Package 'ChAMP'

October 12, 2016

Type Package

Title Chip Analysis Methylation Pipeline for Illumina
HumanMethylation450 and EPIC

Version 1.10.0

Date 2016-04-25

Description The package includes quality control metrics, a selection
of normalization methods and novel methods to identify
differentially methylated regions and to highlight copy number
alterations. In addition there is a method to help calculate hmC using BS and oxBS samples.

License GPL-3

Depends R (>= 3.2), minfi, ChAMPdata, Illumina450ProbeVariants.db

Imports sva, IlluminaHumanMethylation450kmanifest, limma,RPMM,
DNAcopy, preprocessCore, impute, marray, wateRmelon, plyr,
GenomicRanges,RefFreeEWAS,qvalue,isva,doParallel,bumphunter,quadprog

biocViews Microarray, MethylationArray, Normalization, TwoChannel,
CopyNumber, DNAMethylation

Author Tiffany Morris [cre, aut],
Yuan Tian [aut],
Lee Stirling [ctb],
Andrew Feber [ctb],
Andrew Teschendorff [ctb],
Ankur Chakravarthy [ctb]

Maintainer Yuan Tian, Tiffany Morris <champ450k@gmail.com>

NeedsCompilation no

R topics documented:

ChAMP-package   *ChAMP-Chip Analysis Methylation Pipeline*

Description

A pipeline that enables pre-processing of 450k data, a selection of normalization methods and novel methods for downstream analysis including Probe Lasso DMR Hunter and Copy Number Alteration analysis.

Details

Package:	ChAMP
Type:	Package
Version:	1.9.2
Date:	2016-03-13
License:	GPL-3

The full analysis pipeline can be run with all defaults using champ.process()

Alternatively, it can be run in steps using all functions separately.

Author(s)

Tiffany Morris, Lee Stirling, Andy Feber, Andrew Teschendorff, Ankur Chakravarthy, Yuan Tian, Stephen Beck
Maintainer: Yuan Tian, Tiffany Morris <champ450k@gmail.com>

Examples

```
directory=system.file('extdata',package='ChAMPdata')
champ.process(directory=directory)
myLoad=champ.load()
myNorm=champ.norm()
    myRefBase=champ.refbase()
champ.SVD()
batchNorm=champ.runCombat()
limma=champ.MVP()
```

```
myDMR=champ.DMR()
champ.CNA()
```

---

champ.CNA   *Inference of Copy Number Abberrations from intensity values.*

Description

This function enables CNA profiles to be built using methylation data from Illumina HumanMethylation450 BeadChips.

Usage

```
champ.CNA(intensity = myLoad$intensity, pd = myLoad$pd, loadFile = FALSE, batchCorrect = TRUE,
file = "intensity.txt", resultsDir = paste(getwd(), "resultsChamp", sep = "/"),
sampleCNA=TRUE, plotSample=TRUE, filterXY = TRUE, groupFreqPlots=TRUE,freqThreshold=0.3,
control=TRUE,controlGroup="Control",arraytype="450K")
```

Arguments

intensity	A matrix of intensity values for each sample. The default assumes you ran champ.load and saved the output to "myLoad".
pd	This data.frame includes the information from the sample sheet. The default assumes you ran champ.load and saved the output to "myLoad".
loadFile	If loadFile=TRUE, intensity data will be loaded from a separate file. Default is FALSE.
batchCorrect	If batchCorrect=TRUE ComBat will be run on the data to correct for batch effects due to sentrixID/slide number. Default is TRUE.
file	If loadFile=T this is the name of the file with the intensity values. Default is "intensity.txt".
resultsDir	Directory where results will be saved. Default is a folder in the current working directory called "resultsChamp".
sampleCNA	If sampleCNA=TRUE, then . Default is TRUE.
plotSample	If sampleCNA=TRUE and plotSample=TRUE, then CNA plots will be saved for each sample. Default is TRUE.
filterXY	Probes from X and Y chromosomes are removed. Default is TRUE.
groupFreqPlots	If groupFreqPlots=T, then
freqThreshold	If groupFreqPlots=T, then freqThreshold will be used as the cutoff for calling a gain or loss. Default is 0.03.
control	If control=T, then the samples defined by the controlGroup identifier will be used as the baseline for CNA calculations. Default is TRUE.

controlGroup   If Control=T, then controlGroup will be used as the baseline for CNA calcu-
                 lations. The default is "Control". Control samples must be labelled with this
                 identifier in the Sample_Group column of the pd file. If this doesn't exist in
                 your dataset then ChAMP will revert to using the internal blood controls "cham-
                 pCtls"'

arraytype   Choose microarray type is 450K or EPIC.

