GEVACO Introduction

Sydney Manning
Department of Pharmacotherapy
University of North Texas Health Science Center – Fort Worth

Dec 8, 2021

```
library(GEVACO) # load the library
```

Data requirements

At minimum to run this analysis you need a file storing genotype information and a covariate/trait file.

The covariate/trait file should be text based and can have as many columns/covariates as desired, but the first few must be in a specific order.

- Column 1: Phenotype
- Column 2: Environmental factor
- Columns 3+: Additional covariates

```
data(cov_example) # read in the example covariate data

head(cov_example) # view the format of the covariate file
#>            y   BMI age sex
#> 1  2.4062485 29.83  40   1
#> 2  1.7048071 28.57  33   2
#> 3  0.4293889 21.58  31   2
#> 4  5.0239420 29.69  40   2
#> 5  4.9903090 35.70  31   2
#> 6 11.4007632 33.45  51   2
```

If you have PLINK bed files or files in other format, you will need to convert the genotypes to a genotype matrix and filter it for your SNPs of interest. We chose to filter our dataframe by finding the location of all SNPs meeting our desired threshold of minor allele frequency.

```
data(geno_example) # read in the example genomic file

                   # we included the first 20 SNPs to meet our criteria, each column
head(geno_example) # is a different SNP
#>    exm210 exm340 exm2264981 exm2253593 exm596 exm773 exm912 exm1110 exm1542
#> 3       1      0          0          1      1      0      1       0       0
#> 6       0      0          1          0      0      1      0       0       0
#> 10      0      0          0          0      0      0      0       0       0
```

```
#> 11      0      1           1           0      0      0      1       0       0
#> 16      0      1           0           1      1      1      1       0       1
#> 19      0      0           1           0      0      0      1       0       0
#>     exm1649 exm1654 exm1952 exm2070 exm2110 exm2183 exm2250 exm2270 exm2941
#> 3         0       0       1       0       0       0       0       0       0
#> 6         0       0       0       1       0       0       0       0       0
#> 10        0       0       0       0       1       2       0       0       1
#> 11        0       0       1       1       0       0       1       1       0
#> 16        0       0       1       0       1       1       0       0       0
#> 19        0       0       0       0       2       2       0       0       2
#>     exm3098 exm3203
#> 3         1       1
#> 6         0       0
#> 10        1       0
#> 11        0       0
#> 16        0       1
#> 19        2       0
```

Performing the screening test

Using the filtered genomic data, all that's left is to input it with the covariate information into the final function. The default number of simulation iterations is 1E5, and the default number of knots is 7.

```
results <- GxEscreen(dat=cov_example, geno=geno_example, nsim = 1e5, K=7) # run the screen
              # view the first few results: these are the p.values for each SNP in the final
head(results) # genomic file
#> [1] 0.85062 0.28910 0.90202 1.00000 1.00000 0.12141
```

The final output is a vector containing the p-value of each SNP used in the simulation.