

GLAD package : Gain and Loss Analysis of DNA

Philippe Hupe^{1,2} and Emmanuel Barillot²

October 22, 2008

1. UMR 144 CNRS/Institut Curie, Institut Curie, 26, rue d'Ulm, Paris, 75248 cedex 05, France
2. Service Bioinformatique, Institut Curie, 26, rue d'Ulm, Paris, 75248 cedex 05, France
glad@curie.fr <http://bioinfo.curie.fr>

Contents

1	Overview	1
2	Data	1
2.1	Public data set	1
2.2	Bladder cancer data	2
3	GLAD classes	2
3.1	arrayCGH	2
3.2	profileCGH and profileChr	2
4	Analysis of array CGH profile	3
4.1	The <i>glad</i> function	3
4.2	The <i>daglad</i> function	4
4.3	Tuning parameters	6
5	Graphical functions	6
5.1	Plot of raw array data	6
5.2	Plot of genomic profile	9

1 Overview

This document presents an overview of the GLAD package (Gain and Loss Analysis of DNA). This package is devoted to the analysis of Array Comparative Genomic Hybridization (array CGH) (Pinkel et al., 1998; Snijders et al., 2001; Solinas-Toldo et al., 1997; Ishkanian et al., 2004) . The methodology for detecting the breakpoints delimiting altered regions in genomic patterns and assigning a status (normal, gained or lost) to each chromosomal region described in the paper Hupé et al. (2004) is implemented in this package. Some graphical functions are provided as well.

2 Data

2.1 Public data set

We used the public data set described in Snijders et al. (2001). The data correspond to 15 human cell strains with known karyotypes (12 fibroblast cell strains, 2 chorionic villus cell strains, 1 lymphoblast

cell strain) from the NIGMS Human Genetics Cell Repository (<http://locus.umdj.edu/nigms>). Each cell strain has been hybridized with an array CGH of 2276 BAC's, spotted in triplicate. Two array CGH profiles from the data obtained by Veltman et al. (2003) are available.

2.2 Bladder cancer data

Bladder cancer data from tumors collected at Henri Mondor Hospital (Créteil, France) (Billerey et al., 2001) have been hybridized on arrays CGH composed of 2464 BACs (Radvanyi, Pinkel et al., unpublished results). In this data, only the log-ratios are provided and no information about clones is available since the data are not yet published. These data allow only some graphical functionalities to be shown and will be used as a support to illustrate some functions for array normalization (not yet available in the current version of the package).

3 GLAD classes

3.1 arrayCGH

This class stores raw values after images analysis. The object `arrayCGH` is a list with at least a `data.frame` named `arrayValues` and a vector named `arrayDesign`. The `data.frame` `arrayValues` must contain the following fields:

Col Vector of columns coordinates.

Row Vector of rows coordinates.

... Other elements can be added.

The vector `arrayDesign` is composed of 4 values : `c(arrayCol, arrayRow, SpotCol, SpotRow)`. The array CGH is represented by `arrayRow*arrayCol` blocs and each bloc is composed of `SpotRow*SpotCol` spots. N.B. : `Col` takes the values in `1:arrayRow*SpotRow` and `Row` takes the values in `1:arrayCol*SpotCol`

3.2 profileCGH and profileChr

This class stores synthetic values related to each clone available onto the `arrayCGH`. The object `profileChr` corresponds to data of only one chromosome. Objects `profileCGH` and `profileChr` are composed of a list with the first element `profileValues` which is a `data.frame` with the following columns names:

LogRatio Test over Reference log-ratio.

PosOrder The rank position of each clone on the genome.

PosBase The base position of each clone on the genome.

Chromosome Chromosome name.

Clone The name of the corresponding clone.

... Other elements can be added.

`LogRatio`, `Chromosome` and `PosOrder` are compulsory.

To create those objects you can use the function *as.profileCGH*.

