

Introduction

Introduction

These instructions are aimed at people familiar with R and familiar with TCGA/GDC platforms and data types. They are intended to introduce the reader to producing the given assessment. These instructions will only rarely, if ever, touch on the appropriateness of the assessment algorithm or interpretation of output. See MBatch_01_InstallLinux for instructions on downloading test data.

Algorithm

EBNPlus_TrainAndValidateFromVector_Structures creates and returns a training and validation set and corrected data, based on the training and validation set and provided vectors of samples to use for training and validation, instead of using replicates and percentages.

Output

The primary output method for MBatch is to view results in the Batch Effects Website. Correction algorithms generally do not create graphical output and instead create TSV output files.

Usage

```
EBNPlus_TrainAndValidateFromVector_Structures(theDataMatrix1, the-  
DataMatrix2, theBatchId1, theBatchId2, theEBNP_PsuedoReplicates1Train,  
theEBNP_PsuedoReplicates2Train, theEBNP_PsuedoReplicates1Validation,  
theEBNP_PsuedoReplicates2Validation, theEBNP_BatchWithZero,  
theEBNP_FixDataSet, theEBNP_CorrectForZero, theEBNP_ParametricPriorsFlag,  
theEBNP_TestRatio=0, theSeed=NULL, theTestSeed=NULL, theEBNP_PriorPlotsFile  
= NULL, theEBNP_MinSampleNum = 3)
```

Arguments

##theDataMatrix1

A matrix for data set 1 containing numeric values with columns being sample ids and rows being feature (like gene or probe) ids.

##theDataMatrix2

A matrix for data set 2 containing numeric values with columns being sample ids and rows being feature (like gene or probe) ids.

##theBatchId1

The group name for data set 1 (for example, RNASeqV2)

##theBatchId2

The group name for data set 2 (for example, Agilent4502)

##theEBNP_PsuedoReplicates1Train

A vector of sample ids to use as replicates for data set 1 for training.

##theEBNP_PsuedoReplicates2Train

A vector of sample ids to use as replicates for data set 2 for training.

##theEBNP_PsuedoReplicates1Validation

A vector of sample ids to use as replicates for data set 1 for validation. Use NULL if no validation is desired. This means corrections will be done.

##theEBNP_PsuedoReplicates2Validation A vector of sample ids to use as replicates for data set 2 for validation. Use NULL if no validation is desired. This means corrections will be done.

##theEBNP_BatchWithZero Batch With Zero indicates which data set contains zero values. This is a string “1”, “2”, or “both”. These values will be changed in a future release.

##theEBNP_FixDataSet

Fix Data Set indicates which data set should be set as invariate during the corrections. Value are numeric, 1 or 2 for the appropriate data set, or as.numeric(NA) for neither.

##theEBNP_CorrectForZero

Correct For Zero indicates whether or not data sets marked as “Batch With Zero” should be modified from zero (TRUE or FALSE)

##theEBNP_ParametricPriorsFlag

Use parametric adjustments for corrections (TRUE or FALSE)

##theSeed

NULL or a seed to use. Defaults to NULL. This is passed to the standard R set.seed function.

##theTestSeed

NULL or a seed to use. Defaults to NULL. This is passed to the standard R

set.seed function. This is used only when creating Test Set.

##theEBNP_PriorPlotsFile

Defaults to NULL and does not create PNG. Non-NULL should be the full path and filename for where to create the prior plots PNG.

##theEBNP_MinSampleNum

Defaults to 3. Any row (gene) with less than this number of samples is dropped.

