# sequenza possible SNP-array usage example

Francesco Favero*, Aron C. Eklund

September 27, 2013

## Contents



# 1 Working with SNP array data

```
> library(sequenza)
```

## 1.1 Preparing the data

```
> data(BAF)
> data(logR)

> sample.i <- data.frame(chromosome = BAF$chrs, n.base = BAF$pos,
+                        Bf = BAF$S1, adjusted.ratio = logR$S1,
+                        depth.sample = 1, good.s.reads = 1,
+                        ref.zygosity = 'hom', stringsAsFactors = FALSE)
```

*favero@cbs.dtu.dk

### 1.1.1 Correcting logR with a normal sample, or with the mean logR value

Without a reference sample (normal germline sample) we can try to divide for the mean value. It would be correct to use the germline logR.

```
> #sample.i$adjusted.ratio <- 2^(sample.i$adjusted.ratio)
> #sample.i$adjusted.ratio <- sample.i$adjusted.ratio / mean(sample.i$adjusted.ratio
> sample.i$adjusted.ratio <- 2^(sample.i$adjusted.ratio/0.55)
>
```

### 1.1.2 Retrive the homozygous position

It should be available a germline sample to get the heterozygours SNP, doing in the same sample it's a risk if the sample is pure. A treshold around 0.25 or 0.35 can be picked to subset the heterozygous position on the germline. In the example we are lowering the treshold while taking the SNP from the same aberrant sample.

```
> het.lim <- 0.2
> is.het <- sample.i$Bf >= het.lim & sample.i$Bf <= 1 - het.lim
> sample.i$ref.zygosity[is.het] <- 'het'
> sample.i$Bf[sample.i$Bf >= 0.5] <- 1 - sample.i$Bf[sample.i$Bf >= 0.5]
> sample.het.i <- sample.i[is.het, ]
```

## 1.2 Windowing logR values.

```
> snp.r.win <- windowValues(x = sample.i$adjusted.ratio,
+                           positions = sample.i$n.base,
+                           chromosomes = sample.i$chromosome,
+                           window = 1e6, overlap = 1)
```

## 1.3 Windowing B-allele frequencies values.

```
> snp.b.win <- windowValues(x = sample.het.i$Bf,
+                           positions = sample.het.i$n.base,
+                           chromosomes = sample.het.i$chromosome,
+                           window = 1e6, overlap = 1)
```

## 1.4 Chromosome view without mutation

```
> chromosome.view(baf.windows = snp.b.win[[1]],
+                 ratio.windows = snp.r.win[[1]],
+                  min.N.ratio = 1)
```

