

```

1 #set working directory
2 setwd("~/04_Champion_Lab/02_N-terminal_Acetylation/OnePot/NAT5")
3
4 #import libraries
5 library(ggplot2)
6 library(dplyr)
7 library(stringr)
8 library(tidyr)
9 library(ggrepel)
10 library(QFeatures)
11 library(corrplot)
12 library(limma)
13 library(tibble)
14
15 #set colors
16 red <- hue_pal()(3)[1]
17 green <- hue_pal()(3)[2]
18 blue <- hue_pal()(3)[3]
19
20 #import protein quant data and filter contaminants
21 proteins <- read.csv("lfq.proteins.csv")
22 proteins <- proteins[!grepl("CONTAM", proteins$Accession),]
23
24 #parse accession and gene info, save as dictionary
25 proteins$Accession2 <- str_remove(str_extract(proteins$Accession, "^[^\\|]*\\|"), "\\|")
26 proteins$gene <- str_remove(str_remove(str_extract(proteins$Accession,
27 "^[^\\|]*\\|[^\\|]*\\|"), "^[^\\|]*\\|"), "\\|")
28 geneDict <- as.list(proteins$gene)
29 names(geneDict) <- proteins$Accession2
30
31 #clean column names
32 names(proteins)[which(grepl("tr[0-9].Area", names(proteins)))] <- str_remove(str_extract(
33 names(proteins)[which(grepl("tr[0-9].Area", names(proteins)))]
34 "_[^_]+_br[0-9]_tr[0-9]"), "^_")
35 #set missing values to NA
36 proteins[proteins == 0] <- NA
37
38 #import to Qfeatures and set metadata
39 Qprot <- readQFeatures(table = proteins,
40                       ecol = which(grepl("tr[0-9]", names(proteins))),
41                       fnames = "Accession2",
42                       name = "raw_proteins")
43 colnames(Qprot[[1]])
44 Qprot$bioRep <- str_extract(colnames(Qprot[[1]]), "br[0-9]+")
45 Qprot$techRep <- str_extract(colnames(Qprot[[1]]), "tr[0-9]+")
46 Qprot$condition <- str_extract(colnames(Qprot[[1]]), "^[^_]*")
47 Qprot$Inj <- colnames(Qprot[[1]])
48 colData(Qprot)
49
50 #log transform, base 2 so that later FC is in base 2
51 Qprot <- addAssay(Qprot,
52                  logTransform((Qprot[["raw_proteins"]]),
53                               base = 2),
54                  name = "log_proteins")
55 #normalize
56 Qprot <- addAssay(Qprot,
57                  normalize(Qprot[["log_proteins"]],
58                           method = "center.median"),
59                  name = "norm_proteins")
60 #save figure of log and normalizations
61 png("LogNormalization.png", units = "px", width = 1500, height = 1000)
62 par(mfrow = c(1,3))
63 limma::plotDensities(assay(Qprot[[1]]), legend = F)
64 limma::plotDensities(assay(Qprot[[2]]), legend = F)
65 limma::plotDensities(assay(Qprot[[3]]), legend = F)
66 dev.off()

```

```

67
68
69 #add in old column data to normalized data
70 colData(Qprot[["norm_proteins"]]) <- colData(Qprot)
71
72 #save as a summary dataframe
73 brSummaryWide <- Qprot[["norm_proteins"]] %>% assay() %>% data.frame() %>%
74   rownames_to_column("Accession")
75 proteinsBR <- Qprot[["norm_proteins"]]
76 #set variables as factor, and clarify the reference as WT
77 proteinsBR$condition <- factor(Qprot$condition)
78 proteinsBR$condition <- relevel(proteinsBR$condition, ref = "WT")
79
80 #create model based on values in condition
81 model_design <- model.matrix(~proteinsBR$condition)
82
83 #fit proteins to linear model with lmFit and the matrix we just created
84 fitted_lm <- proteinsBR %>%
85   assay() %>%
86   lmFit(design = model_design) %>%
87   eBayes()
88
89 #pull WT vs deletion
90 DE_DEL <- topTable(fit = fitted_lm,
91   adjust.method = "BH",
92   number = Inf,
93   coef = "proteinsBR$conditiond1839",
94   sort.by = "p",
95   confint = T)
96
97 DE_DEL$Accession <- rownames(DE_DEL)
98
99
100
101 #calculate corrected z score based on B-H adujsted p-value. Altman,
102   https://doi.org/10.1136/bmj.d2090
103 DE_DEL$corrected_z <- sqrt(0.743 - (2.404*log(DE_DEL$adj.P.Val))) - 0.862
104 #calculate the one directional 95% confidence interval
105 DE_DEL$error <- abs(DE_DEL$logFC / DE_DEL$corrected_z) * 1.96
