rawtoClust

Jo Hardin

6/12/2018

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.5
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## 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] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] mclust_5.4                  amap_0.8-16
##  [3] DESeq2_1.20.0               SummarizedExperiment_1.10.1
##  [5] DelayedArray_0.6.0          BiocParallel_1.14.1
##  [7] matrixStats_0.53.1          Biobase_2.40.0
##  [9] GenomicRanges_1.32.3        GenomeInfoDb_1.16.0
## [11] IRanges_2.14.10             S4Vectors_0.18.3
## [13] BiocGenerics_0.26.0         cluster_2.0.7-1
## [15] forcats_0.3.0               stringr_1.3.1
## [17] dplyr_0.7.5                 purrr_0.2.5
## [19] readr_1.1.1                 tidyr_0.8.1
## [21] tibble_1.4.2                ggplot2_2.2.1
## [23] tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
##  [1] nlme_3.1-137           bitops_1.0-6           bit64_0.9-7
##  [4] lubridate_1.7.4        RColorBrewer_1.1-2     httr_1.3.1
##  [7] rprojroot_1.3-2        tools_3.5.0            backports_1.1.2
## [10] R6_2.2.2               rpart_4.1-13           DBI_1.0.0
## [13] Hmisc_4.1-1            lazyeval_0.2.1         colorspace_1.3-2
## [16] nnet_7.3-12            tidyselect_0.2.4       gridExtra_2.3
## [19] mnormt_1.5-5           bit_1.1-14             compiler_3.5.0
## [22] cli_1.0.0              rvest_0.3.2            htmlTable_1.12
## [25] xml2_1.2.0             scales_0.5.0           checkmate_1.8.5
## [28] psych_1.8.4            genefilter_1.62.0      digest_0.6.15
## [31] foreign_0.8-70         rmarkdown_1.9          XVector_0.20.0
## [34] base64enc_0.1-3        pkgconfig_2.0.1        htmltools_0.3.6
## [37] htmlwidgets_1.2        rlang_0.2.1            readxl_1.1.0
## [40] RSQLite_2.1.1          rstudioapi_0.7         bindr_0.1.1
## [43] jsonlite_1.5           acepack_1.4.1          RCurl_1.95-4.10
## [46] magrittr_1.5           GenomeInfoDbData_1.1.0 Formula_1.2-3
## [49] Matrix_1.2-14          Rcpp_0.12.17           munsell_0.4.3
```

