

peaks: get statistics

Maria Firulyova

02/11/2020

Load libs

```
library(data.table)
library(dplyr)
library(plyr)
```

Prepare function

```
get_stats <- function(path) {
  df <- fread(path) %>% select(c(V2, V3))
  data.frame(min_len = min(df$V3 - df$V2),
             max_len = max(df$V3 - df$V2),
             mean_len = floor(mean(df$V3 - df$V2)),
             median_len = floor(median(df$V3 - df$V2)))
}
```

Get input filenames (located in script directory)

```
broad_peaks <-
  list.files(recursive = T, full.names = T)[grepl('broad',
                                                  list.files(recursive = T,
                                                            full.names = T))]

narrow_peaks <-
  list.files(recursive = T, full.names = T)[grepl('narrow',
                                                  list.files(recursive = T,
                                                            full.names = T))]
```

Broad peaks: get stats

```
knitr::kable(ldply(broad_peaks, get_stats) %>%
  magrittr::set_rownames(gsub('.*|_peaks.*', '', broad_peaks)))
```

	min_len	max_len	mean_len	median_len
GSM1102782_CD14_H3K27ac_hg19_broad_0.1	295	25362	2002	1361
GSM1102785_CD14_H3K27me3_hg19_broad_0.1	299	63045	3257	1617
GSM1102788_CD14_H3K36me3_hg19_broad_0.1	298	138296	7737	2800
GSM1102793_CD14_H3K4me1_hg19_broad_0.1	297	40476	2454	1482
GSM1102797_CD14_H3K4me3_hg19_broad_0.1	298	10527	1577	1360

Narrow peaks: get stats

```
knitr::kable(ldply(narrow_peaks, get_stats) %>%
  magrittr::set_rownames(gsub('.*|_peaks.*', '', narrow_peaks)))
```

	min_len	max_len	mean_len	median_len
GSM1102782_CD14_H3K27ac_hg19_q0.05	295	9419	1000	647
GSM1102785_CD14_H3K27me3_hg19_q0.05	299	40641	1221	559
GSM1102788_CD14_H3K36me3_hg19_q0.05	298	31226	1380	681
GSM1102793_CD14_H3K4me1_hg19_q0.05	297	14367	1237	815
GSM1102797_CD14_H3K4me3_hg19_q0.05	298	6009	1238	977

Validate using JBR statistics (“about track” window)

#	Title	Min length	Max length	Mean length	Median length
1	GSM1102782_CD14_H3K27ac_hg19_broad_0.1_peaks.broadPeak	295	25 362	2 002	1 361
2	GSM1102782_CD14_H3K27ac_hg19_q0.05_peaks.narrowPeak	295	9 419	1 000	647
3	GSM1102785_CD14_H3K27me3_hg19_broad_0.1_peaks.broadPeak	299	63 045	3 257	1 617
4	GSM1102785_CD14_H3K27me3_hg19_q0.05_peaks.narrowPeak	299	40 641	1 221	559
5	GSM1102788_CD14_H3K36me3_hg19_broad_0.1_peaks.broadPeak	298	138 296	7 737	2 800
6	GSM1102788_CD14_H3K36me3_hg19_q0.05_peaks.narrowPeak	298	31 226	1 380	681
7	GSM1102793_CD14_H3K4me1_hg19_broad_0.1_peaks.broadPeak	297	40 476	2 454	1 482
8	GSM1102793_CD14_H3K4me1_hg19_q0.05_peaks.narrowPeak	297	14 367	1 237	815
9	GSM1102797_CD14_H3K4me3_hg19_broad_0.1_peaks.broadPeak	298	10 527	1 577	1 360
10	GSM1102797_CD14_H3K4me3_hg19_q0.05_peaks.narrowPeak	298	6 009	1 238	977

Figure 1: “JBR: peaks statistics”