

# FIG\_3

Compiled: May 15, 2022

## scAllele: a versatile tool for the detection and analysis of variants in single-cell RNA-sequencing.

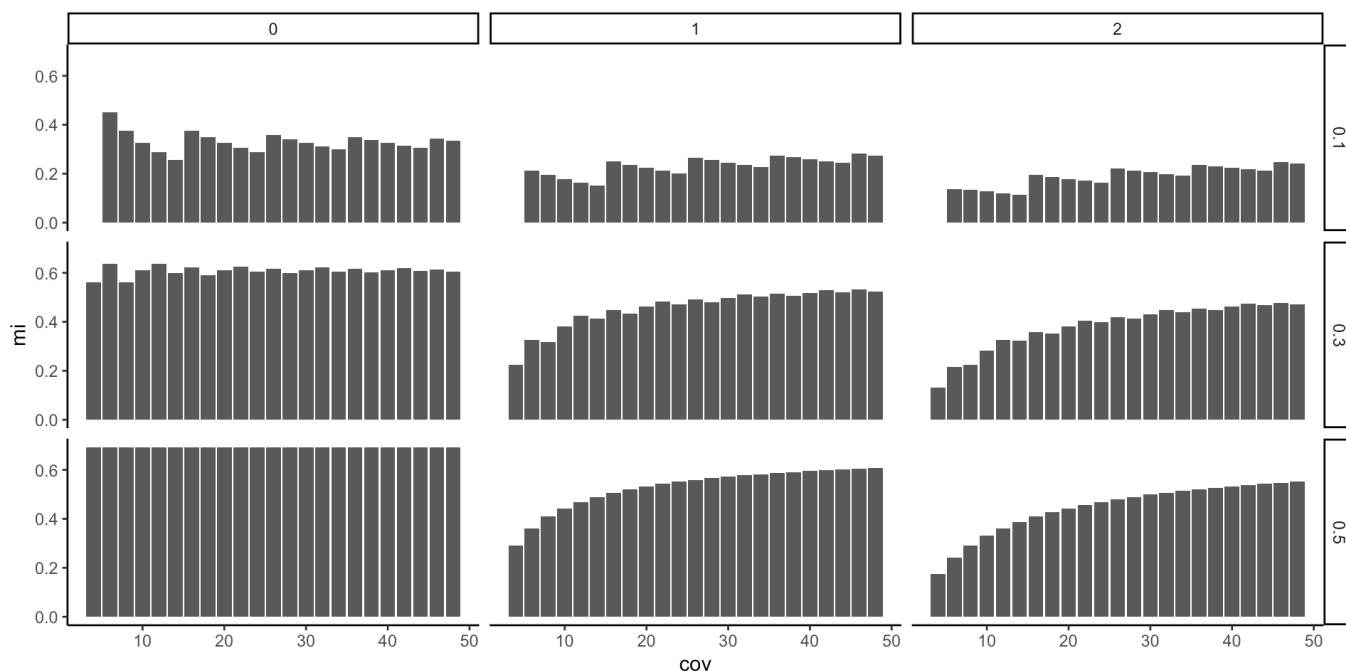
```
library(dplyr)
library(ggplot2)
library(reshape2)
library(RColorBrewer)
```

### Coverage test

```
df <- read.csv("mutual_information_cov_test.txt", sep = " ", header = F)
names(df) <- c("cov", "AB", "error", "mi")
head(df)
```

cov	AB	error	mi
4	0.1	0	0.0000000
6	0.1	0	0.4505612
8	0.1	0	0.3767702
10	0.1	0	0.3250830
12	0.1	0	0.2868360
14	0.1	0	0.2573186

```
ggplot(df, aes(x = cov, y = mi)) + geom_bar(stat = "identity") + facet_grid(as.factor(AB) ~ as.factor(error)) +
  theme_classic()
```