```
## [52] stringi_1.2.2          yaml_2.1.19           zlibbioc_1.26.0
## [55] plyr_1.8.4             blob_1.1.1            grid_3.5.0
## [58] crayon_1.3.4           lattice_0.20-35       haven_1.1.1
## [61] splines_3.5.0          annotate_1.58.0       hms_0.4.2
## [64] locfit_1.5-9.1         knitr_1.20            pillar_1.2.3
## [67] geneplotter_1.58.0     reshape2_1.4.3        XML_3.98-1.11
## [70] glue_1.2.0             evaluate_0.10.1       latticeExtra_0.6-28
## [73] data.table_1.11.4      modelr_0.1.2          cellranger_1.1.0
## [76] gtable_0.2.0           assertthat_0.2.0      xtable_1.8-2
## [79] broom_0.4.4            survival_2.42-3       memoise_1.1.0
## [82] AnnotationDbi_1.42.1   bindrcpp_0.2.2
```

Step 1. Decide which samples to use. (The countfunc does pull in the 100% genes, but we don't use them here.)

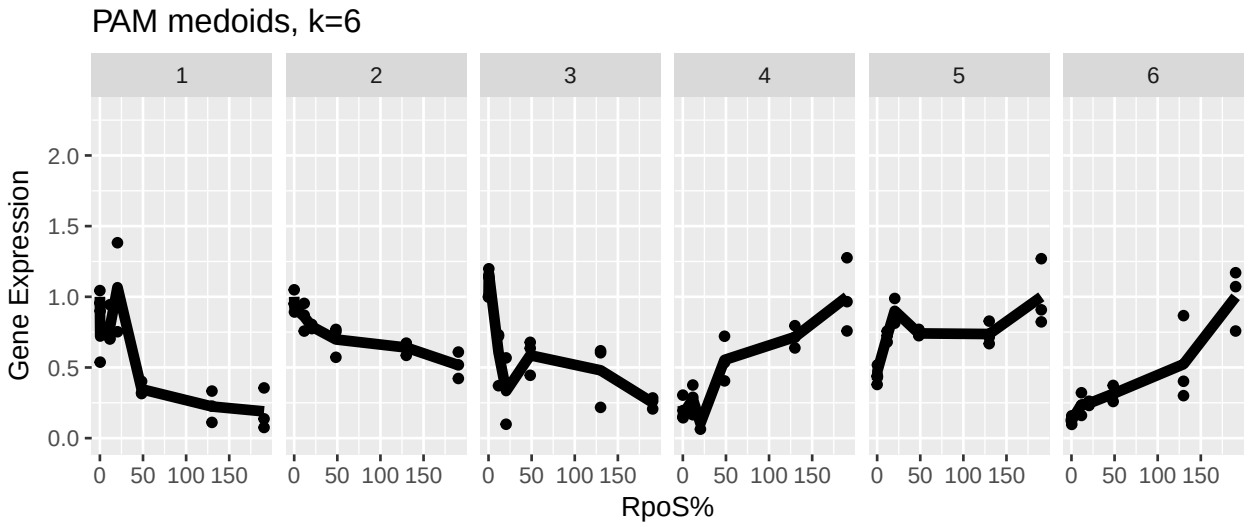
Step 2. Use those samples to input the dataset and create count data and tidy count data

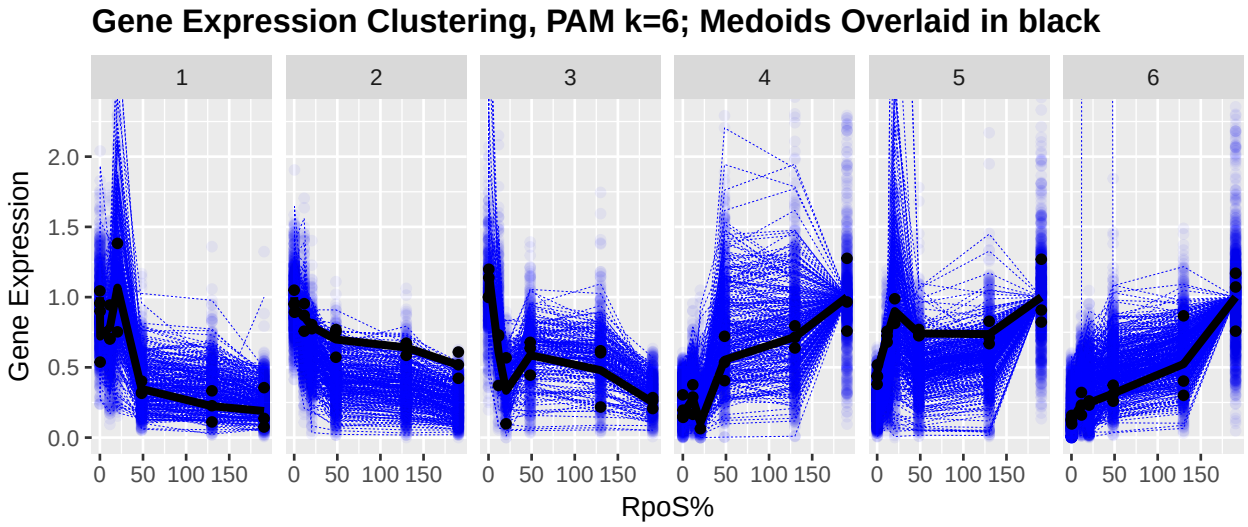
Step 3. Normalize the full dataset

Step 4. Find the DE genes

Step 5. Cluster the normalized counts that are significant for DE

Step 6. Plot the clusters




Gene Expression Clustering, PAM k=6; Medoids Overlaid in black
1
2
3
4
5
6
Gene Expression
2.0
1.5
1.0
0.5
0.0
0
50
100
150
RpoS%