BIOS 7659 Homework 7

Tim Vigers

15 November 2020

# 1. DNA Methylation QC and Normalization (Illumina 450K)

Load the data:

baseDir = "C:/Users/tim/Dropbox/Documents/School/Statistical Genomics/Homework Files/HW7/idats"  
targets = read.metharray.sheet(baseDir)

## [1] "C:/Users/tim/Dropbox/Documents/School/Statistical Genomics/Homework Files/HW7/idats/SampleSheet.csv"

rgSet = read.metharray.exp(targets = targets)  
annotation(rgSet)

## array annotation   
## "IlluminaHumanMethylation450k" "ilmn12.hg19"

## a) Table 1

df = as.data.frame(pData(rgSet))  
df = df[df$Status == "cancer",]  
t1 = tableby(~ patient.age\_at\_initial\_pathologic\_diagnosis +  
 patient.height + patient.weight + Sex + patient.race,  
 data = df)  
summary(t1,labelTranslations =   
 list(patient.age\_at\_initial\_pathologic\_diagnosis =   
 "Age at Diagnosis",  
 patient.height = "Height",patient.weight = "Weight",  
 patient.race = "Race"))

|  |  |
| --- | --- |
|  | Overall (N=3) |
| **Age at Diagnosis** |  |
| Mean (SD) | 76.667 (7.234) |
| Range | 72.000 - 85.000 |
| **Height** |  |
| Mean (SD) | 166.300 (15.934) |
| Range | 151.000 - 182.800 |
| **Weight** |  |
| Mean (SD) | 61.900 (8.080) |
| Range | 52.600 - 67.200 |
| **Sex** |  |
| FEMALE | 2 (66.7%) |
| MALE | 1 (33.3%) |
| **Race** |  |
| BLACK OR AFRICAN AMERICAN | 1 (33.3%) |
| WHITE | 2 (66.7%) |

There are three unique subjects in this dataset, each with two samples (one primary tumor sample and one from normal solid tissue).

## b) Type I and II probes

getManifest(rgSet)

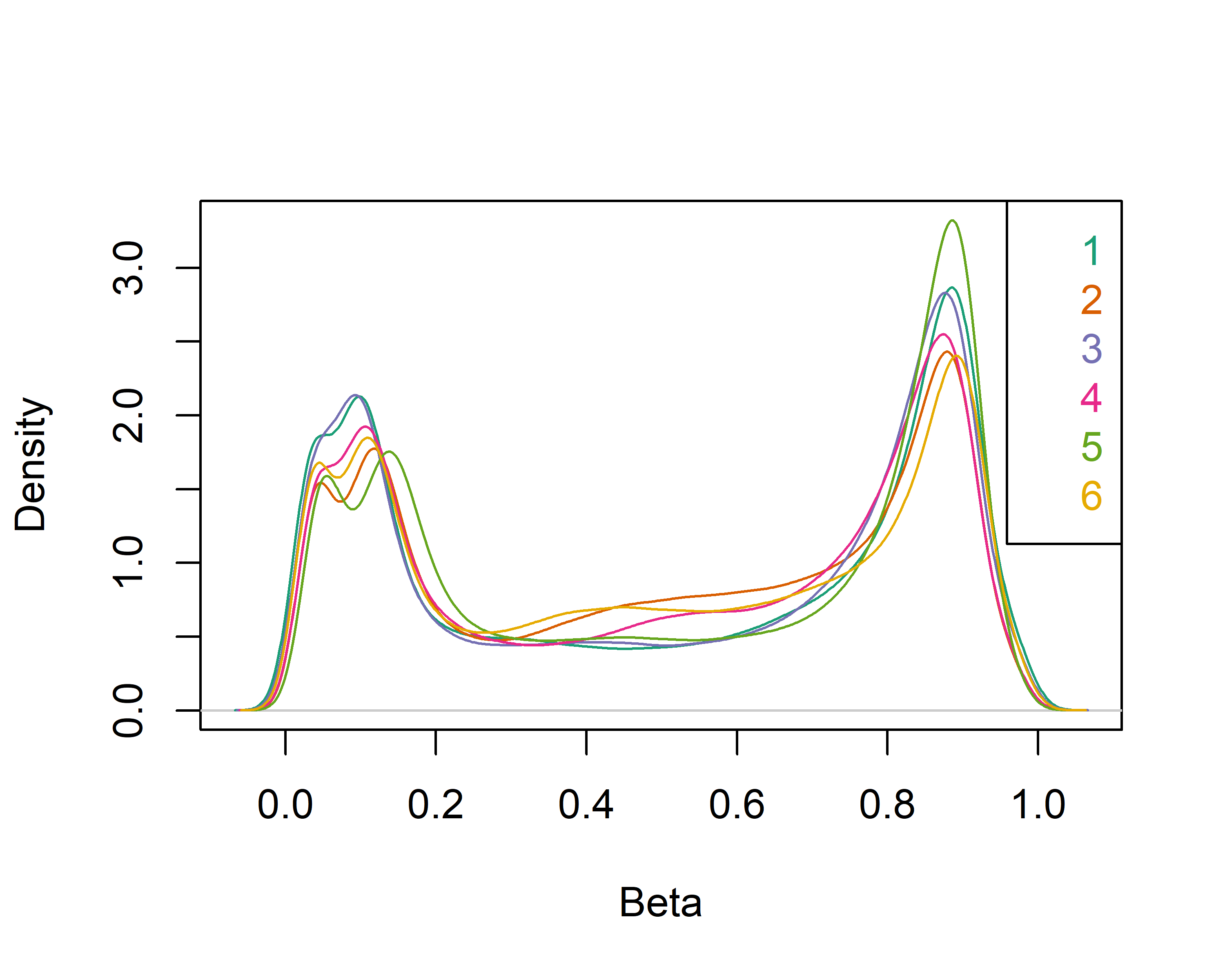
## IlluminaMethylationManifest object  
## Annotation  
## array: IlluminaHumanMethylation450k  
## Number of type I probes: 135476   
## Number of type II probes: 350036   
## Number of control probes: 850   
## Number of SNP type I probes: 25   
## Number of SNP type II probes: 40