106
107 #pull only values with non-missing expression data
108 DE_DEL <- DE_DEL[!is.na(DE_DEL$AveExpr),]
109 DE_DEL$CountWT <- NA
110 DE_DEL$CountExp <- NA
111
112 #count number of valid measurments for each strain
113 for (i in 1:nrow(DE_DEL)) {
114   CountDF <- brSummaryWide[brSummaryWide$Accession == DE_DEL$Accession[i],]
115   DE_DEL$CountWT[i] <- sum(!is.na(CountDF[,grepl("WT", names(CountDF))][1,]))
116   DE_DEL$CountExp[i] <- sum(!is.na(CountDF[,grepl("d1839", names(CountDF))][1,]))
117 }
118
119 #set appropriate negative and positive infinite changes for totally missing values
120 DE_DEL$logFC[DE_DEL$CountWT == 0] <- Inf
121 DE_DEL$logFC[DE_DEL$CountExp == 0] <- -Inf
122
123 #set filters
124 adjSigCutoff <- 0.05
125 NonAdjSigCutoff <- 0.05
126 LFCcutoff <- 1
127
128
129 #option to plot either adj or non-adj p values
130 adj <- T
131
132 if (adj) {
133   proteins_sig <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff]))

```

```

134 proteins_sig2 <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
135                                     abs(DE_DEL$logFC) > LFCcutoff]))
136 proteins_label_pos <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
137                                     DE_DEL$logFC > LFCcutoff]))
138 proteins_label_neg <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
139                                     DE_DEL$logFC < -LFCcutoff]))
140 } else {
141 proteins_sig <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff]))
142 proteins_sig2 <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
143                                     abs(DE_DEL$logFC) > LFCcutoff]))
144 proteins_label_pos <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
145                                     DE_DEL$logFC > LFCcutoff]))
146 proteins_label_neg <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
147                                     DE_DEL$logFC < -LFCcutoff]))
148 }
149
150
151
152
153
154 #add in gene and protein name info from saved dictionaries
155 DE_DEL$gene <- sapply(DE_DEL$Accession, function(Accession) geneDict[[Accession]])
156
157 #set colors
158 colorsUPDOWN <- c(blue, red)
159 names(colorsUPDOWN) <- c(T, F)
160
161 #add extra area above and below plot for error bars
162 plotBufferFactor <- 1.5
163 maxY <- max(DE_DEL$logFC[is.finite(DE_DEL$logFC)]) * plotBufferFactor
164 minY <- min(DE_DEL$logFC[is.finite(DE_DEL$logFC)]) * plotBufferFactor
165
166 #plot MA
167 ggplot() +
168   theme_bw(base_size = 25) +
169   theme(panel.grid = element_blank(),
170         legend.position = "none") +
171   labs(y = "Log2 Fold Change", x = "Normalized Average Expression (Log)",
172        title = expression("WT vs " * Delta * italic(empl))) +
173   geom_point(data = DE_DEL[!DE_DEL$Accession %in% proteins_sig &
174                           is.finite(DE_DEL$logFC)],,
175             aes(x = AveExpr, y = logFC), color = "grey50",
176             alpha = 0.8, size = 0.8) +
177   geom_hline(yintercept = 0, linetype = 1) + #add line at LFC = 0
178   geom_hline(yintercept = c(-LFCcutoff, LFCcutoff), linetype = 2) +
179   geom_point(data = DE_DEL[DE_DEL$Accession %in% proteins_sig],,
180             aes(x = AveExpr, y = logFC, color = logFC > 0), size = 0.8) +
181   scale_color_manual(values = colorsUPDOWN) +
182   scale_fill_manual(values = colorsUPDOWN) +
183   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2],,
184             aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 1) +
185   #add 95% CI lines
186   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2],,
187             aes(x = AveExpr, y = logFC - error, yend = logFC + error,
188               color = logFC > 0), lwd = 0.5) +
189   geom_point(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2 &
190                           is.finite(DE_DEL$logFC)],,
