ARTP(Adaptive Rank Truncated Product) Package

April 5, 2018

```
> library(ARTP)
```

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

[1] "/spin1/scratch/wheelerwi/RtmpxTCecE/Rinst18cc118f6469/ARTP/sampleData/pheno_data.txt"

> print(geno_file)

[1] "/spin1/scratch/wheelerwi/RtmpxTCecE/Rinst18cc118f6469/ARTP/sampleData/geno_data.txt"
```

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"))
```

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] "/spin1/scratch/wheelerwi/RtmpxTCecE/Rbuild18cc42c653bd/ARTP/vignettes"
```

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_SNP_data.txt", package="ARTP")
> print(gs_file)
```

```
[1] "/spin1/scratch/wheelerwi/RtmpxTCecE/Rinst18cc118f6469/ARTP/sampleData/gene_SNP_data.t
```

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 runPermutations and ARTP_pathway functions

Define the names of the 2 output files that will store the observed p-values and permutated p-values.

```
> obs.outfile  <- paste(out.dir, "/", "obs.txt", sep="")
> perm.outfile <- paste(out.dir, "/", "perm.txt", sep="")
```

Set up the options list. Let us run 50 permutations and choose to generate a new response vector for each permutation (perm.method=2).

```
> nperm   <- 50
> op.list <- list(nperm=nperm, obs.outfile=obs.outfile, perm.outfile=perm.outfile, perm.me
```

Run the permutations. The base (NULL) model summary will be printed to the console.

```
> runPermutations(geno.list, pheno.list, 1, op=op.list)
```

```
Call:
glm(formula = response0 ~ phenoData0[, -snpcol] - 1, family = family,
    model = FALSE, x = TRUE, y = TRUE)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.168  -1.128  -1.097   1.227   1.272

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
phenoData0[, -snpcol]x1  -0.1115     0.2396  -0.465    0.642
phenoData0[, -snpcol]x2   0.1024     0.3087   0.332    0.740
phenoData0[, -snpcol]x3  -0.1211     0.3168  -0.382    0.702

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 693.15  on 500  degrees of freedom
```

```
Residual deviance: 691.08  on 497  degrees of freedom
AIC: 697.08

Number of Fisher Scoring iterations: 3

NULL
```

Now we have the observed p-values and permutated p-values stored in the files obs.outfile and perm.outfile so that we can compute the gene and pathway p-values by using the default parameters for op (see the manual for details).

```
> set.seed(76523)
> ret <- ARTP_pathway(obs.outfile, perm.outfile, nperm, out.dir, gene.list=gs.list)
> print(ret)

$pathway.pvalue
[1] 0.01960784

$gene.table
    Gene N.SNP    Pvalue
1 Gene_3    17 1.0000000
2 Gene_4    12 0.0980392
3 Gene_1     9 0.0196078
4 Gene_2    12 0.1764710

$nperm
[1] 50
```

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 <- ARTP_pathway(obs.outfile, perm.outfile, nperm, out.dir)
> print(ret)

$pathway.pvalue
[1] 0.2156863

$gene.table
  Gene N.SNP   Pvalue
1 gene    50 0.215686

$nperm
[1] 50
```