

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

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

December 15, 2016

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].

Contents

1	Setup	2
2	Affymetrix data	3
3	Illumina data	7
A	Session information	7
B	Citing <code>jointseg</code>	11

Please see Appendix B for citing `jointseg`.

```
library(jointseg)
```

1 Setup

The parameters are defined as follows:

```
n <- 1e4                                ## signal length
bkp <- 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, bkp=bkp, regions=regions, regData=regDat)
  dat <- sim$profile
  wHet <- which(dat$genotype==1/2)
  colGG <- colG
  colGG[wHet] <- hetCol
  plotSeg(dat, sim$bkp, ylims=ylims, col=colGG)
}
```

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

```
## Error in ylims[, cc]: indice hors limites
```

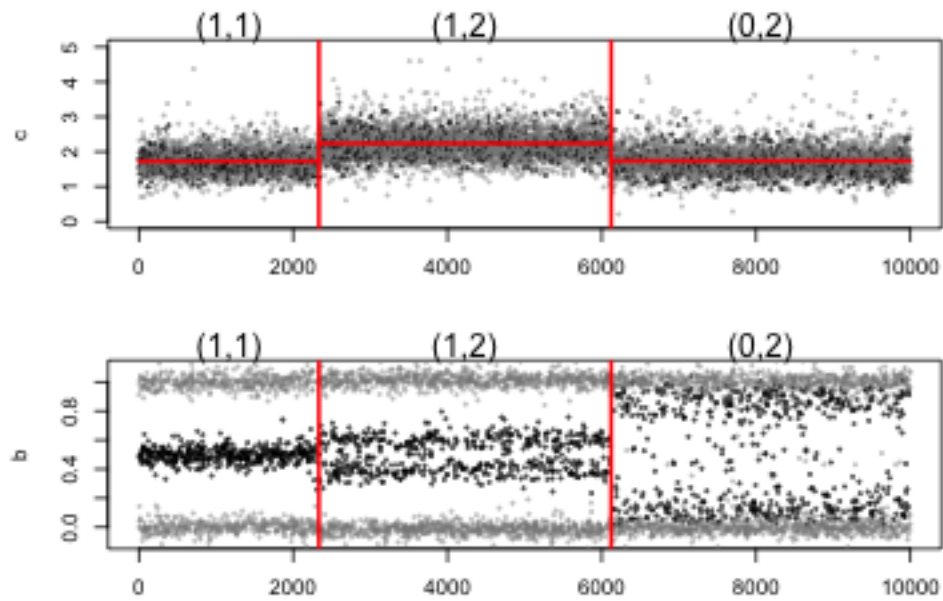


Figure 1: Data set GSE29172, 100% tumor cells

2 Affymetrix data

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

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

```
## Error in ylims[, cc]: indice hors limites
```

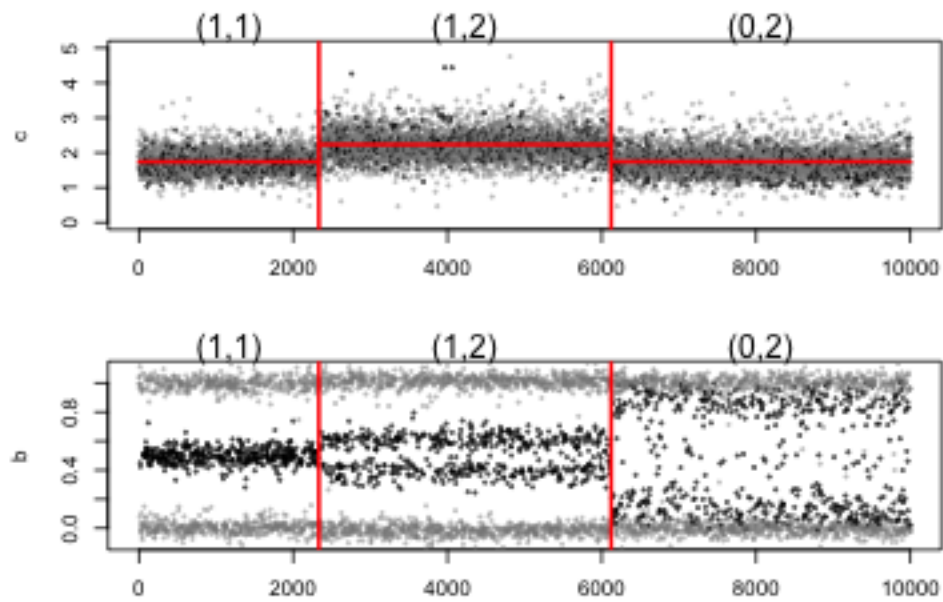


Figure 2: Data set GSE29172, 100% tumor cells (another resampling)