4 Analysis of array CGH profile

4.1 The *glad* function

A result of the GLAD methodology on cell line gm13330 (Snijders et al., 2001) is presented in Figure 1.

```
[1] "Have fun with GLAD"  
  
> data(snijders)  
> profileCGH <- as.profileCGH(gm13330)  
> res <- glad(profileCGH, mediancenter = FALSE, smoothfunc = "lawsglad",  
+   bandwidth = 10, round = 1.5, model = "Gaussian", lkern = "Exponential",  
+   qlambda = 0.999, base = FALSE, lambdabreak = 8, lambdacluster = 8,  
+   lambdaclusterGen = 40, type = "tricubic", param = c(d = 6),  
+   alpha = 0.001, msize = 5, method = "centroid", nmax = 8,  
+   verbose = FALSE)  
  
[1] "You can set bandwidth to 1 to decrease computation time"
```

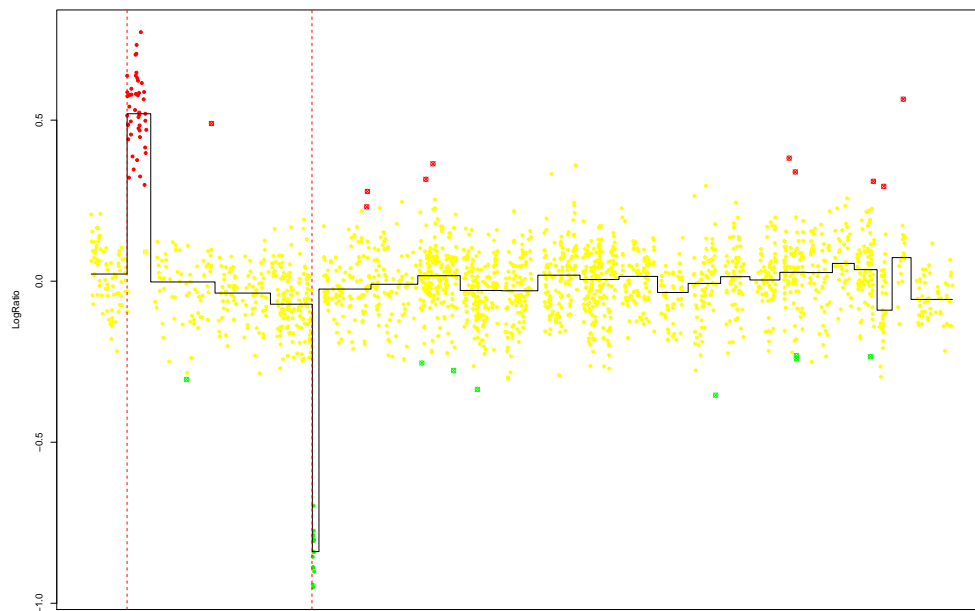


Figure 1: Results of glad on cell line gm13330 (Snijders data).

4.2 The *daglad* function

The algorithm implemented in this function is a slightly modified version of the GLAD algorithm.

```
> data(veltman)
> profileCGH <- as.profileCGH(P9)
> profileCGH <- daglad(profileCGH, mediancenter = FALSE, normalrefcenter = FALSE,
+   genomestep = FALSE, smoothfunc = "lawsglad", lkern = "Exponential",
+   model = "Gaussian", qlambda = 0.999, bandwidth = 10, base = FALSE,
+   round = 1.5, lambdabreak = 8, lambdaclusterGen = 40, param = c(d = 6),
+   alpha = 0.001, msize = 5, method = "centroid", nmin = 1,
+   nmax = 8, amplicon = 1, deletion = -5, deltaN = 0.2, forceGL = c(-0.3,
+   0.3), nbsigma = 3, MinBkpWeight = 0.35, CheckBkpPos = TRUE)
```

```
[1] "You can set bandwidth to 1 to decrease computation time"
[1] "Smoothing for each Chromosome"
[1] "Optimization of the Breakpoints"
[1] "Check Breakpoints Position"
```

