A practical example of tomato QTL mapping using a RIL population

R

http://www.r-project.org/

R/QTL

http://www.rqtl.org/

Dr. Hamid Ashrafi

UC Davis
Seed Biotechnology Center

Breeding with Molecular Markers

Feb 16-17, 2010

Tutorial Objectives

Where to download R and R/QTL?
Why use R?
How to load R/QTL library and settings?
How to read data file?
How to plot genotypic and phenotypic data?
How to perform simple interval mapping (SIM), composite interval mapping (CIM), and multiple interval mapping (MIM)

Notes:
R/QTL commands start with '<' and are given in blue
R/QTL output is given in black

The R Project for Statistical Computing - Windows Internet Explorer

http://www.r-project.org/

File Edit View Favorites Tools Help

Google R Search Sidewiki Bookmarks Check Translate AutoFill R Sign In

WEB SEARCH My Apps

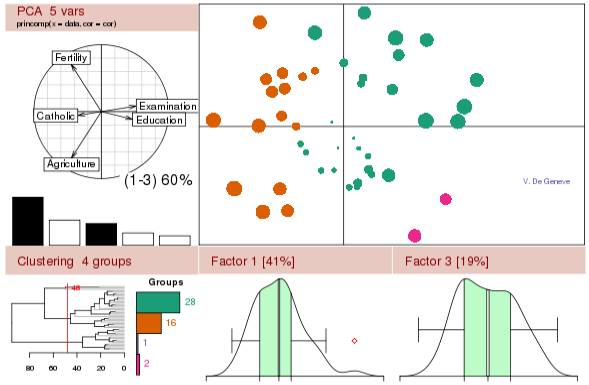
Favorites Suggested Sites Get More Add-ons Getting Started Customize Links

The R Project for... New Tab

Home Feeds (J) Read Mail Print Page Safety Tools Help

The R Project for Statistical Computing

Download package from here



About R
What is R?
Contributors
Screenshots
What's new?

Download, Packages
CRAN

R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search

Documentation
Manuals
FAQs
The R Journal
Wiki
Books
Certification
Other

Misc
Bioconductor
Related Projects
User Groups
Links

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News:

- **R version 2.10.1** has been released on 2009-12-14. The source code will first become available in this directory, and eventually via all of CRAN. Binaries will arrive in due course (see download instructions above).
- The first issue of The R Journal is now available
- **useR! 2010**, the R user conference, will be held at NIST, Gaithersburg, Maryland, USA, July 21-23, 2010.
- We have started to collect information about local UseR Groups in the R Wiki.

This server is hosted by the Institute for Statistics and Mathematics of the WU Wien.

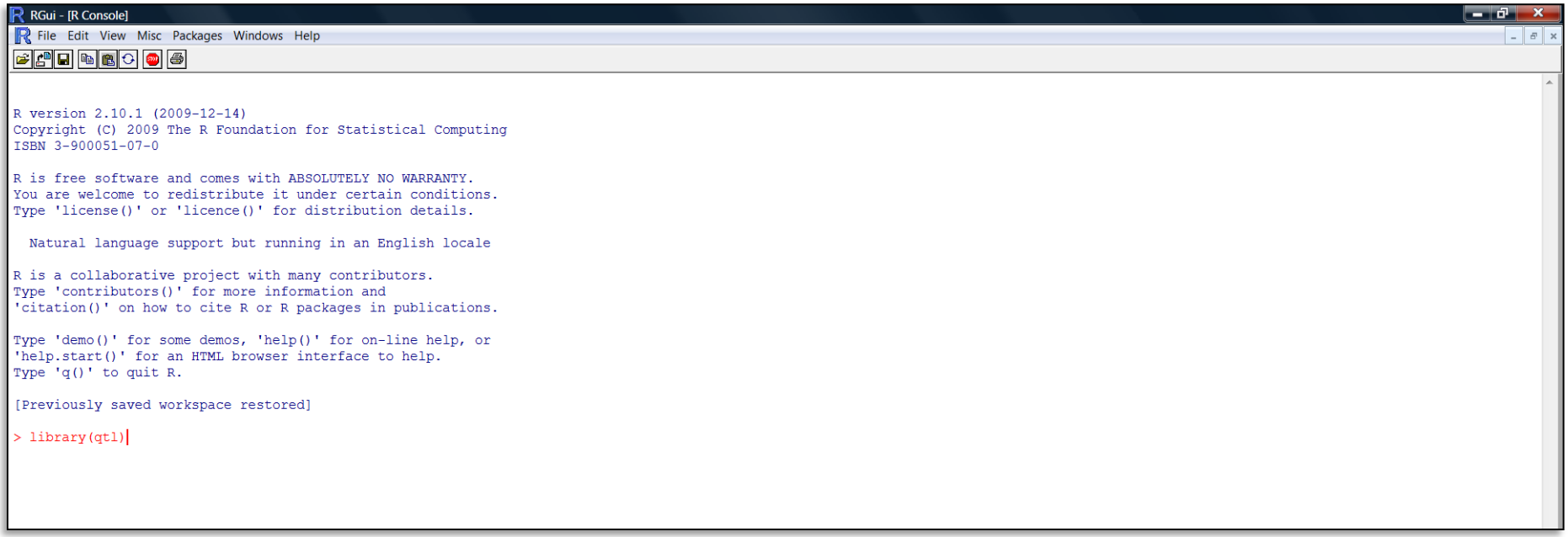
Internet | Protected Mode: On 100%

www.r-project.org

Why R?

It is free
It is written in an object oriented programming style
It is open source, therefore anybody can contribute to write a package
It can be run on Windows/MAC or Linux based operating systems
Can handle large number of markers, compatible with high throughput SNP discovery

