

```

1  setwd("~/04_Champion_Lab/02_N-terminal_Acetylation/OnePot/NAT5")
2
3  #import libraries
4  library(ggplot2)
5  library(stringr)
6  library(dplyr)
7  library(tidyr)
8  library(ggrepel)
9  library(scales)
10
11  #set colors
12  red <- hue_pal()(3)[1]
13  green <- hue_pal()(3)[2]
14  blue <- hue_pal()(3)[3]
15
16  #import peptides
17  peptides <- read.csv("db.protein-peptides.csv")
18
19  #filter out contaminants and clean sequences
20  peptides <- peptides[!grepl("CONTAM", peptides$Accession),]
21  peptides$Peptide <- str_remove(peptides$Peptide, "^\\.\\.")
22  peptides$Peptide <- str_remove(peptides$Peptide, "\\..$")
23
24  #identify peptides as light, heavy or unlabelled
25  peptides$NtermAcet_L <- grepl("^[A-Z]\\(\\+42\\.01\\)", peptides$Peptide)
26  peptides$NtermAcet_H <- grepl("^[A-Z]\\(\\+45\\.03\\)", peptides$Peptide)
27  peptides$UnLabelled <- !peptides$NtermAcet_H & !peptides$NtermAcet_L
28
29  #convert to long data format (one row per inj/pep)
30  peptidesLong <- peptides %>% pivot_longer(which(grepl("^Area.L", names(peptides))),
31                                           values_to = "Area",
32                                           names_to = "sample")
33
34  #remove extra columns
35  drops <- c("USED", "Quality", "Significance", "Avg..ppm", "X1.K0", "Avg..Area",
36            "Sample.Profile..Ratio.", "Area.WT", "Area.Del",
37            "Area.Comp", "Group.Profile..Ratio.", "Max.Ratio", "X.Vector",
38            names(peptides)[grepl("Intensity.", names(peptides))],
39            names(peptides)[grepl("X.Spec", names(peptides))],
40            names(peptides)[grepl("Area\\.\\.", names(peptides))])
41  peptidesLong <- peptidesLong[,!names(peptidesLong) %in% drops]
42
43  #parse meta data
44  peptidesLong$Inj <- str_remove(peptidesLong$sample, "Area.L_x[0-9]+_")
45  peptidesLong$strain <- str_remove(str_extract(peptidesLong$Inj, "^[^_]+_"), "_")
46  peptidesLong$Biorep <- str_extract(peptidesLong$Inj, "br[0-9]")
47  peptidesLong$Techrep <- str_extract(peptidesLong$Inj, "tr[0-9]")
48
49  #pull N-terminal peptides
50  NacetpeptidesLong <- peptidesLong[peptidesLong$Start <= 2,]
51
52  #parse accession and gene info
53  NacetpeptidesLong$Accession2 <- str_remove(str_extract(NacetpeptidesLong$Accession,
54                                                         "^[^\\|]*\\|"), "\\|")
55  NacetpeptidesLong$gene <- str_remove(str_remove(str_extract(NacetpeptidesLong$Accession,
56                                                         "^[^\\|]*\\|"), "\\|"),
57                                         "^[^\\|]*\\|", "\\|")
58
59  #pull out list of accessions with light N-terminal acetylation
60  accessionsWithNtermAcet <- unique(NacetpeptidesLong$Accession2[NacetpeptidesLong$
61                                NtermAcet_L])
62  NacetpeptidesLong <- NacetpeptidesLong[NacetpeptidesLong$Accession2 %in%
63                                accessionsWithNtermAcet,]
64
65  #create sequence without heavy or light for peptide pair matching
66  NacetpeptidesLong$acetSeq[NacetpeptidesLong$NtermAcet_L] <- str_replace(NacetpeptidesLong
67                                $Peptide[NacetpeptidesLong$NtermAcet_L],
68                                "\\(\\+42\\.01\\)",
69                                "",

```

```

