

Knockoff report for example dataset: #22

In this report, we look at the results from knockoff procedure for our example dataset, Dataset22. This is an example dataset, selected from the results we had on May 26, 2020. In this dataset, both IBD based methods (dCorN and dCorIG) localize the peak in the risk region. In addition, none of the non-IBD based methods are able to localize the peak in the risk region. The profiles are demonstrated in [non_IBD_methods.PDF], [IBD_methods.PDF] for non-IBD and IBD-based methods, respectively.

1. We first load all the data that is necessary:

- (i) dCorN.RData: holding the Naive dCor profile along the genome.
- (ii) dCorIG.RData: holding the dCorIG profile along the genome.
- (iii) dCorNIGko.csv: Results from JG's scripts (fitknock.R and dCorKnockoff.R). This csv file records a dataframe with 2 columns (dCorN and dCorIG) and 1000 rows. Each row of this dataframe records the maximum value of the dCor (either N or IG) along the genome when we use knockoff samples.

```
load('dCorN.RData')
load('dCorIG.RData')
ko_csv = read.csv(file = 'dCorNIGko.csv', header = TRUE)
dim(ko_csv)
```

```
## [1] 1000    2
```

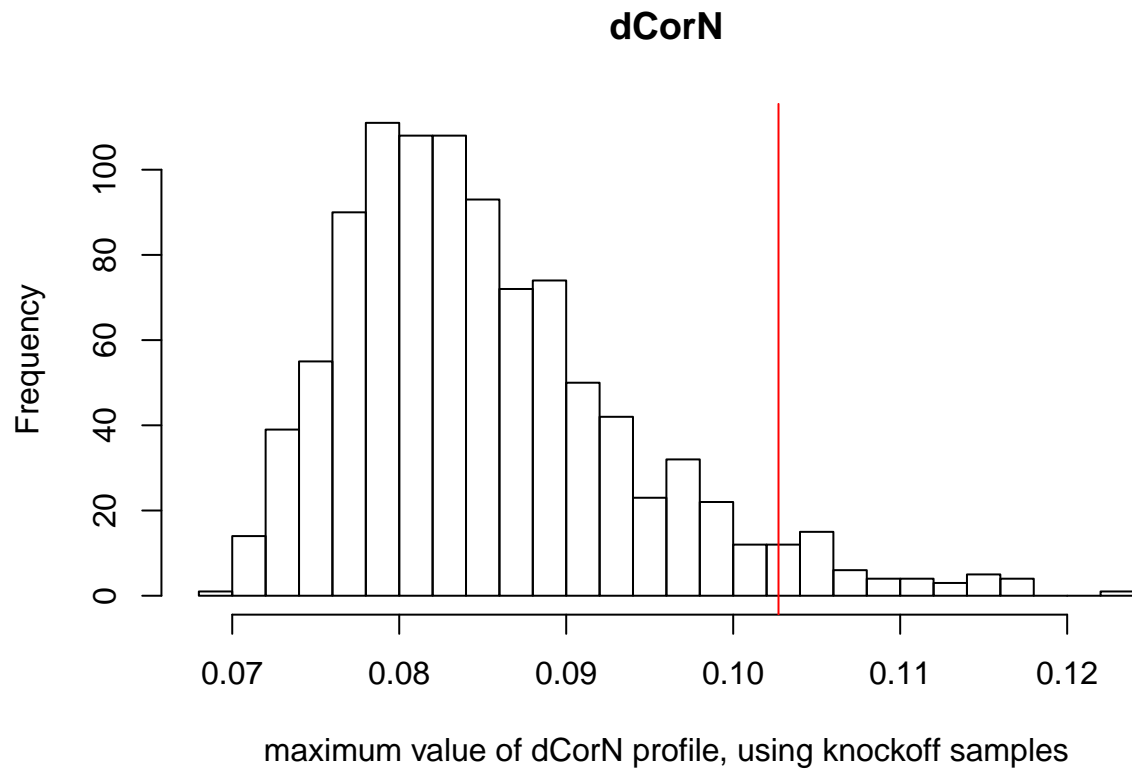
```
head(ko_csv)
```

```
##           dCorN      dCorIG
## 1 0.07901653 0.10761417
## 2 0.08348457 0.09718819
## 3 0.08865651 0.11496311
## 4 0.07916223 0.10766378
## 5 0.09400178 0.09646201
## 6 0.09193200 0.09780205
```

