Pseudotime for clusters 16.0 and 16.1

Ramiro Lorenzo

April 8, 2020

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
```

```r
length(DataClusters)
```

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

Convert DataClusters to a matrix format for input to Monocle

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

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

```r
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

```r
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

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


cds <- cds[,names(ClusterPseudotime[ClusterPseudotime %in% c("16.0","16.1")])]


colData(cds)
```

```
## DataFrame with 202 rows and 2 columns
##                            Cluster.IDs        Size_Factor
##                              <numeric>          <numeric>
## cele-001-009.ATCCGTTAGC             16  1.94844648879067
## cele-001-017.CTGAAGAGAC             16 0.575984181443893
## cele-001-017.CCATCGGACC             16  2.11044203982177
## cele-001-026.TCCGGTAATC             16  2.06994315206399
## cele-001-041.TTCAAGAATC             16  1.23296613840333
## ...                               ...                ...
## cele-010-037.GCAGCGGACT           16.1  4.10838716920527
## cele-010-043.ACTCGACGCC           16.1  4.64837233930892
## cele-010-053.TGCCTAACTT           16.1  5.83633971353695
## cele-010-065.CGCATCCATC           16.1  4.34688061933438
## cele-010-065.TGCGCGATGC           16.1  5.84983934278954
```

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

```
## DataFrame with 202 rows and 2 columns
##                            Cluster.IDs        Size_Factor
##                               <factor>          <numeric>
## cele-001-009.ATCCGTTAGC             16  1.94844648879067
## cele-001-017.CTGAAGAGAC             16 0.575984181443893
## cele-001-017.CCATCGGACC             16  2.11044203982177
## cele-001-026.TCCGGTAATC             16  2.06994315206399
## cele-001-041.TTCAAGAATC             16  1.23296613840333
## ...                               ...                ...
## cele-010-037.GCAGCGGACT           16.1  4.10838716920527
## cele-010-043.ACTCGACGCC           16.1  4.64837233930892
## cele-010-053.TGCCTAACTT           16.1  5.83633971353695
## cele-010-065.CGCATCCATC           16.1  4.34688061933438
## cele-010-065.TGCGCGATGC           16.1  5.84983934278954
```

Step 1: Normalize and pre-process the data

```r
cds <- preprocess_cds(cds, num_dim = 100)