63                                                                                                     "(ac)")
64 NacetylpeptidesLong$acetSeq[NacetylpeptidesLong$NtermAcet_H] <- str_replace(NacetylpeptidesLong
65 $Peptide[NacetylpeptidesLong$NtermAcet_H],
                                                                                                     "\\(\\+45\\.03\\)\\)
                                                                                                     ",
66                                                                                                     "(ac)")
67
68 NacetylpeptidesLong$acetSeq[NacetylpeptidesLong$Unlabelled] <- NacetylpeptidesLong$Peptide[
69 NacetylpeptidesLong$Unlabelled]
70 NacetylpeptidesLong$noAcetSeq <- str_remove(NacetylpeptidesLong$acetSeq, "\\(ac\\)")
71 #strain variables
72 STRAIN <- "1839"
73 DELSTRAIN <- paste0("d", STRAIN)
74 COMPSTRAIN <- paste0("c", STRAIN)
75
76
77 #pull out peptides with these strains of interest (search was with other injections as
78 well)
79 peps <- NacetylpeptidesLong[NacetylpeptidesLong$strain == "WT" |
80 grepl(STRAIN, NacetylpeptidesLong$strain),]
81 #categorize peptides
82 peps$category <- ifelse(peps$Unlabelled, "Unlabelled",
83 ifelse(peps$NtermAcet_H, "Heavy", "Light"))
84 #group peptides and calculate the mean area for each one in each category
85 peps2 <- peps %>% group_by(Peptide, category) %>%
86 mutate(rows = n(),
87 IDs = sum(!is.na(Area)),
88 meanArea = mean(Area, na.rm = T))
89
90 #only light peptides
91 lights <- peps2[peps2$NtermAcet_L,]
92
93 #decide if the peptide is the best flier for each accession based on largest mean area
94 peps3 <- lights %>% group_by(Accession2, category) %>%
95 mutate(bestFlierArea = max(meanArea, na.rm = T),
96 bestFlier = bestFlierArea == meanArea)
97 #pull out best fliers
98 bestFliers <- peps3[peps3$bestFlier &
99 !is.na(peps3$bestFlier),]
100
101
102
103 #find the heavy version of each best flier and pull the area.
104 bestFliers$HeavyArea <- NA
105 for (i in 1:nrow(bestFliers)) {
106 temp <- peps[peps$Inj == bestFliers$Inj[i] &
107 peps$acetSeq == bestFliers$acetSeq[i] &
108 peps$NtermAcet_H &
109 peps$Accession2 == bestFliers$Accession2[i],]
110 if (nrow(temp) == 1) {
111 bestFliers$HeavyArea[i] <- temp$Area[1]
112 } else if (nrow(temp) < 1) {
113 bestFliers$HeavyArea[i] <- NA
114 } else {
115 print(nrow(temp))
116 print(i)
117 print(bestFliers$acetSeq[i])
118 }
119 }
120
121 #calculat pct acet (light / (light + heavy )) * 100
122 bestFliers$pctAcet <- NA
123 bestFliers$pctAcet[!is.na(bestFliers$Area) &
124 !is.na(bestFliers$HeavyArea)] <- 100 *
125 (bestFliers$Area[!is.na(bestFliers$Area) &
126 !is.na(bestFliers$HeavyArea)] / (bestFliers$Area[!is.na(bestFliers$

```

```

127                                     Area) &                                     !is.na(bestFliers
128                                     $HeavyArea)] +
129                                     bestFliers$HeavyArea[!is.na(
130                                     bestFliers$Area) &                                     !is.na(
131                                     bestFliers$HeavyArea))])
132
133 #for NAs, set 0 and 100 % depending on whether heavy or light is NA
134 bestFliers$pctAcet[!is.na(bestFliers$Area) &
