

# HEM

November 11, 2009

## R topics documented:

am.tran.half . . . . .	2
am.tran . . . . .	2
base.ASE.Olig . . . . .	2
base.error.Olig.quanOnly . . . . .	2
base.error.Olig . . . . .	3
base.PSE.Olig . . . . .	3
boot.base.ASE.Olig . . . . .	3
boot.base.error.Olig . . . . .	3
boot.base.PSE.Olig . . . . .	4
fixbound.predict.smooth.spline . . . . .	4
hem.eb.prior . . . . .	4
hem.fdr . . . . .	6
hem.null.no . . . . .	8
hem.null.one . . . . .	8
hem.null.two . . . . .	8
hem.preproc . . . . .	9
hem . . . . .	9
mubcp . . . . .	12
nonpar.error.Olig . . . . .	12
nonpar.no.error.Olig . . . . .	12
nonpar.rep.error.Olig . . . . .	13
par.error.Olig . . . . .	13
par.no.error.Olig . . . . .	13
par.rep.error.Olig . . . . .	13
pbrain . . . . .	14
permut . . . . .	14
quant.normal2 . . . . .	14
quant.normalize . . . . .	15
quant.normal . . . . .	15
quant.norm . . . . .	15
remove.sig.genes . . . . .	15
<b>Index</b>	<b>16</b>

---

`am.tran.half`*AM transformation for LPE*

---

**Description**

Computes AM for LPE

**Author(s)**HyungJun Cho and Jae K. Lee

---

`am.tran`*AM transformation for LPE*

---

**Description**

Computes AM for LPE

**Author(s)**HyungJun Cho and Jae K. Lee

---

`base.ASE.Olig`*Baseline ASE estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**HyungJun Cho and Jae K. Lee

---

`base.error.Olig.quantOnly`*Baseline error estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

base.error.Olig      *Baseline error estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

base.PSE.Olig      *Baseline PSE estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

boot.base.ASE.Olig      *Baseline error bootstrap estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error using bootstrap samples for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

boot.base.error.Olig      *Baseline error bootstrap estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error using bootstrap samples for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

```
boot.base.PSE.Olig
```

*Baseline error bootstrap estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error using bootstrap samples for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

```
fixbound.predict.smooth.spline
```

*Prediction using smoothing spline*

---

**Description**

Makes predictions using smoothing spline

**Author(s)**

HyungJun Cho and Jae K. Lee

---

```
hem.eb.prior
```

*Empirical Bayes (EB) Prior Specification*

---

**Description**

Estimates experimental and biological variances by LPE and resampling

**Usage**

```
hem.eb.prior(dat, n.layer, design,
             method.var.e="neb", method.var.b="peb", method.var.t="neb",
             rep=TRUE, baseline.var="LPE", p.remove=0, max.chip=4,
             q=0.01, B=25, n.digits=10, print.message.on.screen=TRUE)
```

**Arguments**

dat	data
n.layer	number of layers
design	design matrix
method.var.e	prior specification method for experimental variance; "peb"=parametric EB prior specification, "neb"=nonparametric EB prior specification
method.var.b	prior specification method for biological variance; "peb"=parametric EB prior specification

```

method.var.t prior specification method for total variance; "peb"=parametric EB prior speci-
              fication, "neb"=nonparametric EB prior specification
rep          no replication if FALSE
baseline.var baseline variance estimation method: LPE for replicated data and BLPE, PSE,
              or ASE for unreplicated data
p.remove     percent of removed rank-variance genes for BLPE
max.chip     maximum number of chips to estimate errors
q           quantile for partitioning genes based on expression levels
B           number of iterations for resampling
n.digits     number of digits
print.message.on.screen
              if TRUE, process status is shown on screen.