plot_pc_variance_explained(cds)
```
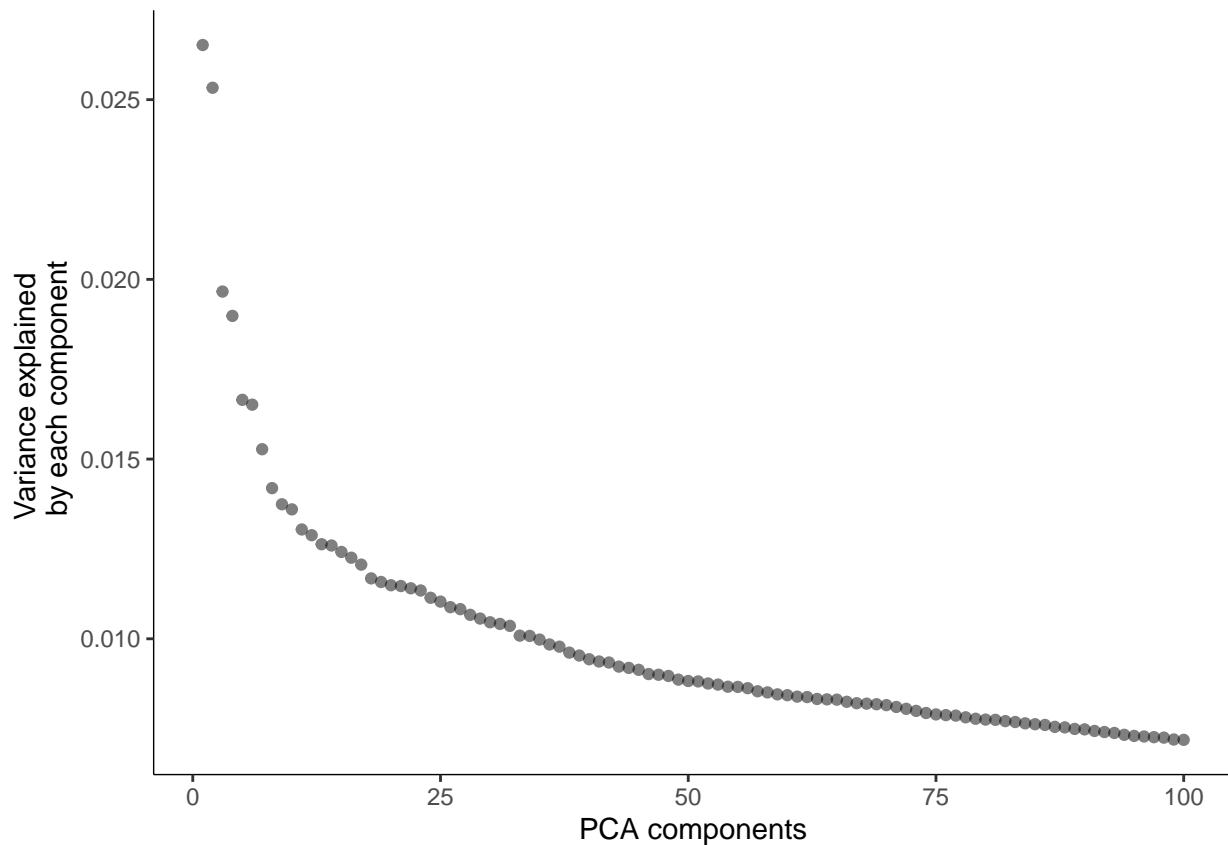
```
cds <- preprocess_cds(cds, num_dim = 20)
```

## 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 'un

```
plot_cells(cds, color_cells_by = "Cluster.IDs", group_label_size = 4, cell_size = 0.5)
```
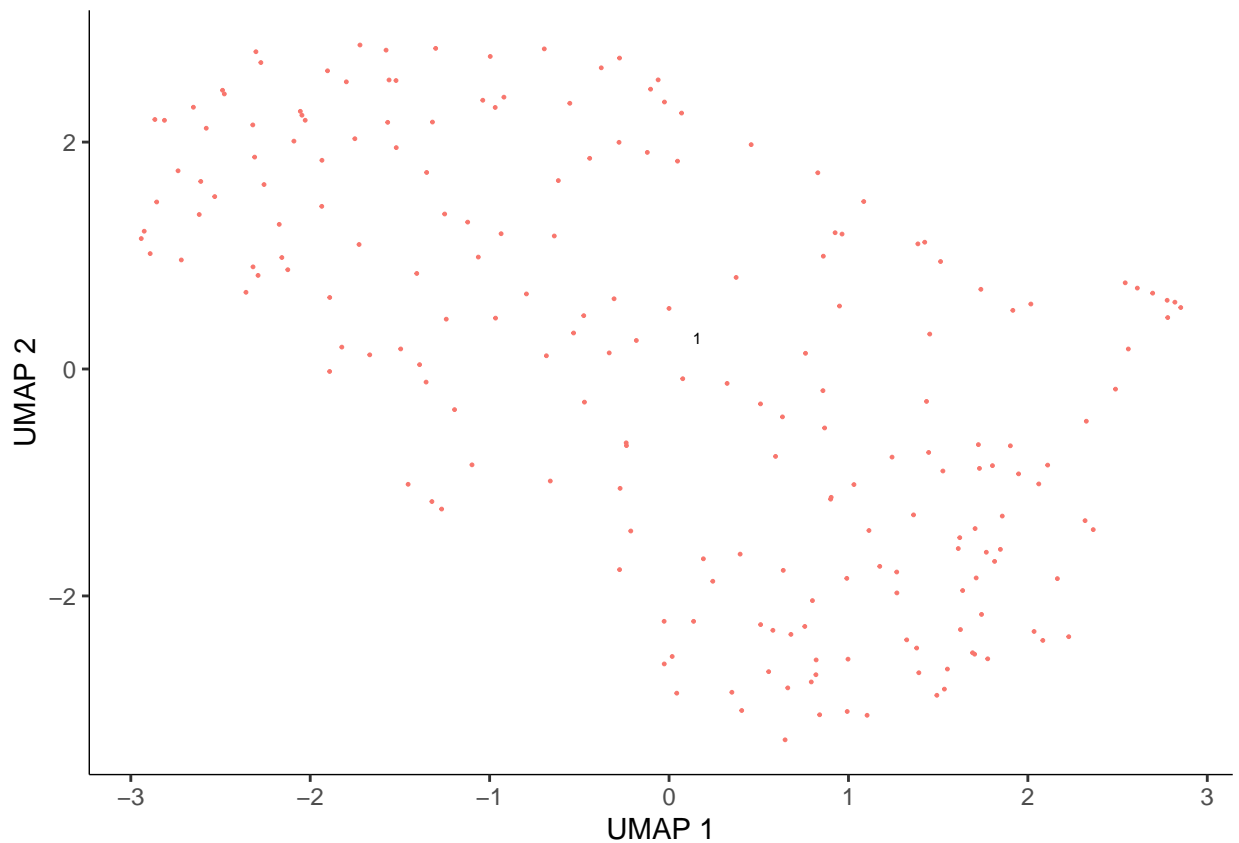
No trajectory to plot. Has learn_graph() been called yet?
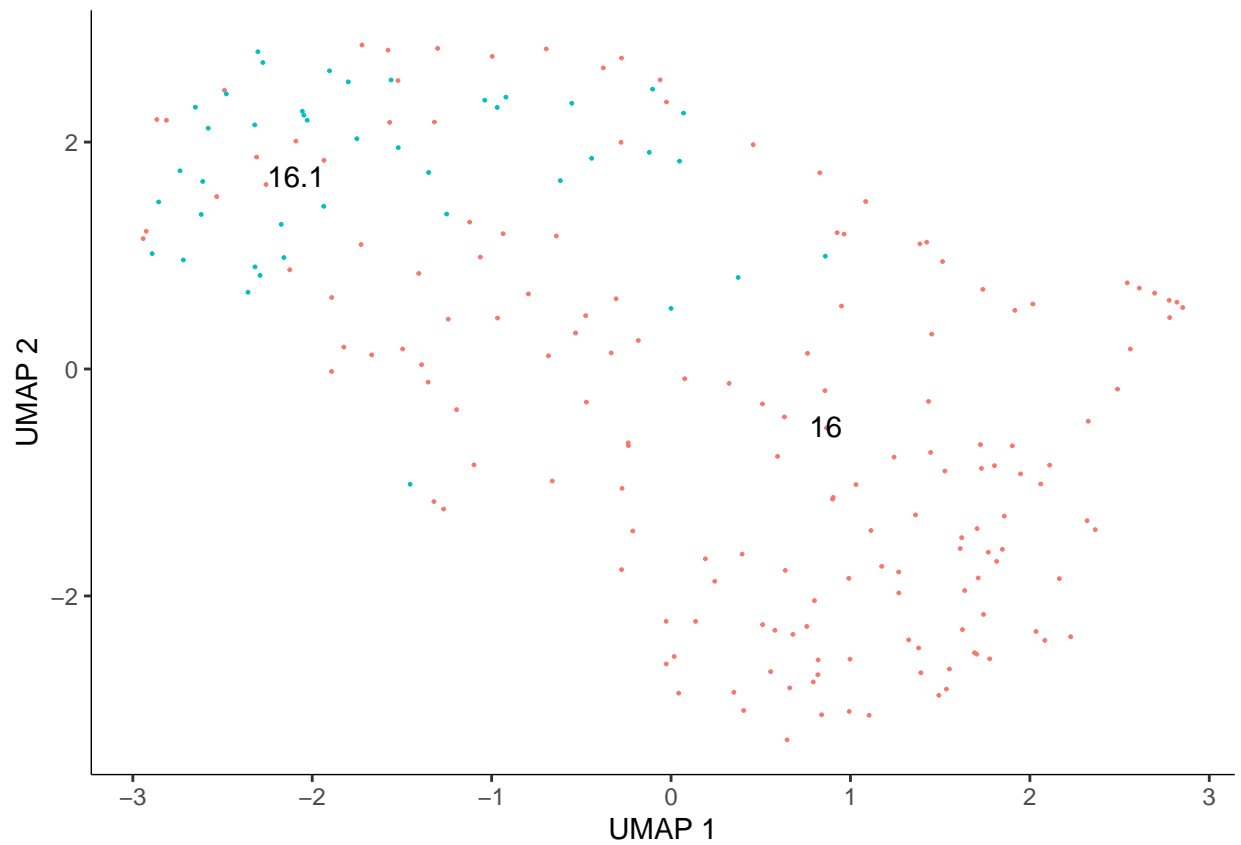
Step 3: Cluster the cells

```
# cds = cluster_cells(cds, resolution=0.02)
cds = cluster_cells(cds, resolution=0.02)
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)
```

```
## 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

print(cds@principal_graph_aux$UMAP$root_pr_nodes)
```
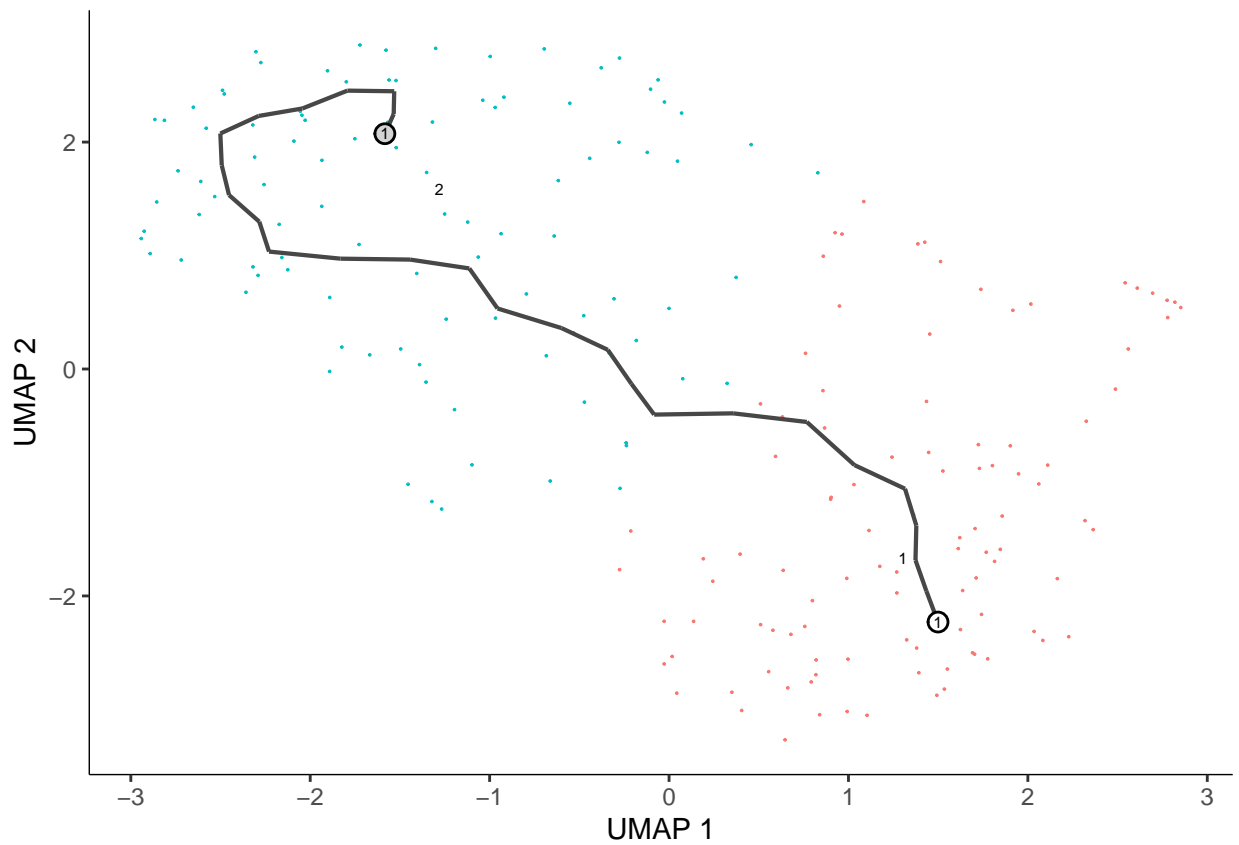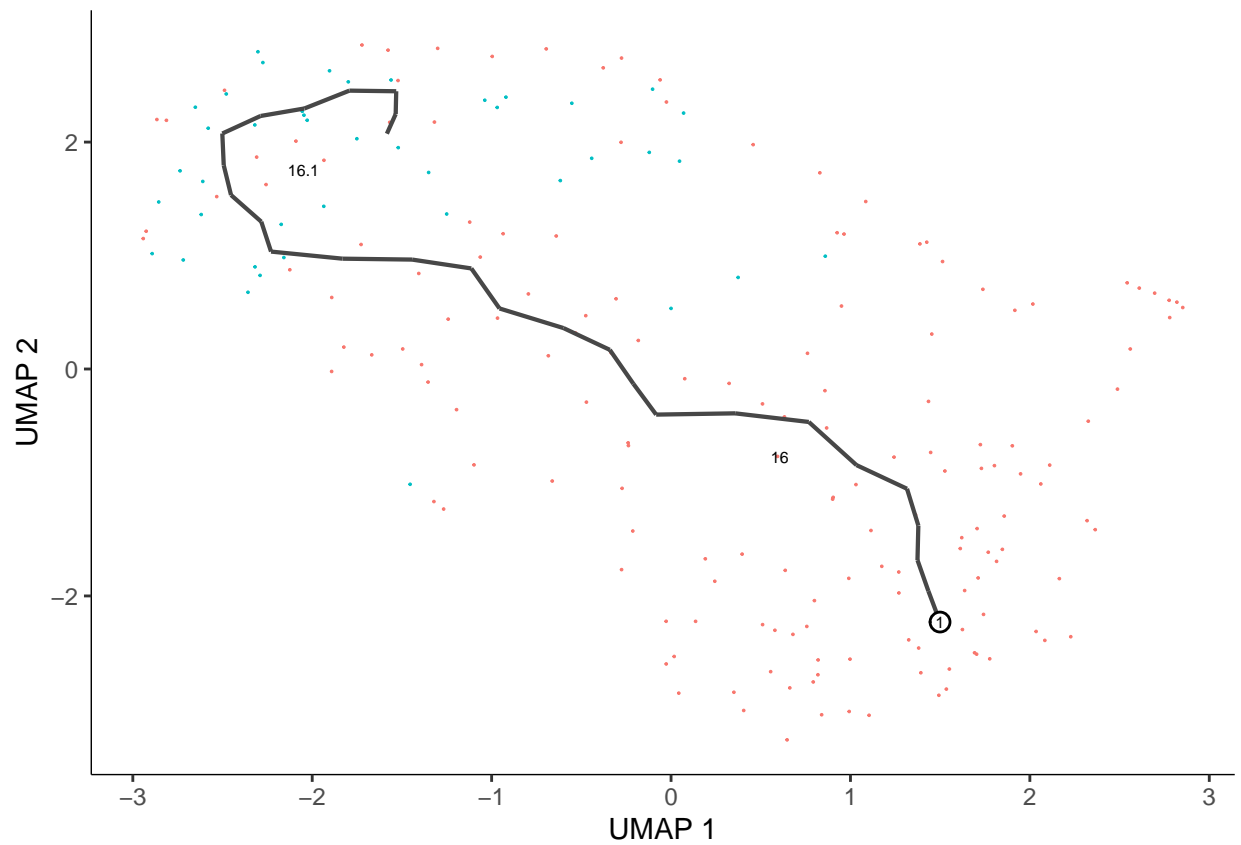
```
## [1] "Y_11"
```

```
plot_cells(cds)
```
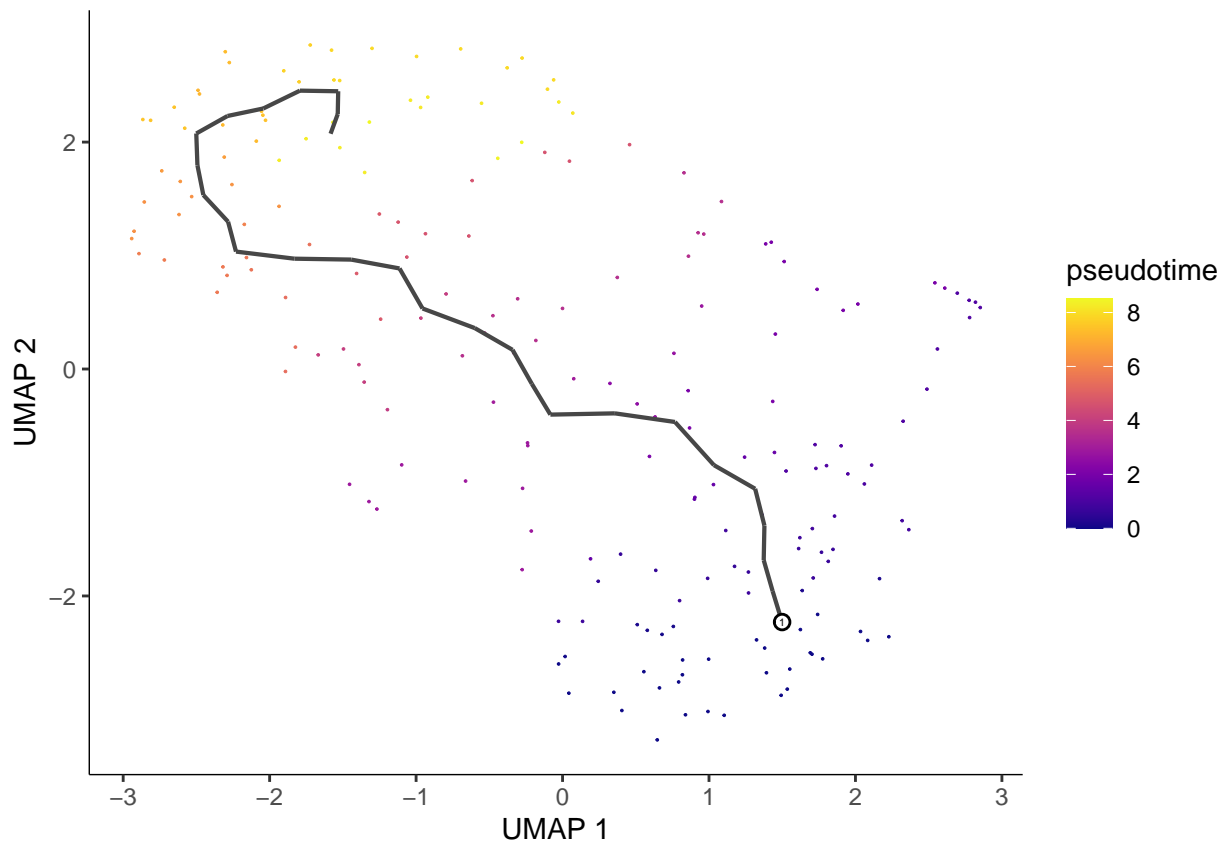
Plot the pseudotime graph

```
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)
```
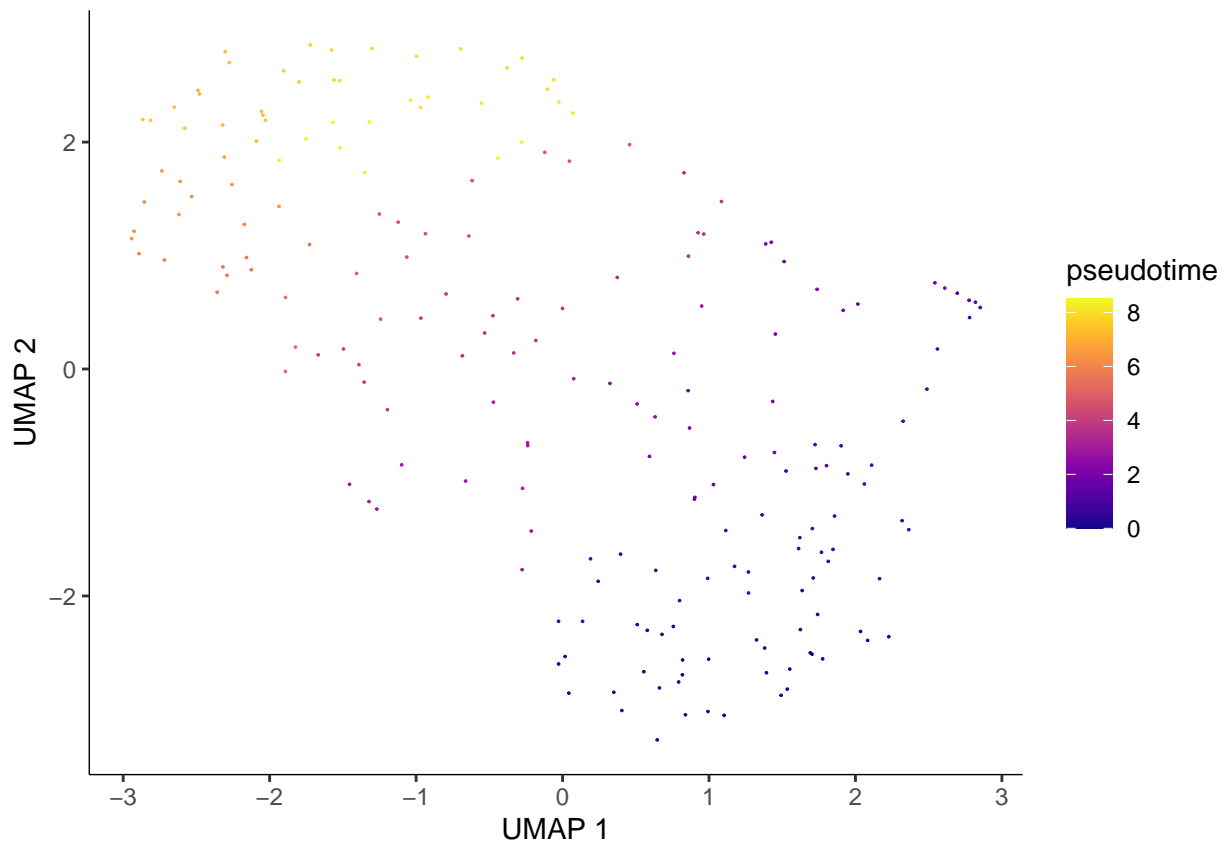
Correlation Analysis

Cluster 16

```r
cluster16.cellNames <- rownames(pData(cds))[pData(cds)$Cluster.IDs %in% c(16, 16.1)]

cds_16 <- cds[,cluster16.cellNames]

cds16_pg <- graph_test(cds_16, neighbor_graph="principal_graph", cores=4, verbose = F)