Author(s)

Feber, A
adapted by Morris, T

References

Feber, A et. al. (2014). CNA profiling using high density DNA methylation arrays. Genome
Biology.

Examples

```
data(testDataSet)
data(champBloodCtls)
myLoad=testDataSet
champ.CNA(batchCorrect=FALSE,sampleCNA=FALSE,groupFreqPlots=FALSE)
```

---

champ.DMR   *Applying Bumphunter or ProbeLasso Algorithms to detect Different*
                                 Methylation Regions in a beta valued Methylation Dataset.

Description

Applying Bumphunter or ProbeLasso Algorithms to Estimate regions for which a genomic profile
deviates from its baseline value. Originally implemented to detect differentially methylated ge-
nomic regions between two populations. By default, we recommend user do champ.DMR on nor-
malized beta value on two populations, like case to control. The function will return detected DMR
and estimated p value. The two algorithms specified in this function is different, while Bumphunter
calcuated averaged candidate bumps methylation value between case and control, ProbeLasso need
Different Methylated Probes (DMP) from champ.MVP as input parameter and find DMRs around
those DMPs. Thus parameters is different for two algorithms.

Usage

```
champ.DMR(betaNorm=myNorm$beta,design=myLoad$pd$Sample_Group,maxGap=300,
        cutoff=0.5,minProbes=7,smooth=TRUE,smoothFunction=loessByCluster,
        useWeights=FALSE,permutations=NULL,B=250,pickCutoff=FALSE,
    pickCutoffQ=0.99,nullMethod="bootstrap",verbose=TRUE,cores=3,arraytype="450K",
    method = "Bumphunter",resultsFile=mylimma,meanLassoRadius=375,minSigProbesLasso=3,minDmrSe
        minDmrSize=50,adjPvalProbe=0.05,adjPvalDmr=0.05,pData=myLoad$pd)
```

Arguments

	Since there are two function incoporated to detect DMRs, user may specify each function to do DMR detection, Bumphunter or ProbeLasso. Both methods are available for both 450K and EPIC beadarray. But they are controled by different parameters, thus users shall be careful when they specify parameters for corresponding algorithm. Parameters shared by two algorithms:
beta	Methylation beta valueed dataset user want to do RefFreeEWAS. We recommend to use normalized beta value. In champ.DMR function, beta value will be transformed to M value. NA value is NOT allowed into this function, thus user may need to do some imputation work beforehead. This parameter is essential for both two algorithms.
arraytype	Assign types of manifest shall be used to do DMR detection, "EPIC" or "450K" are provided. Annotation of probes is playing an important role in clustering probes across genome. This parameter is both important and available for both Bumphunter and ProbeLasso functions.
method	Specify the method users want to use to do DMR detection. There are two options: "Bumphunter" or "ProbeLasso". The default value is "Bumphunter".
	Parameters specified for Bumphunter algorithm:
design	Design vector (subjects x covariates). This parameter MUST be a vecter. Though characters, factor and numeric are all allowed because inside the function, character covariates will be transformed into numeric, we still recommend user input numeric deigned covariates vector. This parameter is specified for Bumphunter algorithm.
cutoff	A numeric value. Values of the estimate of the genomic profile above the cutoff or below the negative of the cutoff will be used as candidate regions. It is possible to give two separate values (upper and lower bounds). If one value is given, the lower bound is minus the value. This parameter is specified for Bumphunter.
pickCutoff	Should bumphunter attempt to pick a cutoff using the permutation distribution? This parameter is specified for Bumphunter.
pickCutoffQ	The quantile used for picking the cutoff using the permutation distribution. This parameter is specified for Bumphunter.
maxGap	If cluster is not provided this maximum location gap will be used to define cluster via the clusterMaker function. This parameter is specified for Bumphunter.
minProbes	Threshold to filtering clusters with too few probes in it. After region detection, champ.DMR will only select probes in clusters contain more than minsample probes to continue the program. This parameter is specified for Bumphunter.
nullMethod	Method used to generate null candidate regions, must be one of 'bootstrap' or 'permutation' (defaults to 'permutation'). However, if covariates in addition to the outcome of interest are included in the design matrix (ncol(design)>2), the 'permutation' approach is not recommended. See vignette and original paper for more information. This parameter is specified for Bumphunter.
smooth	A logical value. If TRUE the estimated profile will be smoothed with the smoother defined by smoothFunction, This parameter is specified for Bumphunter.