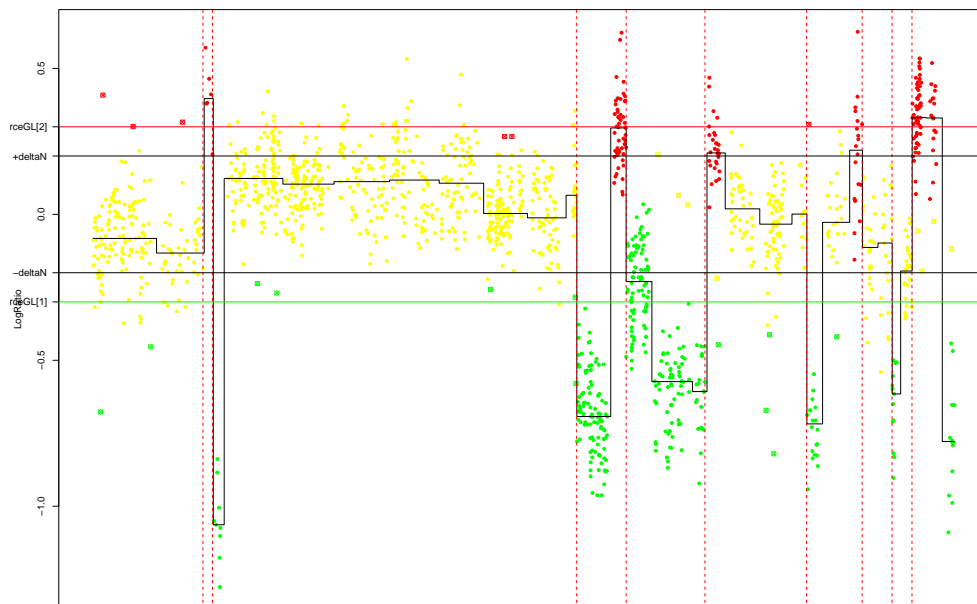


Figure 2: Results of *daglad* on the patient P9 (Veltman data).

The *daglad* function allows to choose some threshold to help the algorithm to identify the status of the genomic regions. The thresholds are given in the following parameters:

- `deltaN`
- `forceGL`
- `deletion`
- `amplicon`

Comparing **Figure 2** and **Figure 3** shows the influence of two different sets of parameters.

```
> data(veltman)
> profileCGH <- as.profileCGH(P9)
> profileCGH <- daglad(profileCGH, mediancenter = FALSE, normalrefcenter = FALSE,
+   genomestep = FALSE, smoothfunc = "lawsglad", lkern = "Exponential",
+   model = "Gaussian", qlambda = 0.999, bandwidth = 10, base = FALSE,
+   round = 1.5, lambdabreak = 8, lambdaclusterGen = 40, param = c(d = 6),
+   alpha = 0.001, msize = 5, method = "centroid", nmin = 1,
+   nmax = 8, amplicon = 1, deletion = -5, deltaN = 0.1, forceGL = c(-0.15,
+   0.15), nbsigma = 3, MinBkpWeight = 0.35, CheckBkpPos = TRUE)

[1] "You can set bandwidth to 1 to decrease computation time"
[1] "Smoothing for each Chromosome"
[1] "Optimization of the Breakpoints"
[1] "Check Breakpoints Position"
```

