

Genome-Wide Association analysis using GenABEL

Yurii Aulchenko, Najaf Amin

March 20, 2007

Abstract

In this exercise, you will become familiar with the **GenABEL** library, designed for GWA analysis. Compared to `dgc.genetics` package, it provides specific facilities for storage and manipulation of large amounts of data, very fast tests for GWA analysis, and special functions to analyse and graphically present the results of GWA analysis (thus "analysis of analysis").

GenABEL is rather new (first public release in mid-2006) and still developing (you can check <http://mga.bionet.nsc.ru/nlru/GenABEL> to see the history); the latest release was done in the beginning of March. This means there may still be (hopefully few) bugs or inconsistencies in the program. We will appreciate your suggestions on improving **GenABEL**.

In the first part of this exercise you will be guided through a GWA analysis of a small data set. In the second part you will investigate a larger data set by yourself, do a verification study and will answer the questions. All data sets used assume a study in a relatively homogeneous population. Try to finish the first part in the morning and the second part in the afternoon.

Though only few thousands of markers located at four chromosomes are used in the scan, we still going to call it Genome-Wide (GW), as the amount of data we will use is approaches the amount to be expected in a real experiment.

Contents

1	Example GWA session	2
1.1	Data descriptives and first round of GWA analysis	2

1 Example GWA session

Copy the file `ge03d2ex.RData` to your desktop and start R by double-clicking on it. Start GenABEL library by typing

```
> library(GenABEL)
```

You can read short overview of the package by asking for `help(GenABEL)`. Investigate the objects loaded by command

```
> ls()
```

```
[1] "ge03d2ex"
```

The `ge03d2ex` is a special data object of the class `gwaa.data`, as can be seeing from

```
> class(ge03d2ex)
```

```
[1] "gwaa.data"  
attr(,"package")  
[1] "GenABEL"
```

As usual, if you are interested in details of this data type, you can get help by using command `help("gwaa.data-class")`; however, it is not strictly necessary to understand details of this data type to do GWA analysis in general and this exercise in particular. The only important thing to remember is that objects of this type contain a slot `phdata` which is a data frame with phenotypic information.

To check the variables in this data frame, you can use

```
> names(ge03d2ex@phdata)
```

```
[1] "id"      "sex"     "age"     "dm2"     "height"  "weight"  "diet"    "bmi"
```

Of course, all standard R procedures will work on this data frame, e.g. we can check the summary for the age variable by

```
> summary(ge03d2ex@phdata$age)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
23.84	38.33	48.71	49.07	58.57	81.57

We can also attach this data frame to the R search path by

```
> attach(ge03d2ex@phdata)
```

1.1 Data descriptives and first round of GWA analysis

Let us investigate what are the traits presented in the data frame loaded and what are the characteristics of the distribution by using specific GenABEL function

```
> descriptives.trait(ge03d2ex)
```

	No	Mean	SD
id	136	NA	NA
sex	136	0.529	0.501
age	136	49.069	12.926
dm2	136	0.632	0.484
height	135	169.440	9.814
weight	135	87.397	25.510
diet	136	0.059	0.236
bmi	135	30.301	8.082

You can see that this frame contains the data on 136 people; the data on sex, age, height, weight, diet and body mass index (BMI) are available. Our trait of interest is `dm2` (type 2 diabetes). Note that every single piece of information in this data set is simulated; however, we tried to keep our simulations in a way we think the control of T2D works.

You can produce a summary for cases and controls separately and compare distributions of the traits by

```
> descriptives.trait(ge03d2ex, by = dm2)
```

	No(<i>by</i> =1)	Mean	SD	No(<i>by</i> =0)	Mean	SD	Ptt	Pkw	Pexact
id	86	NA	NA	50	NA	NA	NA	NA	NA
sex	86	0.593	0.494	50	0.420	0.499	0.053	0.052	0.074
age	86	50.250	12.206	50	47.038	13.971	0.179	0.205	NA
dm2	86	NA	NA	50	NA	NA	NA	NA	NA
height	86	170.448	10.362	49	167.671	8.586	0.097	0.141	NA
weight	86	93.587	27.337	49	76.534	17.441	0.000	0.000	NA
diet	86	0.058	0.235	50	0.060	0.240	0.965	0.965	1.000
bmi	86	32.008	8.441	49	27.304	6.463	0.000	0.001	NA

here, the `by` argument specifies the grouping variable. You can see that cases and controls are different in weight, which is expected, as T2D is associated with obesity.

Similarly, you can produce grand GW descriptives of the marker data by using

```
> descriptives.marker(ge03d2ex)
```