BRIdGE: Bayesian **R**egression for **Id**entifying **G**ene-**E**nvironment interactions

BRIdGE consists of two complementary programs, BRIdGE_bfcalc and BRIdGE_hm. Given genotypes and phenotypic measurements taken under two conditions in paired samples, BRIdGE_bfcalc calculates BFs that correspond to the evidence of genetic association under specific models of gene-environment interaction (or no interaction). BRIdGE_hm uses a hierarchical model to combine information across a large number of related phenotypes (e.g. gene expression measurements from a microarray) in order to estimate parameters (e.g. the proportion of phenotypes with an association following a given interaction model) that are then used to estimate posterior probabilities and false discovery rates. These methods are described in detail in Maranville et al. 2011. The output of BRIdGE_bfcalc is the input for BRIdGE_hm.

BRIdGE_bfcalc:
The command line to calculate the BFs is as follows:
BRIdGE_bfcalc -d phenofile genofile -g grid > BF_output

The grid file is a list of expected effect sizes (σ_a), for example:
0.8
1.0
1.2
1.6

The genofile is formatted as follows:
column1: phenotype ID (e.g. Entrez GeneID)
column2: row identifier (rs#)
columns3 - column(sample size+2): genotypes

Files should be space delimited with no header.

Here is an example of an input genofile:
768329 rs4575638 1.014 0.93 0.778 0.903 1.0 2.0 1.049 0.0 0.0 0.0
768329 rs7254366 1.1 1.057 1.135 1.033 2.0 2.0 1.0 0.0 0.0 1.0
768329 rs11878556 1.001 0.819 1.058 0.949 1.0 2.0 1.0 0.998 0.0 0.0
768329 rs2361806 1.244 1.19 1.268 1.152 1.0 2.0 0.0 0.0 0.0 0.0
768329 rs10426683 1.139 1.122 1.082 1.037 1.0 0.0 1.0 0.0 0.0 1.0

The phenofile is formatted similarly:
column1: phenotype ID (e.g. Entrez GeneID)
column2: treatment condition (pheno1 or pheno2)
columns3 - column(sample size+2): data

Here is an example of an inoput phenofile:
768329 pheno1 0.606986834803486 1.0402029655509 1.30180674888568 -0.716497500177991 -0.774565430336672 0.967421566101701 0.504322046077778 0.716497500177991 -0.835371143572111 0.0880655697240946
768329 pheno2 -0.774565430336672 0.0439901182953111 1.41218757890616 1.0402029655509 0.899434907667234 -0.504322046077778 0.967421566101701 -0.312861399760112 0.454981140306428 0.660751127140255

This program calculates Bayes factors for each polymorphism-phenotype comparison under each interaction model (configuration of genotypic effects in the two conditions) with expected effect size (from the grid file).

An example of the output data is shown below:

```
rs34072018_25888  1   -4.00276e-01  -4.68693e-01  -5.82373e-01  -8.71963e-01
rs34072018_25888  2   -3.87442e-01  -4.55216e-01  -5.68188e-01  -8.57022e-01
rs34072018_25888  3   -7.84782e-01  -9.20826e-01  -1.14732e+00  -1.72557e+00
rs34072018_25888  4   -3.90550e-01  -4.58480e-01  -5.71623e-01  -8.60640e-01
```

Each column represents a different expected effect size.

There are five rows for each polymorphism-phenotype combination, each corresponding to a different model. The models are as follows:
Model 1: Genotypic effect only in treatment condition 2
Model 2: Genotypic effect only in treatment condition 1
Model 3: Genotypic effect in both treatment conditions but of unequal size
Model 4: Identical genotypic effect in both treatment conditions

BRIdGE_hm:
The command line to run the hierarchical model is as follows:
(BRIdGE_hm -d BF_output -s #models -g #values in grid file -t threshold > Posteriors) > EM_results

The threshold (-t) refers to the amount of change in the log likelihood at which the EM algorithm stops.

The EM_results file will record the point estimates of each parameter, the likelihood at each iteration of the EM algorithm and confidence intervals for the parameter estimates.

An example output file is shown below:
iter 217 loglik = 5868.1840 pi0 0.0988 config 0.038 0.050 0.000 0.912 grid 0.988 0.000 0.000 0.012
iter 218 loglik = 5868.1850 pi0 0.0988 config 0.038 0.050 0.000 0.912 grid 0.988 0.000 0.000 0.012
iter 219 loglik = 5868.1860 pi0 0.0988 config 0.038 0.050 0.000 0.912 grid 0.988 0.000 0.000 0.012

Confidence Intervals
pi0: [0.067, 0.129]
config: [0.025, 0.053] [0.037, 0.064] [0.000, 0.001] [0.894, 0.929]

In the record of each iteration, pi0 is an estimate of the proportion of phenotypes with no associated locus and the four numbers after config represent the estimate of the proportion of phenotypes with associations following into each of the four interaction models (in the same order as above).

In the confidence intervals part at the end, the values after "config:" represent the confidence intervals for the point estimates (in the final iteration) of the proportion of phenotypes with associations following interaction models 1-4, respectively.

These estimates are used as priors in calculating the posterior probabilities reported in the Posteriors file. A sample line from a Posterior file is listed below:

2542 0.7713 -0.074 rs569 0.0349 0.908 config * 4 (1) 0.0126 (2) 0.0129 (3) 0.0003 (4) 0.7454

The Posteriors file is formatted as follows:
Column 1: phenotype name
Column 2: the posterior probability of an associated locus for the phenotype
Column 3: mean BF for association for the phenotype across polymorphisms and models
Column 4: most associated polymorphism
Column 5: relative weight of evidence for association at most associated polymorphism (compared to other tested polymorphisms). This is not specific to certain model.
Column 6: BF representing evidence of association between most likely polymorphism and phenotype
Column 9: most likely model
Columns 10-22: The rest of the line is formatted as "(model c)" followed by the posterior probability of an association following model c. Note that this probability is averaged over all

polymorphisms tested for that phenotype and that the summation of the probabilities for each model should give the probability that that phenotype has an associated locus.