135                       is.na(bestFliers$HeavyArea)] <- 100
136
137 bestFliers$pctAcet[is.na(bestFliers$Area) &
138                       !is.na(bestFliers$HeavyArea)] <- 0
139
140
141
142 #combine all bioreps by protein and peptide sequence
143 #calculate mean area of each strain and each heavy/light peptide
144 #count the number of Non NA values for each
145 #calculate sd of area
146 #calculate mean and sd of pct acet for each strain
147 #calculate fold chain betwee del/wt and del/comp
148 #t-tests for each as well, based on pct acet between del/wt and del/comp (only for valid
149 ones)
150 CombinedTandBreps <- bestFliers %>% group_by(Accession2, gene, acetSeq) %>%
151   summarise(mean_HeavyArea_WT = mean(HeavyArea[strain == "WT", na.rm = T],
152   mean_Area_WT = mean(Area[strain == "WT", na.rm = T],
153   mean_HeavyArea_DEL = mean(HeavyArea[strain == DELSTRAIN, na.rm = T],
154   mean_Area_DEL = mean(Area[strain == DELSTRAIN, na.rm = T],
155   mean_HeavyArea_COMP = mean(HeavyArea[strain == COMPSTRAIN, na.rm = T],
156   mean_Area_COMP = mean(Area[strain == COMPSTRAIN, na.rm = T],
157   count_H_WT = sum(!is.na(HeavyArea[strain == "WT"])),
158   count_L_WT = sum(!is.na(Area[strain == "WT"])),
159   count_H_DEL = sum(!is.na(HeavyArea[strain == DELSTRAIN])),
160   count_L_DEL = sum(!is.na(Area[strain == DELSTRAIN])),
161   count_H_COMP = sum(!is.na(HeavyArea[strain == COMPSTRAIN])),
162   count_L_COMP = sum(!is.na(Area[strain == COMPSTRAIN])),
163   sd_HeavyArea_WT = sd(HeavyArea[strain == "WT", na.rm = T],
164   sd_Area_WT = sd(Area[strain == "WT", na.rm = T],
165   sd_HeavyArea_DEL = sd(HeavyArea[strain == DELSTRAIN, na.rm = T],
166   sd_Area_DEL = sd(Area[strain == DELSTRAIN, na.rm = T],
167   sd_HeavyArea_COMP = sd(HeavyArea[strain == COMPSTRAIN, na.rm = T],
168   sd_Area_COMP = sd(Area[strain == COMPSTRAIN, na.rm = T],
169   mean_pctAcet_WT = mean(pctAcet[strain == "WT", na.rm = T],
170   mean_pctAcet_DEL = mean(pctAcet[strain == DELSTRAIN, na.rm = T],
171   mean_pctAcet_COMP = mean(pctAcet[strain == COMPSTRAIN, na.rm = T],
172   sd_pctAcet_WT = sd(pctAcet[strain == "WT", na.rm = T],
173   sd_pctAcet_DEL = sd(pctAcet[strain == DELSTRAIN, na.rm = T],
174   sd_pctAcet_COMP = sd(pctAcet[strain == COMPSTRAIN, na.rm = T],
175   count_pctAcet_WT = sum(!is.na(pctAcet[strain == "WT"])),
176   count_pctAcet_DEL = sum(!is.na(pctAcet[strain == DELSTRAIN])),
177   count_pctAcet_COMP = sum(!is.na(pctAcet[strain == COMPSTRAIN])),
178   FC_pctAcet_WTDEL = mean_pctAcet_WT / mean_pctAcet_DEL,
179   FC_pctAcet_COMPDEL = mean_pctAcet_COMP / mean_pctAcet_DEL,
180   pval_WTDEL = ifelse(is.character(try(t.test(pctAcet[strain == "WT",
181   pctAcet[strain == DELSTRAIN]))[[3]]),
182   silent = T)),
183   NA, t.test(pctAcet[strain == "WT",
184   pctAcet[strain == DELSTRAIN]][[3]]),
185   pval_COMPDEL = ifelse(is.character(try(t.test(pctAcet[strain == COMPSTRAIN],
186   pctAcet[strain == DELSTRAIN]))[[3]]),
187   silent = T)),
188   NA, t.test(pctAcet[strain == COMPSTRAIN],
189   pctAcet[strain == DELSTRAIN]][[3]])

```

```

189
190 )
191
192
193
194 #require at least 2 obs of WT and Comp (out of 6)
195 minWTandCompObs <- 2
196 CombinedTandBreps2 <- CombinedTandBreps[CombinedTandBreps$count_pctAcet_WT >=
197                                     CombinedTandBreps$count_pctAcet_COMP >=