cds16_genes <- cds16_pg %>%
  filter(q_value < 0.05) %>%
  arrange(desc(morans_I)) %>%
  select(gene_short_name)
```

```r
cds16_genes$gene_short_name
```

```
##   [1] apl-1       ptp-4     tba-1      cdh-3      C06B8.7   F41C3.5
##   [7] ric-4       T20F10.2  cutl-26    gpb-1      Y44A6D.2  rhgf-1
##  [13] daf-1       C04F12.1  cmd-1      tct-1      des-2     K08F4.1
##  [19] Y57G11C.43  C45G7.4   F13H10.9   Y47D3A.20  fmi-1     fat-1
##  [25] eat-6       obr-4     T01D3.3    Y46E12BL.2 chdp-1    lgc-12
##  [31] mec-7       plx-2     W05H12.2   T05A8.3    tbcb-1    ftt-2
##  [37] F26F12.3    nsf-1     hlh-11     C01G6.5    F47G9.1   rps-19
##  [43] uba-5       copd-1    gly-5      ret-1      kin-2     lin-39
```

```
##  [49] F25E5.1    zig-11      tbb-2       ceh-88      eva-1       R13A5.9
##  [55] ric-3      rpl-4       arrd-25     clh-1       qns-1       unc-115
##  [61] cdh-1      zig-1       ncx-1       vha-3       cdgs-1      ZC410.5
##  [67] unc-64     hif-1       spc-1       C43H6.4     dig-1       imp-2
##  [73] marc-5     dcap-1      aqp-2       C50E3.6     nfi-1       oig-8
##  [79] T05A12.4   hsp-3       madf-5      ZK688.5     cpsf-1      eat-3
##  [85] kcnl-4     lgc-31      Y57A10A.28  flp-14      