A screen shot of R



Loading the R/qtl library, settings and reading the data file

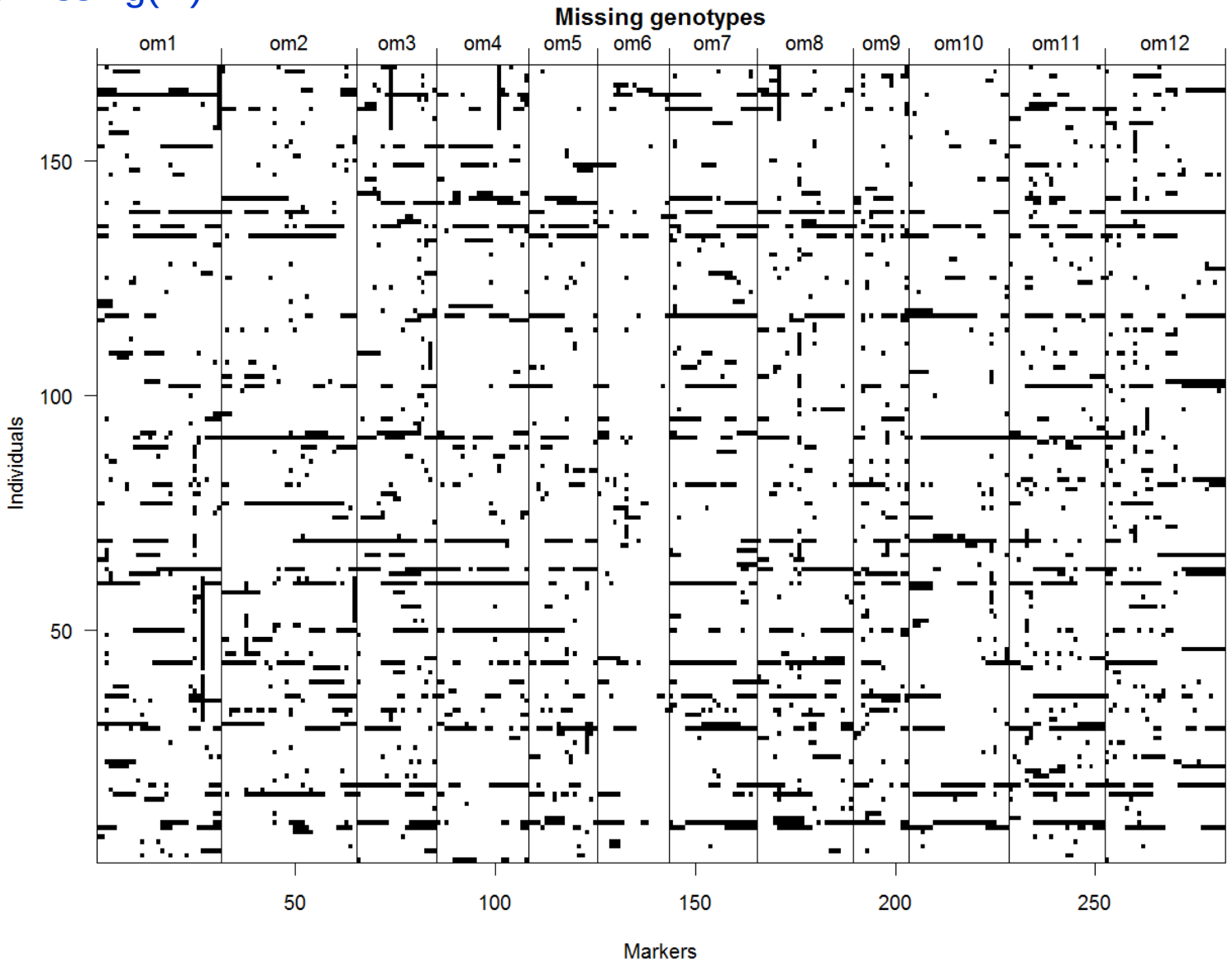
```
> library (qtl)

> setwd("C:/Users/Hamid/Documents/Breeding With Molecular Markers Feb
16_17_2010/R_QTL")

> getwd()
[1] "C:/Users/Hamid/Documents/Breeding With Molecular Markers Feb
16_17_2010/R_QTL“

> ril <- read.cross("mm", dir="", "F7.RAW", "F7.MAP")

> summary(ril)
      Length Class      Mode
geno   12    -none-     list
pheno 113    data.frame list
```

Exploring the data

```
> summary.cross(ril)
  Backcross
    No. individuals:    170

    No. phenotypes:     113
    Percent phenotyped: 90.6 90.6 90.6 89.4 89.4 89.4 89.4 90.6 89.4 90.6 80.6 80.6 80.6 80.6
                        80.6 80.6 80.6 80.6 80.6 80.6 80.6 80.6 80 80 80 80 80 80 80 80 80 80
                        85.3 85.3 85.3 85.3 85.3 85.3 88.2 88.2 88.2 88.2 88.2 88.2 88.2 85.3
                        85.3 92.4 92.4 92.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 99.4 75.9
                        75.9 75.9 75.9 75.9 75.9 75.9 75.9 75.9 89.4 88.8 90.6 89.4 88.8 90.6
                        89.4 88.2 90.6 92.4 99.4 99.4 99.4 70 70 70 70.6 70.6 70.6 70.6 67.6
                        87.6 88.8 66.5 65.9 90 90.6 65.9 90 85.3 99.4 77.6 84.7 68.8 88.8
                        87.1 74.1 91.2 90.6 89.4 87.1 91.8 65.9 90 90.6
    No. chromosomes:    12
        Autosomes:      om1 om2 om3 om4 om5 om6 om7 om8 om9 om10 om11 om12

    Total markers:      282
    No. markers:        31 34 20 23 17 18 22 24 14 25 24 30
    Percent genotyped:  90.5
    Genotypes (%):      AA:50.1  AB:49.9
> nind(ril)
[1] 170
> nchr(ril)
[1] 12
> nphe(ril)
[1] 113
> nmar(ril)
 om1  om2  om3  om4  om5  om6  om7  om8  om9 om10 om11 om12
  31   34   20   23   17   18   22   24   14   25   24   30
> totmar(ril)
[1] 282
```

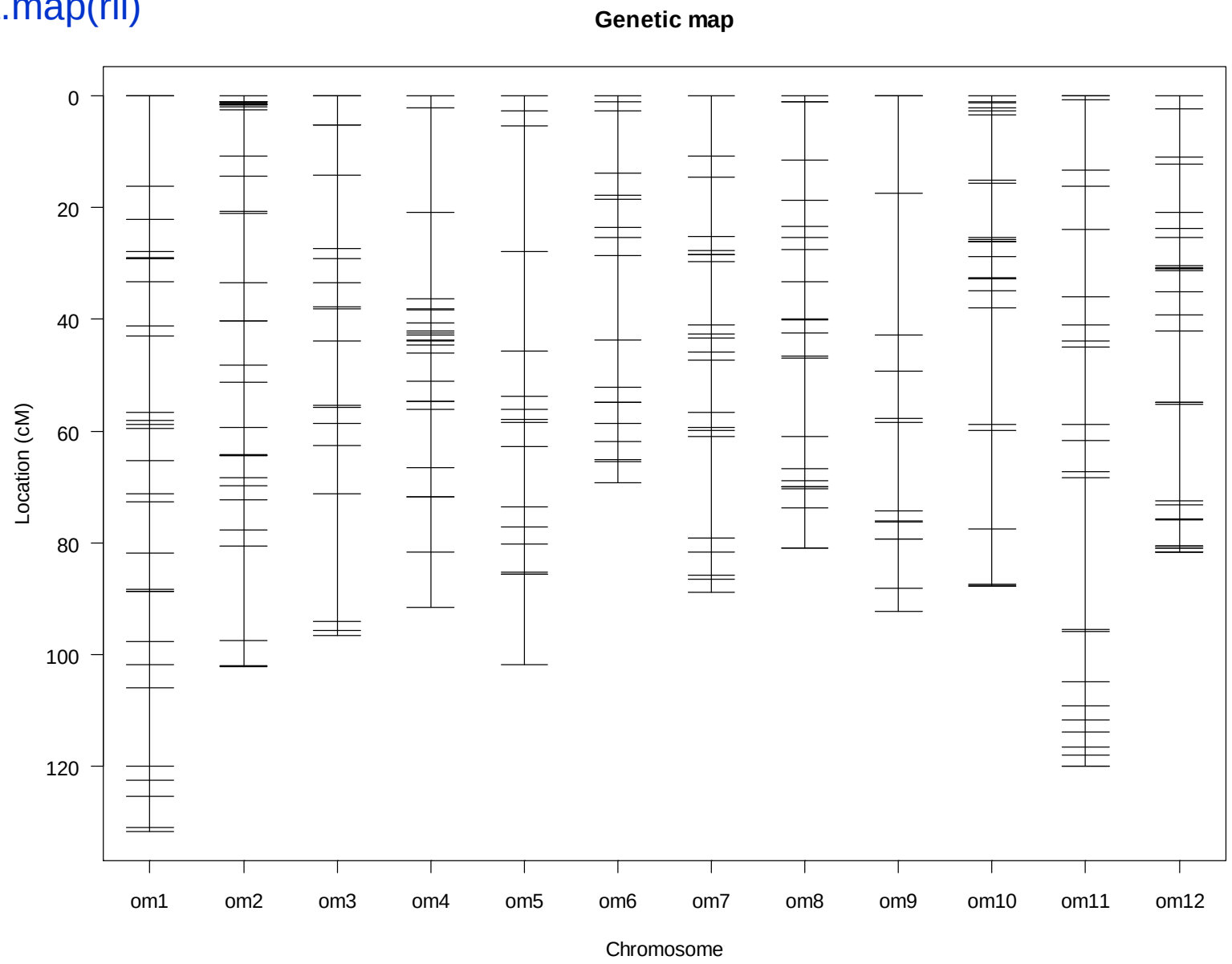
Plot missing data

> plot.missing(ril)