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

```
## Error in ylims[, cc]: indice hors limites
```

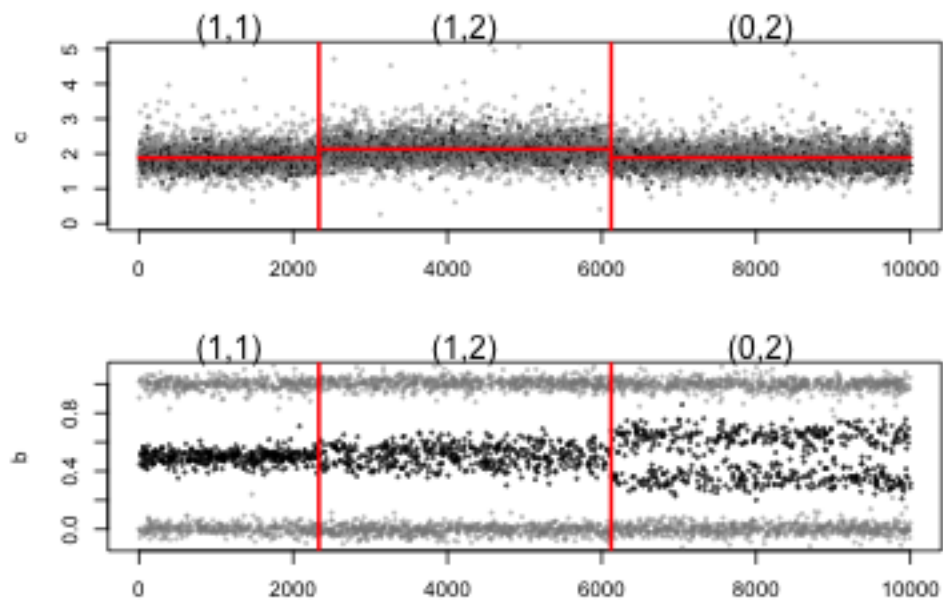


Figure 3: Data set GSE29172, 70% tumor cells

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

```
## Error in ylims[, cc]: indice hors limites
```

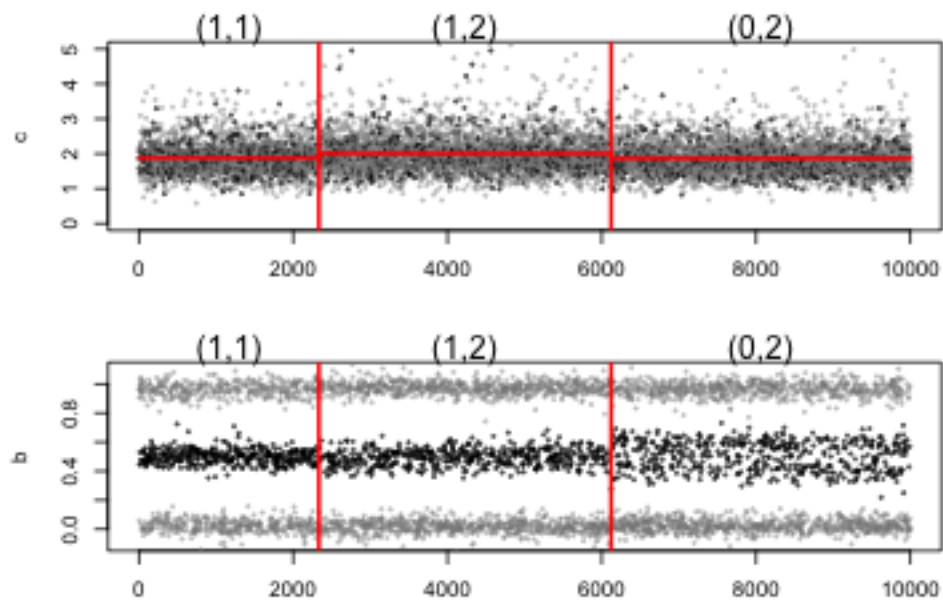


Figure 4: Data set GSE29172, 50% tumor cells

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

```
## Error in ylims[, cc]: indice hors limites
```

