Analyze CRISPRi growth competition data for growth at high CO2 gas feeds, calculate adjusted p values between two conditions

Ute Hoffmann (Science For Life Laboratory (KTH), Stockholm, Sweden)

oktober 18, 2024

Contents

1 Aim of the analysis

Calculation of Wilcoxon rank sum test between sgRNA fitness values between 4% CO2 data and 30% CO2 data.

2 Analysis

In a first step, the results given by the Nextflow pipeline are loaded.

```
load("../results_controlsgRNAs/fitness/result.Rdata")
num_sgRNAs <- read_tsv("../input/number_sgRNAs_per_target.tsv", col_names = c("sgRNA_target", "num_sgRN
DESeq_result_table <- left_join(DESeq_result_table, num_sgRNAs)

count_matrix <- read_tsv("../results_controlsgRNAs/prepare/all_counts.tsv")
count_matrix$Gene <- NULL

df_samplesheet <- readr::read_csv("../input/samplesheet_CRISPRi_CO2_Elena.csv", col_types = cols()) %>%
    select(all_of(c("sample", "condition", "replicate", "time", "group", "reference_group"))) %>%
    dplyr::mutate(group = factor(`group`))
df_samplesheet$name <- paste("gen_", df_samplesheet$time, "_r_", df_samplesheet$replicate, sep="")
```

```
get_controls <- function(condition, sgRNA_target){
  sgRNA_spec = sgRNA_target
  cond_spec = condition
  filter(DESeq_result_table, condition != cond_spec & sgRNA_target == sgRNA_spec)$fitness
}

DESeq_result_table <- dplyr::left_join(
  DESeq_result_table,
  DESeq_result_table %>%
```

```
    dplyr::group_by(sgRNA_target, condition, time) %>%
    dplyr::summarize(
      .groups = "keep",
      # apply Wilcoxon rank sum test against other condition
      p_fitness_condition = stats::wilcox.test(
        x = fitness,
        y = get_controls(condition, sgRNA_target),
        alternative = "two.sided"
        )$p.value,
      # apply Wilcoxon rank sum test against other condition
      p_fitness_condition_2 = stats::wilcox.test(
        x = unique(fitness),
        y = unique(get_controls(condition, sgRNA_target)),
        alternative = "two.sided"
        )$p.value
      ),
  by = c("sgRNA_target", "condition", "time")
  ) %>%
  group_by(condition, time) %>%
  mutate(
    p_fitness_condition_adj = stats::p.adjust(p_fitness_condition, method = "BH"),
    p_fitness_condition_2_adj = stats::p.adjust(p_fitness_condition_2, method = "BH"),
    comb_score = abs(wmean_fitness) * -log10(p_fitness_adj)
    )
```

```
save(DESeq_result_table, file = "../R_results_controlssgRNAs/result.Rdata")
readr::write_tsv(DESeq_result_table, file = "../R_results_controlssgRNAs/result.tsv")
```

3 Add annotation to results tables

In the following, annotation is added to the results table provided by the Nextflow pipeline. Mapping of the sgRNA targets to slr-locus tags is given in this file, downloaded on 24/02/23: https://github.com/m-jahn/R-notebook-crispri-lib/blob/master/sgRNA_library_V2/data/input/mapping_trivial_names.tsv The appended annotation is based on Uniprot and Cyanobase, partially edited manually. The table used for annotation was created beginning of 2021. Therefore, it does not include several genes which were only recently characterized. For a detailed description of all the columns given in the results tables, consult https://mpusp.github.io/nf-core-crispriscreen/output or https://www.biorxiv.org/content/10.1101/2023.02.13.528328v1.full.pdf+htmls

```
mapping_gene_locus <- read_tsv("../input/2023-02-24_mapping_trivial_names.tsv", show_col_types=FALSE)
names(mapping_gene_locus) <- c("sgRNA_target", "locus")
DESeq_result_table <- DESeq_result_table %>% left_join(mapping_gene_locus)
```

```
annotation <- read_tsv("../input/annotation_locusTags_stand13012021.csv", show_col_types = FALSE)
annotation_2 <- annotation[,c(1,2,3)]
names(annotation_2) <- c("locus", "Gene name","Product")
DESeq_result_table <- DESeq_result_table %>% left_join(annotation_2)
```

```
write_tsv(DESeq_result_table, file="../R_results_controlssgRNAs/annotated_DESeq_result_table_comparisons
df_reduced_info <- unique(subset(DESeq_result_table, DESeq_result_table$time==8 | DESeq_result_table$ti
write_tsv(df_reduced_info, file="../R_results_controlssgRNAs/Reduced_annotated_DESeq_result_table_compa
```

```
df_red_wide <- pivot_wider(df_reduced_info, names_from=condition, values_from=c(wmean_fitness, sd_fitnes
df_red_wide$impact_score <- (df_red_wide$wmean_fitness_CO2_30percent - df_red_wide$wmean_fitness_CO2_4p
```

```
write_tsv(df_red_wide, file="../R_results_controlssgRNAs/Wide_DESeq_result_table_comparisonsConditions.t
```