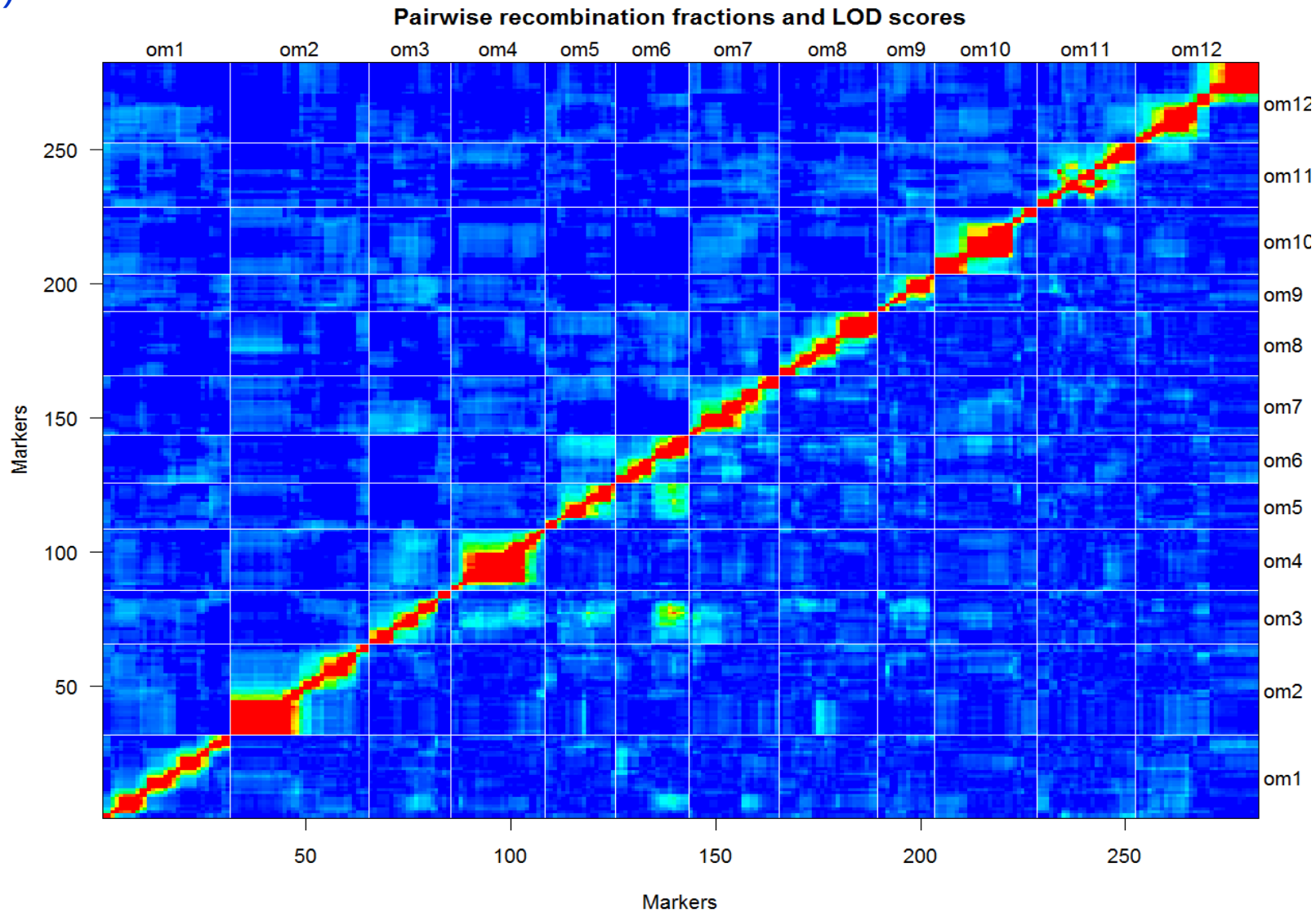
Drawing the genetic map

> plot.map(ril)



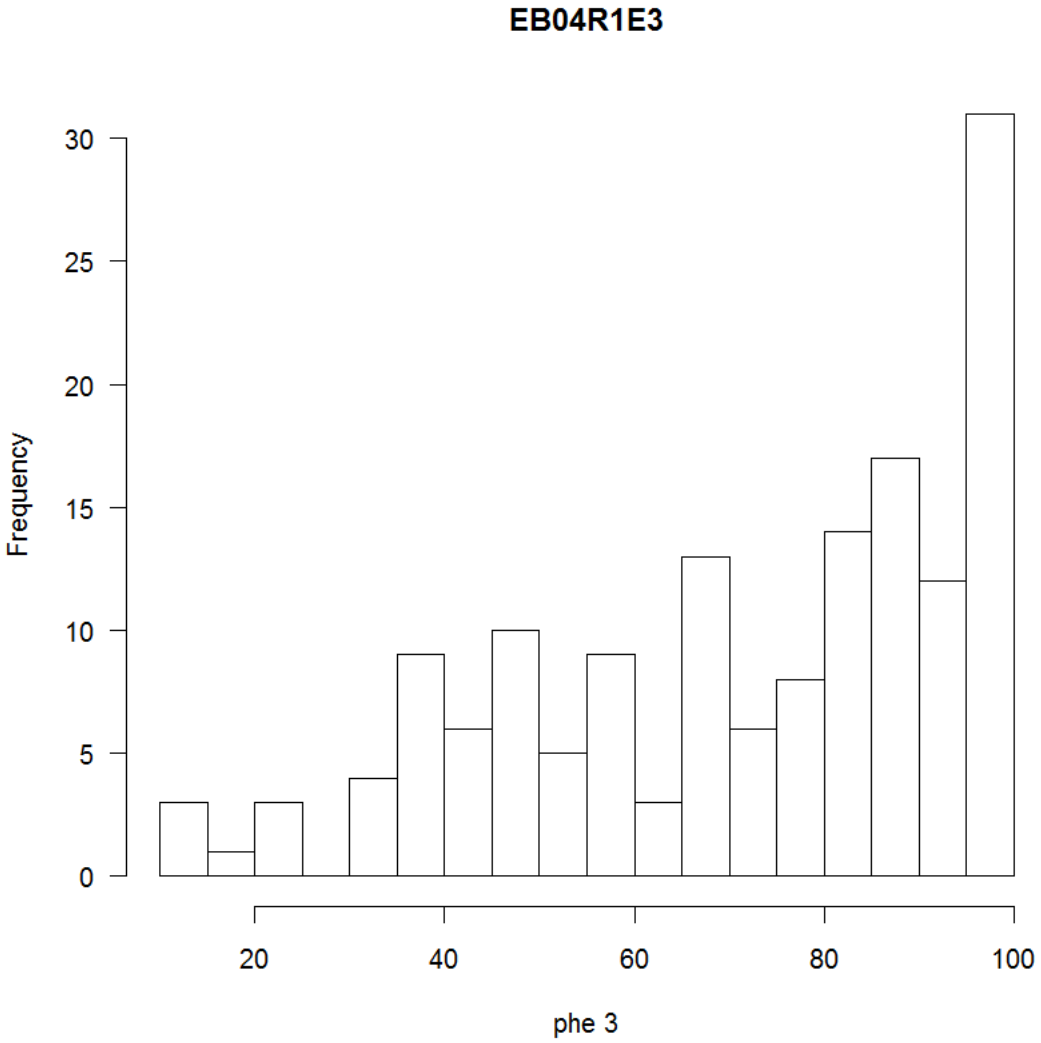
Heat plots showing hot/cold spots for recombination in the genome

```
> est_rf <- est.rf(ril)
> plot.rf(ril)
```