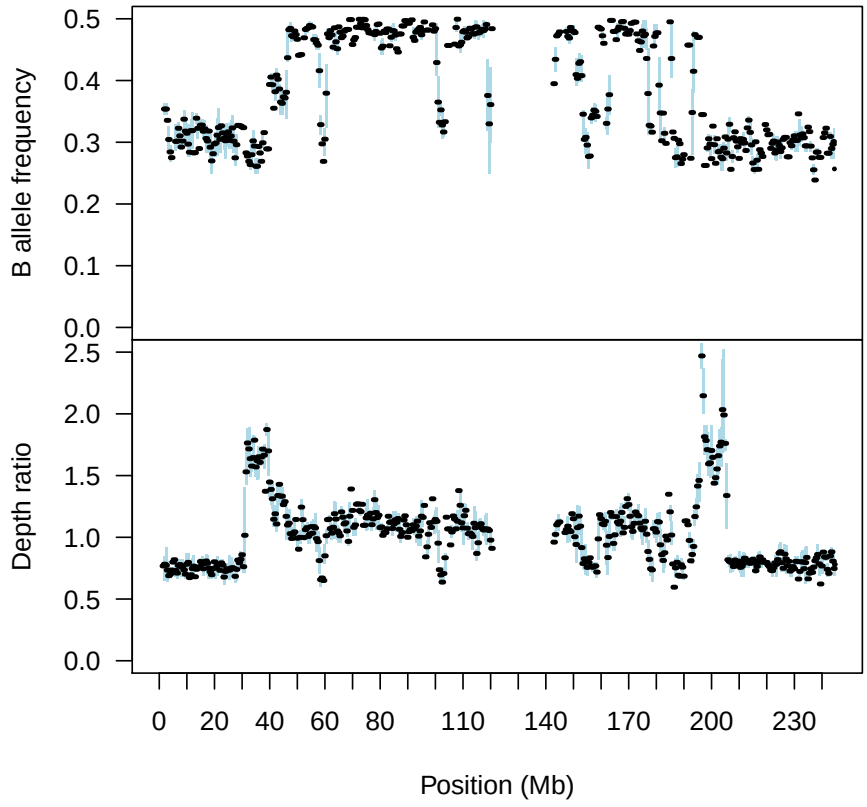


Figure 1: Plots B-allele frequencies (top) and un-logged-logR (bottom) with SNP array data.

## 1.5 Segmenting with the *copynumber* package

```
> breaks <- find.breaks(sample.het.i, gamma = 40, kmin = 20, baf.thres = c(0, 0.5))
> seg.i  <- segment.breaks(sample.i, breaks = breaks)
```

## 1.6 Using the Bayesian inference on segmented SNP arrays

```
> weights.snp   <- 150 + round((seg.i$end.pos - seg.i$start.pos)/1e6 , 0)
> filter.size   <- (seg.i$end.pos - seg.i$start.pos) >= 3e6

> avg.unlogR <- mean(sample.i$adjusted.ratio, na.rm = TRUE)
> CP.snp <- baf.model.fit(Bf = seg.i$Bf[filter.size],
+                         depth.ratio = seg.i$depth.ratio[filter.size],
+                         weight.ratio = weights.snp[filter.size],
+                         weight.Bf = weights.snp[filter.size],
+                         avg.depth.ratio = avg.unlogR,
+                         cellularity = seq(0.1,1,0.01),
+                         dna.index = seq(0.5,3,0.05), mc.cores = 4,
+                         priors.labels = 2, priors.values = 2)

> cint <- get.ci(CP.snp)
> cellularity <- cint$max.y
> dna.index   <- cint$max.x
```