```

**Value**

```

var.b       prior estimate matrix for biological variances (n.layer=2)
var.e       prior estimate matrix for experientnal variances (n.layer=2)
var.t       prior estimate matrix for total variances (n.layer=1)

```

**Author(s)**

HyungJun Cho and Jae K. Lee

**See Also**

[hem](#), [hem.fdr](#)

**Examples**

```

#Example 1: Two-layer HEM with EB prior specification

data(pbrain)

##construct a design matrix
cond <- c(1,1,1,1,1,1,2,2,2,2,2,2)
ind  <- c(1,1,2,2,3,3,1,1,2,2,3,3)
rep  <- c(1,2,1,2,1,2,1,2,1,2,1,2)
design <- data.frame(cond,ind,rep)

##normalization
pbrain.nor <- hem.preproc(pbrain[,2:13])

##take a subset for a testing purpose;
##use all genes for a practical purpose
pbrain.nor <- pbrain.nor[1:1000,]

##estimate hyperparameters of variances by LPE
#pbrain.eb <- hem.eb.prior(pbrain.nor, n.layer=2, design=design,
#                          method.var.e="neb", method.var.b="peb")

#fit HEM with two layers of error

```

```

#using the small numbers of burn-ins and MCMC samples for a testing purpose;
#but increase the numbers for a practical purpose
#pbrain.hem <- hem(pbrain.nor, n.layer=2, design=design, burn.ins=10, n.samples=30,
#               method.var.e="neb", method.var.b="peb",
#               var.e=pbrain.eb$var.e, var.b=pbrain.eb$var.b)

#Example 2: One-layer HEM with EB prior specification

data(mubcp)

##construct a design matrix
cond <- c(rep(1,6), rep(2,5), rep(3,5), rep(4,5), rep(5,5))
ind  <- c(1:6, rep((1:5),4))
design <- data.frame(cond, ind)

##normalization
mubcp.nor <- hem.preproc(mubcp)

##take a subset for a testing purpose;
##use all genes for a practical purpose
mubcp.nor <- mubcp.nor[1:1000,]

##estimate hyperparameters of variances by LPE
#mubcp.eb <- hem.eb.prior(mubcp.nor, n.layer=1, design=design,
#                        method.var.t="neb")

#fit HEM with two layers of error
#using the small numbers of burn-ins and MCMC samples for a testing purpose;
#but increase the numbers for a practical purpose
#mubcp.hem <- hem(mubcp.nor, n.layer=1, design=design, burn.ins=10, n.samples=30,
#                method.var.t="neb", var.t=mubcp.eb$var.t)

```

---

hem.fdr

*FDR Evaluation*


---

## Description

Computes resampling-based False Discovery Rate (FDR)

## Usage

```

hem.fdr(dat, n.layer, design, rep=TRUE, hem.out, eb.out=NULL, n.iter=5, q.trim=
        target.fdr=c(0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.20, 0.30, 0.40, 0.50)
        n.digits=10, print.message.on.screen=TRUE)

```

## Arguments

dat	data
n.layer	number of layers: 1=one-layer EM; 2=two-layer EM
design	design matrix
rep	no replication if FALSE
hem.out	output from hem function

eb.out            output from hem.eb.prior function  
 n.iter            number of iterations  
 q.trim            quantile used for estimating the proportion of true negatives ( $\pi_0$ )  
 target.fdr        Target FDRs  
 n.digits          number of digits  
 print.message.on.screen  
                   if TRUE, process status is shown on screen.

**Value**

fdr                H-values and corresponding FDRs  
 pi0                estimated proportion of true negatives  
 H.null             H-scores from null data  
 targets            given target FDRs, corresponding critical values and numbers of significant genes are provided

**Author(s)**

HyungJun Cho and Jae K. Lee

**See Also**

[hem.eb.prior](#) [hem](#)

**Examples**

```

data(pbrain)

##construct a design matrix
cond <- c(1,1,1,1,1,1,2,2,2,2,2,2)
ind  <- c(1,1,2,2,3,3,1,1,2,2,3,3)
rep  <- c(1,2,1,2,1,2,1,2,1,2,1,2)
design <- data.frame(cond,ind,rep)

##normalization
pbrain.nor <- hem.preproc(pbrain[,2:13])

##take a subset for a testing purpose;
##use all genes for a practical purpose
pbrain.nor <- pbrain.nor[1:1000,]

##estimate hyperparameters of variances by LPE
#pbrain.eb <- hem.eb.prior(pbrain.nor, n.layer=2, design=design,
#                           method.var.e="neb", method.var.b="peb")

##fit HEM with two layers of error
##using the small numbers of burn-ins and MCMC samples for a testing purpose;
##but increase the numbers for a practical purpose
#pbrain.hem <- hem(pbrain.nor, n.layer=2, design=design,burn.ins=10, n.samples=30,
#                  method.var.e="neb", method.var.b="peb",
#                  var.e=pbrain.eb$var.e, var.b=pbrain.eb$var.b)

