# HowTo: Creating HMM objects from BeadStudio-processed Illumina arrays

Robert B. Scharpf

October 22, 2008

This vignette describes how to create an instance of an `RatioSnpSet` from Illumina data.

## 1 Reading in the data

To illustrate, an example of BeadStudio output obtained from the Pevsner website (`http://pevsnerlab.kennedykrieger.org/SNPtrio03.htm`) is included with this package. The Pevsner laboratory provide the following instructions for saving the data using Illumina's BeadStudio in the appropriate format:

1. select the "Full Data Table" tab
2. click on the "Column Chooser" icon
3. in the "Displayed Columns" area, keep "Name", "Chr" and "Position", hide the rest
4. the "Displayed Subcolumns" area, keep "GType" and "Log R Ratio", hide the rest
5. click on "Export Displayed Data to File" icon; finally, save the file

A subset of 1000 SNPs is included with this package and can be loaded by

```
> library(VanillaICE)
> pathToIlluminaData <- system.file("illumina", package = "VanillaICE")
> illuminaEx <- read.table(paste(pathToIlluminaData,
+     "/illuminaEx.txt", sep = ""), sep = "\t", as.is = TRUE)
```

The following code converts this data.frame to an object of class `oligoSnpSet`:

```
> gt <- illuminaEx[, "S1135.GType", drop = FALSE]
> gt[gt == "AA"] <- 1
> gt[gt == "BB"] <- 3
> gt[gt == "AB"] <- 2
> gt[gt == "NC"] <- 4
> gt <- as.matrix(as.integer(gt[[1]]))
> logR <- as.matrix(as.numeric(illuminaEx[, "S1135.Log.R.Ratio"]))
> colnames(gt) <- colnames(logR) <- "S1135"
> rownames(logR) <- rownames(gt) <- illuminaEx[, "Name"]
> fd <- new("AnnotatedDataFrame", data = data.frame(position = illuminaEx[,
+     "Position"], chromosome = integer2chromosome(illuminaEx[,
+     "Chr"]), stringsAsFactors = FALSE), varMetadata = data.frame(labelDescription = c("posi
+     "chromosome")))
> featureNames(fd) <- illuminaEx[, "Name"]
> snpset <- new("oligoSnpSet", copyNumber = logR, calls = gt,
+     phenoData = annotatedDataFrameFrom(logR, byrow = FALSE),
+     featureData = fd, annotation = "Illumina550k")
> chrom <- chromosome2integer(fd$chromosome)
> snpset <- snpset[order(chrom, fd$position), ]
> stopifnot(validObject(snpset))
```

We can now use methods from the R package *SNPchip* to plot the data:

```
> gp <- plotSnp(snpset)
> gp$cex <- 3
> gp$ylab <- "copy number ratio"
> gp$abline <- TRUE
> gp$abline.h <- c(0.5, 1, 3/2)
> gp$abline.col <- "grey20"
> gp$abline.lty <- c(2, 1, 2)

> print(gp)
```

## Session Information

The version number of R and packages loaded for generating the vignette were:

- R version 2.8.0 (2008-10-20), `x86_64-unknown-linux-gnu`
- Locale: `LC_CTYPE=en_US;LC_NUMERIC=C;LC_TIME=en_US;LC_COLLATE=en_US;LC_MONETARY=C;LC_MESSAG`
- Base packages: base, datasets, graphics, grDevices, methods, stats, tools, utils
- Other packages: Biobase 2.2.0, oligoClasses 1.4.0, SNPchip 1.6.0, VanillaICE 1.4.0