

DEPI Day 3 Epistasis

Abigail Seeger

January 8, 2021

Contents

| | |
|---|-----------|
| Load Necessary Packages | 1 |
| Load in Cleaned Data | 2 |
| Functions | 2 |
| Selection Coefficient Calculations | 3 |
| Selection Coefficient Visualizations - All | 4 |
| Selection Coefficient Visualizations - Day 3 | 11 |
| Epistasis Calculations | 18 |
| Epistasis Visualizations - All | 20 |
| Additive Epistasis | 20 |
| Proportional Epistasis | 27 |
| Epistasis Visualizations - Day 3 | 34 |
| Additive | 34 |
| Proportional | 41 |

Load Necessary Packages

```
library(stringr)
library(stringi)
library(dplyr)
library(viridis)
library(ggplot2)
library(extrafont)
library(ggthemes)
library(lemon)
```

Load in Cleaned Data

```
depi_data <- read.table("C:/Users/Owner/Documents/Research/Shiu_Lab/Shiu_Lab_R/Data/Clean_DEPI_Data_V2.1.csv",  
  sep = ",", header = TRUE)
```

Functions

```
add_number <- function(data_frame) {  
  ### First, if the genotype is Col0 (only  
  ### genotype with length 4), assign 0 as  
  ### number Else, assign number as genotype  
  ### with 'mpk' removed Example: mpk1 will  
  ### be 1, mpk1-17 will be 1-17  
  data_frame <- data_frame %>% mutate(number = ifelse(genotype !=  
    "Col0", (stri_sub(genotype, 4, length(genotype))),  
    0))  
  ### Next, for all double mutants, replace  
  ### '-' with '0' Example: 1-17 becomes 1017  
  data_frame$number <- as.numeric(gsub("-",  
    "0", data_frame$number))  
  ### Almost there! There's a problem with  
  ### two single digit double mutants We need  
  ### a four digit number to sort correctly  
  ### Example: mpk1-3 -> 1-3 -> 103, but we  
  ### need it to be 1003 to sort correctly  
  data_frame$number[data_frame$number ==  
    "103"] <- "1003"  
  data_frame$number[data_frame$number ==  
    "506"] <- "5006"  
  data_frame$number[data_frame$number ==  
    "608"] <- "6008"  
  data_frame$number[data_frame$number ==  
    "609"] <- "6009"  
  ### Convert number to a numeric in order  
  ### to sort  
  data_frame$number <- as.numeric(data_frame$number)  
  data_frame <- data_frame %>% arrange(number)  
  data_frame <- data_frame %>% mutate(number_2 = number)  
  data_frame$number_2[nchar(data_frame$number_2) ==  
    4] <- 0  
  data_frame$number_2[nchar(data_frame$number_2) ==  
    5] <- 0  
  return(data_frame)  
}
```

Selection Coefficient Calculations

```
selectionCoef <- data.frame(genotype = rep(NA,
  0), SelectionCoefficient = rep(NA, 0),
  Experiment = rep(NA, 0), Measurement = rep(NA,
  0), Time_Point = rep(NA, 0))

temp_month <- "Jan"
temp_measurement <- "phi2"

temp_data <- depi_data %>% filter(month ==
  temp_month, measurement == temp_measurement)

for (temp_month in c("Dec", "Jan", "Feb")) {
  for (temp_measurement in c("phi2", "npq")) {
    temp_data <- depi_data %>% filter(month ==
      temp_month, measurement == temp_measurement)
    temp_n_row <- ifelse(temp_month ==
      "Dec", 37, 38)
    for (i in unique(temp_data$time_point)) {
      temp_data_2 <- filter(temp_data,
        time_point == i)
      selectionCoefTmp <- data.frame(genotype = rep(NA,
        temp_n_row), SelectionCoefficient = rep(NA,
        temp_n_row), Experiment = rep(NA,
        temp_n_row), Measurement = rep(NA,
        temp_n_row), Time_Point = rep(NA,
        temp_n_row))
      count <- 1
      for (g in unique(temp_data$genotype)) {
        fm <- mean(filter(temp_data_2,
          genotype == g, month ==
            temp_month, measurement ==
            temp_measurement, time_point ==
            i)$normalized_value)
        fwt <- mean(filter(temp_data_2,
          genotype == "Col0", month ==
            temp_month, measurement ==
            temp_measurement, time_point ==
            i)$normalized_value)
        selectionCoefTmp[count, 1] <- g
        selectionCoefTmp[count, 2] <- (fm -
          fwt)/fwt
        selectionCoefTmp[count, 3] <- temp_month
        selectionCoefTmp[count, 4] <- temp_measurement
        selectionCoefTmp[count, 5] <- i

        count <- count + 1
      }
      selectionCoef <- rbind(selectionCoef,
        selectionCoefTmp)
    }
  }
}
```

```
}
}
```

Pick a point and calculate the selection coefficient outside the loop to ensure that the loop is correct:

```
m_test_1 <- mean(filter(depi_data, month ==
  "Dec", measurement == "phi2", time_point ==
  "0", genotype == "mpk1")$normalized_value)
wt_test_1 <- mean(filter(depi_data, month ==
  "Dec", measurement == "phi2", time_point ==
  "0", genotype == "Col0")$normalized_value)

((m_test_1 - wt_test_1)/wt_test_1)
```

```
## [1] 0.01996154
```

```
filter(selectionCoef, Experiment == "Dec",
  Measurement == "phi2", Time_Point ==
  "0", genotype == "mpk1")
```

```
##   genotype SelectionCoefficient Experiment Measurement Time_Point
## 1      mpk1           0.01996154      Dec      phi2           0
```

Selection Coefficient Visualizations - All

```
selectionCoef <- add_number(selectionCoef)
selectionCoef$genotype <- reorder(selectionCoef$genotype,
  selectionCoef$number)
selectionCoef <- add_day_col(selectionCoef)
```

```
for (temp_measurement in c("phi2", "npq")) {

  set_bounds <- selectionCoef %>% filter(Measurement ==
    temp_measurement)

  lower_bound <- round(min(set_bounds$SelectionCoefficient) -
    0.05, 2)
  upper_bound <- round(max(set_bounds$SelectionCoefficient) +
    0.05, 2)

  for (temp_month in c("Dec", "Jan", "Feb")) {
    temp_plot_data <- filter(selectionCoef,
      Experiment == temp_month, Measurement ==
      temp_measurement)

    temp_title <- ifelse(temp_month ==
      "Dec", "December", ifelse(temp_month ==
      "Jan", "January", "February"))
```

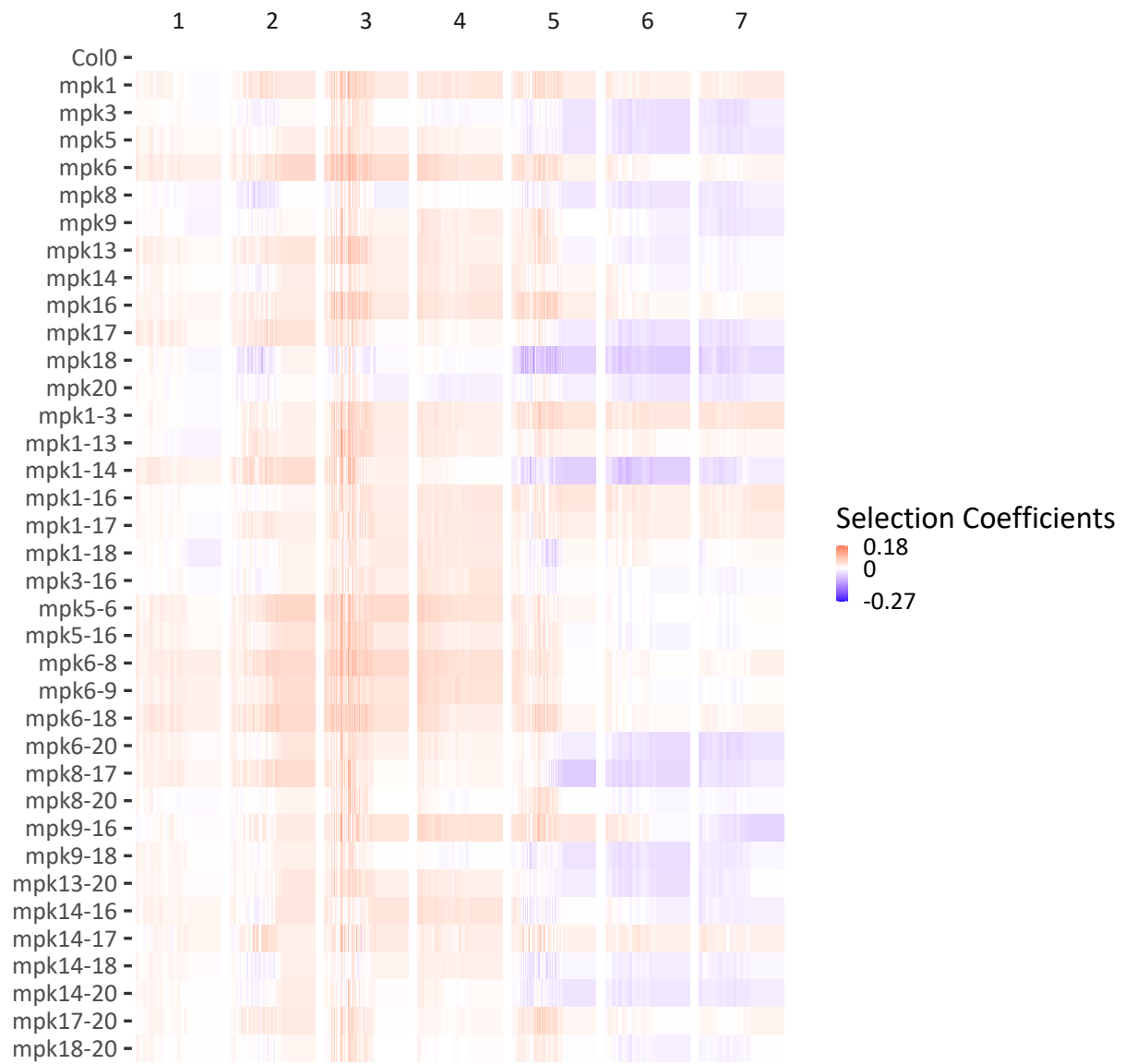
```

plot <- ggplot(data = temp_plot_data,
  aes(x = Time_Point, y = genotype,
    fill = SelectionCoefficient)) +
  labs(fill = "Selection Coefficients",
    x = "Hours", y = NULL, title = paste(temp_title,
      ".", temp_measurement,
      " Selection Coefficient",
      sep = "")) + geom_tile(width = ifelse(temp_measurement ==
"leafarea", 16, 10), height = 30) +
  facet_grid(genotype ~ day, scales = "free",
    switch = "y") + theme_tufte(base_family = "Calibri",
  base_size = 50) + theme(strip.background.y = element_blank(),
  strip.text.y = element_blank(),
  axis.title.x = element_blank(),
  axis.text.x = element_blank(),
  axis.ticks.x = element_blank(),
  panel.spacing = unit(0, "lines")) +
  scale_fill_gradient2(low = "blue",
    high = "red", mid = "white",
    midpoint = 0, limits = c(lower_bound,
      upper_bound), breaks = c(lower_bound,
      0, upper_bound), labels = c(as.character(lower_bound),
      "0", as.character(upper_bound)))

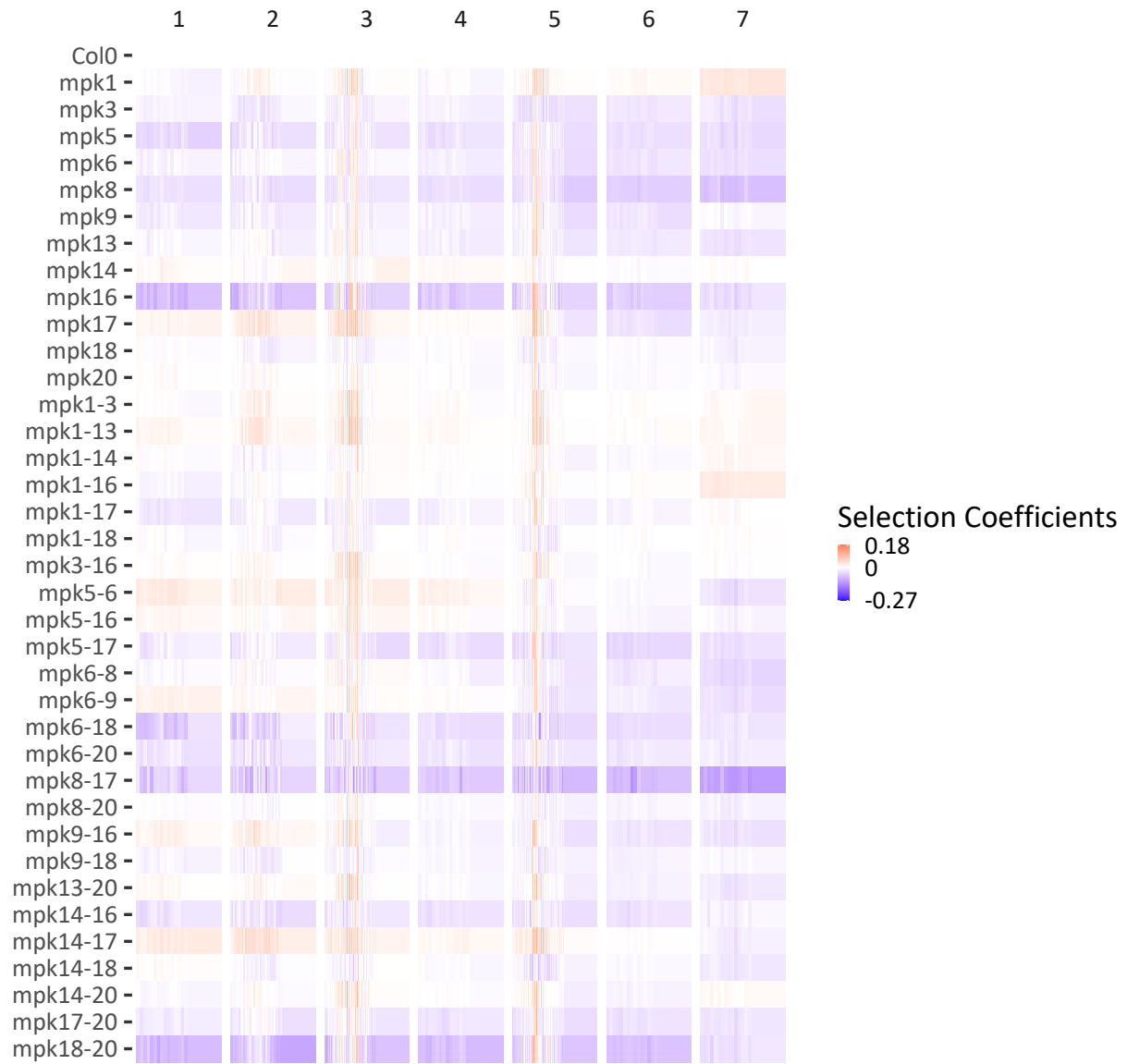
print(plot)
}

