Graphs for Amita's paper - with data from U Laval

Karen Cristine Goncalves dos Santos

4/9/2020

```r
org.adj.pVal <- read.delim("qval_zscore_stat.txt", header = T, stringsAsFactors = F)

qval <- org.adj.pVal[, c("Accession", grep("qvalue", colnames(org.adj.pVal), value = T))]
colnames(qval) <-  gsub("_qvalue", "", colnames(qval))

zscore_ratio <- org.adj.pVal[, c("Accession", grep("ZscRatio", colnames(org.adj.pVal), value = T))]
colnames(zscore_ratio) <- gsub("_ZscRatio", "", colnames(zscore_ratio))

status <- org.adj.pVal[, c("Accession", grep("stat", colnames(org.adj.pVal), value = T))]
colnames(status) <- gsub("_statVariant", "", colnames(status))
```

```r
melt_qval <- melt(qval, id = "Accession") %>% cbind(Variable = "qValue")
melt_zscore <- melt(zscore_ratio, id = "Accession") %>% cbind(Variable = "zscore_ratio")
melt_status <- melt(status, id = "Accession") %>% cbind(Variable = "status")
colnames(melt_qval)[2] <- colnames(melt_zscore)[2] <- colnames(melt_status)[2] <- "Test"

for_plots <- data.frame(melt_qval, melt_zscore %>% select(value), melt_status %>% select(value))
colnames(for_plots) <- c("Accession", "Test", "qValue", "variable", "Zscore_ratio", "Status")
for_plots$variable = NULL

Total <- Down_regulated <- Up_regulated <- c()
summary_plot <- data.frame(matrix(nrow = 0, ncol = 3))
dap <- list()
for (i in levels(for_plots$Test)) {
  dap[[i]] <- status[status[, i] == "VARIANT", "Accession"]
  Total <- length(dap[[i]])
