AdaJoint(Adaptive Joint Test) Package

December 5, 2014

```
> library(AdaJoint)
```

Detailed examples of computing the gene and pathway p-values

We will start with the sample data of SNPs and sample phenotype data to generate the observed and permutation p-values for each SNP in the pathway. First, lets get the paths to the phenotype and genotype data

```
> pheno_file <- system.file("sampleData", "pheno_data.txt", package = "AdaJoint")
> geno_file <- system.file("sampleData", "geno_data.txt", package = "AdaJoint")
> print(pheno_file)
```

```
[1] "C:/Users/wheelerb/AppData/Local/Temp/Rtmpp7FD3i/Rinst39f076ee/AdaJoint/sampleData/phe
```

```
> print(geno_file)
```

```
[1] "C:/Users/wheelerb/AppData/Local/Temp/Rtmpp7FD3i/Rinst39f076ee/AdaJoint/sampleData/gen
```

The phenotype file is tab-delimited text file and has columns, "ID", "Y", "X1", and "X2", where "ID" is the subject id, "Y" is the case-control status, "X1" and "X2" are continuous variables. Define the list that describes the phenotype data:

```
> pheno.list <- list(file = pheno_file, delimiter = "\t", header = 1,
+     id.var = "ID", response.var = "Y", main.vars = c("X1", "X2"),
+     family = "binomial")
```

The genotype file is also a tab-delimited text file of type 2 where row 1 has the string "ldat" followed by the subject ids. The first column of this file has the SNP ids. Define the list that describes the genotype data:

```
> geno.list <- list(file = geno_file, delimiter = "\t", file.type = 2)
```

We need to choose a directory that has write access to serve as the directory where the output files will be created. For this example, let this directory be the working directory.

```
> out.dir <- getwd()
> print(out.dir)
```

```
[1] "C:/Users/wheelerb/AppData/Local/Temp/Rtmpp7FD3i/Rbuild66a91944/AdaJoint/inst/doc"
```

We also need a file that gives the SNPs belonging to each gene. Let us use the sample gene-SNP file which is a tab-delimited text file with columns "SNP" and "Gene".

```
> gs_file <- system.file("sampleData", "gene_data.txt", package = "AdaJoint")
> print(gs_file)
```

```
[1] "C:/Users/wheelerb/AppData/Local/Temp/Rtmpp7FD3i/Rinst39f076ee/AdaJoint/sampleData/gen
```

Define the list that describes this file:

```
> gs.list <- list(file = gs_file, snp.var = "SNP", gene.var = "Gene",
+     delimiter = "\t", header = 1)
```

Calling the adajoint function

Set up the options list. Let us run 100 permutations.

```
> nperm <- 100
> op.list <- list(nperm = nperm, out.dir = out.dir)
```

Set a seed and run the adajoint function

```
> set.seed(76523)
> ret <- adajoint(geno.list, pheno.list, gs.list, op = op.list)
> print(ret)
```

```
$gene.table
    Gene N.SNP    Pvalue
1 Gene_1     9 0.0297030
2 Gene_2    12 0.0792079
3 Gene_3    17 0.9801980
4 Gene_4    12 0.0495050

$pathway.pvalue
[1] 0.02970297

$most.sig.snps
     Gene     Best_1toN SNP.1      SNP.2      Cut1.Pvalue Cut2.Pvalue
[1,] "Gene_1" "2"       "rs556112" "rs460053" "0.029703"  "0.019802"
[2,] "Gene_2" "1"       "rs342947" "rs776914" "0.0693069" "0.0891089"
[3,] "Gene_3" "1"       "rs693822" "rs023502" "0.970297"  "0.990099"
[4,] "Gene_4" "1"       "rs869720" "rs970602" "0.049505"  "0.049505"

$nperm
[1] 100

$most.sig.genes
[1] "Gene_1" "Gene_2" "Gene_4"
```

```
$gene_snp
      SNP        Gene     SNP.Pvalue
 [1,] "rs651810" "Gene_1" "0.960207593955798"
 [2,] "rs601084" "Gene_1" "0.237659510218214"
 [3,] "rs873041" "Gene_1" "0.855539925151013"
 [4,] "rs483698" "Gene_1" "0.900232462836072"
 [5,] "rs710834" "Gene_1" "0.246500227992909"
 [6,] "rs556112" "Gene_1" "0.0021113488188919"
 [7,] "rs269238" "Gene_1" "0.313515115799359"
 [8,] "rs460053" "Gene_1" "0.0148384955582201"
 [9,] "rs254030" "Gene_1" "0.695369282092093"
[10,] "rs859372" "Gene_2" "0.418691653079777"
[11,] "rs013720" "Gene_2" "0.819550268260843"
[12,] "rs817723" "Gene_2" "0.625558966456398"
[13,] "rs677636" "Gene_2" "0.616350531638951"
[14,] "rs776914" "Gene_2" "0.0618204984510662"
[15,] "rs975021" "Gene_2" "0.741859815623697"
[16,] "rs342947" "Gene_2" "0.0105263558409143"
[17,] "rs620579" "Gene_2" "0.308967566845906"
[18,] "rs711504" "Gene_2" "0.142437421126461"
[19,] "rs262589" "Gene_2" "0.775259167091559"
[20,] "rs257422" "Gene_2" "0.238556154645706"
[21,] "rs014342" "Gene_2" "0.104930856181897"
[22,] "rs467442" "Gene_3" "0.638224462219944"
[23,] "rs657122" "Gene_3" "0.730419599628615"
[24,] "rs902453" "Gene_3" "0.428020828374767"
[25,] "rs023502" "Gene_3" "0.305429359218222"
[26,] "rs540465" "Gene_3" "0.507549112428912"
[27,] "rs693822" "Gene_3" "0.179107937692084"
[28,] "rs045071" "Gene_3" "0.516684210706346"
[29,] "rs081827" "Gene_3" "0.346352744177337"
[30,] "rs683489" "Gene_3" "0.724359542281126"
[31,] "rs350111" "Gene_3" "0.598145192096028"
[32,] "rs845434" "Gene_3" "0.351809496697702"
[33,] "rs247716" "Gene_3" "0.576971587494581"
[34,] "rs711571" "Gene_3" "0.437678658851229"
[35,] "rs568502" "Gene_3" "0.35674867418997"
[36,] "rs274103" "Gene_3" "0.602468646399642"
[37,] "rs943580" "Gene_3" "0.826058793691032"
[38,] "rs907292" "Gene_3" "0.516538307160806"
[39,] "rs869720" "Gene_4" "0.00213342091675384"
[40,] "rs111546" "Gene_4" "0.0969369097824907"
[41,] "rs679512" "Gene_4" "0.695955480507909"
[42,] "rs523523" "Gene_4" "0.440556562474894"
[43,] "rs789349" "Gene_4" "0.241719651601861"
[44,] "rs265436" "Gene_4" "0.758957618007028"
[45,] "rs115265" "Gene_4" "0.859669627000873"
[46,] "rs304539" "Gene_4" "0.601450111630443"
[47,] "rs970602" "Gene_4" "0.031269538377777"
[48,] "rs707881" "Gene_4" "0.343469044541856"
```

```
[49,] "rs732255" "Gene_4" "0.0391709509640392"
[50,] "rs090634" "Gene_4" "0.527107010530066"

$remove.n.subjects
[1] 0
```