191             aes(x = AveExpr, y = logFC, fill = logFC > 0), size = 3,
192             shape = 21, stroke = 1.5) +
193   geom_point(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC)],,
194             shape = 23, aes(x = AveExpr, size = CountExp, fill = logFC > 0), y = maxY,
195             stroke = 1, alpha = 0.6) +
196   geom_point(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC)],,
197             shape = 23, aes(x = AveExpr, size = CountWT, fill = logFC > 0), y = minY,
198             stroke = 1, alpha = 0.6) +
199   coord_cartesian(ylim = c(minY, maxY)) +
200   scale_size(range = c(1, 3)) +
201   geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteins_label_pos],,
202                   aes(x = AveExpr, y = logFC, label = gene), nudge_y = 1, nudge_x = -1.2

```

```

200     ) +
201     geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteins_label_neg,],
202                     aes(x = AveExpr, y = logFC, label = gene), nudge_y = -1, nudge_x = -
203                       1.2) +
204     geom_label_repel(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC) & DE_DEL$
205       CountExp > 1,],
206                     aes(x = AveExpr, label = gene), y = maxY, nudge_y = -1, nudge_x = -1)
207     +
208     geom_label_repel(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC) & DE_DEL$
209       CountWT > 1,],
210                     aes(x = AveExpr, label = gene), y = minY, nudge_y = 1, nudge_x = 1)
211
212
213
214
215 ggsave("WTvsDEL_MApplot.png", width = 15, height = 12)
216
217 write.table(DE_DEL, "WTvsDELprotein.tsv", sep = "\t", row.names = F)
218
219
220
221 #add in Acetylation changers
222 acetylation <- read.csv("1839_acetylation_reported.tsv", sep = "\t")
223
224 proteinList <- acetylation$Accession2
225
226
227 ggplot() +
228   theme_bw(base_size = 25) +
229   theme(panel.grid = element_blank(),
230         legend.position = "none") +
231   labs(y = "Log2 Fold Change", x = "Normalized Average Expression (Log)",
232        title = expression("WT vs " * Delta * italic(empl))) +
233   geom_point(data = DE_DEL,
234             aes(x = AveExpr, y = logFC), color = "grey50",
235             alpha = 0.8, size = 0.8) +
236   geom_hline(yintercept = 0, linetype = 1) + #add line at LFC = 0
237   geom_hline(yintercept = c(-LFCcutoff, LFCcutoff), linetype = 2) +
238   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
239               aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 1) +
240   #add 95% CI lines
241   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
242               aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 0.5,
243               color = green) +
244   geom_point(data = DE_DEL[DE_DEL$Accession %in% proteinList &
245     is.finite(DE_DEL$logFC),],
246             aes(x = AveExpr, y = logFC), size = 3,
247             fill = green,
248             shape = 21, stroke = 1.5) +
249   geom_point(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC),],
250             shape = 23, aes(x = AveExpr), y = maxY, stroke = 1, alpha = 0.6) +
251   geom_point(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC),],
252             shape = 23, aes(x = AveExpr), y = minY, stroke = 1, alpha = 0.6) +
253   coord_cartesian(ylim = c(minY, maxY)) +
254   scale_size(range = c(1, 3)) +
255   geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
256                   aes(x = AveExpr, y = logFC, label = gene), nudge_y = 3, nudge_x = -1.2
257                   ,
258                   max.overlaps = 40)
259
260
261 ggsave("WTvsDEL_MApplot_acetylationChangers.png", width = 15, height = 12)
262
263
264 #log transform, base 10 for galaxy plots
265 Qprot <- addAssay(Qprot,
266                  logTransform((Qprot[["raw_proteins"]]),
267                              base = 10),
