

Using MBatch Corrections: EBNPlus__TrainAndValidateReplicates__Structures

Tod Casasent

2019-10-10

1 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.

2 Algorithm

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

3 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.

4 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)

5 Arguments

5.1 theDataMatrix1

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

5.2 theDataMatrix2

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

5.3 theBatchId1

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

5.4 theBatchId2

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

5.5 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.

5.6 theEBNP__FixDataSet

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

5.7 theEBNP__CorrectForZero

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

5.8 theEBNP__ParametricPriorsFlag

Use parametric adjustments for corrections (TRUE or FALSE)

5.9 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.

5.10 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.

5.11 theSeed

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

5.12 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.

5.13 theEBNP_MinSampleNum

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

5.14 theEBNP_AddData1Rows

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

5.15 theEBNP_AddData2Rows

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

5.16 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.

6 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)

  theDataFile1="/bea_testing/MATRIX_DATA/brca_rnaseq2_matrix_data.tsv"
  theDataFile2="/bea_testing/MATRIX_DATA/brca_agi4502_matrix_data.tsv"
  theOutputDir="/bea_testing/output/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)
{
  indexOfDuplicats <- which(duplicated(colnames(theMatrix)))
  if (length(indexOfDuplicats) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[ ,-indexOfDuplicats]
  }
  return(theMatrix)
}

# remove duplicates from rows (genes/probes)
removeDuplicatesFromRows <- function(theMatrix)
{
  indexOfDuplicats <- which(duplicated(rownames(theMatrix)))
  if (length(indexOfDuplicats) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[-indexOfDuplicats, ]
  }
  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"
## 2019 10 10 11:19:40.781 DEBUG megazone23 starting BeaEBNplus
## 2019 10 10 11:19:40.782 DEBUG megazone23 MBatch Version: 2019-09-04-1100
## 2019 10 10 11:19:40.782 DEBUG megazone23 BeaEBNplus theEBNP_AddData1Rows= FALSE
## 2019 10 10 11:19:40.782 DEBUG megazone23 BeaEBNplus theEBNP_AddData2Rows= FALSE

```

```

## 2019 10 10 11:19:40.783 DEBUG megazone23 EBNPlus theEBNP_AddData1Rows= FALSE
## 2019 10 10 11:19:40.783 DEBUG megazone23 EBNPlus theEBNP_AddData2Rows= FALSE
## 2019 10 10 11:19:40.783 DEBUG megazone23 EBNPlus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2019 10 10 11:19:40.783 DEBUG megazone23 dim(theData1)=10000 1215
## 2019 10 10 11:19:40.784 DEBUG megazone23 dim(theData2)=10000 600
## 2019 10 10 11:19:40.784 DEBUG megazone23 remove unknown genes, that start with ?
## 2019 10 10 11:19:40.889 DEBUG megazone23 check on adding missing rows
## 2019 10 10 11:19:40.890 DEBUG megazone23 dim(theData1)=10000 1215
## 2019 10 10 11:19:40.890 DEBUG megazone23 dim(theData2)=10000 600
## 2019 10 10 11:19:40.890 DEBUG megazone23 make EBNplus
## 2019 10 10 11:19:40.890 DEBUG megazone23 before callNextMethod
## 2019 10 10 11:19:40.894 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:19:40.894 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:19:40.894 DEBUG megazone23 removing column duplicates
## 2019 10 10 11:19:40.894 DEBUG megazone23 removeDuplicatesFromColumns
## 2019 10 10 11:19:40.895 DEBUG megazone23 removeDuplicatesFromColumns
## 2019 10 10 11:19:40.895 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:19:40.895 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:19:40.896 DEBUG megazone23 removing row duplicates
## 2019 10 10 11:19:40.896 DEBUG megazone23 removeDuplicatesFromRows
## 2019 10 10 11:19:40.896 DEBUG megazone23 removeDuplicatesFromRows
## 2019 10 10 11:19:40.897 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:19:40.897 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:19:40.897 DEBUG megazone23 after callNextMethod
## 2019 10 10 11:19:40.897 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:19:40.897 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:19:40.898 DEBUG megazone23 after EBNplus
## 2019 10 10 11:19:40.898 DEBUG megazone23 dim(ebObj@mData1)=10000 1215
## 2019 10 10 11:19:40.898 DEBUG megazone23 dim(ebObj@mData2)=10000 600
## 2019 10 10 11:19:40.900 DEBUG megazone23 getBiComOrder
## 2019 10 10 11:19:40.901 DEBUG megazone23 makeCommonRows
## 2019 10 10 11:19:40.989 DEBUG megazone23 dim(mData1)=7785 1215
## 2019 10 10 11:19:40.990 DEBUG megazone23 dim(mData2)=7785 600
## 2019 10 10 11:19:40.990 DEBUG megazone23 makeCommonCols
## 2019 10 10 11:19:41.205 DEBUG megazone23 dim(mData1)=7785 586
## 2019 10 10 11:19:41.205 DEBUG megazone23 dim(mData2)=7785 586
## 2019 10 10 11:19:41.206 DEBUG megazone23 asSameOrder
## 2019 10 10 11:19:41.206 DEBUG megazone23 asSameOrder before
## 2019 10 10 11:19:41.206 DEBUG megazone23 asSameOrder after if 1
## 2019 10 10 11:19:41.214 DEBUG megazone23 asSameOrder after if 2
## 2019 10 10 11:19:41.214 DEBUG megazone23 row.names(mat1)
## 2019 10 10 11:19:41.215 DEBUG megazone23 rownames(mat1)
## 2019 10 10 11:19:41.215 DEBUG megazone23 row.names(mat2)
## 2019 10 10 11:19:41.215 DEBUG megazone23 rownames(mat2)
## 2019 10 10 11:19:41.216 DEBUG megazone23 m.i
## 2019 10 10 11:19:41.225 DEBUG megazone23 asSameOrder after if 3
## 2019 10 10 11:19:41.225 DEBUG megazone23 dim(ebObj@mat1Com)=7785 586
## 2019 10 10 11:19:41.225 DEBUG megazone23 dim(ebObj@mat2Com)=7785 586
## 2019 10 10 11:19:41.226 DEBUG megazone23 getValidationSet
## 2019 10 10 11:19:41.226 INFO megazone23 getValidationSet seed= 314
## 2019 10 10 11:19:41.226 DEBUG megazone23 validationRatio=0.3
## 2019 10 10 11:19:41.226 DEBUG megazone23 numSamples=586
## 2019 10 10 11:19:41.252 DEBUG megazone23 dim(ebObj@mat1Validation)=7785 175
## 2019 10 10 11:19:41.252 DEBUG megazone23 dim(ebObj@mat2Validation)=7785 175

```

