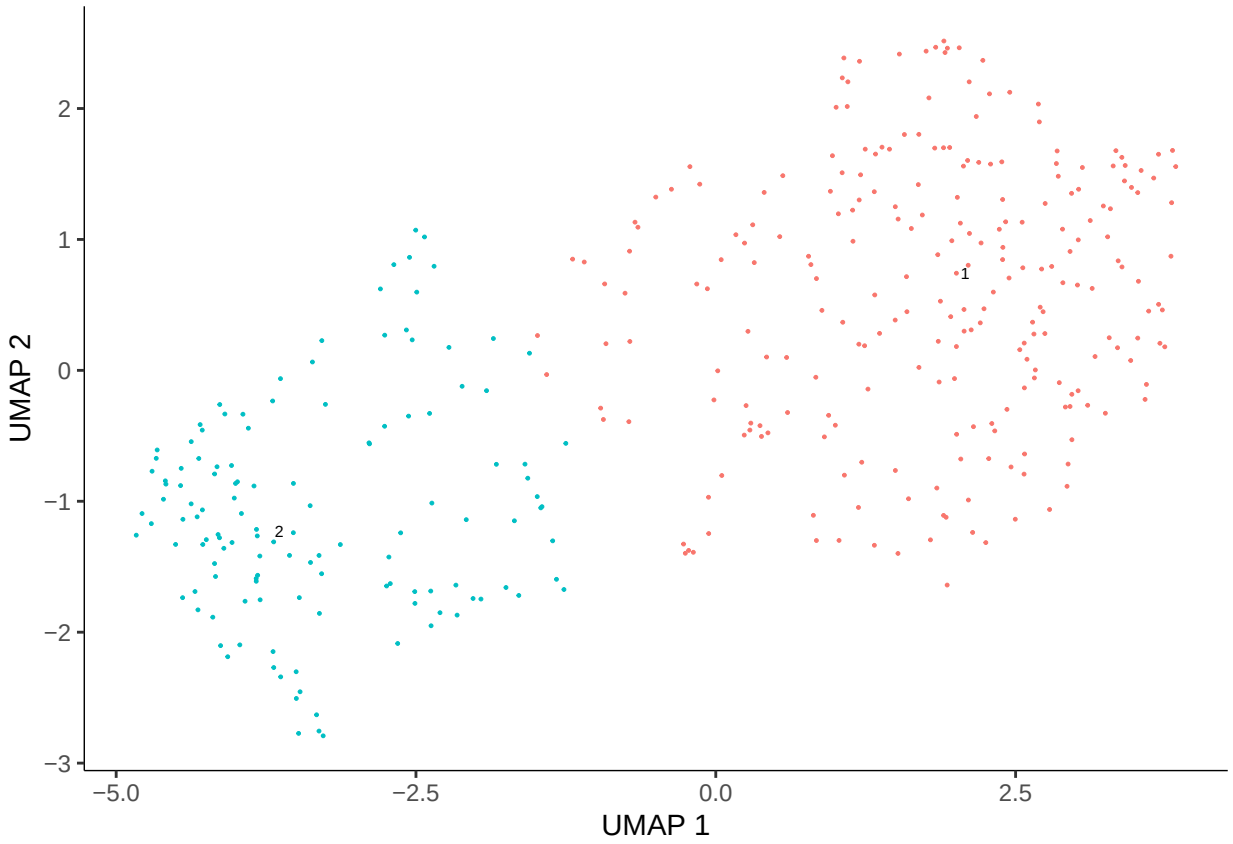
```
## No trajectory to plot. Has learn_graph() been called yet?
```



##Step 3: Cluster the cells

```
# cds = cluster_cells(cds, resolution=0.02)
cds = cluster_cells(cds, resolution=0.01)
plot_cells(cds, cell_size = 0.5)
```

```
## No trajectory to plot. Has learn_graph() been called yet?
```



