

# Phase 1 - HackBio Data Contest

Tasneem Juzer

Hello, this R notebook contains all the tasks given in phase one of the data contest. It's a compilation of all the R codes and their results

```
knitr::opts_chunk$set(fig.width=8, fig.height=8)
```

For figure 1a - 1f

```
par(mfrow = c(2,3))

data <- read.csv("fig_one_a_e.csv", header=TRUE)
data_color <- c("red", "blue", "green", "orange", "purple", "pink",
               "brown", "gray", "cyan", "magenta")
technology1 <- unique(data$tech)
# fig_1a
plot(0,0,
     type = "n",
     xlim = c(5,40),
     ylim = c(0,12000),
     xlab = "Sequencing Depth",
     ylab = "Unique Fragment in Peaks",
     main = "Fig 1a",
     las = 1)
for (i in 1:10) {
  subset_data <- data [data$tech == technology1[i],]
  lines(subset_data$depth, subset_data$Unique_nr_frag_in_regions,
        col = data_color[i], pch = 19, lwd = 3, lty = 2)
}
#fig 1b
plot(0,0,
     type = "n",
     xlim = c(5,40),
     ylim = c(0,40),
     xlab = "Sequencing Depth",
     ylab = "Unique Fragment in Peaks",
     main = "Fig 1b",
     las = 1)

for (i in 1:10) {
  subset_data <- data [data$tech == technology1[i],]
```

```

    lines(subset_data$depth, subset_data$TSS_enrichment,
          col = data_color[i], pch = 23, lty = 2, lwd = 3)
}

#fig c

plot(0,0,
     type = "n",
     xlim = c(5,40),
     ylim = c(0,0.5),
     xlab = "Sequencing Depth",
     ylab = "Sequencing Efficiency",
     main = "Fig 1c",
     las = 1)

for (i in 1:10) {
  subset_data <- data [data$tech == technology1[i],]
  lines(subset_data$depth, subset_data$X._unique_nr_frag_in_regions_in_cells,
        col = data_color[i], pch = 23,lty = 2, lwd = 3)
}

#fig d

plot(0,0,
     type = "n",
     xlim = c(5,40),
     ylim = c(0,0.9),
     xlab = "Sequencing Depth",
     ylab = "Median Seurat Score",
     main = "Fig 1d",
     las = 1)

for (i in 1:10) {
  subset_data <- data [data$tech == technology1[i],]
  lines(subset_data$depth, subset_data$median_cell_type_pred_score,
        col = data_color[i], pch = 23, lwd = 3)
}

#fig e

plot(0,0,
     type = "n",
     xlim = c(5,40),
     ylim = c(0,35),
     xlab = "Sequencing Depth",
     ylab = "BCell Strength",
     main = "Fig 1e",
     las = 1)

for (i in 1:10) {

