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_TrainAndValidateReplicates_Structures creates and returns a training and validation set and corrected data, based on the training and validation set.

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_TrainAndValidateReplicates_Structures(theDataMatrix1, theDataMatrix2, theBatchId1, theBatchId2, theEBNP_BatchWithZero, theEBNP_FixDataSet, theEBNP_CorrectForZero, theEBNP_ParametricPriorsFlag, theEBNP_ValidationRatio, theEBNP_TestRatio=0, theSeed = NULL, theEBNP_PriorPlotsFile=NULL, theEBNP_MinSampleNum = 3, theEBNP_AddData1Rows = FALSE, theEBNP_AddData2Rows = FALSE, theTestSeed=NULL)

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_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)

##theEBNP_ValidationRatio
A numeric. If 0, do not use a training set. For this function, the value should be less than 1. 0.33 means 33 percent of the replicates available are used for validation with 67 percent for training.

##theEBNP_TestRatio
A numeric. If 0, do not use a test set. For this function, the value should be less than 1. 0.33 means 33 percent of the replicates available are used for the test set with 67 percent for training.

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

##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.

##theEBNP_AddData1Rows
Defaults to FALSE. TRUE indicates rows in data set 2 not in 1 should be added as all NA.

##theEBNP_AddData2Rows
Defaults to FALSE. TRUE indicates rows in data set 1 not in 2 should be added as all NA.

##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.

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])
        }
```

```r
    })
  ))
  foo
}

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

# remove duplicates from rows (genes/probes)
removeDuplicatesFromRows <- function(theMatrix)
{
  indexOfDuplicates <- which(duplicated(rownames(theMatrix)))
  if (length(indexOfDuplicates) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[-indexOfDuplicates, ]
  }
  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))
print("EBNPlus_TrainAndValidateReplicates_Structures")
resultsList <- EBNPlus_TrainAndValidateReplicates_Structures(
  theDataMatrix1, theDataMatrix2, theBatchId1, theBatchId2,
  theEBNP_BatchWithZero="1",
  theEBNP_FixDataSet=as.numeric(NA),
  theEBNP_CorrectForZero=TRUE,
  theEBNP_ParametricPriorsFlag=TRUE,
  theEBNP_ValidationRatio=0.3,
  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 16:20:36.148 DEBUG ab7c64738d52 starting BeaEBNPlus
## 2020 11 18 16:20:36.148 DEBUG ab7c64738d52 MBatch Version: BEA_VERSION_TIMESTAMP
## 2020 11 18 16:20:36.148 DEBUG ab7c64738d52 BeaEBNPlus theEBNP_AddData1Rows= FALSE
## 2020 11 18 16:20:36.148 DEBUG ab7c64738d52 BeaEBNPlus theEBNP_AddData2Rows= FALSE
## 2020 11 18 16:20:36.149 DEBUG ab7c64738d52 EBNPlus theEBNP_AddData1Rows= FALSE
## 2020 11 18 16:20:36.149 DEBUG ab7c64738d52 EBNPlus theEBNP_AddData2Rows= FALSE
## 2020 11 18 16:20:36.150 DEBUG ab7c64738d52 EBNPlus theEBNP_PriorPlotsFile= /builds/BatchF
## 2020 11 18 16:20:36.150 DEBUG ab7c64738d52 dim(theData1)=10000 1215
## 2020 11 18 16:20:36.150 DEBUG ab7c64738d52 dim(theData2)=10000 600
## 2020 11 18 16:20:36.150 DEBUG ab7c64738d52 remove unknown genes, that start with ?
## 2020 11 18 16:20:36.271 DEBUG ab7c64738d52 check on adding missing rows
## 2020 11 18 16:20:36.271 DEBUG ab7c64738d52 dim(theData1)=10000 1215
## 2020 11 18 16:20:36.272 DEBUG ab7c64738d52 dim(theData2)=10000 600
## 2020 11 18 16:20:36.272 DEBUG ab7c64738d52 make EBNplus
## 2020 11 18 16:20:36.272 DEBUG ab7c64738d52 before callNextMethod
## 2020 11 18 16:20:36.275 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:36.276 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:36.276 DEBUG ab7c64738d52 removing column duplicates
## 2020 11 18 16:20:36.276 DEBUG ab7c64738d52 removeDuplicatesFromColumns
## 2020 11 18 16:20:36.276 DEBUG ab7c64738d52 removeDuplicatesFromColumns
## 2020 11 18 16:20:36.277 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:36.277 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:36.277 DEBUG ab7c64738d52 removing row duplicates
## 2020 11 18 16:20:36.277 DEBUG ab7c64738d52 removeDuplicatesFromRows
## 2020 11 18 16:20:36.278 DEBUG ab7c64738d52 removeDuplicatesFromRows
## 2020 11 18 16:20:36.278 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:36.279 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:36.279 DEBUG ab7c64738d52 after callNextMethod
## 2020 11 18 16:20:36.279 DEBUG ab7c64738d52 dim(mData1)=10000 1215
## 2020 11 18 16:20:36.279 DEBUG ab7c64738d52 dim(mData2)=10000 600
## 2020 11 18 16:20:36.279 DEBUG ab7c64738d52 after EBNplus
## 2020 11 18 16:20:36.280 DEBUG ab7c64738d52 dim(ebObj@mData1)=10000 1215
## 2020 11 18 16:20:36.280 DEBUG ab7c64738d52 dim(ebObj@mData2)=10000 600
## 2020 11 18 16:20:36.281 DEBUG ab7c64738d52 getBiComOrder
## 2020 11 18 16:20:36.282 DEBUG ab7c64738d52 makeCommonRows
```