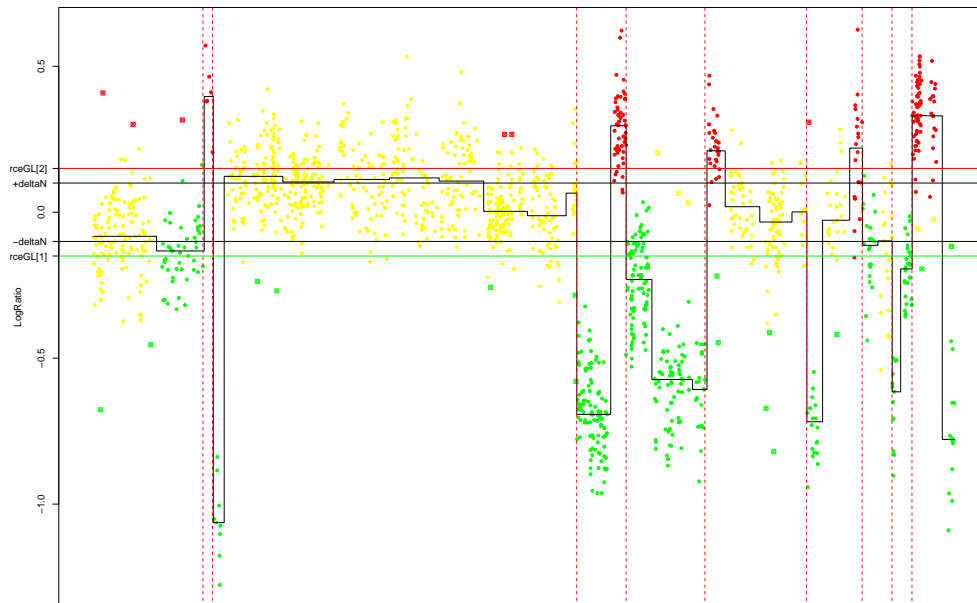


Figure 3: Results of daglad on the patient P9 (Veltman data) - Influence of the thresholds.

The *daglad* function allows a smoothing step over the whole genome (if *genomestep=TRUE*) where all the chromosomes are concatenated together. During this step, the cluster which corresponds to the Normal DNA level is identified: the thresholds used in the function (*deltaN*, *forceGL*, *amplicon*, *deletion*) are then compared to the median of this cluster.

4.3 Tuning parameters

The most important parameters are:

- *lambdabreak*
- *lambdacluster*
- *lambdaclusterGen*
- *param* $c(d = 6)$

Decreasing those parameters will lead to a higher number of breakpoints identified. For arrays experiments with very small Signal to Noise ratio it is recommended to use a small value of *param* like $d = 3$ or less.

5 Graphical functions

5.1 Plot of raw array data

```
> data(arrayCGH)
> array <- list(arrayValues = array2, arrayDesign = c(4, 4, 21,
+ 22))
> class(array) <- "arrayCGH"
```

```
> arrayPlot(array, "Log2Rat", bar = "none")
```

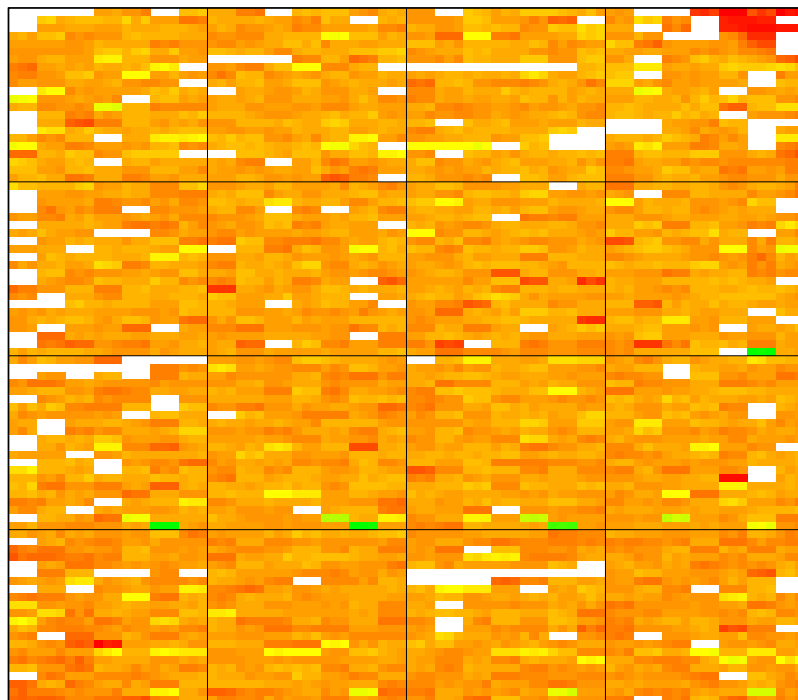


Figure 4: Spatial image of array CGH

```
> arrayPersp(array, "Log2Rat", box = FALSE, theta = 110, phi = 40,  
+           bar = FALSE)
```