mafr-1      sec-61
##  [91] rpl-26     mbk-2       F54G2.1     anc-1       ZC449.5     Y38C1AA.1
##  [97] frm-1      M04G7.3     gex-3       sto-2       cct-7       cct-5
## [103] ain-2      F32D8.2     lim-9       Y18D10A.2   Y97E10C.1   rap-2
## [109] gbb-2      usp-3       C39F7.1     cap-2       let-60      rps-11
## [115] unc-44     let-611     unc-43      hpo-30      gdi-1       D1022.9
## [121] eef-1A.1   emb-4       sea-2       Y105E8A.20  his-24      C03A3.1
## [127] D1007.5    rpt-3       dma-1       nhr-78      Y53F4B.9    npa-1
## [133] B0416.5    Y54E2A.4    gbb-1       rps-10      hda-2       F55A3.3
## [139] Y54G2A.26  gck-3       pes-7       eef-1A.2    glb-18      mcd-1
## [145] W03F8.4    cab-1       spe-39      lron-3      F32B4.4     acy-2
## [151] Y39H10A.6  hlh-13      dhhc-2      F40F8.11    deg-3       set-27
## [157] gst-7      gei-4       rpl-19      tbc-12      flp-1       praf-3
## [163] T03G6.3    csn-5       tbc-1       F53A2.9     F36D4.5     F53F4.14
## [169] unc-1      rom-4       dnj-28      F15A8.4     cap-1       epi-1
## [175] atl-1      cls-2       Y43F8B.2    rpm-1       ben-1       unc-86
## [181] rps-14     F55C12.5    egl-21      ZK1320.9    T26C12.1    ser-2
## [187] flp-6      