268                  name = "log_proteins_10")
269
270 Qprot <- addAssay(Qprot,

```

```

262         normalize(Qprot[["log_proteins_10"]],
263                 method = "center.median"),
264         name = "norm_proteins_10")
265
266
267
268 #add in old column data to normalized data
269 colData(Qprot[["norm_proteins_10"]]) <- colData(Qprot)
270
271 #pull log 10 data
272 loglog <- Qprot[["norm_proteins_10"]] %>% assay() %>% data.frame() %>% rownames_to_column(
  ("Accession"))
273
274 #pivot longer
275 loglogLong <- loglog %>% pivot_longer(2:ncol(loglog))
276 loglogLong$strain <- str_extract(loglogLong$name, "^[^_]+")
277
278 #average area and count of non na values for each
279 loglogGrouped <- loglogLong %>% group_by(Accession, strain) %>%
280   summarise(Area = mean(value, na.rm = T),
281             count = sum(!is.na(value)))
282
283 #pivot wider for plotting
284 loglogWide <- loglogGrouped %>% pivot_wider(id_cols = Accession,
285                                             values_from = Area,
286                                             names_from = strain)
287
288 #plot
289 ggplot(loglogWide) +
290   geom_point(aes(x = WT, y = d1839), color = "grey50",
291             size = 0.8, alpha = 0.8) +
292   theme_bw(base_size = 25) +
293   theme(panel.grid = element_blank(),
294         legend.position = "none") +
295   geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
296   geom_point(data = loglogWide[loglogWide$Accession %in% proteins_sig2,],
297             aes(x = WT, y = d1839, fill = WT < d1839), size = 3,
298             shape = 21, stroke = 1.5) +
299   scale_fill_manual(values = colorsUPDOWN) +
300   labs(title = expression("WT vs " * Delta * italic(emp1)),
301        x = "Normalized WT expression (log)",
302        y = expression("Normalized " * Delta * italic(emp1) * " expression (log)"))
303
304 ggsave("galaxyWTvsDelProteinDE.png", width = 10, height = 10)
305
306 ggplot(loglogWide) +
307   geom_point(aes(x = WT, y = d1839), color = "grey50",
308             size = 0.8, alpha = 0.8) +
309   theme_bw(base_size = 25) +
310   theme(panel.grid = element_blank(),
311         legend.position = "none") +
312   geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
313   geom_point(data = loglogWide[loglogWide$Accession %in% proteinList,],
314             aes(x = WT, y = d1839), fill = green, size = 3,
315             shape = 21, stroke = 1.5) +
316   scale_fill_manual(values = colorsUPDOWN) +
317   labs(title = expression("WT vs " * Delta * italic(emp1)),
318        x = "Normalized WT expression (log)",
319        y = expression("Normalized " * Delta * italic(emp1) * " expression (log)")) +
320   geom_label_repel(data = loglogWide[loglogWide$Accession %in% proteinList,],
321                   aes(x = WT, y = d1839, label = Accession), nudge_y = 1, nudge_x = -1,
322                   max.overlaps = 40)
323
324
325 ggsave("galaxyWTvsDelAcetylationChangers_labs.png", width = 10, height = 10)
326
327 ggplot(loglogWide) +
328   geom_point(aes(x = WT, y = d1839), color = "grey50",
329             size = 0.8, alpha = 0.8) +

```

```

330 theme_bw(base_size = 25) +
331 theme(panel.grid = element_blank(),
332       legend.position = "none") +
333 geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
334 geom_point(data = loglogWide[loglogWide$Accession %in% proteinList,],
335           aes(x = WT, y = d1839), fill = green, size = 3,
336           shape = 21, stroke = 1.5) +
337 geom_point(data = loglogWide[loglogWide$Accession %in% proteins_sig2,],
338           aes(x = WT, y = d1839, fill = WT < d1839), size = 3,
339           shape = 21, stroke = 1.5) +
340 scale_fill_manual(values = colorsUPDOWN) +
341 labs(title = expression("WT vs " * Delta * italic(empl)),
342      x = "Normalized WT expression (log)",
