Overview of binQTL

Wen Yao, Guangwei Li, Yanru Cui, Yiming Yu, Qifa Zhang and Shizhong Xu

2019-05-14

Contents

1 Introduction

Composite interval mapping (CIM) is the dominating algorithm used in QTL analysis of various phenotypes in experimental populations. With the development of next-generation sequening, genotyping of SNPs became much easier, which leads to the development of binmap. In binmap, the distance between adjacent markers were zero. As a result, the concept of "interval mapping" doesn't make sense in binmap. The **binQTL** package is designed to perform QTL analysis with binmap. In this vignette, we will rely on three simple, illustrative example datasets to explain the usage of **binQTL**.

To use the **binQTL** package, we need to load it into R first.

```
library(binQTL)
```

2 Use of binQTL with example dataset of 210 RILs

The phenotype in the first example dataset is the 1000-grain weight of 210 recombinant inbred lines (RIL) in the year of 1999. The genetic map of the 210 RILs is a binmap composed of 1619 bins.

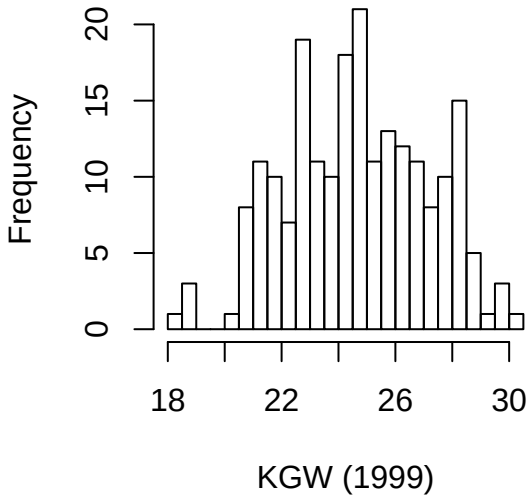


Figure 1: KGW of 210 RILs

2.1 Read in the phenotype file of 210 RILs

The phenotype file for the **binQTL** package should be a text file containing two columns. The first column is the identifier of all accessions while the second column contains the phenotypic value of specific trait for each accession. The column names of this file can be any characters specified by the User.

```
ril.phe <- read.csv(system.file("examples/ril.phe.csv", package="binQTL"), as.is=T)
dim(ril.phe)
```

```
## [1] 210   2
```

```
head(ril.phe)
```

```
##    ril  kgw
## 1 R001 24.4
## 2 R002 24.6
## 3 R003 24.5
## 4 R004 22.8
## 5 R005 25.2
## 6 R006 21.4
```

```
par(mar = c(5, 4, 1, 1))
hist(ril.phe[, 2], breaks=40, xlab="KGW (1999)", main="")
```

2.2 Read in the genotype file of 210 RILs

The genotype file for the **binQTL** package should be a text file containing multiple columns and rows. Each row of this file is a bin. The first four columns contain the information of each bin including the identifier, the chromosome, the start coordinate and the end coordinate. The names of the first four columns should be "Bin", "Chr", "Start" and "Stop". Each of the rest columns gives the genotype of each accession at different bins. The names of the rest columns can be any characters specified by the User.

```
ril.geno <- read.csv(system.file("examples/ril.geno.csv", package="binQTL"), as.is=T)
dim(ril.geno)
```

```
## [1] 1619  214
```

```
ril.geno[1:2, 1:9]
```

```
##    Bin   Chr Start  Stop R001 R002 R003 R004 R005
## 1 Bin1 chr01 0.000 0.565   P1   P2   P2   P2   P1
## 2 Bin2 chr01 0.565 0.599   P1   P2   P2   P2   P1
```

```
## Not run
# par(mar = c(3, 4, 1, 1))
# plotBinmap(ril.geno, xlab="Genomic position", ylab="RILs", cex.axis=0.6)
## Not run
```

2.3 QTL mapping of RILs using ANOVA

```
qtl.res.1 <- aovQTL(phenotype = ril.phe, genotype = ril.geno)
head(qtl.res.1)
```

```
##    Bin   Chr Start  Stop         p     -logp
## 1 Bin1 chr01 0.000 0.565 0.2563842 0.5911088
## 2 Bin2 chr01 0.565 0.599 0.2143561 0.6688641
## 3 Bin3 chr01 0.599 0.922 0.2102838 0.6771941
## 4 Bin4 chr01 0.922 1.075 0.1410645 0.8505823
## 5 Bin5 chr01 1.075 1.147 0.1202287 0.9199918
## 6 Bin6 chr01 1.147 1.221 0.1240959 0.9062425
```

