

# Generating realistic data with known truth using the `jointseg` package

M. Pierre-Jean, G. Rigaiil, P. Neuvial

September 9, 2014

## Abstract

This vignette illustrates how the `jointseg` package may be used to generate a variety of copy-number profiles from the same biological “truth”. Such profiles have been used to compare the performance of segmentation methods [1].

```
library(jointseg)
```

## Contents

1	Setup	1
2	Affymetrix data	2
3	Illumina data	6
A	Session information	6

## 1 Setup

The parameters are defined as follows:

```
n <- 1e4                ## signal length
bcp <- c(2334, 6121)    ## breakpoint positions
regions <- c("(1,1)", "(1,2)", "(0,2)") ## copy number regions
```

```
ylims <- cbind(c(0, 5), c(-0.1, 1.1))
colG <- rep("#88888855", n)
hetCol <- "#00000088"
```

For convenience we define a custom plot function for this vignette:

```
plotFUN <- function(dataSet, tumorFraction) {
  regDat <- loadCnRegionData(dataSet=dataSet, tumorFraction=tumorFraction)
  sim <- getCopyNumberDataByResampling(n, bcp=bcp, regions=regions, regData=regDat)
  dat <- sim$profile
  wHet <- which(dat$genotype==1/2)
  colGG <- colG
  colGG[wHet] <- hetCol
  plotSeg(dat, sim$bcp, ylims=ylims, col=colGG)
}
```

```
tf <- 1
plotFUN(ds, tf)
```