There are 135,476 type I probes and 350,036 type II probes. Type I probes have two different sequences per CpG site, one for methylated and one for unmethylated CpGs. Type II probes use a two-color channel, which allows each probe to measure both methylated and unmethylated CpGs. As a result, type II probes take up half the physical space of type I probes. However, they have a lower dynamic range than type I probes, and are also more biased and less reproducible.

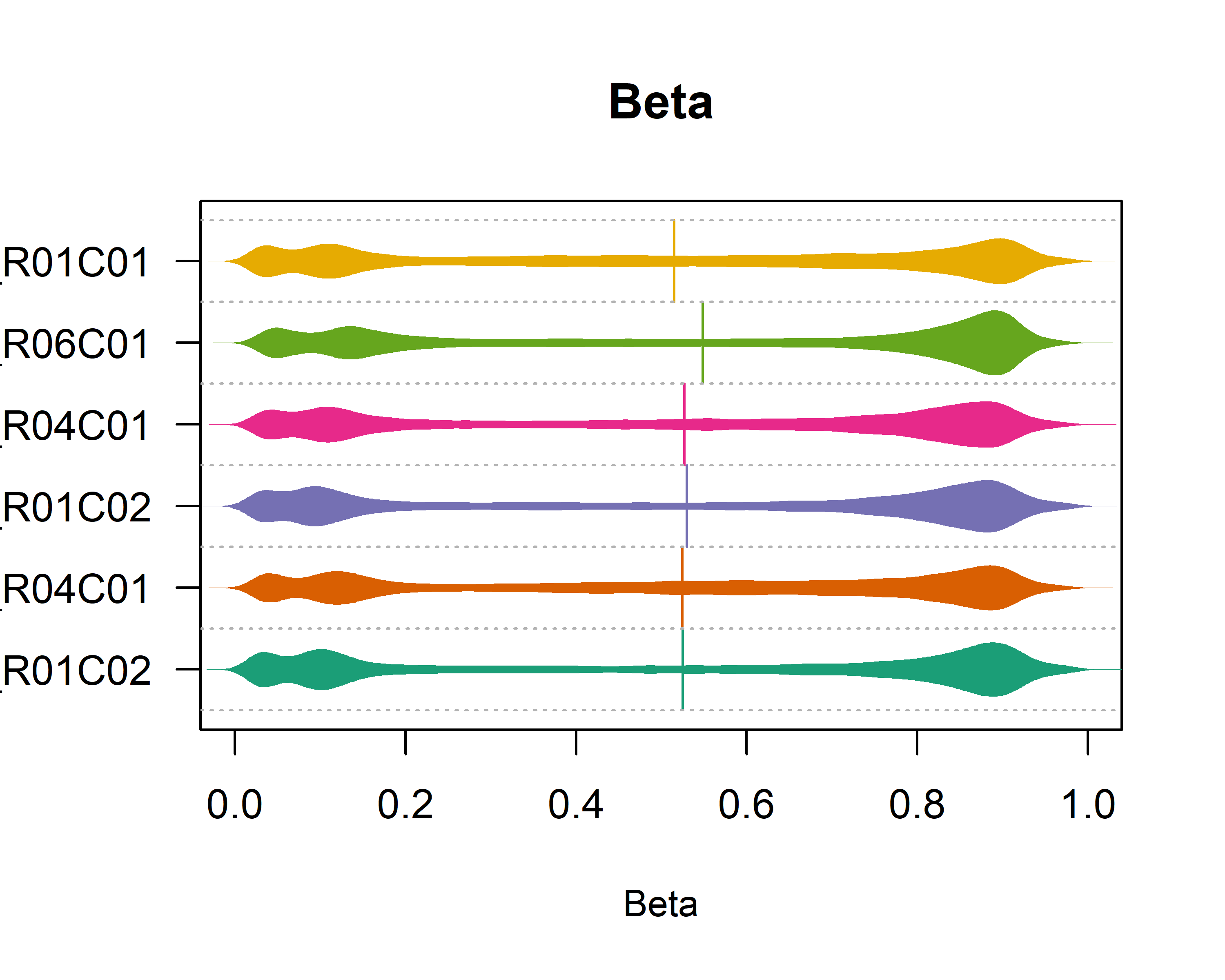
## c) QC Plots

### By ID

id = pData(rgSet)$id  
densityPlot(rgSet,sampGroups = id)