```

```
##Estimate FDR based on resampling
#pbrain.fdr <- hem.fdr(pbrain.nor, n.layer=2, design=design,
#                      hem.out=pbrain.hem, eb.out=pbrain.eb)
```

---

hem.null.no                    *Generation of null data*

---

**Description**

Generates null data by resampling

**Author(s)**

HyungJun Cho and Jae K. Lee

---

hem.null.one                    *Generation of null data*

---

**Description**

Generates null data by resampling

**Author(s)**

HyungJun Cho and Jae K. Lee

---

hem.null.two                    *Generation of null data*

---

**Description**

Generates null data by resampling

**Author(s)**

HyungJun Cho and Jae K. Lee



---

hem.preproc	<i>Preprocessing</i>
-------------	----------------------

---

**Description**

Performs IQR normalization, thesholding, and log2-transformation

**Usage**

```
hem.preproc(x, data.type = "MAS5")
```

**Arguments**

x	data
data.type	data type: MAS5 or MAS4

**Author(s)**

HyungJun Cho and Jae K. Lee

**See Also**

[hem](#), [hem.eb.prior](#), [hem.fdr](#)

**Examples**

```
library(HEM)

data(pbrain)
pbrain.nor <- hem.preproc(pbrain[,2:13])
```

---

hem	<i>Heterogeneous Error Model for Identification of Differential Expressed Genes Under Multiple Conditions</i>
-----	---

---

**Description**

Fits an error model with heterogeneous experimental and biological variances.

**Usage**

```
hem(dat, probe.ID=NULL, n.layer, design, burn.ins=1000, n.samples=3000,
     method.var.e="gam", method.var.b="gam", method.var.t="gam",
     var.e=NULL, var.b=NULL, var.t=NULL, var.g=1, var.c=1, var.r=1,
     alpha.e=3, beta.e=.1, alpha.b=3, beta.b=.1, alpha.t=3, beta.t=.2,
     n.digits=10, print.message.on.screen=TRUE)
```

**Arguments**

<code>dat</code>	data
<code>probe.ID</code>	a vector of probe set IDs
<code>n.layer</code>	number of layers; 1=one-layer EM, 2=two-layer EM
<code>design</code>	design matrix
<code>burn.ins</code>	number of burn-ins for MCMC
<code>n.samples</code>	number of samples for MCMC
<code>method.var.e</code>	prior specification method for experimental variance; "gam"=Gamma(alpha,beta), "peb"=parametric EB prior specification, "neb"=nonparametric EB prior specification
<code>method.var.b</code>	prior specification method for biological variance; "gam"=Gamma(alpha,beta), "peb"=parametric EB prior specification
<code>method.var.t</code>	prior specification method for total variance; "gam"=Gamma(alpha,beta), "peb"=parametric EB prior specification, "neb"=nonparametric EB prior specification
<code>var.e</code>	prior estimate matrix for experimental variance
<code>var.b</code>	prior estimate matrix for biological variance
<code>var.t</code>	prior estimate matrix for total variance
<code>var.g</code>	$N(0, \text{var.g})$ ; prior parameter for gene effect
<code>var.c</code>	$N(0, \text{var.c})$ ; prior parameter for condition effect
<code>var.r</code>	$N(0, \text{var.r})$ ; prior parameter for interaction effect of gene and condition
<code>alpha.e, beta.e</code>	$\text{Gamma}(\text{alpha.e}, \text{alpha.e})$ ; prior parameters for inverse of experimental variance
<code>alpha.b, beta.b</code>	$\text{Gamma}(\text{alpha.b}, \text{alpha.b})$ ; prior parameters for inverse of biological variance
<code>alpha.t, beta.t</code>	$\text{Gamma}(\text{alpha.b}, \text{alpha.b})$ ; prior parameters for inverse of total variance
<code>n.digits</code>	number of digits
<code>print.message.on.screen</code>	if TRUE, process status is shown on screen.