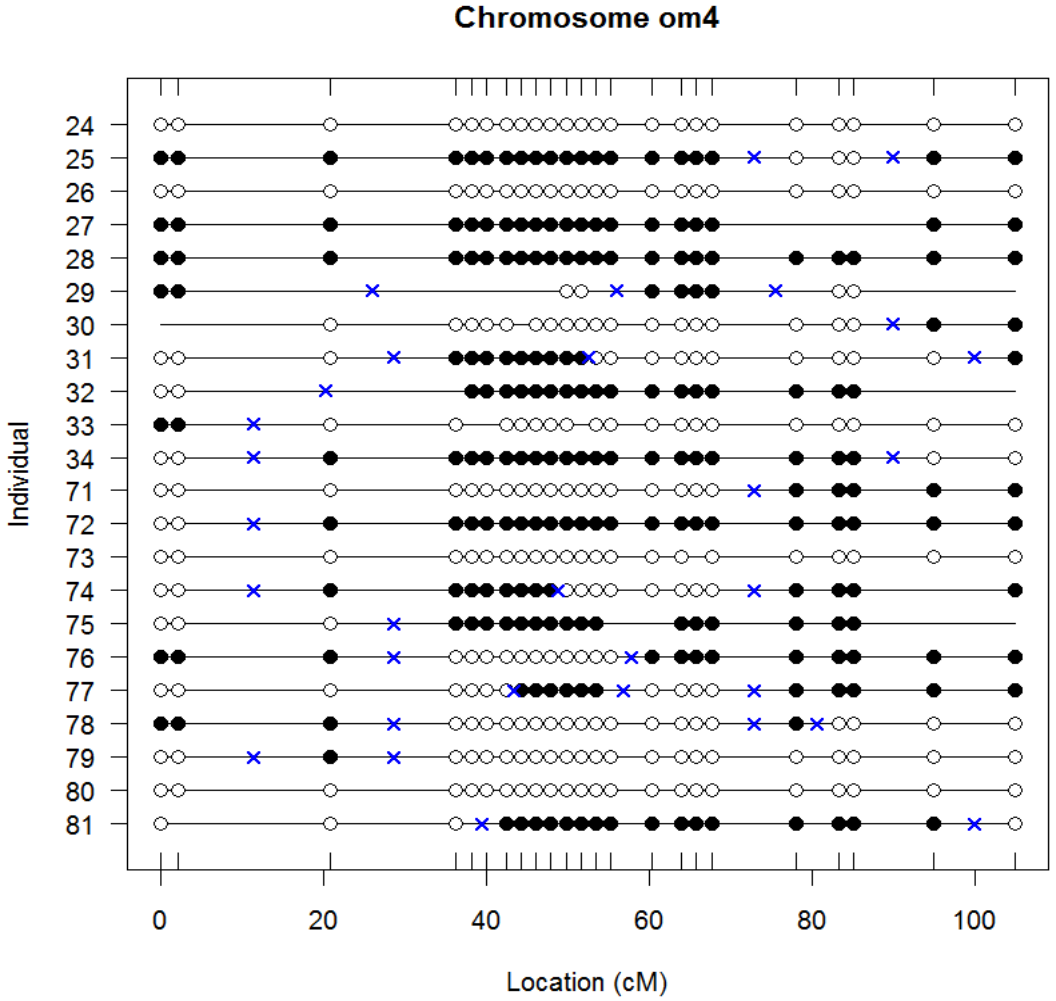
Plotting the phenotypic data

> plot.pheno(ril, pheno.col=3)



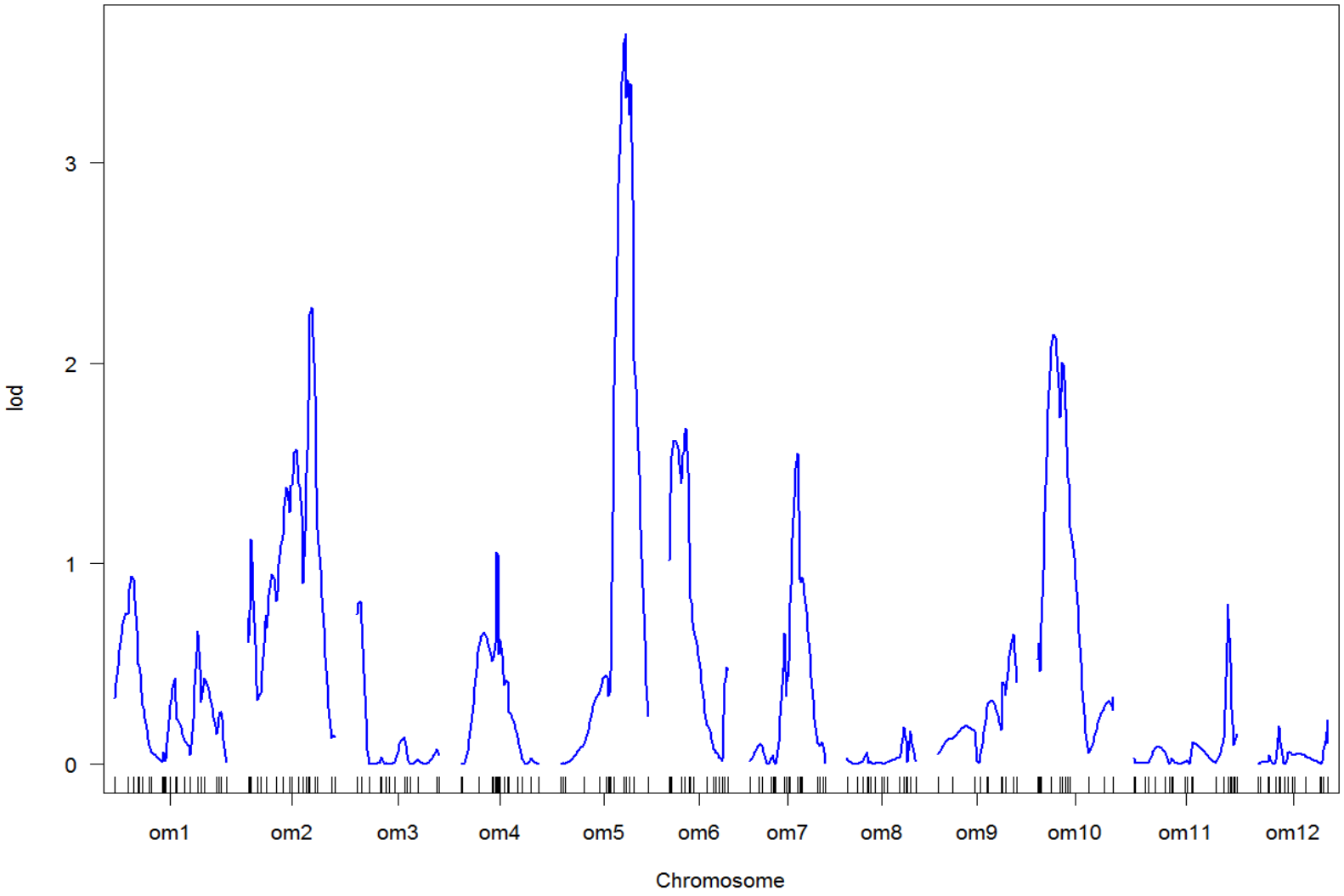
A few intermediate calculations

```
> ril <- calc.errorlod(ril, error.prob=0.01)
> ril <- calc.genoprob(ril, step=2.0, error.prob=0.01)
> plot.geno(ril, chr="om4", ind=c(24:34, 71:81), cutoff=4, min.sep=2, cex=1.2)
```