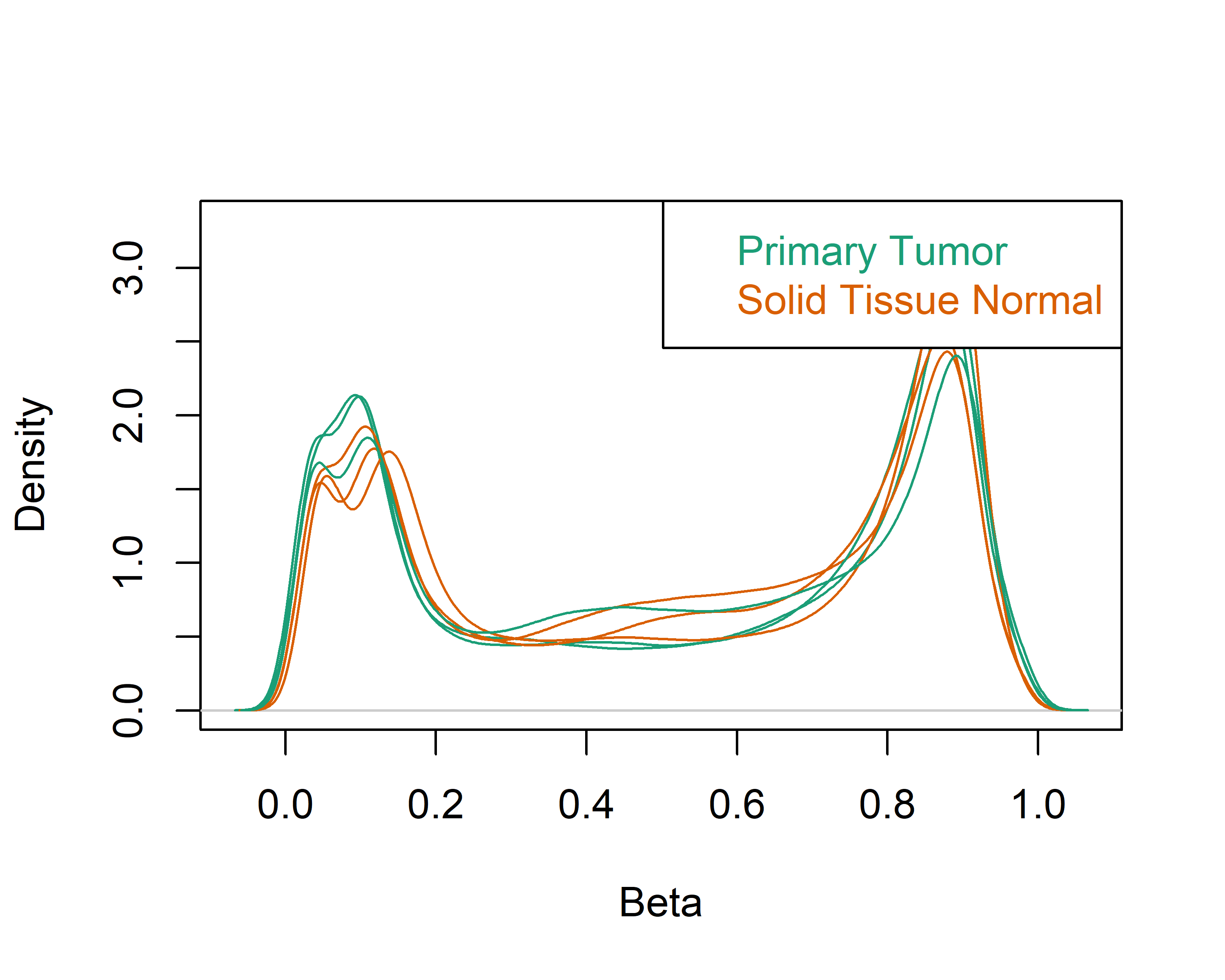


densityBeanPlot(rgSet,sampGroups = id)