```
## 2020 11 18 16:20:36.398 DEBUG ab7c64738d52 dim(mData1)=7785 1215
## 2020 11 18 16:20:36.399 DEBUG ab7c64738d52 dim(mData2)=7785 600
## 2020 11 18 16:20:36.399 DEBUG ab7c64738d52 makeCommonCols
## 2020 11 18 16:20:36.675 DEBUG ab7c64738d52 dim(mData1)=7785 586
## 2020 11 18 16:20:36.676 DEBUG ab7c64738d52 dim(mData2)=7785 586
## 2020 11 18 16:20:36.676 DEBUG ab7c64738d52 asSameOrder
## 2020 11 18 16:20:36.676 DEBUG ab7c64738d52 asSameOrder before
## 2020 11 18 16:20:36.676 DEBUG ab7c64738d52 asSameOrder after if 1
## 2020 11 18 16:20:36.701 DEBUG ab7c64738d52 asSameOrder after if 2
## 2020 11 18 16:20:36.701 DEBUG ab7c64738d52 row.names(mat1)
## 2020 11 18 16:20:36.701 DEBUG ab7c64738d52 rownames(mat1)
## 2020 11 18 16:20:36.701 DEBUG ab7c64738d52 row.names(mat2)
## 2020 11 18 16:20:36.701 DEBUG ab7c64738d52 rownames(mat2)
## 2020 11 18 16:20:36.702 DEBUG ab7c64738d52 m.i
## 2020 11 18 16:20:36.726 DEBUG ab7c64738d52 asSameOrder after if 3
## 2020 11 18 16:20:36.726 DEBUG ab7c64738d52 dim(ebObj@mat1Com)=7785 586
## 2020 11 18 16:20:36.727 DEBUG ab7c64738d52 dim(ebObj@mat2Com)=7785 586
## 2020 11 18 16:20:36.727 DEBUG ab7c64738d52 getValidationSet
## 2020 11 18 16:20:36.727 INFO ab7c64738d52 getValidationSet seed= 314
## 2020 11 18 16:20:36.727 DEBUG ab7c64738d52 validationRatio=0.3
## 2020 11 18 16:20:36.728 DEBUG ab7c64738d52 numSamples=586
## 2020 11 18 16:20:36.757 DEBUG ab7c64738d52 dim(ebObj@mat1Validation)=7785 175
## 2020 11 18 16:20:36.758 DEBUG ab7c64738d52 dim(ebObj@mat2Validation)=7785 175
## 2020 11 18 16:20:36.758 DEBUG ab7c64738d52 dim(ebObj@mat1Train)=7785 411
## 2020 11 18 16:20:36.758 DEBUG ab7c64738d52 dim(ebObj@mat2Train)=7785 411
## 2020 11 18 16:20:36.758 DEBUG ab7c64738d52 before Test Set
## 2020 11 18 16:20:36.780 DEBUG ab7c64738d52 after Test Set
## 2020 11 18 16:20:36.841 DEBUG ab7c64738d52 train
## 2020 11 18 16:20:36.842 DEBUG ab7c64738d52 inside train
## 2020 11 18 16:20:36.842 DEBUG ab7c64738d52 train theEBNP_PriorPlotsFile= /builds/BatchEff
## 2020 11 18 16:20:36.867 DEBUG ab7c64738d52 colnames(Object@mat1Train)=TCGA-A1-A0SH-01A-11
## 2020 11 18 16:20:36.867 DEBUG ab7c64738d52 colnames(Object@mat2Train)=TCGA-A1-A0SH-01A-11
## 2020 11 18 16:20:36.867 DEBUG ab7c64738d52 inside train, call getData4EB
## 2020 11 18 16:20:37.294 DEBUG ab7c64738d52 data4EB <- as.matrix
## 2020 11 18 16:20:37.294 DEBUG ab7c64738d52 dim(data4EB)=7785 576
## 2020 11 18 16:20:37.295 DEBUG ab7c64738d52 cbinds and rbinds
## 2020 11 18 16:20:37.295 DEBUG ab7c64738d52 Object@DF1batch= RNASeqV2
## 2020 11 18 16:20:37.295 DEBUG ab7c64738d52 Object@DF2batch= Agilent4502
## 2020 11 18 16:20:37.296 DEBUG ab7c64738d52 row.names(sampBatch)
## 2020 11 18 16:20:37.296 DEBUG ab7c64738d52 TCGA-A1-A0SH-01A-11R-A084-07.RNASeqV2 TCGA-A1-
## 2020 11 18 16:20:37.296 DEBUG ab7c64738d52 colnames(sampBatch)
## 2020 11 18 16:20:37.296 DEBUG ab7c64738d52 sample batch
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 inside train, after getData4EB
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 call EB MBatch
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 TDC HERE trainEB mbatch
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 EBNplus args Obj
```