2. Histogram

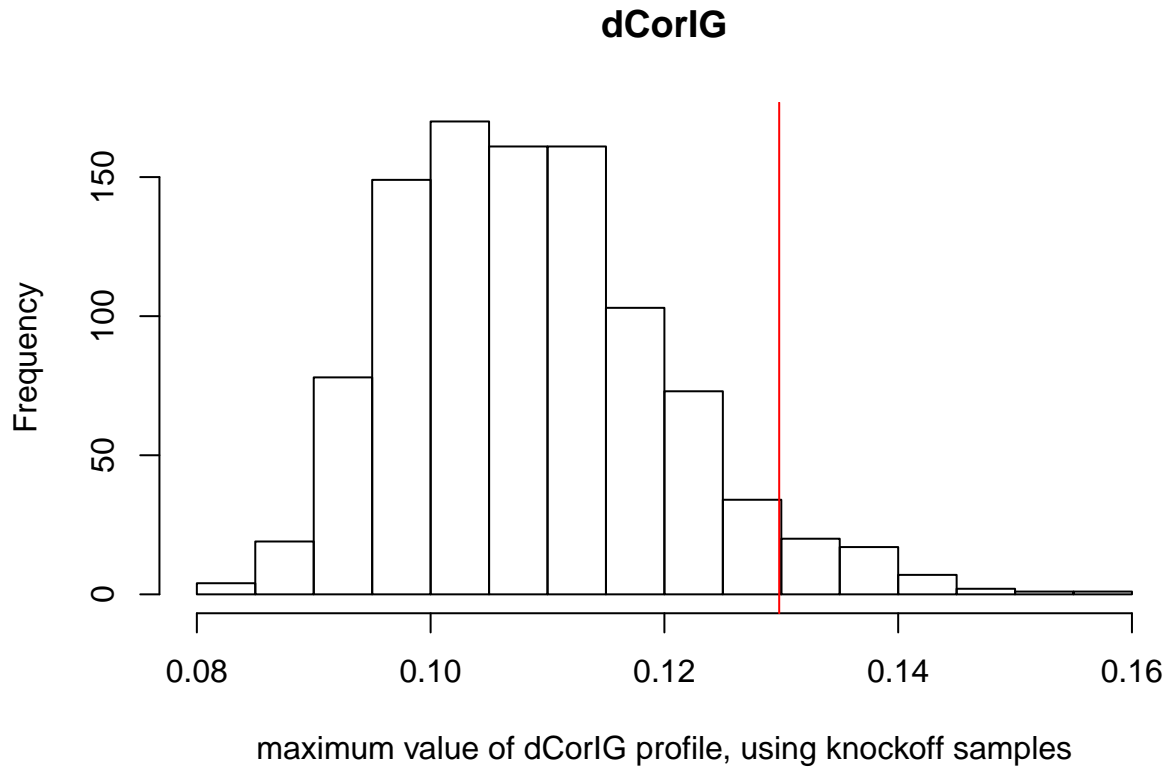
(i) dCorN: The red vertical line represents the 95-th percentile of the distribution.

```
hist( x = ko_csv$dCorN, main = 'dCorN', xlab = 'maximum value of dCorN profile, using knockoff samples', col = 'red')
abline(v = quantile(ko_csv$dCorN, probs = 0.95), col = "red")
```



(ii) dCorIG: The red vertical line represents the 95-th percentile of the distribution.

```
hist( x = ko_csv$dCorIG, main = 'dCorIG', xlab = 'maximum value of dCorIG profile, using knockoff samples',  
      abline(v = quantile(ko_csv$dCorIG, probs = 0.95), col = "red"))
```



3. Calculate pvalue

The maximum of dCorN is 0.127885 and the maximum of dCorIG is 0.1852596.

```
mean( ko_csv[, "dCorN" ] >= max(dCorN) )
```

```
## [1] 0
```

```
mean( ko_csv[, "dCorIG" ] >= max(dCorIG_Profile) )
```

```
## [1] 0
```