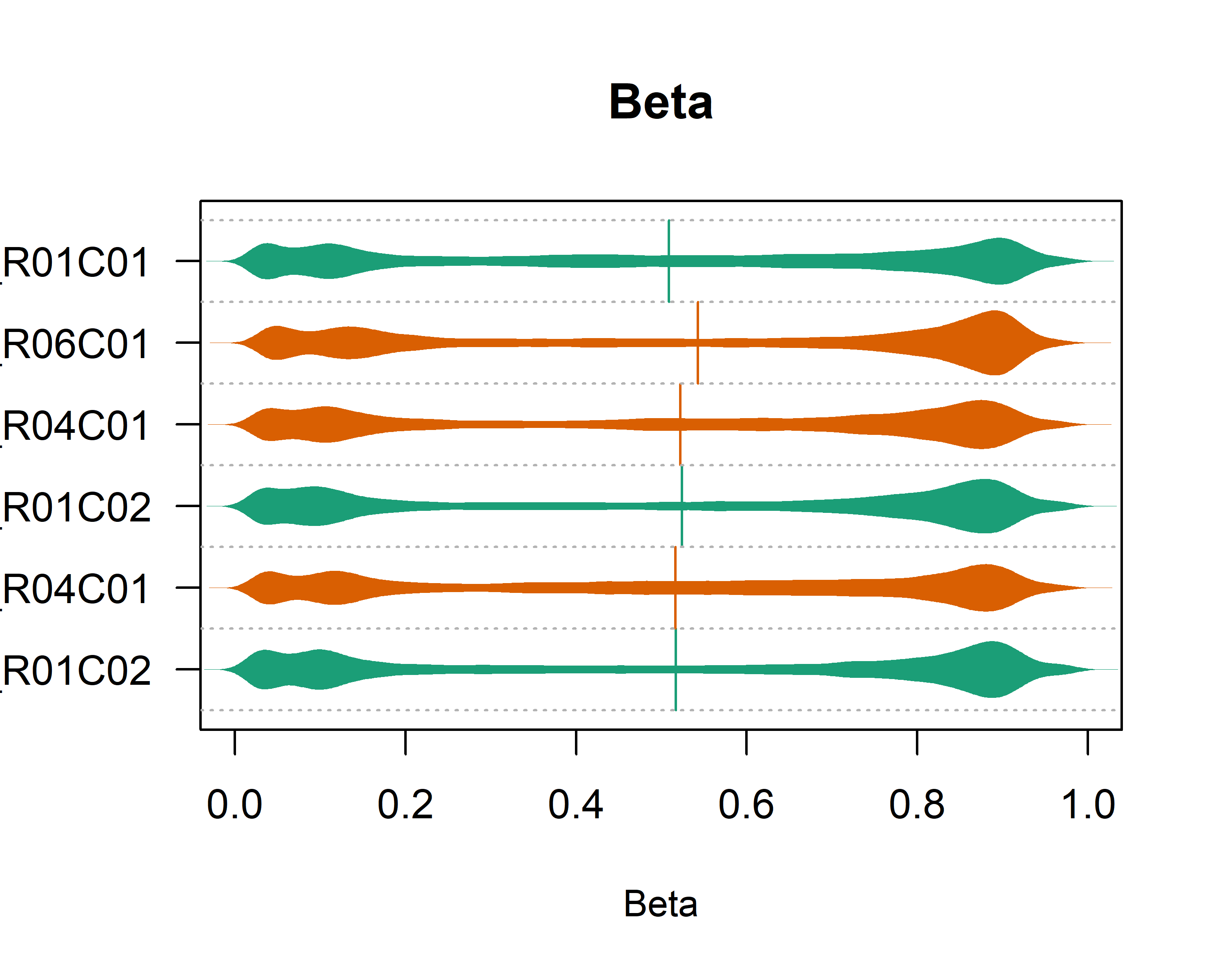


### By sample type

stype = pData(rgSet)$sample\_type  