                                     minWTandCompObs,]
198
199 #manually set a new variable for pvalue filtering, for values of 100%WT and 0 % Del set
to arbitrary significant
200 CombinedTandBreps2$pvalCategoryWTDEL <- CombinedTandBreps2$pval_WTDEL
201 CombinedTandBreps2$pvalCategoryWTDEL[CombinedTandBreps2$mean_pctAcet_WT == 100 &
202                                     CombinedTandBreps2$mean_pctAcet_DEL == 0] <- 0.001
203
204 #calculate %RSD of pct acet
205 #if there are missing values, use appropriate area to calculate instead of pct acet
206 CombinedTandBreps2$RSD_pctAcet_WT <- CombinedTandBreps2$sd_pctAcet_WT /
CombinedTandBreps2$mean_pctAcet_WT
207 for (i in 1:nrow(CombinedTandBreps2)) {
208   if ((CombinedTandBreps2$RSD_pctAcet_WT[i] == 0 &
209       !is.na(CombinedTandBreps2$RSD_pctAcet_WT[i])) |
210       (CombinedTandBreps2$mean_pctAcet_WT[i] == 0 &
211       !is.na(CombinedTandBreps2$mean_pctAcet_WT[i]))) {
212     if (CombinedTandBreps2$mean_pctAcet_WT[i] == 100) {
213       CombinedTandBreps2$RSD_pctAcet_WT[i] <- CombinedTandBreps2$sd_Area_WT[i] /
214       CombinedTandBreps2$mean_Area_WT[i]
215     } else if (CombinedTandBreps2$mean_pctAcet_WT[i] == 0) {
216       CombinedTandBreps2$RSD_pctAcet_WT[i] <- CombinedTandBreps2$sd_HeavyArea_WT[i] /
217       CombinedTandBreps2$mean_HeavyArea_WT[i]
218     }
219   }
220 }
221
222 CombinedTandBreps2$RSD_pctAcet_DEL <- CombinedTandBreps2$sd_pctAcet_DEL /
CombinedTandBreps2$mean_pctAcet_DEL
223 for (i in 1:nrow(CombinedTandBreps2)) {
224   if ((CombinedTandBreps2$RSD_pctAcet_DEL[i] == 0 &
225       !is.na(CombinedTandBreps2$RSD_pctAcet_DEL[i])) |
226       (CombinedTandBreps2$mean_pctAcet_DEL[i] == 0 &
227       !is.na(CombinedTandBreps2$mean_pctAcet_DEL[i]))) {
228     if (CombinedTandBreps2$mean_pctAcet_DEL[i] == 100) {
229       CombinedTandBreps2$RSD_pctAcet_DEL[i] <- CombinedTandBreps2$sd_Area_DEL[i] /
230       CombinedTandBreps2$mean_Area_DEL[i]
231     } else if (CombinedTandBreps2$mean_pctAcet_DEL[i] == 0) {
232       CombinedTandBreps2$RSD_pctAcet_DEL[i] <- CombinedTandBreps2$sd_HeavyArea_DEL[i] /
233       CombinedTandBreps2$mean_HeavyArea_DEL[i]
234     }
235   }
236 }
237
238 CombinedTandBreps2$RSD_pctAcet_COMP <- CombinedTandBreps2$sd_pctAcet_COMP /
CombinedTandBreps2$mean_pctAcet_COMP
239 for (i in 1:nrow(CombinedTandBreps2)) {
240   if ((CombinedTandBreps2$RSD_pctAcet_COMP[i] == 0 &
241       !is.na(CombinedTandBreps2$RSD_pctAcet_COMP[i])) |
242       (CombinedTandBreps2$mean_pctAcet_COMP[i] == 0 &
243       !is.na(CombinedTandBreps2$mean_pctAcet_COMP[i]))) {
244     if (CombinedTandBreps2$mean_pctAcet_COMP[i] == 100) {
245       CombinedTandBreps2$RSD_pctAcet_COMP[i] <- CombinedTandBreps2$sd_Area_COMP[i] /
246       CombinedTandBreps2$mean_Area_COMP[i]
247     } else if (CombinedTandBreps2$mean_pctAcet_COMP[i] == 0) {
248       CombinedTandBreps2$RSD_pctAcet_COMP[i] <- CombinedTandBreps2$sd_HeavyArea_COMP[i] /
249       CombinedTandBreps2$mean_HeavyArea_COMP[i]
250     }
251   }
252 }

```

```

252 }
253
254 #calculate the absolute percentage change between WT and del
255 CombinedTandBreps2$absPctChange_WTvDel <- abs(CombinedTandBreps2$mean_pctAcet_WT -
CombinedTandBreps2$mean_pctAcet_DEL)
256
257 #export