343      y = expression("Normalized " * Delta * italic(empl) * " expression (log)"))
344
345 ggsave("galaxyWTvsDelAcetylationChangers_andDE.png", width = 10, height = 10)
346
347
348 #SAME as above but for complement instea of Del
349 #fit proteins to linear model with lmFit and the matrix we just created
350 fitted_lm <- proteinsBR %>%
351   assay() %>%
352   lmFit(design = model_design) %>%
353   eBayes()
354
355 DE_DEL <- topTable(fit = fitted_lm,
356                  adjust.method = "BH",
357                  number = Inf,
358                  coef = "proteinsBR$conditionc1839",
359                  sort.by = "p",
360                  confint = T)
361
362
363 DE_DEL$Accession <- rownames(DE_DEL)
364
365
366
367 #calculate corrected z score based on B-H adujsted p-value. Altman,
368 #https://doi.org/10.1136/bmj.d2090
369 DE_DEL$corrected_z <- sqrt(0.743 - (2.404*log(DE_DEL$adj.P.Val))) - 0.862
370 #calculate the one directional 95% confidence interval
371 DE_DEL$error <- abs(DE_DEL$logFC /DE_DEL$corrected_z) * 1.96
372
373
374 DE_DEL <- DE_DEL[!is.na(DE_DEL$AveExpr),]
375 DE_DEL$CountWT <- NA
376 DE_DEL$CountExp <- NA
377
378 for (i in 1:nrow(DE_DEL)) {
379   CountDF <- brSummaryWide[brSummaryWide$Accession == DE_DEL$Accession[i],]
380   DE_DEL$CountWT[i] <- sum(!is.na(CountDF[,grepl("WT", names(CountDF))][1,]))
381   DE_DEL$CountExp[i] <- sum(!is.na(CountDF[,grepl("c1839", names(CountDF))][1,]))
382 }
383
384 DE_DEL$logFC[DE_DEL$CountWT == 0] <- Inf
385 DE_DEL$logFC[DE_DEL$CountExp == 0] <- -Inf
386
387
388 adjSigCutoff <- 0.05
389 NonAdjSigCutoff <- 0.05
390 LFCcutoff <- 1
391
392
393
394 adj <- T
395
396 if (adj) {
397   proteins_sig <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff]))

```

```

398 proteins_sig2 <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
399                               abs(DE_DEL$logFC) > LFCcutoff]))
400 proteins_label_pos <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
401                               DE_DEL$logFC > LFCcutoff]))
402 proteins_label_neg <- na.omit(unique(DE_DEL$Accession[DE_DEL$adj.P.Val < adjSigCutoff &
403                               DE_DEL$logFC < -LFCcutoff]))
404 } else {
405 proteins_sig <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff]))
406 proteins_sig2 <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
407                               abs(DE_DEL$logFC) > LFCcutoff]))
408 proteins_label_pos <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
409                               DE_DEL$logFC > LFCcutoff]))
410 proteins_label_neg <- na.omit(unique(DE_DEL$Accession[DE_DEL$P.Value < adjSigCutoff &
411                               DE_DEL$logFC < -LFCcutoff]))
412 }
413
414
415
416
417
418 #add in gene and protein name info from saved dictionaries
419 DE_DEL$gene <- sapply(DE_DEL$Accession, function(Accession) geneDict[[Accession]])
420
421
422
423 plotBufferFactor <- 1.5
424 maxY <- max(DE_DEL$logFC[is.finite(DE_DEL$logFC)]) * plotBufferFactor
425 minY <- min(DE_DEL$logFC[is.finite(DE_DEL$logFC)]) * plotBufferFactor
426
427
428 ggplot() +
429   theme_bw(base_size = 25) +
430   theme(panel.grid = element_blank(),
431         legend.position = "none") +
432   labs(y = "Log2 Fold Change", x = "Normalized Average Expression (Log)",
433        title = expression("WT vs " * Delta * italic(empl) * "/"comp")) +
434   geom_point(data = DE_DEL[!DE_DEL$Accession %in% proteins_sig &
435                             is.finite(DE_DEL$logFC),],
436             aes(x = AveExpr, y = logFC), color = "grey50",
437             alpha = 0.8, size = 0.8) +
438   geom_hline(yintercept = 0, linetype = 1) + #add line at LFC = 0
439   geom_hline(yintercept = c(-LFCcutoff, LFCcutoff), linetype = 2) +
440   geom_point(data = DE_DEL[DE_DEL$Accession %in% proteins_sig,],