  Up_regulated <- length(zscore_ratio[ zscore_ratio$Accession %in% dap[[i]] & zscore_ratio[, i] < -1.96
  Down_regulated <- length(zscore_ratio[ zscore_ratio$Accession %in% dap[[i]] & zscore_ratio[, i]  > 1.9
  summary_plot <- rbind(summary_plot, c(Total, Up_regulated, Down_regulated)) %>% as.data.frame()
  colnames(summary_plot) <- c("Total", "Up_regulated", "Down_regulated")

}
rownames(summary_plot) <- summary_plot$Test <- levels(for_plots$Test)
```

```r
plot_summary <- reshape2::melt(summary_plot, id = "Test")
plot_summary$Test <- factor(plot_summary$Test,
                            levels = c("IFN.Control", "Virus.Control", "IFN_Virus.Control",
                                       "IFN_Virus.IFN", "IFN_Virus.Virus"))
ggplot(plot_summary, aes(Test, value, fill = variable, color = variable)) +
  geom_col(position = "dodge2") +
  geom_text(aes( label = value, y = value + 10), position = position_dodge(0.9), color = "black") +
```

```
scale_color_manual(name = "", values = c("grey", "red", "blue"))  +
scale_fill_manual(name = "", values = c("grey", "red", "blue")) +
labs(y = "Number of variant proteins", x = "Comparison") +
theme_bw() + theme(axis.text.x = element_text(angle = 90, color = "black"))
```
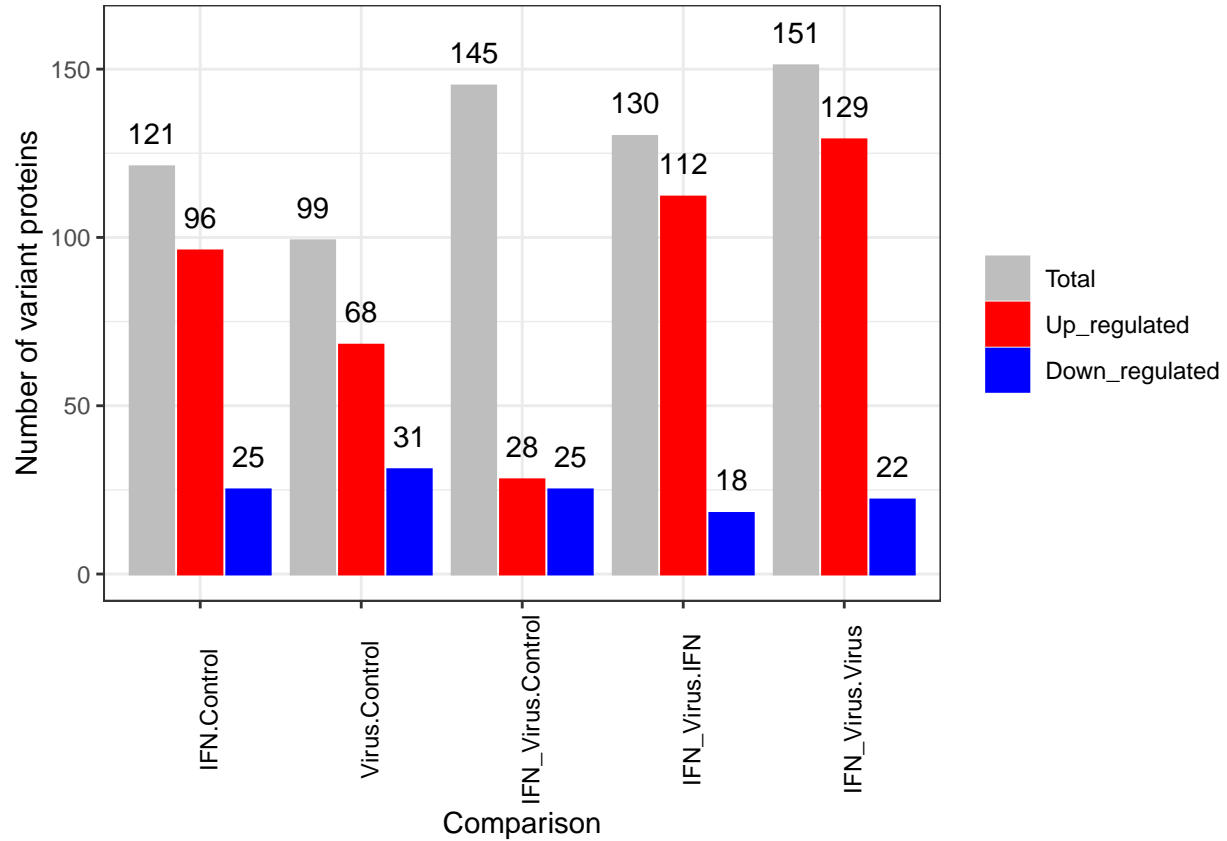

To do the boxplot of the data before normalization, I got the normalization factors from page 3 of the report as I do not have the actual data before normalization. However, as the data was re-analysed, maybe those are not the right normalization factors.