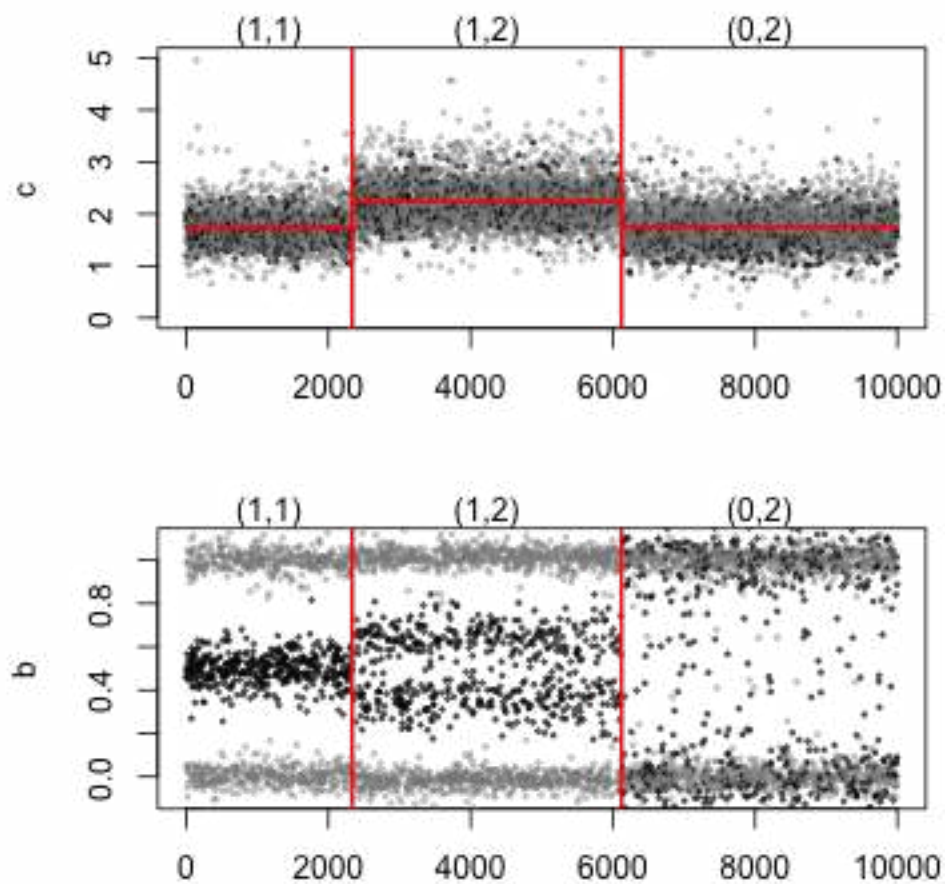


Figure 1: Data set GSE29172, 100% tumor cells

## 2 Affymetrix data

```
ds <- "GSE29172"
```

```
tf <- 1
plotFUN(ds, tf)
```

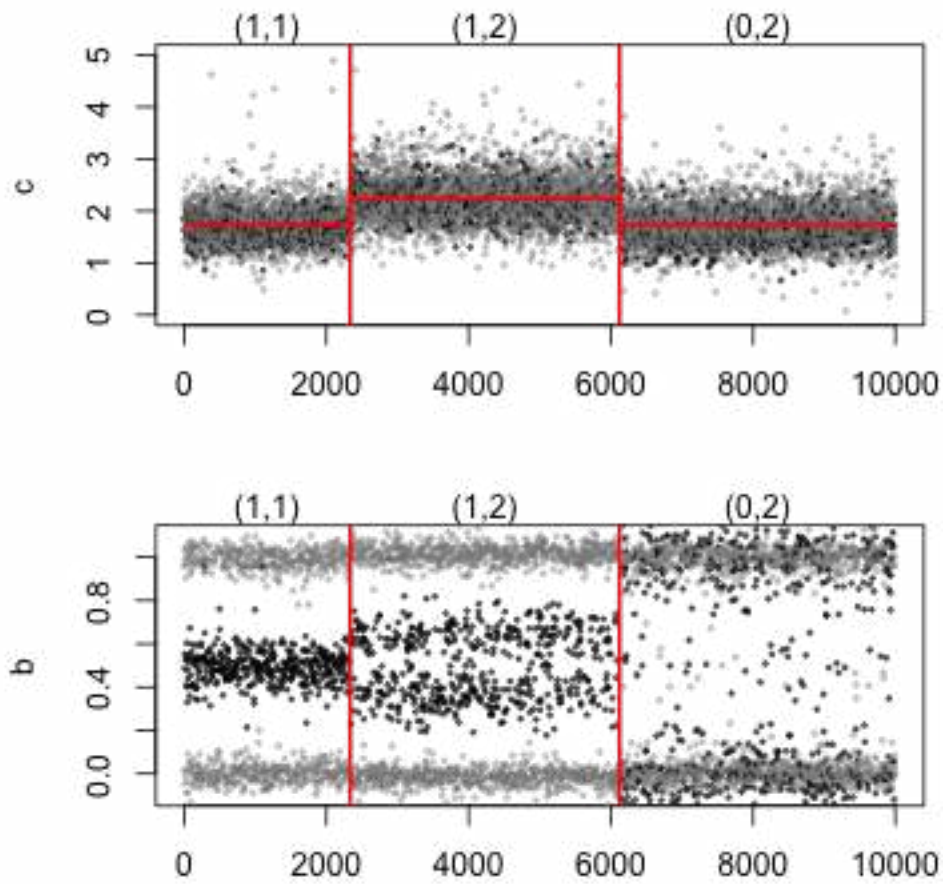


Figure 2: Data set GSE29172, 100% tumor cells (another resampling)

```
tf <- .7  
plotFUN(ds, tf)
```

