library(GenomicRanges)

library(dplyr)

library(ggplot2)

# Load peak data

load("/Users/yulongqiu/Desktop/Biostats/Master\_Thesis/Peaks\_allsample/All\_sample\_Narrowpeaks/combined\_peaks2.RData")

load("/Users/yulongqiu/Desktop/Biostats/Master\_Thesis/Peaks\_bysample/Combined\_and\_reduced\_Granges/combined\_peaks.RData")

# Create data frame with length and strategy labels

all\_sample\_df <- data.frame(length = width(combined.peaks2), strategy = "All Sample")

by\_sample\_df <- data.frame(length = width(combined.peaks), strategy = "By Sample")

peak\_data <- rbind(all\_sample\_df, by\_sample\_df)

# Set bin width and assign bins

bin\_width <- 100

peak\_data <- peak\_data %>%

mutate(bin\_start = floor(length / bin\_width) \* bin\_width)

# Compute percentage for each bin

peak\_summary <- peak\_data %>%

group\_by(strategy, bin\_start) %>%

summarise(count = n(), .groups = "drop") %>%

group\_by(strategy) %>%

mutate(percent = count / sum(count) \* 100)

#Define the output folder for saving plots

output\_folder <- "/Users/yulongqiu/Desktop/Biostats/Master\_Thesis/Output"

peak\_summary$strategy <- factor(peak\_summary$strategy, levels = c("All Sample", "By Sample"))

p\_combined<-ggplot(peak\_summary, aes(x = bin\_start, y = percent, fill = strategy)) +

geom\_col(alpha = 0.8, width = bin\_width \* 0.9, show.legend = TRUE) +

facet\_grid(. ~ strategy) + # 横向排列分组图

scale\_fill\_manual(

name = "Strategy",

values = c("All Sample" = "skyblue", "By Sample" = "salmon"),

labels = c("All-samples method", "By-sample method")

) +

scale\_x\_continuous(

name = "Peak Length (bp)",

breaks = seq(0, 3000, by = 1000)

) +

ylab("Percentage of Peaks (%)") +

ggtitle("Peak Length Distribution by Strategy (Percentage)") +

theme\_minimal() +

theme(

strip.text = element\_blank(), # 取消 facet 上方标签

plot.title = element\_text(hjust = 0.5),

legend.title = element\_text(size = 12, face = "bold"),

legend.text = element\_text(size = 10)

)

ggsave(filename = file.path(output\_folder, "peak\_length\_distribution\_Percentage\_combined2.png"),

plot = p\_combined, width = 8, height = 6, dpi = 300)