Step 4: Learn a graph and order cells

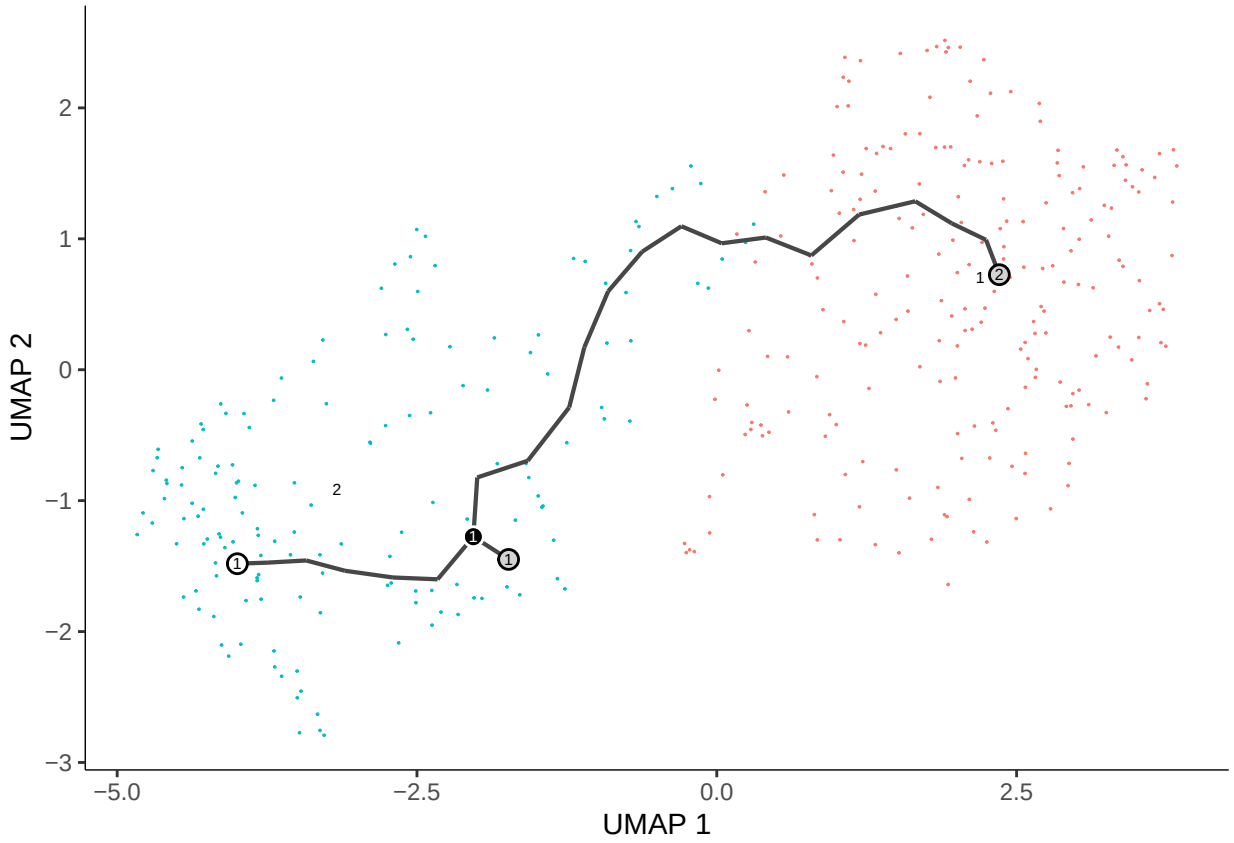
```
cds <- learn_graph(cds)
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
## With Shiny
# cds <- order_cells(cds)
# #
# save(file = "cdsRoots.rda", cds) #Save object state after selection of roots

## Without Shiny
load("cdsRoots.rda") #Load object

## Get info by: cds@principal_graph_aux$UMAP$root_pr_nodes

# cds <- order_cells(cds, root_pr_nodes = c("Y_8", "Y_9", "Y_12", "Y_14", "Y_15", "Y_31", "Y_34", "Y_43
# cds <- order_cells(cds, root_pr_nodes = c("Y_4",   "Y_17",  "Y_30",  "Y_33",  "Y_46",  "Y_84",  "Y_12

print(cds@principal_graph_aux$UMAP$root_pr_nodes)
```

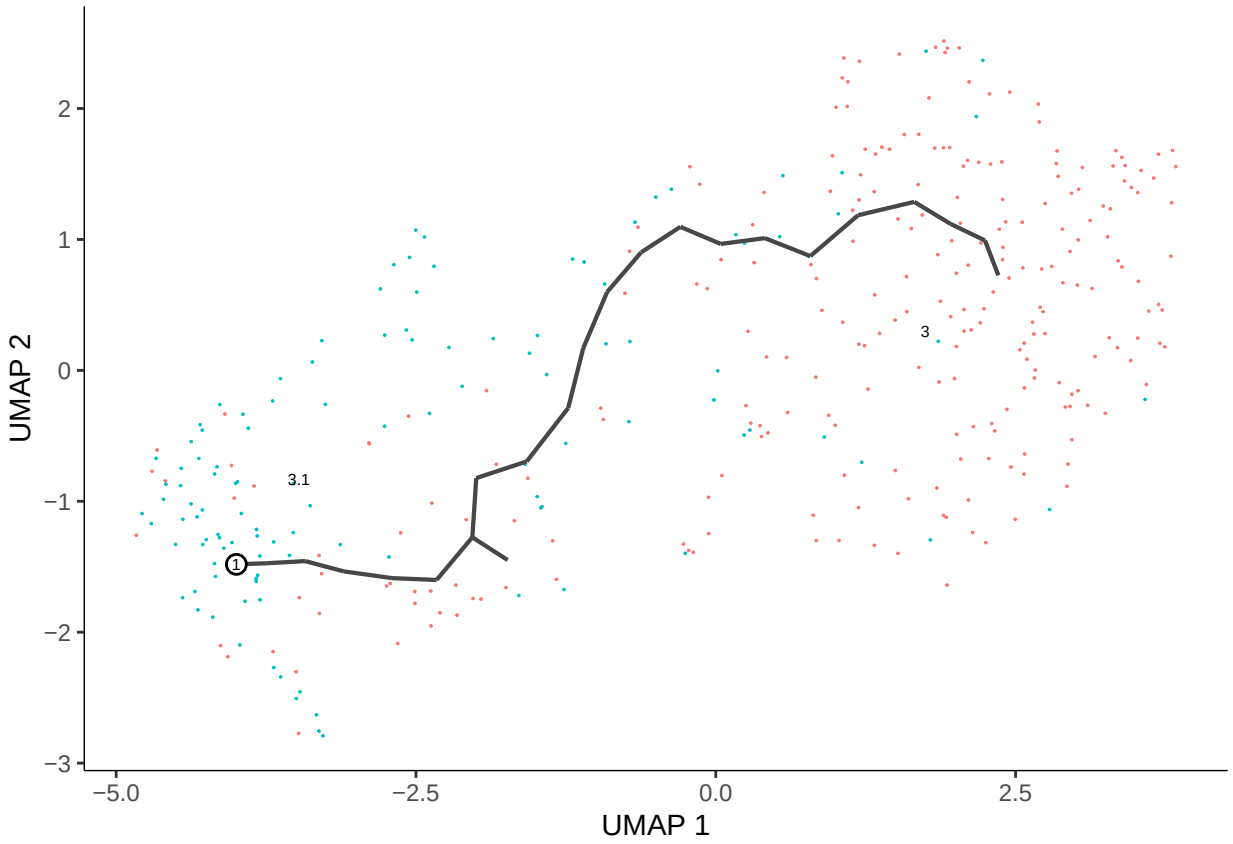
```
## [1] "Y_22"
```

```
plot_cells(cds)
```



