

CHARM

Getting started

CHARM is a program written in *R* source code, though not yet a full *R* package. The source code can be obtained from:

- <http://niallcardin.com/CHARM/>

To run CHARM, first download code.zip, and unzip it. Make sure you know the directory where you have now placed all of the .r files. When you run *R*, either make sure that these files are in the working directory, or ‘source’ each file individually. If running *R* from Linux or mac terminals then if *R* is loaded from the directory containing the files then this will happen automatically, otherwise use the ‘setwd’ function in *R*.

Running CHARM

Then run the command

```
source("CHARM.r")
```

To run CHARM on some random, very simple, test data (with no signal) then use the line:

```
CHARMresults = runCHARMonTestData()
```

By reading the code of the function runCHARMonTestData in the file CHARM.r you will be able to see how this was done, look at the format of the objects used and returned by CHARM and then replace input objects with those of your own.

The basic inputs are:

1. rGen: a named matrix of genotypes for the region. Each row should be an individual and each column should be a variant.
2. caseControlStatus: a named (using the same names as the rows of the genotypes matrix) binary vector indicating the cases (1) or control (0) status of each individual in the genotypes matrix.

Example line one might run:

```
CHARMmain(  
  rGen          = yourGenotypeMatrix,  
  caseControlStatus = yourCaseControlStatusList  
)
```

Outputs

CHARM returns a list of objects (some of which are lists). For most users this is far more information than is required and most of the returned value can be immediately discarded. The first two elements of the list give the basic result in the form an analog to a Bayes Factor, the cross validated Bayes Factor described in the paper. To pull out and display the Bayes Factor you might use:

```
log10cB = CHARMresults[["Log10BayesFactor"]]
cat("  \"Cross Validated Bayes Factor\" supporting association = ",
    round(10^log10cB, 2), "\n", sep="")
```

In principle these statistics mimic Bayes Factors, see e.g. http://en.wikipedia.org/wiki/Bayes_factor. However this is an approximation, so although this may be a useful rule of thumb, they will not have the correct properties and cannot be used as is for publication.

There are multiple ways one can use these for publication:

1. With small enough data, or with a compute cluster: use a permutation test on each region. Saving time by ending unpromising permutation tests early.
2. Run CHARM for each region of interest. Focusing on the regions showing most promise (highest Bayes Factors) run permutation tests only on these. Note: If assessing significance using a multiple testing correction, then the total number of regions examined must be used, not just those on which permutation was applied!
3. Genome Wide Data: Break the genome into regions. Run CHARM on each region. Permute the case control labels and run CHARM on the genome again. Use the permuted genome wide data as a null distribution of Bayes Factors.

Note, for option 3 if the genome has been split into regions of varying size then a little caution must be used. The null distribution of Bayes Factors depends somewhat on the region size. In our simulations we broke regions up into

- *small* = fewer than 30 variants
- *medium* = 30 to 100 variants
- *large* = more than 101 variants