```

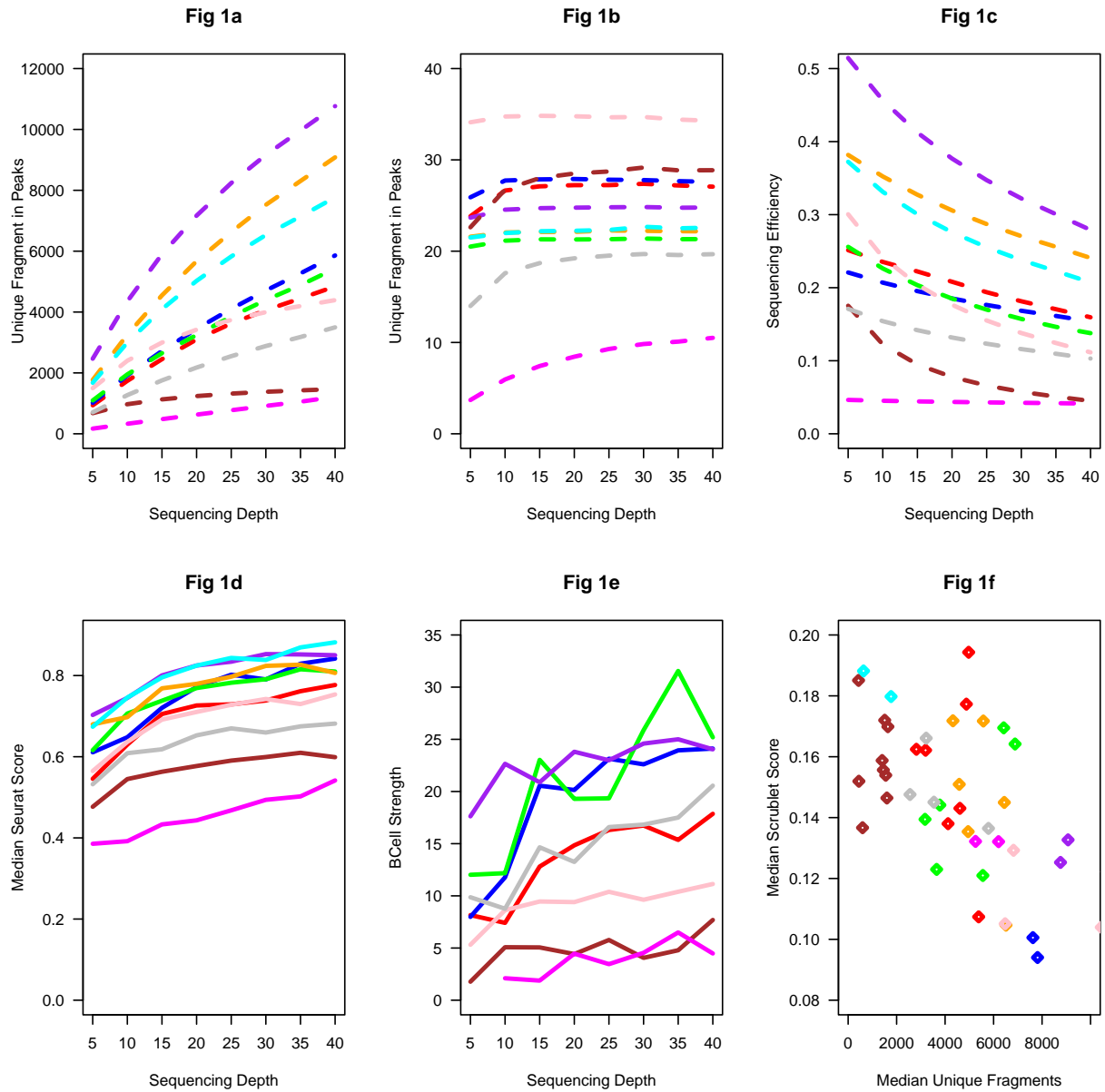
```

subset_data <- data [data$tech == technology1[i],]
lines(subset_data$depth, subset_data$fc__B_cell,
      col = data_color[i], lwd =3)
}

# fig f
data_f <- read.csv("fig_one_f.csv", header=TRUE)
technology2 <- unique(data_f$technology)
data_color1 <- c("red", "blue", "green", "orange", "purple", "pink", "brown", "gray", "cyan", "magenta")
plot(0,0,
     type = "n",
     xlim = c(0, 10000),
     ylim = c(0.08, 0.2),
     xlab = "Median Unique Fragments",
     ylab = "Median Scrublet Score",
     main = "Fig 1f",
     las = 1)

for (i in 1:10) {
  subset_data1 <- data_f [data_f$technology == technology2[i],]
  points(subset_data1$Median_Unique_nr_frag_in_regions,
        subset_data1$Mean_scrublet_doublet_scores_fragments,
        col = data_color1[i], pch = 23, lwd =3)
}