smoothFunction   A function to be used for smoothing the estimate of the genomic profile. Two
                 functions are provided by the package: loessByCluster and runmedByCluster.
                 This parameter is specified for Bumphunter.

useWeights   A logical value. If TRUE then the standard errors of the point-wise estimates of
                 the profile function will be used as weights in the loess smoother loessByCluster.
                 If the runmedByCluster smoother is used this argument is ignored. This param-
                 eter is specified for Bumphunter.

B   An integer denoting the number of resamples to use when computing null distri-
                 butions. This defaults to 0. If permutations is supplied that defines the number
                 of permutations/bootstraps and B is ignored. The default value is 250, This pa-
                 rameter is specified for Bumphunter.

permutations   is a matrix with columns providing indexes to be used to scramble the data and
                 create a null distribution when nullMethod is set to permutations. If the boot-
                 strap approach is used this argument is ignored. If this matrix is not supplied and
                 B>0 then these indexes are created using the function sample. This parameter is
                 specified for Bumphunter.

verbose   logical value. If TRUE, it writes out some messages indicating progress. If FALSE
                 nothing should be printed. This parameter is specified for Bumphunter.

cores   The embeded DMR detection function, bumphunter, could automatically use
                 more parallel to accelerate the program. User may assgin number of cores
                 could be used on users's computer. The default value is 3. User may use
                 detectCore() function to detect number of cores in total. This parameter is
                 specified for Bumphunter.

                 Parameters specified for ProbeLasso algorithm:

resultsFile   Different Methylated Probes (DMP) detected from champ.MVP() function, which
                 used limma function to find all CpGs show significant different methylation
                 value. It's a MUST provided parameter for ProbeLasso algorithm.

meanLassoRadius
                 Radius around each DMP to detect DMR, the default value is 375.

minSigProbesLasso
                 The minimum number of significant probes to be captured in lasso, default = 3.

minDmrSep   The minimum seperation (bp) between neighbouring DMRs, default = 1000.

minDmrSize   The minimum DMR size (bp), default = 50.

adjPvalProbe   The minimum threshold of significance for probes to be includede in DMRs,
                 default = 0.05

adjPvalDmr   This is the significance threshold for including DMRs in the final DMR list.

pData   This data.frame includes the information from the sample sheet. The default
                 assumes you ran champ.load and saved the output to "myLoad".

Value

myDmrs   Different Methylation Regions detected by champ.DMR. For different algo-
                 rithms, myDmrs would be in different structure.

myDmrProbes   Different Methylated Probes detected in DMRs.

Author(s)

Butcher, L,Aryee MJ, Irizarry RA, Andrew Teschendorff, Yuan Tian

References

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD and Irizarry RA (2014). "Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays." Bioinformatics, 30(10), pp. 1363-1369. http://doi.org/10.1093/bioinformatics/btu049. Butcher LM and Beck S (2015). "Probe Lasso: A novel method to rope in differentially methylated regions with 450K DNA methylation data." Methods, 72, pp. 21-28. http://doi.org/10.1016%2Fj.ymeth.2014.10.036.

Examples

```
myLoad <- champ.load(directory=system.file("extdata",package="ChAMPdata"))
    myNorm <- champ.norm(norm="NONE")
    myDMR <- champ.DMR(B=10)
```

---

champ.load   *Upload of raw HumanMethylation450 data from IDAT files.*

Description

Function that loads data from IDAT files to calculate intensity and produce quality control images.