## 1.7 Cellularity and DNA-index plot for SNP array

```
> cp.plot(CP.snp)
```

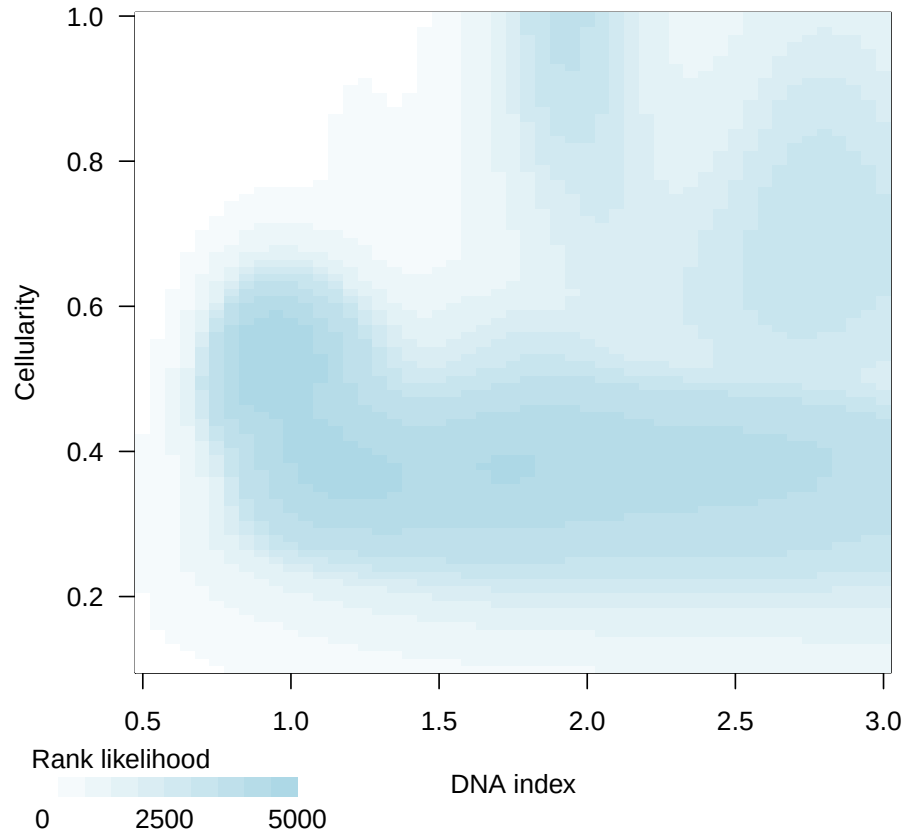


Figure 2: Result from the Bayesian inference over the defined range of cellularity and DNA-index from artificial SNP array data. The color indicate the log-likelihood of the corresponding cellularity/DNA-index values.

```
>     cp.plot.contours(CP.snp)
```

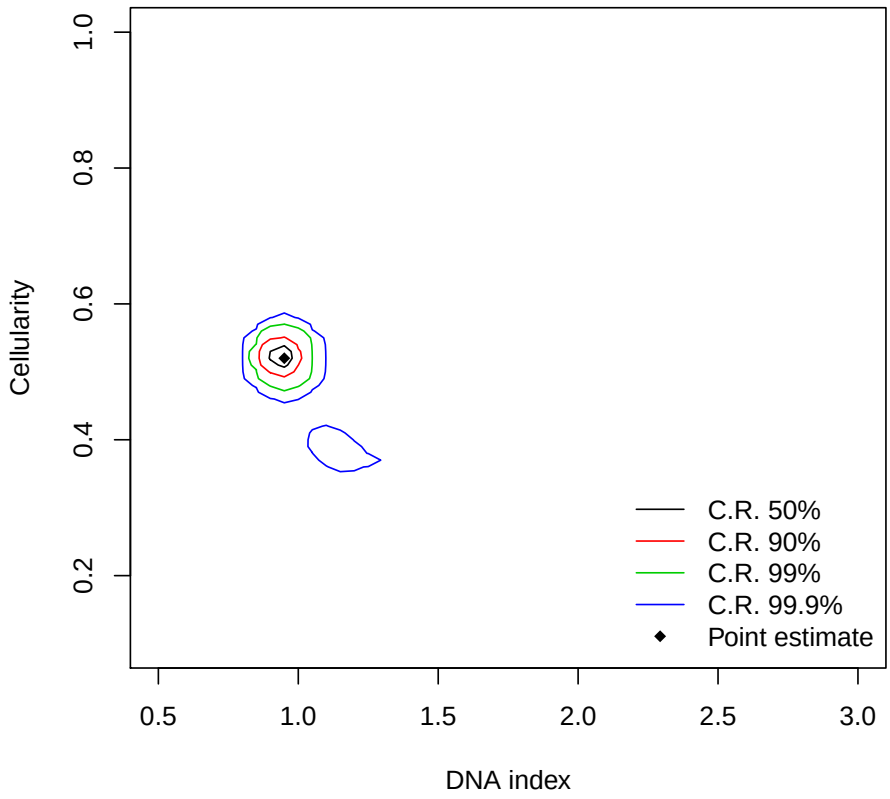


Figure 3: Plot of the log likelihood with respective cellularity and DNA-index probability distribution and confidence intervals.

## 1.8 Call for copy number variation using inferred parameters.

```
> snp.seg.cn <- baf.bayes(Bf = seg.i$Bf,
+                         depth.ratio = seg.i$depth.ratio,
+                         avg.depth.ratio = avg.unlogR,
+                         cellularity = cellularity,
+                         weight.ratio = 2 * 300,
+                         weight.Bf = 300, ratio.priority = FALSE,
+                         dna.index = dna.index, CNt.max = 10)
> segmented.snp <- cbind(seg.i, snp.seg.cn)
> head(segmented.snp[segmented.snp$chromosome == 1, ])
```