Sample	Normalization factors
CTRL1	1.13
CTRL2	0.96
CTRL3	1.05
IFN1	1.01
IFN2	0.97
IFN3	0.95
VI1	1.08
VI2	0.97
VI3	1.01
NL1	1
NL2	1
NL3	1

```r
normalized_abundances <- read.delim("normalized_abundances.txt", header = T, stringsAsFactors = F)
normalized_abundances$gene_name <- gsub(".*GN=(\\w+).*", "\\1", normalized_abundances$Description)
norm_abundances <- data.frame(normalized_abundances[,3:14] %>% log2(),
                              row.names = normalized_abundances$Accession)

plot_boxplot <- melt(norm_abundances)


## No id variables; using all as measure variables

plot_boxplot$Test <- plot_boxplot$variable <- as.character(plot_boxplot$variable)

for (i in 1:nrow(plot_boxplot)) {
  if (grepl("CTRL", plot_boxplot$Test[i])) {
    plot_boxplot$Test[i] = "Control"
  } else if (grepl("^IFN\\d$", plot_boxplot$Test[i])) {
    plot_boxplot$Test[i] = "IFN"
  } else if (grepl("^VIRUS\\d$", plot_boxplot$Test[i])) {
    plot_boxplot$Test[i] = "NL"
    plot_boxplot$variable[i] = gsub("VIRUS", "NL", plot_boxplot$variable[i])
  } else {
    plot_boxplot$Test[i] = "VI"
    plot_boxplot$variable[i] = gsub("IFN_VIRUS", "VI", plot_boxplot$variable[i])
  }
}
plot_boxplot$Normalization <- "After Normalization"


normalization_factors <- c(
  "CTRL1" = 1.13, "CTRL2" = 0.96, "CTRL3" = 1.05,
  "IFN1" = 1.01, "IFN2" = 0.97, "IFN3" = 0.95,
  "VI1" = 1.08, "VI2" = 0.97, "VI3" = 1.01,
  "NL1" = 1, "NL2" = 1, "NL3" = 1)

org_abundances <- data.frame(0,0)

for (i in names(normalization_factors)) {
  if (grepl("VI", i)){
    a <- gsub("VI", "IFN_VIRUS", i)
    org_abundances <- data.frame(cbind(org_abundances, (normalized_abundances[, a] * normalization_fact
    colnames(org_abundances)[ncol(org_abundances)] <- a

  } else if (grepl("NL", i)){
    a <- gsub("NL", "VIRUS", i)
    org_abundances <- data.frame(cbind(org_abundances, (normalized_abundances[, a] * normalization_fact
    colnames(org_abundances)[ncol(org_abundances)] <- a
  } else {
    org_abundances <- data.frame(cbind(org_abundances, (normalized_abundances[, i] * normalization_fact
    colnames(org_abundances)[ncol(org_abundances)] <- i
  }
}
org_abundances$X0 <- org_abundances$X0.1 <- NULL

rownames(org_abundances) <- normalized_abundances$Accession
```

```r
plot_boxplot_org <- melt(org_abundances)
```

```
## No id variables; using all as measure variables
```

```r
plot_boxplot_org$Test <- plot_boxplot_org$variable <- as.character(plot_boxplot_org$variable)

for (i in 1:nrow(plot_boxplot_org)) {
  if (grepl("CTRL", plot_boxplot_org$Test[i])) {
    plot_boxplot_org$Test[i] = "Control"
  } else if (grepl("^IFN\\d$", plot_boxplot_org$Test[i])) {
    plot_boxplot_org$Test[i] = "IFN"
  } else if (grepl("^VIRUS\\d$", plot_boxplot_org$Test[i])) {
    plot_boxplot_org$Test[i] = "NL"
    plot_boxplot_org$variable[i] = gsub("VIRUS", "NL", plot_boxplot_org$variable[i])
  } else {
    plot_boxplot_org$Test[i] = "VI"
    plot_boxplot_org$variable[i] = gsub("IFN_VIRUS", "VI", plot_boxplot_org$variable[i])
  }
}
plot_boxplot_org$Normalization <- "Before Normalization"
```

```r
plot_boxplot_org$variable <- factor(plot_boxplot_org$variable,
                             levels = c(paste0("CTRL", 1:3), paste0("IFN", 1:3),
                                        paste0("VI", 1:3), paste0("NL", 1:3)))
plot_boxplot_org$Test <- factor(plot_boxplot_org$Test, levels = c("Control", "IFN", "VI", "NL"))

joined_plots <- rbind(plot_boxplot, plot_boxplot_org)
joined_plots$Normalization <- factor(joined_plots$Normalization,
                               levels = c("Before Normalization", "After Normalization"))

ggplot(joined_plots, aes(variable, value, fill = Test)) +
  geom_boxplot() + scale_fill_manual(values = c("#db8823", "#106909", "#8f2424", "#1082b3"))+
  labs(x = "", y = expression(log[2](Intensity))) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, color = "black", hjust = 1, vjust = 0.5),
        legend.position = "none") +
  facet_wrap(~Normalization, nrow = 2)
```