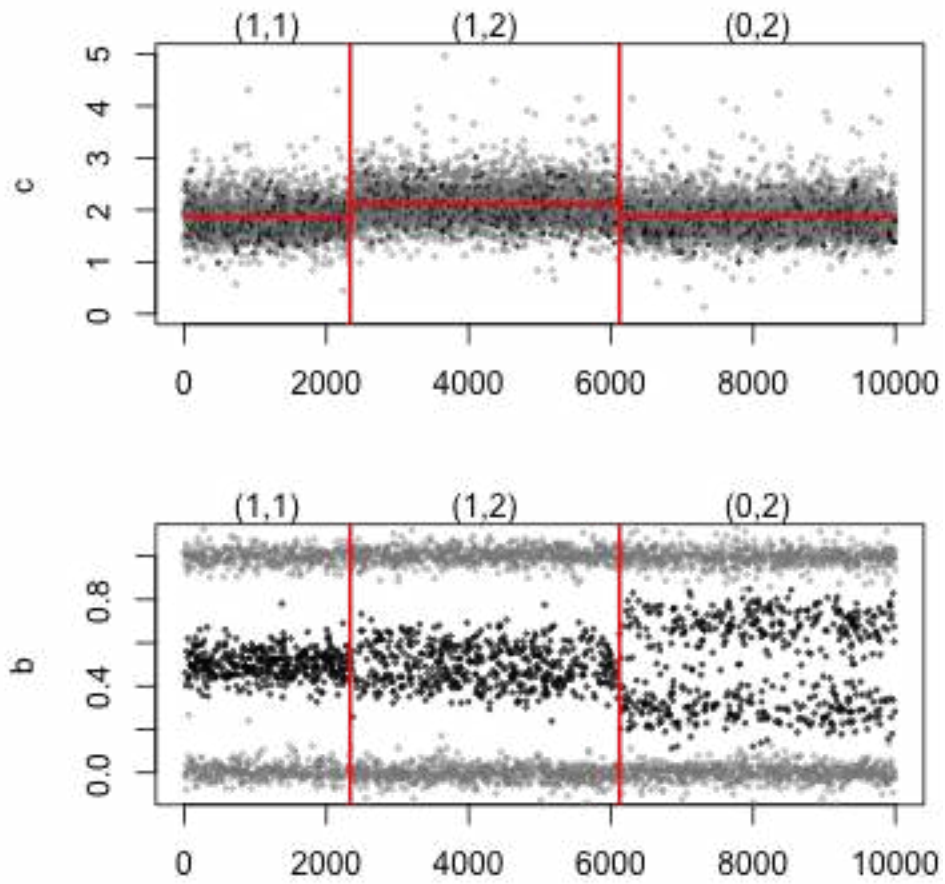


Figure 3: Data set GSE29172, 70% tumor cells

```
tf <- .5  
plotFUN(ds, tf)
```

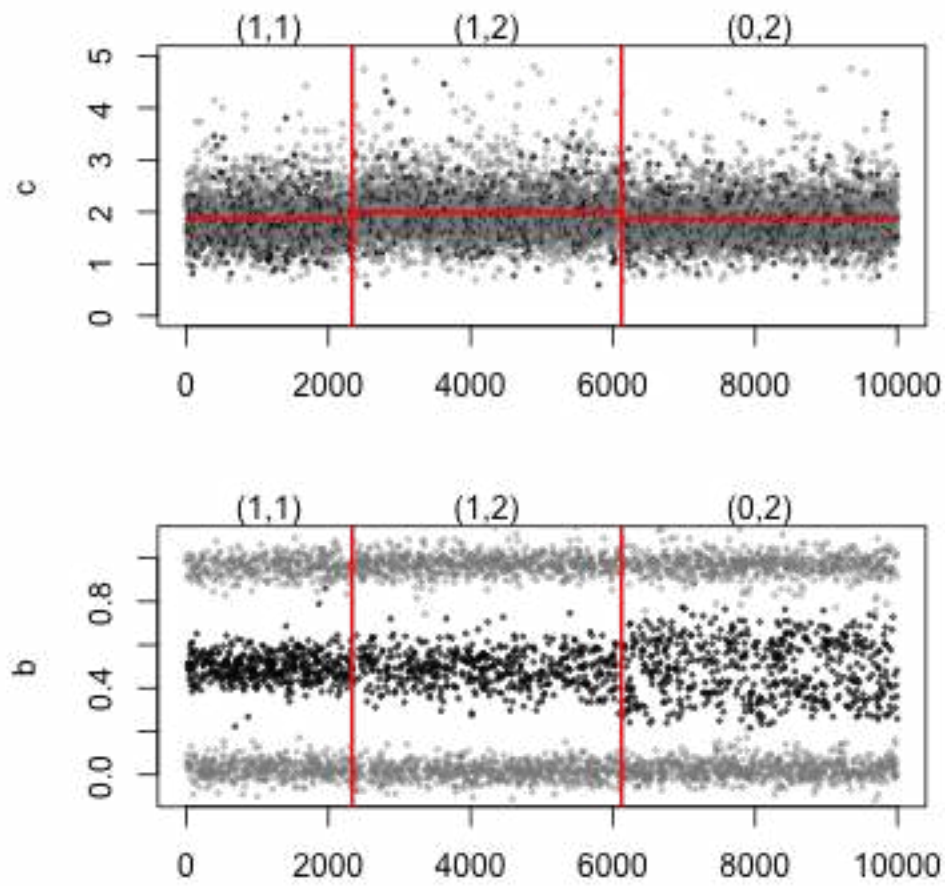


Figure 4: Data set GSE29172, 50% tumor cells

```
tf <- 1
plotFUN(ds, tf)
```

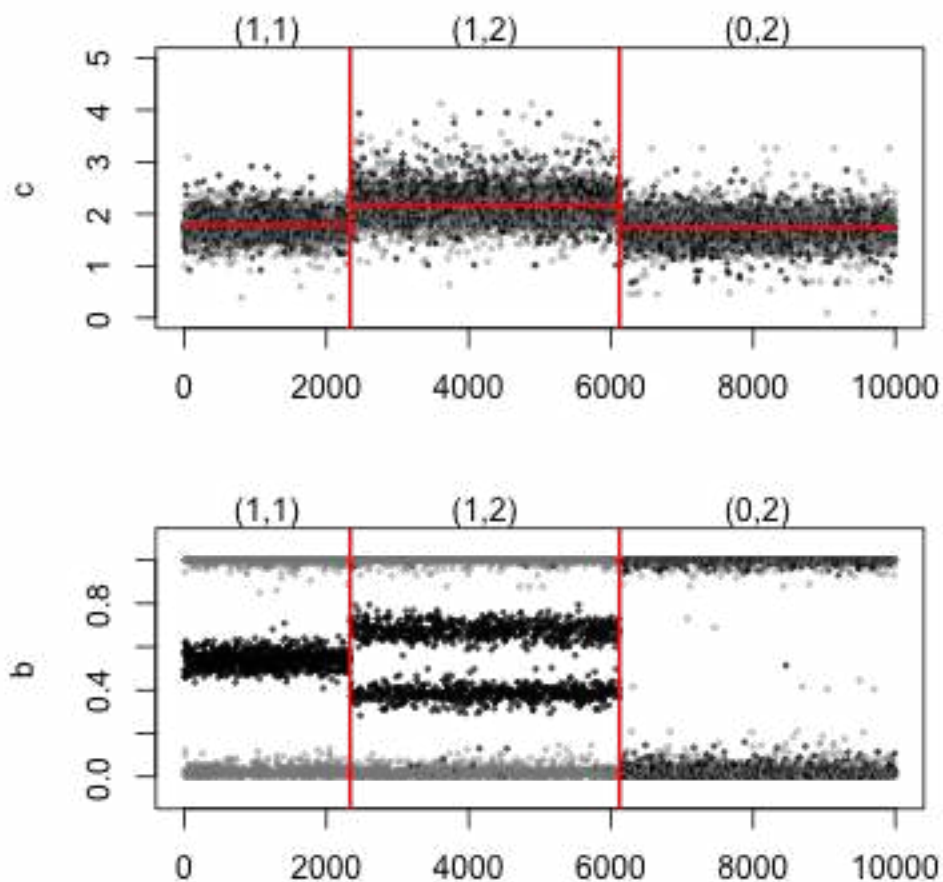


Figure 5: Data set GSE11976, 100% tumor cells

### 3 Illumina data

```
ds <- "GSE11976"
```

### References

- [1] Morgane Pierre-Jean, Guillem J Rigaiil, and Pierre Neuvial. Performance evaluation of DNA copy number segmentation methods. *Briefings in Bioinformatics*, to appear.

### A Session information

```
sessionInfo()
```

```
## R version 3.1.1 (2014-07-10)
```

```
tf <- 1
plotFUN(ds, tf)
```