Simple interval mapping – plotting one trait

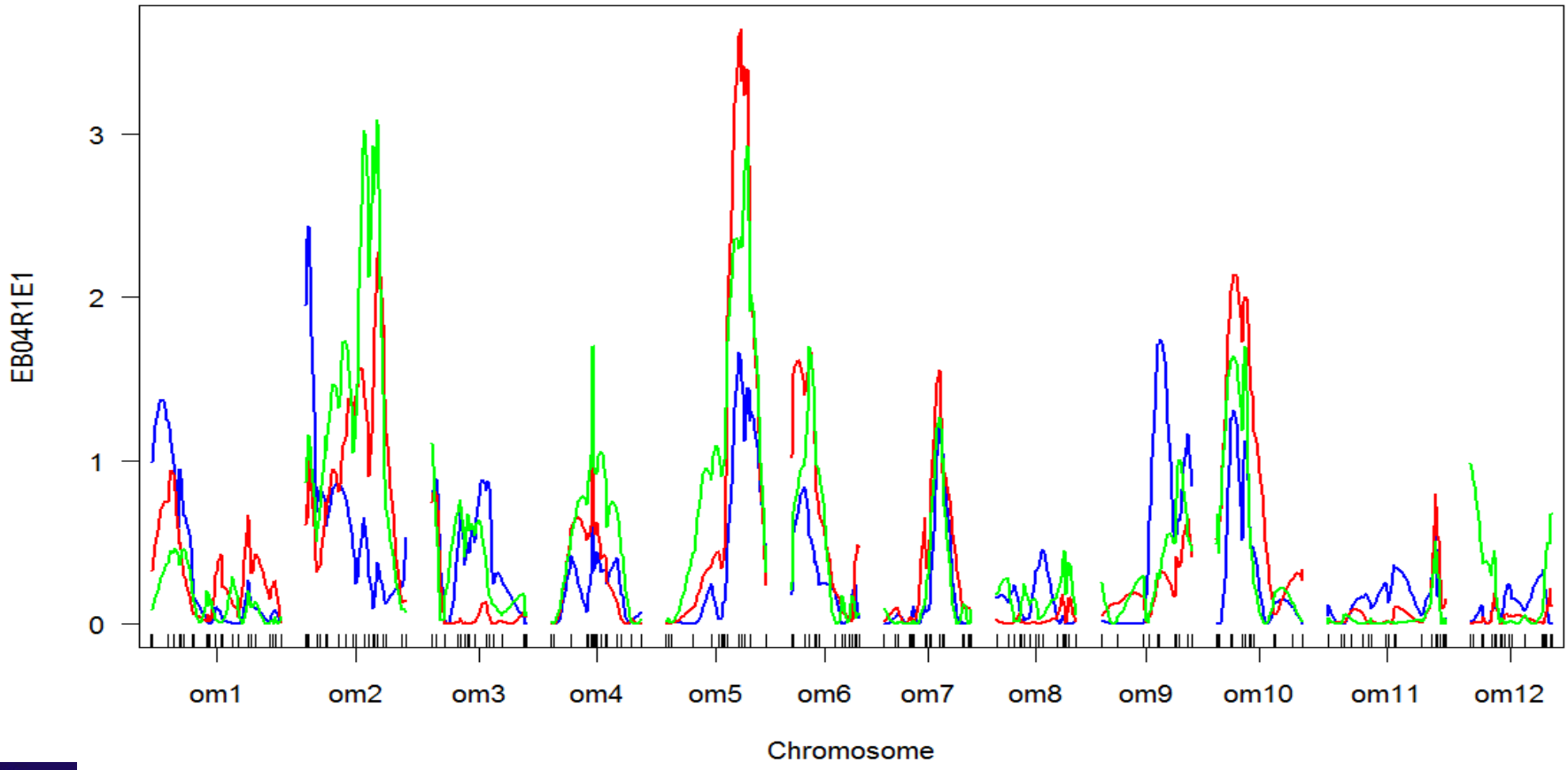
```
> out.pheno3.em <- scanone (ril, pheno.col=c(3),method="em")

> plot(out.pheno3.em, lodcolumn=c(1), lty=1, col=c("blue"))
```



Simple interval mapping – plotting more than one trait

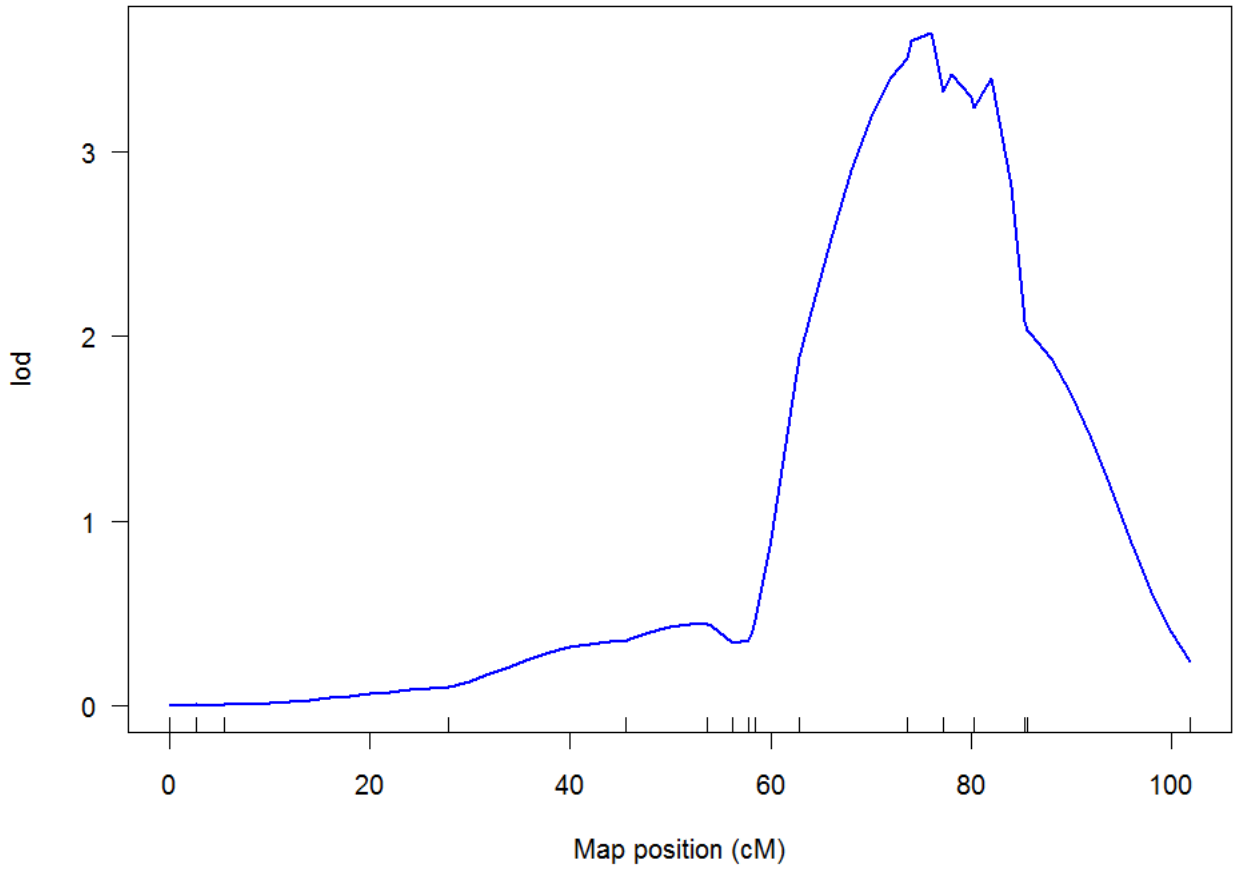
> out.pheno1_3_6.em <- scanone (ril, pheno.col=c(1,3,6),method="em")

> plot(out.pheno1_3_6.em, lodcolumn=c(1,2,3),lty=1, col=c("blue", "red", "green"))