Example Call

The following code is adapted from the tests/EB_withNonParametricPriors.R file. Data used is from the testing data as per the MBatch_01_InstallLinux document. In the future, we plan to make the output from MBatch more user friendly, but currently, this produces the following output at the command line.

```
{
library(MBatch)

inputDir <- getTestInputDir()
outputDir <- getTestOutputDir()
compareDir <- getTestCompareDir()

theDataFile1=file.path(inputDir, "brca_rnaseq2_matrix_data.tsv")
theDataFile2=file.path(inputDir, "brca_agi4502_matrix_data.tsv")
theOutputDir=file.path(outputDir, "EBNPlus_TrainAndValidateReplicates_Structures")
theBatchId1="RNASeqV2"
theBatchId2="Agilent4502"
theRandomSeed=314

# trim genes to get just gene symbols from standardized data
trimGenes <- function(theGenes)
{
  foo <- as.vector(unlist(
    sapply(theGenes, function(theGene)
    {
      # keep the same if it starts with ?
      if (TRUE==grepl("^[?]+", theGene))
      {
        return(theGene)
      }
      else
      {
        # split on the | and take the first argument
        # this makes no change if no pipe
        return(strsplit(theGene, "|", fixed=TRUE)[[1]][1])
      }
    }
  ))
}
```

```

    }
  })
})
foo
}

# remove duplicates from columns (samples)
removeDuplicatesFromColumns <- function(theMatrix)
{
  indexOfDuplications <- which(duplicated(colnames(theMatrix)))
  if (length(indexOfDuplications) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[ , -indexOfDuplications]
  }
  return(theMatrix)
}

# remove duplicates from rows (genes/probes)
removeDuplicatesFromRows <- function(theMatrix)
{
  indexOfDuplications <- which(duplicated(rownames(theMatrix)))
  if (length(indexOfDuplications) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[-indexOfDuplications, ]
  }
  return(theMatrix)
}

printMatrix <- function(theMatrix)
{
  print(is.matrix(theMatrix))
  print(dim(theMatrix))
  rowMax <- dim(theMatrix)[1]
  colMax <- dim(theMatrix)[2]
  rowMax <- min(rowMax, 4)
  colMax <- min(colMax, 4)
  print(theMatrix[1:rowMax, 1:colMax])
}

if ((!dir.exists(theDataFile1))&&(!dir.exists(theDataFile2)))
{
  warnLevel<-getOption("warn")
  on.exit(options(warn=warnLevel))
}

```

```

# warnings are errors
options(warn=3)
# if there is a warning, show the calls leading up to it
options(showWarnCalls=TRUE)
# if there is an error, show the calls leading up to it
options(showErrorCalls=TRUE)
#
unlink(theOutputDir, recursive=TRUE)
dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)
# read the files in. This can be done however you want
print("read the files")
theDataMatrix1 <- readAsGenericMatrix(theDataFile1)
theDataMatrix2 <- readAsGenericMatrix(theDataFile2)
# this is the reduce genes to just gene symbols, handling those from standardized data
print("reduce to gene symbols")
rownames(theDataMatrix1) <- trimGenes(rownames(theDataMatrix1))
rownames(theDataMatrix2) <- trimGenes(rownames(theDataMatrix2))
# remove any duplicates (this is a requirement for EBNplus)
print("remove duplicates")
theDataMatrix1 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix1))
theDataMatrix2 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix2))
# pseudoreplicate vectors
theEBNP_PsuedoReplicates1Train <- c("TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07", "TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07")
theEBNP_PsuedoReplicates2Train <- c("TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07", "TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07")
theEBNP_PsuedoReplicates1Validation <- c("TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-11R-A115-07", "TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-11R-A115-07")
theEBNP_PsuedoReplicates2Validation <- c("TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-11R-A115-07", "TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-11R-A115-07")
# call function
print("EBNPlus_TrainAndValidateReplicates_Structures")
resultsList <- EBNPlus_TrainAndValidateFromVector_Structures(theDataMatrix1, theDataMatrix2,
  theBatchId1, theBatchId2,
  theEBNP_PsuedoReplicates1Train,
  theEBNP_PsuedoReplicates2Train,
  theEBNP_PsuedoReplicates1Validation,
  theEBNP_PsuedoReplicates2Validation,
  theEBNP_BatchWithZero="1",
  theEBNP_FixDataSet=as.numeric(NA),
  theEBNP_CorrectForZero=TRUE,
  theEBNP_ParametricPriorsFlag=TRUE,
  theEBNP_TestRatio=0.3,
  theSeed=theRandomSeed,
  theTestSeed=theRandomSeed,
  theEBNP_PriorPlotsFile=file.path(theOutputDir, "priorplots.PNG"))
print("TestSet1")
printMatrix(resultsList$TestSet1)
print("TestSet2")
printMatrix(resultsList$TestSet2)

