

Sequencing Depth (work of MH)

Jo Hardin

6/26/2018

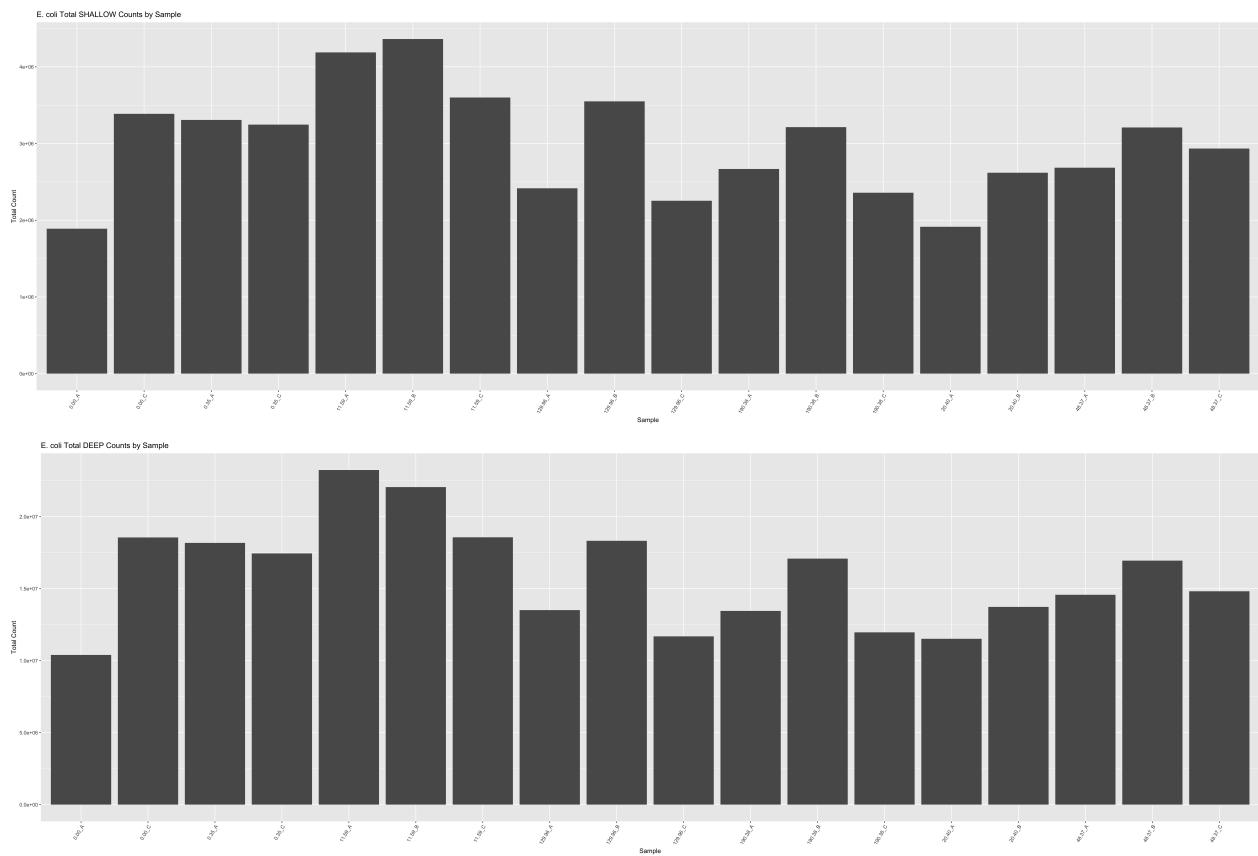
```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.5
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## 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 tidyverse_1.2.1
## [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_1.10 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
## [55] plyr_1.8.4             blob_1.1.1
## [58] crayon_1.3.4           lattice_0.20-35
## [61] splines_3.5.0           annotate_1.58.0
## [64] locfit_1.5-9.1          knitr_1.20
## [67] geneplotter_1.58.0      reshape2_1.4.3
## [70] glue_1.2.0               evaluate_0.10.1
## [73] data.table_1.11.4       modelr_0.1.2
## [76] gtable_0.2.0              assertthat_0.2.0
## [79] broom_0.4.4              survival_2.42-3
## [82] AnnotationDbi_1.42.1     bindrcpp_0.2.2