Simple interval mapping – zooming in on a particular chromosome

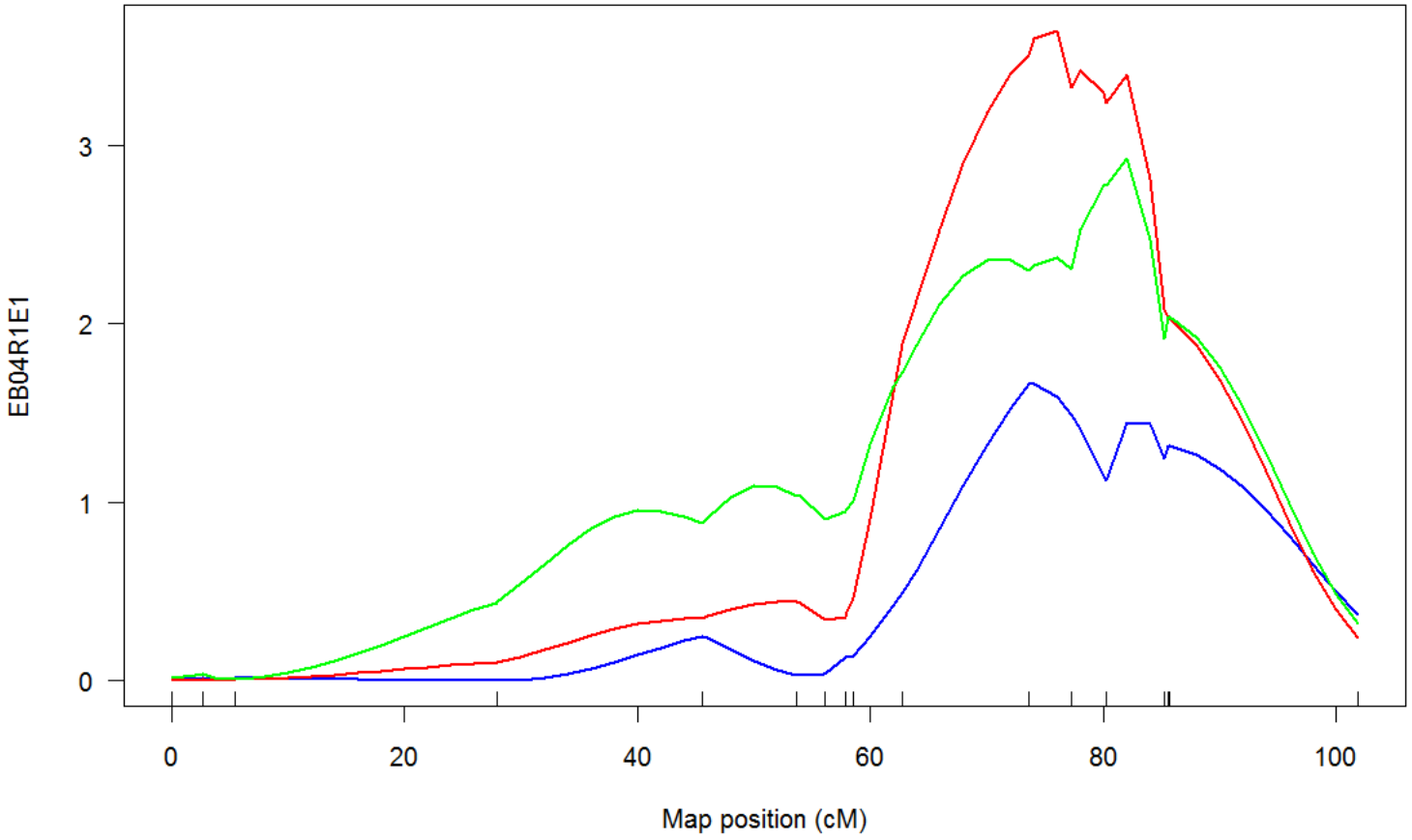
```
> out.pheno3.em <- scanone (ril, pheno.col=c(3),method="em")

> plot(out.pheno3.em, lodcolumn=c(1),chr=c("om5"), lty=1, col=c("blue"))
```



Simple interval mapping – zooming in on a particular chromosome, plotting more than one trait

> out.pheno1_3_6.em <- scanone (ril, pheno.col=c(1,3,6),method="em")

> plot(out.pheno1_3_6.em, lodcolumn=c(1,2,3),lty=1, chr=c("om5"), col=c("blue", "red", "green")
)



One may use scanone to perform a permutation test to get a genome-wide LOD significance threshold.

Haley-Knott regression method is performed (Haley and Knott, 1992)

```
> operm.hk <- scanone(ril, method="hk", n.perm=1000)
Doing permutation in batch mode ...
Warning message:
In checkcovar(cross, pheno.col, addcovar, intcovar, perm.strata,  :
  Dropping 16 individuals with missing phenotypes.


> summary(operm.hk, alpha=0.05)
LOD thresholds (1000 permutations)
    lod
5% 2.90

> save.image()
```

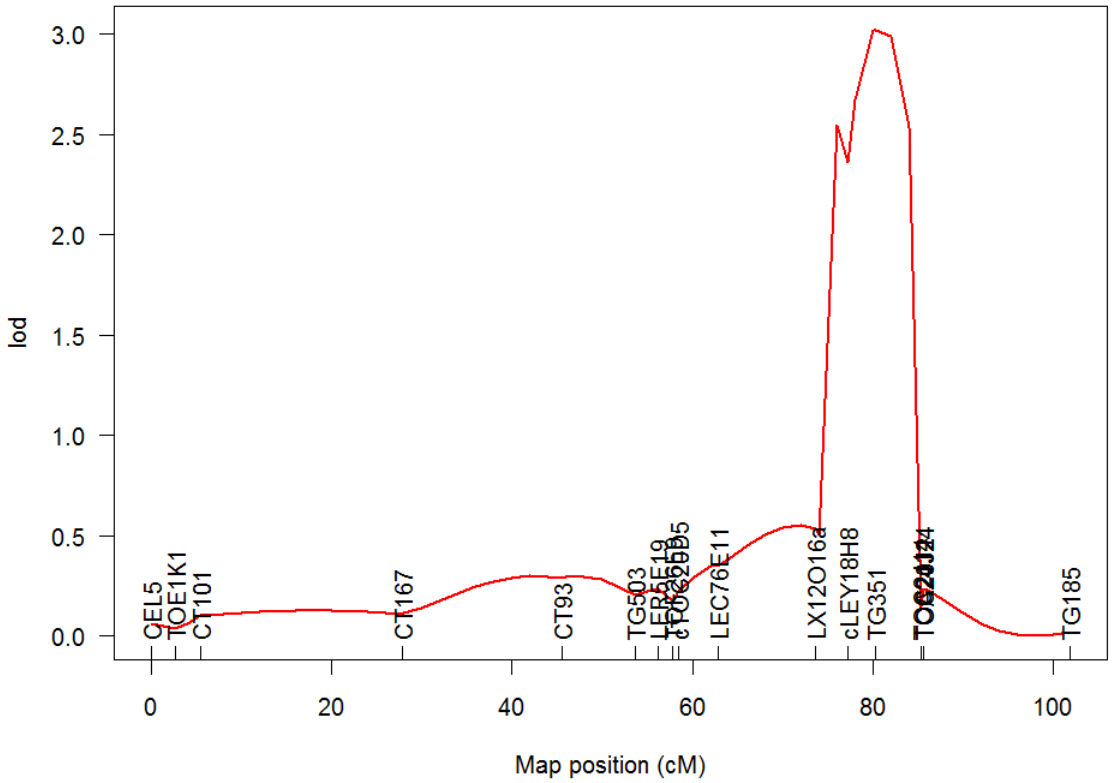
One may use scanone to perform a permutation test to get a genome-wide LOD significance threshold.

```
> operm.em <- scanone(ril, method="em", n.perm=1000)
Permutation 20
Permutation 40
            .
            .
            .
Permutation 980
Permutation 1000
Warning message:
In checkcovar(cross, pheno.col, addcovar, intcovar, perm.strata,  :
  Dropping 16 individuals with missing phenotypes.

> summary(out.em, perms=operm.em, alpha=0.05, pvalues=TRUE)
    There were no LOD peaks above the threshold.
> summary(operm.em, alpha=0.05)
LOD thresholds (1000 permutations)
    lod
5% 2.87
```