```

December:phi2 Selection Coefficient



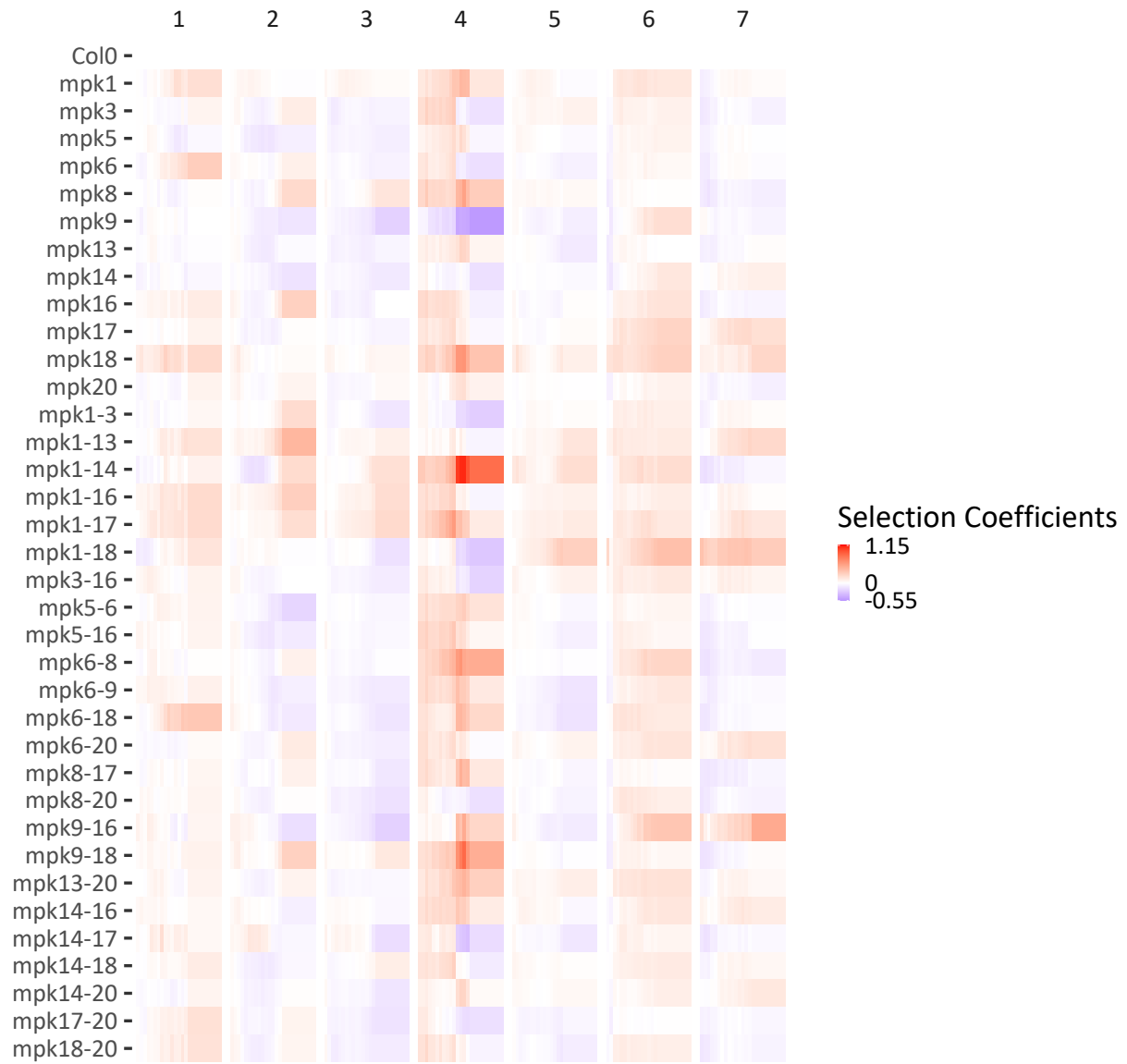
January:phi2 Selection Coefficient

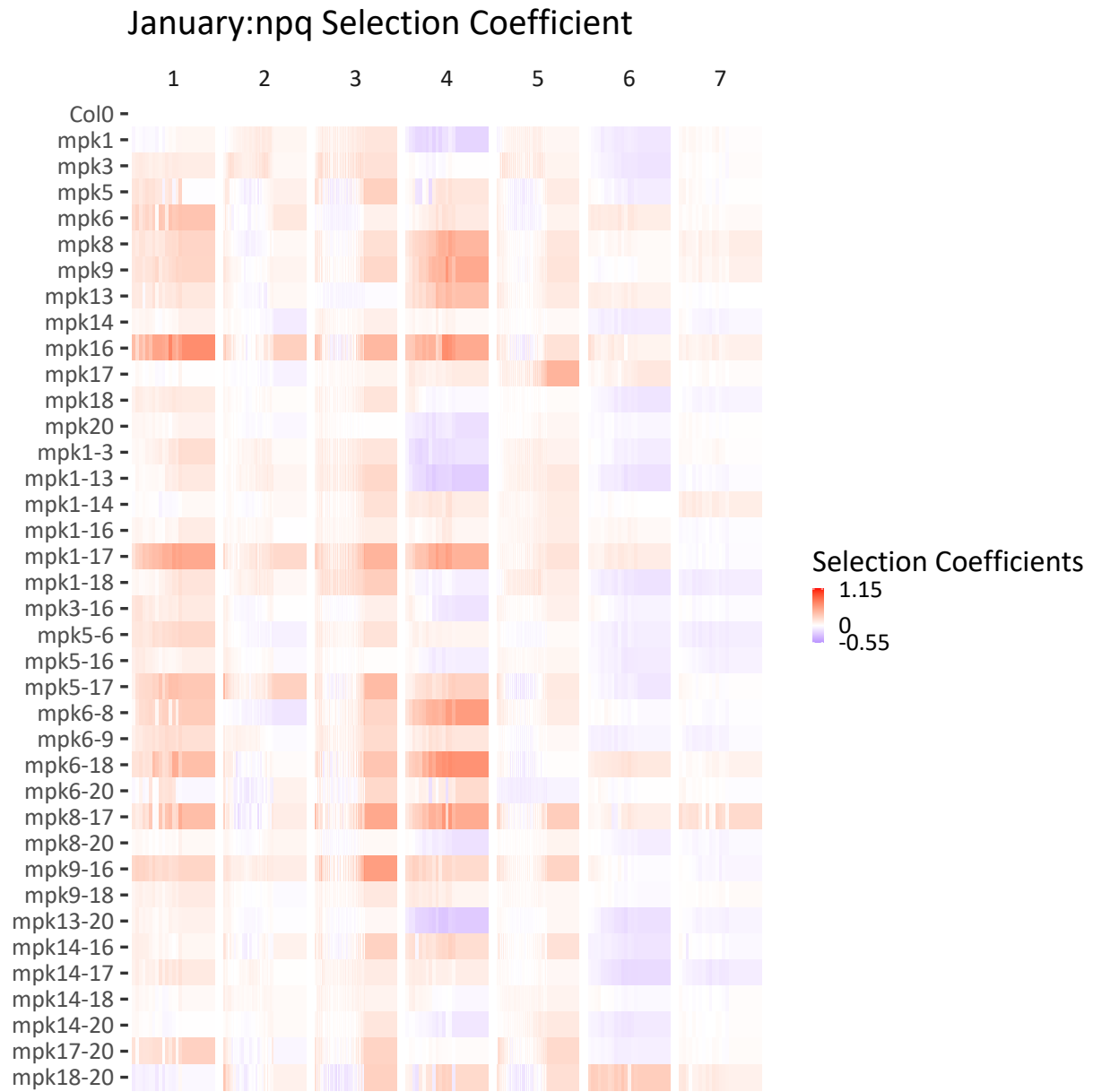


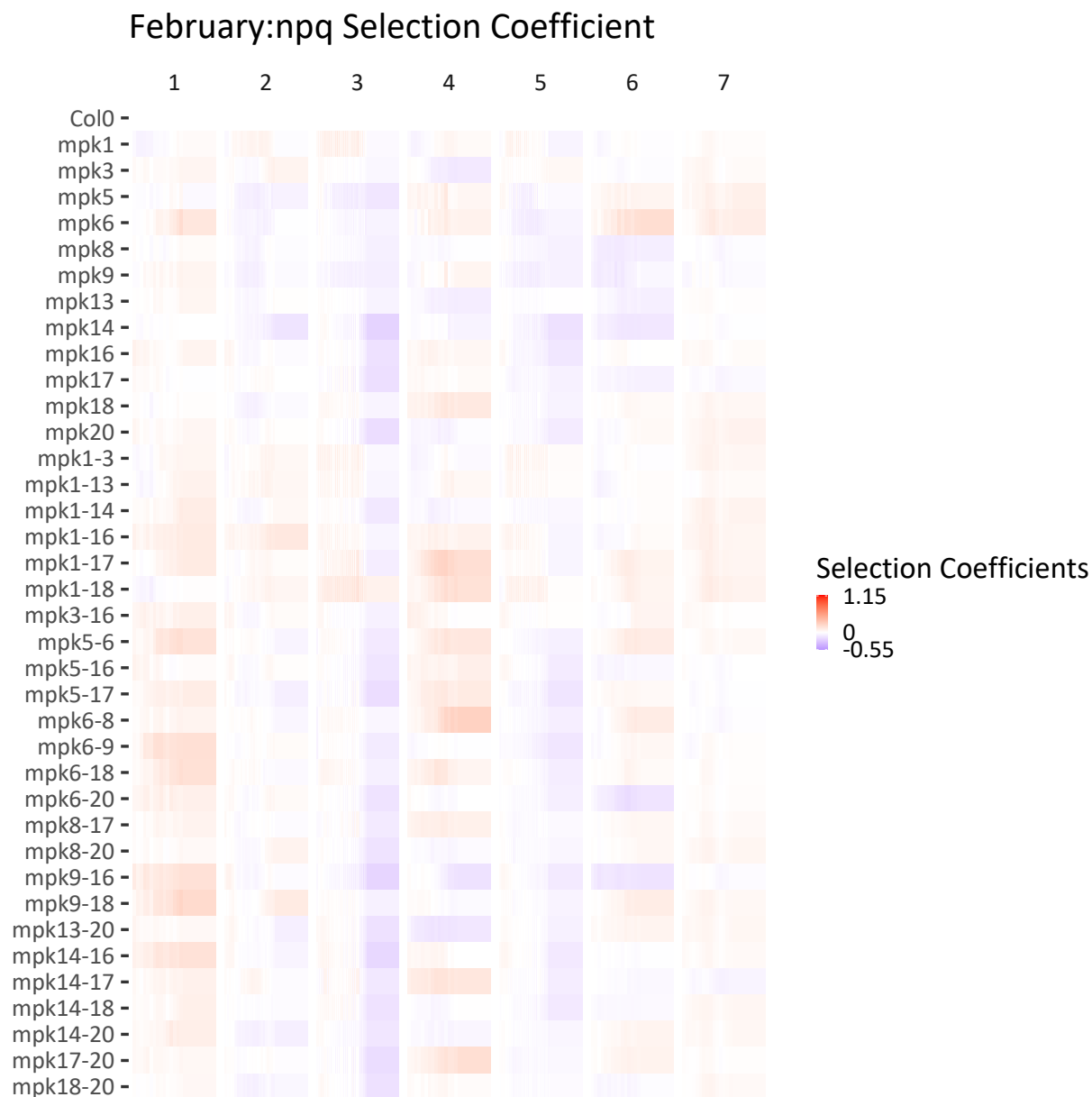
February:phi2 Selection Coefficient



December:npq Selection Coefficient







Selection Coefficient Visualizations - Day 3

```
selectionCoef_day3 <- selectionCoef %>% filter(day ==
  "3")
```

```
for (temp_measurement in c("phi2", "npq")) {

  set_bounds <- selectionCoef_day3 %>%
    filter(Measurement == temp_measurement)

  lower_bound <- round(min(set_bounds$SelectionCoefficient) -
```

```

    0.05, 2)
upper_bound <- round(max(set_bounds$SelectionCoefficient) +
    0.05, 2)

for (temp_month in c("Dec", "Jan", "Feb")) {

  temp_plot_data <- filter(selectionCoef_day3,
    Experiment == temp_month, Measurement ==
      temp_measurement)
  temp_title <- ifelse(temp_month ==
    "Dec", "December", ifelse(temp_month ==
    "Jan", "January", "February"))
  plot <- ggplot(data = temp_plot_data,
    aes(x = Time_Point, y = genotype,
      fill = SelectionCoefficient)) +
    labs(fill = "Selection Coefficients",
      x = "Hours", y = NULL, title = paste(temp_title,
        ":", temp_measurement,
        " Selection Coefficient",
        sep = "")) + geom_tile(width = ifelse(temp_measurement ==
    "leafarea", 16, 10), height = 30) +
    facet_grid(genotype ~ day, scales = "free",
      switch = "y") + theme_tufte(base_family = "Calibri",
    base_size = 50) + theme(strip.background.y = element_blank(),
    strip.text.y = element_blank(),
    axis.title.x = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    panel.spacing = unit(0, "lines")) +
    scale_fill_gradient2(low = "blue",
      high = "red", mid = "white",
      midpoint = 0, limits = c(lower_bound,
        upper_bound), breaks = c(lower_bound,
        0, upper_bound), labels = c(as.character(lower_bound),
        "0", as.character(upper_bound)))

  print(plot)
}
}