densityPlot(rgSet,sampGroups = stype)

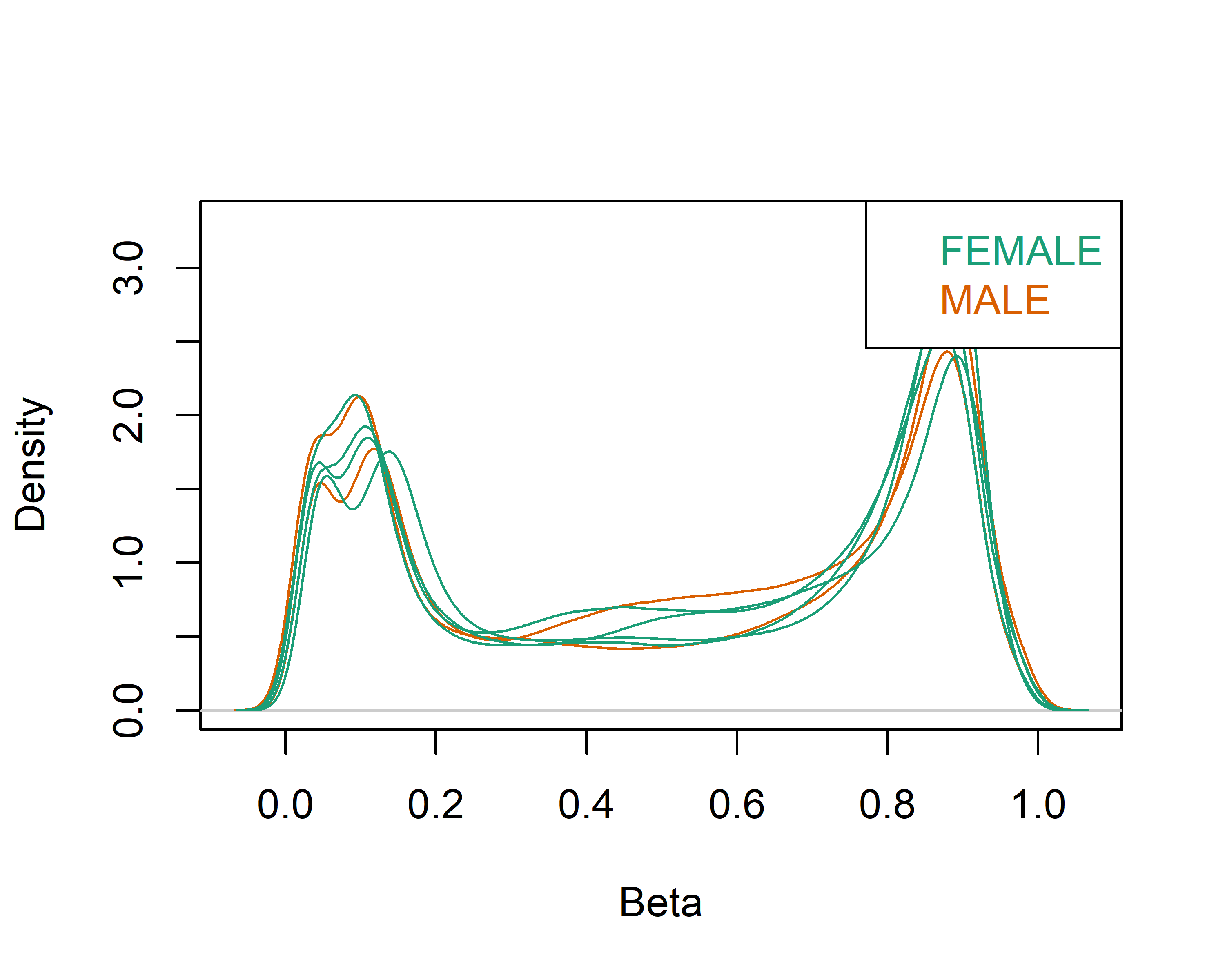


densityBeanPlot(rgSet,sampGroups = stype)

