1 Introduction

1.1 Synopsis

Package `cgmisc` contains miscellaneous functions, hopefully useful for extending genome-wide association study (GWAS) analyses.

1.2 Getting help

Like every other R function, the functions provided in this package are documented in the standard R-help (Rd) format and can be easily accessed by issuing `help()` or its shorter version, `?` function. For instance, if you want to get more information on how to use the `clump.markers()` function, type either `help(clumpmarkers)` or `?clump.markers` and press return/enter. To see this document from within R you type `vignette('cgmisc')`.

1.3 Purpose of this document

This document aims at presenting how to use functions provided in this package in a typical GWAS data analyses workflow. It is, however, not pretending to be a GWAS tutorial as such.

1.4 Conventions

- All R commands are written in terminal type: `myfun(foo=T, bar=54)`
- In the above example: `myfun` is a *function* and both `foo` and `bar` are its *arguments*

2 Working with `cgmisc`

2.1 Installation

In order to install `cgmisc`, you either use one of the R GUIs (native R GUI, RStudio etc.) or type the following command:

```
install.packages("cgmisc", repos = "")
```

Functions in the `cgmisc` package often complement or use `GenABEL` `??` package functions and data structures. `GenABEL` is an excellent and widely-used R package for performing genome-wide association studies and much more... Therefore `GenABEL` will be loaded automagically when loading `cgmisc`. If for some mysterious reason this does not happen, you can install and load `GenABEL` by typing:

```
install.packages("GenABEL")
require("GenABEL")
```

```
## Loading required package:  GenABEL
## Loading required package:  MASS
## GenABEL v.  1.7-4 (February 22, 2013) loaded
```

You load `cgmisc` package in exactly the same way (both `require` and `library` will do):

```
require("cgmisc")

## Loading required package:  cgmisc

library("cgmisc")  # Alternative to require
```

After having loaded the package it is time to load some data:

```
data(cgmiscdat1)
```

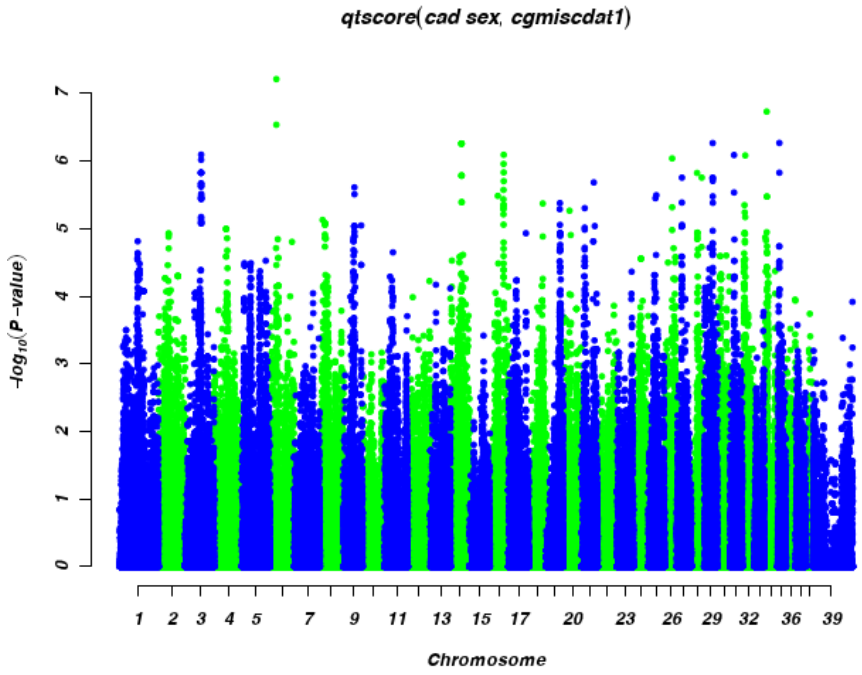
3 Using `plot.Manhattan.LD` function

The `plot.Manhattan.LD` function allows you to visualize the LD pattern in a genome fragment on an enchanced Manhattan plot. You select one marker, typically the one with the strongest association to the analysed trait and all other markers in the region are coloured according to the degree of linkage disequilibrium with this index marker. You need to begin by running a standard GWAS analyses:

```
# Run association analyses
an <- qtscore(cad ~ sex, data = cgmiscdat1)

## Warning:  binomial trait is analysed as gaussian
## Warning:  27 observations deleted due to missingness

plot(an, pch = 19, cex = 0.5)  # Plot standard Manhattan
```