258 write.table(CombinedTandBreps2, "AllMeasuredAcetylation1839.tsv", sep = "\t", row.names =
F)
259
260 #create plotting dataframe requiring more acetlation in WT and comp than del
261 plot <- CombinedTandBreps2[CombinedTandBreps2$FC_pctAcet_WTDEL > 1 &
262 !is.na(CombinedTandBreps2$FC_pctAcet_WTDEL) &
263 CombinedTandBreps2$FC_pctAcet_COMPDEL > 1 &
264 !is.na(CombinedTandBreps2$FC_pctAcet_COMPDEL),]
265
266 #filters. At least 20 absolute pct diff between WT and del
267 #max p val of 0.05
268 abs_pct_change_min <- 20
269 pval_cutoff <- 0.05
270
271 #filtering
272 plot <- plot[plot$absPctChange_WTvDel >= abs_pct_change_min,]
273 plot <- plot[plot$pvalCategoryWTDEL <= pval_cutoff,]
274
275 #pivot to longer format for plotting
276 plotLong <- plot %>% pivot_longer(cols = (starts_with("mean_pctAcet_") |
277 starts_with("count_pctAcet_") |
278 starts_with("sd_pctAcet_") |
279 starts_with("RSD_pctAcet_")),
280 names_to = c(".value", "strain"),
281 names_pattern = "(.*)_pctAcet_(.*)")
282
283 #set order
284 plotLong$strain <- factor(plotLong$strain, levels = c("WT", "DEL", "COMP"))
285
286 #set NA RSDs to high (1)
287 plotLong$RSD[is.na(plotLong$RSD)] <- 1
288
289 #plot
290 ggplot(plotLong) +
291   geom_line(aes(x = strain, y = mean,
292                 group = acetSeq, color = gene),
293             lwd = 1.2) +
294   geom_point(aes(x = strain, y = mean,
295                  group = acetSeq, fill = gene,
296                  size = ifelse(RSD < 0.2,
297                                "<20%", ifelse(RSD < 0.5,
298                                                  "<50%", ">50%")),
299                  shape = ifelse(pvalCategoryWTDEL <= 0.001,
300                                "<= 0.001", ifelse(pvalCategoryWTDEL <= 0.01,
301                                                       "<= 0.01", ifelse(pvalCategoryWTDEL <=
302                                                       0.05,
303                                                       "<= 0.05", "> 0.05"
304                                                       )))),
305             stroke = 1.4) +
306   theme_bw(base_size = 20) +
307   theme(legend.position = "right",
308         panel.grid = element_blank()) +
309   labs(y = "Percent N-Acetylation",
310        x = element_blank(),
311        shape = "p-value\n(WT vs DEL)",
312        size = "%RSD",
313        title = paste0(STRAIN, " (2 Bio Rep x 3 Tech Rep)")) +
314   scale_colour_discrete(guide = "none") +
315   scale_fill_discrete(guide = "none") +
316   scale_size_manual(values = c(5, 3.5, 2)) +
317   scale_shape_manual(values = c(21, 22, 23, 24)) +
318   geom_label_repel(data = plotLong[plotLong$strain == "WT",],

```

```

317         aes(x = strain, y = mean,
318             label = gene, color = gene),
319         nudge_x = -1.3, alpha = 0.9,
320         size = 3.5,
321         max.overlaps = 50) +
322 coord_cartesian(ylim = c(0,120))
323
324
325
326
327
328 #plot
329 ggplot(plotLong) +
330   geom_line(aes(x = strain, y = mean,
331                 group = acetSeq, color = ifelse(pvalCategoryWTDEL <= 0.001,
332                                                  "<= 0.001", ifelse(pvalCategoryWTDEL <=
333                                                                0.01,
334                                                                "<= 0.01", ifelse(
