Package: RLMM (via r-universe)

November 21, 2026

Version 1.74.0

Date 2005-09-02

Title A Genotype Calling Algorithm for Affymetrix SNP Arrays

Author Nusrat Rabbee <nrabbee@post.harvard.edu>, Gary Wong <wongg62@berkeley.edu>

Maintainer Nusrat Rabbee <nrabbee@post.harvard.edu>

Description A classification algorithm, based on a multi-chip, multi-SNP approach for Affymetrix SNP arrays. Using a large training sample where the genotype labels are known, this aglorithm will obtain more accurate classification results on new data. RLMM is based on a robust, linear model and uses the Mahalanobis distance for classification. The chip-to-chip non-biological variation is removed through normalization. This model-based algorithm captures the similarities across genotype groups and probes, as well as thousands other SNPs for accurate classification. NOTE: 100K-Xba only at for now.

License LGPL (>= 2)

Depends R (>= 2.1.0)

Imports graphics, grDevices, MASS, stats, utils

SystemRequirements Internal files Xba.CQV, Xba.regions (or other regions file)

URL http://www.stat.berkeley.edu/users/nrabbee/RLMM

biocViews Microarray, OneChannel, SNP, GeneticVariability

Repository Bioconductor 3.23

Date/Publication 2026-04-28 12:31:25 UTC

RemoteUrl https://github.com/bioc/RLMM

RemoteRef RELEASE_3_23

RemoteSha e65c627bd302e89232111d74283069985ac2a463

Contents

Classify 2
create_Thetafile 3
normalize_Rawfiles 3
plot_theta 4

Index **5**

`Classify` *Classification of SNPs based on theta estimates*

Description

This function entails classification of SNPs based on the theta estimates (thetafile), genotype information (A regions file), and some internal files. Currently, this algorithm works for the Affymetrix 100K - Xba dataset.

Usage

```
Classify(genotypefile = "",
         regionsfile = "",
  thetafile = "",
  callrate = 100)
```

Arguments

`genotypefile`	Name of the classified SNPs with the genotypes (required)
`regionsfile`	Character string specifying the directory AND name of regionsfile - e.g., "Xba.regions" (required)
`thetafile`	Character string specifying the directory AND name of thetafile (required)
`callrate`	Call Rate percentage; The user can specify any number from the list: 80,82,84,86,88,90,92,94,96,98,100. Default is 100%(optional)

Details

For each SNP, Mahalanobis distances from each chip's (theta A, theta B) ordered pair to the genotype centers is calculated. Each chip is assigned the genotype of the cluster which it is closest to (ie: AA, AB, BB).

Author(s)

Nusrat Rabbee <nrabbee@post.harvard.edu>, Gary Wong <wongg62@gmail.com>

create_Thetafile *Calculating Parameter Estimates*

Description

Assuming that the *.norm files are created, this step of the data analysis will calculate estimates of theta A and theta B values for each SNP and chip based on normalized probe intensity data from the *.norm files. The theta values are produced from fitting a probe-level additive model to the log2 A probe intensities and the B intensities separately.

Usage

```
create_Thetafile(probefiledir = getwd(),
                 start = 1,
                 end = -1,
                 thetafile = "")
```

Arguments

probefiledir	Character string specifying the directory with the *.norm files (optional)
start	An integer value specifying which SNP number we should start at when calculating the theta values (optional)
end	An integer value specifying which SNP number we should stop at when calculating the theta values (optional)
thetafile	A character string specifying the name the theta file will be saved as (optional)

Author(s)

Nusrat Rabbee <nrabbee@post.harvard.edu>, Gary Wong <wongg62@gmail.com>

normalize_Rawfiles *Normalize PM Intensity values*

Description

Given a directory with *.raw files, it will normalize the PMA and PMB intensities in each file using Xba.CQV (composite quantile vector) and return the normalized values written to *.norm files corresponding to its *.raw files. EG: If two *.raw files are used, two *.norm files will be returned. This normalization simply puts the probe data on the same scale as the training data.

Usage

```
normalize_Rawfiles(cqvfile = "",
                   probefiledir = getwd())
```

Arguments

cqvfile	Character string specifying the CQV filename (e.g., Xba.CQV) (required)
probefiledir	Character string specifying location of the *.raw files and *.norm files (optional)

Author(s)

Nusrat Rabbee <nrabbee@post.harvard.edu>, Gary Wong <wongg62@gmail.com>

---

plot_theta *Allele Summary Plot*

Description

Creates an Allele Summary plot (allele B vs. allele A) for each SNP specified in snpsfilename. The points in the plot are the (theta A, theta B) ordered pairs for all the samples of the SNP. If a plotfilename is specified, it will save the plot as a .ps file, otherwise the plot is shown on screen.

Usage

```
plot_theta(genotypefile = "Xba.rlmm",
    thetafile = "Xba.theta",
    Pick.Obj = "FALSE",
    plotfile = "plots.ps",
    snpsfile = "snps.lst")
```

Arguments

genotypefile	Character string specifying the directory AND name of the .rlmm file created by `Classify` (optional)
thetafile	Character string specifying the directory AND name of the .rlmm file created by `Create_Thetafile` (optional)
Pick.Obj	At this point, it should always be left as the default FALSE, ie: it is for development purposes only (optional)
plotfile	The name where to store the plot as a .ps file, if blank such as "", it will display on screen instead (optional)
snpsfile	A list of SNPs to plot, with one SNP following another on a newline (optional)

Author(s)

Nusrat Rabbee <nrabbee@post.harvard.edu>, Gary Wong <wongg62@gmail.com>

Index

∗ **classif**
    Classify, 2
∗ **hplot**
    plot_theta, 4
∗ **manip**
    normalize_Rawfiles, 3
∗ **methods**
    create_Thetafile, 3

Classify, 2
create_Thetafile, 3

normalize_Rawfiles, 3

plot_theta, 4