Once this is done, you might be interested in checking the top associated marker. This can be done using the following call:

```
summary(an, top = 5)  # List top 5 markers
```

```
## Summary for top 5 results, sorted by P1df
##                 Chromosome Position Strand A1 A2   N    effB se_effB
## BICF2P453669             6 23340000      u  C  T 180 -0.2947 0.05443
## BICF2P1299812           34 10230234      u  G  A 180 -0.2729 0.05234
## BICF2P564616             6 23257649      u  C  T 180 -0.2621 0.05110
## BICF2G630770115         35  8408042      u  G  A 180 -0.3082 0.06150
## BICF2G630627595         29 29951256      u  G  A 180  0.2493 0.04975
##                 chi2.1df      P1df   effAB   effBB chi2.2df      P2df
## BICF2P453669       29.32 6.125e-08 -0.3111 -0.5754    29.40 4.123e-07
## BICF2P1299812      27.18 1.854e-07 -0.2806 -0.5447    27.19 1.245e-06
## BICF2P564616       26.31 2.902e-07 -0.2596 -0.5254    26.32 1.931e-06
## BICF2G630770115    25.12 5.395e-07 -0.2944 -0.6416    25.18 3.402e-06
## BICF2G630627595    25.11 5.416e-07  0.3391  0.4561    26.92 1.429e-06
##                    Pc1df
## BICF2P453669    0.0001053
## BICF2P1299812   0.0001888
## BICF2P564616    0.0002392
## BICF2G630770115 0.0003318
## BICF2G630627595 0.0003325
```

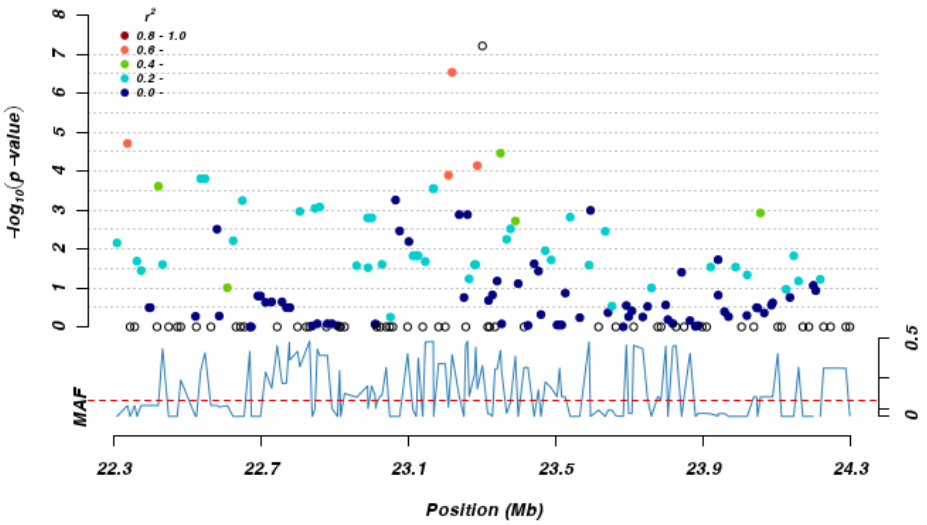


Figure 1: Here the black, empty circles denote monomorphic markers – we did not perform any quality control before.

We can see that `BICF2P453669` on chromosome 6 (23.34Mb) is the strongest association. Now time to visualize a short fragment of chromosome 6, boyh 1.0Mb downstream and 1.0Mb upstream of the top-associated marker.

```
plot.manhattan.LD(data=cgmiscdat1, gwas.result=an, chr=6,
                  region=c(22340000, 24340000),
                  index.snp="BICF2P453669",
                  p.value=0.05, bonferroni=F,
                  mafThreshold=.1)
```

4 Using `clump.markers` function

The `clump.markers` functions implements clumping algorithm as described in PLINK documentation **??**. In short, the clumping algorithm consists of the following steps:

- First step