```

Pulling in data

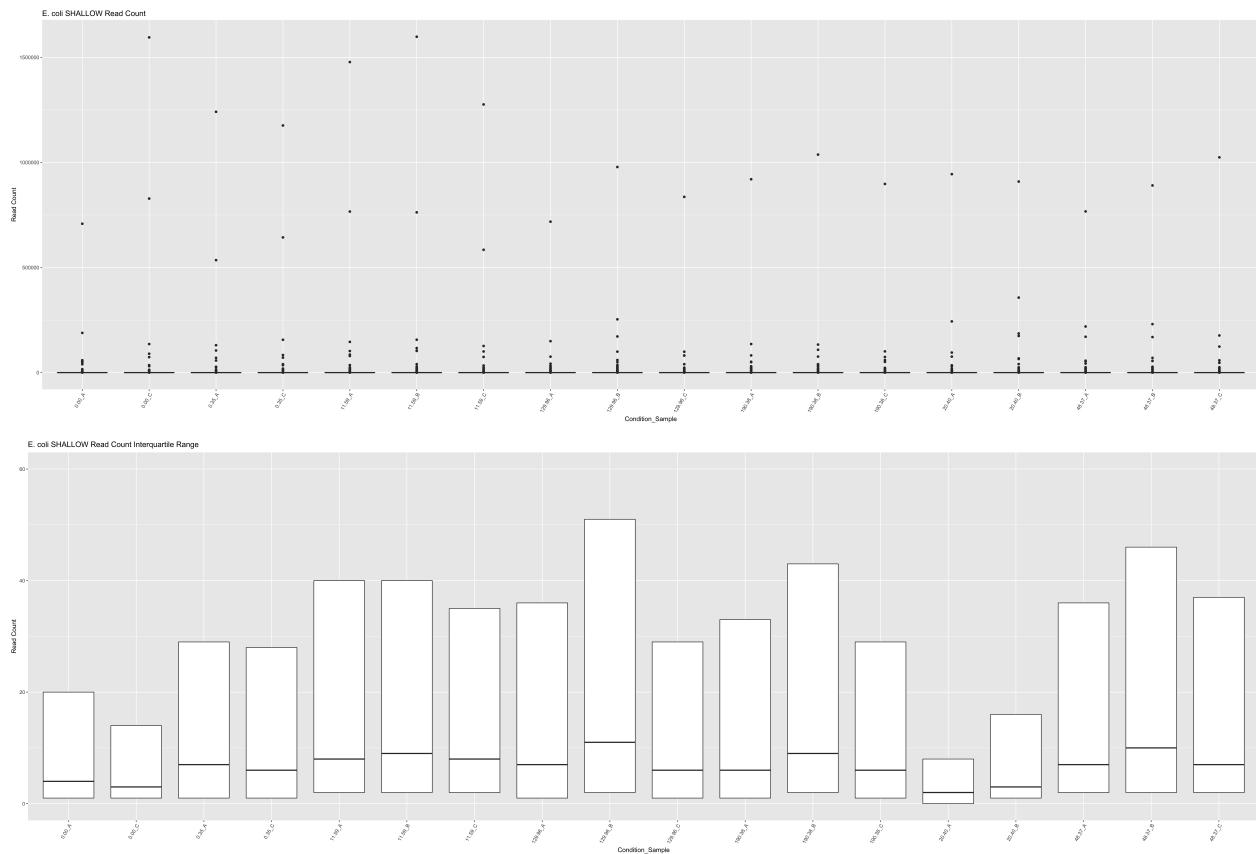


Normalizing data for sequencing depth

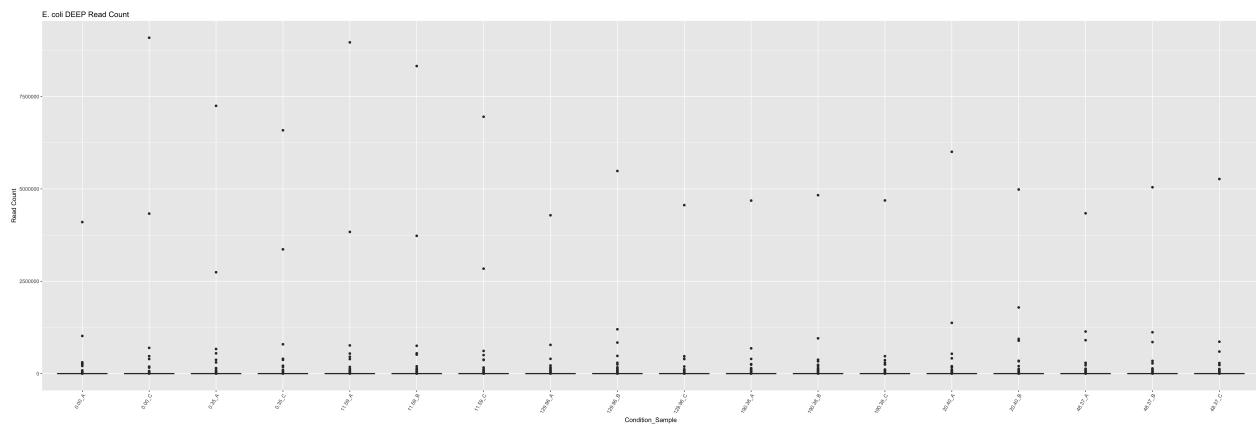
```
## # A tibble: 18 x 6
##   condition    Q1 median    Q3 sizeFactor totalCount
##   <chr>      <dbl>  <dbl> <dbl>      <dbl> <chr>
## 1 0.00        1     4     20     0.736 1889658
## 2 0.00        1     3     14     0.508 3386987
## 3 0.35        1     7     29     1.09  3308303
## 4 0.35        1     6     28     1.03  3246960
## 5 11.59       2     8     40     1.42  4187736
## 6 11.59       2     9     40     1.42  4362489
## 7 11.59       2     8     35     1.27  3600378
## 8 20.40       0     2     8      0.294 1915087
## 9 20.40       1     3     16     0.595 2619360
## 10 48.37      2     7     36     1.21  2684491
## 11 48.37      2    10     46     1.57  3209777
## 12 48.37      2     7     37     1.27  2934345
## 13 129.96     1     7     36     1.23  2416381
## 14 129.96     2    11     51     1.75  3550286
## 15 129.96     1     6     29     0.984 2254070
## 16 190.38     1     6     33     1.10  2668335
## 17 190.38     2     9     43     1.46  3214298
## 18 190.38     1     6     29     0.945 2358646

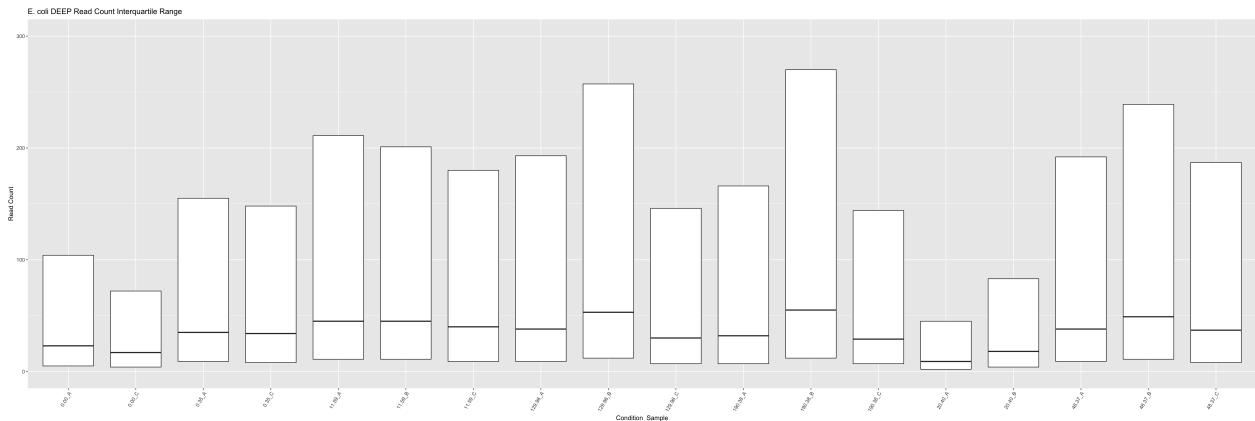
## # A tibble: 18 x 6
##   condition    Q1 median    Q3 sizeFactor totalCount
##   <chr>      <dbl>  <dbl> <dbl>      <dbl> <chr>
## 1 0          5     23    104     0.741 10395764
## 2 0          4     17    72      0.531 18554337
## 3 0.35      9     35    155     1.12  18179978
## 4 