```
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 EBNplus args par.prior
## 2020 11 18 16:20:37.297 DEBUG ab7c64738d52 EBNplus args minSampleNum
## 2020 11 18 16:20:37.298 DEBUG ab7c64738d52 EBNplus theEBNP_PriorPlotsFile= /builds/BatchB
## 2020 11 18 16:20:37.298 DEBUG ab7c64738d52 dat
## [1] 7785  576
## 2020 11 18 16:20:37.298 DEBUG ab7c64738d52 saminfo
## [1] 576    2
## 2020 11 18 16:20:37.298 DEBUG ab7c64738d52 check column and row names
## 2020 11 18 16:20:37.465 DEBUG ab7c64738d52 in design.mat_plus
## 2020 11 18 16:20:37.466 DEBUG ab7c64738d52 in build.design_plus
## 2020 11 18 16:20:37.466 DEBUG ab7c64738d52 after build.design_plus loop
## 2020 11 18 16:20:37.467 DEBUG ab7c64738d52 design
## 2020 11 18 16:20:37.467 DEBUG ab7c64738d52 list.batch_plus(saminfo)
## 2020 11 18 16:20:37.467 DEBUG ab7c64738d52 in list.batch_plus
## 2020 11 18 16:20:37.467 DEBUG ab7c64738d52 list.batch_plus tmp1
## 2020 11 18 16:20:37.468 DEBUG ab7c64738d52 list.batch_plus uniTmp[i]
## 2020 11 18 16:20:37.468 DEBUG ab7c64738d52 list.batch_plus uniTmp[i]
## 2020 11 18 16:20:37.468 DEBUG ab7c64738d52 list.batch_plus batches
## 2020 11 18 16:20:37.468 DEBUG ab7c64738d52 after list.batch_plus(saminfo)
## 2020 11 18 16:20:37.468 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:37.469 DEBUG ab7c64738d52 n.array
## 2020 11 18 16:20:38.980 DEBUG ab7c64738d52 missbatch matrix
## 2020 11 18 16:20:39.010 DEBUG ab7c64738d52 There are  46  genes that were removed because
## 2020 11 18 16:20:39.935 DEBUG ab7c64738d52 B.hat
## 2020 11 18 16:20:39.935 DEBUG ab7c64738d52 grand.mean
## 2020 11 18 16:20:40.227 DEBUG ab7c64738d52 var.pooled
## 2020 11 18 16:20:40.249 DEBUG ab7c64738d52 stand.mean 1
## 2020 11 18 16:20:40.341 DEBUG ab7c64738d52 stand.mean 2
## 2020 11 18 16:20:40.392 DEBUG ab7c64738d52 s.data
## 2020 11 18 16:20:40.992 DEBUG ab7c64738d52 gamma.hat
## 2020 11 18 16:20:41.373 DEBUG ab7c64738d52 plot priors
## 2020 11 18 16:20:41.373 DEBUG ab7c64738d52 plotPrior priorPlotsFile= /builds/BatchEffects