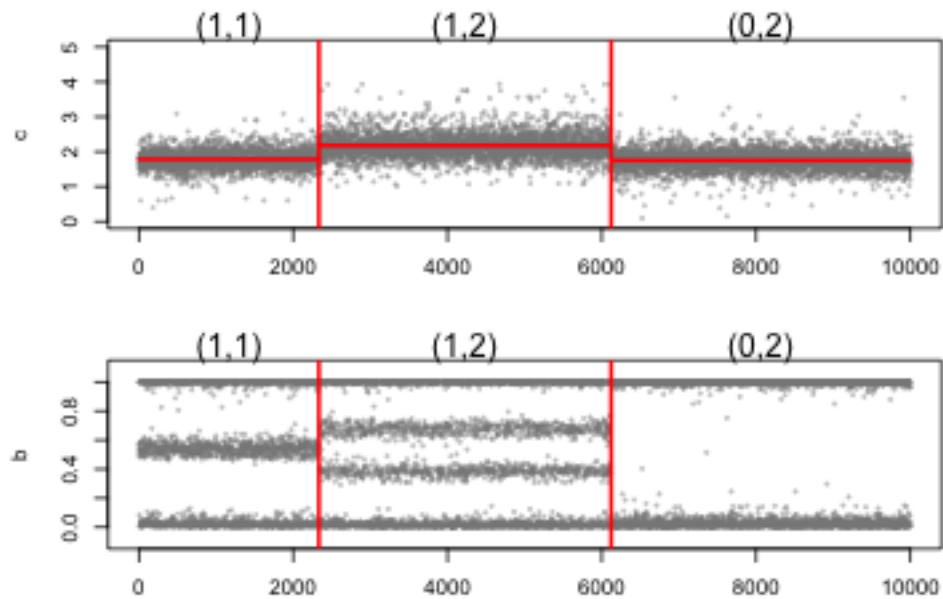


Figure 5: Data set GSE11976, 100% tumor cells

3 Illumina data

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

A Session information

```
sessionInfo()

## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X El Capitan 10.11.6
##
## locale:
```

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

```
## Error in ylims[, cc]: indice hors limites
```

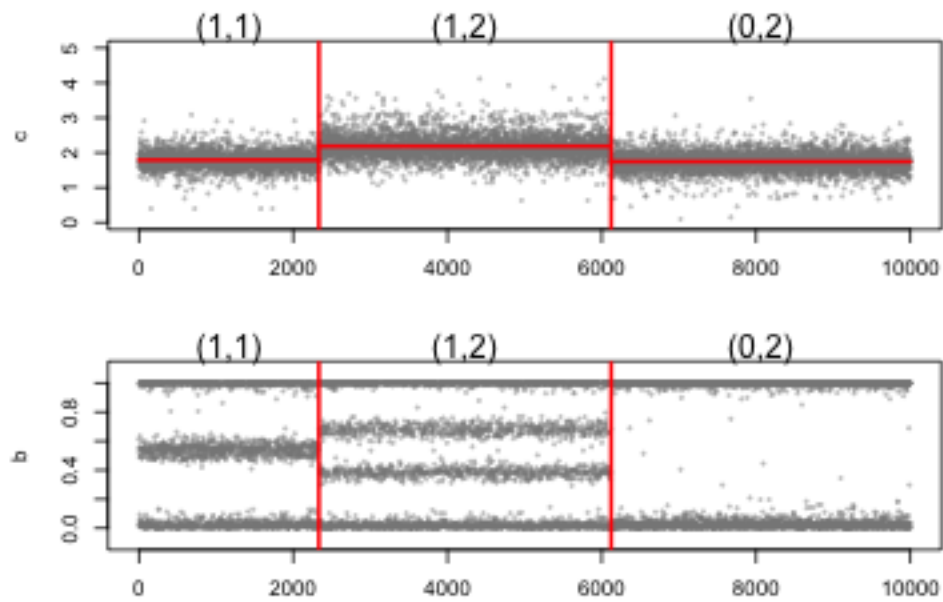


Figure 6: Data set GSE11976, 100% tumor cells (another resampling)


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

```
## Error in ylims[, cc]: indice hors limites
```