```

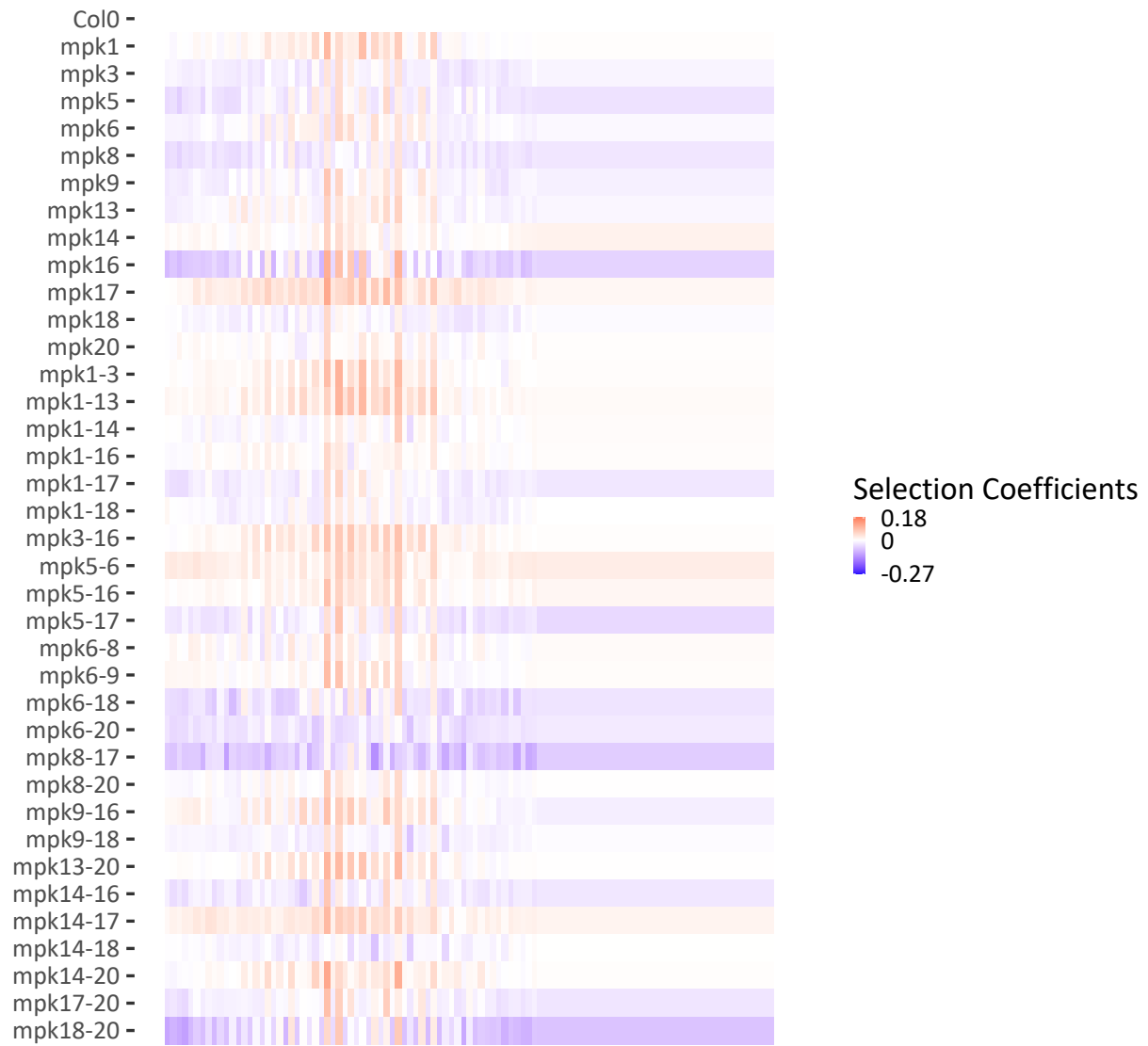
December:phi2 Selection Coefficient

3



January:phi2 Selection Coefficient

3



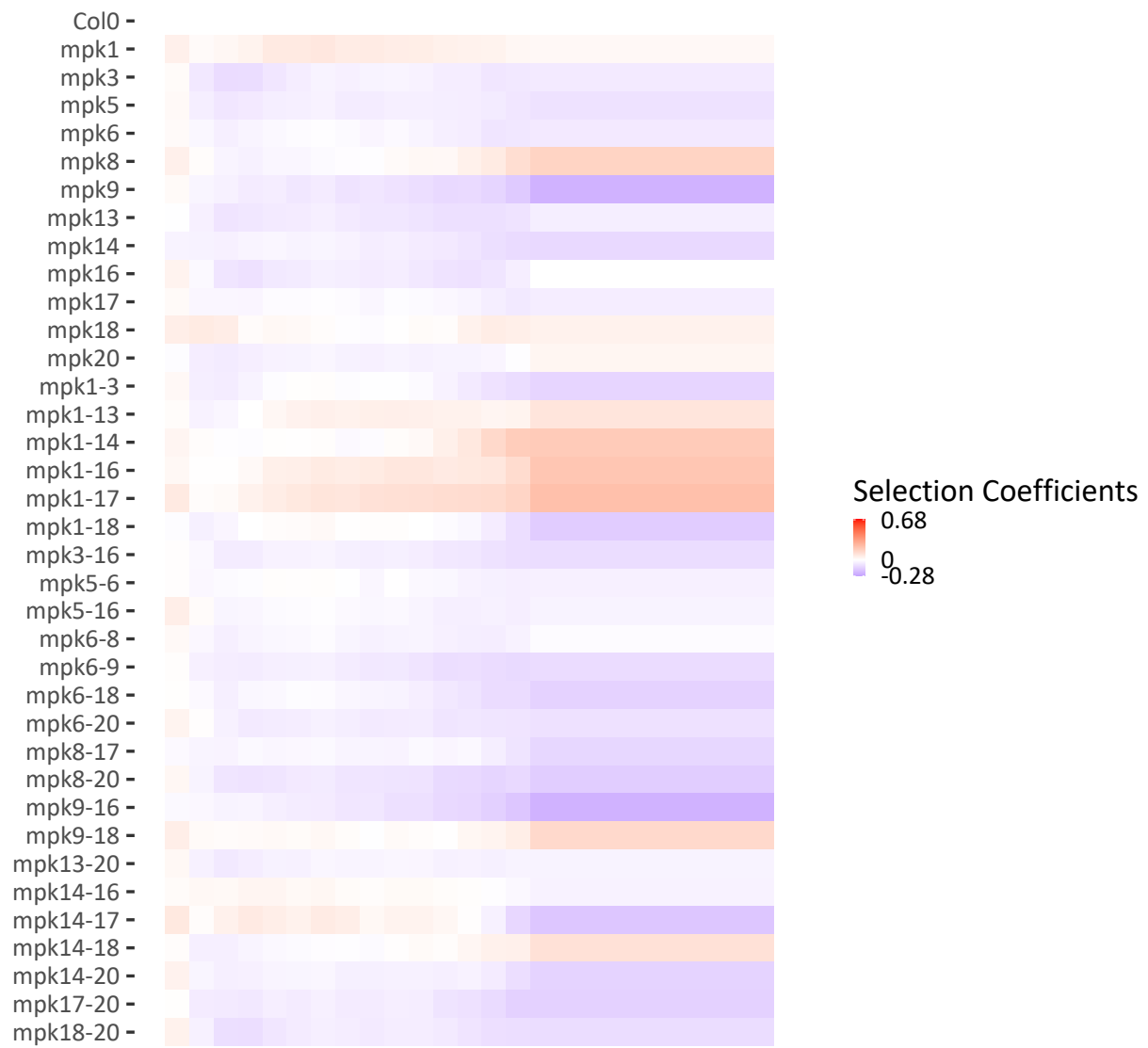
February:phi2 Selection Coefficient

3



December:npq Selection Coefficient

3



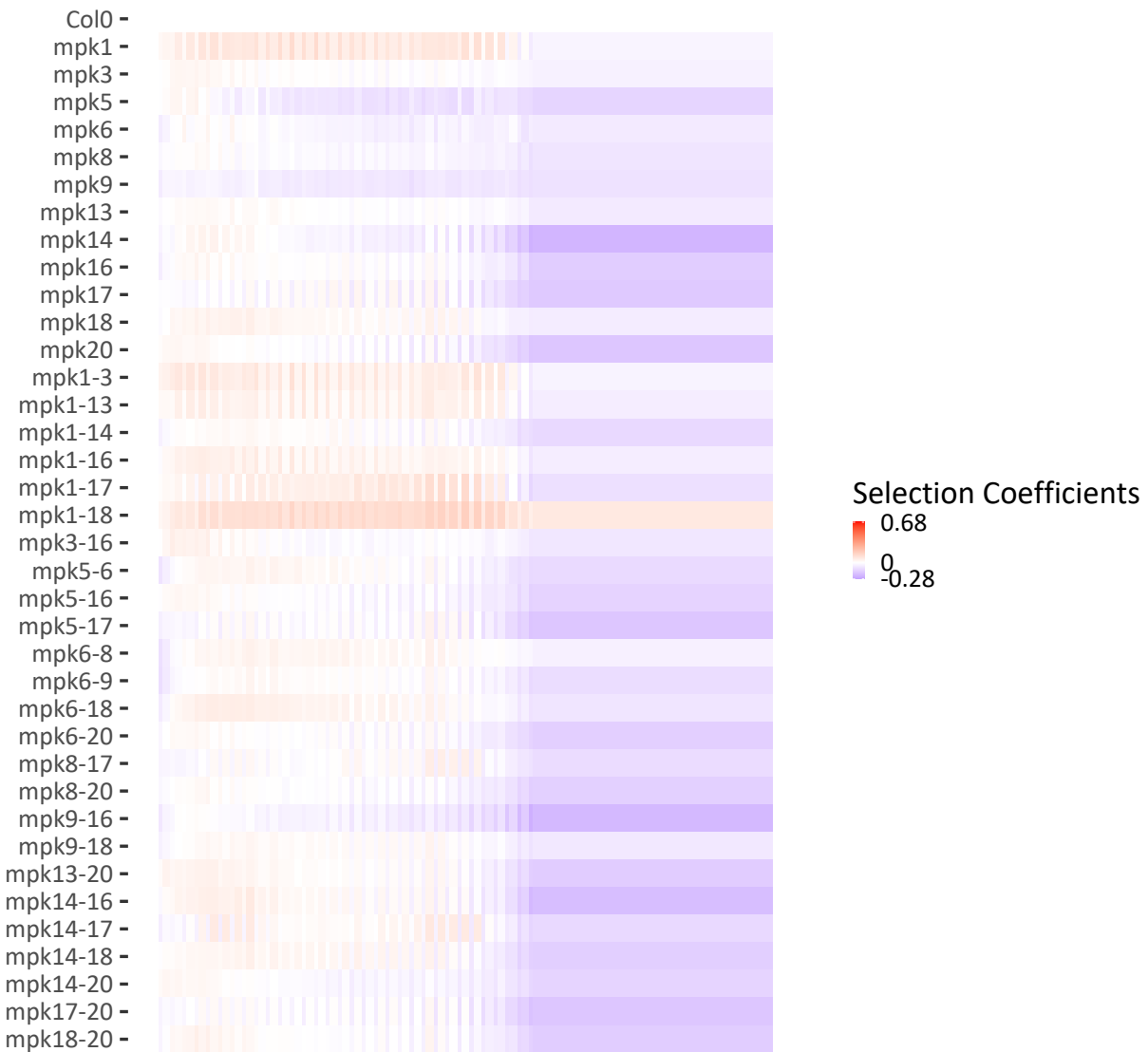
January:npq Selection Coefficient

3



February:npq Selection Coefficient

3



Epistasis Calculations

```
all_double_mutants = list()
for (gen in unique(depi_data$genotype)) {
  if (str_detect(gen, "-") == T) {
    all_double_mutants = c(all_double_mutants,
                           gen)
  }
}
```

Epistasis Calculations: Initialize an

```

### empty data frame to populate with
### information:
geneticInteractions <- data.frame(genotype = rep(NA,
  0), MutantA = rep(NA, 0), MutantB = rep(NA,
  0), AdditiveEpistasis = rep(NA, 0), ProportionalEpistasis = rep(NA,
  0), Experiment = rep(NA, 0), Measurement = rep(NA,
  0), Time_Point = rep(NA, 0))

### Loop through each experiment and
### measurement:
for (e in c("Dec", "Jan", "Feb")) {
  for (m in c("phi2", "npq")) {
    temp_data <- depi_data %>% filter(month ==
      e, measurement == m)
    temp_nrow <- 0
    for (i in unique(temp_data$time_point)) {
      ### Filter to each specific experiment and
      ### measurement
      tempData <- temp_data %>% filter(time_point ==
        i)
      ### Create an empty data frame to fill with
      ### the information and calculations:
      geneticInteractionsTmp <- data.frame(genotype = rep(NA,
        temp_nrow), MutantA = rep(NA,
        temp_nrow), MutantB = rep(NA,
        temp_nrow), AdditiveEpistasis = rep(NA,
        temp_nrow), ProportionalEpistasis = rep(NA,
        temp_nrow), Experiment = rep(NA,
        temp_nrow), Measurement = rep(NA,
        temp_nrow), Time_Point = rep(NA,
        temp_nrow))
      ### Initialize a row count to use to
      ### populate the data frame
      rowCount <- 1
      ### For each of the double mutants:
      for (dm in unlist(all_double_mutants)) {
        ### Extract the single mutants from the
        ### double mutant
        ma <- unlist(strsplit(dm,
          "-"))[1]
        mb <- paste("mpk", unlist(strsplit(dm,
          "-"))[2], sep = "")
        ### Calculate the fitness of the dm, ma,
        ### mb, and wt
        fdm <- mean(filter(tempData,
          genotype == dm)$normalized_value)
        fwt <- mean(filter(tempData,
          genotype == "Col0")$normalized_value)
        fma <- mean(filter(tempData,
          genotype == ma)$normalized_value)
        fmb <- mean(filter(tempData,
          genotype == mb)$normalized_value)
        ### Calculate Additive and Proportional

```

```

    ### Epistasis
    AddEp <- fdm + fwt - (fma +
      fmb)
    PropEp <- log((fdm * fwt)/(fma *
      fmb))
    ### Populate the data frame with this
    ### information:
    geneticInteractionsTmp[rowCount,
      1] <- dm
    geneticInteractionsTmp[rowCount,
      2] <- ma
    geneticInteractionsTmp[rowCount,
      3] <- mb
    geneticInteractionsTmp[rowCount,
      4] <- AddEp
    geneticInteractionsTmp[rowCount,
      5] <- PropEp
    geneticInteractionsTmp[rowCount,
      6] <- e
    geneticInteractionsTmp[rowCount,
      7] <- m
    geneticInteractionsTmp[rowCount,
      8] <- i
    rowCount <- rowCount + 1
  }
  ### Add the rows of the temporary genetic
  ### interaction information to the main
  ### data frame
  geneticInteractions <- rbind(geneticInteractions,
    geneticInteractionsTmp)
}
}
}

```

Epistasis Visualizations - All

Additive Epistasis

```

geneticInteractions <- add_number(geneticInteractions)
geneticInteractions$genotype <- reorder(geneticInteractions$genotype,
  geneticInteractions$number)
geneticInteractions <- add_day_col(geneticInteractions)

```

```

for (temp_measurement in c("phi2", "npq")) {
  set_bounds <- geneticInteractions %>%
    filter(Measurement == temp_measurement)
  lower_bound <- round(min(set_bounds$AdditiveEpistasis) -
    0.05, 2)
  upper_bound <- round(max(set_bounds$AdditiveEpistasis) +
    0.05, 2)
}

```

```

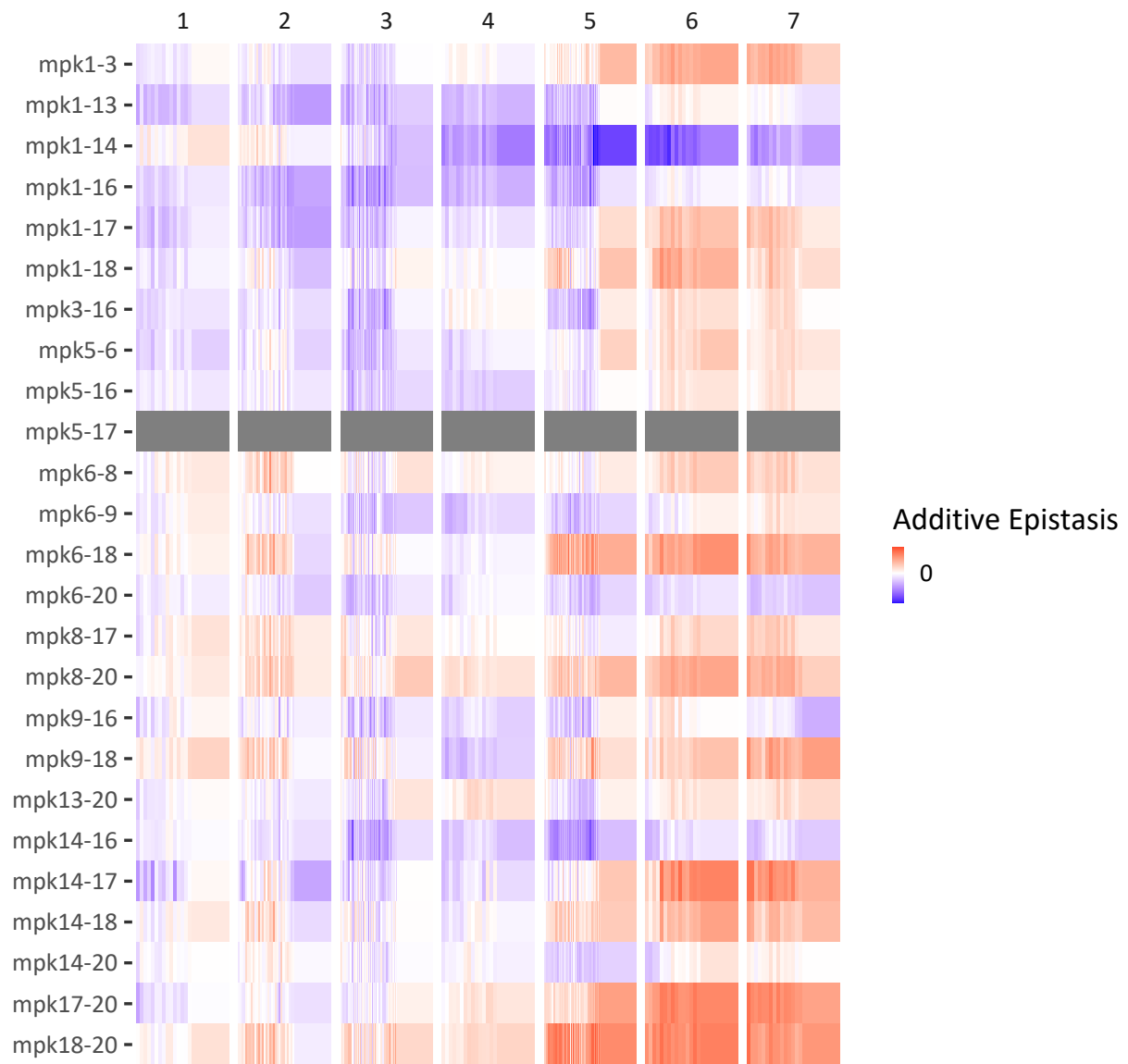
for (temp_month in c("Dec", "Jan", "Feb")) {

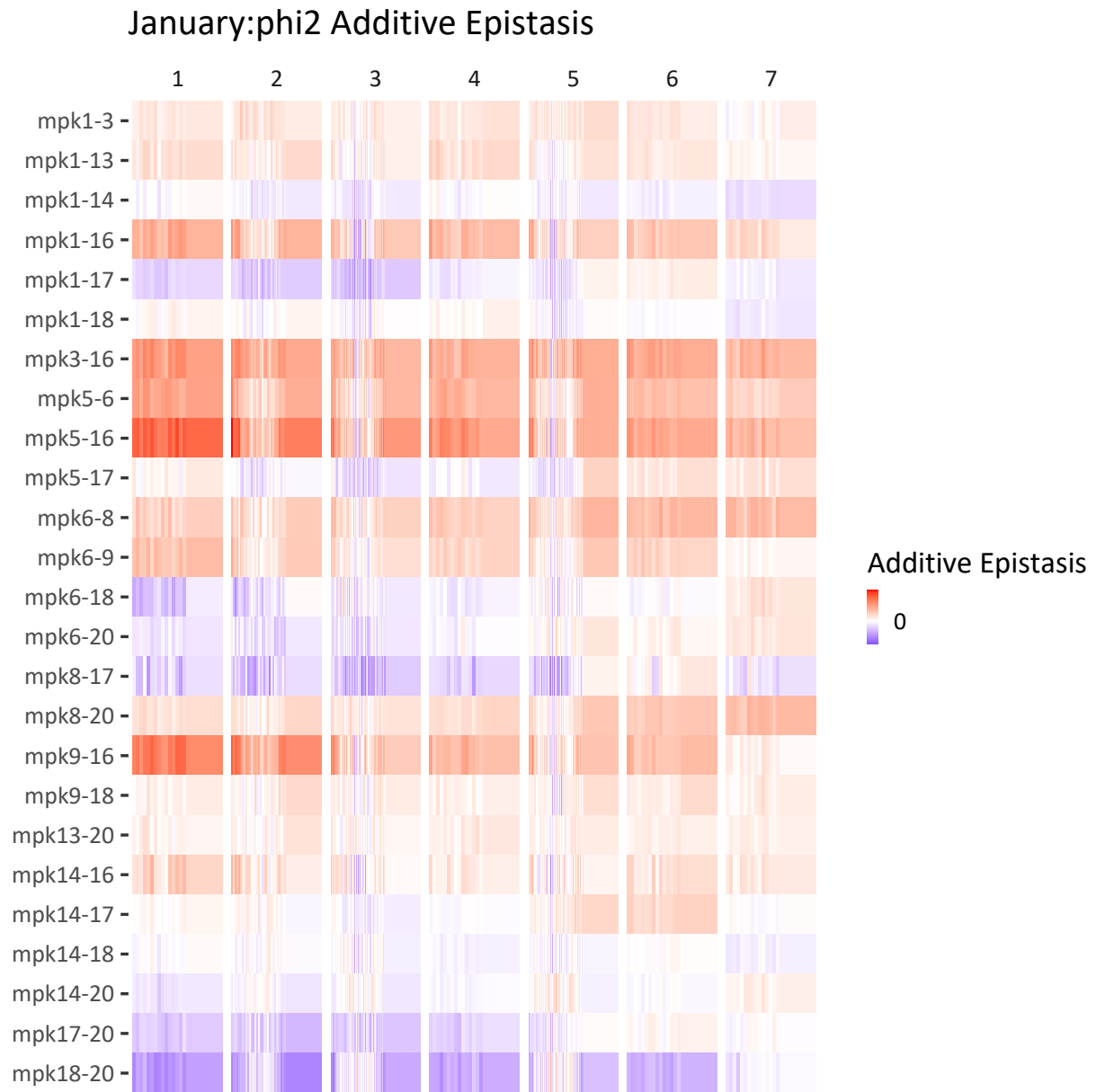
  temp_plot_data <- filter(geneticInteractions,
    Experiment == temp_month, Measurement ==
      temp_measurement)

  temp_title <- ifelse(temp_month ==
    "Dec", "December", ifelse(temp_month ==
      "Jan", "January", "February"))
  plot <- ggplot(data = temp_plot_data,
    aes(x = Time_Point, y = genotype,
      fill = AdditiveEpistasis)) +
    labs(fill = "Additive Epistasis",
      x = "Hours", y = NULL, title = paste(temp_title,
        ":", temp_measurement,
        " Additive Epistasis",
        sep = "")) + geom_tile(width = ifelse(temp_measurement ==
      "leafarea", 16, 10), height = 30) +
    facet_grid(genotype ~ day, scales = "free",
      switch = "y") + theme_tufte(base_family = "Calibri",
      base_size = 50) + theme(strip.background.y = element_blank(),
      strip.text.y = element_blank(),
      axis.title.x = element_blank(),
      axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),
      panel.spacing = unit(0, "lines")) +
    scale_fill_gradient2(low = "blue",
      high = "red", mid = "white",
      midpoint = 0, limits = c(lower_bound,
        upper_bound), breaks = c(lower_bound,
        0, upper_bound), labels = c(as.character(lower_bound),
        "0", as.character(upper_bound)))
  print(plot)
}
}

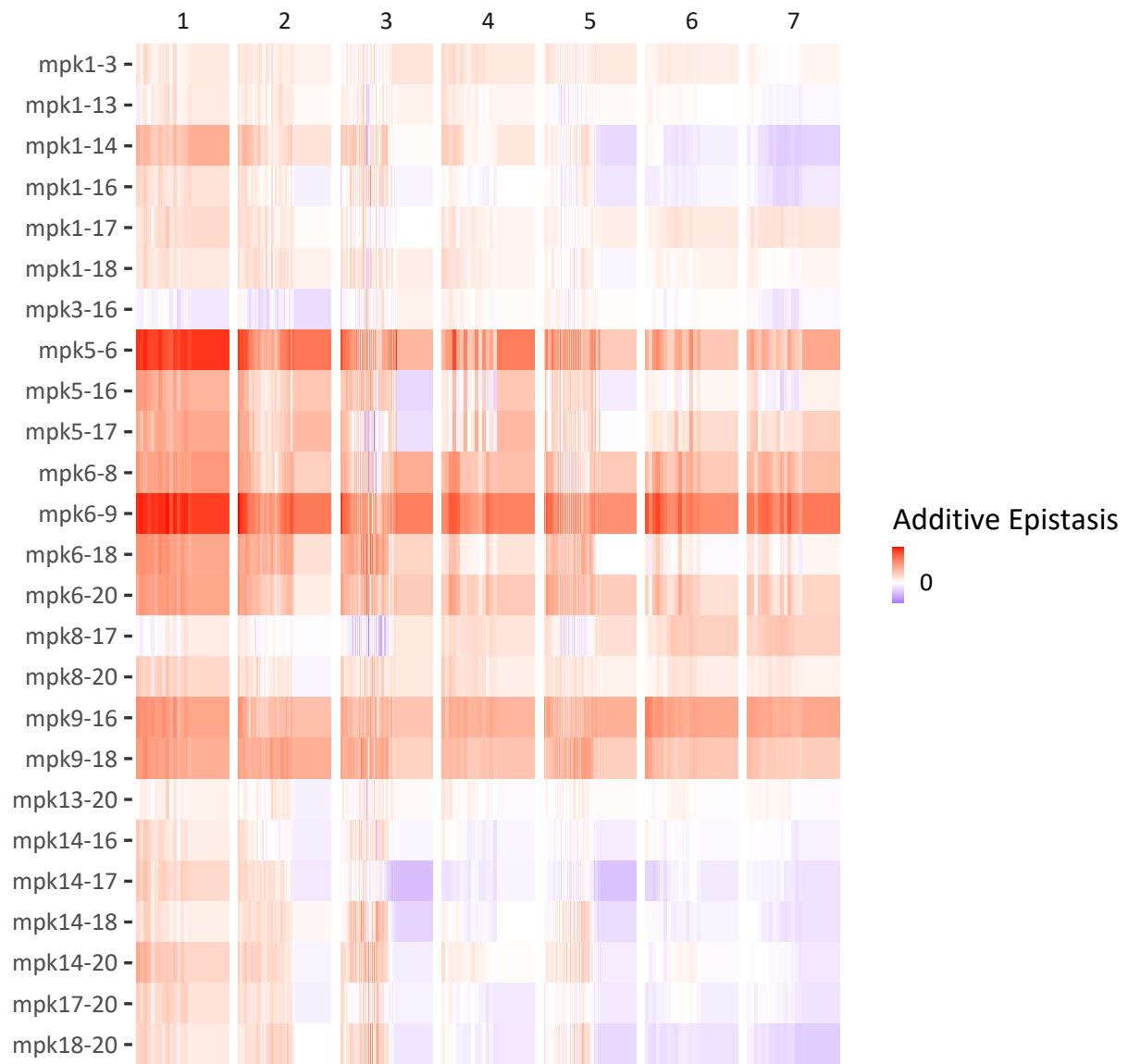
```