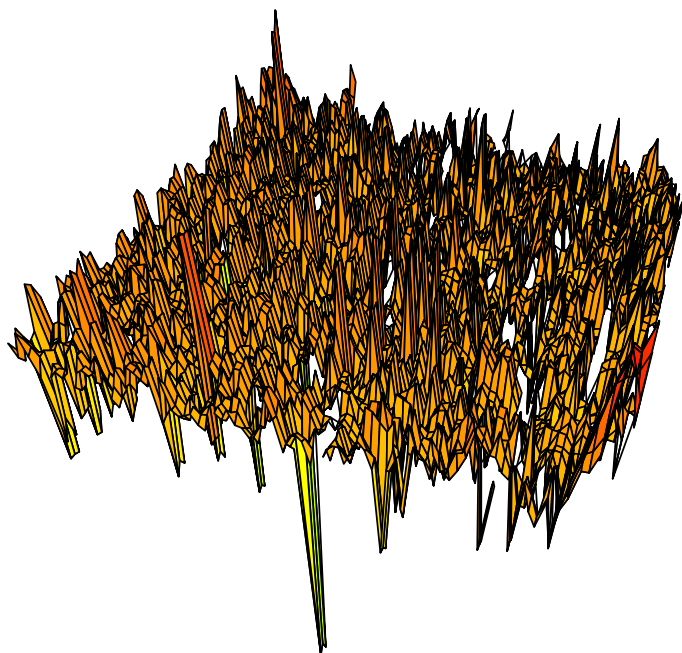


Figure 5: Perspective image of array CGH

5.2 Plot of genomic profile

```
[1] "You can set bandwidth to 1 to decrease computation time"
```

```
> plotProfile(res, unit = 3, Bkp = TRUE, labels = FALSE, Smoothing = "Smoothing",  
+ plotband = FALSE)
```

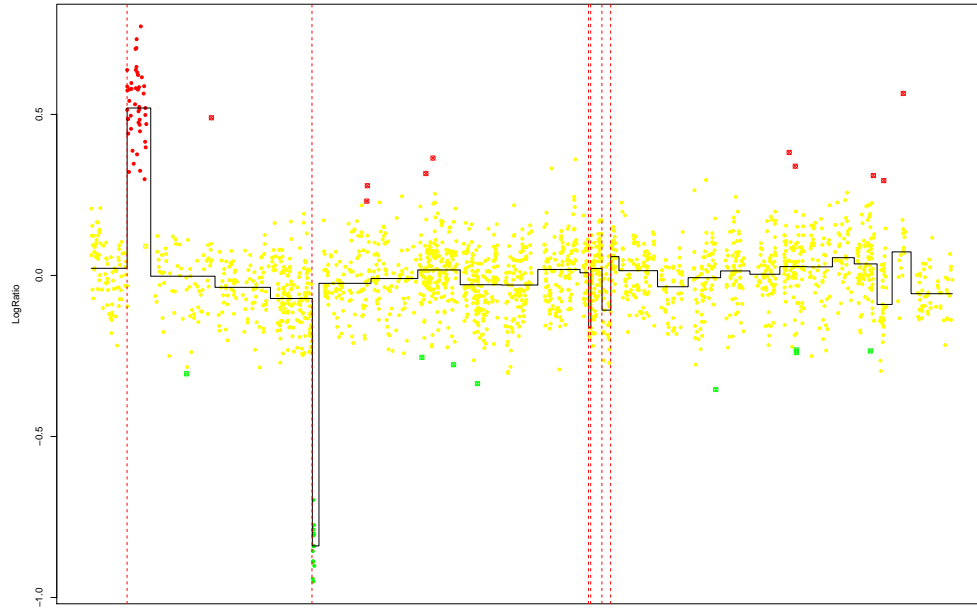


Figure 6: Genomic profile on the whole genome

```
> plotProfile(res, unit = 3, Bkp = TRUE, labels = FALSE, Smoothing = "Smoothing")
```

