Monocle: Differential expression and time-series analysis for single-cell RNA-Seq and qPCR experiments

Cole Trapnell

University of Washington,
Seattle, Washington, USA
coletrap@uw.edu

Davide Cacchiarelli

Harvard University,
Cambridge, Massachussetts, USA
davide@broadinstitute.org

November 26, 2015

Abstract

Single cell gene expression studies enable profiling of transcriptional regulation during complex biological processes and within highly hetergeneous cell populations. These studies allow discovery of genes that identify certain subtypes of cells, or that mark a particular intermediate states during a biological process. In many single cell studies, individual cells are executing through a gene expression program in an unsynchronized manner. In effect, each cell is a snapshot of the transcriptional program under study. The package *monocle* provides tools for analyzing single-cell expression experiments. It performs differential gene expression and clustering to identify important genes and cell states. It is designed for RNA-Seq studies, but can be used with qPCR or other targeted assays. For more information on the algorithm at the core of *monocle*, or to learn more about how to use single cell RNA-Seq to study a complex biological process, see Trapnell and Cacchiarelli *et al* [1]

Contents

1 BEAM

```
baseLoc <- system.file(package="monocle")
baseLoc <- './inst'
extPath <- file.path(baseLoc, "extdata")
load(file.path(extPath, "lung_phenotype_data.RData"))
```

```
## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed file './inst/extdata/lung_phenc
probable reason 'No such file or directory'
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
```

```
load(file.path(extPath, "lung_exprs_data.RData"))
```

```
## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed file './inst/extdata/lung_exprs
probable reason 'No such file or directory'
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
```

```
load(file.path(extPath, "lung_feature_data.RData"))
```

```
## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed file './inst/extdata/lung_featu
probable reason 'No such file or directory'
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
```

```
lung_exprs_data <- lung_exprs_data[,row.names(lung_phenotype_data)]
```

```
## Error in eval(expr, envir, enclos):  object 'lung_exprs_data' not found
```

```
pd <- new("AnnotatedDataFrame", data = lung_phenotype_data)
```

```
## Error in value[[3L]](cond):  object 'lung_phenotype_data' not found
##  AnnotatedDataFrame 'initialize' could not update varMetadata:
##  perhaps pData and varMetadata are inconsistent?
```

```
fd <- new("AnnotatedDataFrame", data = lung_feature_data)
```

```
## Error in value[[3L]](cond):  object 'lung_feature_data' not found
##  AnnotatedDataFrame 'initialize' could not update varMetadata:
##  perhaps pData and varMetadata are inconsistent?
```

```
# Now, make a new CellDataSet using the RNA counts
lung <- newCellDataSet(as.matrix(lung_exprs_data),
                       phenoData = pd,
                       featureData = fd,
                       lowerDetectionLimit=1,
                       expressionFamily=negbinomial())
```

```
## Error in as.matrix(lung_exprs_data):  object 'lung_exprs_data' not found
```

```
lung <- estimateSizeFactors(lung)
```

```
## Error in estimateSizeFactors(lung):  error in evaluating the argument 'object' in selecting
a method for function 'estimateSizeFactors':  Error:  object 'lung' not found
```

```
lung <- estimateDispersions(lung)
```

```
## Error in estimateDispersions(lung):  error in evaluating the argument 'object' in selecting
a method for function 'estimateDispersions':  Error:  object 'lung' not found
```

```
pData(lung)$Total_mRNAs <- colSums(exprs(lung))
```

```
## Error in exprs(lung):  error in evaluating the argument 'object' in selecting a method for func-
tion 'exprs':  Error:  object 'lung' not found
```

```
lung <- detectGenes(lung, min_expr = 1)
```

```
## Error in exprs(cds):  error in evaluating the argument 'object' in selecting a method for func-
tion 'exprs':  Error:  object 'lung' not found
```

```
expressed_genes <- row.names(subset(fData(lung), num_cells_expressed >= 120))
```

```
## Error in fData(lung):  error in evaluating the argument 'object' in selecting a method for func-
tion 'fData':  Error:  object 'lung' not found
```

```
ordering_genes <- expressed_genes
```

```
## Error in eval(expr, envir, enclos):  object 'expressed_genes' not found
```

```
lung <- setOrderingFilter(lung, ordering_genes)
```

```
## Error in fData(cds):  error in evaluating the argument 'object' in selecting a method for func-
tion 'fData':  Error:  object 'lung' not found
```

```
lung <- reduceDimension(lung, use_vst = F, pseudo_expr = 1)
```

```
## Error in exprs(cds):  error in evaluating the argument 'object' in selecting a method for func-
tion 'exprs':  Error:  object 'lung' not found
```

```
lung <- orderCells(lung, num_paths=2)
```

```
## Error in t(cds@reducedDimS): object 'lung' not found
```

```
plot_spanning_tree(lung, color_by="Time")
```

```
## Error in pData(cds):  error in evaluating the argument 'object' in selecting a method for func-
tion 'pData':  Error:  object 'lung' not found
```

```
BEAM_res <- BEAM(lung)
```

```
## Error in eval(expr, envir, enclos):  could not find function "BEAM"
```

```
head(BEAM_res)
```

```
## Error in head(BEAM_res):  object 'BEAM_res' not found
```

2 Session Info

```
sessionInfo()

## R version 3.2.1 (2015-06-18)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.10.1 (Yosemite)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
##  [1] splines   stats4    parallel  stats     graphics  grDevices utils
##  [8] datasets  methods   base
##
## other attached packages:
## [1] monocle_1.1.7          VGAM_0.9-8             HSMMSingleCell_0.101.5
## [4] ggplot2_1.0.1          reshape2_1.4.1         Biobase_2.28.0
## [7] BiocGenerics_0.14.0    knitr_1.11
##
## loaded via a namespace (and not attached):
##  [1] igraph_1.0.1       Rcpp_0.11.6        cluster_2.0.2
##  [4] magrittr_1.5       MASS_7.3-42        munsell_0.4.2
##  [7] lattice_0.20-31    colorspace_1.2-6   R6_2.0.1
## [10] dplyr_0.4.2        stringr_1.0.0      highr_0.5.1
## [13] plyr_1.8.3         tools_3.2.1        grid_3.2.1
## [16] gtable_0.1.2       irlba_1.0.3        DBI_0.3.1
## [19] matrixStats_0.14.2 assertthat_0.1     digest_0.6.8
## [22] Matrix_1.2-1       formatR_1.2.1      evaluate_0.8
## [25] limma_3.24.12      fastICA_1.2-0      stringi_0.5-5
## [28] scales_0.2.5       combinat_0.0-8     proto_0.3-10
```

References

[1] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 2014.