December:phi2 Additive Epistasis

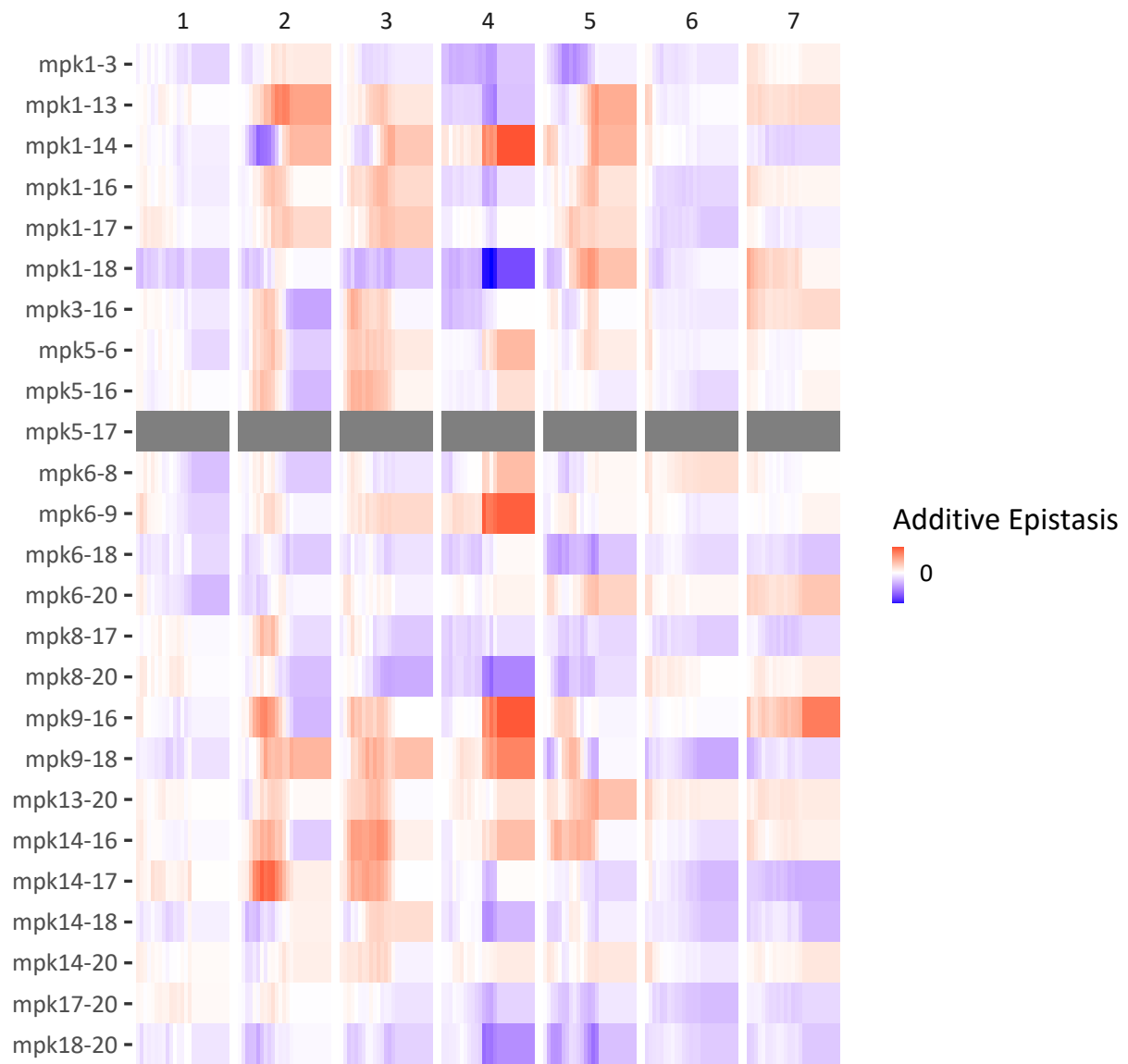




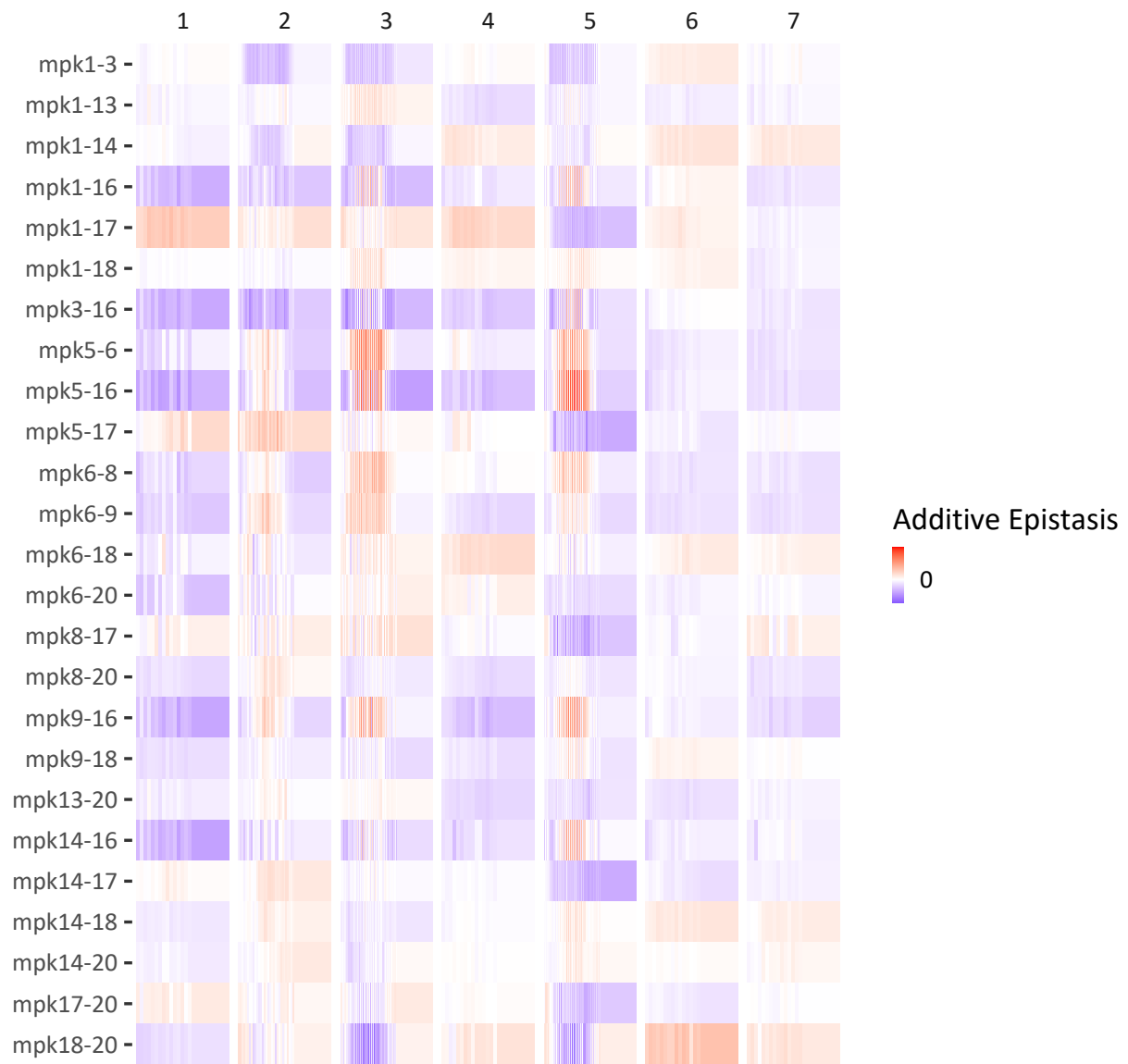
February:phi2 Additive Epistasis



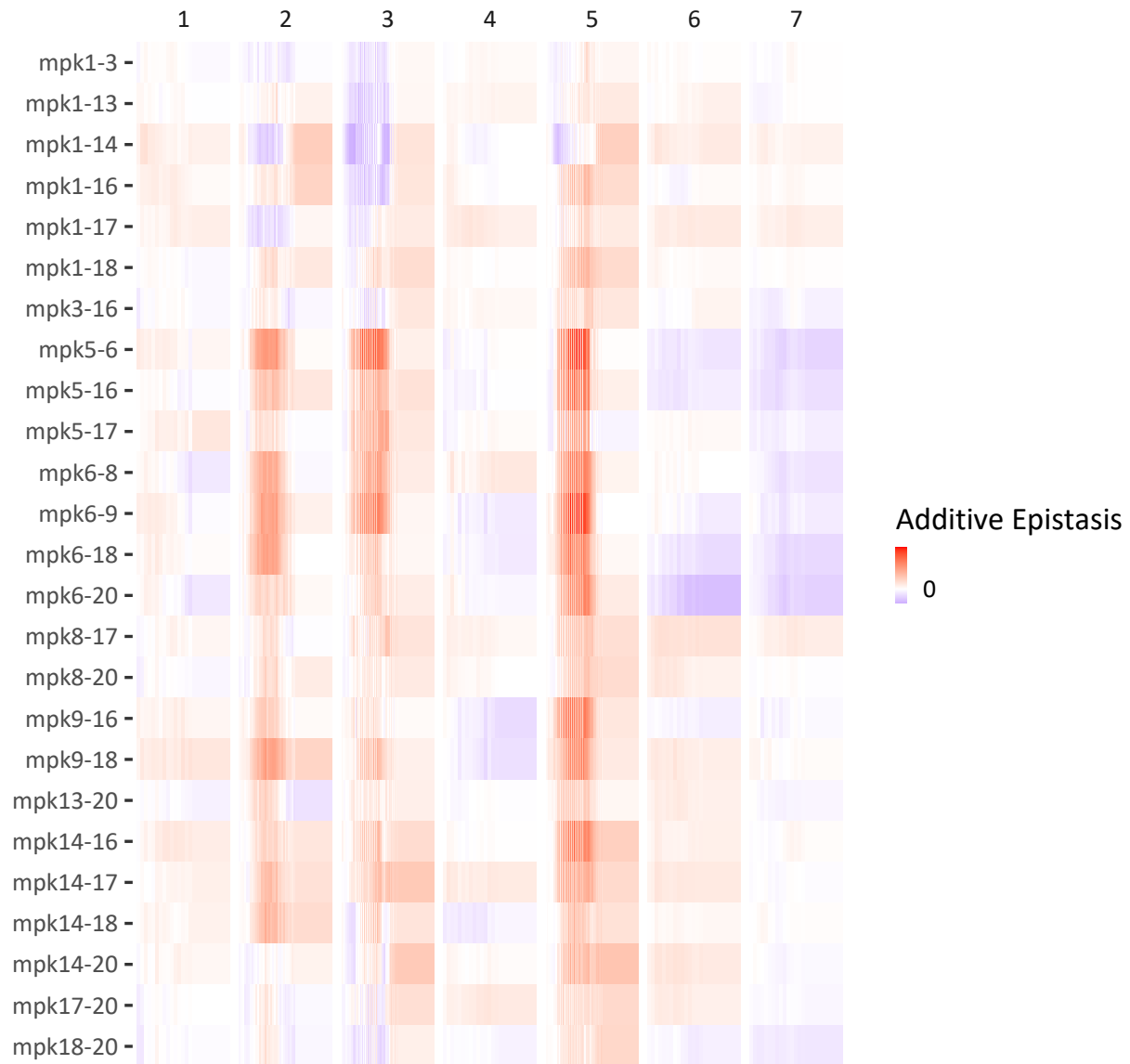
December:npq Additive Epistasis



January:npq Additive Epistasis



February:npq Additive Epistasis



Proportional Epistasis

```
for (temp_measurement in c("phi2", "npq")) {

  set_bounds <- geneticInteractions %>%
    filter(Measurement == temp_measurement)
  lower_bound <- round(min(set_bounds$ProportionalEpistasis) -
    0.05, 2)
  upper_bound <- round(max(set_bounds$ProportionalEpistasis) +
    0.05, 2)

  for (temp_month in c("Dec", "Jan", "Feb")) {
```