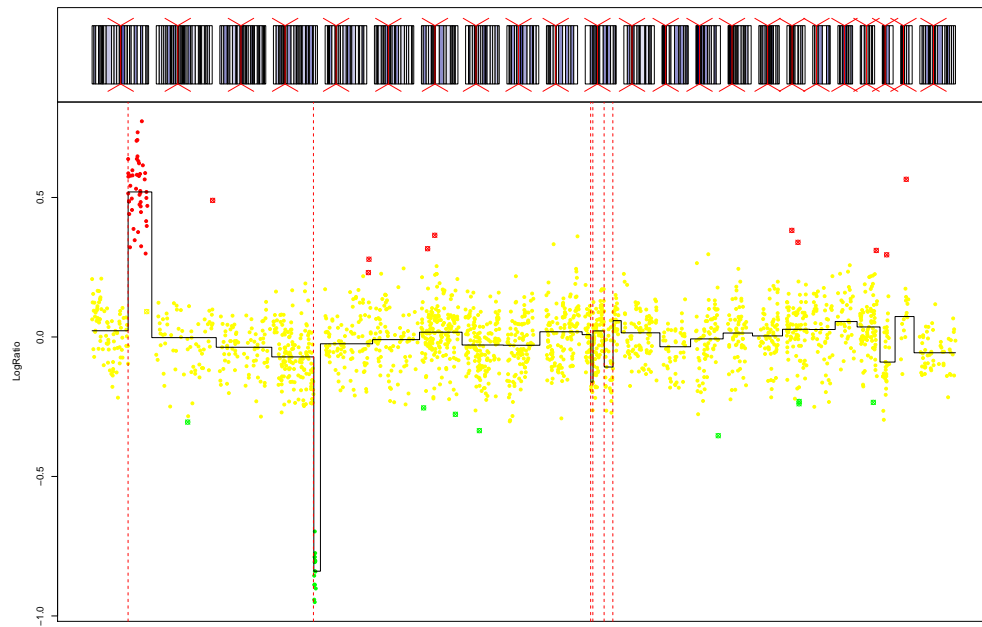


Figure 7: Genomic profile on the whole genome and cytogenetic banding

```
> text <- list(x = c(90000, 2e+05), y = c(0.15, 0.3), labels = c("NORMAL",  
+ "GAIN"), cex = 2)  
> plotProfile(res, unit = 3, Bkp = TRUE, labels = TRUE, Chromosome = 1,  
+ Smoothing = "Smoothing", plotband = FALSE, text = text)
```

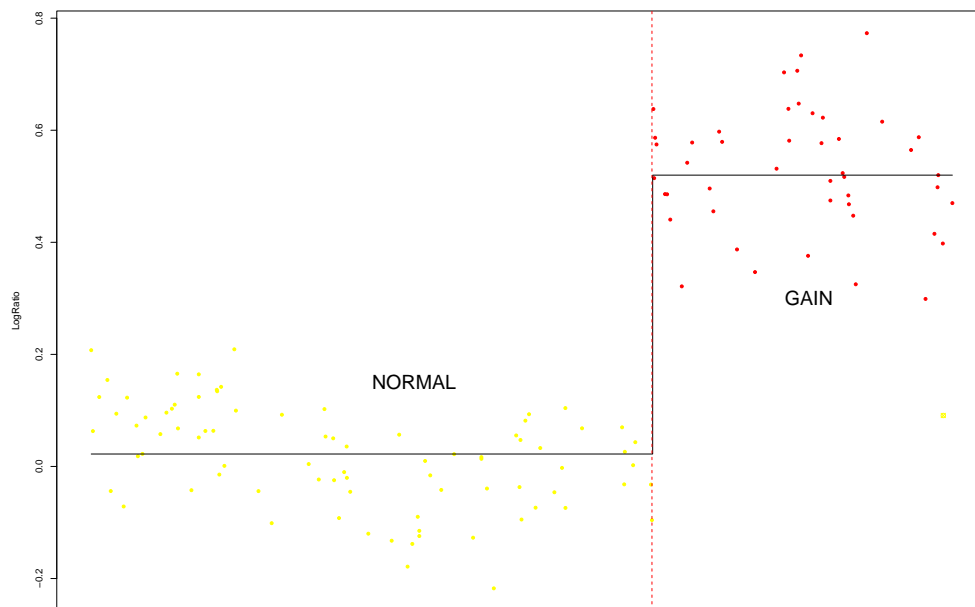


Figure 8: Genomic profile for chromosome 1

```

> text <- list(x = c(90000, 2e+05), y = c(0.15, 0.3), labels = c("NORMAL",
+   "GAIN"), cex = 2)
> plotProfile(res, unit = 3, Bkp = TRUE, labels = TRUE, Chromosome = 1,
+   Smoothing = "Smoothing", text = text, main = "Chromosome 1")