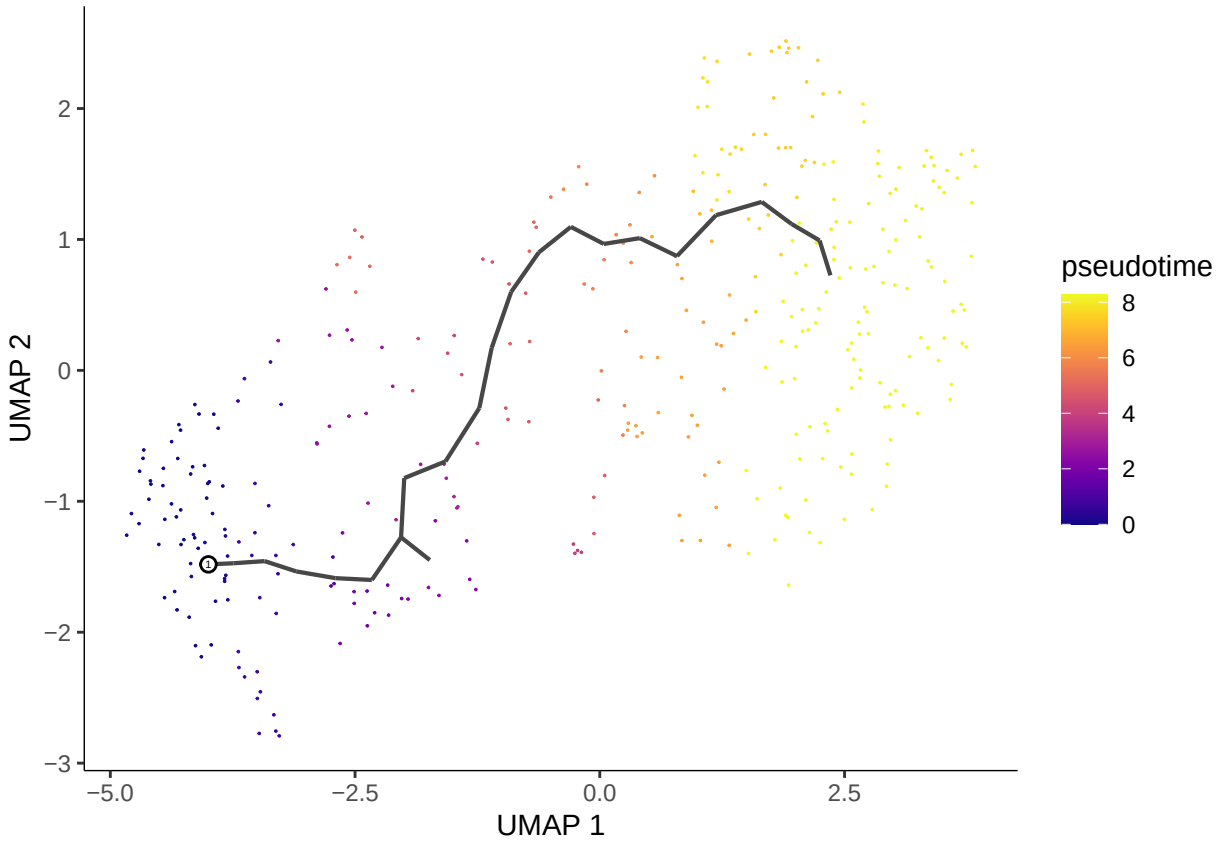
##Plot the pseudotime graph

```
plot_cells(cds,
           color_cells_by = "Cluster.IDs",
           label_groups_by_cluster=FALSE,
           label_leaves=FALSE,
           label_branch_points=FALSE)
```



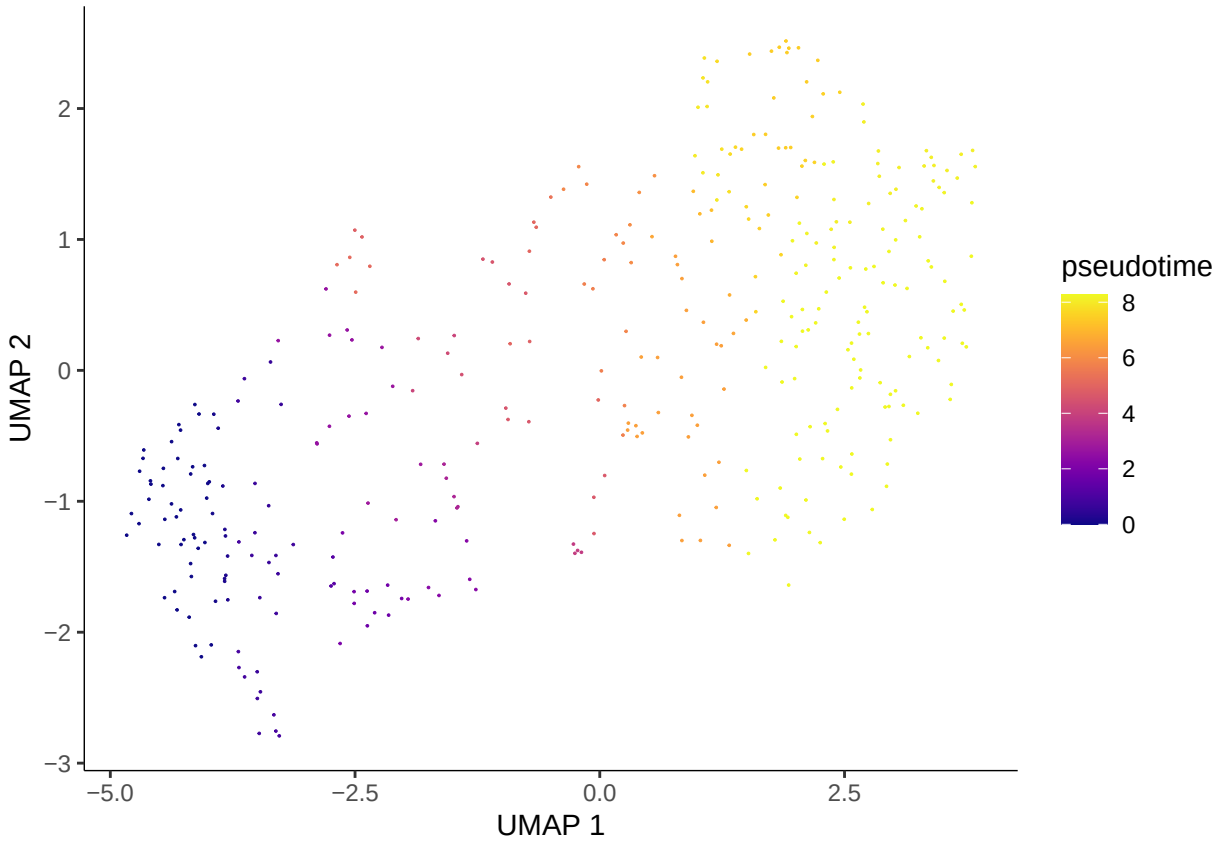
##Plot the pseudotime graph

```
plot_cells(cds,
           color_cells_by = "pseudotime",
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           label_branch_points=FALSE,
           graph_label_size=1.5)
```