2.4 QTL mapping of RILs using binQTL

```
qtl.res.2 <- binQTLScan(phenotype = ril.phe, genotype = ril.geno, population="RIL")
head(qtl.res.2)
```

```
##    Bin   Chr Start  Stop p -logp
## 1 Bin1 chr01 0.000 0.565 1     0
## 2 Bin2 chr01 0.565 0.599 1     0
## 3 Bin3 chr01 0.599 0.922 1     0
## 4 Bin4 chr01 0.922 1.075 1     0
## 5 Bin5 chr01 1.075 1.147 1     0
## 6 Bin6 chr01 1.147 1.221 1     0
```

2.5 Comparison of different algotrithms

Let's get a glimpse of the QTL mapping results of different algorithms.

```
par(mar = c(3, 4, 2, 1))
par(mfrow=c(2, 1))
plotQTL(qtl.res.1[, c(1:4, 6)], main="ANOVA", cex.axis=0.6, ylab=expression(-log[10](p)))
plotQTL(qtl.res.2[, c(1:4, 6)], main="binQTL", cex.axis=0.6, ylab=expression(-log[10](p)))
```

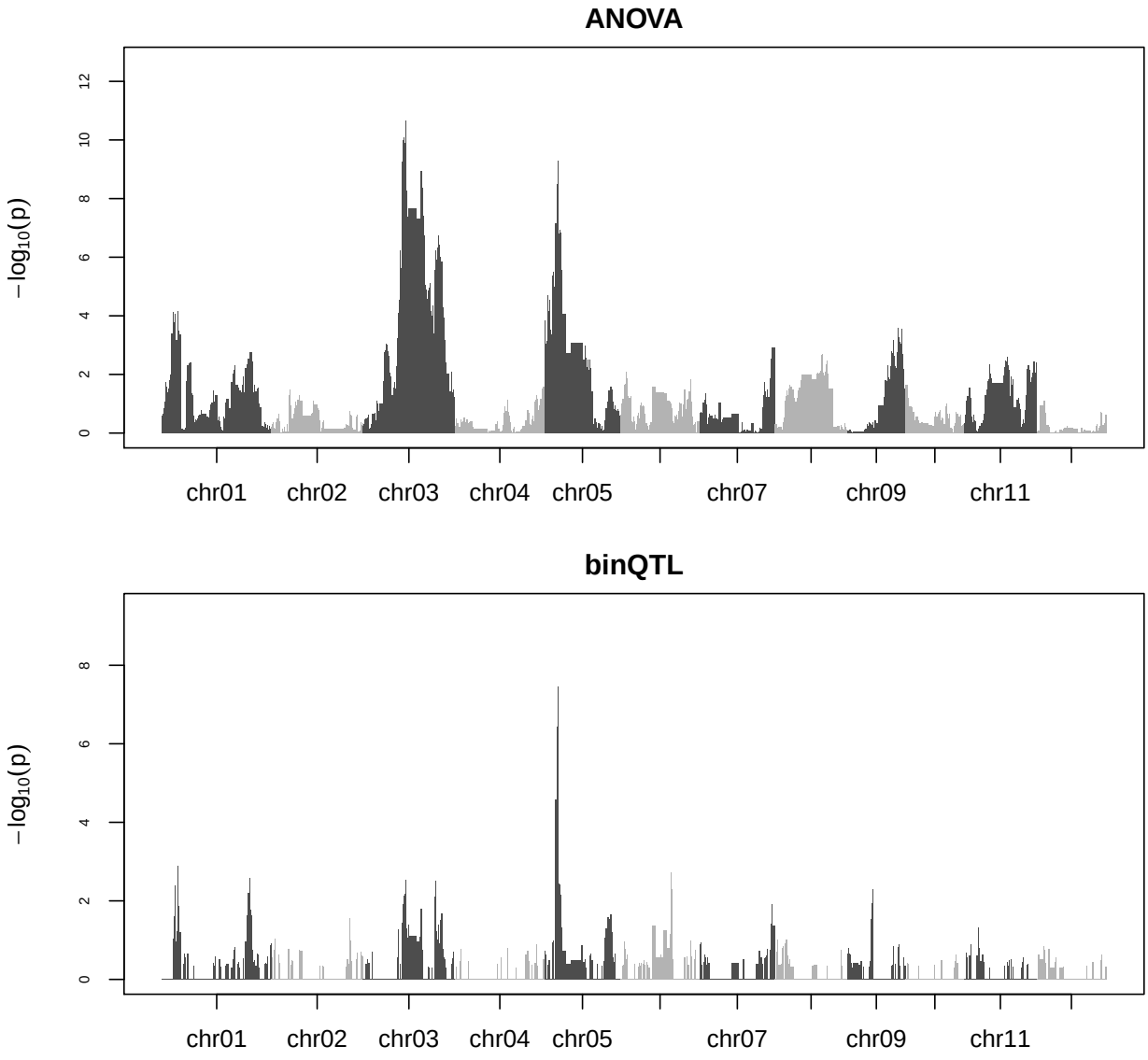


Figure 2: QTL mapping of 210 RILs using different algorithms.

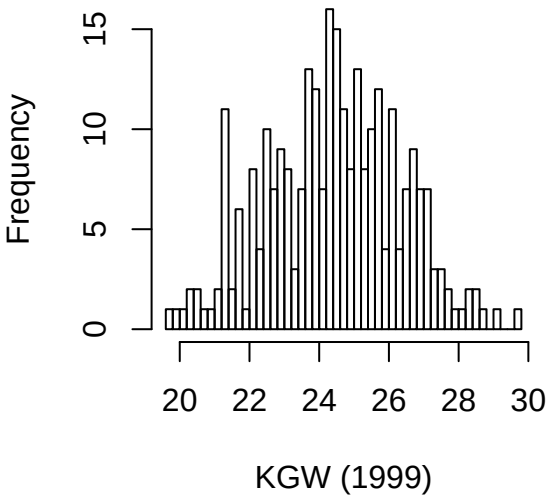


Figure 3: KGW of 278 IMF2s

3 Use of binQTL with example dataset of 278 IMF2s

The phenotype in the second example dataset is the 1000-grain weight of 278 immortalized F2 (IMF2) in the year of 1999. The genetic map of the 278 IMF2s is a binmap composed of 1619 bins.

3.1 Read in the phenotype file of 278 IMF2s

```
imf2.phe <- read.csv(system.file("examples/imf2.phe.csv", package="binQTL"), as.is=T)