npp-3       rilp-1      egl-30      col-107     gdh-1
## [193] rpl-3      F43C9.2     ets-6       T19A6.1     ddl-3       atg-16.2
## [199] C04F5.9    F56G4.4     pmr-1       Y54E10BL.3  ced-10      aakb-1
## [205] tat-1      rpl-35      Y48G10A.6   ZK524.4     C53D5.1     gly-20
## [211] rps-1      hgap-1      T05C12.11   snt-6       chp-1       C25H3.8
## [217] taf-12     R07E5.1     mrck-1      F46H5.3     cct-4       hda-3
## [223] odr-4      flap-1      cyn-5       ham-3       pyp-1       sec-22
## [229] rib-1      C03A3.2     aagr-4      ZK154.6     Y111B2A.12  hpk-1
## [235] cdh-4      sbds-1      tol-1       mkk-4       C05D10.4    dgn-1
## [241] pde-2      rod-1       mrps-12     dnj-19      cey-1       ZK1073.1
## [247] ZK154.4    F48E8.4     rps-6       gmeb-1      Y47D9A.1    K02D10.4
## [253] syd-1      cdka-1      myrf-2      rps-0       C43H6.3     unc-41
## [259] ril-1      lin-40      itsn-1      Y71F9AL.9   tlp-1       sdc-2
## [265] Y53F4B.21  hinf-1      cct-3       bath-43     rps-25      W08E3.2
## [271] kpc-1      F49E10.4    H18N23.2    pop-1       F40E10.6    rps-28
## [277] prp-6      dnj-7       sec-15      F32D1.11    Y52B11A.3   C25A1.4
## [283] dhhc-14    hgap-2      R02F2.1     msi-1       