##Plot the pseudotime graph without trajectory

```
plot_cells(cds,
           color_cells_by = "pseudotime",
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           label_branch_points=FALSE,
           graph_label_size=1.5,
           show_trajectory_graph = FALSE)
```



Correlation Analysis

Cluster 3

```
cluster3.cellNames <- rownames(pData(cds))[pData(cds)$Cluster.IDs %in% c(3, 3.1)]

cds_3 <- cds[,cluster3.cellNames]

# cds_3 <- choose_cells(cds)

cds3_pg <- graph_test(cds_3, neighbor_graph="principal_graph", cores=4, verbose = F)
cds3_genes <- cds3_pg %>%
  filter(q_value < 0.05) %>%
  arrange(morans_I) %>%
  select(gene_short_name)
```

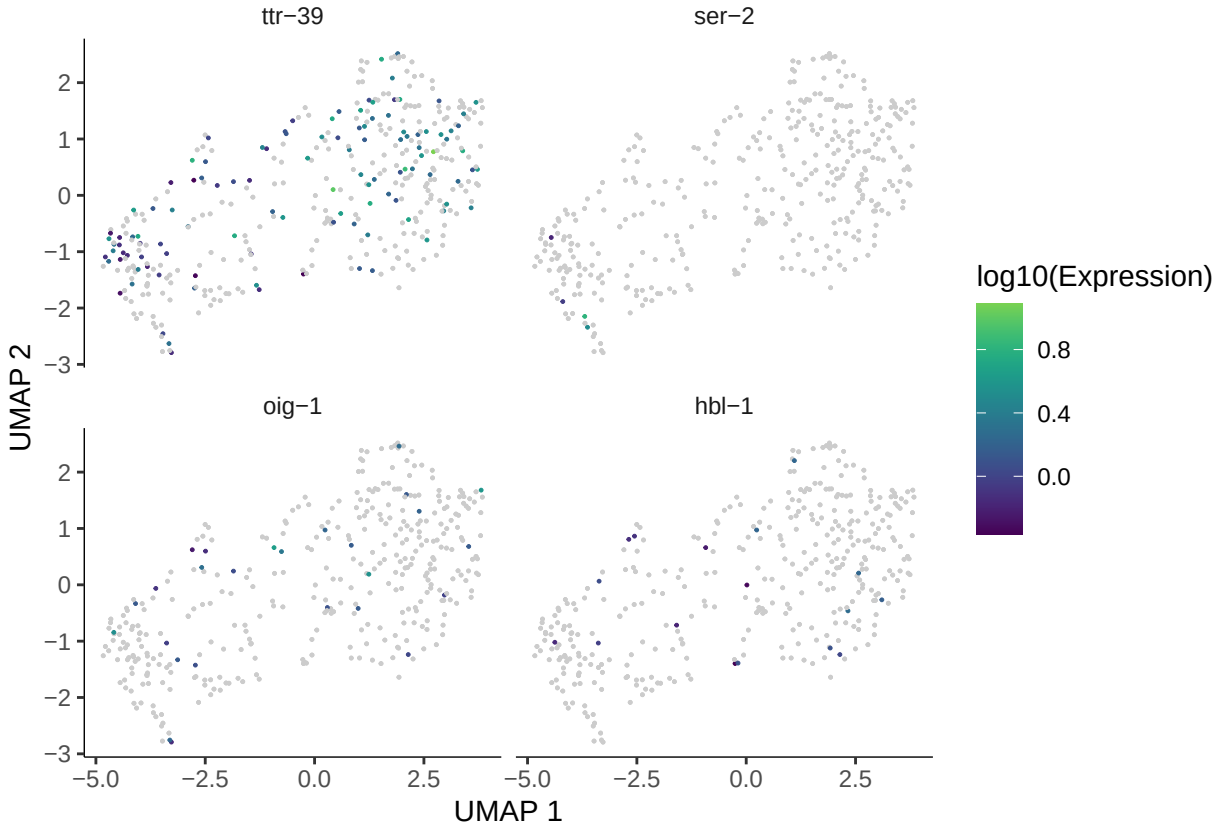
```
cds3_genes$gene_short_name
```

```
##   [1] T07A9.10   alh-11     inx-10     npp-3      stg-2      dgn-1
##   [7] B0035.15   dpt-1      Y105E8A.2  Y76B12C.3  ubc-22     ddx-35
##  [13] K03H1.11   hot-8      parg-1     lin-61     F01E11.18  R10E4.1
##  [19] lmtr-3     T26C11.4   F56D1.2    nlp-6      K06H7.2    mes-2
##  [25] haf-2      W03D8.3    glrx-22    F46F11.8   agt-1      acd-3
##  [31] F54D5.12   C46H3.2    cebp-1     gly-20     pfd-6      srp-7
##  [37] mms-19     dhhc-8     nft-1      riok-3     gab-1      F40F11.3
```

