

Calculation of Mapping for RNAseq samples

Dave Bridges

April 2, 2021

Contents

1	Purpose	1
2	Experimental Details	1
3	Summary of Mapping	2
4	Session Information	2

1 Purpose

To calculate amount of mapping for each sample

2 Experimental Details

Ran samples through salmon, this script analyzes their log files

```
quant.directories <- 'quants'
sample.directories <- list.dirs(quant.directories,
                                recursive=F) #identify directories that contain salmon output
sample.directories <- sample.directories[grep('NEH', sample.directories)] #only noura samples
log.files <- file.path(sample.directories,"aux_info/meta_info.json") #locate meta_info files for each d
log.files <- log.files[file.exists(log.files)]

file <- log.files[1]
library(jsonlite)
mapping.data <- data.frame()
mapping.data <- setNames(data.frame(matrix(ncol = 4, nrow = 0)), c("file", "processed", "mapped", "perce
  mutate(file=as.character(),
         processed=as.integer(),
         mapped=as.integer(),
         percent.mapped=as.numeric()))

for (file in log.files) {
  data <- fromJSON(file)
  mapping.data <- add_row(mapping.data,
                          file=file,
                          processed=data$num_processed,
                          mapped=data$num_mapped,
```

```

    percent.mapped=data$percent_mapped)
}

```

3 Summary of Mapping

```

mapping.data %>%
  separate(file, into=c('Folder', 'Sample', 'Aux', 'Meta'), sep="/") %>%
  select(-Folder, -Aux, -Meta) %>%
  kable(caption="Sample level mapping results")

```

Table 1: Sample level mapping results

Sample	processed	mapped	percent.mapped
NEH-Sample_1415-NEH-1_quant	68797630	36919689	53.7
NEH-Sample_1415-NEH-10_quant	53074920	29967599	56.5
NEH-Sample_1415-NEH-11_quant	60844806	33393597	54.9
NEH-Sample_1415-NEH-2_quant	60837388	33004083	54.2
NEH-Sample_1415-NEH-3_quant	46190369	25817975	55.9
NEH-Sample_1415-NEH-4_quant	46879516	26061239	55.6
NEH-Sample_1415-NEH-5_quant	56732424	30472454	53.7
NEH-Sample_1415-NEH-6_quant	69409104	37894237	54.6
NEH-Sample_1415-NEH-7_quant	60796832	34385159	56.6
NEH-Sample_1415-NEH-8_quant	56688381	30427723	53.7
NEH-Sample_1415-NEH-9_quant	50846057	26973283	53.0

```

mapping.data %>%
  select(-file) %>%
  summarize_all(.funs=list(Average=mean,Min=min,Max=max)) %>%
  pivot_longer(everything(), names_sep="_", names_to=c('Measure', 'Stat')) %>%
  pivot_wider(everything(), names_from='Stat', values_from='value') %>%
  kable(caption="Summary statistics for salmon mapping.")

```

Table 2: Summary statistics for salmon mapping.

Measure	Average	Min	Max
processed	57372493.4	46190369	69409104.0
mapped	31392458.0	25817975	37894237.0
percent.mapped	54.8	53	56.6

4 Session Information

```

sessionInfo()

## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS 10.16
##

```

```

## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] jsonlite_1.7.2 dplyr_1.0.5   tidyr_1.1.3   knitr_1.31
##
## loaded via a namespace (and not attached):
## [1] magrittr_2.0.1   tidyselect_1.1.0 R6_2.5.0        rlang_0.4.10
## [5] fansi_0.4.2      highr_0.8        stringr_1.4.0   tools_4.0.2
## [9] xfun_0.22        utf8_1.2.1       DBI_1.1.1       htmltools_0.5.1.1
## [13] ellipsis_0.3.1   assertthat_0.2.1 yaml_2.2.1      digest_0.6.27
## [17] tibble_3.1.0     lifecycle_1.0.0  crayon_1.4.1    purrr_0.3.4
## [21] vctrs_0.3.7      glue_1.4.2       evaluate_0.14   rmarkdown_2.7
## [25] stringi_1.5.3    compiler_4.0.2   pillar_1.5.1    generics_0.1.0
## [29] pkgconfig_2.0.3

```