**Value**

<code>n.gene</code>	number of genes
<code>n.chip</code>	number of chips
<code>n.cond</code>	number of conditions
<code>design</code>	design matrix
<code>burn.ins</code>	number of burn-ins for MCMC
<code>n.samples</code>	number of samples for MCMC
<code>priors</code>	prior parameters
<code>m.mu</code>	estimated mean expression intensity for each gene under each condition
<code>m.x</code>	estimated unobserved expression intensity for each combination of genes, conditions, and individuals (n.layer=2)
<code>m.var.b</code>	estimated biological variances (n.layer=2)
<code>m.var.e</code>	estimated experimental variances (n.layer=2)
<code>m.var.t</code>	estimated total variances (n.layer=1)
<code>H</code>	H-scores

**Author(s)**

HyungJun Cho and Jae K. Lee

**References**

Cho, H. and Lee, J.K. (2004) Bayesian Hierarchical Error Model for Analysis of Gene Expression Data, *Bioinformatics*, 20: 2016-2025.

**See Also**

[hem.eb.prior](#), [hem.fdr](#)

**Examples**

```
#Example 1: Two-layer HEM

data(pbrain)

##construct a design matrix
cond <- c(1,1,1,1,1,1,2,2,2,2,2,2) #condition
ind <- c(1,1,2,2,3,3,1,1,2,2,3,3) #biological replicate
rep <- c(1,2,1,2,1,2,1,2,1,2,1,2) #experimental replicate
design <- data.frame(cond,ind,rep)

##normalization
pbrain.nor <- hem.preproc(pbrain[,2:13])

##fit HEM with two layers of error
##using the small numbers of burn-ins and MCMC samples for a testing purpose;
##but increase the numbers for a practical purpose
#pbrain.hem <- hem(pbrain.nor, n.layer=2, design=design,
#                 burn.ins=10, n.samples=30)

##print H-scores
#pbrain.hem$H

#Example 2: One-layer HEM

data(mubcp)

##construct a design matrix
cond <- c(rep(1,6),rep(2,5),rep(3,5),rep(4,5),rep(5,5))
ind <- c(1:6,rep((1:5),4))
design <- data.frame(cond,ind)

##construct a design matrix
mubcp.nor <- hem.preproc(mubcp)

##fit HEM with one layers of error
##using the small numbers of burn-ins and MCMC samples for a testing purpose;
##but increase the numbers for a practical purpose
#mubcp.hem <- hem(mubcp.nor, n.layer=1,design=design, burn.ins=10, n.samples=30)

##print H-scores
#mubcp.hem$H
```

```
###NOTE: Use 'hem.fdr' for FDR evaluation
###NOTE: Use 'hem.eb.prior' for Empirical Bayes (EB) prior sepecification
###NOTE: Use EB-HEM ('hem' after 'hem.eb.prior') for small data sets
```

---

mubcp

*Gene expression data for mouse B cell development*


---

### Description

This data set consists of gene expression of the five consecutive stages (pre-B1, large pre-B2, small pre-B2, immature B, and mature B cells) of mouse B cell development. The data were obtained with high-density oligonucleotide arrays, Affymetrix Mu11k GeneChips, from flow-cytometrically purified cells.

### Usage

data(mubcp)

### Format

A matrix containing 13,207 probe sets and 26 chips; first 6 chips for pre-B1 cell and next 20 chips for other stages (5 chips for each)

### Source

Hoffmann, R., Seidl, T., Neeb, M., Rolink, A. and Melchers, F. (2002). Changes in gene expression profiles in developing B cells of murine bone marrow, *Genome Research* 12:98-111.