arx-3       ZK1067.4
## [289] pck-1      F49C12.12   glr-8       C26D10.6    kin-19      gem-1
## [295] F55F8.2    C53C11.2    H34C03.2    cnt-1       emc-5       sur-6
## [301] taf-4      immt-1      Y111B2A.10  T22H9.1     C17H12.10   ZC101.1
## [307] gob-1      aqp-7       C49H3.9     rbbp-5      Y37E11AM.2  F25F8.1
## [313] C10B5.3    sel-9       ceh-86      mnat-1      C24H10.2    dpy-28
## [319] M01F1.3    Y54F10BM.9  F07F6.8     K07A12.8    F56D1.2     rad-23
## [325] vps-39     ccg-1       M03F8.5     F58H7.1     Y92H12BR.7  Y53G8AL.2
## [331] fasn-1     arch-1      F08G12.3    Y48E1B.3    hsp-75      zyg-12
## [337] T16H12.3   apn-1       mtx-2       Y66H1A.8    rtel-1      snt-7
## [343] har-1      Y17G7B.3    rde-10      C04E7.4     C45B2.6     sto-5
## [349] gro-1      T08G2.2     Y18D10A.3   gtbp-1      ttr-8       Y97E10AR.4
## [355] fbxa-137   Y71H2AM.20
## 20271 Levels: 2L52.1 2RSSE.1 4R79.2 6R55.2 aagr-1 aagr-2 aagr-3 aagr-4 ... zyx-1
```
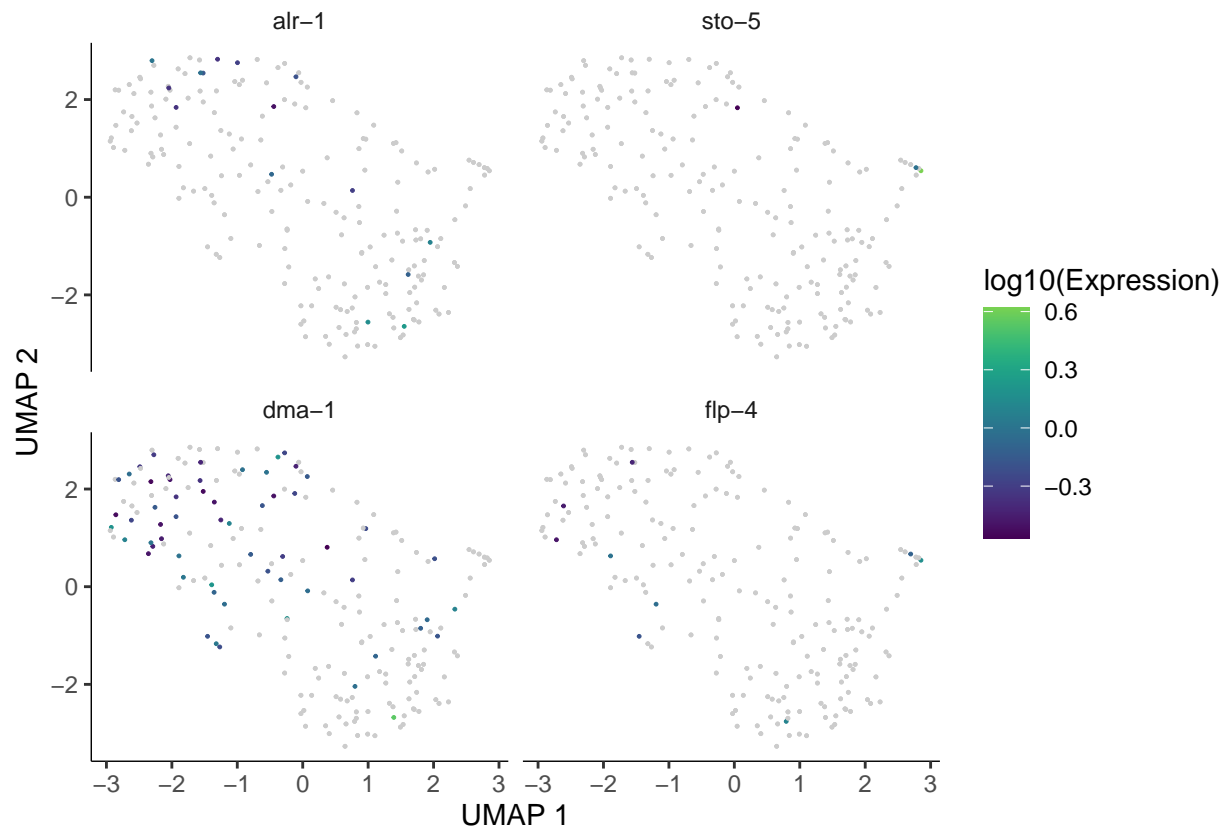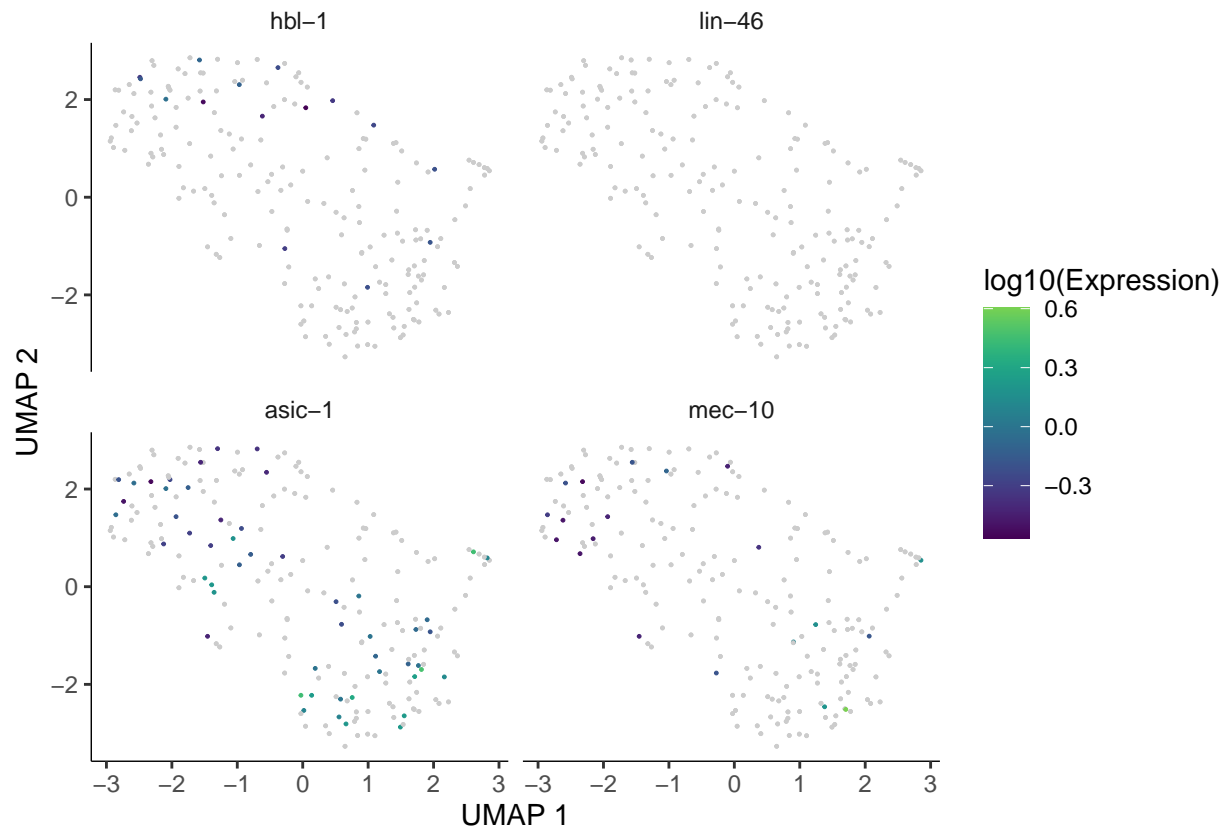
```
# Plot a few genes
plot_cells(cds_16, genes=c("Y48G10A.6", "egl-44", "egl-46", "zag-1"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = 0.5)
```