335                                                                    pvalCategoryWTDEL <=
336                                                                        0.05,
337                                                                        "<=
338                                                                0.05", "> 0.05")))),
339           lwd = 1.2) +
340   geom_point(aes(x = strain, y = mean,
341                  group = acetSeq,
342                  size = ifelse(RSD < 0.2,
343                               "<20%", ifelse(RSD < 0.5,
344                                               "<50%", ">50%")),
345                  fill = ifelse(pvalCategoryWTDEL <= 0.001,
346                                "<= 0.001", ifelse(pvalCategoryWTDEL <= 0.01,
347                                                       "<= 0.01", ifelse(pvalCategoryWTDEL <=
348                                                           0.05,
349                                                           "<= 0.05", "> 0.05"
350                                                           )))),
351              stroke = 1.4, shape = 21) +
352   theme_bw(base_size = 30) +
353   theme(legend.position = "right",
354         panel.grid = element_blank()) +
355   labs(y = "Percent N-Acetylation",
356        x = element_blank(),
357        color = "p-value\n(WT vs DEL)",
358        fill = "p-value\n(WT vs DEL)",
359        size = "%RSD",
360        title = element_blank()) +
361   scale_color_discrete(guide = "none") +
362   scale_size_manual(values = c(5,3.5,2)) +
363   geom_label_repel(data = plotLong[plotLong$strain == "WT",],
364                   aes(x = strain, y = mean,
365                       label = gene, color = ifelse(pvalCategoryWTDEL <= 0.001,
366                                                       "<= 0.001", ifelse(pvalCategoryWTDEL
367                                                           <= 0.01,
368                                                           "<= 0.01", ifelse(
369                                                               pvalCategoryWTDEL
370                                                                   <= 0.05,
371                                                                   "<= 0.05", "> 0.05")))),
372                   nudge_x = -1.3,
373                   size = 3.5,
374                   max.overlaps = 50) +
375   coord_cartesian(ylim = c(0,120)) +
376   scale_y_continuous(breaks = c(0, 25, 50, 75, 100))
377
378 ggsave("1839_acetylation_change_20pctAbs_pval0_05.png", width = 15, height = 12)
379 #export plotted table
380 write.table(plot, "1839_acetylation_reported.tsv", sep = "\t", row.names = F)
381
382 #categorize rsd
383 plotLong$RSDCat <- ifelse(plotLong$RSD < 0.2,
384                            "<20%", ifelse(plotLong$RSD < 0.5,

```

```

376                                     "<50%", ">50%"))
377 #plot one color
378 ggplot(plotLong) +
379   geom_hline(yintercept = c(0,100), linetype = 2, lwd = 1, color = "grey") +
380   geom_line(aes(x = strain, y = mean,
381                 group = acetSeq),
382             lwd = 1.4, color = "black") +
383   geom_line(aes(x = strain, y = mean,
384                 group = acetSeq),
385             lwd = 0.8, color = green) +