---

nonpar.error.Olig *Baseline error nonparametric estimation for oligonucleotide arrays*


---

### Description

Estimates baseline error for oligonucleotide arrays

### Author(s)

HyungJun Cho and Jae K. Lee

---

nonpar.no.error.Olig *Baseline error nonparametric estimation for oligonucleotide arrays*


---

### Description

Estimates baseline error for oligonucleotide arrays

### Author(s)

HyungJun Cho and Jae K. Lee

---

nonpar.rep.error.Olig *Baseline error nonparametric estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

par.error.Olig *Baseline error parametric estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

par.no.error.Olig *Baseline error parametric estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

par.rep.error.Olig *Baseline error parametric estimation for oligonucleotide arrays*

---

**Description**

Estimates baseline error for oligonucleotide arrays

**Author(s)**

HyungJun Cho and Jae K. Lee

---

pbrain

*Gene expression data for primate brains*

---

### Description

This data set consists of gene expression of primate brains (Affymetrix U95A GeneChip). The frozen brains of three humans (H1, H2, H3) and three chimpanzees (C1, C2, C3) were used to take the postmortem tissue samples, and two independent tissue samples for each individual were taken.

### Usage

```
data(pbrain)
```

### Format

A matrix containing 12,600 probe sets and 12 chips (H1,H1,H2,H2,H3,H3,C1,C1,C2,C2,C3,C3); the first column is probe set ID

### Source

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.R., and Paabo, S. (2002) Intra- and interspecific variation in primate gene expression patterns, *Science* 296:340-343

---

permut

*Permutation*

---

### Description

Permute

### Author(s)

HyungJun Cho and Jae K. Lee

---

quant.normal2

*Normalization*

---

### Description

Normalization

### Author(s)

HyungJun Cho and Jae K. Lee

---

quant.normalize      *Quantile normalization*

---

**Description**

Performs quantile normalization

**Author(s)**

HyungJun Cho and Jae K. Lee

---

quant.normal      *Normalization*

---

**Description**

Normalization

**Author(s)**

HyungJun Cho and Jae K. Lee

---

quant.norm      *Quantile normalization*

---

**Description**

Performs quantile normalization

**Author(s)**

HyungJun Cho and Jae K. Lee

---

remove.sig.genes      *Remove significant genes*

---

**Description**

Remove significant genes

**Author(s)**

HyungJun Cho and Jae K. Lee

# Index

## \*Topic **datasets**

mubcp, 11  
pbrain, 13

## \*Topic **models**

am.tran, 1  
am.tran.half, 1  
base.ASE.Olig, 1  
base.error.Olig, 2  
base.error.Olig.quanOnly, 2  
base.PSE.Olig, 2  
boot.base.ASE.Olig, 2  
boot.base.error.Olig, 3  
boot.base.PSE.Olig, 3  
fixbound.predict.smooth.spline,  
3  
hem, 9  
hem.eb.prior, 4  
hem.fdr, 6  
hem.null.no, 7  
hem.null.one, 7  
hem.null.two, 8  
hem.preproc, 8  
nonpar.error.Olig, 12  
nonpar.no.error.Olig, 12  
nonpar.rep.error.Olig, 12  
par.error.Olig, 12  
par.no.error.Olig, 13  
par.rep.error.Olig, 13  
permut, 14  
quant.norm, 15  
quant.normal, 14  
quant.normal2, 14  
quant.normalize, 14  
remove.sig.genes, 15

am.tran, 1  
am.tran.half, 1

base.ASE.Olig, 1  
base.error.Olig, 2  
base.error.Olig.quanOnly, 2  
base.PSE.Olig, 2  
boot.base.ASE.Olig, 2  
boot.base.error.Olig, 3

boot.base.PSE.Olig, 3

fixbound.predict.smooth.spline,  
3

hem, 5, 6, 8, 9  
hem.eb.prior, 4, 6, 8, 10  
hem.fdr, 5, 6, 8, 10  
hem.null.no, 7  
hem.null.one, 7  
hem.null.two, 8  
hem.preproc, 8

mubcp, 11

nonpar.error.Olig, 12  
nonpar.no.error.Olig, 12  
nonpar.rep.error.Olig, 12

par.error.Olig, 12  
par.no.error.Olig, 13  
par.rep.error.Olig, 13  
pbrain, 13  
permut, 14

quant.norm, 15  
quant.normal, 14  
quant.normal2, 14  
quant.normalize, 14

remove.sig.genes, 15