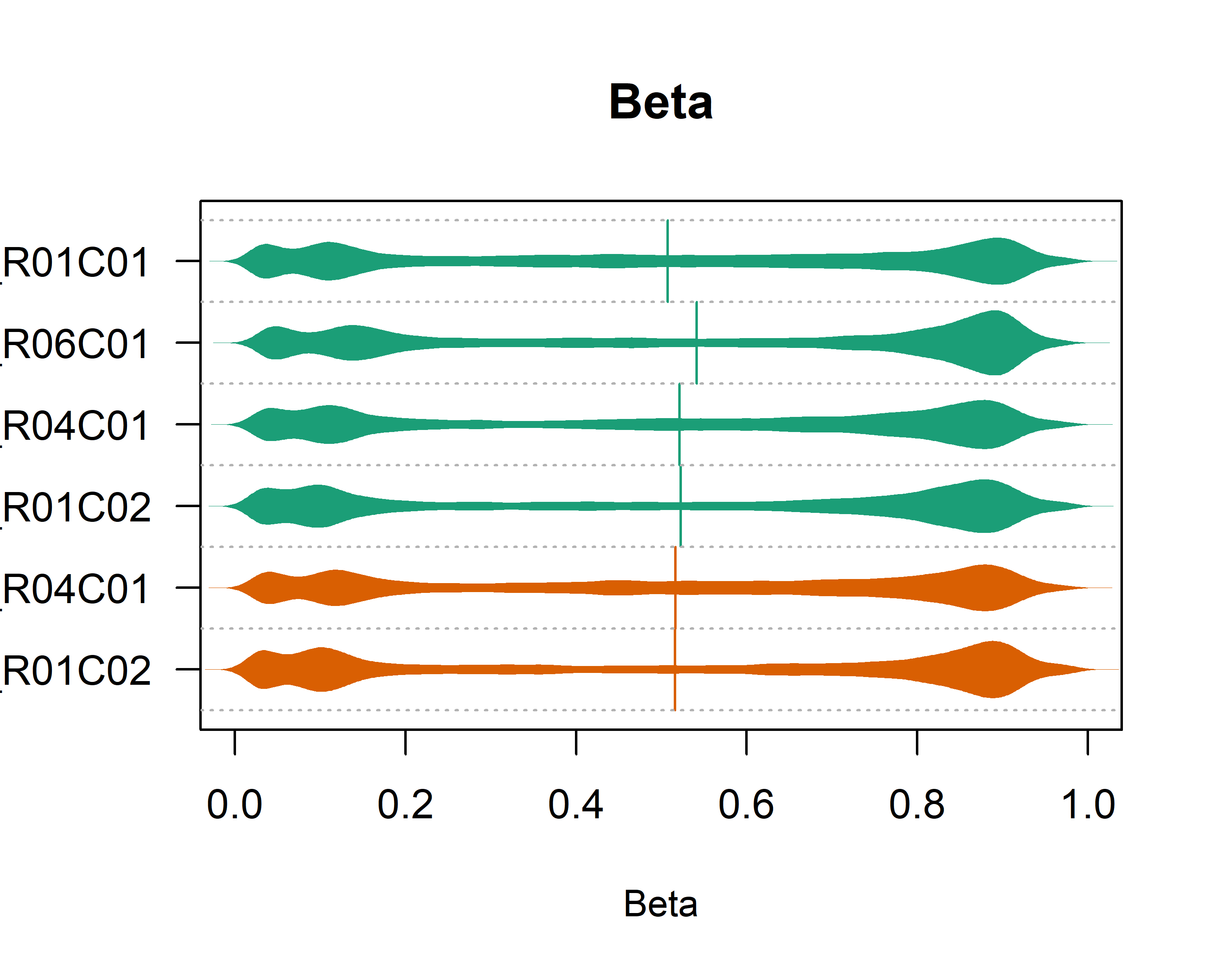


### By sex

sex = pData(rgSet)$Sex  
densityPlot(rgSet,sampGroups = sex)

