**R code for Genome wide plots**

library(plotly)

library(dplyr)

library(readxl)

var\_plot <- function(path) {

vardetail <- read\_excel(path)

vardetail$Strand <- as.character(vardetail$Strand)

vardetail$Impact <- factor(vardetail$Impact)

p <- ggplot(vardetail, aes(x = Position, y = reorder(Chromosome, as.numeric(Chromosome)),

shape = `Variant Type`, color = `Region of Variation`, size = Impact)) +

geom\_point() +

labs(x = "Genomic Position", y = "Chromosome", shape = "Variant Type", color = "Region of Variation", size = "Impact") +

ggtitle("Genome-wide Plot of Variant Positions") +

geom\_point(aes(text = paste("Position: ", Position, "<br>",

"Alternate: ", Alternate, "<br>",

"Reference: ", Reference, "<br>",

"Read Depth: ", ifelse(is.na(`Read Depth`), "NA", as.character(`Read Depth`)), "<br>",

"Variant Type: ", ifelse(is.na(`Variant Type`), "NA", as.character(`Variant Type`)), "<br>",

"Region of Variation: ", ifelse(is.na(`Region of Variation`), "NA", as.character(`Region of Variation`)), "<br>",

"Gene\_ID: ", `Gene Name`, "<br>",

"Impact: ", Impact, "<br>",

"Start: ", `Start`, "<br>",

"Description: ", Description, "<br>",

"End: ", `End`, "<br>",

"Strand: ", Strand)),

color = "transparent") +

theme\_minimal() +

scale\_size\_manual(values = c("HIGH" = 4, "LOW" = 2, "MODERATE" = 3, "MODIFIER" = 1)) +

scale\_shape\_manual(values = c("snp" = 22, "mnp" = 24, "del" = 25, "ins" = 23, "complex" = 21))

p <- ggplotly(p, tooltip = "text")

p <- p %>% layout(legend = list(title = "Region of Variation"))

return(p)

}

final\_plot <- var\_plot("E:/602uniquesnps/final611describe.xlsx")

final\_plot

**R code for scatter plot:**

library(readxl)

library(ggplot2)

vard <- read\_excel("E:/distance plot/final611describe.xlsx")

vard <- vard[complete.cases(vard[c("Gene Start", "Position", "Gene End")]), ]

vard[c("Gene Start", "Gene End", "Position")] <- lapply(vard[c("Gene Start", "Gene End", "Position")], as.numeric)

vard$Variant\_3prime\_Distance <- ifelse(vard$Strand == "+",

vard$`Gene End` - vard$Position,

vard$Position - vard$`Gene Start`)

vard$Variant\_5prime\_Distance <- ifelse(vard$Strand == "+",

vard$Position - vard$`Gene Start`,

vard$`Gene End` - vard$Position)

negative\_strand\_indices <- vard$Strand == "-"

vard$Variant\_3prime\_Distance[negative\_strand\_indices] <- ifelse(!is.na(vard$`Gene End`[negative\_strand\_indices]) & !is.na(vard$Position[negative\_strand\_indices]),

vard$Position[negative\_strand\_indices] - vard$`Gene End`[negative\_strand\_indices],

NA)

vard$Variant\_5prime\_Distance[negative\_strand\_indices] <- ifelse(!is.na(vard$`Gene Start`[negative\_strand\_indices]) & !is.na(vard$Position[negative\_strand\_indices]),

vard$`Gene Start`[negative\_strand\_indices] - vard$Position[negative\_strand\_indices],

NA)

ggplot(vard, aes(x = Variant\_3prime\_Distance, y = Variant\_5prime\_Distance, color = Chromosome)) +

geom\_point() +

labs(x = "Variant 3' Distance to Next Gene", y = "Variant 5' Distance from Next Gene",

title = "Af\_6\_1\_1 Variants Distance to Next Gene", color = "Chromosome") +

theme\_minimal()