Copy number estimation and genotype calling with crlmm

Rob Scharpf

Oct, 2009

```
R> library(crlmm)
R> crlmm:::validCdfNames()

[1] "genomewidesnp6"  "genomewidesnp5"  "human370v1c"
[4] "human370quadv3c" "human550v3b"     "human650v3a"
[7] "human610quadv1b" "human660quadv1a" "human1mduov3b"

R> outdir <- "/thumper/ctsa/snpmicroarray/rs/data/hapmap/illumina/HumanCNV370-Duo"
R> datadir <- "/thumper/ctsa/snpmicroarray/illumina/IDATS/370k"
R> cdfName <- "human370v1c"

R> samplesheet = read.csv(file.path(datadir, "HumanHap370Duo_Sample_Map.csv"),
     header = TRUE, as.is = TRUE)
R> samplesheet <- samplesheet[-c(28:46, 61:75, 78:79),
     ]
R> arrayNames <- file.path(datadir, unique(samplesheet[,
     "SentrixPosition"]))
R> grnfiles = all(file.exists(paste(arrayNames, ".Grn.dat",
     sep = "")))
R> redfiles = all(file.exists(paste(arrayNames, ".Red.dat",
     sep = "")))
```

Alternatively, arguments to the readIdatFiles can be passed through the `...` argument of the R function `crlmmWrapper`.

```
R> crlmmWrapper(sampleSheet = samplesheet, arrayNames = arrayNames,
     arrayInfoColNames = list(barcode = NULL, position = "SentrixPosition"),
     saveDate = TRUE, cdfName = cdfName, load.it = TRUE,
     save.it = TRUE, intensityFile = file.path(outdir,
         "normalizedIntensities.rda"), crlmmFile = file.path(outdir,
         "snpsetObject.rda"), rgFile = file.path(outdir,
         "rgFile.rda"))

[1] "load.it is TRUE and 'crlmmSetList_<CHR>.rda' objects found. Nothing to do..."
```

This creates a `crlmmSetList` object in the `outdir` directory. The first element of this object contains the quantile-normalized A and B intensities. The second element in the list contains the crlmm genotype calls.

```
R> CHR <- 1
R> filename <- paste(outdir, "/crlmmSetList_", CHR,
     ".rda", sep = "")
R> load(filename)
R> hist(crlmmSetList[[1]]$SNR)
```

Run update on the CrlmmSetList object to obtain copy number estimates. Estimate copy number for chromosome 1.

```
R> update(filename)
```

```
Processing  /thumper/ctsa/snpmicroarray/rs/data/hapmap/illumina/HumanCNV370-Duo/crlmmSetList
```

Samples with low signal to noise ratios tend to have a lot of variation in the point estimates of copy number. One may want to exclude these samples, or smooth after filtering outliers. Here we load the crlmmSetList object. See the copynumber.Rnw vignette for example plots.

```
R> load(filename)
R> cn <- copyNumber(crlmmSetList)
```

1 Session information

```
R> toLatex(sessionInfo())
```

- R version 2.10.0 (2009-10-26), `x86_64-unknown-linux-gnu`
- Locale: `LC_CTYPE=en_US.iso885915`, `LC_NUMERIC=C`, `LC_TIME=en_US.iso885915`, `LC_COLLATE=en_US.iso885915`, `LC_MONETARY=C`, `LC_MESSAGES=en_US.iso885915`, `LC_PAPER=en_US.iso885915`, `LC_NAME=C`, `LC_ADDRESS=C`, `LC_TELEPHONE=C`, `LC_MEASUREMENT=en_US.iso885915`, `LC_IDENTIFICATION=C`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: Biobase 2.6.0, crlmm 1.4.5, human370v1cCrlmm 1.0.0
- Loaded via a namespace (and not attached): affyio 1.14.0, annotate 1.24.0, AnnotationDbi 1.8.1, Biostrings 2.14.8, DBI 0.2-4, ellipse 0.3-5, genefilter 1.28.1, IRanges 1.4.8, mvtnorm 0.9-8, oligoClasses 1.8.0, preprocessCore 1.8.0, RSQLite 0.7-3, SNPchip 1.10.0, splines 2.10.0, survival 2.35-7, xtable 1.5-6