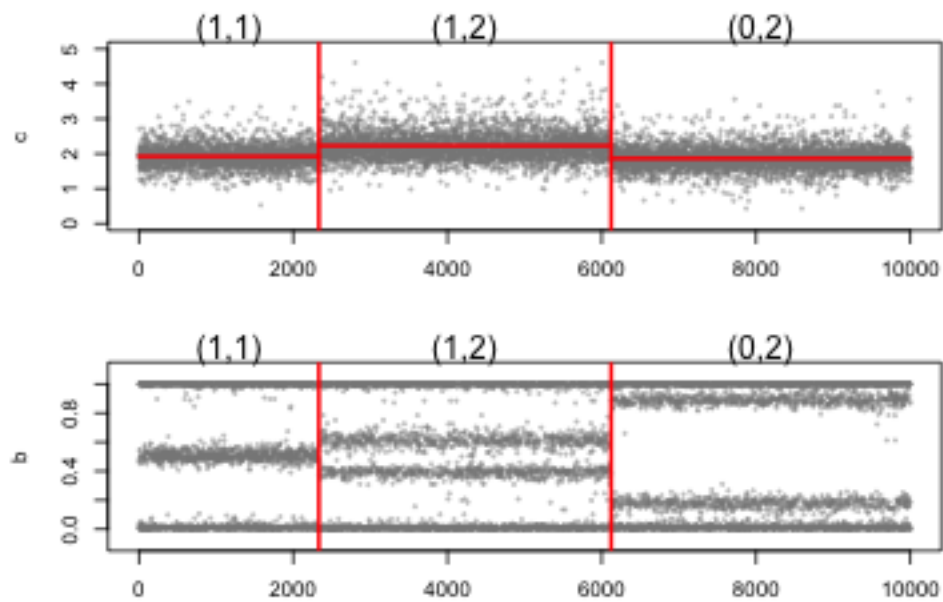


Figure 7: Data set GSE11976, 79% tumor cells

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

```
## Error in ylims[, cc]: indice hors limites
```

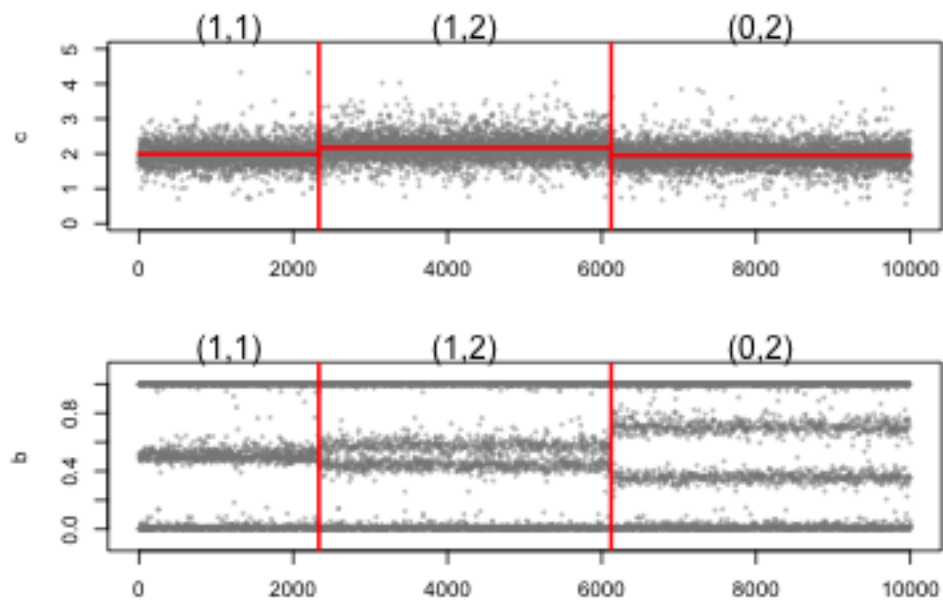


Figure 8: Data set GSE11976, 50% tumor cells

```
## [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  methods   base
##
## other attached packages:
## [1] jointseg_0.7.1 acnr_0.2.8      knitr_1.15.1
##
## loaded via a namespace (and not attached):
## [1] magrittr_1.5      DNACopy_1.46.0    tools_3.3.2       stringi_1.1.2
## [5] highr_0.6         stringr_1.1.0     matrixStats_0.51.0 evaluate_0.10
```

B Citing jointseg

```
citation("jointseg")

##
## To cite package 'jointseg' in publications, please use the following references:
##
## Morgane Pierre-Jean and Guillem Rigai11 and Pierre Neuvial (2016). jointseg:
## Joint segmentation of multivariate (copy number) signals. R package version
## 0.7.1.
##
## A BibTeX entry for LaTeX users is
##
## @Manual{pierre-jean14jointseg,
##   title = {jointseg: Joint segmentation of multivariate (copy number) signals},
##   author = {Morgane Pierre-Jean and Guillem Rigai11 and Pierre Neuvial},
##   year = {2016},
##   note = {R package version 0.7.1},
## }
##
## Morgane Pierre-Jean and Guillem Rigai11 and Pierre Neuvial (2014). Performance
## evaluation of DNA copy number segmentation methods. Brief Bioinform (2015) 16
## (4): 600-615.
##
## A BibTeX entry for LaTeX users is
##
## @Article{pierre-jean14performance,
##   title = {Performance evaluation of {DNA} copy number segmentation methods},
##   author = {Morgane Pierre-Jean and Guillem Rigai11 and Pierre Neuvial},
##   year = {2015},
##   institution = {Laboratoire de Math\`ematiques et Mod\`elisation d'\`Evry, Universit\`e d'\`Evry V
##   url = {http://bib.oxfordjournals.org/content/16/4/600.abstract},
##   journal = {Briefings in Bioinformatics},
## }
```

References

- [1] Morgane Pierre-Jean, Guillem J Rigai11, and Pierre Neuvial. Performance evaluation of DNA copy number segmentation methods. *Briefings in Bioinformatics*, to appear.