Usage

```
champ.load(directory = getwd(), methValue = "B", resultsDir = paste(getwd(),
 "resultsChamp", sep = "/"), filterXY = TRUE, QCimages = TRUE, filterDetP = TRUE,
 detPcut = 0.01,  removeDetP = 0, filterBeads=TRUE, beadCutoff=0.05, filterNoCG=FALSE,filterSNPs=TRUE
```

Arguments

arraytype	Choose microarray type is 450K or EPIC.
directory	Location of IDAT files, default is current working directory.
methValue	Indicates whether you prefer m-values M or beta-values B.
resultsDir	Directory where results will be saved.
QCimages	If QCimages=T, then images will be saved.
filterDetP	If filter = T, then probes above the detPcut will be filtered out.
filterXY	If filterXY=TRUE, probes from X and Y chromosomes are removed. Default is TRUE.
detPcut	The detection p-value threshhold. Probes about this cutoff will be filtered out. Default is 0.01
removeDetP	The removeDetP parameter represents the fraction of samples that can contain a detection p-value above the detPcut. Default is 0.

filterBeads	If filterBeads=TRUE, probes with a beadcount less than 3 will be removed depending on the beadCutoff value. Default is TRUE.
filterCutoff	The beadCutoff represents the fraction of samples that must have a beadcount less than 3 before the probe is removed. Default is 0.05 or 5% of samples.
filterNoCG	If filterNoCG=TRUE, non-cg probes are removed. Default is FALSE.
filterSNPs	If filterSNPs=TRUE, probes in which the probed CpG falls near a SNP as defined in Nordlund et al are removed. Default is TRUE.
filterMultiHit	If filterMultiHit=TRUE, probes in which the probe aligns to multiple locations with bwa as defined in Nordlund et al are removed.. Default is TRUE.

Value

mset	mset object
rgSet	rgset object
pd	pd file of all sample information from Sample Sheet
intensity	A matrix of intensity values for all probes and all samples.
beta	A matrix of methylation scores (M or beta values) for all probes and all samples.
detP	A matrix of detection p-values for all probes and all samples.

Author(s)

Morris, T

Examples

```
myLoad=champ.load(directory=system.file("extdata",package="ChAMPdata"),filterBeads=TRUE)
```

---

champ.MVP   *Identify Most Variable Positions in Illumina HumanMethylation450 data.*

Description

This function

Usage

```
champ.MVP(beta.norm = myNorm$beta, pd = myLoad$pd, adjPVal = 0.05, adjust.method = "BH",
compare.group = c("C", "T"), resultsDir = paste(getwd(), "resultsChamp", sep = "/"),
bedFile = TRUE,arraytype="450K")
```

Arguments

beta.norm	A matrix of values representing the methylation scores for each sample (M or B). The default assumes you ran champ.norm and saved the output to "norm".
pd	This data.frame includes the information from the sample sheet. The default assumes you ran champ.load and saved the output to "myLoad".
adjPVal	The minimum threshold of significance for probes to be considered an MVP, default = 0.05
adjust.method	The p-value adjustment method to be used for the limma analyis, default= BH (Benjamini-Hochberg)
compare.group	Not yet implemented
resultsDir	Directory where results will be saved. Default is a folder in the current working directory called "resultsChamp".
bedFile	If bedFile=TRUE, the MVPs will be saved in bedfile format for downstream analysis.
arraytype	Choose microarray type is 450K or EPIC.

Value

results.file	A matrix of all probes with an adjusted p-value for significance of differential methylation containing columns for probeID, logFC, AveExpr, t, P.Value, adjusted p-value, B, chromosome, map info, chromosome arm, closest gene.1, gene.2, gene.3, gene.4, closest feature.1, feature.2, feature.3, feature.4, UCSC_CpG_ISLANDS_NAME, Relation to UCSC CpG Island, Phantom, DMR, Enhancer, HMM_Island, regulatory feature name, regulatory feature group, feature relation, average of first sample group, average of second sample group, delta beta

Author(s)

Morris, T

---

champ.norm	*Normalization of HumanMethylation450 data*

Description

Option to normalize data with a selection of normalization methods.