## 2020 11 18 16:20:41.769 DEBUG ab7c64738d52 parametric adjustments
## 2020 11 18 16:20:42.332 DEBUG ab7c64738d52 parametric adjustments
## 2020 11 18 16:20:42.619 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:42.619 DEBUG ab7c64738d52 dim(s.data)=7739 576
## 2020 11 18 16:20:42.619 DEBUG ab7c64738d52 dim(stand.mean)=7739 576
## 2020 11 18 16:20:42.748 DEBUG ab7c64738d52 dim(bayesdata)=7739 576
## 2020 11 18 16:20:42.799 DEBUG ab7c64738d52 Second dim(bayesdata)=7739 576
## 2020 11 18 16:20:42.799 DEBUG ab7c64738d52 keep the dim of Original.dat, keep the gene wi
## 2020 11 18 16:20:43.068 DEBUG ab7c64738d52 resultsDat
## 2020 11 18 16:20:43.069 DEBUG ab7c64738d52 after EB MBatch
## 2020 11 18 16:20:43.069 DEBUG ab7c64738d52 inside TEST branch MBatch
## 2020 11 18 16:20:43.073 DEBUG ab7c64738d52 test
## 2020 11 18 16:20:43.073 DEBUG ab7c64738d52 dim(mat1)=7785 123
```

```
## 2020 11 18 16:20:43.074 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:43.074 DEBUG ab7c64738d52 dim(mat2)=7785 123
## 2020 11 18 16:20:43.074 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:43.136 DEBUG ab7c64738d52 dim(dat)=7785 246
## 2020 11 18 16:20:43.136 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:43.144 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:43.144 DEBUG ab7c64738d52 dim(dat)=7785 246
## 2020 11 18 16:20:43.144 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:43.145 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:43.150 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:43.150 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:43.151 DEBUG ab7c64738d52 dim(dat)=7739 246
## 2020 11 18 16:20:43.179 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:43.180 DEBUG ab7c64738d52 dim(stand.data)=7739 246
## 2020 11 18 16:20:43.180 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:43.180 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:43.180 DEBUG ab7c64738d52 dim(batch.design)=246 2
## 2020 11 18 16:20:43.181 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:43.181 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:43.181 DEBUG ab7c64738d52 dim(s.data)=7739 246
## 2020 11 18 16:20:43.181 DEBUG ab7c64738d52 dim(stand.mean)=7739 246
## 2020 11 18 16:20:43.238 DEBUG ab7c64738d52 dim(bayesdata)=7739 246
## 2020 11 18 16:20:43.259 DEBUG ab7c64738d52 Second dim(bayesdata)=7739 246
## 2020 11 18 16:20:43.259 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:43.259 DEBUG ab7c64738d52 dim(bayesData)=7739 246
## 2020 11 18 16:20:43.260 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:43.277 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:43.277 DEBUG ab7c64738d52 Check the adjusted test set
## 2020 11 18 16:20:43.277 DEBUG ab7c64738d52 after train
## 2020 11 18 16:20:43.277 DEBUG ab7c64738d52 EBadj
## 2020 11 18 16:20:43.278 DEBUG ab7c64738d52 EBadj validation for validation results
## 2020 11 18 16:20:43.344 DEBUG ab7c64738d52 validation
## 2020 11 18 16:20:43.359 DEBUG ab7c64738d52 dim(mat1)=7785 175
## 2020 11 18 16:20:43.360 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:43.360 DEBUG ab7c64738d52 dim(mat2)=7785 175
## 2020 11 18 16:20:43.360 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:43.473 DEBUG ab7c64738d52 dim(dat)=7785 350
## 2020 11 18 16:20:43.473 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:43.484 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:43.484 DEBUG ab7c64738d52 dim(dat)=7785 350
## 2020 11 18 16:20:43.484 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:43.485 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:43.492 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:43.492 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:43.492 DEBUG ab7c64738d52 dim(dat)=7739 350
## 2020 11 18 16:20:43.539 DEBUG ab7c64738d52 after getStandData
```