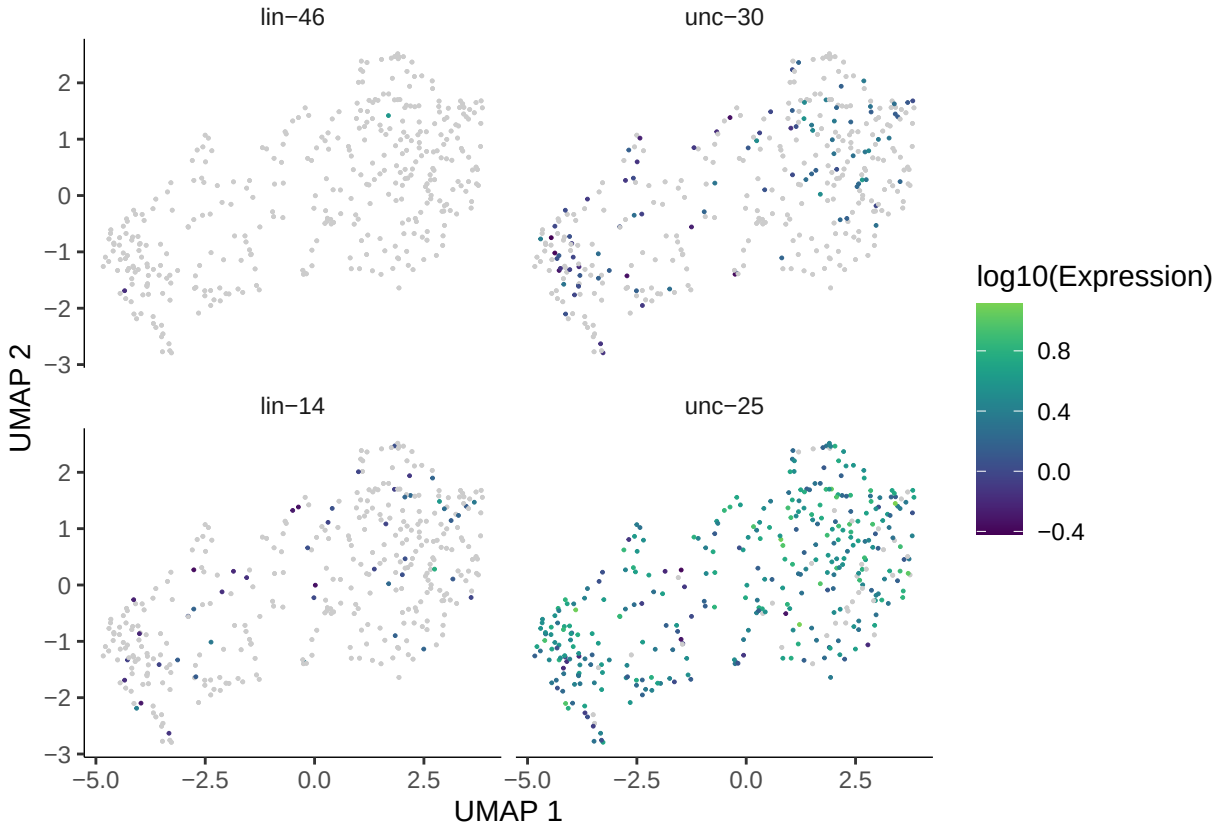
```
##  [43] F25F8.1    F54D5.7    col-17     atg-4.2    pcaf-1     nape-1
##  [49] slx-1      Y57A10A.26 sym-2      T10B11.7   W09C5.1    K05F1.5
##  [55] mup-4      lin-15A    pyc-1      acl-1      C53B4.2    R10E11.6
##  [61] Y66D12A.9  pfd-3      Y39A1A.24  T14B4.8    Y105E8A.13 gfi-2
##  [67] F13B12.6   nuo-1      B0416.5    K07C5.6    C18A3.2    sqv-8
##  [73] snap-1     spat-2     lin-66     rps-27     Y105E8A.11 ribo-1
##  [79] hsp-3      mev-1      pan-1      gbb-2      rpl-7      C24A3.2
##  [85] lin-39     dip-2      xbp-1      Y54F10AM.5 cyn-16     F21F3.6
##  [91] rps-16     glb-25     pam-1      F31C3.3    nra-1      ekl-5
##  [97] rpl-39     rpl-11.2   tag-80     Y71H2AM.5  lmn-1      gdi-1
## [103] klp-7      unc-115    sacy-1     ZC395.10   dnj-5      hmr-1
## [109] ser-4      F32B4.5    des-2      rbm-28     cul-1      chc-1
## [115] lin-10     inf-1      F46F11.10  F52C9.3    hsp-1      let-716
## [121] B0252.8    psd-1      set-15     gars-1     cnnm-1     atp-5
## [127] rpl-5      ceh-44     sur-6      C08B6.8    Y53F4B.18  ZK1307.8
## [133] ZK673.2    fkh-9      let-504    sas-4      fln-1      brp-1
## [139] C03A3.1    T04F8.2    fbxl-1     F14F9.5    eak-7      M03E7.1
## [145] odr-2      ftt-2      rpl-27     hmg-11     vha-1      ztf-13
## [151] top-2      imp-2      usp-46     let-70     tba-2      nlr-1
## [157] gopc-1     F29B9.11   lgc-49     his-35     exp-2      gop-3
## [163] rgef-1     F55C12.4   lin-53     elf-1      tag-196    dnj-18
## [169] nkb-1      emb-5      rpl-30     pab-1      ZC434.7    idhg-1
## [175] rpl-13     somi-1     aldo-1     cmd-1      R106.5     acr-14
## [181] nsf-1      rpl-26     T27C4.1    C09G1.5    ubl-1      trap-4
## [187] hex-2      rpt-6      ras-1      Y94H6A.8   mlc-4      lpd-7
## [193] C26D10.6   rps-25     F44E7.4    mrpl-24    lin-52     clr-1
## [199] K08F4.1    vig-1      Y119C1B.3  gon-1      rpl-16     rps-9
## [205] mthf-1     daf-21     lfe-2      tmc-1      ensa-1     F36A2.7
## [211] ncx-1      W09D10.1   ykt-6      rpl-34     inst-1     T20F5.7
## [217] Y105E8A.3  R02E12.5   smc-3      cyn-3      ama-1      madd-4
## [223] C25H3.11   rpl-6      W05B5.1    atp-2      nono-1     cct-8
## [229] sup-1      hipr-1     his-72     npp-7      cct-4      mig-22
## [235] rps-8      mel-11     Y66D12A.10 unc-13     K10B2.4    anc-1
## [241] frpr-4     rpl-3      F02A9.1    F28E10.1   mai-1      vha-14
## [247] F12F6.1    cct-1      brf-1      mma-1      mex-3      uig-1
## [253] rpy-1      C36B7.5    rps-0      rpl-43     lin-25     oaz-1
## [259] tbc-9      Y57G11C.9  flp-13     rpl-29     set-20     vha-13
## [265] vha-2      eif-3.E    Y97E10C.1  rpl-38     rps-14     rpl-25.2
## [271] prpf-4     C27A7.5    tbb-2      ser-2      rps-12     F40F8.1
## [277] F19G12.2   skr-1      spe-15     tin-44     eef-2      Y39B6A.36
## [283] nlp-9      Y55F3AM.5  unc-40     T25F10.3   pas-4      mom-5
## [289] ZC190.4    F26H9.2    R13A5.9    trap-1     unc-83     hse-5
## [295] unc-25     chdp-1     his-37     rpl-22     R12B2.6    tba-1
## [301] T19D12.6   rps-1      rpl-32     Y110A2AR.3 rpl-9      T10C6.6
## [307] rpl-41.2   W05E10.5   rpc-2      pcca-1     act-4      eef-1A.1
## [313] clic-1     rps-13     igcm-4     marc-4     iff-2      rps-11
## [319] rpl-10     rpn-12     gpb-1      patr-1     rla-1      K05B2.2
## [325] snrp-200   nlp-21     Y37E3.8    aqp-7      cdh-8      rsp-5
## [331] sir-2.2    C53H9.2    Y44A6D.2   far-1      rla-0      rpl-18
## [337] rpl-4      B0041.5    baf-1
## 20271 Levels: 2L52.1 2RSSE.1 4R79.2 6R55.2 aagr-1 aagr-2 aagr-3 aagr-4 ... zyx-1
```

```
# Plot a few genes
plot_cells(cds_3, genes= c("ttr-39","ser-2", "oig-1", "hbl-1"),
```