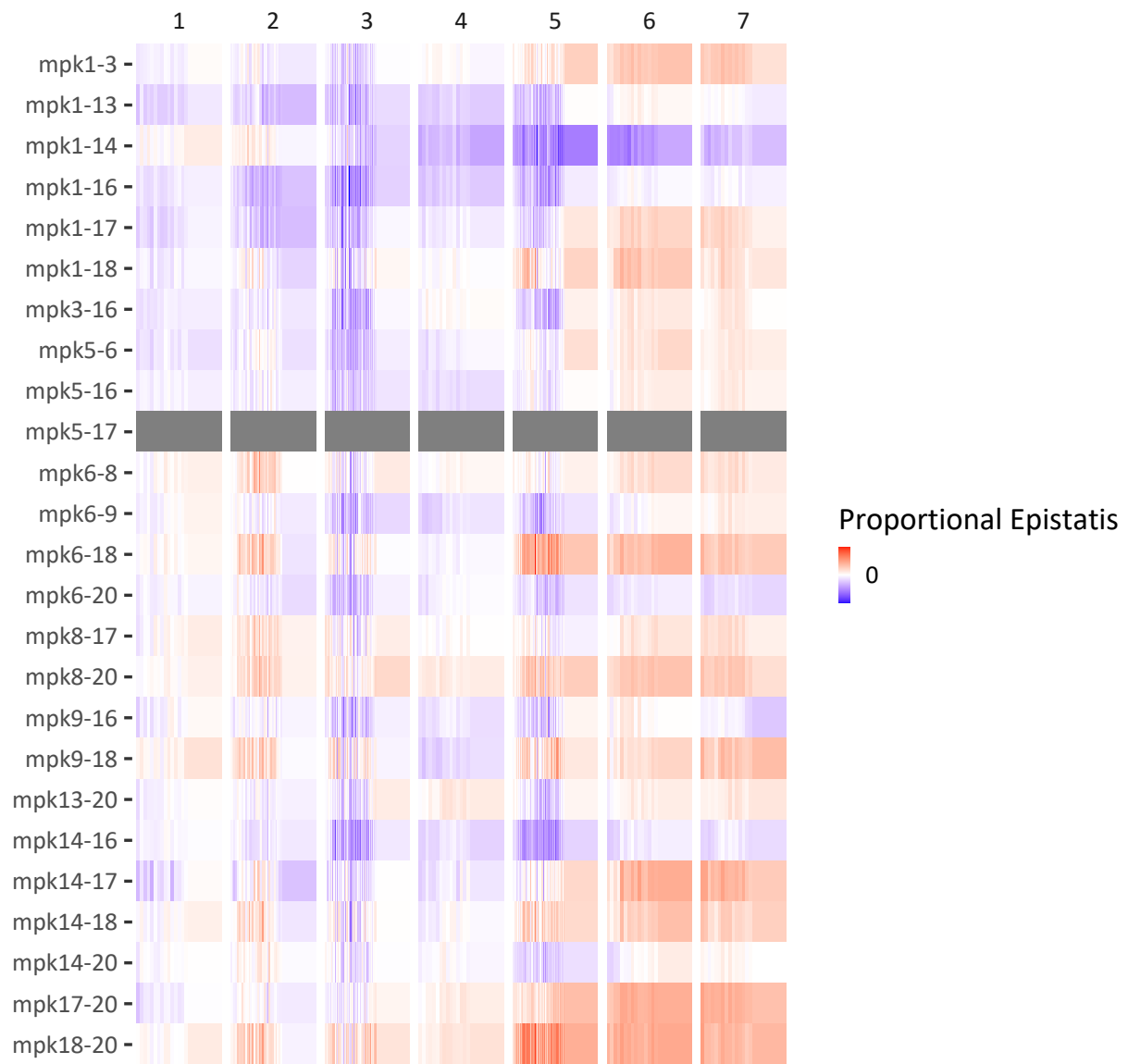
```

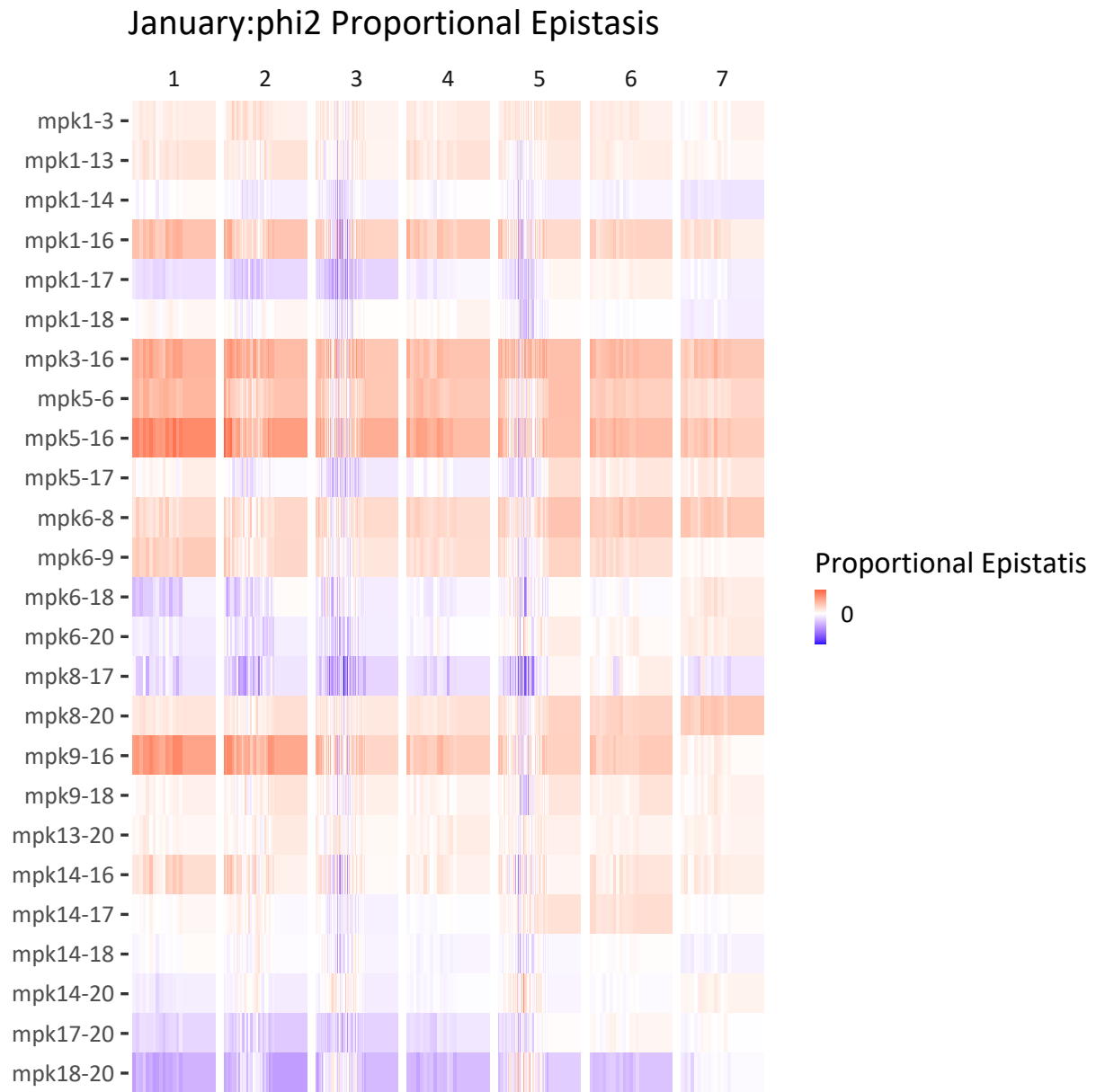
temp_plot_data <- filter(geneticInteractions,
  Experiment == temp_month, Measurement ==
    temp_measurement)

temp_title <- ifelse(temp_month ==
  "Dec", "December", ifelse(temp_month ==
    "Jan", "January", "February"))
plot <- ggplot(data = temp_plot_data,
  aes(x = Time_Point, y = genotype,
    fill = ProportionalEpistasis)) +
  labs(fill = "Proportional Epistasis",
    x = "Hours", y = NULL, title = paste(temp_title,
      ".", temp_measurement,
      " Proportional Epistasis",
      sep = "")) + geom_tile(width = ifelse(temp_measurement ==
    "leafarea", 16, 10), height = 30) +
  facet_grid(genotype ~ day, scales = "free",
    switch = "y") + theme_tufte(base_family = "Calibri",
    base_size = 50) + theme(strip.background.y = element_blank(),
    strip.text.y = element_blank(),
    axis.title.x = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    panel.spacing = unit(0, "lines")) +
  scale_fill_gradient2(low = "blue",
    high = "red", mid = "white",
    midpoint = 0, limits = c(lower_bound,
      upper_bound), breaks = c(lower_bound,
        0, upper_bound), labels = c(as.character(lower_bound),
          "0", as.character(upper_bound)))
print(plot)
}

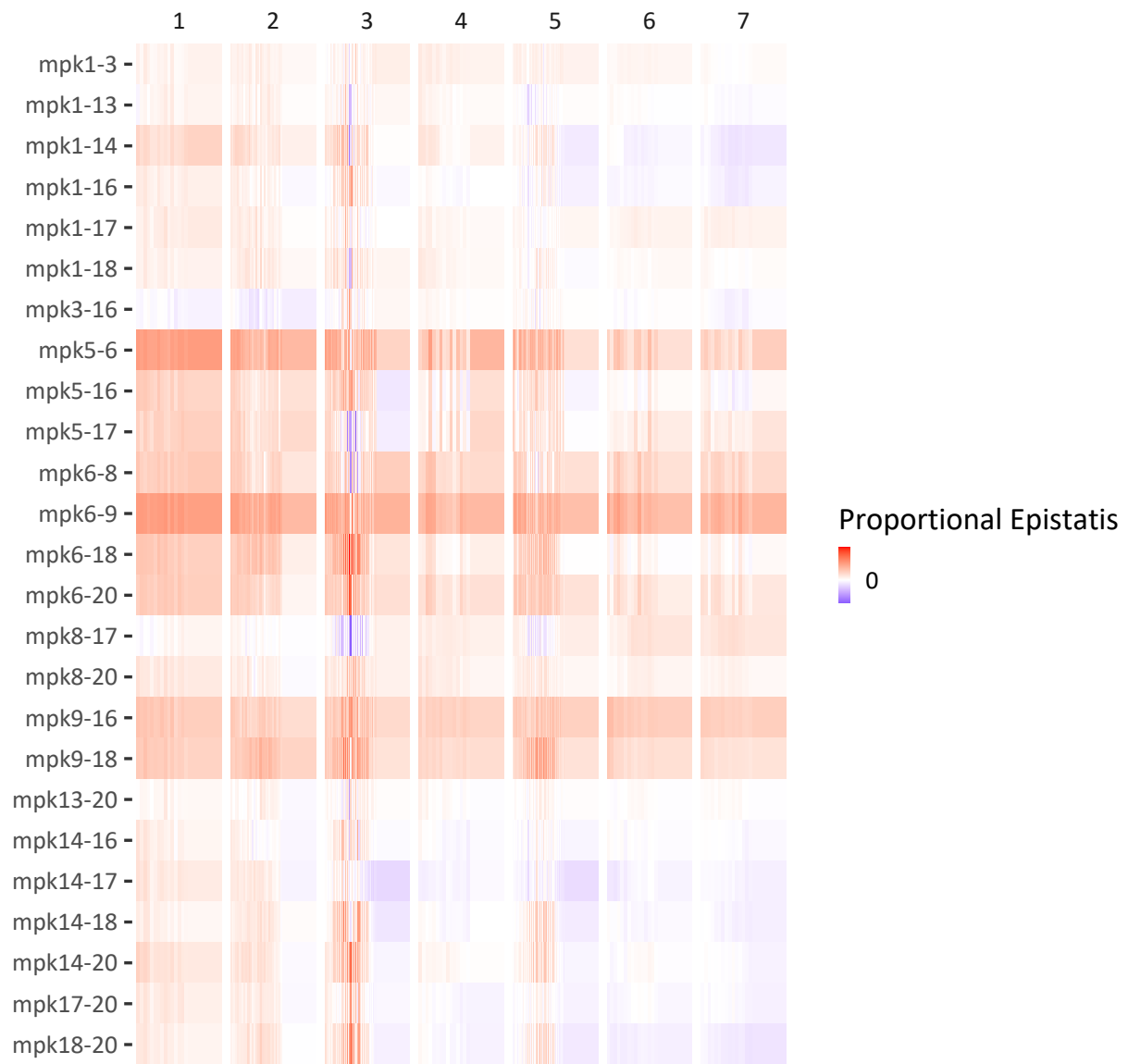
```