```

```

    print("TrainingSet1")
    printMatrix(resultsList$TrainingSet1)
    print("TrainingSet2")
    printMatrix(resultsList$TrainingSet2)
    print("TrainingResults")
    printMatrix(resultsList$TrainingResults)
    print("ValidationSet1")
    printMatrix(resultsList$ValidationSet1)
    print("ValidationSet2")
    printMatrix(resultsList$ValidationSet2)
    print("ValidationResults")
    printMatrix(resultsList$ValidationResults)
    print("CorrectedResults")
    printMatrix(resultsList$CorrectedResults)
  }
}

## [1] "read the files"
## [1] "reduce to gene symbols"
## [1] "remove duplicates"
## [1] "EBNPlus_TrainAndValidateReplicates_Structures"
## 2020 11 18 16:20:49.324 DEBUG ab7c64738d52 EBNPlus_TrainAndValidateFromVector_Structures
## 2020 11 18 16:20:49.324 DEBUG ab7c64738d52 removeDuplicatesFromRows
## 2020 11 18 16:20:49.325 DEBUG ab7c64738d52 removeDuplicatesFromRows
## 2020 11 18 16:20:49.325 DEBUG ab7c64738d52 removeDuplicatesFromColumns
## 2020 11 18 16:20:49.326 DEBUG ab7c64738d52 removeDuplicatesFromColumns
## 2020 11 18 16:20:49.326 DEBUG ab7c64738d52 EBNPlus theEBNP_PriorPlotsFile= /builds/Batch
## 2020 11 18 16:20:49.326 DEBUG ab7c64738d52 dim(theDataMatrix1)=10000 1215
## 2020 11 18 16:20:49.326 DEBUG ab7c64738d52 dim(theDataMatrix2)=10000 600
## 2020 11 18 16:20:49.327 DEBUG ab7c64738d52 remove unknown genes, that start with ?
## 2020 11 18 16:20:49.463 DEBUG ab7c64738d52 dim(theDataMatrix1)=10000 1215
## 2020 11 18 16:20:49.463 DEBUG ab7c64738d52 dim(theDataMatrix2)=10000 600
## 2020 11 18 16:20:49.463 DEBUG ab7c64738d52 make EBNplus
## 2020 11 18 16:20:49.467 DEBUG ab7c64738d52 before callNextMethod
## 2020 11 18 16:20:49.467 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 after callNextMethod
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 after EBNplus
## 2020 11 18 16:20:49.468 DEBUG ab7c64738d52 dim(ebObj@mData1)=10000 1215
## 2020 11 18 16:20:49.469 DEBUG ab7c64738d52 dim(ebObj@mData2)=10000 600
## 2020 11 18 16:20:49.470 DEBUG ab7c64738d52 getBiComOrder
## 2020 11 18 16:20:49.471 DEBUG ab7c64738d52 makeCommonRows
## 2020 11 18 16:20:49.523 DEBUG ab7c64738d52 dim(mData1)=7785 1215
## 2020 11 18 16:20:49.523 DEBUG ab7c64738d52 dim(mData2)=7785 600

