Package 'GAprediction'

May 8, 2026

Date 2016-07-04

Type Package

Title Prediction of gestational age with Illumina HumanMethylation450 data

Version 1.39.0

Author Jon Bohlin

Maintainer Jon Bohlin <jon.bohlin@gmail.com>

Description [GAprediction] predicts gestational age using Illumina HumanMethylation450 CpG data.

License GPL (>=2)

LazyData TRUE

Depends R (>= 3.3)

Imports glmnet, stats, utils, Matrix

biocViews ImmunoOncology, DNAMethylation, Epigenetics, Regression, BiomedicalInformatics

Suggests knitr, rmarkdown

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/GAprediction

git_branch devel

git_last_commit fa3e70b

git_last_commit_date 2026-04-28

Repository Bioconductor 3.24

Date/Publication 2026-05-08

Contents

`extractSites` *Extract CpG sites for gestational age prediction*

Description

The function allows the user to extract CpG sites used for gestational age prediction with the function `predictGA`.

Usage

```
extractSites(type="se")
```

Arguments

The argument type=c("se", "min", "all") can be used to specify which CpGs are to be extracted. "se" designates the CpGs needed by the predictGA function if the penalty term lambda is to be set to one standard error within the minimum, "min" specifies the minimum lambda, while "all" returns the complete sets of CpGs in the `UL.mod.cv` object.

`type`	- a string that can be "se" (default), "min" or "all", depending on which CpGs is wanted by the user.

Details

Use this function if predictGA fails due to missing predictor CpGs, or to see which CpGs are used by `predictGA` for gestational age prediction.

Value

Returns a vector with the requested CpG sites.

Author(s)

Jon Bohlin

See Also

`predictGA`, `UL.mod.cv`

Examples

```
CpGs <- extractSites( type="se" )
```

predictGA *Predict gestational age in days from conception*

Description

The function predictGA takes a matrix with Illumina HumanMethylation450 type DNA methylation data. Column names must designate CpG sites (i.e. 'cgXXXXXX', X=number) and row names samples IDs.

Usage

```
predictGA(mldat, transp=TRUE, se=TRUE)
```

Arguments

mldat	A matrix containing DNA methylation beta values (0<=beta<=1)
transp	If TRUE (default), the transpose is automatically taken if the number of rows is greater than the number of columns.
se	If se=TRUE, the estimated coefficients are based on the prediction model with the lambda penalty term being allowed to vary up to one standard error within the minimum. If se=FALSE, the minimum lambda is assumed.

Details

The minimum lambda (se=FALSE) may result in slightly better predictions, however substantially more CpG sites are needed for estimation. Since the prediction difference is hardly noticeable se=TRUE is the default option.

Value

The function returns estimated gestational age predictions, together with samples IDs as row names, in a data.frame object.

Note

Requires quite a bit of memory due to the large DNA methylation matrix required for the prediction model.

Author(s)

Jon Bohlin

References

Jon Bohlin, Siri E. Haaberg, Per Magnus, et al. (2016). Prediction of gestational age based on genome-wide differentially methylated regions. Genome Biology (in review)

Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL http://www.jstatsoft.org/v33/i01/.

Examples

```
## Make a mock Illumina HumanMethylation450 type DNA methylation matrix
cpgs <- extractSites( type="se" )
allcpgs <- extractSites( type="all" )
numsamples <- 100
mlmatr <- matrix( NA, ncol=length( allcpgs ), nrow=numsamples )
mlmatr <- data.frame( mlmatr )
for( i in cpgs )
  mlmatr[,i] <- runif( numsamples, min=0, max=1 )
## Perform gestational age prediction
mypred <- predictGA( mlmatr )
```

`UL.mod.cv` *A glmnet-object trained to perform gestational age prediction.*

Description

The glmnet-object consists of a Lasso-regression model 'trained' to perform gestational age predictions. It is called by the wrapper function `predictGA`, which is more user-friendly.

Details

The trained Lasso-model contains cross-validated estimates of the penalty term lambda that regulates the number of CpG sites needed for gestational age prediction. It is called by the glmnet-inherited predict function with a matrix of CpG betas (with values between 0 and 1) that conforms to the Illumina HumanMethylation450 platform. The gestational age estimates used to train the regression model were taken from the MoBa cohort and are based on ultrasound.

Source

Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C, MoBa Study Group. Cohort profile: the Norwegian mother and child cohort study (MoBa). International journal of epidemiology. 2006 Oct 1;35(5):1146-50.

References

Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL http://www.jstatsoft.org/v33/i01/.

Examples

```
## Extract all non-zero regression coefficients
temp <- as.matrix( coef( UL.mod.cv ) )
allNonZeroCoefs <- rownames( temp )[ temp[,1]!=0 ]
allNonZeroCoefs[ -1 ]
```

Index