```



For fig k

```
par(mfrow = c(2,5))

data_k <- read.csv("fig_one_k.csv", header=TRUE)
k1<-subset(data_k, tech == "ddseq")
boxplot(k1$seurat_cell_type_pred_score~ k1$sample_id, notch = TRUE, col = "red",
        xast = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "ddseq")

k2<-subset(data_k, tech == "mtscatacfacs")
boxplot(k2$seurat_cell_type_pred_score~ k2$sample_id, notch = TRUE, col = "blue",
        xast = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "mtscatacfacs")
```

```

k3<-subset(data_k, tech == "10xmultiome")
boxplot(k3$seurat_cell_type_pred_score~ k3$sample_id, notch = TRUE, col = "green",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "10xmultiome")

k4<-subset(data_k, tech == "10xv11")
boxplot(k4$seurat_cell_type_pred_score~ k4$sample_id, notch = TRUE, col = "yellow",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "10xv11")

k5<-subset(data_k, tech == "10xv11c")
boxplot(k5$seurat_cell_type_pred_score~ k5$sample_id, notch = TRUE, col = "violet",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "10xv11c")

k6<-subset(data_k, tech == "10xv2")
boxplot(k6$seurat_cell_type_pred_score~ k6$sample_id, notch = TRUE, col = "grey",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "10xv2")

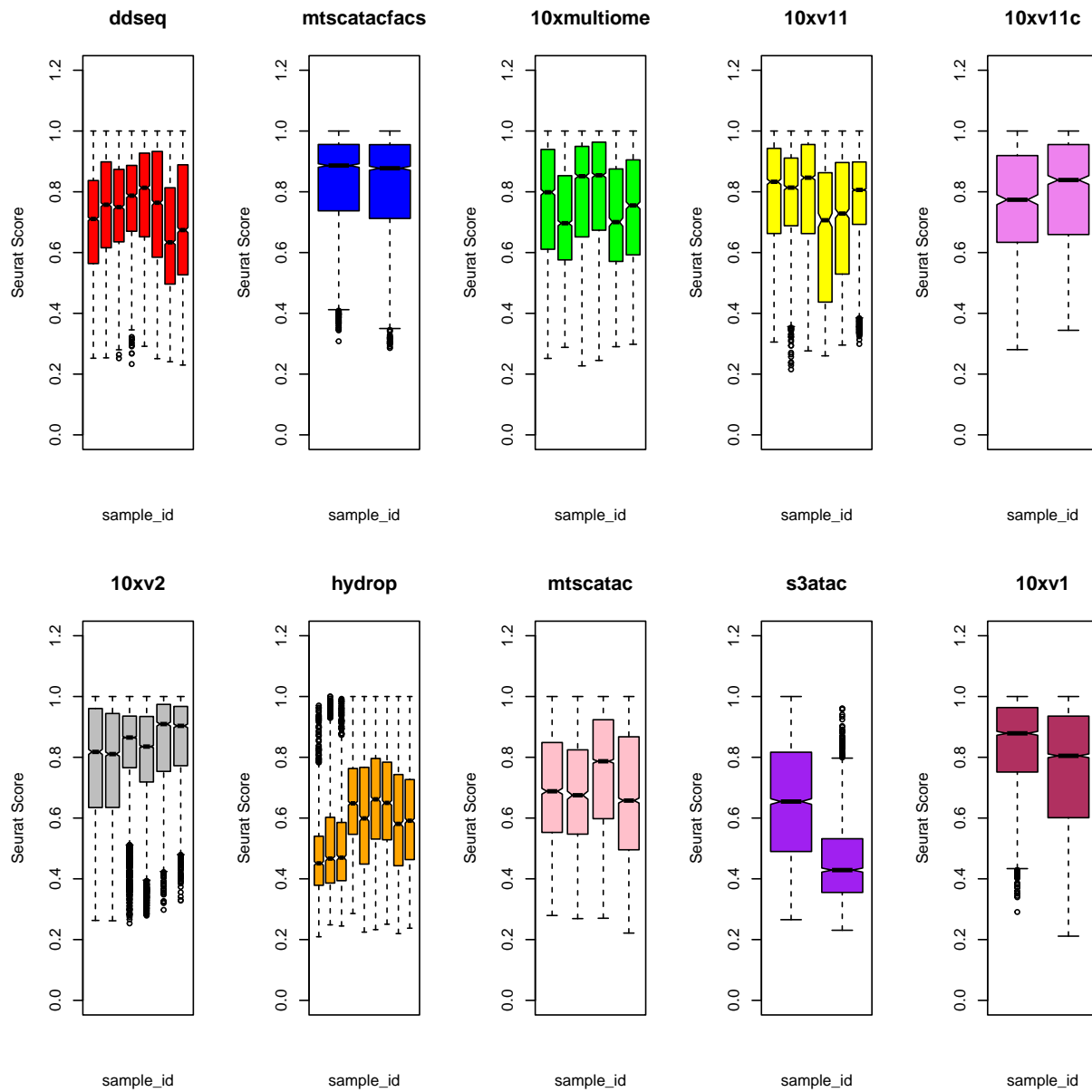
k7<-subset(data_k, tech == "hydrop")
boxplot(k7$seurat_cell_type_pred_score~ k7$sample_id, notch = TRUE, col = "orange",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "hydrop")

k8<-subset(data_k, tech == "mtscatac")
boxplot(k8$seurat_cell_type_pred_score~ k8$sample_id, notch = TRUE, col = "pink",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "mtscatac")

k9<-subset(data_k, tech == "s3atac")
boxplot(k9$seurat_cell_type_pred_score~ k9$sample_id, notch = TRUE, col = "purple",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "s3atac")

k10<-subset(data_k, tech == "10xv1")
boxplot(k10$seurat_cell_type_pred_score~ k10$sample_id, notch = TRUE, col = "maroon",
        xaxt = "n", ylim = c(0,1.2), xlab = "sample_id", ylab = "Seurat Score", main = "10xv1")