dim(imf2.phe)
```

```
## [1] 278   2
```

```
head(imf2.phe)
```

```
##   cross  kgw99
## 1  F001 25.334
## 2  F002 22.432
## 3  F003 26.411
## 4  F005 25.029
## 5  F006 23.573
## 6  F008 26.463
```

```
par(mar = c(5, 4, 1, 1))
hist(imf2.phe[, 2], breaks=40, xlab="KGW (1999)", main="")
```

3.2 Read in the genotype file of 278 IMF2s

```
imf2.geno <- read.csv(system.file("examples/imf2.geno.csv", package="binQTL"), as.is=T)
dim(imf2.geno)
```

```
## [1] 1619  282
```

```
imf2.geno[1:2, 1:9]
```

```
##    Bin   Chr Start  Stop F001 F002 F003 F005 F006
## 1 Bin1 chr01 0.000 0.565   P1   P2   P2    H   P2
## 2 Bin2 chr01 0.565 0.599   P1   P2   P2    H   P2
```

```
## Not run
# par(mar = c(3, 4, 1, 1))
# plotBinmap(imf2.geno, xlab="Genomic position", ylab="IMF2s", cex.axis=0.6)
## Not run
```

3.3 QTL mapping of IMF2s using ANOVA

```
qtl.res.3 <- aovQTL(phenotype = imf2.phe, genotype = imf2.geno)
head(qtl.res.3)
```

```
##    Bin   Chr Start  Stop           p    -logp
## 1 Bin1 chr01 0.000 0.565 0.011152487 1.952628
## 2 Bin2 chr01 0.565 0.599 0.007299795 2.136689
## 3 Bin3 chr01 0.599 0.922 0.008330228 2.079343
## 4 Bin4 chr01 0.922 1.075 0.010369449 1.984244
## 5 Bin5 chr01 1.075 1.147 0.004231273 2.373529
## 6 Bin6 chr01 1.147 1.221 0.004798506 2.318894
```