Now compute the pathway p-value assuming all the SNPs belong to the same gene. Note that if gene.list is NULL, then the program assumes all SNPs belong to the same gene.

```
> set.seed(76523)
> ret <- adajoint(geno.list, pheno.list, NULL, op = op.list)
> print(ret)

$gene.table
    Gene N.SNP   Pvalue
1 gene_1    50 0.108911

$pathway.pvalue
[1] 0.1089109

$most.sig.snps
     Gene     Best_1toN SNP.1      SNP.2      Cut1.Pvalue Cut2.Pvalue
[1,] "gene_1" "2"       "rs556112" "rs869720" "0.138614"  "0.0792079"

$nperm
[1] 100

$most.sig.genes
[1] "gene_1"

$gene_snp
        SNP   Gene Group  SNP.Pvalue
1  rs013720 gene_1     1 0.819550268
2  rs014342 gene_1     1 0.104930856
3  rs023502 gene_1     1 0.305429359
4  rs045071 gene_1     1 0.516684211
5  rs081827 gene_1     1 0.346352744
6  rs090634 gene_1     1 0.527107011
7  rs111546 gene_1     1 0.096936910
8  rs115265 gene_1     1 0.859669627
9  rs247716 gene_1     1 0.576971587
10 rs254030 gene_1     1 0.695369282
11 rs257422 gene_1     1 0.238556155
12 rs262589 gene_1     1 0.775259167
13 rs265436 gene_1     1 0.758957618
14 rs269238 gene_1     1 0.313515116
15 rs274103 gene_1     1 0.602468646
16 rs304539 gene_1     1 0.601450112
17 rs342947 gene_1     1 0.010526356
18 rs350111 gene_1     1 0.598145192
```

```
19 rs460053 gene_1     1 0.014838496
20 rs467442 gene_1     1 0.638224462
21 rs483698 gene_1     1 0.900232463
22 rs523523 gene_1     1 0.440556562
23 rs540465 gene_1     1 0.507549112
24 rs556112 gene_1     1 0.002111349
25 rs568502 gene_1     1 0.356748674
26 rs601084 gene_1     1 0.237659510
27 rs620579 gene_1     1 0.308967567
28 rs651810 gene_1     1 0.960207594
29 rs657122 gene_1     1 0.730419600
30 rs677636 gene_1     1 0.616350532
31 rs679512 gene_1     1 0.695955481
32 rs683489 gene_1     1 0.724359542
33 rs693822 gene_1     1 0.179107938
34 rs707881 gene_1     1 0.343469045
35 rs710834 gene_1     1 0.246500228
36 rs711504 gene_1     1 0.142437421
37 rs711571 gene_1     1 0.437678659
38 rs732255 gene_1     1 0.039170951
39 rs776914 gene_1     1 0.061820498
40 rs789349 gene_1     1 0.241719652
41 rs817723 gene_1     1 0.625558966
42 rs845434 gene_1     1 0.351809497
43 rs859372 gene_1     1 0.418691653
44 rs869720 gene_1     1 0.002133421
45 rs873041 gene_1     1 0.855539925
46 rs902453 gene_1     1 0.428020828
47 rs907292 gene_1     1 0.516538307
48 rs943580 gene_1     1 0.826058794
49 rs970602 gene_1     1 0.031269538
50 rs975021 gene_1     1 0.741859816

$remove.n.subjects
[1] 0
```