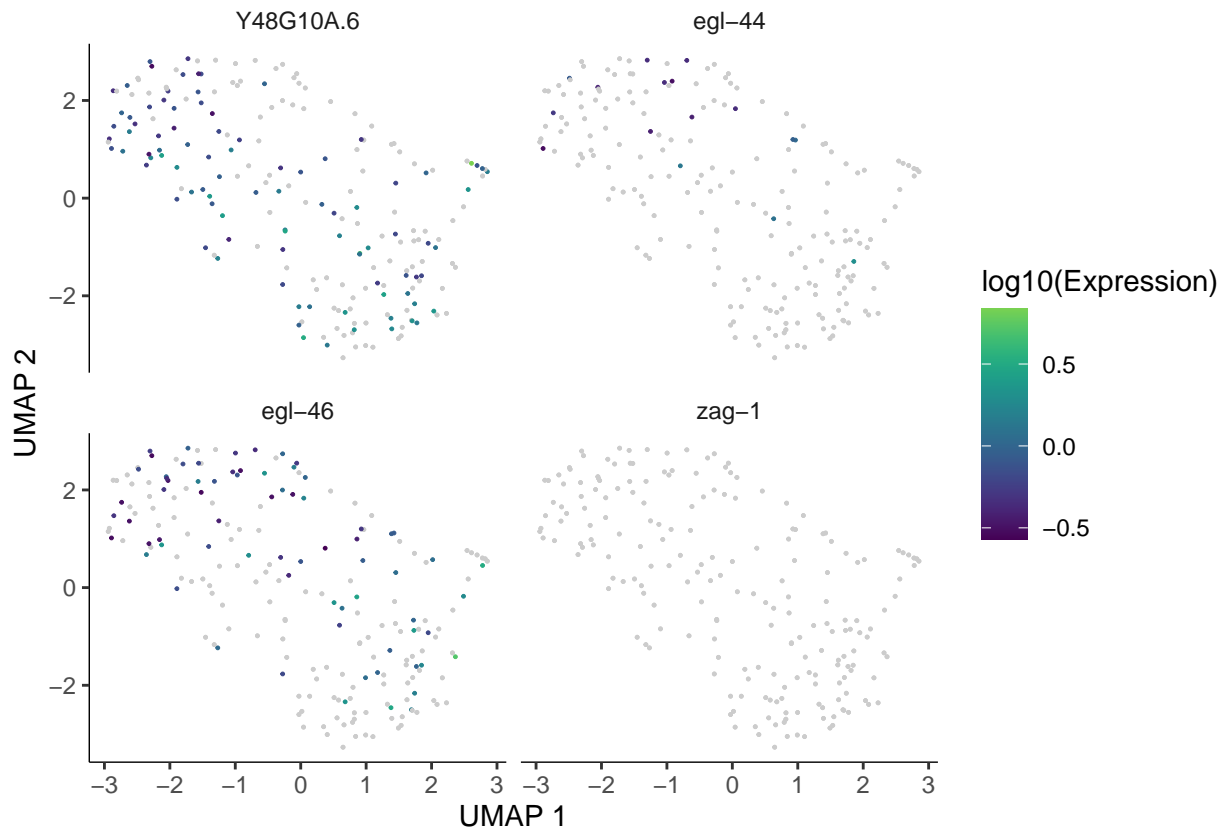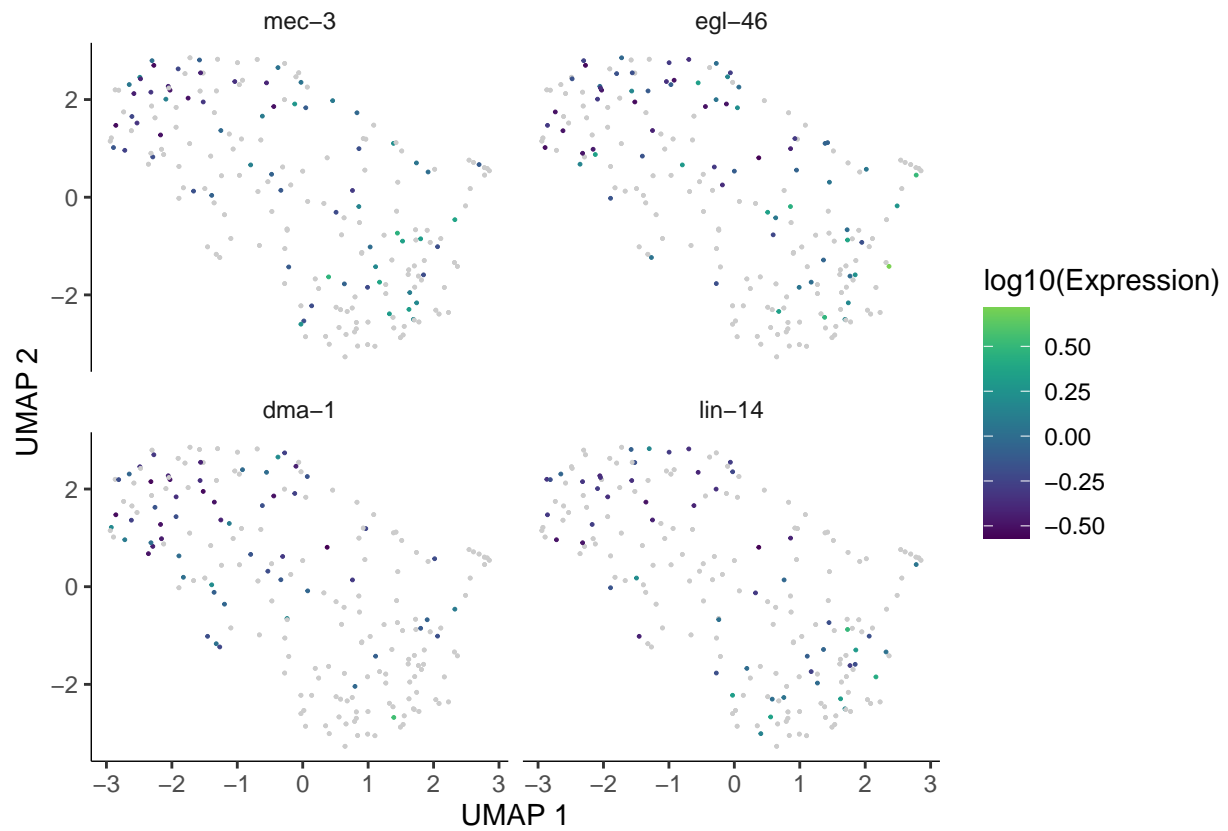


```
# Plot a few genes
plot_cells(cds_16, genes=c("alr-1", "sto-5", "dma-1", "flp-4"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = 0.5)
```

```r
# Plot a few genes
plot_cells(cds_16, genes=c("hbl-1", "lin-46","asic-1", "mec-10"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = 0.5)
```

```
# Plot a few genes
plot_cells(cds_16, genes=c("mec-3", "egl-46", "dma-1", "lin-14"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = 0.5)
```
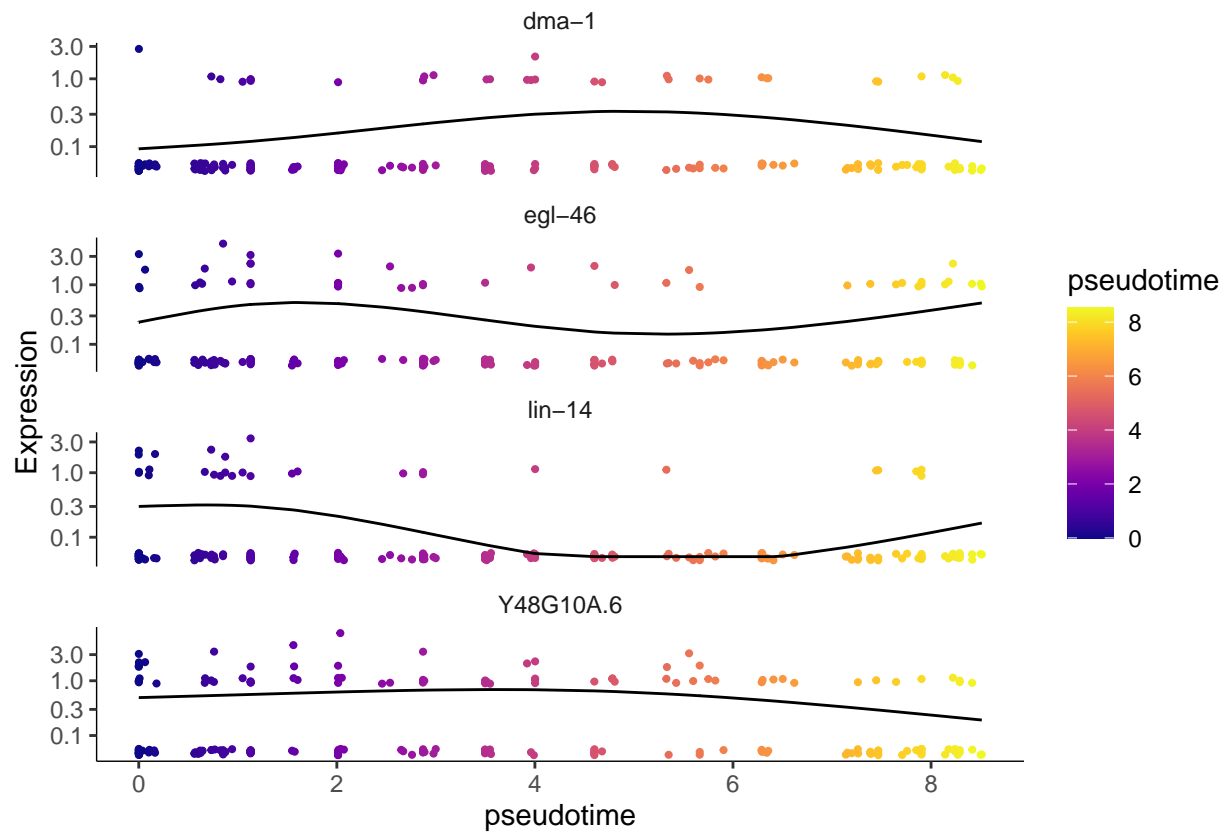
```r
# Plot a few genes
plot_cells(cds_16, genes=c("mec-3", "egl-46", "F49H12.4", "ser-2"),
           show_trajectory_graph=FALSE,
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           cell_size = 0.5)
```
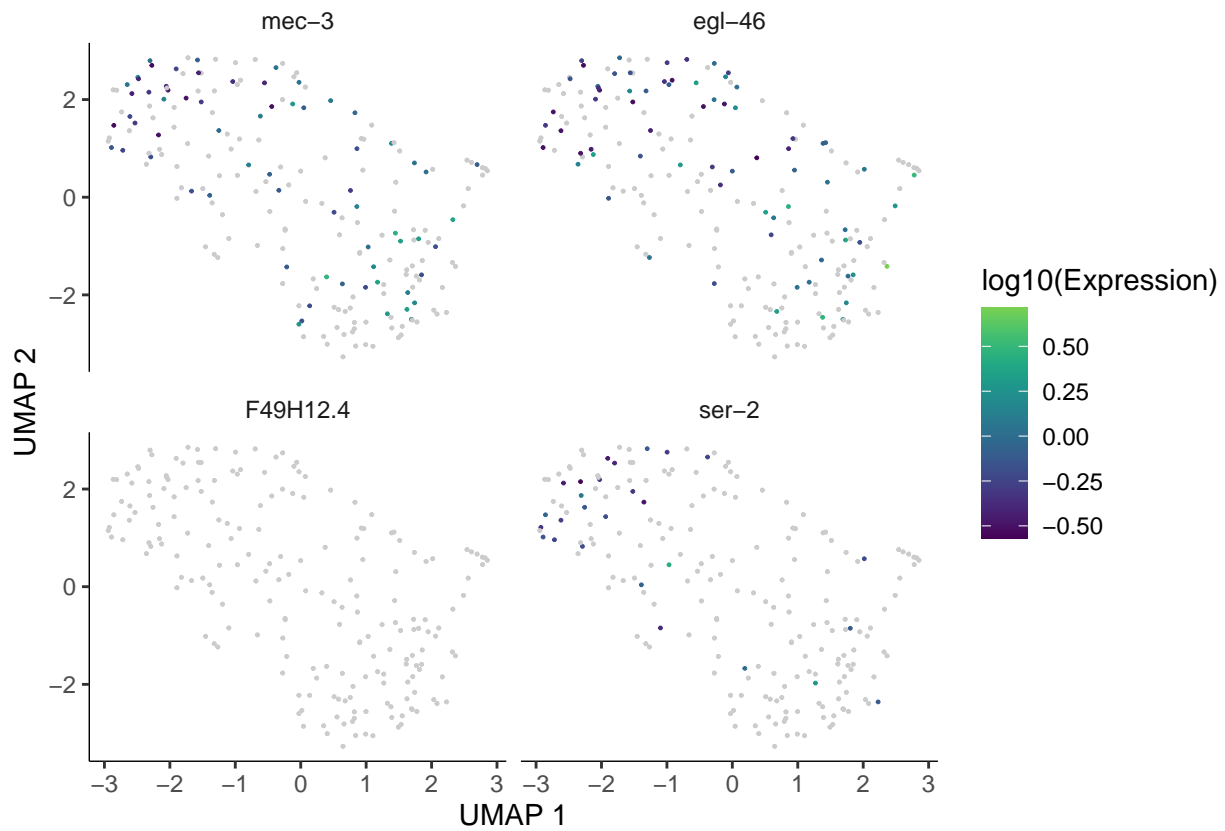
Clustering Genes by Pseudotemporal Expression Pattern

```r
cds_16_lineage_cds <- cds_16[rowData(cds_16)$gene_short_name %in% c("Y48G10A.6","dma-1", "lin-14", "egl-
plot_genes_in_pseudotime(cds_16_lineage_cds,
                         min_expr=0.05)
```

```
cds_16_lineage_cds <- cds_16[rowData(cds_16)$gene_short_name %in% c("mec-3", "mec-10", "asic-1","F49H12
plot_genes_in_pseudotime(cds_16_lineage_cds,
                         min_expr=0.05)
```