Package 'ChAMP'

December 12, 2014

Type Package

Title Chip Analysis Methylation Pipeline for Illumina HumanMethylation450

Version 1.4.0

Date 2014-09-10

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 aberrations.

License GPL-3

Depends R (>= 3.0.1), minfi, ChAMPdata, Illumina450ProbeVariants.db

Imports sva, IlluminaHumanMethylation450kmanifest, limma, RPMM, DNAcopy, preprocessCore, impute, marray, wateRmelon,plyr

biocViews Microarray, MethylationArray, Normalization, TwoChannel, CopyNumber

Author Tiffany Morris [cre, aut],
Lee Butcher [aut],
Andrew Feber [aut],
Andrew Teschendorff [aut],
Ankur Chakravarthy [aut]

Maintainer Tiffany Morris <tiffany.morris@ucl.ac.edu>

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 Aberration analysis.

Details

Package:	ChAMP
Type:	Package
Version:	1.3.14
Date:	2014-09-10
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 Butcher, Andy Feber, Andrew Teschendorff, Ankur Chakravarthy, Stephen Beck
Maintainer: Tiffany Morris <tiffany.morris@ucl.ac.uk>

Examples

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

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 calculations. 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 "champCtls"'

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.lasso *Probe Lasso DMR Hunter*

Description

A method for identifying DMRs (differentially methylated regions) using a feature based dynamic window. Also offers the option to filter SNPs based on data from the 1000 Genomes Project.

Usage

```
champ.lasso(fromFile = FALSE, uploadResults = FALSE, uploadFile = "limma.txt", limma,
beta.norm = myNorm$beta, pd = myLoad$pd, filterXY = TRUE, image = TRUE, mafPol.lower = 0,
mafPol.upper = 0.05, popPol = "eur", lassoStyle = "max", lassoRadius = 2000,
minSigProbesLasso = 3, minDmrSep = 1000, minDmrSize = 0, adjPVal = 0.05,
adjust.method = "BH", resultsDir = paste(getwd(), "resultsChamp", sep = "/"),
bedFile = TRUE, DMRpval = 0.05, batchDone = FALSE, normSave)
```

Arguments

fromFile	if
uploadResults	Set uploadResults=TRUE if you haven't loaded data from .idat files and need to upload the limma file
uploadFile	If uploadResults=TRUE this is the file name
limma	If
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".
filterXY	If filterXY=T, probes from the X and Y chromosomes are removed.

image	If image=T, images will be saved as a pdf file in the resultsDir.
mafPol.lower	The lower limit for the minor allele frequencies of included polymorphisms
mafPol.upper	The upper limit for the minor allele frequencies of included polymorphisms
popPol	Indicates the population on which to base the polymorphic frequency Asian (asn), American (amr), African (afr) or Northern European (eur)
lassoStyle	Determines whether lassoRadius is the minimum (min) or maximum (max) lasso size, default = "max"
lassoRadius	The lasso size, default = 2000
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 = 0
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)
resultsDir	Directory where results will be saved. Default is to create a folder called "resultsChamp"in the current working directory.
bedFile	If bedFile=TRUE, the DMRs will be saved in bedfile format for downstream analysis. Default is TRUE.
DMRpval	This is the significance threshold for including DMRs in the final DMR list.
batchDone	Internal variable to indicate if combat batch correction was performed.
normSave	Internal variable to store normalized, not-batch corrected beta values.

Value

dmrList	A matrix of DMRs is returned containing columns for probeID, deltaBeta, adjusted p-value, chromosome, map info, chromosome arm, nearest feature, SNP allele frequency on forward strand, SNP allele frequence on reverse strand, distance of nearest probe, radius of lasso that captured DMR, DMR number, DMR start, DMR end, DMR size, p-value for DMR

Author(s)

Butcher, L

Examples

```
data(testDataSet)
myLoad=testDataSet
myNorm=champ.norm(norm="NONE")
```

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)
```

Arguments

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.
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% of samples.
filterNoCG	If filterNoCG=TRUE, non-cg probes are removed. Default is FALSE.

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)
```

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.

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

Examples

```
data(testDataSet)
myLoad=testDataSet
myNorm=champ.norm(norm="NONE")
```

---

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 = TRUE, plotBMIQ = TRUE)
```

Arguments

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.

Examples

```
myLoad=champ.load(directory=system.file("extdata",package="ChAMPdata"))
myNorm=champ.norm(norm="NONE")
```

---

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 = TRUE, DMRpval = 0.05,
sampleCNA=TRUE,plotSample = TRUE,groupFreqPlots=TRUE,freqThreshold=0.3, bedFile
= FALSE, methProfile = FALSE, controlProfile = FALSE)
```

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 MethylationProbeProfile 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.

Author(s)

Morris, T

Examples

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

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

Examples

```
data(testDataSet)
myLoad=testDataSet
myNorm=champ.norm(norm="NONE")
```

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 "myLoad".
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 MethylationProbeProfile file from Genome Studio
methFile	If methProfile=TRUE then the beta values will be uploaded using the MethylationProbeProfile 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

Examples

```
data(testDataSet)
myLoad=testDataSet
myNorm=champ.norm(norm="NONE")
champ.SVD()
```

Index