```

```

## 2020 11 18 16:20:49.524 DEBUG ab7c64738d52 makeCommonCols
## 2020 11 18 16:20:49.572 DEBUG ab7c64738d52 dim(mData1)=7785 586
## 2020 11 18 16:20:49.572 DEBUG ab7c64738d52 dim(mData2)=7785 586
## 2020 11 18 16:20:49.572 DEBUG ab7c64738d52 asSameOrder
## 2020 11 18 16:20:49.590 DEBUG ab7c64738d52 asSameOrder before
## 2020 11 18 16:20:49.591 DEBUG ab7c64738d52 asSameOrder after if 1
## 2020 11 18 16:20:49.614 DEBUG ab7c64738d52 asSameOrder after if 2
## 2020 11 18 16:20:49.614 DEBUG ab7c64738d52 row.names(mat1)
## 2020 11 18 16:20:49.614 DEBUG ab7c64738d52 rownames(mat1)
## 2020 11 18 16:20:49.614 DEBUG ab7c64738d52 row.names(mat2)
## 2020 11 18 16:20:49.615 DEBUG ab7c64738d52 rownames(mat2)
## 2020 11 18 16:20:49.615 DEBUG ab7c64738d52 m.i
## 2020 11 18 16:20:49.638 DEBUG ab7c64738d52 asSameOrder after if 3
## 2020 11 18 16:20:49.638 DEBUG ab7c64738d52 dim(ebObj@mat1Com)=7785 586
## 2020 11 18 16:20:49.639 DEBUG ab7c64738d52 dim(ebObj@mat2Com)=7785 586
## 2020 11 18 16:20:49.639 DEBUG ab7c64738d52 get Validation and Training sets from vectors
## 2020 11 18 16:20:49.665 DEBUG ab7c64738d52 dim(ebObj@mat1Validation)=10000 175
## 2020 11 18 16:20:49.665 DEBUG ab7c64738d52 dim(ebObj@mat2Validation)=10000 175
## 2020 11 18 16:20:49.666 DEBUG ab7c64738d52 dim(ebObj@mat1Train)=10000 288
## 2020 11 18 16:20:49.666 DEBUG ab7c64738d52 dim(ebObj@mat2Train)=10000 288
## 2020 11 18 16:20:49.666 DEBUG ab7c64738d52 before Test Set
## 2020 11 18 16:20:49.701 DEBUG ab7c64738d52 after Test Set
## 2020 11 18 16:20:49.743 DEBUG ab7c64738d52 train
## 2020 11 18 16:20:49.759 DEBUG ab7c64738d52 inside train
## 2020 11 18 16:20:49.759 DEBUG ab7c64738d52 train theEBNP_PriorPlotsFile= /builds/BatchEff
## 2020 11 18 16:20:49.767 DEBUG ab7c64738d52 colnames(Object@mat1Train)=TCGA-A1-A0SJ-01A-11
## 2020 11 18 16:20:49.767 DEBUG ab7c64738d52 colnames(Object@mat2Train)=TCGA-A1-A0SJ-01A-11
## 2020 11 18 16:20:49.767 DEBUG ab7c64738d52 inside train, call getData4EB
## 2020 11 18 16:20:50.158 DEBUG ab7c64738d52 data4EB <- as.matrix
## 2020 11 18 16:20:50.158 DEBUG ab7c64738d52 dim(data4EB)=7785 404
## 2020 11 18 16:20:50.159 DEBUG ab7c64738d52 cbinds and rbinds
## 2020 11 18 16:20:50.159 DEBUG ab7c64738d52 Object@DF1batch= RNASeqV2
## 2020 11 18 16:20:50.159 DEBUG ab7c64738d52 Object@DF2batch= Agilent4502
## 2020 11 18 16:20:50.159 DEBUG ab7c64738d52 row.names(sampBatch)
## 2020 11 18 16:20:50.160 DEBUG ab7c64738d52 TCGA-A1-A0SJ-01A-11R-A084-07.RNASeqV2 TCGA-A2
## 2020 11 18 16:20:50.160 DEBUG ab7c64738d52 colnames(sampBatch)
## 2020 11 18 16:20:50.160 DEBUG ab7c64738d52 sample batch
## 2020 11 18 16:20:50.160 DEBUG ab7c64738d52 inside train, after getData4EB
## 2020 11 18 16:20:50.160 DEBUG ab7c64738d52 call EB MBatch
## 2020 11 18 16:20:50.233 DEBUG ab7c64738d52 TDC HERE trainEB mbatch
## 2020 11 18 16:20:50.233 DEBUG ab7c64738d52 EBNplus args Obj
## 2020 11 18 16:20:50.233 DEBUG ab7c64738d52 EBNplus args par.prior
## 2020 11 18 16:20:50.233 DEBUG ab7c64738d52 EBNplus args minSampleNum
## 2020 11 18 16:20:50.233 DEBUG ab7c64738d52 EBNplus theEBNP_PriorPlotsFile= /builds/BatchEff
## 2020 11 18 16:20:50.234 DEBUG ab7c64738d52 dat
## [1] 7785 404