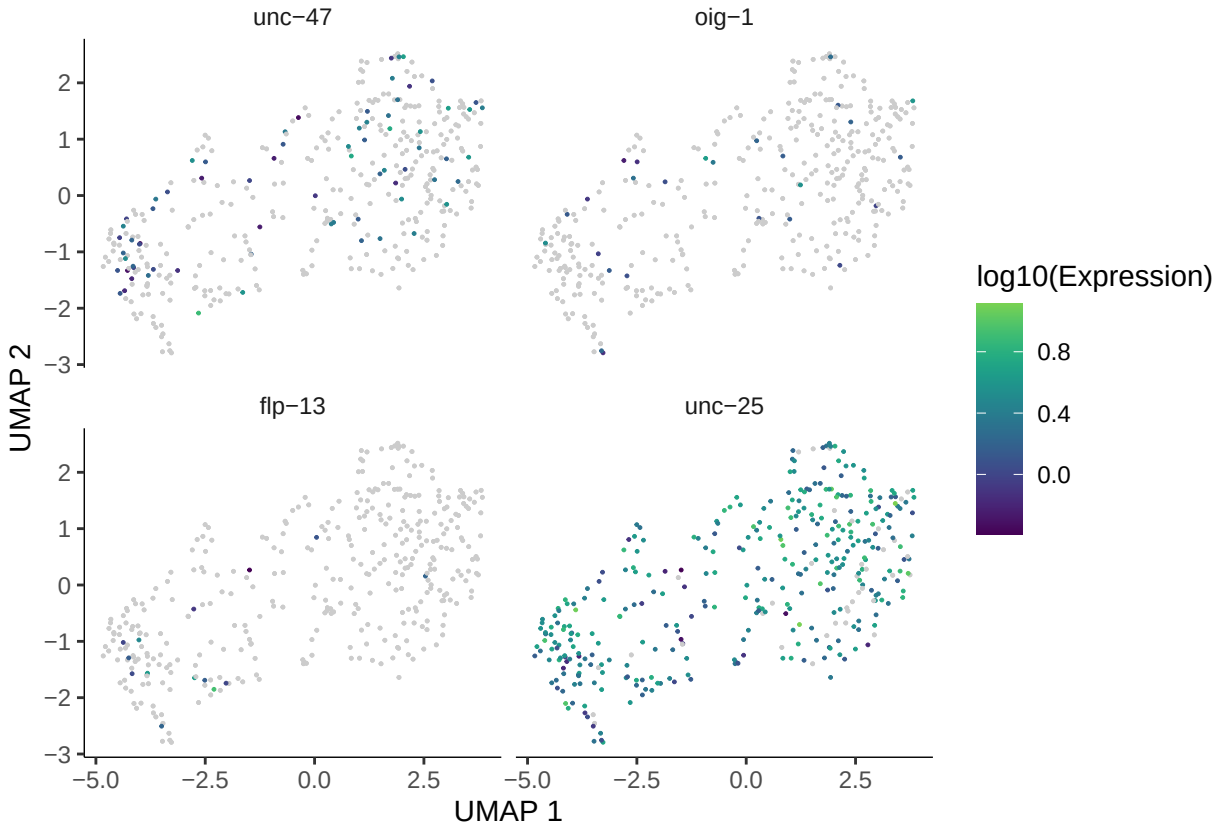
```
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = .5
          )
```