### Fig. 3a,b - Mutual information vs common reads

```
df3 <- read.csv("mutual_information.GM12878_GSE77288.pairs.txt", sep = " ", header = F)
names(df3) <- c("var1", "var2", "mi", "cov", "label")

df3$label[df3$label == 1] <- 0

df3$cov_bin <- cut(df3$cov, breaks = c(0, 5, 10, 15, 20, Inf), right = F)

PAL <- brewer.pal(7, "Set2")

df4 <- df3[df3$label != 0, ]
df4c <- df4 %>%
  group_by(cov_bin, label) %>%
  summarise(n = n(), x = mean(mi), y = mean(cov)/3)

g1 <- ggplot(df4, aes(x = mi)) + geom_density(aes(color = cov_bin), size = 1, adjust = 1) + geom_text(data = df4c,
  aes(color = cov_bin, label = n, x = x, y = y)) + scale_color_manual(values = PAL[2:7]) + theme_classic()

df4 <- df3[df3$label == 0, ]
df4c <- df4 %>%
  group_by(cov_bin, label) %>%
  summarise(n = n(), x = mean(mi), y = mean(cov)/3)

g2 <- ggplot(df4, aes(x = mi)) + geom_density(aes(color = cov_bin), size = 1, adjust = 1) + geom_text(data = df4c,
  aes(color = cov_bin, label = n, x = x, y = y)) + scale_color_manual(values = PAL[2:7]) + theme_classic()

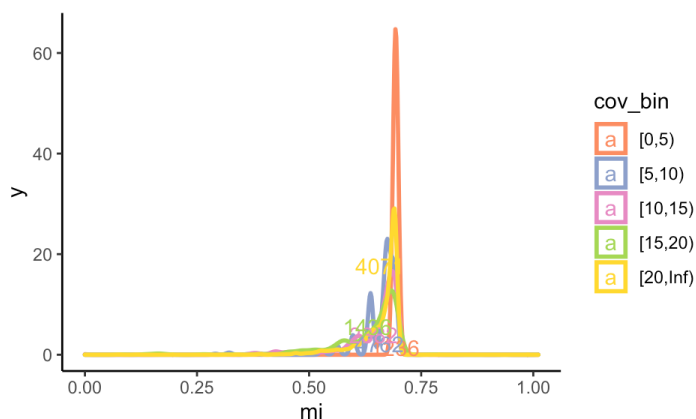
pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/3a.pdf")
g1
dev.off()
```

```
## quartz_off_screen
##                2
```

```
pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/3b.pdf")
g2
dev.off()
```

```
## quartz_off_screen
##                2
```

```
g1
```



```
g2
```

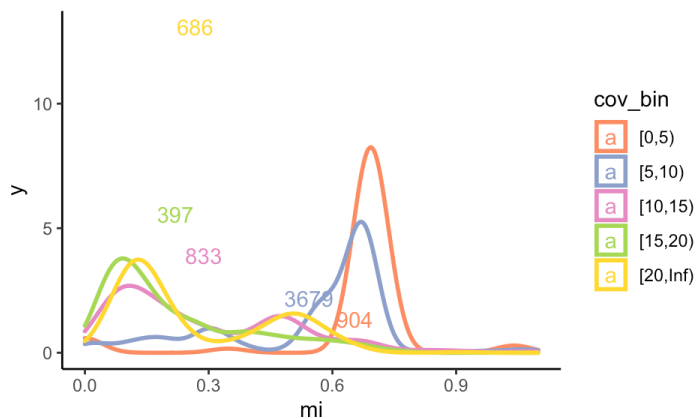


Fig. 3c - Detremining MI cutoffs

```
FPv = 0.9
FNv = 0.05
cov_cutoff = 10

y <- df3[df3$label == 0 & df3$cov >= cov_cutoff, ]$mi
y <- sort(y)
cumy <- ecdf(y)(y)
for (i in 1:length(cumy)) {
  if (y[i] > 0.515) {
    print(cumy[i])
    FP <- y[i]
    break
  }
}
```

```
## [1] 0.8653445
```

```
y <- df3[df3$label == 2 & df3$cov >= cov_cutoff, ]$mi
y <- sort(y)
cumy <- ecdf(y)(y)
for (i in 1:length(cumy)) {
  if (cumy[i] > FNv) {
    print(y[i])
    FN <- y[i]
    break
  }
}
```