Usage

```
champ.norm(beta = myLoad$beta, rgSet = myLoad$rgSet, pd = myLoad$pd, mset = myLoad$mset,
sampleSheet = "sampleSheet.txt", resultsDir = paste(getwd(), "resultsChamp",
sep = "/"), methValue = "B", fromIDAT = TRUE, norm = "BMIQ", fromFile = FALSE, betaFile,
filter = TRUE, filterXY = TRUE, QCimages = FALSE, plotBMIQ = FALSE,arraytype="450K")
```

Arguments

arraytype	Choose microarray type is 450K or EPIC.
beta	A matrix of values representing the methylation scores for each sample (M or B). The default assumes you ran champ.load and saved the output to "myLoad".
rgSet	An rgSet object that was created when data was loaded the data from the .idat files. The default assumes you ran champ.load and saved the output to "myLoad".
pd	This data.frame includes the information from the sample sheet. The default assumes you ran champ.load and saved the output to "myLoad".
mset	Loads an mset object that was created when data was loaded from the .idat files. The default assumes you ran champ.load and saved the output to "myLoad".
sampleSheet	If the data has not been loaded from .idat files and fromFile=TRUE then this points to the required sampleSheet. Default is "sampleSheet.txt".
resultsDir	Directory where results will be saved. Default is a folder in the current working directory called "resultsChamp".
methValue	Indicates whether you prefer the methylation scores to be calculated as m-values (M) or beta-values (B). Default is B.
fromIDAT	If fromIDAT=T,
norm	This specifies which normalization method will be used. Values can be BMIQ (by default), PBC, SWAN or NONE.
fromFile	If loadFile=TRUE, then the beta values and sample sheet need to be uploaded.
betaFile	If
filter	Not yet implemented. If fromFile=T and this is from a genome studio file, probes that have a detection p-value below detPcut are filtered out. Default is TRUE.
filterXY	If fromFile=True, probes from X and Y chromosomes are removed. Default is TRUE.
QCimages	If QCimages=TRUE, then quality control images are saved to the resultsDir. Default is TRUE.
plotBMIQ	If plotBMIQ=TRUE and norm="BMIQ", BMIQ plots will be saved. Default is TRUE.

Value

beta	A matrix of normalised methylation scores (M or beta values) for all probes and all samples.

Author(s)

Morris, T. wrote the wrappers

References

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. Bioinformatics. 2013 Jan 15;29(2):189-96.

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F.Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011,Dec;3(6):771-84.

Touleimat N, Tost J. Complete pipeline for Infinium Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012 Jun;4(3):325-41.

---

champ.process | *Process function to run all methods in ChAMP pipeline.*

Description

This function allows the user to run the entire pipeline in one function. Arguments allow user to select functions if desired.

Usage

```
champ.process(fromIDAT = TRUE, fromFile = FALSE, directory = getwd(), resultsDir =
paste(getwd(), "resultsChamp", sep = "/"), methValue = "B", filterDetP = TRUE,
detPcut = 0.01, filterXY = TRUE, removeDetP = 0, filterBeads = TRUE, beadCutoff =
0.05, filterNoCG = FALSE, QCimages = TRUE, batchCorrect = TRUE, runSVD =
TRUE, studyInfo = FALSE, infoFactor = c(), norm = "BMIQ", adjust.method = "BH",
adjPVal = 0.05, runDMR = TRUE,  runCNA = TRUE, plotBMIQ = FALSE, DMRpval = 0.05,
sampleCNA=TRUE,plotSample = TRUE,groupFreqPlots=TRUE,freqThreshold=0.3, bedFile
= FALSE, methProfile = FALSE, controlProfile = FALSE,arraytype="450K")
```