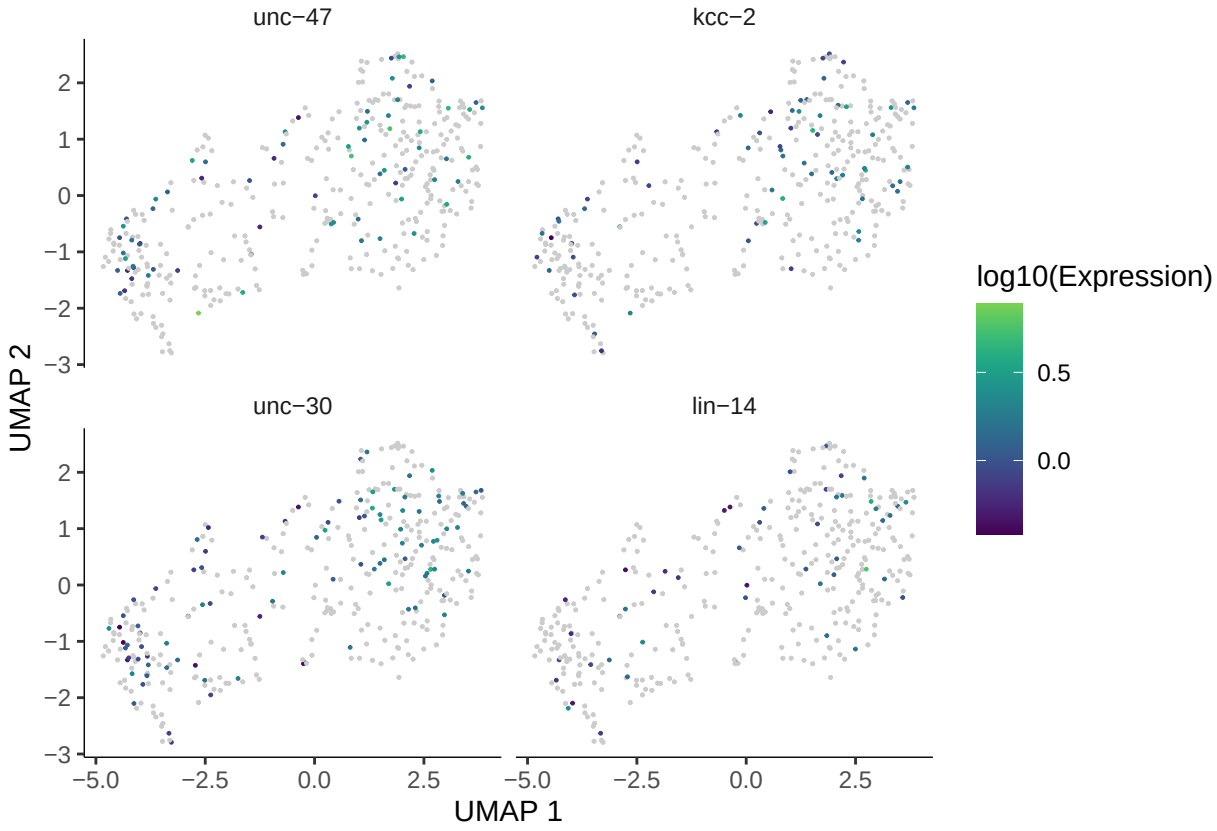

```
# Plot a few genes
plot_cells(cds_3, genes= c("lin-46", "unc-30", "lin-14","unc-25"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = .5
          )
```



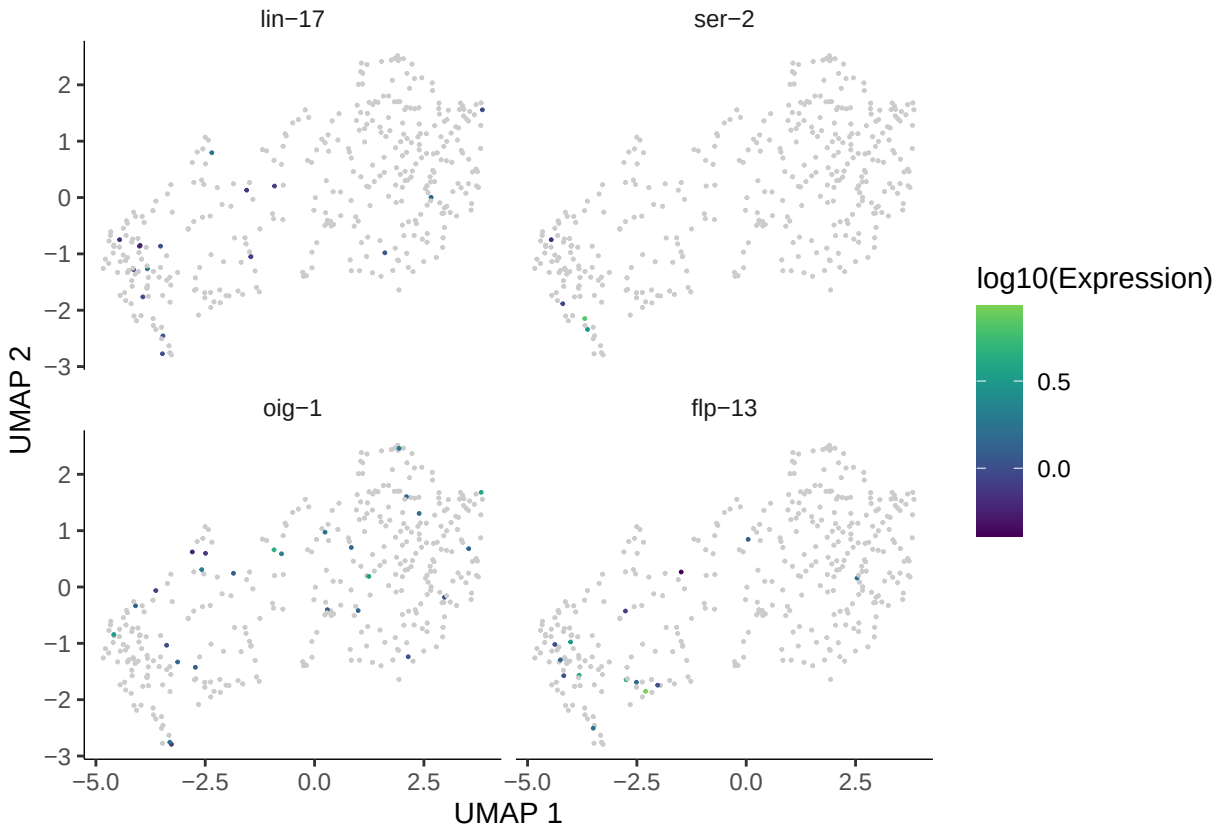

```
# Plot a few genes
plot_cells(cds_3, genes= c( "unc-47","oig-1", "flp-13", "unc-25"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = .5
          )
```