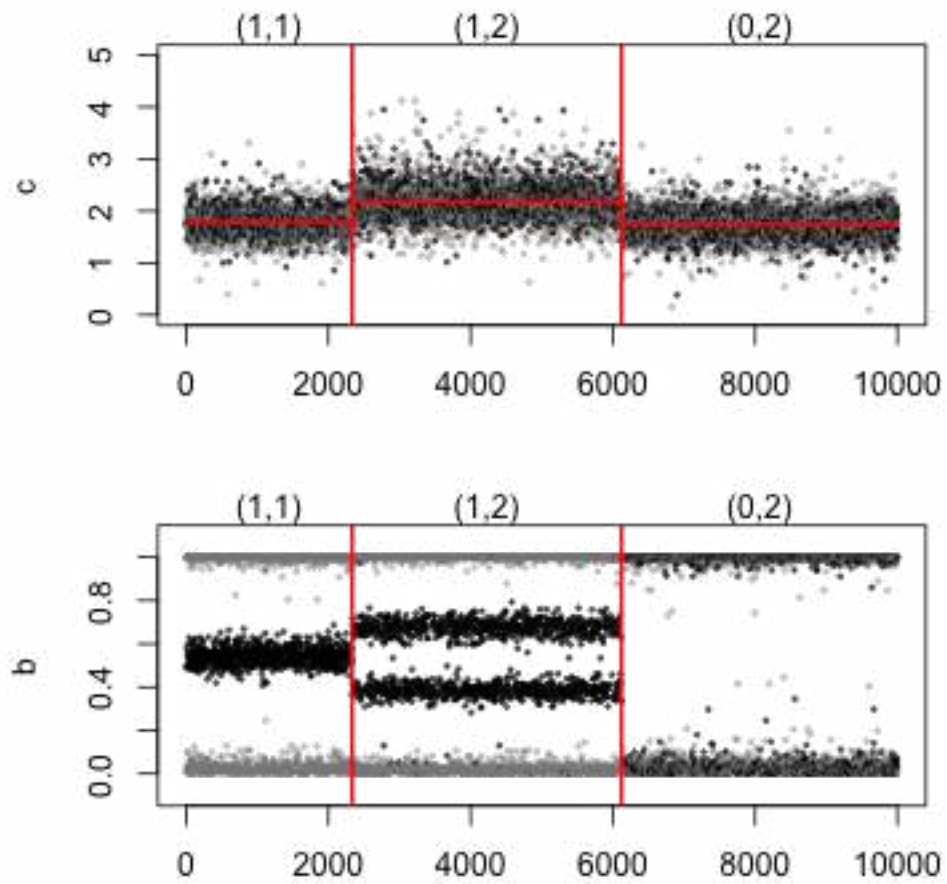


Figure 6: Data set GSE11976, 100% tumor cells (another resampling)

```
tf <- .79  
plotFUN(ds, tf)
```