386   geom_point(aes(x = strain, y = mean,
387                  group = acetSeq,
388                  size = RSDCat,
389                  shape = RSDCat),
390              stroke = 1.4,
391              fill = green) +
392   theme_bw(base_size = 30) +
393   theme(legend.position = "right",
394         panel.grid = element_blank()) +
395   labs(y = "Percent N-Acetylation",
396        x = element_blank(),
397        color = "p-value\n(WT vs DEL)",
398        fill = "p-value\n(WT vs DEL)",
399        size = "%RSD",
400        shape = "%RSD",
401        title = element_blank()) +
402   scale_color_discrete(guide = "none") +
403   scale_size_manual(values = c(6,4,2)) +
404   geom_label_repel(data = plotLong[plotLong$strain == "WT",],
405                   aes(x = strain, y = mean,
406                      label = gene),
407                   nudge_x = -1.3,
408                   size = 3.5,
409                   max.overlaps = 50) +
410   coord_cartesian(ylim = c(0,120)) +
411   scale_y_continuous(breaks = c(0, 25, 50, 75, 100)) +
412   scale_shape_manual(values = c(21,22,23)) +
413   scale_x_discrete(labels = c("WT", expression(Delta * italic(empl)),
414                               expression(Delta * italic(empl)* "/comp"))))
415
416 ggsave("1839_acetylation_change_20pctAbs_pval0_05_ONECOLOR.png", width = 15, height = 12)
417
418
419 #nonCanonicalN-terms
420 #pull accessions and genes
421 peptidesLong$Accession2 <- str_remove(str_extract(peptidesLong$Accession, "[^\\|]*\\|"),
422   "\\|")
423
424 peptidesLong$gene <- str_remove(str_remove(str_extract(peptidesLong$Accession,
425   "[^\\|]*\\|"),
426                                     "[^\\|]*\\|"), "\\|")
427
428 #pull acetylated peptides starting after position 2
429 nonCanonicals <- peptidesLong[!peptidesLong$UnLabelled,]
430 nonCanonicals <- nonCanonicals[nonCanonicals$Start > 2,]
431
432 #filter out extra strains
433 nonCanonicals <- nonCanonicals[nonCanonicals$strain != "dCb",]
434
435 #light peptides
436 LightNonCans <- nonCanonicals[nonCanonicals$NtermAcet_L,]
437 #same acetylation string as above
438 LightNonCans$acetSeq <- str_replace(LightNonCans$Peptide,
439   "\\(\\+42\\.01\\)",
440   "(ac)")
441
442 #heavy peptides
443 HeavyNonCans <- nonCanonicals[nonCanonicals$NtermAcet_H,]
444 HeavyNonCans$acetSeq <- str_replace(HeavyNonCans$Peptide,
445   "\\(\\+45\\.03\\)",
446   "(ac)")

```



```

443
444 #combine heavy and lights
445 NonCans2 <- LightNonCans %>% left_join(HeavyNonCans %>% select(c("Accession2", "acetSeq"
, "strain", "Inj", "Area"))),
446                                     by = c("Accession2", "acetSeq", "strain", "Inj"),
447                                     suffix = c("_L", "_H"))
448
449 #set missing data to 0 for calc
450 NonCans2$Area_H[is.na(NonCans2$Area_H)] <- 0
451 NonCans2$Area_L[is.na(NonCans2$Area_L)] <- 0
452
453 #percent acet
454 NonCans2$pctAcet <- NonCans2$Area_L / (NonCans2$Area_H + NonCans2$Area_L)
455
456 #count number of valid values and mean + sd of pct acet
457 NonCans3 <- NonCans2 %>% group_by(Accession2, acetSeq, strain, Start) %>%
458   summarise(countValid = sum(!is.na(pctAcet)),
459             meanPctAcet = mean(pctAcet, na.rm = T),
460             sdPctAcet = sd(pctAcet, na.rm = T))
461
462 #set order