```



For figures g- j

```
par(mfrow = c(2,2))
data_g <- read.csv("fig_one_g.csv", header=TRUE)
technology3 <- unique(data_g$technology)
data_color2 <-c ("red", "blue", "green", "orange", "purple", "pink", "brown",
                "gray", "cyan", "magenta")
plot(0,0,
     type = "n",
     xlim = c(0, 25000),
     ylim = c(0, 400),
     xlab = "Median Unique Fragment",
```

```

ylab = "Median Freemuxlet",
main = "Fig 1g",
las = 1)

for (i in 1:10) {
  subset_data2 <- data_g [data_g$technology == technology3[i],]
  points(subset_data2$Median_unique_nr_frag , subset_data2$fmix_delta_donor_llk,
         col = data_color2[i],pch = 23, lwd =3)
}

data_h <- read.csv("fig_one_h.csv", header=TRUE)
data_color_h <-c("red", "blue", "green", "orange", "purple", "pink",
                 "brown", "gray", "cyan", "magenta")
technology_h <- unique(data_h$technology)

# fig_1h
plot(0,0,
     type = "n",
     xlim = c(2.5,4.0),
     ylim = c(0.4,0.9),
     xlab = "Median Unique Fragment",
     ylab = "Median Seurat Score",
     main = "Fig 1h",
     las = 1)

for (i in 1:10) {
  subset_data_h <- data_h [data_h$technology == technology_h[i],]
  points(subset_data_h$log_median_unique_nr_frag_in_regions,
         subset_data_h$seurat_score, col = data_color_h[i],pch = 23, lwd =3)
}

data_i <- read.csv("fig_one_i.csv", header=TRUE)
data_color_i <-c("red", "blue", "green", "orange", "purple", "pink",
                 "brown", "gray", "cyan", "magenta")
technology_i <- unique(data_i$technology)

# fig_1i
plot(0,0,
     type = "n",
     xlim = c(5,35),
     ylim = c(0.1,0.7),
     xlab = "Median TSS Enrichment",
     ylab = "Median FRIP",
     main = "Fig 1i",
     las = 1)

for (i in 1:10) {
  subset_data_i <- data_i [data_i$technology == technology_i[i],]
  points(subset_data_i$Median_tss_enrichment,subset_data_i$Median_frip,
         col = data_color_h[i], pch = 15,lwd =3)
}