441             aes(x = AveExpr, y = logFC, color = logFC > 0), size = 0.8) +
442   scale_color_manual(values = colorsUPDOWN) +
443   scale_fill_manual(values = colorsUPDOWN) +
444   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2,],
445             aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 1) +
446   #add 95% CI lines
447   geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2,],
448             aes(x = AveExpr, y = logFC - error, yend = logFC + error,
449                 color = logFC > 0), lwd = 0.5) +
449   geom_point(data = DE_DEL[DE_DEL$Accession %in% proteins_sig2 &
450                             is.finite(DE_DEL$logFC),],
451             aes(x = AveExpr, y = logFC, fill = logFC > 0), size = 3,
452             shape = 21, stroke = 1.5) +
453   geom_point(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC),],
454             shape = 23, aes(x = AveExpr, size = CountExp, fill = logFC > 0), y = maxY,
455             stroke = 1, alpha = 0.6) +
456   geom_point(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC),],
457             shape = 23, aes(x = AveExpr, size = CountWT, fill = logFC > 0), y = minY,
458             stroke = 1, alpha = 0.6) +
459   coord_cartesian(ylim = c(minY, maxY)) +
460   scale_size(range = c(1, 3)) +
461   geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteins_label_pos,],
462                   aes(x = AveExpr, y = logFC, label = gene), nudge_y = 1, nudge_x = -1.2) +
463   geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteins_label_neg,],
464                   aes(x = AveExpr, y = logFC, label = gene), nudge_y = -1, nudge_x = -

```



```

1.2) +
463 geom_label_repel(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC) & DE_DEL$
CountExp > 1,],
464 aes(x = AveExpr, label = gene), y = maxY, nudge_y = -1, nudge_x = -1)
+
465 geom_label_repel(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC) & DE_DEL$
CountWT > 1,],
466 aes(x = AveExpr, label = gene), y = minY, nudge_y = 1, nudge_x = 1)
467
468
469
470 ggsave("WTvsComp_MApplot.png", width = 15, height = 12)
471 write.table(DE_DEL, "WTvsCOMPprotein.tsv", sep = "\t", row.names = F)
472
473
474
475 proteinList <- acetylation$Accession2
476
477
478 ggplot() +
479 theme_bw(base_size = 25) +
480 theme(panel.grid = element_blank(),
481 legend.position = "none") +
482 labs(y = "Log2 Fold Change", x = "Normalized Average Expression (Log)",
483 title = expression("WT vs " * Delta * italic(empl) * "/comp")) +
484 geom_point(data = DE_DEL,
485 aes(x = AveExpr, y = logFC), color = "grey50",
486 alpha = 0.8, size = 0.8) +
487 geom_hline(yintercept = 0, linetype = 1) + #add line at LFC = 0
488 geom_hline(yintercept = c(-LFCcutoff, LFCcutoff), linetype = 2) +
489 geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
490 aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 1) +
491 #add 95% CI lines
492 geom_segment(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
493 aes(x = AveExpr, y = logFC - error, yend = logFC + error), lwd = 0.5,
494 color = green) +
495 geom_point(data = DE_DEL[DE_DEL$Accession %in% proteinList &
496 is.finite(DE_DEL$logFC),],
497 aes(x = AveExpr, y = logFC), size = 3,
498 fill = green,
499 shape = 21, stroke = 1.5) +
500 geom_point(data = DE_DEL[DE_DEL$logFC > 0 & !is.finite(DE_DEL$logFC),],
501 shape = 23, aes(x = AveExpr), y = maxY, stroke = 1, alpha = 0.6) +
502 geom_point(data = DE_DEL[DE_DEL$logFC < 0 & !is.finite(DE_DEL$logFC),],
503 shape = 23, aes(x = AveExpr), y = minY, stroke = 1, alpha = 0.6) +
504 coord_cartesian(ylim = c(minY, maxY)) +
505 scale_size(range = c(1, 3)) +
506 geom_label_repel(data = DE_DEL[DE_DEL$Accession %in% proteinList,],
507 aes(x = AveExpr, y = logFC, label = gene), nudge_y = 3, nudge_x = -1.2
508 ,
509 max.overlaps = 40)
510
511
512
513