Arguments

fromIDAT	If fromIDAT=TRUE, data is imported from .idat files with an associated sample sheet (.csv). If rawdata=FALSE then data is uploaded from a text file (saved as "beta.txt". Default is TRUE.)
fromFile	The
directory	The directory where the .idat files and sample sheet are located, default is current working directory.
resultsDir	Directory where results will be saved. Default is to create a folder called "resultsChamp"in the current working directory.
methValue	Indicates whether you prefer the methylation scores to be calculated as m-values (M) or beta-values (B). Default is B.
filterDetP	If filter=TRUE, probes that have a detection p-value below detPcut are filtered out. Default is TRUE.

detPcut	If filter=TRUE, this value with be used as the significance threshold for filtering out probes based on the detection p-value. Default=0.01.
filterXY	If filterXY=TRUE, probes from X and Y chromosomes are removed. Default is TRUE.
QCimages	If QCimages=TRUE, then quality control images are saved to the resultsDir. Default is TRUE.
removeDetP	The removeDetP parameter represents the fraction of samples that can contain a detection p-value above the detPcut. Default is 0.
filterBeads	If filterBeads=TRUE, probes with a beadcount less than 3 will be removed depending on the beadCutoff value. Default is TRUE.
beadCutoff	The beadCutoff represents the fraction of samples that must have a beadcount less than 3 before the probe is removed. Default is 0.05 or 5 percent of samples.
filterNoCG	If filterNoCG=TRUE, non-cg probes are removed. Default is FALSE.
batchCorrect	If batchCorrect=TRUE, then the ComBat batch correction will be performed on batch effects related to bead chip. Default is TRUE.
runSVD	If runSVD=TRUE, SVD analysis for identifying batch effects will be performed. Default is TRUE.
studyInfo	If runSVD = TRUE, additional study covariate information can be included in the SVD analysis. Default is FALSE.
infoFactor	This
norm	This specifies which normalization method will be used. Values can be BMIQ (by default), PBC, SWAN or NONE.
adjPVal	The minimum threshold of significance for probes to be includede in DMRs, default = 0.05
adjust.method	The p-value adjustment method to be used for the limma analyis, default= BH (Bonferroni-Hochberg)
runDMR	If runDMR=TRUE, runs the probe lasso method for finding DMRs. This will result in an MVP list with p-values and a DMR list with p-values. Default is TRUE.
runCNA	If runCNA=TRUE, copy number abberation analysis will be performed. Default is TRUE.
plotBMIQ	If plotBMIQ=TRUE and norm="BMIQ", BMIQ plots will be saved. Default is TRUE.
DMRpval	If runDMR=TRUE, this value will be used as the cutoff for the DMR p-value. Default is 0.05.
sampleCNA	If sampleCNA=TRUE, then . Default is TRUE.
plotSample	If plotSample=TRUE, CNA plots will be saved. Default is TRUE.
groupFreqPlots	If groupFreqPlots=T, then
freqThreshold	If groupFreqPlots=T, then freqThreshold will be used as the cutoff for calling a gain or loss. Default is 0.03.
bedFile	if bedFile = TRUE. MVP list will be saved as an additional file in bedfile format for downstream analysis. Defaults is TRUE.

methProfile	If methProfile=TRUE then the beta values will be uploaded using the Methylation ProbeProfile file from Genome Studio. Default is FALSE.
controlProfile	If rawdata = FALSE and runSVD = TRUE, then it is useful to have a control probe profile file exported from Genome Studio so that internal control probes can be included in the SVD analyis. Default is FALSE.
arraytype	Choose microarray type is 450K or EPIC.

Author(s)

Morris, T

Examples

```
directory=system.file("extdata",package="ChAMPdata")
champ.process(directory=directory)
```

---

champ.refbase   *Applying References-Base Methold to beta valued methylation data.*

Description

Applying References-Based Methold to correct cell-proportion in a methylation dataset. Reference-based method use purified whole blood cell-type specific methylation value to correct beta value dataset. Cell Proportions for each cell-type will be detected, and lm function will be used to correct beta value for 5 largest cell types. Cell type with smallest cell proportion will not be corrected.

Usage

```
champ.refbase(beta=myLoad$beta,arraytype="450K")
```

Arguments

beta	whole blood beta methylation dataset user want to correct.
arraytype	There are two types of purified cell-type specific references can be chosen, "450K" and "27K". By default, 450K value will be used, but user may choose 27K as well.