December:phi2 Proportional Epistasis

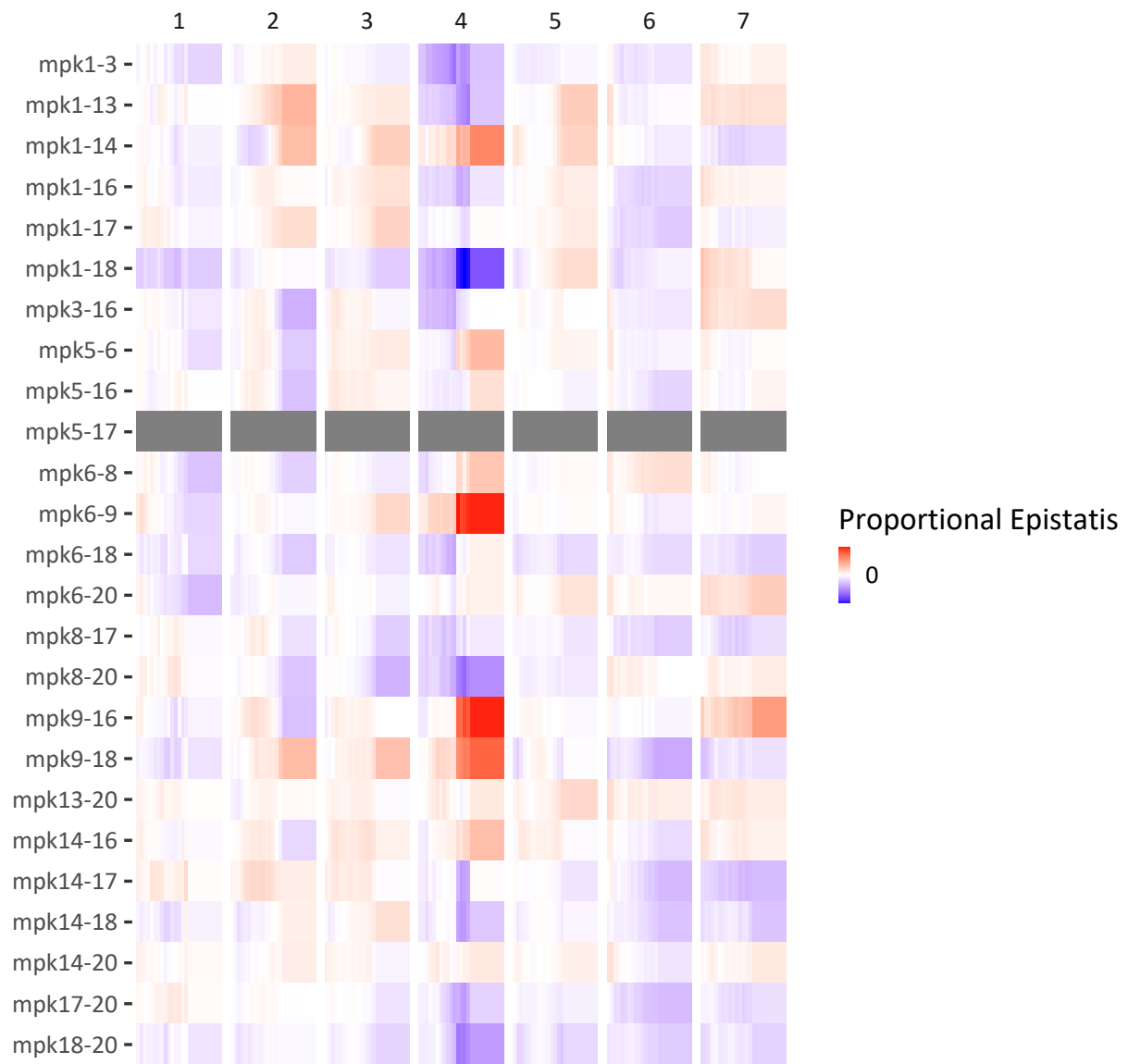


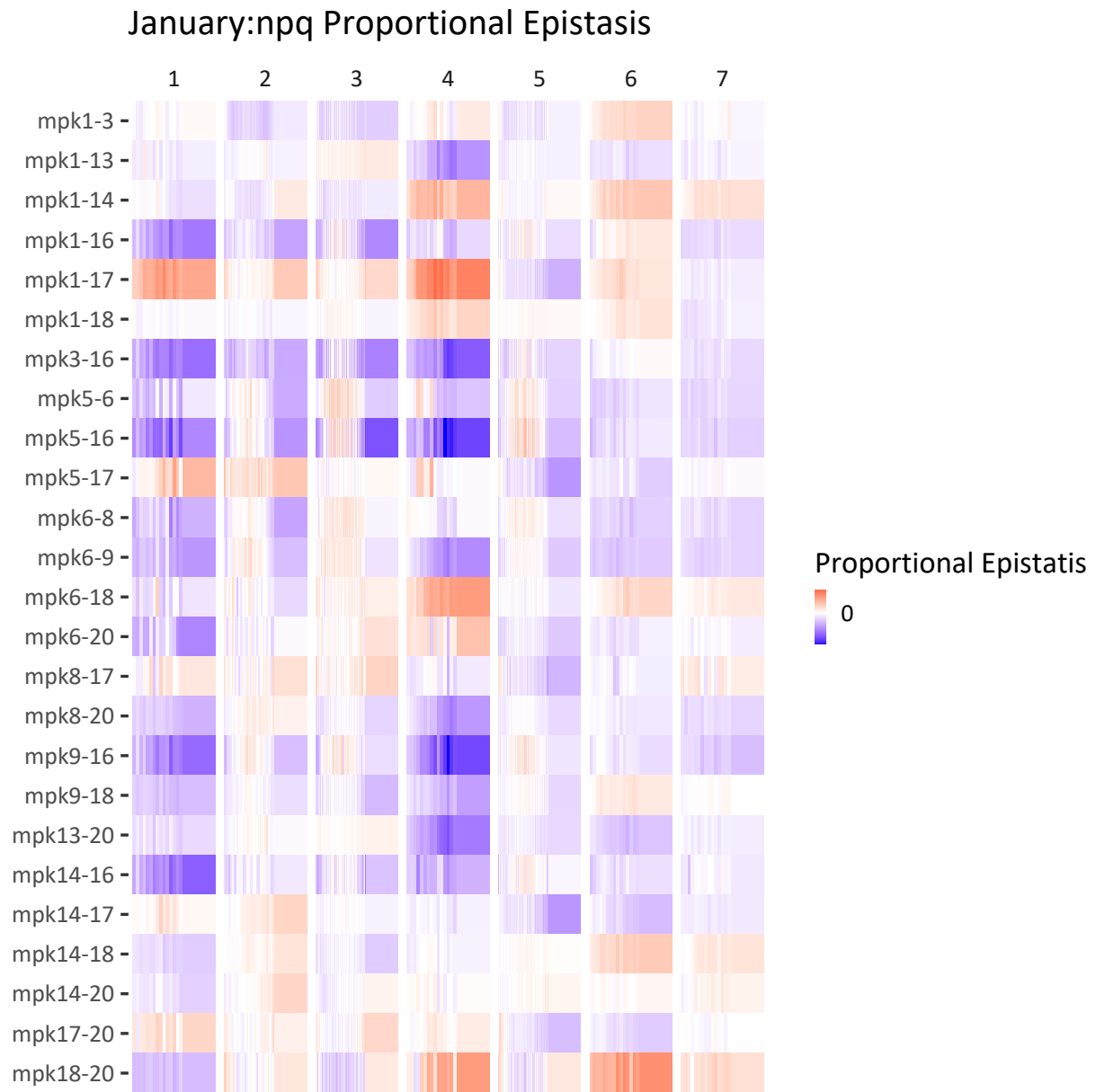


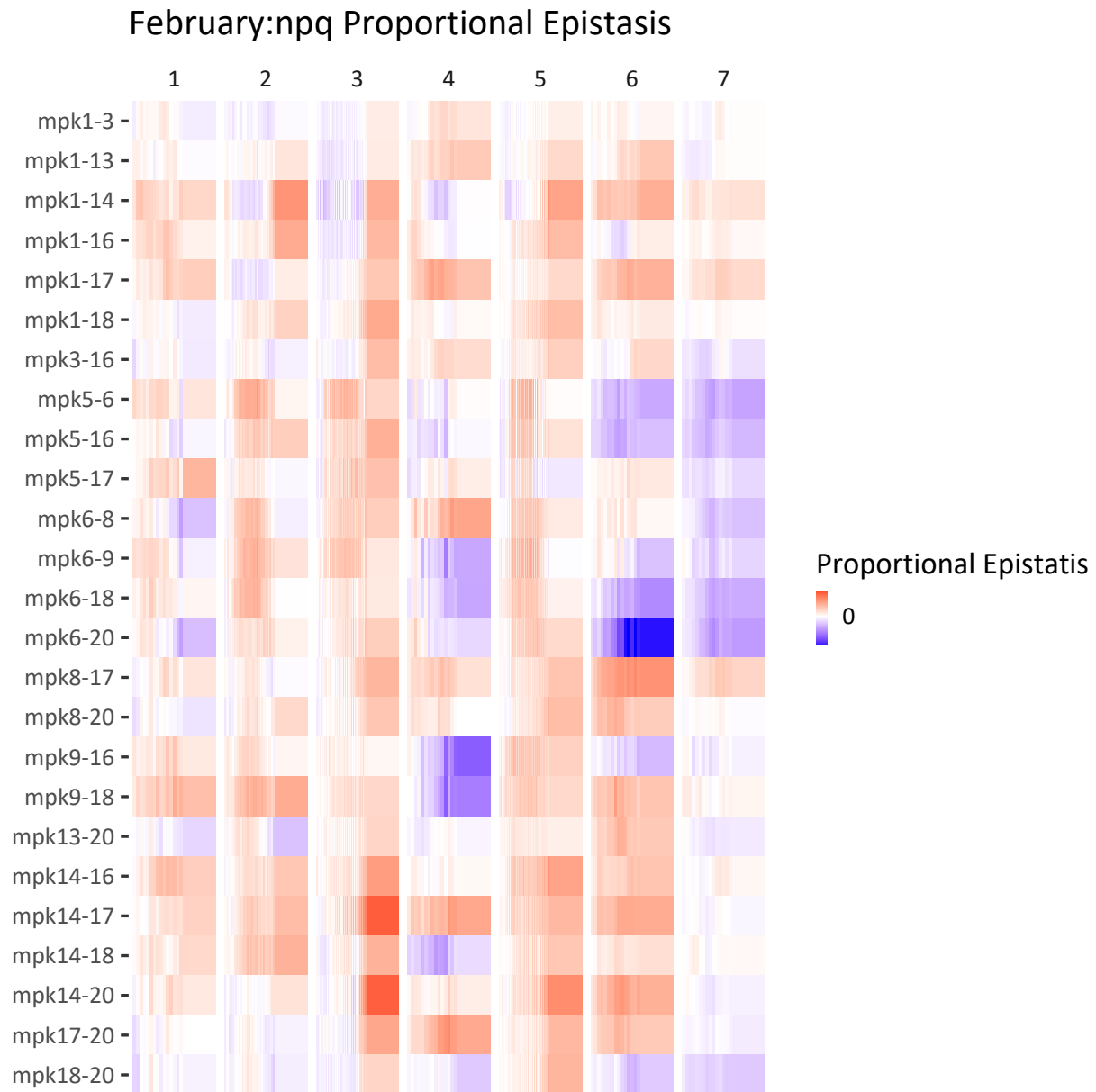
February:phi2 Proportional Epistasis



December:npq Proportional Epistasis







Epistasis Visualizations - Day 3

```
geneticInteractions_day3 <- geneticInteractions %>%
  filter(day == "3")
```

Additive

```
for (temp_measurement in c("phi2", "npq")) {
  set_bounds <- geneticInteractions_day3 %>%
```

```

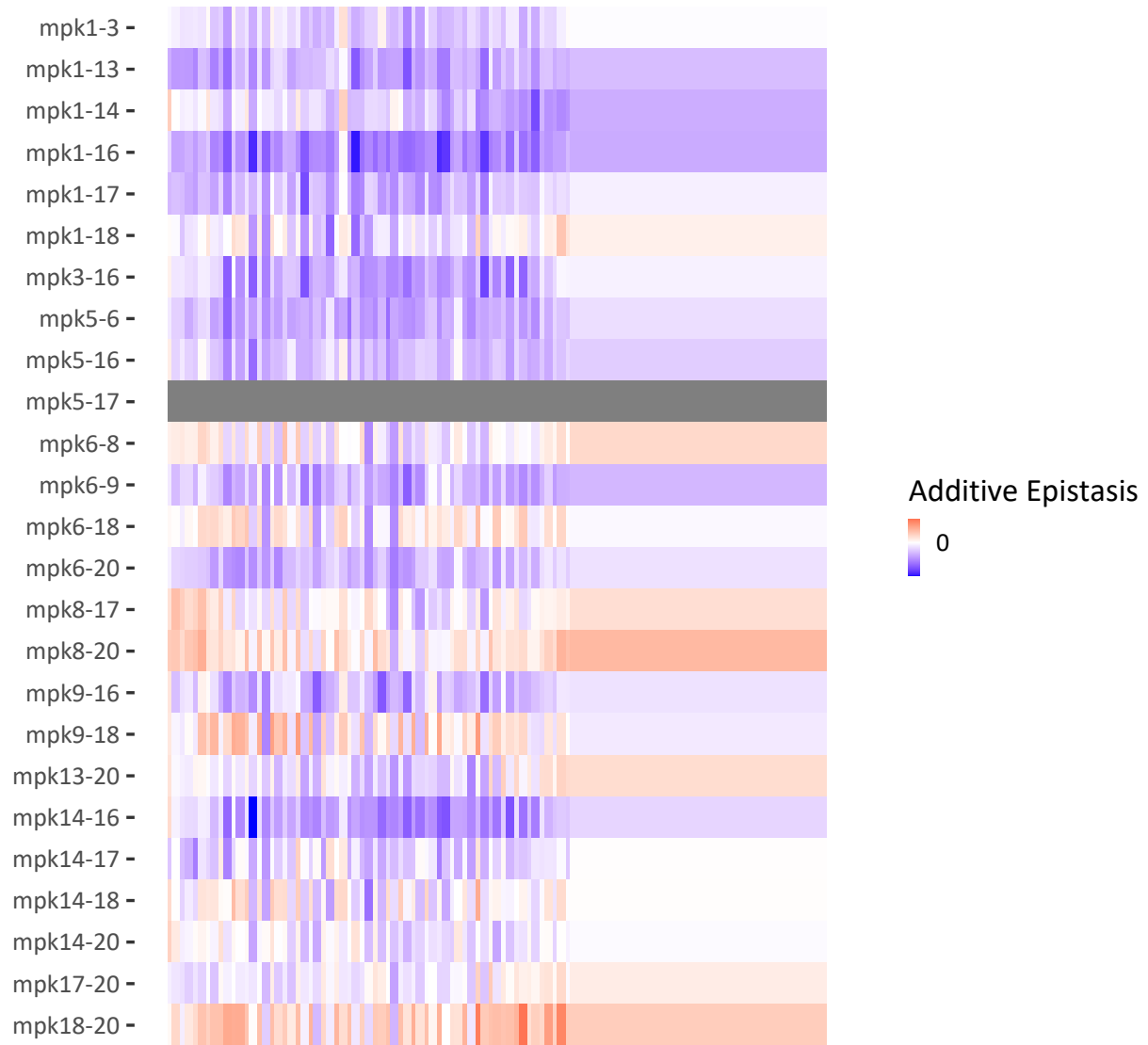
    filter(Measurement == temp_measurement)
lower_bound <- round(min(set_bounds$AdditiveEpistasis) -
0.05, 2)
upper_bound <- round(max(set_bounds$AdditiveEpistasis) +
0.05, 2)
for (temp_month in c("Dec", "Jan", "Feb")) {
  temp_plot_data <- filter(geneticInteractions_day3,
    Experiment == temp_month, Measurement ==
    temp_measurement)

  temp_title <- ifelse(temp_month ==
    "Dec", "December", ifelse(temp_month ==
    "Jan", "January", "February"))
  plot <- ggplot(data = temp_plot_data,
    aes(x = Time_Point, y = genotype,
    fill = AdditiveEpistasis)) +
    labs(fill = "Additive Epistasis",
    x = "Hours", y = NULL, title = paste(temp_title,
    ":", temp_measurement,
    " Additive Epistasis",
    sep = "")) + geom_tile(width = ifelse(temp_measurement ==
    "leafarea", 16, 10), height = 30) +
    facet_grid(genotype ~ day, scales = "free",
    switch = "y") + theme_tufte(base_family = "Calibri",
    base_size = 50) + theme(strip.background.y = element_blank(),
    strip.text.y = element_blank(),
    axis.title.x = element_blank(),
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    panel.spacing = unit(0, "lines")) +
    scale_fill_gradient2(low = "blue",
    high = "red", mid = "white",
    midpoint = 0, limits = c(lower_bound,
    upper_bound), breaks = c(lower_bound,
    0, upper_bound), labels = c(as.character(lower_bound),
    "0", as.character(upper_bound)))
  print(plot)
}
}

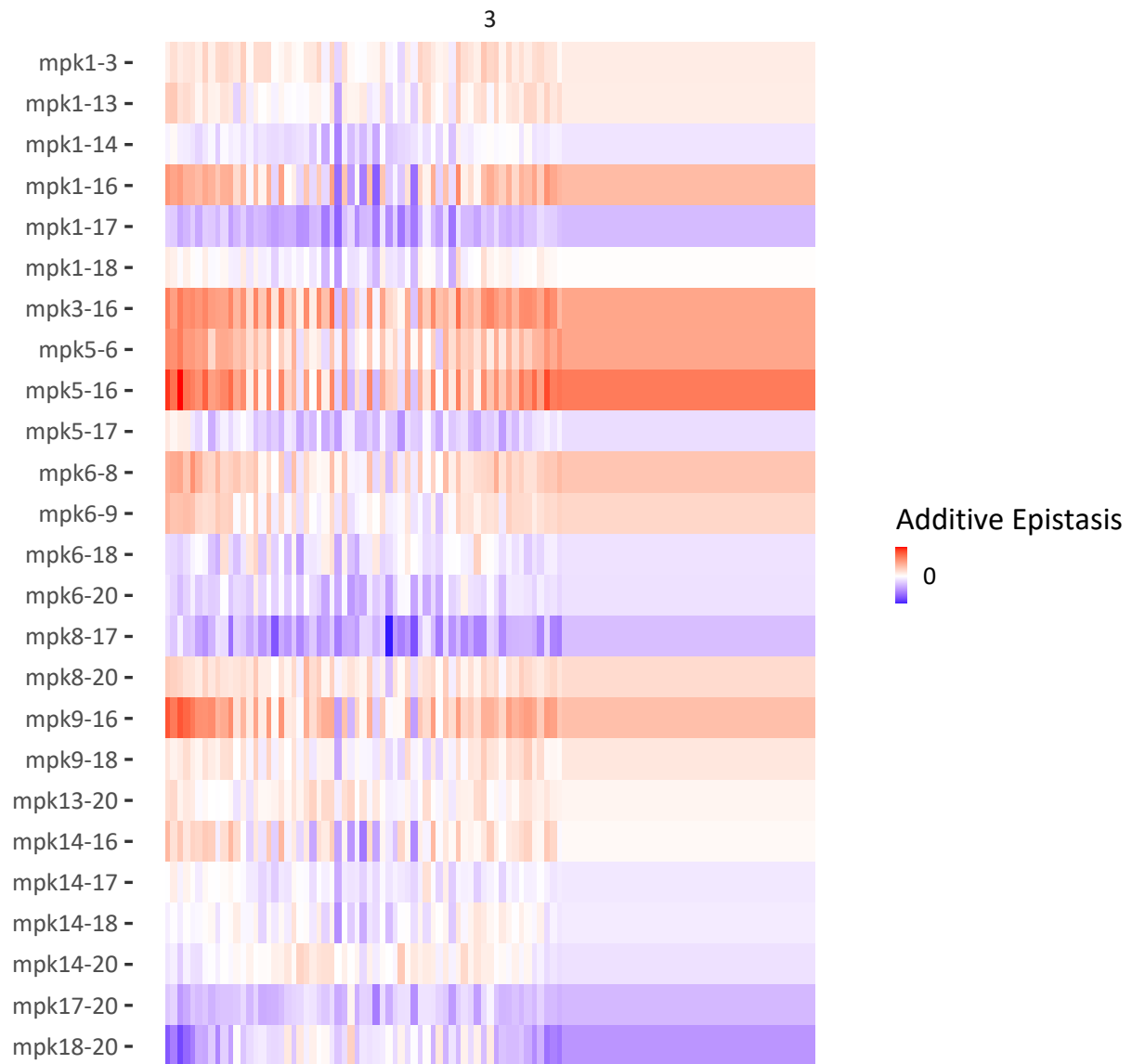
```