```

```

## 2020 11 18 16:20:50.234 DEBUG ab7c64738d52 saminfo
## [1] 404 2
## 2020 11 18 16:20:50.234 DEBUG ab7c64738d52 check column and row names
## 2020 11 18 16:20:50.301 DEBUG ab7c64738d52 in design.mat_plus
## 2020 11 18 16:20:50.302 DEBUG ab7c64738d52 in build.design_plus
## 2020 11 18 16:20:50.302 DEBUG ab7c64738d52 after build.design_plus loop
## 2020 11 18 16:20:50.303 DEBUG ab7c64738d52 design
## 2020 11 18 16:20:50.303 DEBUG ab7c64738d52 list.batch_plus(saminfo)
## 2020 11 18 16:20:50.303 DEBUG ab7c64738d52 in list.batch_plus
## 2020 11 18 16:20:50.303 DEBUG ab7c64738d52 list.batch_plus tmp1
## 2020 11 18 16:20:50.304 DEBUG ab7c64738d52 list.batch_plus uniTmp[i]
## 2020 11 18 16:20:50.304 DEBUG ab7c64738d52 list.batch_plus uniTmp[i]
## 2020 11 18 16:20:50.304 DEBUG ab7c64738d52 list.batch_plus batches
## 2020 11 18 16:20:50.304 DEBUG ab7c64738d52 after list.batch_plus(saminfo)
## 2020 11 18 16:20:50.304 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:50.305 DEBUG ab7c64738d52 n.array
## 2020 11 18 16:20:51.665 DEBUG ab7c64738d52 missbatch matrix
## 2020 11 18 16:20:51.677 DEBUG ab7c64738d52 There are 63 genes that were removed because
## 2020 11 18 16:20:52.176 DEBUG ab7c64738d52 B.hat
## 2020 11 18 16:20:52.177 DEBUG ab7c64738d52 grand.mean
## 2020 11 18 16:20:52.395 DEBUG ab7c64738d52 var.pooled
## 2020 11 18 16:20:52.411 DEBUG ab7c64738d52 stand.mean 1
## 2020 11 18 16:20:52.470 DEBUG ab7c64738d52 stand.mean 2
## 2020 11 18 16:20:52.504 DEBUG ab7c64738d52 s.data
## 2020 11 18 16:20:52.988 DEBUG ab7c64738d52 gamma.hat
## 2020 11 18 16:20:53.307 DEBUG ab7c64738d52 plot priors
## 2020 11 18 16:20:53.308 DEBUG ab7c64738d52 plotPrior priorPlotsFile= /builds/BatchEffects

## 2020 11 18 16:20:53.713 DEBUG ab7c64738d52 parametric adjustments
## 2020 11 18 16:20:54.037 DEBUG ab7c64738d52 parametric adjustments
## 2020 11 18 16:20:54.224 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:54.224 DEBUG ab7c64738d52 dim(s.data)=7722 404
## 2020 11 18 16:20:54.225 DEBUG ab7c64738d52 dim(stand.mean)=7722 404
## 2020 11 18 16:20:54.318 DEBUG ab7c64738d52 dim(bayesdata)=7722 404
## 2020 11 18 16:20:54.353 DEBUG ab7c64738d52 Second dim(bayesdata)=7722 404
## 2020 11 18 16:20:54.353 DEBUG ab7c64738d52 keep the dim of Original.dat, keep the gene w
## 2020 11 18 16:20:54.380 DEBUG ab7c64738d52 resultsDat
## 2020 11 18 16:20:54.381 DEBUG ab7c64738d52 after EB MBatch
## 2020 11 18 16:20:54.381 DEBUG ab7c64738d52 inside TEST branch MBatch
## 2020 11 18 16:20:54.385 DEBUG ab7c64738d52 test
## 2020 11 18 16:20:54.385 DEBUG ab7c64738d52 dim(mat1)=10000 86
## 2020 11 18 16:20:54.385 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:54.385 DEBUG ab7c64738d52 dim(mat2)=10000 86
## 2020 11 18 16:20:54.385 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:54.665 DEBUG ab7c64738d52 dim(dat)=7785 172
## 2020 11 18 16:20:54.666 DEBUG ab7c64738d52 dim(dat)=

