Genotyping with the crlmm Package

Benilton Carvalho

January, 2009

1 Quick intro to crlmm

The crlmm package contains a new implementation for the CRLMM algorithm (Carvalho et. al. 2007). Our focus is on efficient genotyping of SNP 5.0 and 6.0 Affymetrix arrays, although extensions of the method are under development for similar platforms.

This implementation, compared to the previous one (in oligo), offers improved confidence scores, quality scores for SNP's and batches, higher accuracy on different datasets and better performance.

Additionally, this package does not use the pd.genomewidesnp packages created via pdInfoBuilder for oligo. Instead, it uses different annotation packages (genomewidesnp.5 and genomewidesnp.6), which use simple R objects to store only the information needed for genotyping. This allowed us to improve the speed of the method, as SQL queries are no longer performed here.

It is also our priority to make the package simple to use. Below we demonstrate how to get genotype calls with the 'new' CRLMM. We use 3 samples on SNP 5.0 made available via the hapmapsnp5 package.

```
R> library(crlmm)
R> library(hapmapsnp5)
R> path <- system.file("celFiles", package = "hapmapsnp5")
R> celFiles <- list.celfiles(path, full.names = TRUE)
R> system.time(crlmmResult <- crlmm(celFiles, verbose = FALSE))

   user  system elapsed
141.273   3.549 145.215
```

The `crlmmResult` is a list with the following components:

- `calls`: genotype calls (1 - AA; 2 - AB; 3 - BB);
- `confs`: confidence scores, which can be translated to probabilities by using:

$$1 - 2^{-}(\text{confs}/1000),$$

  although we prefer this representation as it saves a significant amount of memory;

- `SNPQC`: SNP quality score;
- `batchQC`: Batch quality score;
- `SNR`: Signal-to-noise ratio.

```
R> crlmmResult[["calls"]][1:10, ]
```

```
             NA06985_GW5_C.CEL NA06991_GW5_C.CEL
SNP_A-1780520                 3                 3
SNP_A-1780618                 3                 2
SNP_A-1780632                 3                 3
SNP_A-1780654                 1                 1
SNP_A-4192495                 3                 3
SNP_A-4192498                 3                 3
SNP_A-1780732                 3                 3
SNP_A-1780848                 2                 2
SNP_A-1780985                 3                 3
SNP_A-1781022                 1                 1
             NA06993_GW5_C.CEL
SNP_A-1780520                 3
SNP_A-1780618                 2
SNP_A-1780632                 3
SNP_A-1780654                 1
SNP_A-4192495                 2
SNP_A-4192498                 2
SNP_A-1780732                 3
SNP_A-1780848                 3
SNP_A-1780985                 2
SNP_A-1781022                 1
```

```
R> crlmmResult[["confs"]][1:10, ]
```

```
             NA06985_GW5_C.CEL NA06991_GW5_C.CEL
SNP_A-1780520             15571             15841
SNP_A-1780618             10684              4501
SNP_A-1780632             16146             16097
SNP_A-1780654             20078             18479
SNP_A-4192495             16026             16760
SNP_A-4192498             14773             18508
SNP_A-1780732             15220             15549
SNP_A-1780848             10617             11545
SNP_A-1780985             13088             14968
SNP_A-1781022              7547              7525
             NA06993_GW5_C.CEL
SNP_A-1780520             14251
SNP_A-1780618              4443
```

```
SNP_A-1780632             11703
SNP_A-1780654             18577
SNP_A-4192495             10371
SNP_A-4192498             12506
SNP_A-1780732             13142
SNP_A-1780848             13520
SNP_A-1780985              9697
SNP_A-1781022              7631
```

2 Details

This document was written using:

```
R> sessionInfo()

R version 2.8.0 (2008-10-20)
x86_64-unknown-linux-gnu

locale:
LC_CTYPE=en_US.UTF-8;LC_NUMERIC=C;LC_TIME=en_US.UTF-8;LC_COLLATE=en_US.UTF-8;LC_MONETARY=C;L

attached base packages:
[1] tools     stats     graphics  grDevices utils
[6] datasets  methods   base

other attached packages:
[1] genomewidesnp5Crlmm_1.0 hapmapsnp5_1.3
[3] crlmm_1.27              preprocessCore_1.4.0
[5] affyio_1.10.1
```