| | chromosome | start.pos | end.pos | Bf | N.BAF | depth.ratio | N.ratio | CNt | A | B |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2189662 | 28792900 | 0.3076600 | 85 | 0.7514180 | 130 | 1 | 1 | 0 |
| 2 | 1 | 29582868 | 40285096 | 0.2899053 | 19 | 1.5631176 | 43 | 4 | 3 | 1 |
| 3 | 1 | 40630391 | 46296225 | 0.3774100 | 20 | 1.2782663 | 29 | 3 | 2 | 1 |
| 4 | 1 | 46437972 | 57009803 | 0.4780000 | 31 | 1.0506103 | 43 | 2 | 1 | 1 |
| 5 | 1 | 57301533 | 64307493 | 0.4303579 | 19 | 0.9826375 | 28 | 2 | 1 | 1 |
| 6 | 1 | 65068455 | 100351185 | 0.4793541 | 61 | 1.1143564 | 107 | 2 | 1 | 1 |

| | L |
|---|---|
| 1 | -14.85780 |
| 2 | -16.02514 |
| 3 | -15.07017 |
| 4 | -14.56029 |
| 5 | -17.16355 |
| 6 | -14.83473 |

## 1.9 Graphical representation of copy number with SNP arrays

```
> chromosome.view(baf.windows = snp.b.win[[1]],
+                 ratio.windows = snp.r.win[[1]],  min.N.ratio = 1,
+                 segments = segmented.snp[segmented.snp$chromosome == "1", ],
+                 cellularity = cellularity, dna.index = dna.index,
+                 avg.depth.ratio = avg.unlogR, main = "1")
```