```



```

## 2020 11 18 16:20:54.672 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:54.672 DEBUG ab7c64738d52 dim(dat)=7785 172
## 2020 11 18 16:20:54.672 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:54.673 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:54.677 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:54.677 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:54.677 DEBUG ab7c64738d52 dim(dat)=7722 172
## 2020 11 18 16:20:54.697 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:54.698 DEBUG ab7c64738d52 dim(stand.data)=7722 172
## 2020 11 18 16:20:54.698 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:54.698 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:54.698 DEBUG ab7c64738d52 dim(batch.design)=172 2
## 2020 11 18 16:20:54.698 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:54.699 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:54.699 DEBUG ab7c64738d52 dim(s.data)=7722 172
## 2020 11 18 16:20:54.699 DEBUG ab7c64738d52 dim(stand.mean)=7722 172
## 2020 11 18 16:20:54.724 DEBUG ab7c64738d52 dim(bayesdata)=7722 172
## 2020 11 18 16:20:54.732 DEBUG ab7c64738d52 Second dim(bayesdata)=7722 172
## 2020 11 18 16:20:54.732 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:54.732 DEBUG ab7c64738d52 dim(bayesData)=7722 172
## 2020 11 18 16:20:54.733 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:54.740 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:54.740 DEBUG ab7c64738d52 Check the adjusted test set
## 2020 11 18 16:20:54.740 DEBUG ab7c64738d52 after train
## 2020 11 18 16:20:54.740 DEBUG ab7c64738d52 EBadj
## 2020 11 18 16:20:54.740 DEBUG ab7c64738d52 EBadj train
## 2020 11 18 16:20:54.741 DEBUG ab7c64738d52 train
## 2020 11 18 16:20:54.741 DEBUG ab7c64738d52 dim(mat1)=10000 202
## 2020 11 18 16:20:54.741 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:54.741 DEBUG ab7c64738d52 dim(mat2)=10000 202
## 2020 11 18 16:20:54.741 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:54.859 DEBUG ab7c64738d52 dim(dat)=7785 404
## 2020 11 18 16:20:54.860 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:54.872 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:54.872 DEBUG ab7c64738d52 dim(dat)=7785 404
## 2020 11 18 16:20:54.872 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:54.873 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:54.887 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:54.887 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:54.887 DEBUG ab7c64738d52 dim(dat)=7722 404
## 2020 11 18 16:20:54.934 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:54.935 DEBUG ab7c64738d52 dim(stand.data)=7722 404
## 2020 11 18 16:20:54.935 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:54.935 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:54.935 DEBUG ab7c64738d52 dim(batch.design)=404 2
## 2020 11 18 16:20:54.936 DEBUG ab7c64738d52 n.batches

```

```

## 2020 11 18 16:20:54.936 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:54.936 DEBUG ab7c64738d52 dim(s.data)=7722 404
## 2020 11 18 16:20:54.936 DEBUG ab7c64738d52 dim(stand.mean)=7722 404
## 2020 11 18 16:20:55.004 DEBUG ab7c64738d52 dim(bayesdata)=7722 404
## 2020 11 18 16:20:55.031 DEBUG ab7c64738d52 Second dim(bayesdata)=7722 404
## 2020 11 18 16:20:55.032 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:55.032 DEBUG ab7c64738d52 dim(bayesData)=7722 404
## 2020 11 18 16:20:55.032 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:55.059 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:55.059 DEBUG ab7c64738d52 objafterEB@trainEB
## 2020 11 18 16:20:55.059 DEBUG ab7c64738d52 EBadj validation
## 2020 11 18 16:20:55.060 DEBUG ab7c64738d52 validation
## 2020 11 18 16:20:55.078 DEBUG ab7c64738d52 dim(mat1)=10000 175
## 2020 11 18 16:20:55.078 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:55.078 DEBUG ab7c64738d52 dim(mat2)=10000 175
## 2020 11 18 16:20:55.079 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:55.175 DEBUG ab7c64738d52 dim(dat)=7785 350
## 2020 11 18 16:20:55.175 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:55.186 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:55.186 DEBUG ab7c64738d52 dim(dat)=7785 350
## 2020 11 18 16:20:55.187 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:55.187 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:55.194 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:55.194 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:55.194 DEBUG ab7c64738d52 dim(dat)=7722 350
## 2020 11 18 16:20:55.252 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:55.252 DEBUG ab7c64738d52 dim(stand.data)=7722 350
## 2020 11 18 16:20:55.252 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:55.253 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:55.253 DEBUG ab7c64738d52 dim(batch.design)=350 2
## 2020 11 18 16:20:55.253 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:55.253 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:55.253 DEBUG ab7c64738d52 dim(s.data)=7722 350
## 2020 11 18 16:20:55.254 DEBUG ab7c64738d52 dim(stand.mean)=7722 350
## 2020 11 18 16:20:55.313 DEBUG ab7c64738d52 dim(bayesdata)=7722 350
## 2020 11 18 16:20:55.330 DEBUG ab7c64738d52 Second dim(bayesdata)=7722 350
## 2020 11 18 16:20:55.330 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:55.330 DEBUG ab7c64738d52 dim(bayesData)=7722 350
## 2020 11 18 16:20:55.331 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:55.351 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:55.351 DEBUG ab7c64738d52 objafterEB@validationEB
## 2020 11 18 16:20:55.351 DEBUG ab7c64738d52 EBadj train for test results
## 2020 11 18 16:20:55.351 DEBUG ab7c64738d52 test
## 2020 11 18 16:20:55.352 DEBUG ab7c64738d52 dim(mat1)=10000 86
## 2020 11 18 16:20:55.352 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:55.352 DEBUG ab7c64738d52 dim(mat2)=10000 86