```
## 2020 11 18 16:20:43.540 DEBUG ab7c64738d52 dim(stand.data)=7739 350
## 2020 11 18 16:20:43.540 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:43.540 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:43.540 DEBUG ab7c64738d52 dim(batch.design)=350 2
## 2020 11 18 16:20:43.541 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:43.541 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:43.541 DEBUG ab7c64738d52 dim(s.data)=7739 350
## 2020 11 18 16:20:43.541 DEBUG ab7c64738d52 dim(stand.mean)=7739 350
## 2020 11 18 16:20:43.629 DEBUG ab7c64738d52 dim(bayesdata)=7739 350
## 2020 11 18 16:20:43.660 DEBUG ab7c64738d52 Second dim(bayesdata)=7739 350
## 2020 11 18 16:20:43.660 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:43.660 DEBUG ab7c64738d52 dim(bayesData)=7739 350
## 2020 11 18 16:20:43.661 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:43.686 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:43.686 DEBUG ab7c64738d52 objafterEB@validationEB
## 2020 11 18 16:20:43.686 DEBUG ab7c64738d52 EBadj train for training results
## 2020 11 18 16:20:43.686 DEBUG ab7c64738d52 train
## 2020 11 18 16:20:43.687 DEBUG ab7c64738d52 dim(mat1)=7785 288
## 2020 11 18 16:20:43.687 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:43.687 DEBUG ab7c64738d52 dim(mat2)=7785 288
## 2020 11 18 16:20:43.687 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:43.821 DEBUG ab7c64738d52 dim(dat)=7785 576
## 2020 11 18 16:20:43.822 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:43.839 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:43.839 DEBUG ab7c64738d52 dim(dat)=7785 576
## 2020 11 18 16:20:43.839 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:43.840 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:43.850 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:43.850 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:43.851 DEBUG ab7c64738d52 dim(dat)=7739 576
## 2020 11 18 16:20:43.959 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:43.959 DEBUG ab7c64738d52 dim(stand.data)=7739 576
## 2020 11 18 16:20:43.959 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:43.960 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:43.960 DEBUG ab7c64738d52 dim(batch.design)=576 2
## 2020 11 18 16:20:43.960 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:43.960 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:43.960 DEBUG ab7c64738d52 dim(s.data)=7739 576
## 2020 11 18 16:20:43.961 DEBUG ab7c64738d52 dim(stand.mean)=7739 576
## 2020 11 18 16:20:44.072 DEBUG ab7c64738d52 dim(bayesdata)=7739 576
## 2020 11 18 16:20:44.101 DEBUG ab7c64738d52 Second dim(bayesdata)=7739 576
## 2020 11 18 16:20:44.101 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:44.101 DEBUG ab7c64738d52 dim(bayesData)=7739 576
## 2020 11 18 16:20:44.101 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:44.139 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:44.139 DEBUG ab7c64738d52 trainingResults@trainEB
```