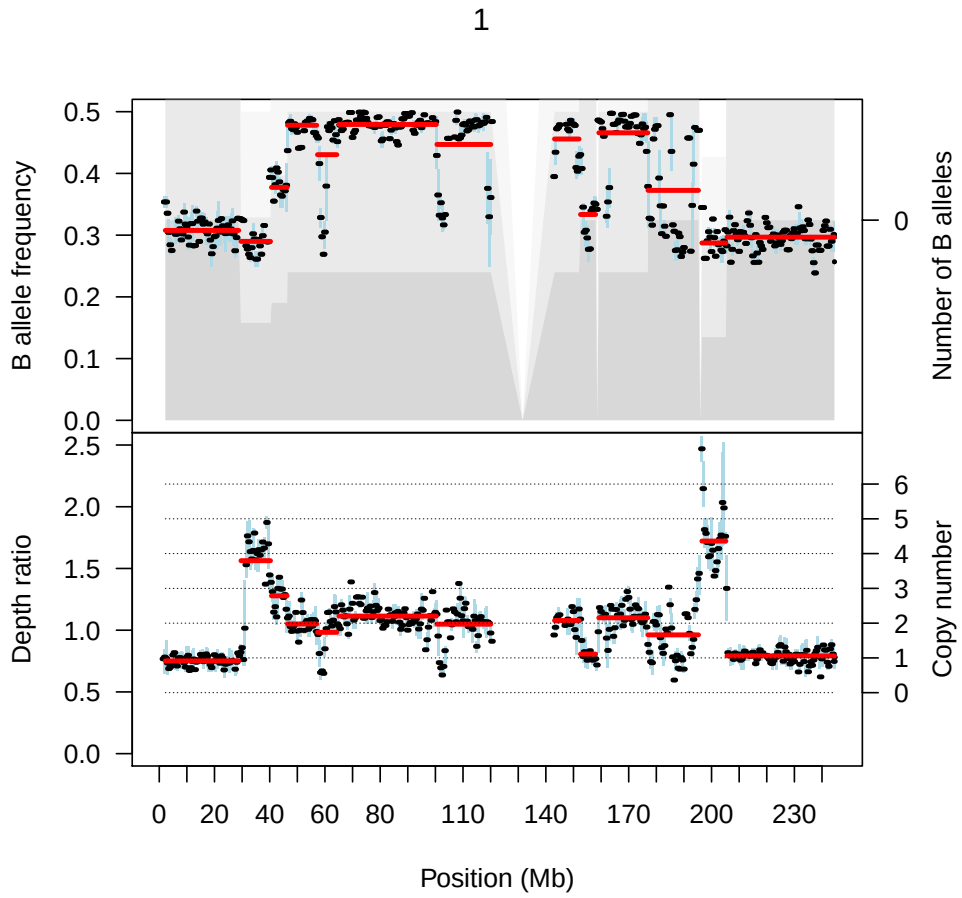


Figure 4: Plots B-allele frequencies (top) and un-logged-logR (bottom) with SNP array data. Chromosome 16. Horizontal dotted line indicate different copy number/ allelic state.

```
> chromosome.view(baf.windows = snp.b.win[[16]],
+                 ratio.windows = snp.r.win[[16]],  min.N.ratio = 1,
+                 segments = segmented.snp[segmented.snp$chromosome == "16", ],
+                 cellularity = cellularity, dna.index = dna.index,
+                 avg.depth.ratio = avg.unlogR, main = "16")
```

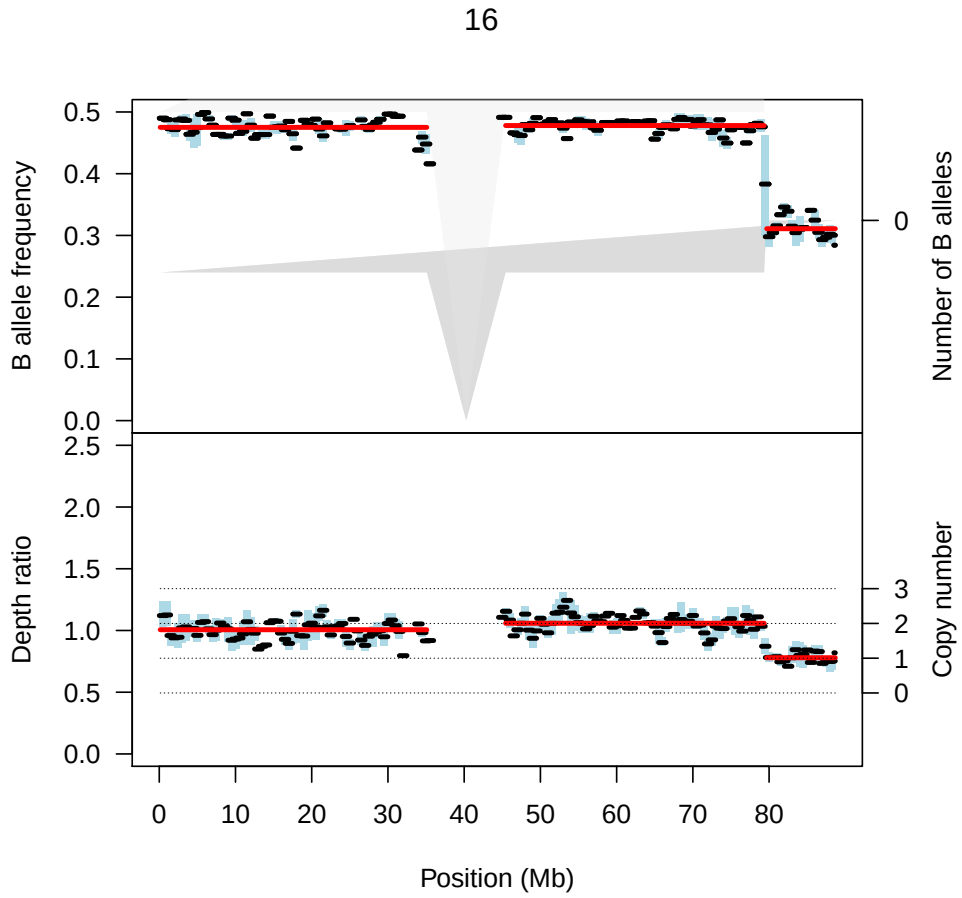


Figure 5: Plots B-allele frequencies (top) and un-logged-logR (bottom) with SNP array data. Chromosome 16. Horizontal dotted line indicate different copy number/ allelic state.