```

```

## 2020 11 18 16:20:55.352 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:55.410 DEBUG ab7c64738d52 dim(dat)=7785 172
## 2020 11 18 16:20:55.411 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:55.416 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:55.417 DEBUG ab7c64738d52 dim(dat)=7785 172
## 2020 11 18 16:20:55.417 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:55.418 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:55.426 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:55.427 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:55.427 DEBUG ab7c64738d52 dim(dat)=7722 172
## 2020 11 18 16:20:55.447 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:55.447 DEBUG ab7c64738d52 dim(stand.data)=7722 172
## 2020 11 18 16:20:55.448 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:55.448 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:55.448 DEBUG ab7c64738d52 dim(batch.design)=172 2
## 2020 11 18 16:20:55.448 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:55.448 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:55.449 DEBUG ab7c64738d52 dim(s.data)=7722 172
## 2020 11 18 16:20:55.449 DEBUG ab7c64738d52 dim(stand.mean)=7722 172
## 2020 11 18 16:20:55.474 DEBUG ab7c64738d52 dim(bayesdata)=7722 172
## 2020 11 18 16:20:55.482 DEBUG ab7c64738d52 Second dim(bayesdata)=7722 172
## 2020 11 18 16:20:55.482 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:55.482 DEBUG ab7c64738d52 dim(bayesData)=7722 172
## 2020 11 18 16:20:55.483 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:55.490 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:55.490 DEBUG ab7c64738d52 testResults@testEB
## [1] "TestSet1"
## [1] TRUE
## [1] 10000      86
##          TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A2-A04X-01A-21R-A034-07
## A1BG                                7.4960                        5.431
## A1CF                                -2.0160                       -2.016
## A2BP1                               -0.7169                       -2.016
## A2LD1                                7.0290                        5.904
##          TCGA-A2-A0CU-01A-12R-A034-07 TCGA-A2-A0CX-01A-21R-A00Z-07
## A1BG                                6.707                        6.4130
## A1CF                                -2.016                       -0.6197
## A2BP1                               -2.016                       -2.0160
## A2LD1                                5.635                        5.9360
## [1] "TestSet2"
## [1] TRUE
## [1] 10000      86
##          TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A2-A04X-01A-21R-A034-07
## 15E1.2                             -1.9440000                   -0.6882500
## 2'-PDE                             -0.0295625                    0.3431250
## 7A5                                1.8550000                    2.1730000