Session info

```
## R version 4.4.1 (2024-06-14)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=sv_SE.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=sv_SE.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=sv_SE.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=sv_SE.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Stockholm
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] pheatmap_1.0.12        enrichplot_1.24.0      clusterProfiler_4.12.0
##  [4] magrittr_2.0.3         lubridate_1.9.3        forcats_1.0.0
##  [7] stringr_1.5.1          dplyr_1.1.4            purrr_1.0.2
## [10] readr_2.1.5            tidyr_1.3.1            tibble_3.2.1
## [13] tidyverse_2.0.0        ggpubr_0.6.0           ggrepel_0.9.5
## [16] ggplot2_3.5.1          knitr_1.47
##
## loaded via a namespace (and not attached):
##   [1] DBI_1.2.2               gson_0.1.0              shadowtext_0.1.3
##   [4] gridExtra_2.3           rlang_1.1.3             DOSE_3.30.1
##   [7] compiler_4.4.1          RSQLite_2.3.7           png_0.1-8
##  [10] vctrs_0.6.5             reshape2_1.4.4          pkgconfig_2.0.3
##  [13] crayon_1.5.2            fastmap_1.2.0           backports_1.5.0
##  [16] XVector_0.44.0          ggraph_2.2.1            utf8_1.2.4
##  [19] HDO.db_0.99.1           rmarkdown_2.27          tzdb_0.4.0
##  [22] UCSC.utils_1.0.0        bit_4.0.5               xfun_0.44
##  [25] zlibbioc_1.50.0         cachem_1.1.0            aplot_0.2.2
##  [28] GenomeInfoDb_1.40.1     jsonlite_1.8.8          blob_1.2.4
##  [31] tweenr_2.0.3            BiocParallel_1.38.0     broom_1.0.6
##  [34] parallel_4.4.1          R6_2.5.1                RColorBrewer_1.1-3
##  [37] stringi_1.8.4           car_3.1-2               GOSemSim_2.30.0
##  [40] Rcpp_1.0.12             IRanges_2.38.0          Matrix_1.6-5
##  [43] splines_4.4.1           igraph_2.0.3            timechange_0.3.0
##  [46] tidyselect_1.2.1        viridis_0.6.5           qvalue_2.36.0
##  [49] rstudioapi_0.16.0       abind_1.4-5             yaml_2.3.8
##  [52] codetools_0.2-19        lattice_0.22-5          plyr_1.8.9
```

```
##  [55] treeio_1.28.0          Biobase_2.64.0          withr_3.0.0
##  [58] KEGGREST_1.44.0        evaluate_0.23           gridGraphics_0.5-1
##  [61] scatterpie_0.2.2       polyclip_1.10-6         Biostrings_2.72.0
##  [64] ggtree_3.12.0          pillar_1.9.0            carData_3.0-5
##  [67] stats4_4.4.1           ggfun_0.1.5             generics_0.1.3
##  [70] vroom_1.6.5            S4Vectors_0.42.0        hms_1.1.3
##  [73] tidytree_0.4.6         munsell_0.5.1           scales_1.3.0
##  [76] glue_1.7.0             lazyeval_0.2.2          tools_4.4.1
##  [79] data.table_1.15.4      fgsea_1.30.0            ggsignif_0.6.4
##  [82] graphlayouts_1.1.1     fs_1.6.4                fastmatch_1.1-4
##  [85] tidygraph_1.3.1        cowplot_1.1.3.9000      grid_4.4.1
##  [88] ape_5.8                AnnotationDbi_1.66.0    colorspace_2.1-1
##  [91] nlme_3.1-165           patchwork_1.2.0         GenomeInfoDbData_1.2.12
##  [94] ggforce_0.4.2          cli_3.6.2               fansi_1.0.6
##  [97] viridisLite_0.4.2      gtable_0.3.5            rstatix_0.7.2
## [100] yulab.utils_0.1.4      digest_0.6.35           BiocGenerics_0.50.0
## [103] ggplotify_0.1.2        farver_2.1.2            memoise_2.0.1
## [106] htmltools_0.5.8.1      lifecycle_1.0.4         httr_1.4.7
## [109] GO.db_3.19.1           MASS_7.3-61             bit64_4.0.5
```