

# 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[,
  "SentryPosition"]))
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 = "SentryPosition"),
  saveDate = TRUE, cdfName = cdfName, load.it = FALSE,
  save.it = FALSE, intensityFile = file.path(outdir,
  "normalizedIntensities.rda"), crlmmFile = file.path(outdir,
  "snpsetObject.rda"), rgFile = file.path(outdir,
  "rgFile.rda"))

reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585367_A      _Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585376_B      _Grn.idat
```

reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585413_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585415_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585422_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585483_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585506_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585520_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585575_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585596_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585597_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186100_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186125_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186132_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186216_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186219_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186263_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186347_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186513_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186572_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058011_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058059_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058082_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058102_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058127_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058132_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058211_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585498_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585508_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4019585512_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186109_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186167_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186197_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186254_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186332_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186339_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186396_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058010_B	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058210_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058213_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4031058239_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186415_A	_Grn.idat
reading /thumper/ctsa/snpmicroarray/illumina/IDATS/370k/4030186434_B	_Grn.idat

====	8%
=====	17%

=====		25%
=====		33%
=====		42%
=====		50%
=====		58%
=====		67%
=====		75%
=====		83%
=====		92%
=====		100%
=		2%
==		5%
===		7%
====		9%
=====		12%
=====		14%
=====		16%
=====		19%
=====		21%
=====		23%
=====		26%
=====		28%

=====	30%
=====	33%
=====	35%
=====	37%
=====	40%
=====	42%
=====	44%
=====	47%
=====	49%
=====	51%
=====	53%
=====	56%
=====	58%
=====	60%
=====	63%
=====	65%
=====	67%
=====	70%
=====	72%
=====	74%
=====	77%
=====	79%
=====	81%

```

|=====| 84%
|
|=====| 86%
|
|=====| 88%
|
|=====| 91%
|
|=====| 93%
|
|=====| 95%
|
|=====| 98%
|
|=====| 100%
Calling 346451 SNPs for recalibration... Filling out empty centers.....
Calling 346451 SNPs... Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12
Chromosome 13
Chromosome 14
Chromosome 15
Chromosome 16
Chromosome 17
Chromosome 18
Chromosome 19
Chromosome 20
Chromosome 21
Chromosome 22
Chromosome 23
Chromosome 24
NULL

```

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
-----
-           Estimating copy number for chromosome 1
-----
....
```

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)
```