Standard interval mapping (Lander and Botstein 1989): use of a normal model and the EM Algorithm (Dempster et al. 1977)

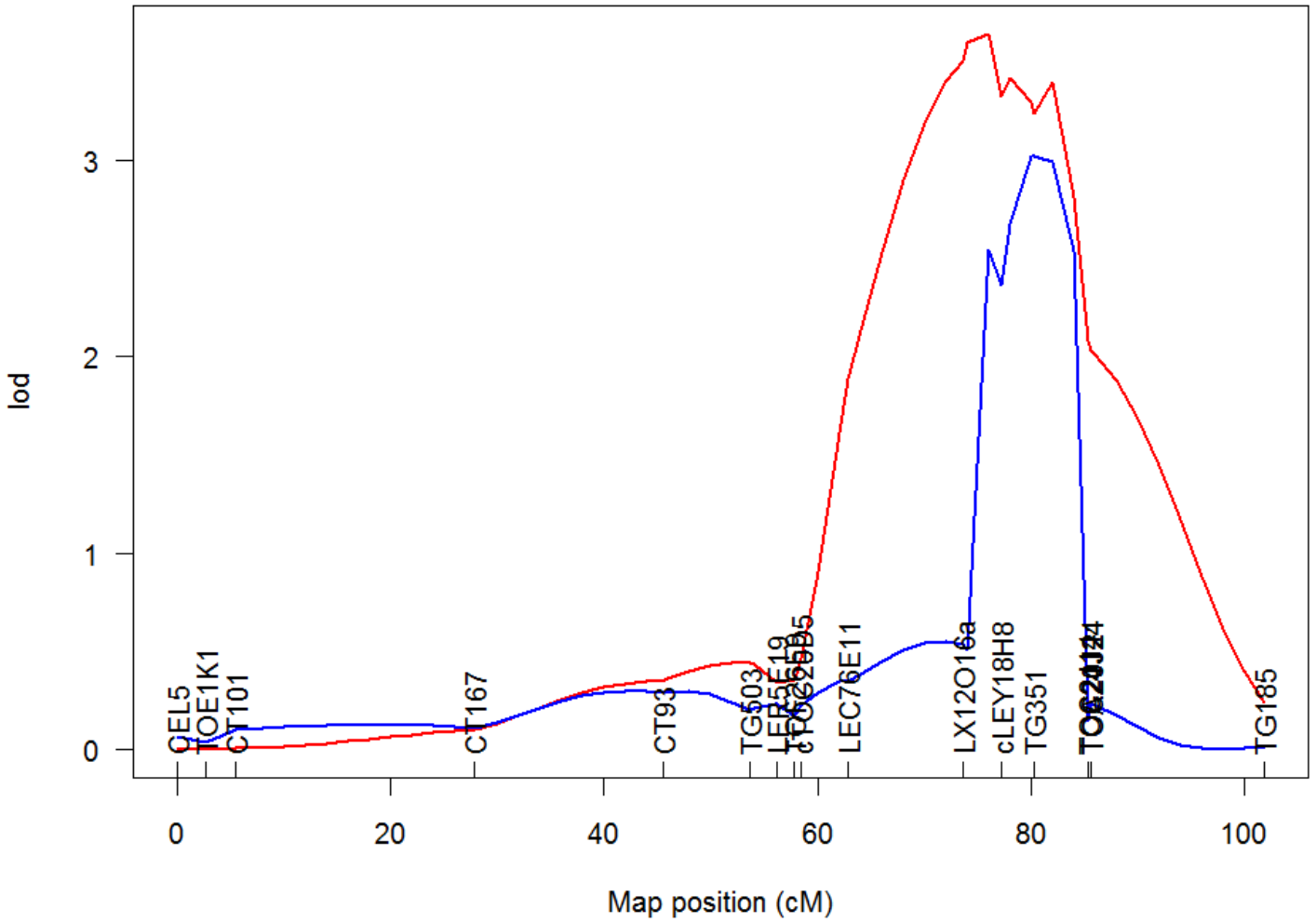
One may also use CIM to perform composite interval mapping

```
> out_pheno3_cim.em <- cim(ril, pheno.col = (6), n.marcovar=3, method=c("em"), map.function=c("kosambi") )

> plot(out_pheno6_cim.em,  chr="om5", show.marker.names=T, col =c("red"))

> attr(out_pheno6_cim.em, "marker.covar")
>[1] "TG454" "TG351" "TG590“

> attr(out_pheno6_cim.em, "marker.covar.pos")
      chr      pos
TG454 om2 59.26381
TG351 om5 80.20932
TG590 om6 18.52876
```



Superimposing SIM and CIM graphs

```
> plot(out.pheno3.em, out_pheno6_cim.em,  chr="om5", show.marker.names=T, col =c("red", "blue"))
```



Two dimensional genome scan to identify interacting QTLs

```
> out_pheno_3_scantwo <- scantwo(ril, chr=5, pheno.col=3, model=c("normal"), method=c("em"), use=c("all.obs"), verbose=TRUE)
> summary(out_pheno_3_scantwo)
```

--Running scanone
--Running scantwo
(om1,om1)
(om1,om2)
(om1,om3)

-
-
-

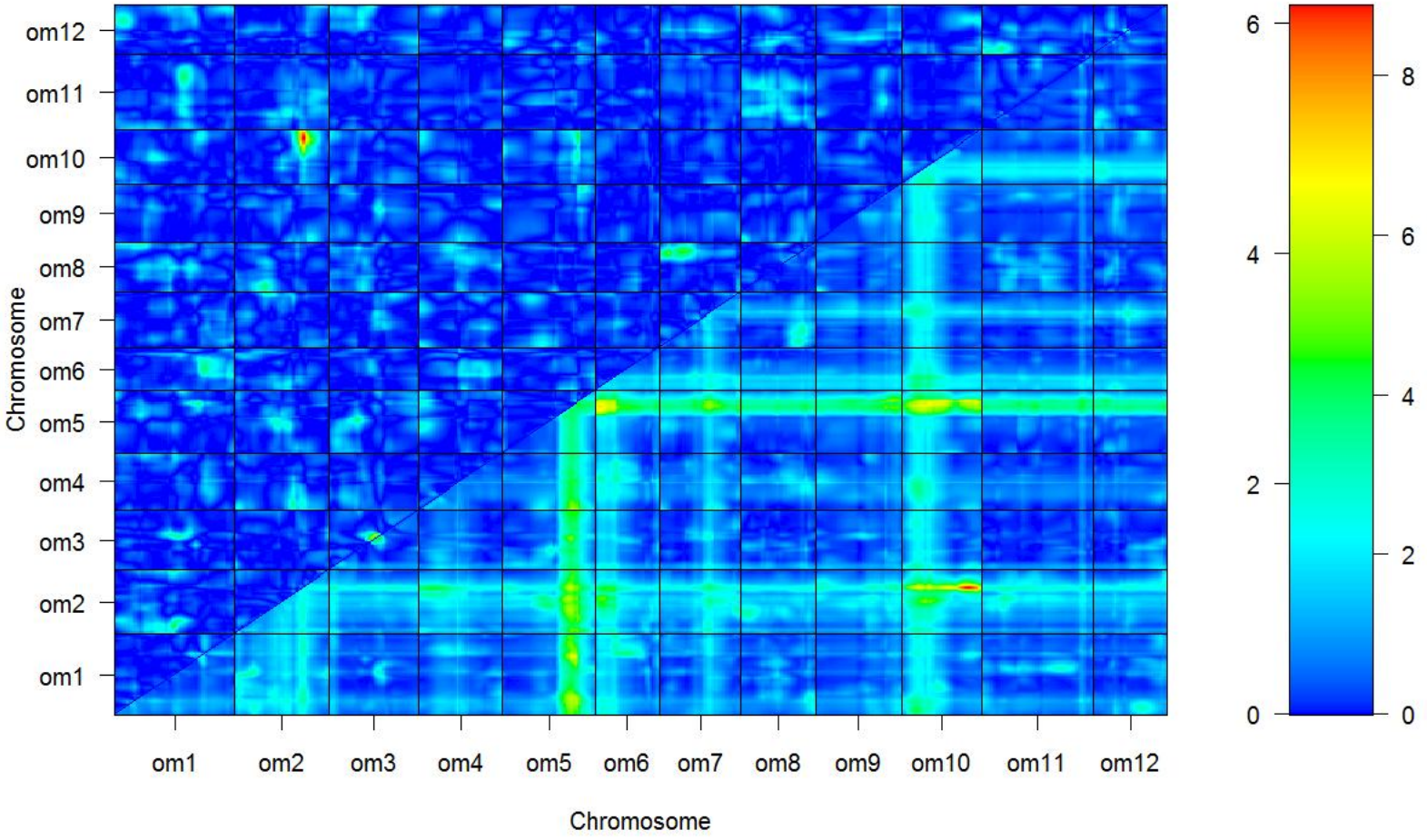
(om10,om12)
(om11,om11)
(om11,om12)
(om12,om12)