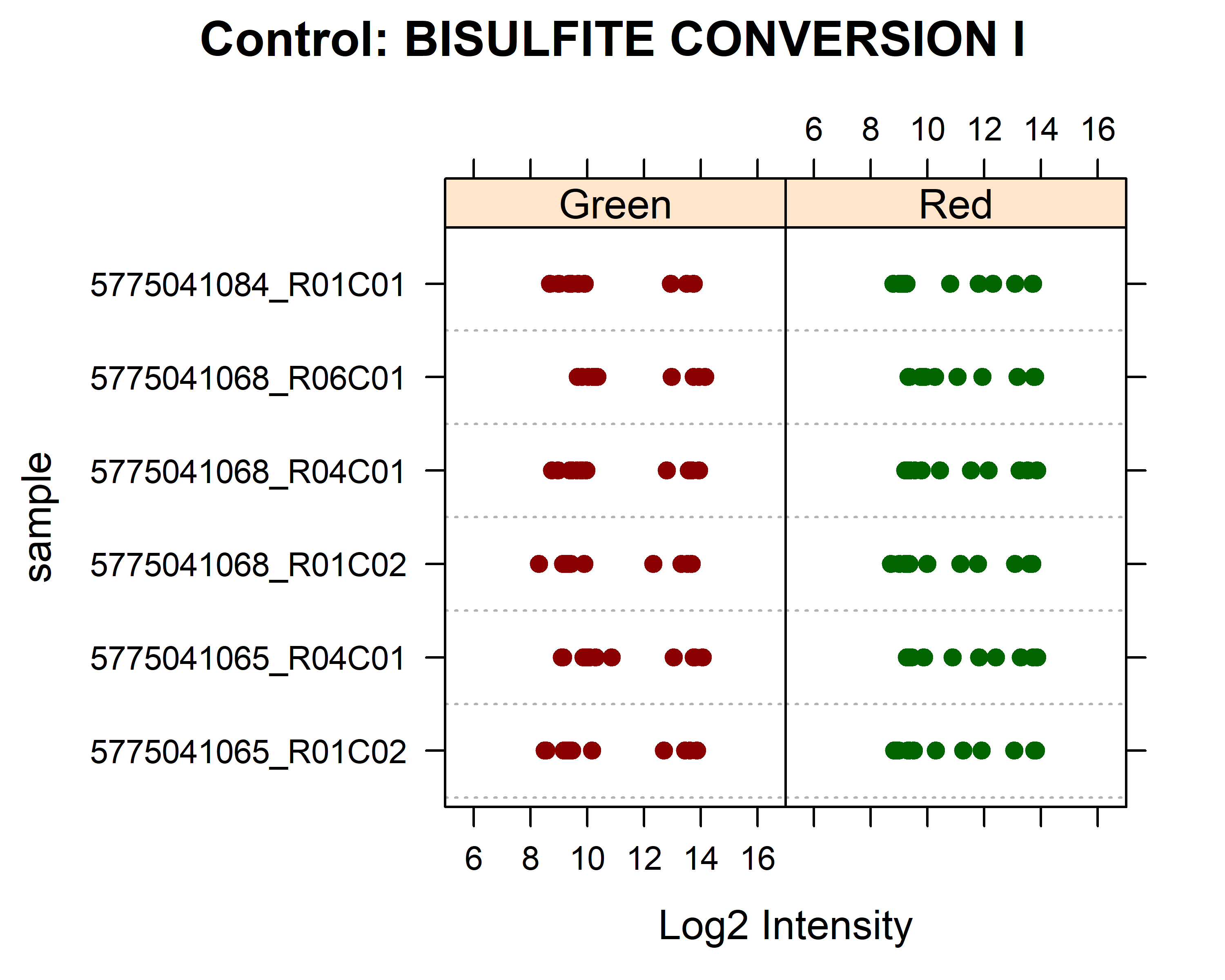


densityBeanPlot(rgSet,sampGroups = sex)

