

# ldblock package: linkage disequilibrium data structures

*Vincent J. Carey, stvjc at channing.harvard.edu*

February 2015

## Contents

1	Introduction . . . . .	2
2	Import of HapMap LD data . . . . .	2
3	A view of the block structure . . . . .	2
4	Collecting SNPs exhibiting linkage to selected SNP . . . . .	3

## 1 Introduction

---

There is a nice vignette in *snpStats* concerning linkage disequilibrium (LD) analysis as supported by software in that package. The purpose of this package is to simplify handling of existing population-level data on LD for the purpose of flexibly defining LD blocks.

## 2 Import of HapMap LD data

---

The `hmld` function imports gzipped tabular data from hapmap's repository [hapmap.ncbi.nlm.nih.gov/downloads/ld\\_data/2009-02\\_phaseIII\\_r2/](http://hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-02_phaseIII_r2/).

```
suppressPackageStartupMessages(library(ldblock))
path = dir(system.file("hapmap", package="ldblock"), full=TRUE)
ceu17 = hmld(path, poptag="CEU", chrom="chr17")
## importing /tmp/Rtmp055tJ5/Rinst2fbd1b2c5941/ldblock/hapmap/ld_chr17-CEU.txt.gz
## done.
ceu17
## ldstruct for population CEU, chrom chr17
## dimensions: 36621 x 36621 ; statistic is Dprime
## object structure:
##      ldmatrix      chrom      genome      allpos      poptag      statInUse
## "dsCMatrix" "character" "character" "numeric" "character" "character"
##      allrs
## "character"
```

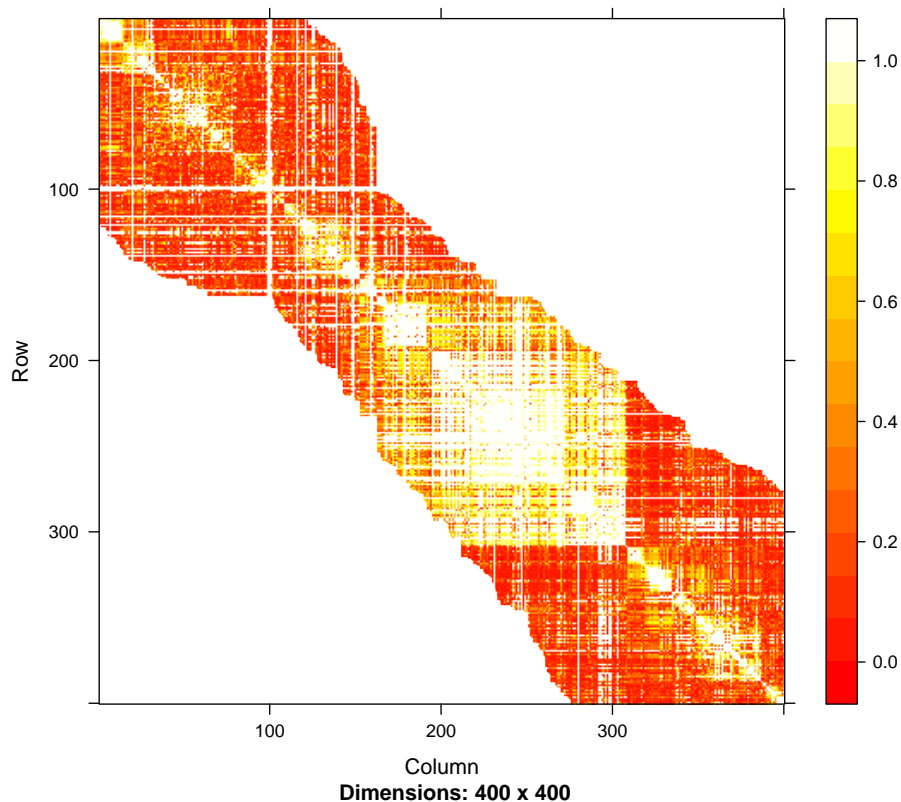
## 3 A view of the block structure

---

For some reason knitr/render will not display this image nicely.

```
library(Matrix)
##
## Attaching package: 'Matrix'
## The following object is masked from 'package:S4Vectors':
##
##      expand
image(ceu17@ldmatrix[1:400,1:400],
      col.reg=heat.colors(120), colorkey=TRUE, useRaster=TRUE)
```

## ldblock package: linkage disequilibrium data structures



This ignores physical distance and MAF. The bright stripes are probably due to SNP with low MAF.

## 4 Collecting SNPs exhibiting linkage to selected SNP

We'll use `ceu17` and the `gwascat` package to enumerate SNP that are in LD with GWAS hits.

```
library(gwascat)
## Registered S3 methods overwritten by 'ffbase':
##   method from
##   [.ff      ff
##   [.ffdf   ff
##   [<-.ff   ff
##   [<-.ffdf ff
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
## gwascat loaded. Use data(ebicat38) for hg38 coordinates;
## data(ebicat37) for hg19 coordinates.
data(ebicat37)
seqlevelsStyle(ebicat37) = "NCBI"
```

## ldblock package: linkage disequilibrium data structures

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =  
## dec, : EOF within quoted string  
e17 = ebicat37[ which(as.character(seqnames(ebicat37)) == "17") ]
```

Some dbSNP names for GWAS hits on chr17 are

```
rsh17 = unique(e17$SNPS)  
head(rsh17)  
## [1] "rs11078895" "rs11891" "rs7501939" "rs9905704" "rs4796793"  
## [6] "rs78378222"
```

We will use `expandSnpSet` to obtain names for SNP that were found in HapMap CEU to have which  $D' > .9$  with any of these hits. These names are added to the input set.

```
length(rsh17)  
## [1] 530  
exset = expandSnpSet( rsh17, ldstruct= ceu17, lb=.9 )  
## Warning in expandSnpSet(rsh17, ldstruct = ceu17, lb = 0.9): dropping 149 rsn  
## not matched in ld matrix  
length(exset)  
## [1] 13209  
all(rsh17 %in% exset)  
## [1] TRUE
```

Not all GWAS SNP are in the HapMap LD resource. You can use your own LD data as long as the format agrees with that of the HapMap distribution.