3.4 QTL mapping of IMF2s using binQTL

```
qtl.res.4 <- binQTLScan(phenotype = imf2.phe, genotype = imf2.geno, population="F2")
head(qtl.res.4)
```

```
##    Bin   Chr Start  Stop p -logp
## 1 Bin1 chr01 0.000 0.565 1     0
## 2 Bin2 chr01 0.565 0.599 1     0
## 3 Bin3 chr01 0.599 0.922 1     0
## 4 Bin4 chr01 0.922 1.075 1     0
## 5 Bin5 chr01 1.075 1.147 1     0
## 6 Bin6 chr01 1.147 1.221 1     0
```

3.5 Comparison of different algotrithms

Let's get a glimpse of the QTL mapping results of different algorithms.

```
par(mar = c(3, 4, 2, 1))
par(mfrow=c(2, 1))
plotQTL(qtl.res.3[, c(1:4, 6)], main="ANOVA", cex.axis=0.6, ylab=expression(-log[10](p)))
plotQTL(qtl.res.4[, c(1:4, 6)], main="binQTL", cex.axis=0.6, ylab=expression(-log[10](p)))
```

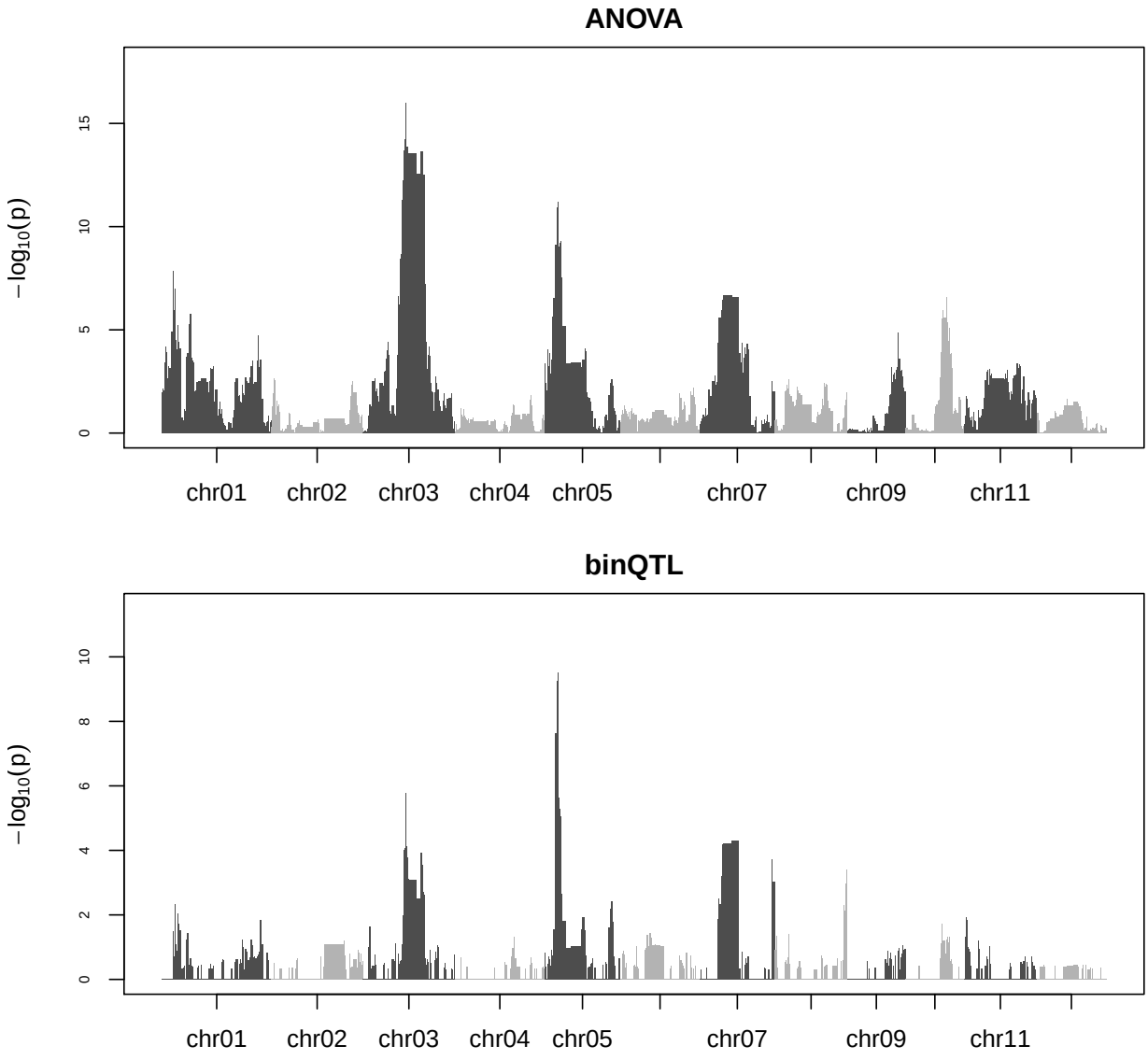


Figure 4: QTL mapping of 278 IMF2s using different algorithms.

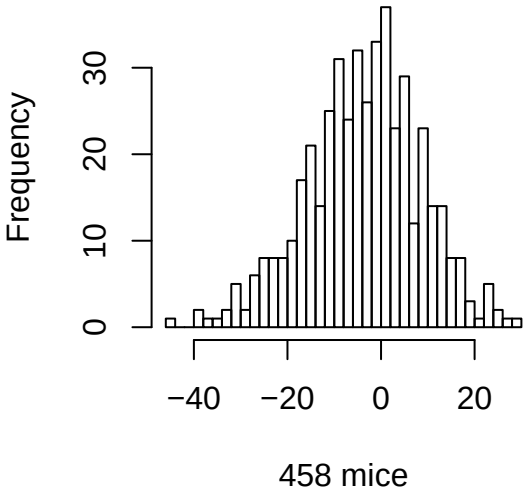


Figure 5: MAGIC population phenotype

4 Use of binQTL with example dataset of a mouse MAGIC population