December:phi2 Additive Epistasis

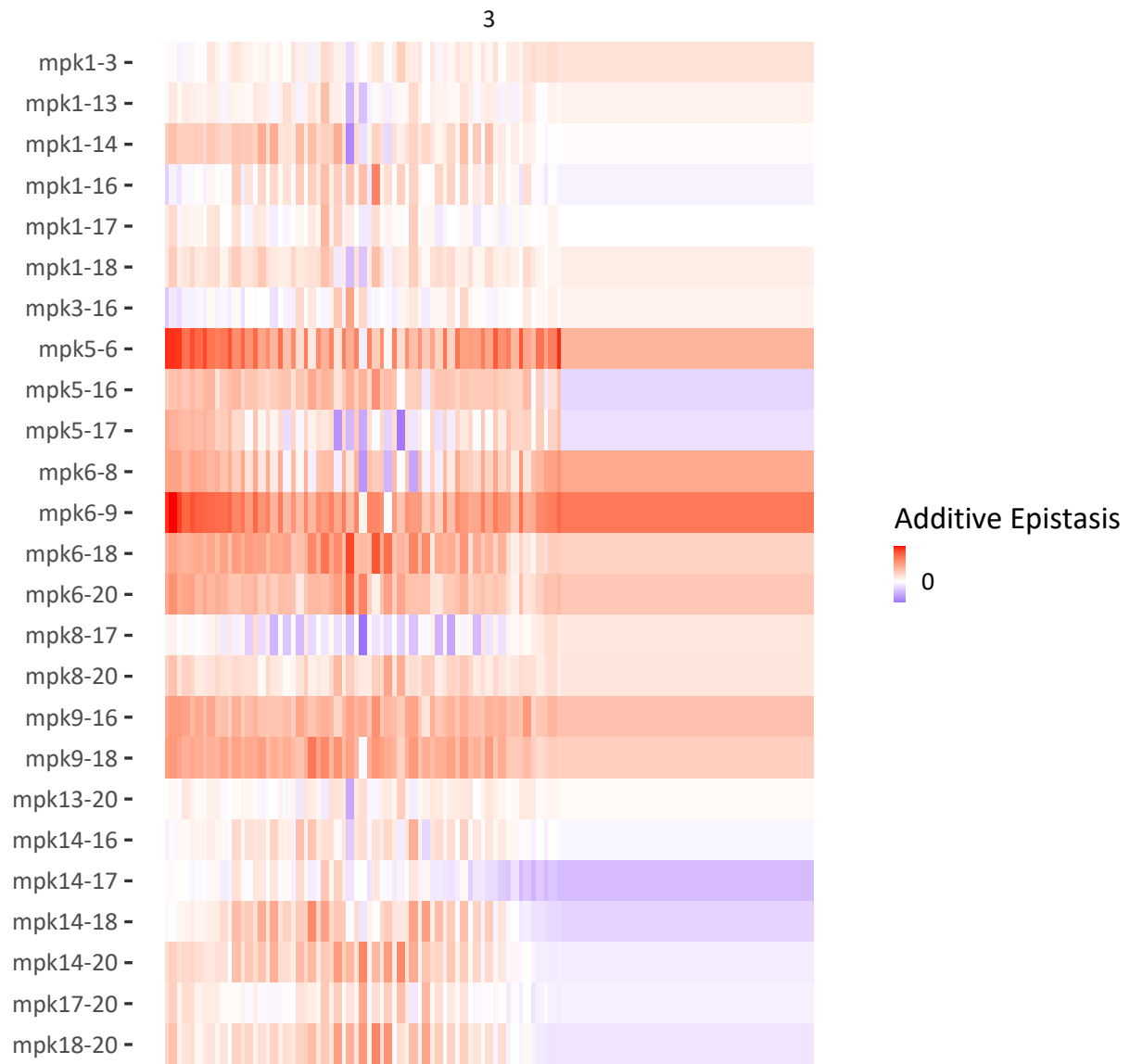
3



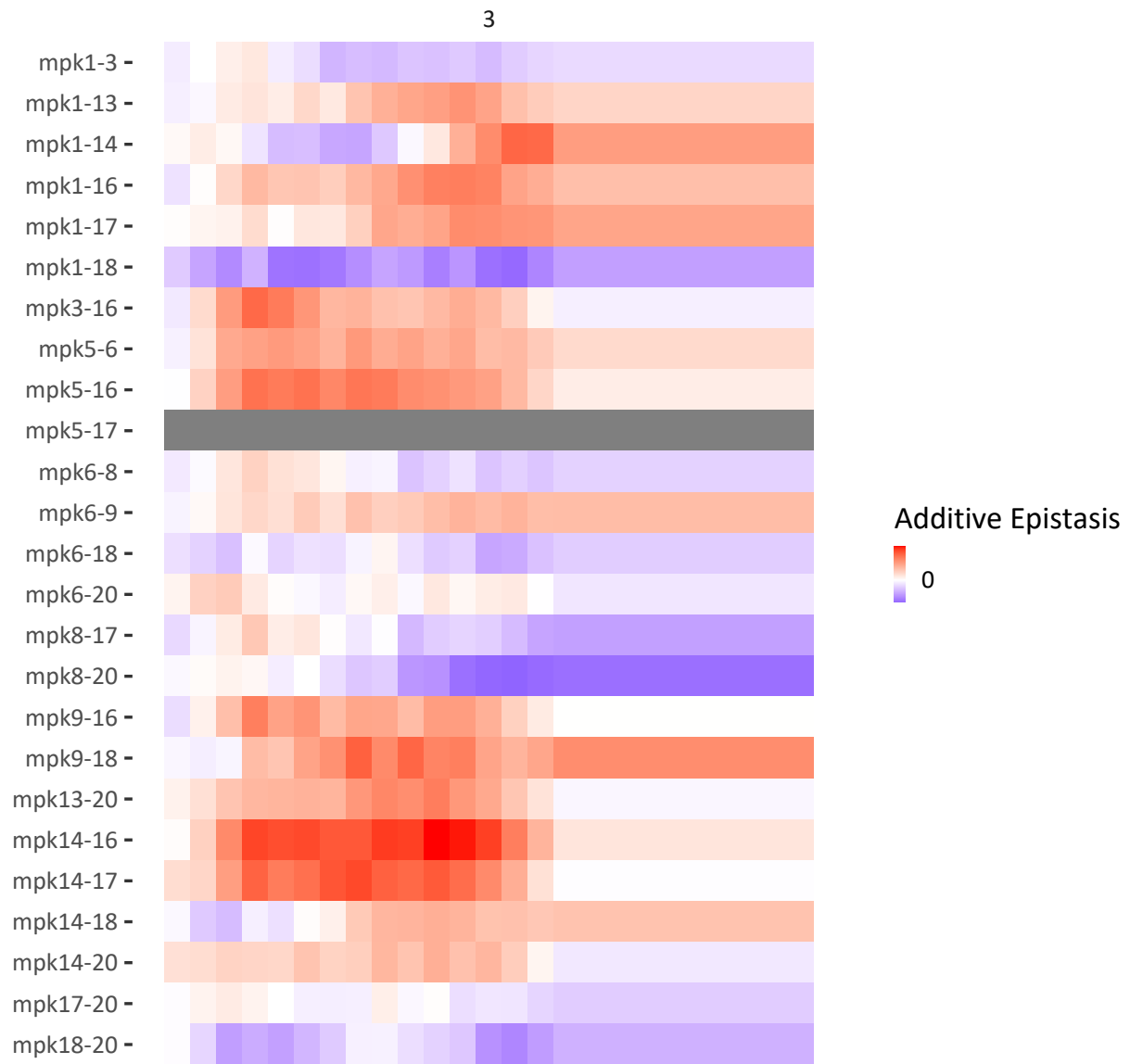
January:phi2 Additive Epistasis



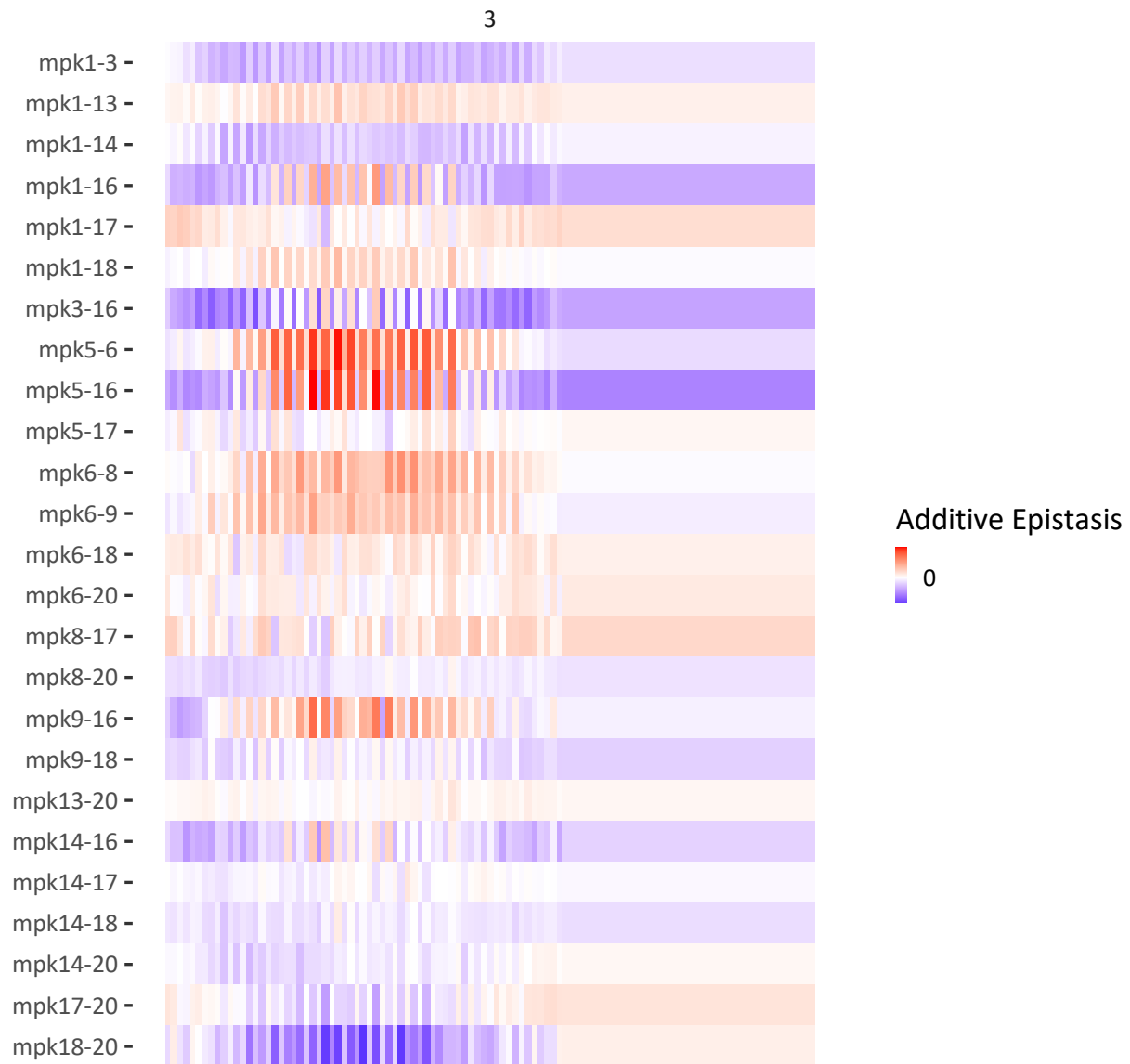
February:phi2 Additive Epistasis



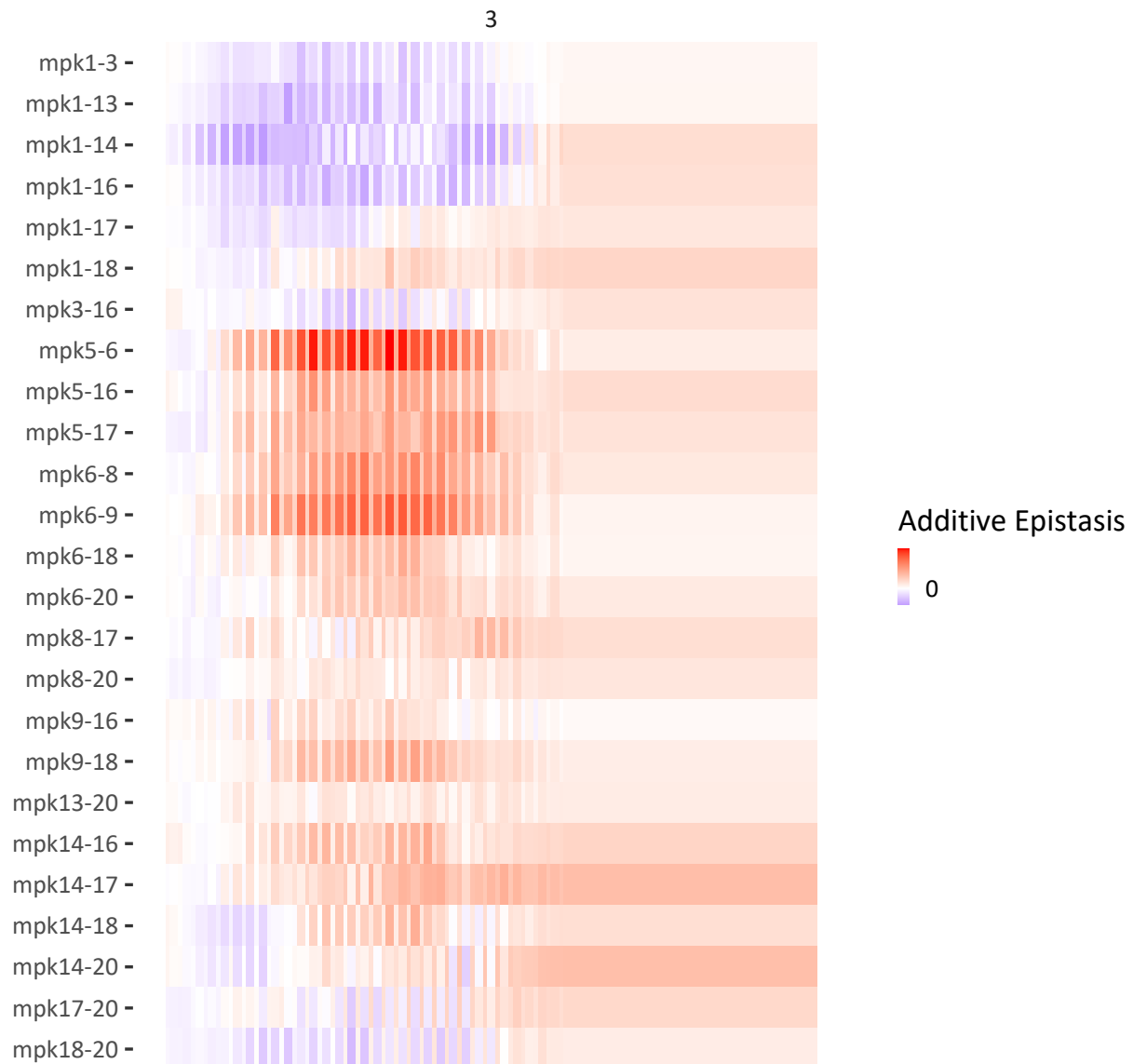
December:npq Additive Epistasis



January:npq Additive Epistasis



February:npq Additive Epistasis



Proportional

```
for (temp_measurement in c("phi2", "npq")) {
  set_bounds <- geneticInteractions_day3 %>%
    filter(Measurement == temp_measurement)
  lower_bound <- round(min(set_bounds$ProportionalEpistatis) -
    0.05, 2)
  upper_bound <- round(max(set_bounds$ProportionalEpistatis) +
    0.05, 2)
  for (temp_month in c("Dec", "Jan", "Feb")) {
    temp_plot_data <- filter(geneticInteractions_day3,
      Experiment == temp_month, Measurement ==
```