> summary(out_pheno_3_scantwo)

	pos1f	pos2f	lod.full	lod.fv1	lod.int	pos1a	pos2a	lod.add	lod.av1
com1:com1	90	96	2.42	1.479	0.764587	22	98	1.653	0.7147
com1:com2	68	20	3.42	1.145	0.529322	22	74	2.895	0.6159
com1:com3	70	54	3.09	2.148	1.310696	20	4	1.775	0.8370
com1:com4	98	42	2.59	1.541	0.414882	20	42	2.171	1.1259
com1:com5	94	76	6.32	2.678	0.859992	20	82	5.460	1.8180
com1:com6	98	32	3.97	2.294	1.481627	20	6	2.484	0.8120
com1:com7	98	56	2.72	1.174	0.587359	20	56	2.133	0.5869

Two dimensional genome scan to indentify interacting QTLs

> plot(out_pheno_3_scantwo)



Two dimensional genome scan is memory intensive

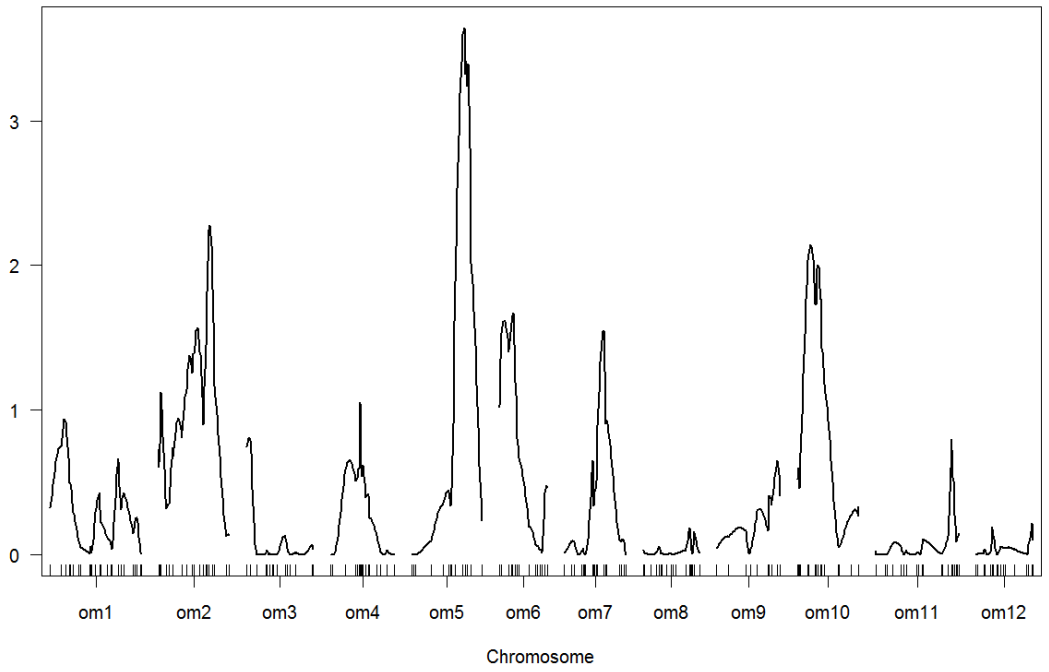
```
> out_pheno_3_scantwodim <- scantwo(ril, chr=5, pheno.col=3,
model=c("normal"),n.perm=1000, method=c("hk"), use=c("all.obs"), verbose=TRUE)
```

```
Doing permutation in batch mode ...
Error: cannot allocate vector of size 1.2 Gb
```

Fit a mutli-QTL model (Multiple Interval Mapping)

```
> out.pheno6.em <- scanone (ril, pheno.col=c(3),method="em")
> plot (out.pheno6.em)
> summary(out.pheno6.em)
```

	chr	pos	lod
com1.loc20	om1	20.0	0.938
com2.loc74	om2	74.0	2.279
com3.loc4	om3	4.0	0.813
LEN13D5	om4	42.0	1.055
com5.loc76	om5	76.0	3.642
com6.loc20	om6	20.0	1.672
TG128	om7	56.7	1.551
TG201	om8	66.7	0.183
CT96	om9	88.1	0.649
com10.loc18	om10	18.0	2.142
TG546	om11	109.1	0.796
LEZ15E8	om12	80.9	0.218




```
> chr <- c("om2", "om5", "om10")
> pos <- c(74, 76, 18)
> qtl <- makeqtl(ril, chr, pos)
> my.formula <- y~Q1 + Q2 + Q3 + Q1:Q2 + Q1:Q3 + Q2:Q3
```

Fit a mutli-QTL model (Multiple Interval Mapping)

```
> out.fitqtl <- fitqtl(ril, qtl=qtl, formula=my.formula)
> summary(out.fitqtl)
```

fitqtl summary

```
Method: multiple imputation
Number of observations : 154

Full model result
----------------------------------
Model formula: y ~ Q1 + Q2 + Q3 + Q1:Q2 + Q1:Q3 + Q2:Q3

       df        SS        MS      LOD     %var Pvalue(Chi2)  Pvalue(F)
Model   6  6863.668 1143.9447 3.275684 9.331036   0.01960520 0.02367131
Error 147 66693.741  453.6989
Total 153 73557.409


Drop one QTL at a time ANOVA table:
----------------------------------
                    df Type III SS     LOD   %var F value Pvalue(Chi2) Pvalue(F)
om2@75.0             3      1005.5 0.50040 1.3670  0.7387        0.512    0.5305
om5@75.0             3      3174.6 1.55505 4.3158  2.3324        0.067    0.0765 .
om10@17.5            3      2637.3 1.29689 3.5854  1.9376        0.113    0.1260
om2@75.0:om5@75.0    1       155.0 0.07763 0.2107  0.3417        0.550    0.5598
om2@75.0:om10@17.5   1       157.1 0.07866 0.2135  0.3462        0.547    0.5572
om5@75.0:om10@17.5   1       167.7 0.08396 0.2279  0.3695        0.534    0.5442
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```