```

```

data_j <- read.csv("fig_one_j.csv", header=TRUE)
data_color_j <-c("red", "blue", "green", "orange", "purple", "pink",
                 "brown", "gray", "cyan", "magenta")
technology_j <- unique(data_j$technology)

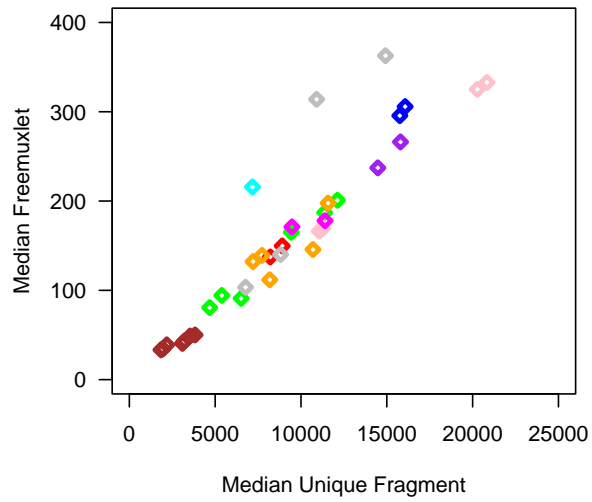
# fig_1j
plot(0,0,
     type = "n",
     xlim = c(12,25),
     ylim = c(1.0,4.0),
     xlab = "Median Strength",
     ylab = "Median Distance to TSS(kpp)",
     main = "Fig 1j",
     las = 1)
data_j$normal_values <- data_j$top2kdars_median_dar_tss_dist/1000

for (i in 1:10) {
  subset_data_j <- data_j [data_j$technology == technology_j[i],]
  points(subset_data_j$normal_values,subset_data_j$top2kdars_median_dar_logfc,
         col = data_color_h[i], pch = 19, lwd =2)
}

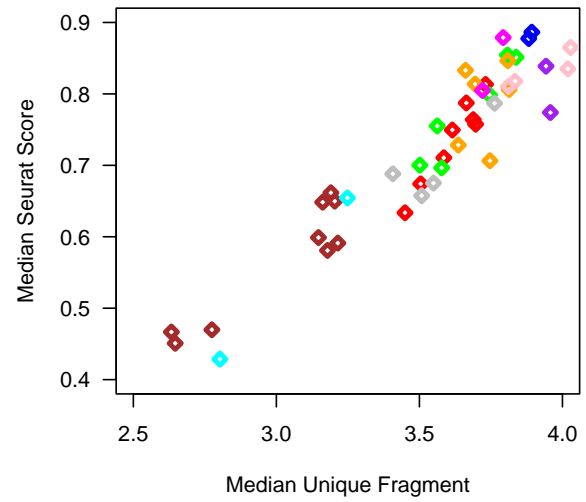
```



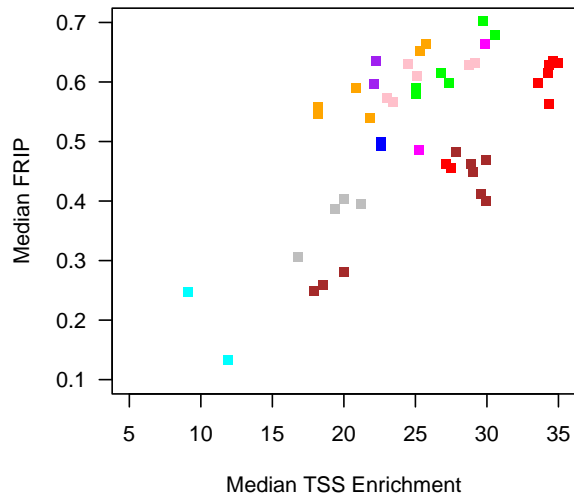
**Fig 1g**



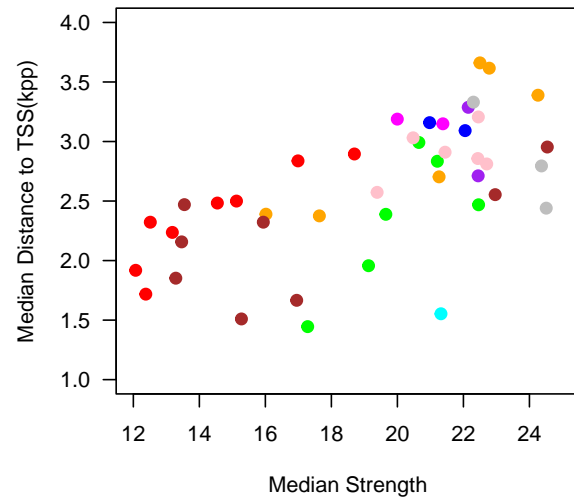
**Fig 1h**



**Fig 1i**



**Fig 1j**



Thank you