scRNAseq-singlecellexperiment

A scRNE-seq study

SingleCellExperiment (SCE) is a S4 class for storing data from single-cell experiments. Load SingleCellExperiment package

```
library(SingleCellExperiment)
library(ggplot2)
data=readRDS("fletcher.rds")
```

Dataset

Mouse Epithelium dataset Stem Cell Differentiation in the Mouse Olfactory Epithelium Cell Stem Cell, Fletcher et al, Deconstructing Olfactory Stem Cell Trajectories at Single Cell Resolution (2017) # create a SingleCellExperiment object sce

```
# create a SingleCellExperiment object sce
sce <- as(data, "SingleCellExperiment")
sce
```

```
FALSE class: SingleCellExperiment
FALSE dim: 100 94
FALSE metadata(0):
FALSE assays(1): counts
FALSE rownames(100): Cyp2g1 Sec14l3 ... Ifitm1 Maged1
FALSE rowData names(0):
FALSE colnames(94): OEL19_N724_S503 OEL19_N719_S502 ... OEL23_N701_S511
FALSE   OEL23_N703_S502
FALSE colData names(20): Experiment Batch ... ERCC_reads colPublishedClusters
FALSE reducedDimNames(0):
FALSE altExpNames(0):
```

```
# find dimensions
mydims <- dim(sce)

# extract cell and gene names
cellNames <- colnames(sce)
geneNames <- rownames(sce)
cellNames
```

```
FALSE  [1] "OEL19_N724_S503" "OEL19_N719_S502" "OEL21_N712_S507" "OEL19_N723_S506"
FALSE  [5] "OEL19_N720_S507" "OEL19_N721_S510" "OEL21_N711_S507" "OEL19_N719_S508"
FALSE  [9] "OEL19_N721_S505" "OEL19_N722_S503" "OEL22_N718_S515" "OEL22_N726_S516"
```

```
FALSE [13] "OEL19_N726_S505" "OEL22_N716_S518" "OEL21_N715_S505" "OEL21_N714_S505"
FALSE [17] "OEL22_N720_S521" "OEL19_N724_S508" "OEL22_N721_S516" "OEL22_N728_S516"
FALSE [21] "OEL19_N720_S505" "OEL19_N724_S505" "OEL21_N712_S505" "OEL22_N726_S515"
FALSE [25] "OEL22_N720_S520" "OEL22_N727_S515" "OEL22_N722_S522" "OEL19_N723_S502"
FALSE [29] "OEL19_N719_S503" "OEL22_N723_S522" "OEL19_N723_S507" "OEL19_N720_S506"
FALSE [33] "OEL22_N718_S513" "OEL22_N723_S513" "OEL19_N721_S503" "OEL19_N724_S510"
FALSE [37] "OEL19_N724_S507" "OEL19_N722_S510" "OEL22_N727_S517" "OEL22_N720_S522"
FALSE [41] "OEL22_N716_S522" "OEL19_N722_S502" "OEL21_N712_S502" "OEL22_N716_S516"
FALSE [45] "OEL22_N719_S517" "OEL19_N720_S511" "OEL19_N723_S505" "OEL21_N715_S510"
FALSE [49] "OEL21_N714_S502" "OEL22_N720_S517" "OEL21_N711_S505" "OEL21_N715_S502"
FALSE [53] "OEL22_N727_S513" "OEL21_N711_S508" "OEL22_N723_S521" "OEL22_N716_S515"
FALSE [57] "OEL19_N719_S510" "OEL19_N723_S511" "OEL19_N719_S507" "OEL22_N722_S517"
FALSE [61] "OEL19_N724_S506" "OEL19_N726_S502" "OEL22_N728_S517" "OEL22_N724_S521"
FALSE [65] "OEL22_N721_S521" "OEL22_N721_S517" "OEL22_N726_S522" "OEL22_N718_S517"
FALSE [69] "OEL23_N706_S507" "OEL23_N705_S511" "OEL23_N705_S507" "OEL23_N705_S508"
FALSE [73] "OEL23_N702_S510" "OEL23_N703_S505" "OEL23_N706_S510" "OEL23_N705_S505"
FALSE [77] "OEL23_N703_S507" "OEL23_N701_S503" "OEL23_N706_S505" "OEL23_N706_S506"
FALSE [81] "OEL23_N707_S502" "OEL23_N705_S503" "OEL23_N702_S502" "OEL23_N706_S503"