The phenotype in the third example dataset is a simulated data value of 458 mice constituting a MAGIC population with 8 parents. The genetic map of the 458 mice is a binmap composed of 6683 bins. Each bin is represented by 8 sequential lines in the genetic map. The genotype code for a heterozygote individual carrying the genome material from the 3rd and the 4th parents should be [0 0 1 1 0 0 0 0]. The genotype code for a homozygote individual carrying both copies of the genome from the first parent should be [2 0 0 0 0 0 0 0].

4.1 Read in the phenotype file of a mouse MAGIC population

```
magic.phe <- read.csv(system.file("examples/magic.phe.csv", package="binQTL"), as.is=T)
dim(magic.phe)
```

```
## [1] 458   2
```

```
head(magic.phe)
```

```
##   magicID       phe
## 1      z1  -1.85704
## 2      z2 -26.41925
## 3      z3  -4.34248
## 4      z4   4.16259
## 5      z5   5.24857
## 6      z6   3.99186
```

```
par(mar = c(5, 4, 1, 1))
hist(magic.phe[, 2], breaks=40, xlab="458 mice", main="")
```

4.2 Read in the genotype file of a mouse MAGIC population

```
load(system.file("examples/magic.geno.RData", package="binQTL"))
dim(magic.geno)
```

```
## [1] 53464   462
```

```
magic.geno[1:9, 1:9]
```