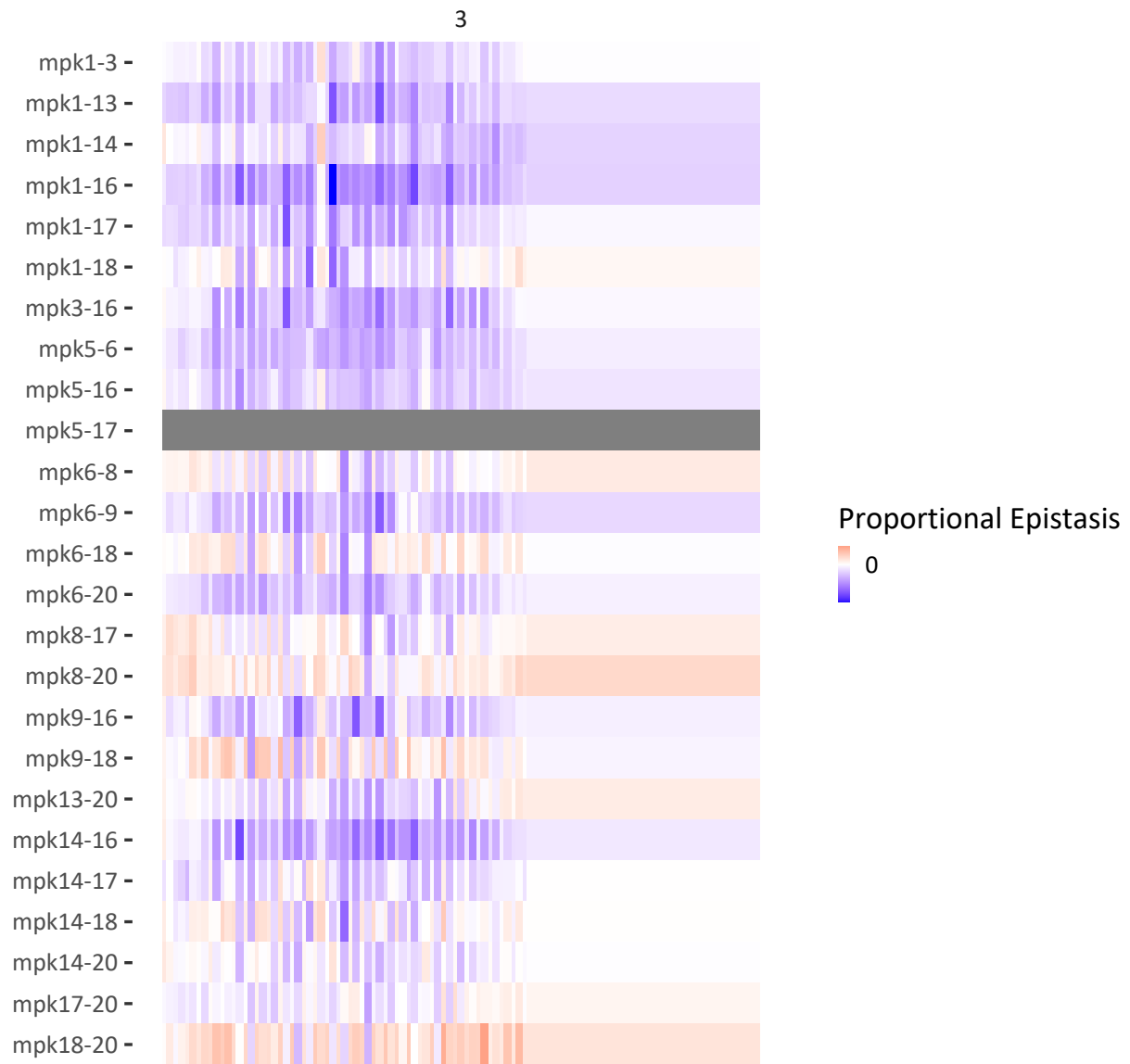
```

temp_measurement)

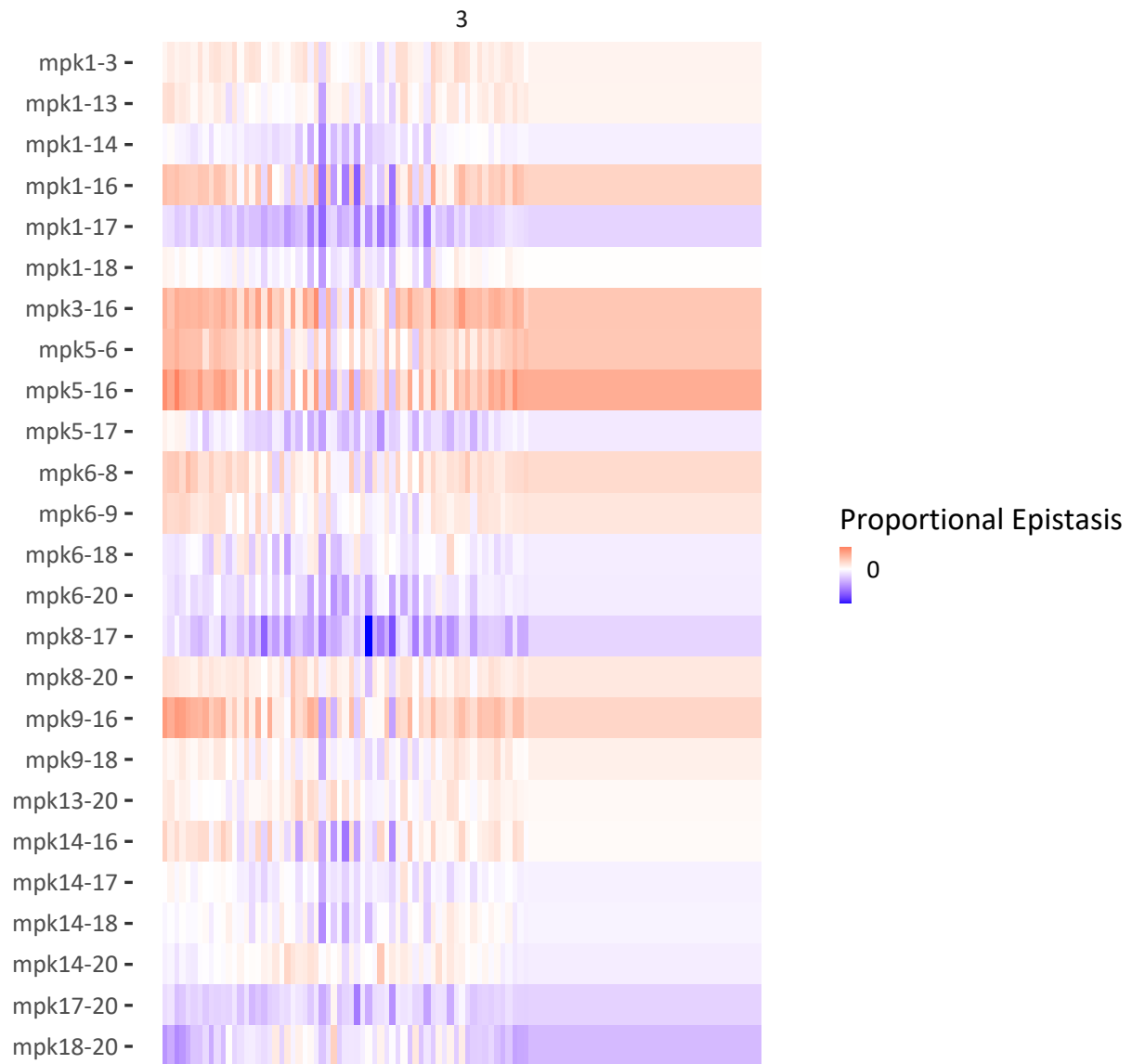
temp_title <- ifelse(temp_month ==
  "Dec", "December", ifelse(temp_month ==
    "Jan", "January", "February"))
plot <- ggplot(data = temp_plot_data,
  aes(x = Time_Point, y = genotype,
    fill = ProportionalEpistasis)) +
labs(fill = "Proportional Epistasis",
  x = "Hours", y = NULL, title = paste(temp_title,
    ":", temp_measurement,
    " Proportional Epistasis",
    sep = "")) + geom_tile(width = ifelse(temp_measurement ==
  "leafarea", 16, 10), height = 30) +
facet_grid(genotype ~ day, scales = "free",
  switch = "y") + theme_tufte(base_family = "Calibri",
  base_size = 50) + theme(strip.background.y = element_blank(),
  strip.text.y = element_blank(),
  axis.title.x = element_blank(),
  axis.text.x = element_blank(),
  axis.ticks.x = element_blank(),
  panel.spacing = unit(0, "lines")) +
scale_fill_gradient2(low = "blue",
  high = "red", mid = "white",
  midpoint = 0, limits = c(lower_bound,
    upper_bound), breaks = c(lower_bound,
    0, upper_bound), labels = c(as.character(lower_bound),
    "0", as.character(upper_bound)))
print(plot)
}
}

```

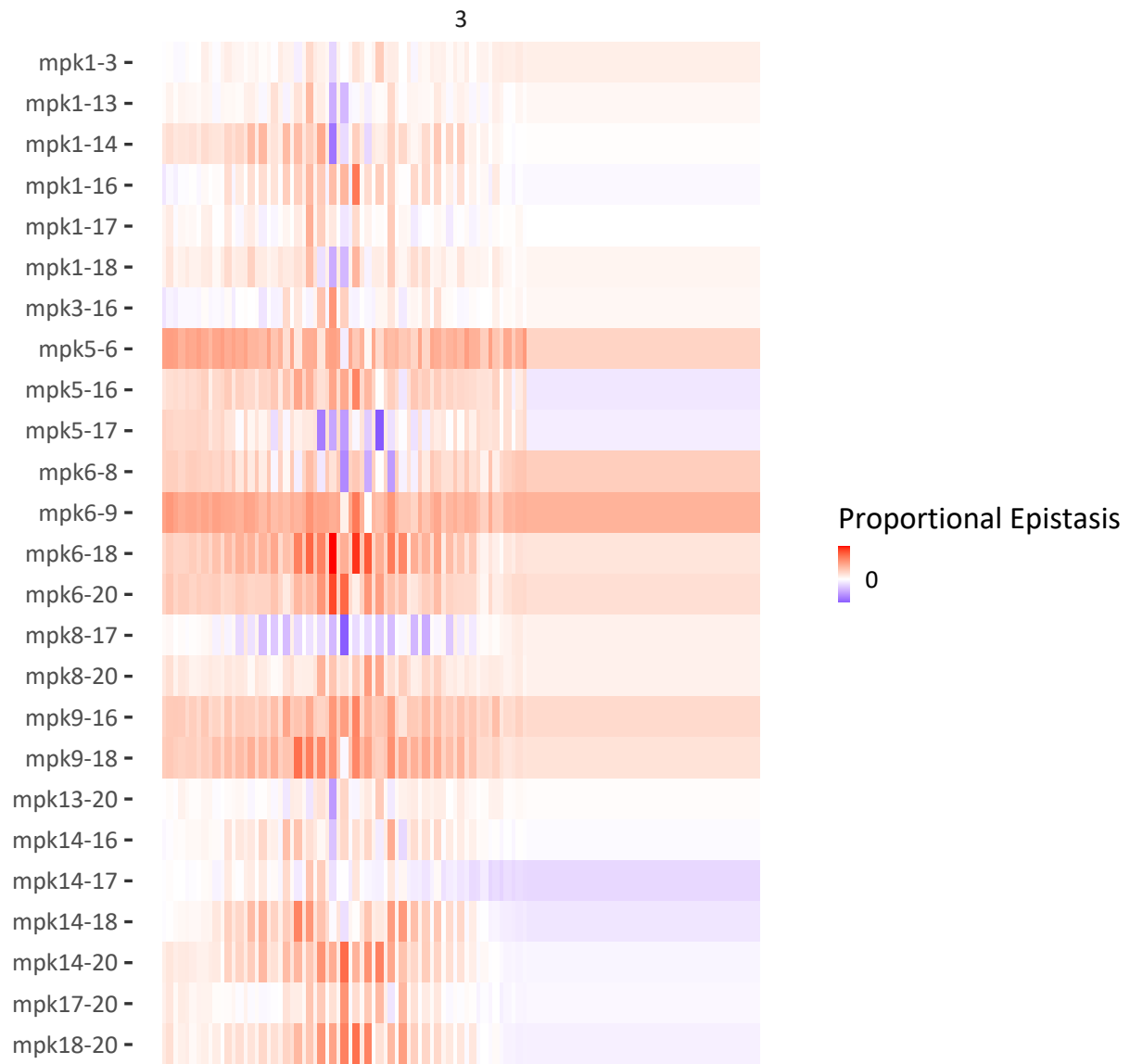
December:phi2 Proportional Epistasis



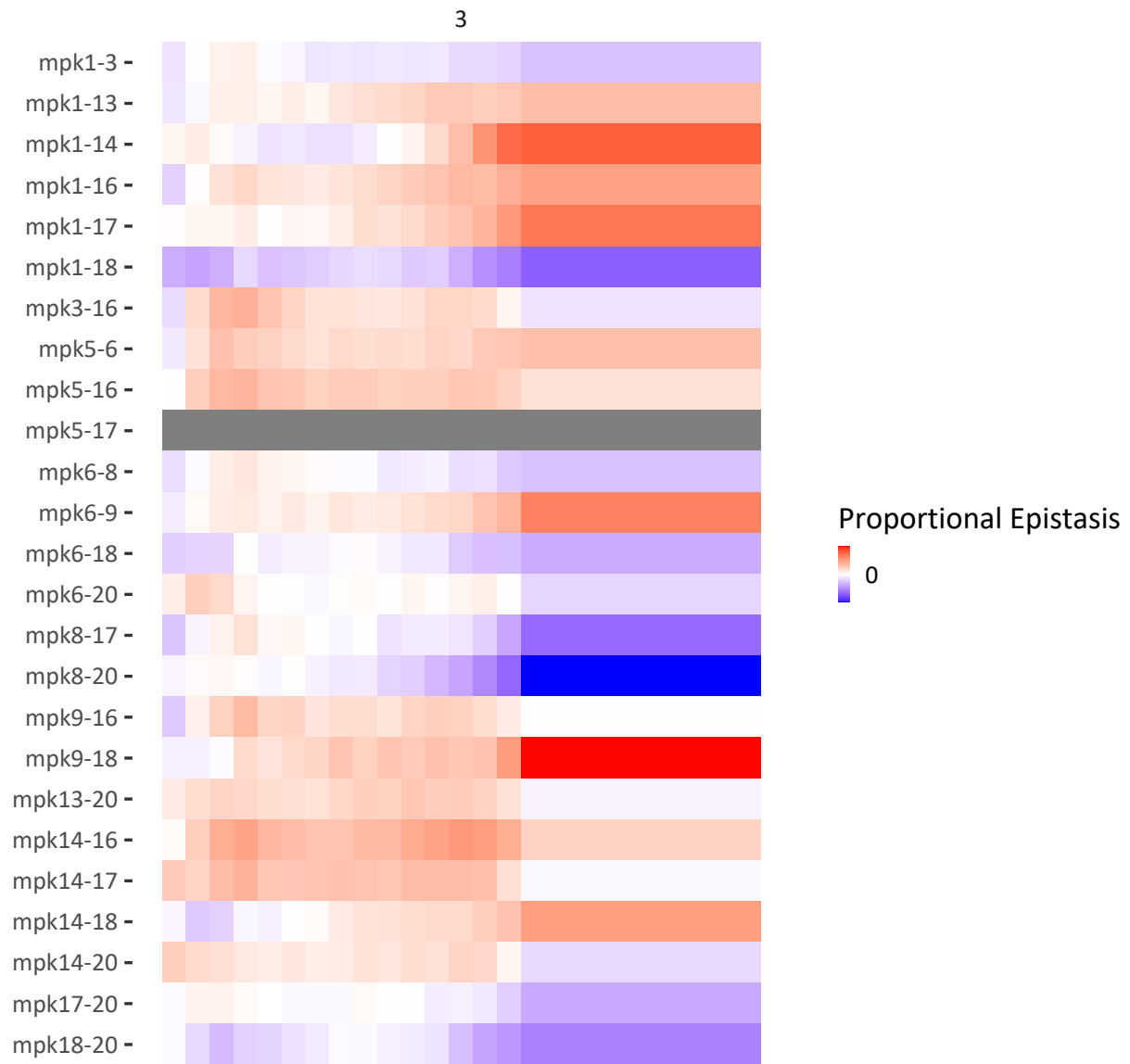
January:phi2 Proportional Epistasis



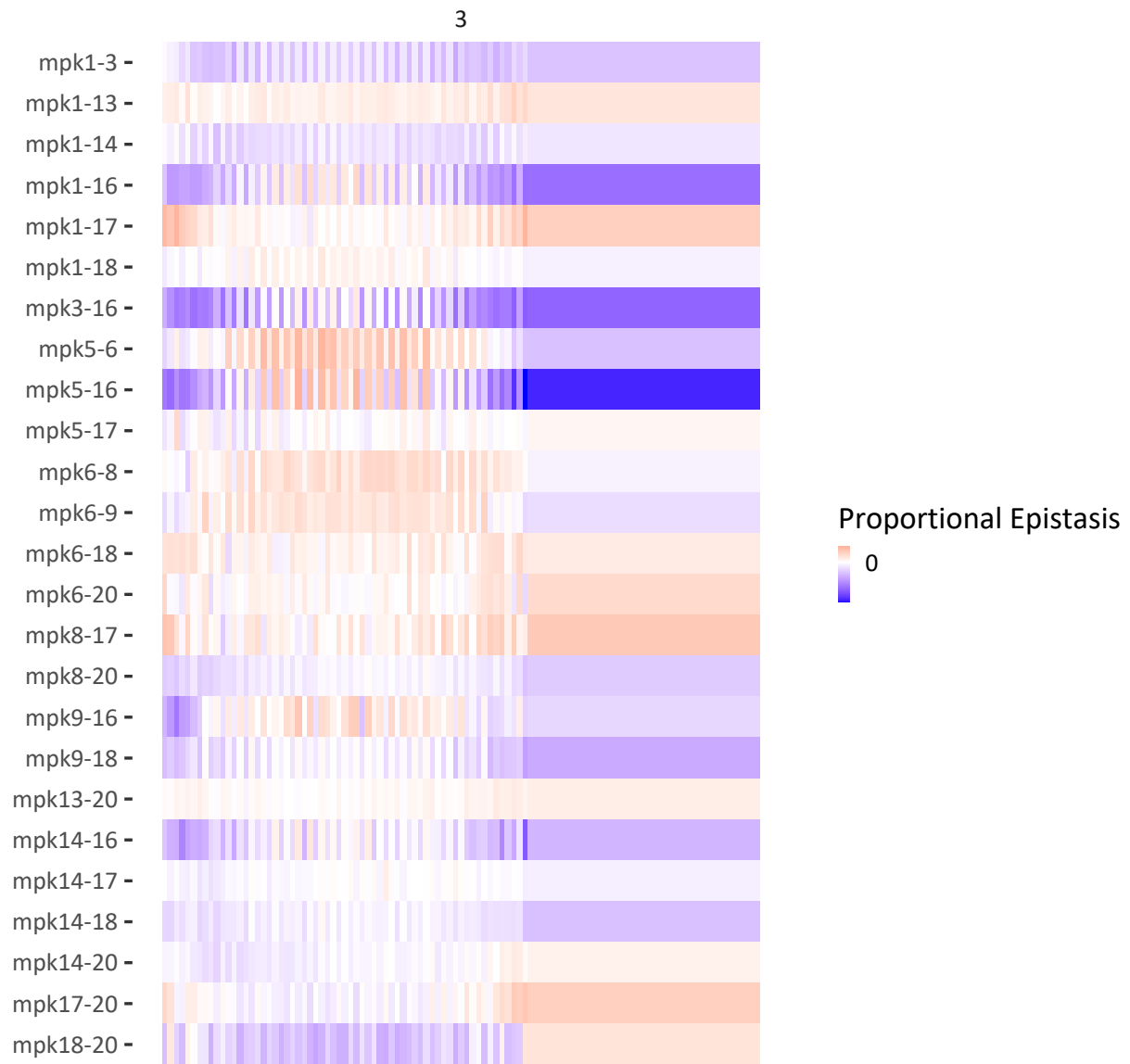
February:phi2 Proportional Epistasis



December:npq Proportional Epistasis



January:npq Proportional Epistasis



February:npq Proportional Epistasis