Value

CorrectedBea	A beta valued matrix, with all value get corrected with RefBaseEWAS method. Be aware, champ.refbase will only correct top 5 cell types with largest mean cell proportions, and leave the cell with smallest mean cell proportion. User may check CellFraction result to find out which cell types are get corrected.
CellFraction	Proportion for each cell type.

Author(s)

Houseman EA, Yuan Tian, Andrew Teschendorff

References

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13: 86. doi: 10.1186/1471-2105-13-86. pmid:22568884

Examples

```
data(testDataSet)
myLoad=testDataSet
myRefBase=champ.refbase(myLoad$beta)
```

---

champ.reffree   *Applying RefFreeEWAS Methold to beta valued methylation data.*

Description

Applying RefFreeEWAS method to beta valued methylation data. This method does not rely on puritied cell reference, thus can be easily used on tissue data set, while RefbaseEWAS can only be used to whole blood samples. Reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This method is similar to surrogate variable analysis (SVA and ISVA), except that it makes additional use of a biological mixture assumption. Returns mixture-adjusted Beta and unadjusted Bstar, as well as estimates of various latent quantities.

Usage

```
champ.reffree(beta=myLoad$beta,design=myLoad$pd$Sample_Group,K=NULL,nboot=10)
```

Arguments

beta	Methylation beta valueed dataset user want to do RefFreeEWAS.
design	Design matrix (subjects x covariates). This parameter MUST be a vecter or a matrix. Though Characters are allowed because inside the function, character covariates will be transformed into numeric, we still recommend user input numeric deigned covariates matrix or vector.
K	Number of latent variable. If this value was ignored, function will use Random Matrix Theory from isva pacakge to estimate latent variables.
nboot	Number for Bootstrap on result of RefFreeEWAS.

Value

RefFreeEWASModel

> RefFreeEWASModel S4 Object from RefFreeEWAS pacakge, contains adjusted beta value and unadjusted beta value (Bstar).

pvBeta
> p value of each covariates, calculated from cell type mixture corrected Beta value.

qvBeta
> q value of each covariates, calculated from cell type mixture corrected Beta value.

Author(s)

Houseman EA, Yuan Tian, Andrew Teschendorff

References

Houseman EA, Kile ML, et al., Reference-free deconvolution of DNA methylation data and mediation by cell composition effects (2016). http://biorxiv.org/content/early/2016/01/23/037671.

Examples

```
myLoad=champ.load(directory=system.file("extdata",package="ChAMPdata"))
    myRefFree=champ.reffree()
```

---

champ.runCombat	*Function that uses ComBat to correct for batch effects related to slide/BeadChip.*

Description

This function formats data to run through ComBat batch correction. If beta values are used the data is first logit transformed.

Usage

```
champ.runCombat(beta.c = myNorm$beta, pd = myLoad$pd, logitTrans = TRUE)
```

Arguments

beta.c
> A matrix of values representing the methylation scores for each sample (M or B). The default assumes you ran champ.norm and saved the output to "norm".

pd
> This data.frame includes the information from the sample sheet. The default assumes you ran champ.load and saved the output to "myLoad".

logitTrans
> If logitTrans=T then your data will be logit transformed before the Combat correction and inverse logit transformed after correction. This is T by default for Beta values but if you have selected M values it will revert to False. It is also False when used with CNA as those are intensity values that don't need to be transformed.

Value

beta   The matrix of values represeting the methylation scores for each sample after
               ComBat batch correction.

Author(s)

T. Morris

---

champ.SVD   *Singular Value Decomposition analysis for batch effects prediciton in*
                              HumanMethylation450 data

Description

Runs Singular Value Decomposition on a dataset to estimate the impact of batch effects.

