

# InbreedingCoeff - Multi-Threaded GenotypeGVCFs

Camden Lopez

3/27/2018

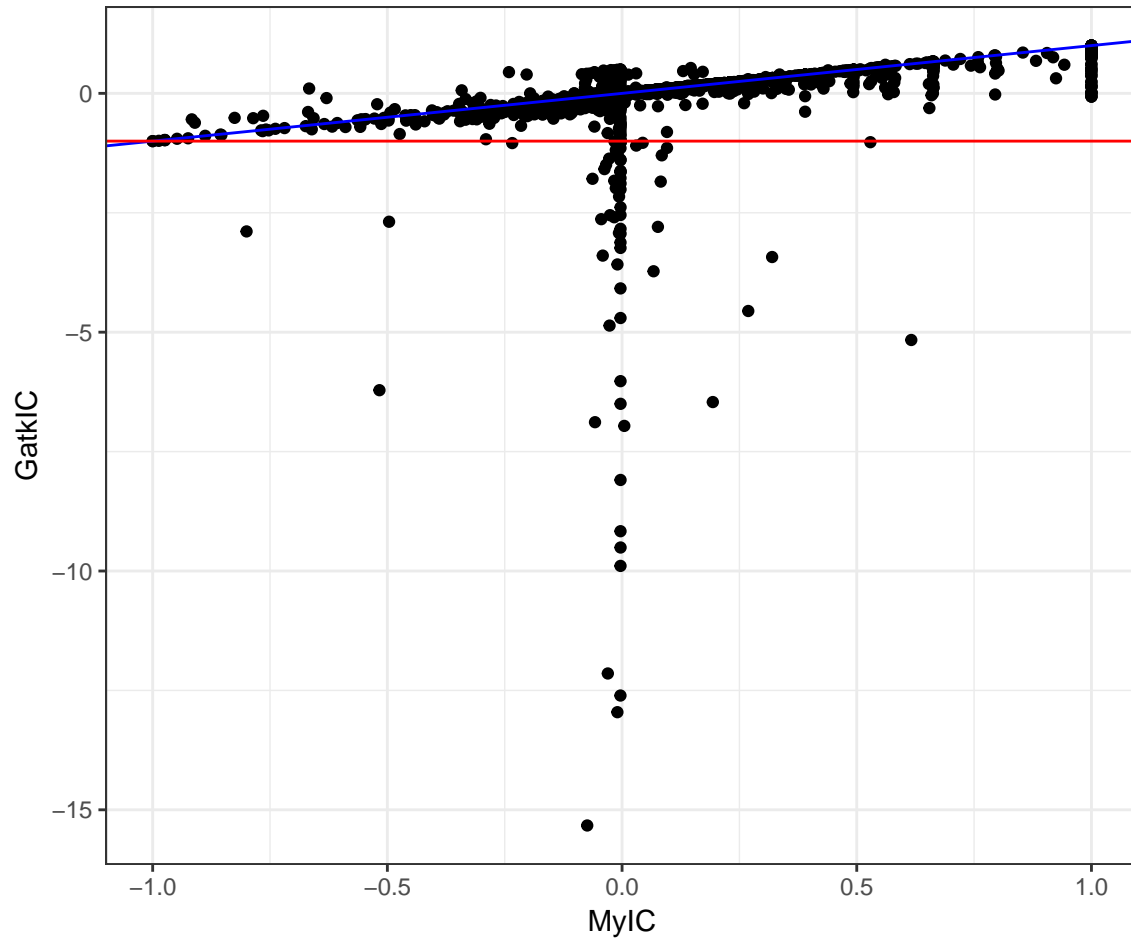
```
library(vcfR)
library(ggplot2)
vcf <- read.vcfR("data/exome.bwa.gatk.multithread.vcf.gz",
                verbose = FALSE)
inbreed <- as.double(extract.info(vcf, "InbreedingCoeff"))
check <- sample(1:length(inbreed), 10000)
gt <- vcf@gt[check,]
GatkIC <- inbreed[check]
sum(GatkIC < -1, na.rm = TRUE)

## [1] 68

gt <- t(apply(gt, 1, function (x) sapply(strsplit(x, ":"),
                                       function (x) x[1])))
gt[1:10,1:10]

##          FORMAT CB-1  CB-10 CB-104 CB-105 CB-106 CB-107 CB-109 CB-11 CB-111
## [1,] "GT"      "0/1"  "0/1"  "0/0"  "0/0"  "0/0"  ". / ." "0/0"  "0/1"  "0/0"
## [2,] "GT"      "0/0"  "0/0"  "0/1"  "0/0"  "0/1"  ". / ." "0/0"  "0/0"  "0/0"
## [3,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"
## [4,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  ". / ." "0/0"  "0/0"  "0/0"
## [5,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  ". / ." "0/0"  "0/0"  "0/0"
## [6,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  ". / ." "0/0"  "0/0"  "0/0"
## [7,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  "0/0"
## [8,] "GT"      ". / ." ". / ." "0/0" ". / ." ". / ." ". / ." ". / ." "0/0"
## [9,] "GT"      "0/0"  "0/0"  "0/0"  "0/0"  "0/0"  ". / ." "0/0"  "0/0"  "0/0"
## [10,] "GT"     "0/0"  "0/0"  "0/0"  "0/0"  "0/1"  ". / ." "0/0"  "0/1"  "0/0"

gt <- gt[, -1]
MyIC <- apply(gt, 1, function (x) {
  ref <- sum(x == "0/0")
  het <- sum(x == "0/1")
  alt <- sum(x == "1/1")
  p <- (2 * ref + het) / (2 * (ref + het + alt))
  q <- 1 - p
  exp <- 2 * p * q * (ref + het + alt)
  1 - het / exp
})
ggplot(mapping = aes(x = MyIC, y = GatkIC)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "blue") +
  geom_hline(yintercept = -1, color = "red") +
  theme_bw()
```



```
smallest <- which(GatkIC %in% sort(GatkIC)[1:3])
GatkIC[smallest]
```

```
## [1] -12.6072 -12.9559 -15.3280
```

```
MyIC[smallest]
```

```
## [1] -0.003412969 -0.009966777 -0.074204947
```

```
apply(gt[smallest,], 1, table)
```

```
## [[1]]
##
## ./ 0/0 0/1 0/2
## 8 146 1 2
##
## [[2]]
##
## ./ 0/0 0/1
## 5 149 3
##
## [[3]]
##
## ./ 0/1 1/1
## 5 21 131
```