514 ggsave("WTvsCOMP_MApplot_acetylationChangers.png", width = 15, height = 12)
515
516
517
518 Qprot <- addAssay(Qprot,
519 logTransform((Qprot[["raw_proteins"]]),
520 base = 10),
521 name = "log_proteins_10")
522
523 Qprot <- addAssay(Qprot,
524 normalize(Qprot[["log_proteins_10"]],
525 method = "center.median"),
526 name = "norm_proteins_10")
527
528 #add in old column data to normalized data

```



```

526 colData(Qprot[["norm_proteins_10"]]) <- colData(Qprot)
527
528
529 loglog <- Qprot[["norm_proteins_10"]] %>% assay() %>% data.frame() %>% rownames_to_column
  ("Accession")
530
531 loglogLong <- loglog %>% pivot_longer(2:ncol(loglog))
532 loglogLong$strain <- str_extract(loglogLong$name, "^[^_]+")
533
534 loglogGrouped <- loglogLong %>% group_by(Accession, strain) %>%
535   summarise(Area = mean(value, na.rm = T),
536     count = sum(!is.na(value)))
537
538
539 loglogWide <- loglogGrouped %>% pivot_wider(id_cols = Accession,
540   values_from = Area,
541   names_from = strain)
542
543 ggplot(loglogWide) +
544   geom_point(aes(x = WT, y = c1839), color = "grey50",
545     size = 0.8, alpha = 0.8) +
546   theme_bw(base_size = 25) +
547   theme(panel.grid = element_blank(),
548     legend.position = "none") +
549   geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
550   geom_point(data = loglogWide[loglogWide$Accession %in% proteins_sig2,],
551     aes(x = WT, y = c1839, fill = WT < c1839), size = 3,
552     shape = 21, stroke = 1.5) +
553   scale_fill_manual(values = colorsUPDOWN) +
554   labs(title = expression("WT vs " * Delta * italic(emp1) * "/comp"),
555     x = "Normalized WT expression (log)",
556     y = expression("Normalized " * Delta * italic(emp1) * "/comp expression (log)"))
557
558 ggsave("galaxyWTvsCOMPPProteinDE.png", width = 10, height = 10)
559
560
561 ggplot(loglogWide) +
562   geom_point(aes(x = WT, y = c1839), color = "grey50",
563     size = 0.8, alpha = 0.8) +
564   theme_bw(base_size = 25) +
565   theme(panel.grid = element_blank(),
566     legend.position = "none") +
567   geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
568   geom_point(data = loglogWide[loglogWide$Accession %in% proteinList,],
569     aes(x = WT, y = c1839), fill = green, size = 3,
570     shape = 21, stroke = 1.5) +
571   scale_fill_manual(values = colorsUPDOWN) +
572   labs(title = expression("WT vs " * Delta * italic(emp1) * "/comp"),
573     x = "Normalized WT expression (log)",
574     y = expression("Normalized " * Delta * italic(emp1) * "/comp expression (log)")) +
575   geom_label_repel(data = loglogWide[loglogWide$Accession %in% proteinList,],
576     aes(x = WT, y = c1839, label = Accession), nudge_y = 1, nudge_x = -1,
577     max.overlaps = 40)
578
579 ggsave("galaxyWTvsCOMPAcetylationChangers_labs.png", width = 10, height = 10)
580
581 ggplot(loglogWide) +
582   geom_point(aes(x = WT, y = c1839), color = "grey50",
583     size = 0.8, alpha = 0.8) +
584   theme_bw(base_size = 25) +
585   theme(panel.grid = element_blank(),
586     legend.position = "none") +
587   geom_abline(slope = 1, intercept = 0, linetype = 2, size = 1) +
588   geom_point(data = loglogWide[loglogWide$Accession %in% proteinList,],
589     aes(x = WT, y = c1839), fill = green, size = 3,
590     shape = 21, stroke = 1.5) +
591   geom_point(data = loglogWide[loglogWide$Accession %in% proteins_sig2,],
592     aes(x = WT, y = c1839, fill = WT < c1839), size = 3,
593     shape = 21, stroke = 1.5) +

```

```
594     scale_fill_manual(values = colorsUPDOWN) +
595     labs(title = expression("WT vs " * Delta * italic(empl) * "/comp"),
596          x = "Normalized WT expression (log)",
597          y = expression("Normalized " * Delta * italic(empl) * "/comp expression (log)"))
598
599 ggsave("galaxyWTvsCOMPAcetylationChangers_andDE.png", width = 10, height = 10)
600
601
602
603
```