Usage

```
champ.SVD(beta = myNorm$beta, rgSet = myLoad$rgSet, detP = myLoad$detP, pd = myLoad$pd,
loadFile = FALSE, betaFile = "beta.txt", sampleSheet = "sampleSheet.txt", methProfile = FALSE,
methFile = "MethylationProbeProfile.txt", controlProfile = FALSE,
controlFile = "ControlProbeProfile.txt", studyInfo = FALSE, studyInfoFile = "studyInfo.txt",
infoFactor = c(), resultsDir = paste(getwd(), "resultsChamp", sep = "/"))
```

Arguments

beta   A matrix of values representing the methylation scores for each sample (M or B).
               The default assumes you ran champ.norm and saved the output to "myNorm".

rgSet   An rgSet object that was created when data was loaded the data from the .idat
               files. The default assumes you ran champ.load and saved the output to "my-
               Load".

detP   A matrix of detection p-values for each sample. The default assumes you ran
               champ.load and saved the output to "myLoad".

pd   This data.frame includes the information from the sample sheet. The default
               assumes you ran champ.load and saved the output to "myLoad".

loadFile   If loadFile=TRUE, then the beta values and sample sheet need to be uploaded

betaFile   If loadFile=T,

sampleSheet   If the data has not been loaded from .idat files and fromFile=TRUE then this
               points to the required sampleSheet. Default is "sampleSheet.txt"

methProfile   If methprofile=TRUE then the beta values will be uploaded using the Methyla-
               tionProbeProfile file from Genome Studio

methFile   If methProfile=TRUE then the beta values will be uploaded using the Methyla-
               tionProbeProfile from Genome Studio. This is the name of the file. Default is
               "MethylationProbeProfile.txt"

controlProfile If rawdata = FALSE and runSVD = TRUE, then it is useful to have a control probe profile file exported from Genome Studio so that internal control probes can be included in the SVD analyis. Default is FALSE.

controlFile If controlProfile = TRUE then the control probe values will be uploaded using the ControlProbeProfile from Genome Studio. This is the name of the file. Default is "ControlProbeProfile.txt"

studyInfo If studyInfo=TRUE, additional study covariate information can be included in the SVD analysis. Default is FALSE.

infoFactor This.

studyInfoFile If studyInfo =T, this file will include the additional study information. Default is "studyInfo.txt".

resultsDir Directory where results will be saved. Default is to create a folder called "resultsChamp"in the current working directory.

Author(s)

Teschendorff, A
adapted by Morris, T

References

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I. J., and Widschwendter, M. (2009). An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One, 4(12), e8274

---

champ.TrueMethyl *Identify Most Variable Positions between oxBS TrueMethyl Samples and BS samples in Illumina HumanMethylation450 data.*

Description

This function

Usage

```
champ.TrueMethyl(beta.norm = myNorm$beta, pd = myLoad$pd, adjPVal = 0.05, adjust.method = "BH",
compare.group = c("oxBS", "BS"), resultsDir = paste(getwd(), "resultsChamp", sep = "/"),
bedFile = TRUE,arraytype="450K")
```

Arguments

beta.norm A matrix of values representing the methylation scores for each sample (M or B). The default assumes you ran champ.norm and saved the output to "norm".

pd This data.frame includes the information from the sample sheet. The default assumes you ran champ.load and saved the output to "myLoad".

`adjPVal`	The minimum threshold of significance for probes to be considered an MVP, default = 0.05
`adjust.method`	The p-value adjustment method to be used for the limma analyis, default= BH (Benjamini-Hochberg)
`compare.group`	Not yet implemented
`resultsDir`	Directory where results will be saved. Default is a folder in the current working directory called "resultsChamp".
`bedFile`	If bedFile=TRUE, the MVPs will be saved in bedfile format for downstream analysis.
`arraytype`	Choose microarray type is 450K or EPIC. The default value is 450K.

Value

`results.file`	A matrix of all probes with an adjusted p-value for significance of differential methylation containing columns for probeID, logFC, AveExpr, t, P.Value, adjusted p-value, B, chromosome, map info, chromosome arm, closest gene.1, gene.2, gene.3, gene.4, closest feature.1, feature.2, feature.3, feature.4, UCSC_CpG_ISLANDS_NAME, Relation to UCSC CpG Island, Phantom, DMR, Enhancer, HMM_Island, regulatory feature name, regulatory feature group, feature relation, average of first sample group, average of second sample group, delta beta

Author(s)

Morris, T

Index