```

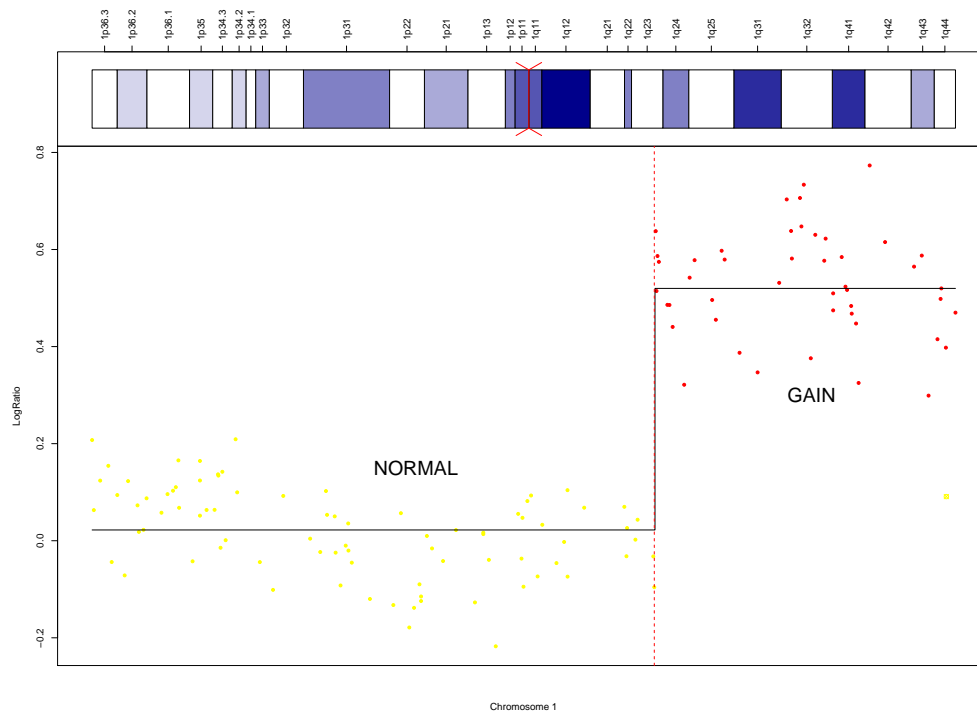


Figure 9: Genomic profile for chromosome 1 and cytogenetic banding with labels

References

- Billerey, C., Chopin, D., Aubriot-Lorton, M. H., Ricol, D., de Medina, S. G. D., Rhijn, B. V., Bralet, M. P., Lefrere-Belda, M. A., Lahaye, J. B., Abbou, C. C., Bonaventure, J., Zafrani, E. S., van der Kwast, T., Thiery, J. P., and Radvanyi, F. (2001). Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *Am. J. Pathol.*, 158:955–1959.
- Hupé, P., Stransky, N., Thiery, J. P., Radvanyi, F., and Barillot, E. (2004). Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, 20:3413–3422.
- Ishkanian, A. S., Malloff, C. A., Watson, S. K., DeLeeuw, R. J., Chi, B., Coe, B. P., Snijders, A., Albertson, D. G., Pinkel, D., Marra, M. A., Ling, V., MacAulay, C., and Lam, W. L. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.*, 36:299–303.
- Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. (1998). High resolution analysis of dna copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, 20:207–211.
- Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., S, S. L., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). Assembly of microarrays for genome-wide measurement of dna copy number. *Nat. Genet.*, 29:263–4.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 20:399–407.
- Veltman, J. A., Fridlyand, J., Pejavar, S., Olshen, A. B., Korkola, J. E., DeVries, S., Carroll, P., Kuo, W.-L., Pinkel, D., Albertson, D., Cordon-Cardo, C., Jain, A. N., and Waldman, F. M. (2003). Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res*, 63:2872–2880.