Reproducing the example in the paper

1. Do the following to create the R workspace smoking.RData.

```
# The data used to illustrate the method implemented in mvama was discussed in:
# Spira, A. et al. (2004). Effects of cigarette smoke on the human
# airway epithelial cell transcriptome. Proc. Natl Acad. Sci. USA,
# 101, 10143-10148.

# The raw data may be obtained from the NCBI at
# http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE994. Only CEL
# files corresponding to 'smokers' and 'never smokers were used in the
# paper; these correspond to Accession IDs GSM15684-15717
# and GSM15718-15740, respectively.

# Once the compressed CEL files (GSM15XYZ.CEL.gz) have been downloaded
# to a directory, they can be read in and rma-corrected as follows:

# Install the following Bioconductor (http://www.bioconductor.org)
# packages if required:
require(affy)
require(limma)
require(hgu133acdf)

data <- ReadAffy(celfile.path = "/path_to_CEL_files", compress = TRUE)

# Carry out rma correction
rma.data <- rma(data)

# Note that this is equivalent to using the function 'expresso' as
# follows:
#rma.data <- expresso(data,
#                     bgcorrect.method = "rma",
#                     normalize.method = "quantiles",
#                     pmcorrect.method = "pmonly",
#                     summary.method = "medianpolish")

# And finally,
rma.data <- t(exprs(rma.data))
# which gives a 57 x 22283 matrix where the first 34 rows correspond
# to 'smokers', and the next 23 correspond to 'never
# nmokers'. Consequently, the vector of class labels y can be
# generated as
y <- rep(1:2, times = c(34, 23))
# generate the design matrix
D<-matrix(0,nrow=57,ncol=2)
D[,1]<-as.numeric(y==2)
D[,2]<-as.numeric(y==1)
```

```
# save to smoking.RData
x<-rma.data
save(x,y,D,file="smoking.RData")
```

The first 5 rows and columns of x should be

```
     X1007_s_at X1053_at  X117_at  X121_at X1255_g_at
[1,]   9.729530 4.169440 5.599752 8.113206   4.256429
[2,]  10.172280 4.381633 5.166447 8.025542   3.807625
[3,]   9.801157 4.335081 5.696982 8.079880   3.982507
[4,]  10.084170 4.025541 5.193149 7.842162   3.951598
[5,]   9.765211 4.142204 5.378281 7.858246   3.582457
```

2. Unpack the file BMCex.zip into the same directory as smoking.RData.
You should see the following files
perm1.script.r
perm2.script.r
perm3.script.r
perm4.script.r
perm5.script.r
perm.collate.r
mvama.RData

3. Get the permutation distributions. Make sure that the mvama library is installed in R on each processor. On separate processors or in a sequence of jobs on a single processor invoke the scripts

```
R --vanilla<perm1.script.r>perm.out1&
R --vanilla<perm2.script.r>perm.out2&
R --vanilla<perm3.script.r>perm.out3&
R --vanilla<perm4.script.r>perm.out4&
R --vanilla<perm5.script.r>perm.out5&
```

or inside R

source("perm1.script.r") , wait until the job completes then
source("perm2.script.r") etc (each job may take a couple of hours)

If necessary move the output files from the above into the same directory as smoking.RData.

3. Start R in the same directory as smoking.RData and source the script perm.collate.r. Plots will most likely have a different layout to those in the paper but are the same otherwise.

Note that the fitted inverse covariance matrix is given in mvama.RData.
The steps to reproduce this can be found on the page for the package sparse.inv.cov at www.bioinformatics.csiro.au/sparse.inv.cov/index.shtml.