0.35      8     34    148     1.06  17446114
## 5 11.59     11    45    211     1.43  23240163
## 6 11.59     11    45    201     1.37  22053814
## 7 11.59     9     40    180     1.23  18570454
## 8 20.4       2     9     45     0.301 11522258
## 9 20.4       4     18    83      0.572 13733256
## 10 48.37     9     38    192     1.22  14574126
## 11 48.37    11    49    239     1.53  16947330
## 12 48.37     8     37    187     1.20  14817865
## 13 129.96    9     38    193     1.24  13515044
## 14 129.96   12    53    257.    1.66  18324358
## 15 129.96    7     30    146     0.949 11690122
## 16 190.38    7     32    166     1.04  13454873
## 17 190.38   12    55    270     1.43  17084795
## 18 190.38    7     29    144     0.901 11968598
```

Shallow Data



Deep Data





Note that the y-axes on both graphs have different scales, but the distribution of read counts per sample looks similar across the two sequencing depths. The samples with majority lower read count genes in the shallow data still display majority lower read count in the deep data. That is, the samples with the lowest Q3 read count in the shallow data are the same as the samples with lowest read count in the deep data: namely, 0.00_B, 0.35_B, and 20.40_A. The rest of the samples in the deep data follow a similar distribution to those in the shallow data.

This result could suggest that the deep data is confirming what the shallow data saw, but gene by gene plots are needed to see if this is true.

Clustering Data