```

```

## A1BG 0.7486667 -0.6243333
## TCGA-A2-A0CU-01A-12R-A034-07 TCGA-A2-A0CX-01A-21R-A00Z-07
## 15E1.2 -1.42200 -0.36475000
## 2'-PDE 0.43875 0.21425000
## 7A5 -0.96550 0.34000000
## A1BG 0.28800 -0.02466667
## [1] "TrainingSet1"
## [1] TRUE
## [1] 10000 202
## TCGA-A1-A0SJ-01A-11R-A084-07 TCGA-A2-A04N-01A-11R-A115-07
## A1BG 7.820 7.521
## A1CF -2.016 -2.016
## A2BP1 -2.016 -2.016
## A2LD1 7.035 6.995
## TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04T-01A-21R-A034-07
## A1BG 7.7130 6.03800
## A1CF -0.7145 -2.01600
## A2BP1 -2.0160 -0.08512
## A2LD1 6.0800 6.25700
## [1] "TrainingSet2"
## [1] TRUE
## [1] 10000 202
## TCGA-A1-A0SJ-01A-11R-A084-07 TCGA-A2-A04N-01A-11R-A115-07
## 15E1.2 -2.0980000 -0.94900000
## 2'-PDE 0.7919375 -0.11818750
## 7A5 1.4080000 2.64550000
## A1BG 0.7500000 0.08566667
## TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04T-01A-21R-A034-07
## 15E1.2 -1.652250 -1.2670000
## 2'-PDE 0.635625 0.2290625
## 7A5 -0.392000 1.2170000
## A1BG 1.278000 -0.0885000
## [1] "TrainingResults"
## [1] TRUE
## [1] 7785 404
## TCGA-A1-A0SJ-01A-11R-A084-07.Agilent4502
## A1BG 4.3247402
## A2BP1 -0.3851765
## A2M 8.0978466
## A2ML1 1.8998344
## TCGA-A1-A0SJ-01A-11R-A084-07.RNASeqV2
## A1BG 4.2896562
## A2BP1 -2.1547612
## A2M 8.0183429
## A2ML1 -0.4772133
## TCGA-A2-A04N-01A-11R-A115-07.Agilent4502

```

```

## A1BG 3.4531507
## A2BP1 -0.2851838
## A2M 6.7163478
## A2ML1 1.9818843
## TCGA-A2-A04N-01A-11R-A115-07.RNASeqV2
## A1BG 4.0392535
## A2BP1 -2.1547612
## A2M 6.8664714
## A2ML1 -0.9104235
## [1] "ValidationSet1"
## [1] TRUE
## [1] 10000 175
## TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A1-A0SK-01A-12R-A084-07
## A1BG 7.1552 6.440
## A1CF -2.0160 -2.016
## A2BP1 -0.7908 -2.016
## A2LD1 6.0680 8.659
## TCGA-A1-A0SM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG 8.265 6.7681
## A1CF -2.016 -2.016
## A2BP1 -0.829 -2.016
## A2LD1 6.686 6.4740
## [1] "ValidationSet2"
## [1] TRUE
## [1] 10000 175
## TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A1-A0SK-01A-12R-A084-07
## 15E1.2 -1.4477500 -1.2850000
## 2'-PDE 0.0153750 -0.1911250
## 7A5 2.4105000 -1.3005000
## A1BG 0.9493333 -0.5471667
## TCGA-A1-A0SM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## 15E1.2 -2.1540000 -1.5990000
## 2'-PDE 0.3263125 -0.3471250
## 7A5 2.1600000 1.7275000
## A1BG 0.5976667 -0.1568333
## [1] "ValidationResults"
## [1] TRUE
## [1] 7785 350
## TCGA-A1-A0SD-01A-11R-A115-07.Agilent4502
## A1BG 4.586261
## A2BP1 2.363897
## A2M 6.632396
## A2ML1 1.648391
## TCGA-A1-A0SD-01A-11R-A115-07.RNASeqV2
## A1BG 3.732908
## A2BP1 -1.327814

```

```

## A2M 8.122790
## A2ML1 1.673291
## TCGA-A1-A0SK-01A-12R-A084-07.Agilent4502
## A1BG 2.622888
## A2BP1 -1.583156
## A2M 5.507855
## A2ML1 1.337396
## TCGA-A1-A0SK-01A-12R-A084-07.RNASeqV2
## A1BG 3.133951
## A2BP1 -2.154761
## A2M 5.653508
## A2ML1 2.874396
## [1] "CorrectedResults"
## [1] FALSE
## NULL
## NULL

```

Example Output

A list of matrices with the following names().

```

##TestSet1
a matrix of the test set from data set 1

##TestSet2
a matrix of the test set from data set 2

##TrainingSet1
a matrix of the training set from data set 1

##TrainingSet2
a matrix of the training set from data set 2

##TrainingResults
a matrix of the corrected training data

##ValidationSet1
a matrix of the validation set from data set 1

##ValidationSet2
a matrix of the validation set from data set 2

##ValidationResults a matrix of the corrected validation data

##CorrectedResults
If Validation Ratio is 0, the validation set values will be 0 x 0 matrices, but this
list member will contain a matrix of corrected values, using all replicates.

This also generates the prior plots output file like this one.

```