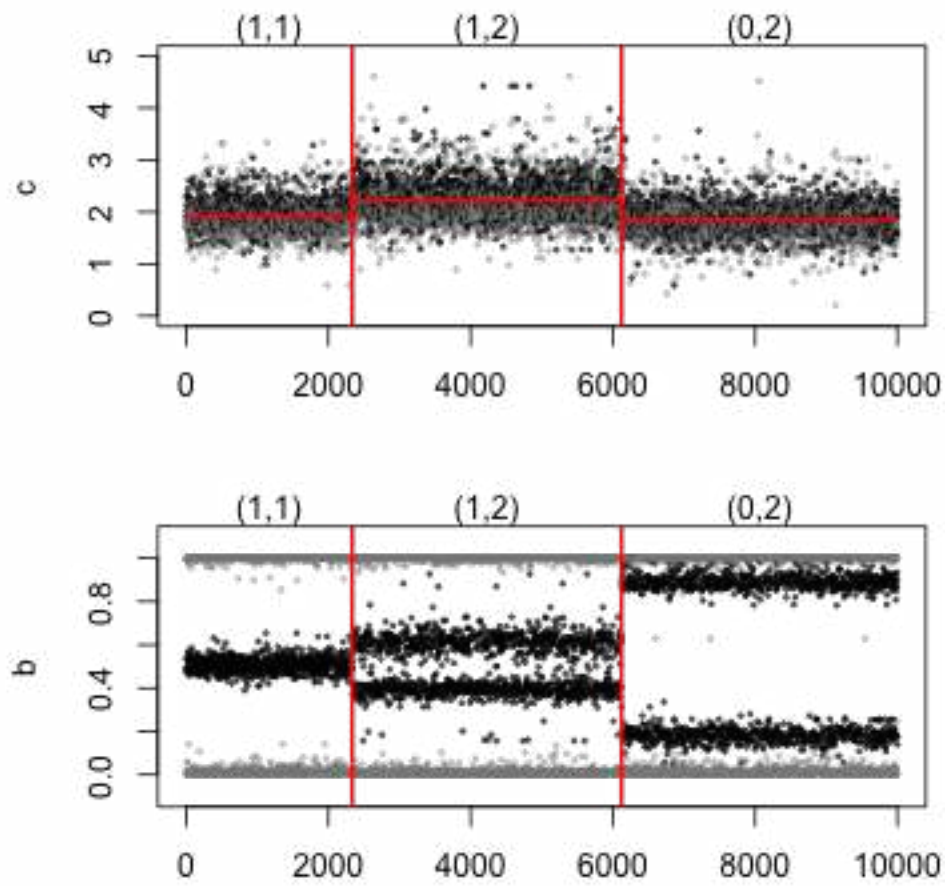


Figure 7: Data set GSE11976, 79% tumor cells



```
tf <- .5  
plotFUN(ds, tf)
```

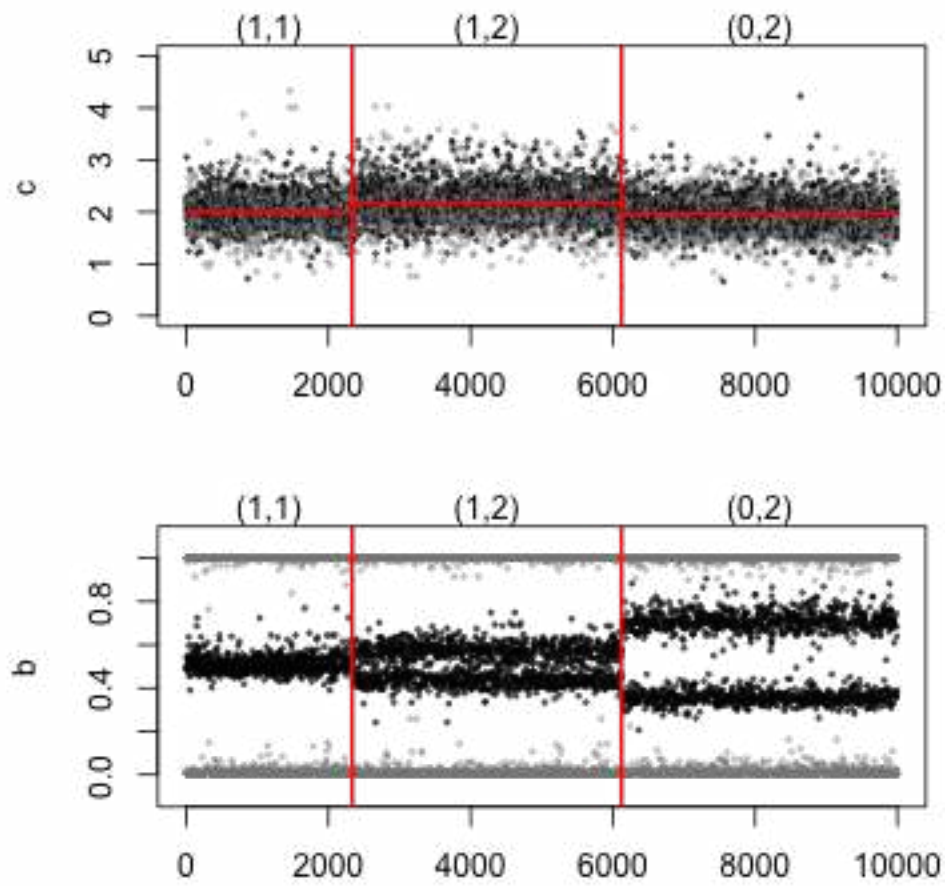


Figure 8: Data set GSE11976, 50% tumor cells

```
## Platform: x86_64-apple-darwin10.8.0 (64-bit)
##
## locale:
## [1] fr_FR.UTF-8/fr_FR.UTF-8/fr_FR.UTF-8/C/fr_FR.UTF-8/fr_FR.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  base
##
## other attached packages:
## [1] jointseg_0.6.1  acnr_0.2.0      R.utils_1.33.0  R.oo_1.18.2
## [5] R.methodsS3_1.6.2 knitr_1.6.10
##
## loaded via a namespace (and not attached):
## [1] DNACopy_1.38.1  evaluate_0.5.5  formatR_1.0      highr_0.3.1
## [5] matrixStats_0.10.1 methods_3.1.1   stringr_0.6.2    tools_3.1.1
```