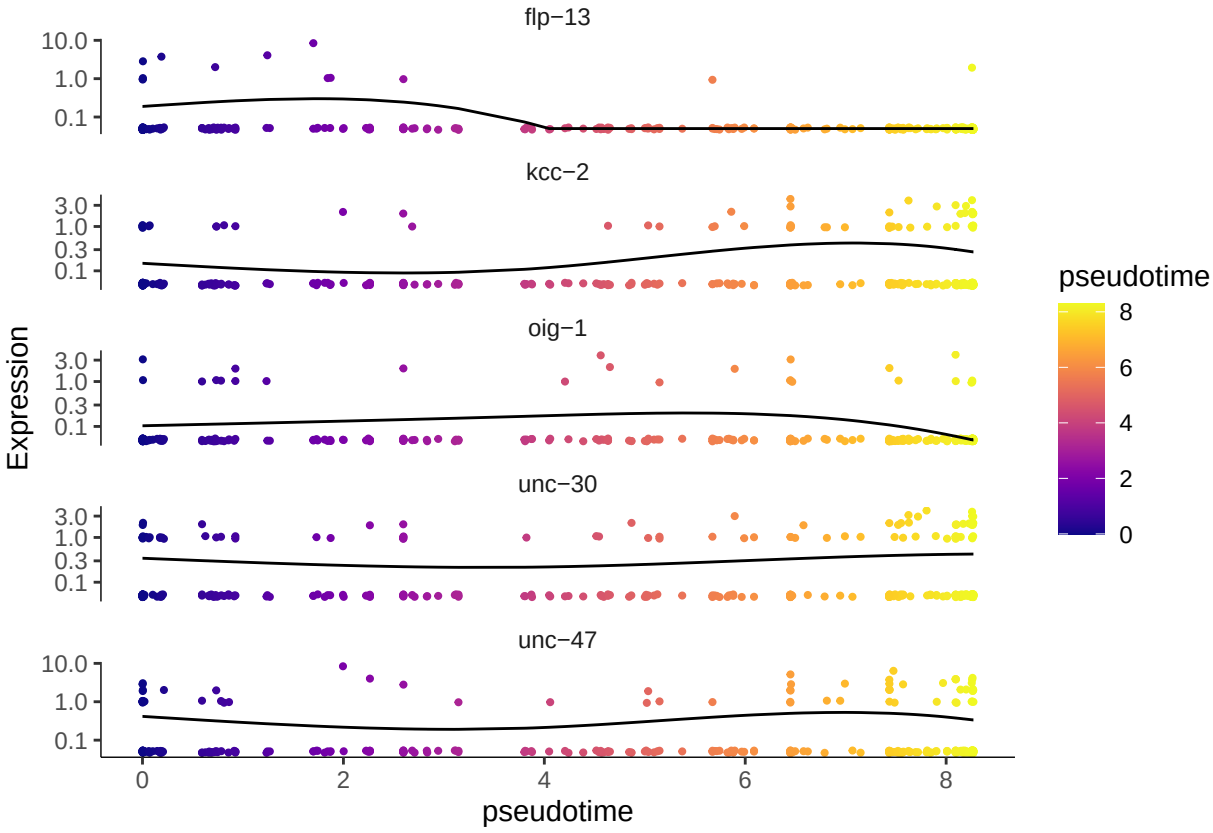

```
# Plot a few genes
plot_cells(cds_3, genes= c("unc-47", "kcc-2", "unc-30", "lin-14"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = .5
          )
```




```
# Plot a few genes
plot_cells(cds_3, genes= c("lin-17","ser-2", "oig-1", "flp-13"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = .5
          )
```




```
cds_3_lineage_cds <- cds_3[rowData(cds_3)$gene_short_name %in% c("unc-30", "unc-47", "kcc-2", "oig-1",
plot_genes_in_pseudotime(cds_3_lineage_cds,
                         # color_cells_by="embryo.time.bin",
                         min_expr=0.05)
```


flp−13
kcc−2
oig−1
unc−30
unc−47
Expression
10.0
1.0
0.1
3.0
1.0
0.3
0.1
0
2
4
6
8
pseudotime
pseudotime