```
## 2020 11 18 16:20:44.139 DEBUG ab7c64738d52 EBadj train for test results
## 2020 11 18 16:20:44.139 DEBUG ab7c64738d52 test
## 2020 11 18 16:20:44.139 DEBUG ab7c64738d52 dim(mat1)=7785 123
## 2020 11 18 16:20:44.140 DEBUG ab7c64738d52 dim(mat1)=
## 2020 11 18 16:20:44.140 DEBUG ab7c64738d52 dim(mat2)=7785 123
## 2020 11 18 16:20:44.140 DEBUG ab7c64738d52 dim(mat2)=
## 2020 11 18 16:20:44.202 DEBUG ab7c64738d52 dim(dat)=7785 246
## 2020 11 18 16:20:44.202 DEBUG ab7c64738d52 dim(dat)=
## 2020 11 18 16:20:44.210 DEBUG ab7c64738d52 after parlist
## 2020 11 18 16:20:44.210 DEBUG ab7c64738d52 dim(dat)=7785 246
## 2020 11 18 16:20:44.210 DEBUG ab7c64738d52 valid.genes
## 2020 11 18 16:20:44.216 DEBUG ab7c64738d52 nonValidGenes
## 2020 11 18 16:20:44.221 DEBUG ab7c64738d52 Non-valid genes were removed before adjustment
## 2020 11 18 16:20:44.222 DEBUG ab7c64738d52 after Non-valid
## 2020 11 18 16:20:44.222 DEBUG ab7c64738d52 dim(dat)=7739 246
## 2020 11 18 16:20:44.251 DEBUG ab7c64738d52 after getStandData
## 2020 11 18 16:20:44.251 DEBUG ab7c64738d52 dim(stand.data)=7739 246
## 2020 11 18 16:20:44.251 DEBUG ab7c64738d52 batches
## 2020 11 18 16:20:44.251 DEBUG ab7c64738d52 after batch.design
## 2020 11 18 16:20:44.251 DEBUG ab7c64738d52 dim(batch.design)=246 2
## 2020 11 18 16:20:44.252 DEBUG ab7c64738d52 n.batches
## 2020 11 18 16:20:44.252 DEBUG ab7c64738d52 in getBayesData
## 2020 11 18 16:20:44.252 DEBUG ab7c64738d52 dim(s.data)=7739 246
## 2020 11 18 16:20:44.252 DEBUG ab7c64738d52 dim(stand.mean)=7739 246
## 2020 11 18 16:20:44.293 DEBUG ab7c64738d52 dim(bayesdata)=7739 246
## 2020 11 18 16:20:44.305 DEBUG ab7c64738d52 Second dim(bayesdata)=7739 246
## 2020 11 18 16:20:44.305 DEBUG ab7c64738d52 after bayesData
## 2020 11 18 16:20:44.305 DEBUG ab7c64738d52 dim(bayesData)=7739 246
## 2020 11 18 16:20:44.306 DEBUG ab7c64738d52 EBadj datNonValid
## 2020 11 18 16:20:44.317 DEBUG ab7c64738d52 EBadj EBadj after non valid
## 2020 11 18 16:20:44.317 DEBUG ab7c64738d52 testResults@testEB
## 2020 11 18 16:20:44.355 DEBUG ab7c64738d52 finishing BeaEBNPlus
## [1] "TestSet1"
## [1] TRUE
## [1] 7785  123
##       TCGA-A1-A0SE-01A-11R-A084-07 TCGA-A1-A0SP-01A-11R-A084-07
## A1BG                        7.5920                        5.210
## A2BP1                      -2.0160                       -2.016
## A2M                        13.9410                       12.357
## A2ML1                       0.3038                        9.738
##       TCGA-A2-A04Y-01A-21R-A034-07 TCGA-A2-A0CP-01A-11R-A034-07
## A1BG                        7.5074                      8.54000
## A2BP1                       1.5477                     -0.12437
## A2M                        13.5140                     14.51200
## A2ML1                      -2.0160                     -0.12437
## [1] "TestSet2"
```

11