FALSE [85] "OEL23_N701_S502" "OEL23_N704_S507" "OEL23_N702_S505" "OEL23_N703_S503"
FALSE [89] "OEL23_N701_S505" "OEL23_N702_S506" "OEL23_N704_S505" "OEL23_N702_S507"
FALSE [93] "OEL23_N701_S511" "OEL23_N703_S502"
```

```
geneNames
```

```
FALSE   [1] "Cyp2g1"        "Sec14l3"       "Rgs5"          "Sdc4"
FALSE   [5] "Cbr2"          "Cyp2f2"        "Krt18"         "Fstl5"
FALSE   [9] "Nupr1"         "Sult1c1"       "Ebf1"          "Flrt1"
FALSE  [13] "Gstm1"         "Ugt2a1"        "Gap43"         "Prdx6"
FALSE  [17] "Prune2"        "Galm"          "Tubb3"         "Cyp2a5"
FALSE  [21] "Stmn2"         "Ncam1"         "Atf3"          "Mgst1"
FALSE  [25] "Rtn1"          "5730409K12Rik" "Ell3"          "Dpysl2"
FALSE  [29] "Tst"           "Ebf2"          "Foxn3"         "Vim"
FALSE  [33] "Fmo6"          "Ly6e"          "Sat1"          "1810011O10Rik"
FALSE  [37] "Sox11"         "Rd3"           "Rbm47"         "Ado"
FALSE  [41] "Sp8"           "Edn2"          "Gstm2"         "Stmn1"
FALSE  [45] "Elf3"          "Pon1"          "Cebpg"         "Zfp36"
FALSE  [49] "Calb2"         "Tmprss13"      "Krt8"          "Epas1"
FALSE  [53] "Enc1"          "Gm11223"       "Scn9a"         "Poldip3"
FALSE  [57] "Ces1d"         "Il33"          "Cyp2j6"        "Gpm6a"
FALSE  [61] "Znf512b"       "Cntn4"         "Anxa2"         "Anxa1"
FALSE  [65] "Rqcd1"         "Stmn3"         "Nrxn1"         "Ak1"
FALSE  [69] "Golt1b"        "Ugt2a2"        "Tuba1a"        "Actr1b"
FALSE  [73] "Btg2"          "Aqp3"          "Hist2h2be"     "Dpysl3"
FALSE  [77] "Lamp2"         "Traf3ip1"      "Gsta4"         "Blvrb"
FALSE  [81] "Rdh13"         "Snap25"        "Efcab14"       "Gm581"
FALSE  [85] "Hmgcs1"        "Zfp36l1"       "Entpd5"        "Dcn"
FALSE  [89] "Mef2b"         "Notch2"        "Spcs3"         "Cpm"
FALSE  [93] "Prdx1"         "P4hb"          "Glo1"          "Impact"
FALSE  [97] "Slc31a1"       "Trim66"        "Ifitm1"        "Maged1"
```

```
mydims
```

```
FALSE [1] 100  94
```

```
# cell data
cData <- colData(sce)
#print column names
colnames(cData)
```

```
FALSE  [1] "Experiment"           "Batch"                "publishedClusters"
FALSE  [4] "NREADS"               "NALIGNED"             "RALIGN"
FALSE  [7] "TOTAL_DUP"            "PRIMER"               "PCT_RIBOSOMAL_BASES"
FALSE [10] "PCT_CODING_BASES"     "PCT_UTR_BASES"        "PCT_INTRONIC_BASES"
FALSE [13] "PCT_INTERGENIC_BASES" "PCT_MRNA_BASES"       "MEDIAN_CV_COVERAGE"
FALSE [16] "MEDIAN_5PRIME_BIAS"   "MEDIAN_3PRIME_BIAS"   "CreER"
FALSE [19] "ERCC_reads"           "colPublishedClusters"
```

```
# subset cData to batch & clusters
cData <- cData[, c('Batch', 'publishedClusters')]
#tabulate cData
table(cData)
```

```
FALSE      publishedClusters
FALSE Batch  7  9 15
FALSE   P11  5 22  1
FALSE   P12 10  8  1
FALSE   P13 15  4  2
FALSE   P14 11 12  3
```