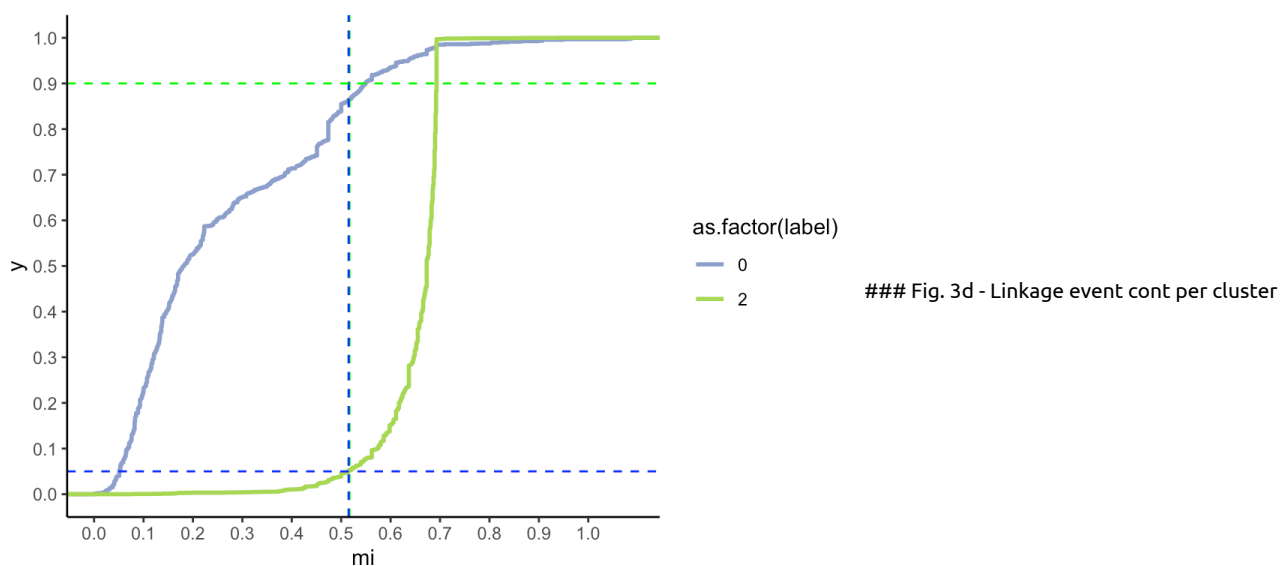
```
## [1] 0.515
```

```
g1 <- ggplot(df3[df3$cov >= cov_cutoff & df3$label != 1, ], aes(x = mi)) + stat_ecdf(aes(color = as.factor(label)),
  size = 1, adjust = 1) + geom_vline(xintercept = FP, linetype = "dashed", color = "green") +
  geom_vline(xintercept = FN, linetype = "dashed", color = "blue") + geom_hline(yintercept = FPv,
  linetype = "dashed", color = "green") + geom_hline(yintercept = FNv, linetype = "dashed", color = "blue") +
  scale_x_continuous(breaks = seq(0, 1, 0.1)) + scale_y_continuous(breaks = seq(0, 1, 0.1)) +
  scale_color_manual(values = c(PAL[3], PAL[5])) + theme_classic()

pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/3c.pdf")
g1
dev.off()
```

```
## quartz_off_screen
## 2
```

g1



```
df3 <- read.csv("mutual_information.GM12878_event_counts.txt", sep = " ", header = F)
names(df3) <- c("cell", "far", "close", "non_testable", "testable")
head(df3)
```

cell	far	close	non_testable	testable
1	5127	0	107	43
5	15060	0	404	290
10	27683	0	1052	561
15	43827	0	2090	1162
20	63927	0	3642	2182
25	88374	0	5696	3740

```
df3 <- melt(df3, id.vars = c("cell"))

df4 <- df3[!df3$variable %in% c("far"), ]
df4 <- df3[df3$variable %in% c("non_testable", "testable"), ]

g1 <- ggplot(df4, aes(x = as.factor(cell), y = value, label = value)) + geom_bar(aes(fill = variable),
  color = "white", position = "dodge", stat = "identity", width = 0.8) + geom_text(aes(y = value,
  color = variable), position = "stack") + scale_fill_brewer(palette = "Set2") + scale_color_brewer(palette = "
Set2") +
  scale_y_log10() + theme_minimal()

pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/3d.pdf")
g1
dev.off()
```

```
## quartz_off_screen
## 2
```

g1