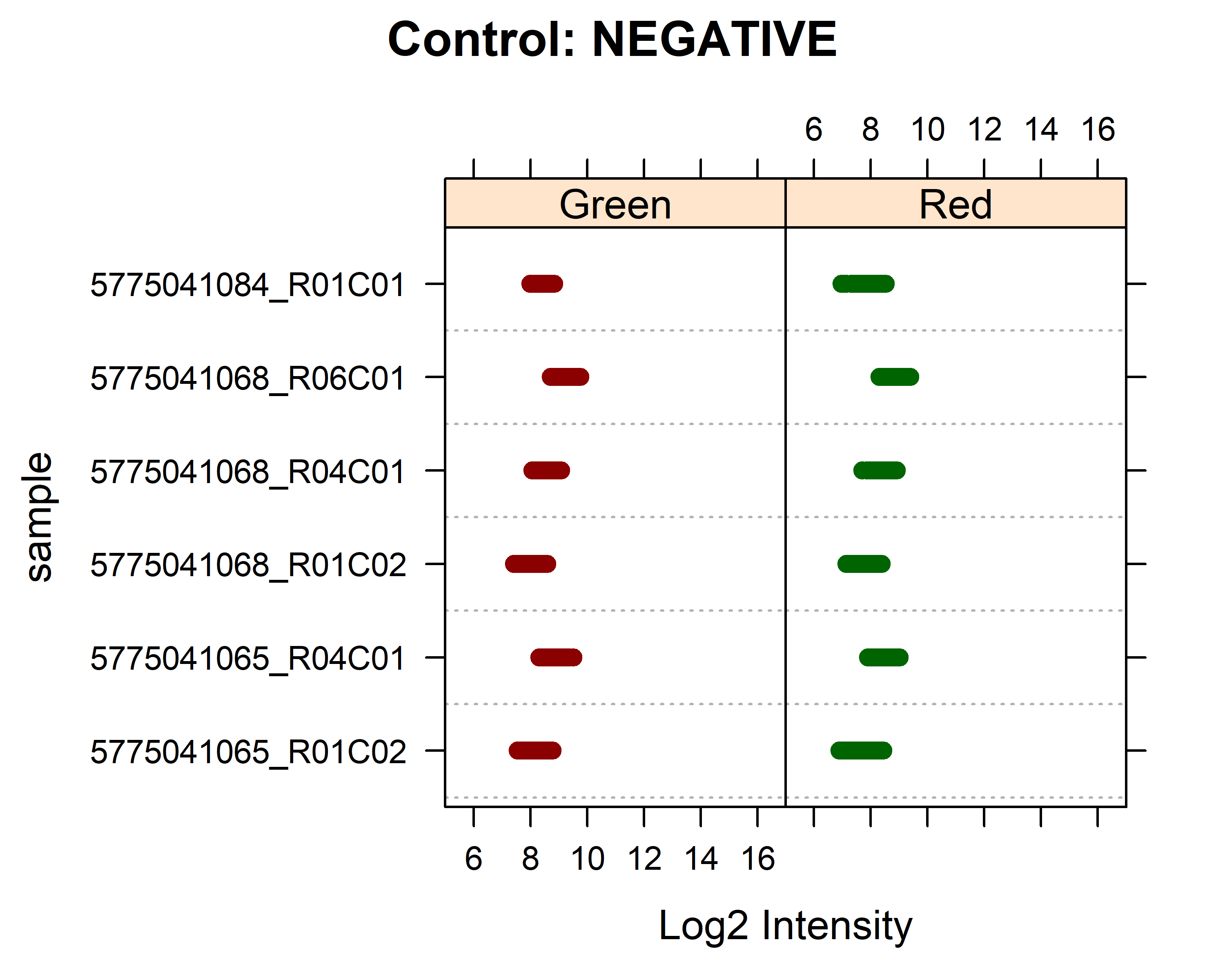


## d) Control probes

controlStripPlot(rgSet,controls = "BISULFITE CONVERSION I")



controlStripPlot(rgSet,controls = "NEGATIVE")



## e) Detection p values

# Count p values >= 0.05 per sample  
detect = detectionP(rgSet)  
colSums(detect >= 0.05)

## 5775041065\_R01C02 5775041065\_R04C01 5775041068\_R01C02 5775041068\_R04C01   
## 133 567 102 493   
## 5775041068\_R06C01 5775041084\_R01C01   
## 681 554

# Row means  
rmeans = rowMeans(detect)

Sample 5775041068\_R06C01 has the most detection p values with 681 (0.14%). Out of the 485512 probes, 853 have a mean p value .