Subset sce

For computational efficacy, retain only the 50 most variable genes of the sce object

```
library(magrittr)
# gene variance
vars <- assay(sce) %>% log1p() %>% rowVars()

#rename vars
names(vars) <- rownames(sce)

#sort vars
vars_2 <- sort(vars, decreasing = TRUE)
head(vars_2)
```

```
FALSE   Cyp2g1  Sec14l3     Rgs5     Sdc4     Cbr2   Cyp2f2
FALSE 15.40474 14.72372 14.08343 13.25418 13.10627 12.68986
```

```
# subset sce
sce_sub <- sce[names(vars_2[1:50]),]
sce_sub
```

```
FALSE class: SingleCellExperiment
FALSE dim: 50 94
FALSE metadata(0):
```

```
FALSE assays(1): counts
FALSE rownames(50): Cyp2g1 Sec14l3 ... Calb2 Tmprss13
FALSE rowData names(0):
FALSE colnames(94): OEL19_N724_S503 OEL19_N719_S502 ... OEL23_N701_S511
FALSE   OEL23_N703_S502
FALSE colData names(20): Experiment Batch ... ERCC_reads colPublishedClusters
FALSE reducedDimNames(0):
FALSE altExpNames(0):
```

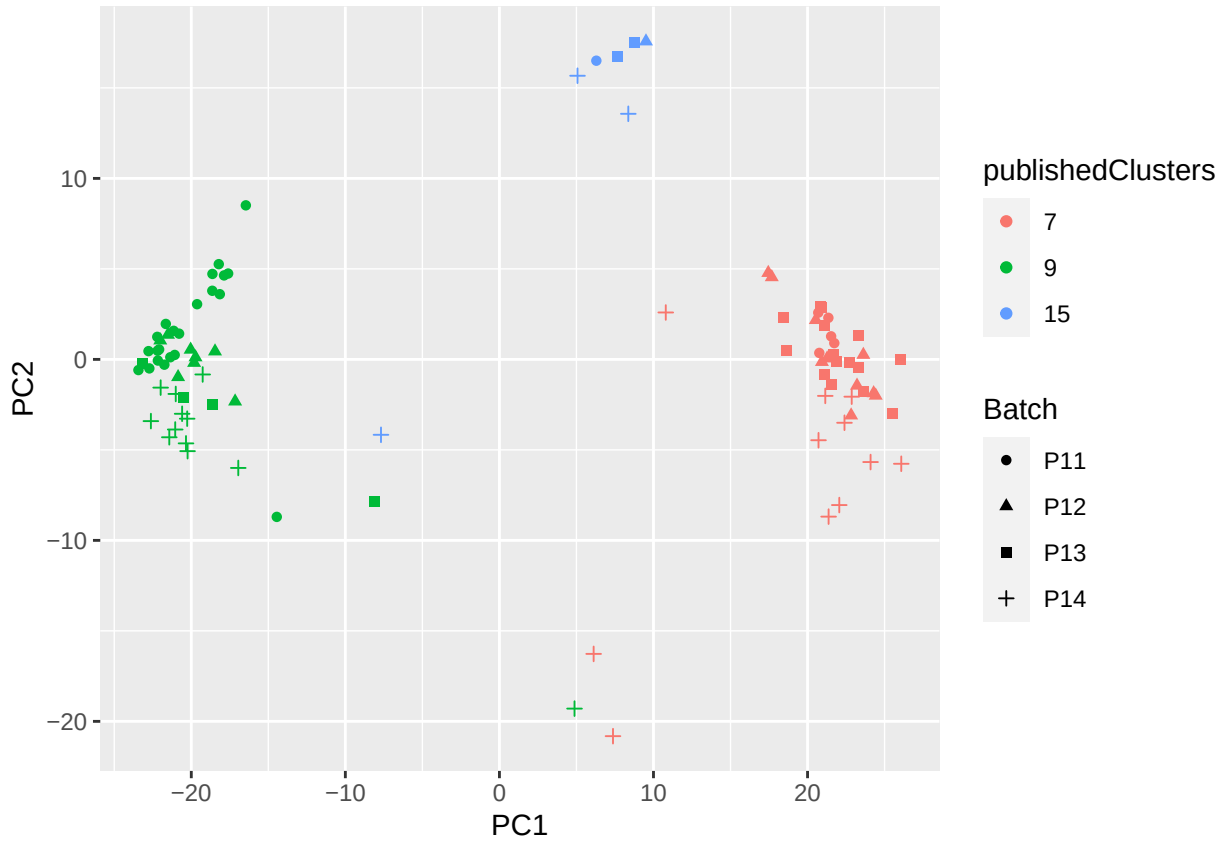
PCA

Perform PCA on log counts and Plot PCA using ggplot

```
# log counts
logcounts <- log1p(assay(sce_sub))

# transpose
tlogcounts <- t(logcounts)

# perform pca
pca <- prcomp(tlogcounts)

# store pca matrix in sce
reducedDims(sce_sub) <- SimpleList(PCA = pca$x)
head(reducedDim(sce_sub, "PCA")[, 1:2])
```

```
FALSE                       PC1        PC2
FALSE OEL19_N724_S503 -18.63651  3.7905674
FALSE OEL19_N719_S502  21.53071  1.2738851
FALSE OEL21_N712_S507  20.93405 -0.1382121
FALSE OEL19_N723_S506 -17.60803  4.7438350
FALSE OEL19_N720_S507  20.69562  2.5815635
FALSE OEL19_N721_S510 -18.62802  4.7205441
```

```
# Extract PC1 and PC2 and create a data frame
pca <- reducedDim(sce_sub, "PCA")[, 1:2]
col_shape <- data.frame(publishedClusters = colData(sce)$publishedClusters, Batch = factor(colData(sce)$
df <- cbind(pca, col_shape)

# plot PC1, PC2
ggplot(df, aes(x = PC1, y = PC2,
               colour = publishedClusters,
               shape = Batch)) +
  geom_point()
```


10
0
−10
−20
PC2
−20
−10
0
10
20
PC1
publishedClusters
7
9
15
Batch
P11
P12
P13
P14