```
## [1] TRUE
## [1] 7785  123
##       TCGA-A1-A0SE-01A-11R-A084-07 TCGA-A1-A0SP-01A-11R-A084-07
## A1BG                     0.2798333                    -1.291167
## A2BP1                   -1.5346667                    -1.133833
## A2M                      1.0385000                    -0.549500
## A2ML1                    0.3715000                     1.332000
##       TCGA-A2-A04Y-01-21R-A034-07 TCGA-A2-A0CP-01A-11R-A034-07
## A1BG                     0.3046667                    0.9571667
## A2BP1                   -0.8158333                   -1.4495000
## A2M                      0.5885000                    1.2132500
## A2ML1                    0.6145000                    1.1390000
## [1] "TrainingSet1"
## [1] TRUE
## [1] 7785  288
##       TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A1-A0SJ-01A-11R-A084-07
## A1BG                        7.4960                     7.820000
## A2BP1                      -0.7169                    -2.016000
## A2M                        13.7700                    14.432000
## A2ML1                       0.7585                    -0.030913
##       TCGA-A1-A0SK-01A-12R-A084-07 TCGA-A1-A0SM-01A-11R-A084-07
## A1BG                         6.440                        8.265
## A2BP1                       -2.016                       -0.829
## A2M                         12.032                       13.120
## A2ML1                        4.652                       -0.829
## [1] "TrainingSet2"
## [1] TRUE
## [1] 7785  288
##       TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A1-A0SJ-01A-11R-A084-07
## A1BG                     0.7486667                     0.750000
## A2BP1                   -1.3048333                    -1.357833
## A2M                      1.1605000                     1.677750
## A2ML1                    0.3575000                     0.518500
##       TCGA-A1-A0SK-01A-12R-A084-07 TCGA-A1-A0SM-01A-11R-A084-07
## A1BG                    -0.5471667                    0.5976667
## A2BP1                   -2.1845000                   -1.1756667
## A2M                     -0.8597500                    0.5745000
## A2ML1                    0.3060000                    0.2635000
## [1] "TrainingResults"
## [1] TRUE
## [1] 7785  576
##      TCGA-A1-A0SH-01A-11R-A084-07.Agilent4502
## A1BG                                 4.2922928
## A2BP1                               -0.3358343
## A2M                                  7.5542896
## A2ML1                                1.4633094
```

12

```
##       TCGA-A1-A0SH-01A-11R-A084-07.RNASeqV2
## A1BG                           4.02642841
## A2BP1                         -1.30645343
## A2M                            7.37907814
## A2ML1                          0.06540906
##       TCGA-A1-A0SJ-01A-11R-A084-07.Agilent4502
## A1BG                           4.2939693
## A2BP1                         -0.4097362
## A2M                            8.0769652
## A2ML1                          1.9279318
##       TCGA-A1-A0SJ-01A-11R-A084-07.RNASeqV2
## A1BG                           4.3031814
## A2BP1                         -2.2213301
## A2M                            8.0374491
## A2ML1                         -0.4942765
## [1] "ValidationSet1"
## [1] TRUE
## [1] 7785  175
##       TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG                        7.1552                       6.7681
## A2BP1                      -0.7908                      -2.0160
## A2M                        14.5380                      13.1800
## A2ML1                       2.9738                      10.6840
##       TCGA-A2-A04V-01A-21R-A034-07 TCGA-A2-A04W-01A-31R-A115-07
## A1BG                        7.2050                        7.185
## A2BP1                      -2.0160                       -2.016
## A2M                        13.2340                       14.490
## A2ML1                       0.9083                        6.947
## [1] "ValidationSet2"
## [1] TRUE
## [1] 7785  175
##       TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG                     0.9493333                   -0.1568333
## A2BP1                    0.5391667                   -0.8194000
## A2M                      0.2420000                    0.2440000
## A2ML1                    0.4235000                    1.9145000
##       TCGA-A2-A04V-01A-21R-A034-07 TCGA-A2-A04W-01A-31R-A115-07
## A1BG                       0.44500                   -0.7923333
## A2BP1                     -1.18800                   -1.3916667
## A2M                        0.17925                    1.9417500
## A2ML1                      0.74500                    0.9930000
## [1] "ValidationResults"
## [1] TRUE
## [1] 7785  350
##       TCGA-A1-A0SD-01A-11R-A115-07.Agilent4502
## A1BG                                  4.544614
```

```
## A2BP1                                   2.235393
## A2M                                     6.626155
## A2ML1                                   1.653776
##       TCGA-A1-A0SD-01A-11R-A115-07.RNASeqV2
## A1BG                                    3.735325
## A2BP1                                  -1.358497
## A2M                                     8.142868
## A2ML1                                   1.636034
##       TCGA-A2-A04P-01A-31R-A034-07.Agilent4502
## A1BG                                    3.1537028
## A2BP1                                   0.3410418
## A2M                                     6.6281763
## A2ML1                                   5.9565837
##       TCGA-A2-A04P-01A-31R-A034-07.RNASeqV2
## A1BG                                    3.404674
## A2BP1                                  -2.221330
## A2M                                     6.792313
## A2ML1                                   7.102485
## [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.

Density Plot

Normal Q-Q Plot

Density Plot

Q-Q Plot