463 NonCans3$strain <- factor(NonCans3$strain, levels = c("WT", "d1839", "c1839"))
464
465 #pull out WT to pick examples
466 lightNonCansWT <- LightNonCans[LightNonCans$strain == 'WT',] %>%
467   group_by(Accession2, acetSeq) %>%
468   summarise(valid = sum(!is.na(Area)))
469
470 #require 2 valid WT measurements
471 lightNonCansWT <- lightNonCansWT[lightNonCansWT$valid > 1,]
472 lightNonCansAccessions <- unique(lightNonCansWT$Accession2)
473
474 #pull out the peptide with the maximum nubmer of valid WT values for each accession
475 lightNonCansWTKeeps <- lightNonCansWT %>% group_by(Accession2) %>%
476   slice_max(valid, with_ties = F) %>%
477   ungroup()
478 NonCans4 <- NonCans3[NonCans3$acetSeq %in% lightNonCansWTKeeps$acetSeq,]
479
480 #create string for plotting
481 NonCans4$startString <- paste0("Start: ", NonCans4$Start)
482 ggplot(NonCans4) +
483   geom_point(aes(x = strain, y = meanPctAcet, color = Accession2,
484                 size = countValid)) +
485   geom_errorbar(aes(x = strain, ymin = meanPctAcet - sdPctAcet,
486                    ymax = meanPctAcet + sdPctAcet, color = Accession2), width = 0.2) +
487   geom_line(aes(x = strain, y = meanPctAcet, color = Accession2, group = acetSeq)) +
488   theme_bw(base_size = 8) +
489   theme(legend.position = "none") +
490   coord_cartesian(ylim = c(0,1.1)) +
491   facet_wrap(c("Accession2", "acetSeq", "startString"))
492
493 ggsave("AllNonCanonicalExamples.png", width = 15, height = 15)
494
495 #pull out examples
496 Examples <- c("MMAR_0359", "MMAR_1369", "MMAR_1730", "MMAR_3090", "MMAR_3818",
497             "MMAR_5064", "MMAR_2201",
498             "MMAR_1025", "MMAR_4849")
499
500 ExamplePlot <- NonCans4[NonCans4$Accession2 %in% Examples,]
501
502 ExamplePlot$countValid <- factor(ExamplePlot$countValid)
503
504 ExamplePlot$Accession2 <- factor(ExamplePlot$Accession2, levels = Examples)
505
506 #plot
507 ggplot(ExamplePlot) +
508   geom_hline(yintercept = c(0,1), linetype = 2, lwd = 1, color = "grey") +
509   geom_errorbar(aes(x = strain, ymin = meanPctAcet - sdPctAcet,
510                    ymax = meanPctAcet + sdPctAcet, width = 0.25, lwd = 1.5) +

```



```

510   geom_errorbar(aes(x = strain, ymin = meanPctAcet - sdPctAcet,
511                     ymax = meanPctAcet + sdPctAcet, color = Accession2), width = 0.2) +
512   geom_line(aes(x = strain, y = meanPctAcet, group = acetSeq), lwd = 3) +
513   geom_line(aes(x = strain, y = meanPctAcet, color = Accession2, group = acetSeq), lwd =
514             1) +
515   geom_point(aes(x = strain, y = meanPctAcet, fill = Accession2,
516                 size = countValid), shape = 21, stroke = 2) +
517   theme_bw(base_size = 20) +
518   theme(legend.position = "none",
519         panel.grid = element_blank()) +
520   coord_cartesian(ylim = c(0, 1.2)) +
521   facet_wrap(c("Accession2", "acetSeq", "startString")) +
522   scale_size_manual(values = c(5, 6, 7, 8)) +
523   scale_y_continuous(breaks = c(0, 0.25, 0.5, 0.75, 1)) +
524   labs(x = element_blank(), y = "Proportion Acetylated") +
525   scale_x_discrete(labels = c("WT", expression(Delta * italic(empl)),
526                               expression(Delta * italic(empl) * "/comp"))))
527
528   ggsave("NonCanonicalN-terminaiExamples.png", width = 15, height = 15)
529
530
531
532

```