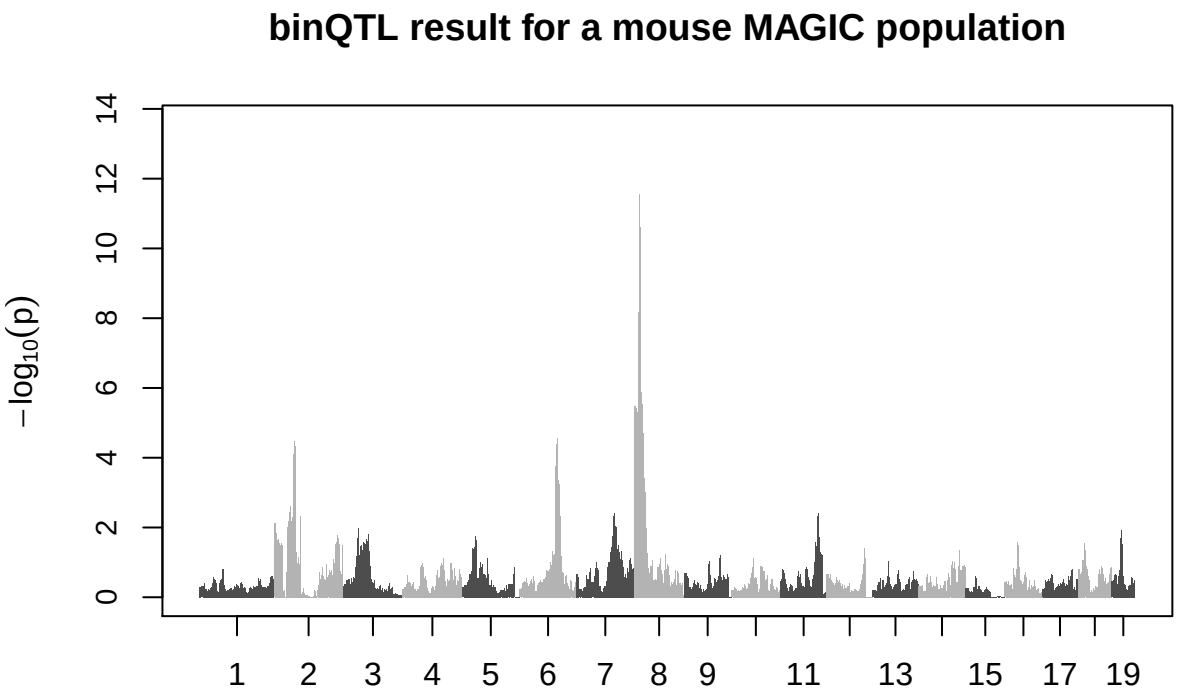
```
##    Bin Chr   Start    Stop z1 z2 z3 z4 z5
## 1 Bin1   1       1 3242877  0  2  2  0  0
## 2 Bin1   1       1 3242877  0  0  0  0  0
## 3 Bin1   1       1 3242877  0  0  0  0  2
## 4 Bin1   1       1 3242877  0  0  0  0  0
## 5 Bin1   1       1 3242877  2  0  0  2  0
## 6 Bin1   1       1 3242877  0  0  0  0  0
## 7 Bin1   1       1 3242877  0  0  0  0  0
## 8 Bin1   1       1 3242877  0  0  0  0  0
## 9 Bin2   1 3242878 6568106  0  2  2  0  0
```

4.3 QTL mapping of a mouse MAGIC population using binQTL

```
qtl.res.4 <- binQTLScan(phenotype = magic.phe, genotype = magic.geno, population="MAGIC")
head(qtl.res.4)
```

```
##     Bin Chr   Start    Stop         p     -logp
## 1  Bin1   1       1 3242877 0.5011497 0.3000325
## 9  Bin2   1 3242878 6568106 0.4907281 0.3091590
## 17 Bin3   1 6568107 8058994 0.4890375 0.3106579
## 25 Bin4   1 8058995 9035641 0.4918058 0.3082064
## 33 Bin5   1 9035642 9391058 0.4919881 0.3080454
## 41 Bin6   1 9391059 9654262 0.4906753 0.3092058
```

```
plotQTL(qtl.res.4[, -5], main="binQTL result for a mouse MAGIC population",
        cex.axis=0.9, ylab=expression(-log[10](p)))
```



5 Session Information

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

```
sessionInfo()
```

```
## R version 3.5.3 (2019-03-11)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.936
## [2] LC_CTYPE=Chinese (Simplified)_China.936
## [3] LC_MONETARY=Chinese (Simplified)_China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.936
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] binQTL_0.1.0  MASS_7.3-51.1
##
```

```
## loaded via a namespace (and not attached):
##  [1] compiler_3.5.3  magrittr_1.5    tools_3.5.3     htmltools_0.3.6
##  [5] yaml_2.2.0      Rcpp_1.0.0      stringi_1.2.4   rmarkdown_1.11
##  [9] highr_0.7       knitr_1.20      stringr_1.3.1   digest_0.6.18
## [13] evaluate_0.12
```