```

## 2019 10 10 11:19:41.253 DEBUG megazone23 dim(ebObj@mat1Train)=7785 411
## 2019 10 10 11:19:41.253 DEBUG megazone23 dim(ebObj@mat2Train)=7785 411
## 2019 10 10 11:19:41.253 DEBUG megazone23 before Test Set
## 2019 10 10 11:19:41.280 DEBUG megazone23 after Test Set
## 2019 10 10 11:19:41.362 DEBUG megazone23 train
## 2019 10 10 11:19:41.362 DEBUG megazone23 inside train
## 2019 10 10 11:19:41.363 DEBUG megazone23 train theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_T
## 2019 10 10 11:19:41.369 DEBUG megazone23 colnames(Object@mat1Train)=TCGA-A1-AOSH-01A-11R-A084-07.RNA
## 2019 10 10 11:19:41.369 DEBUG megazone23 colnames(Object@mat2Train)=TCGA-A1-AOSH-01A-11R-A084-07.Agi
## 2019 10 10 11:19:41.369 DEBUG megazone23 inside train, call getData4EB
## 2019 10 10 11:19:41.742 DEBUG megazone23 data4EB <- as.matrix
## 2019 10 10 11:19:41.742 DEBUG megazone23 dim(data4EB)=7785 576
## 2019 10 10 11:19:41.742 DEBUG megazone23 cbinds and rbinds
## 2019 10 10 11:19:41.742 DEBUG megazone23 Object@DF1batch= RNASeqV2
## 2019 10 10 11:19:41.743 DEBUG megazone23 Object@DF2batch= Agilent4502
## 2019 10 10 11:19:41.744 DEBUG megazone23 row.names(sampBatch)
## 2019 10 10 11:19:41.744 DEBUG megazone23 TCGA-A1-AOSH-01A-11R-A084-07.RNASeqV2 TCGA-A1-AOSJ-01A-11R-
## 2019 10 10 11:19:41.745 DEBUG megazone23 colnames(sampBatch)
## 2019 10 10 11:19:41.745 DEBUG megazone23 sample batch
## 2019 10 10 11:19:41.745 DEBUG megazone23 inside train, after getData4EB
## 2019 10 10 11:19:41.745 DEBUG megazone23 call EB MBatch
## 2019 10 10 11:19:41.745 DEBUG megazone23 TDC HERE trainEB mbatch
## 2019 10 10 11:19:41.746 DEBUG megazone23 EBNplus args Obj
## 2019 10 10 11:19:41.746 DEBUG megazone23 EBNplus args par.prior
## 2019 10 10 11:19:41.746 DEBUG megazone23 EBNplus args minSampleNum
## 2019 10 10 11:19:41.746 DEBUG megazone23 EBNplus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2019 10 10 11:19:41.747 DEBUG megazone23 dat
## [1] 7785 576
## 2019 10 10 11:19:41.747 DEBUG megazone23 saminfo
## [1] 576 2
## 2019 10 10 11:19:41.747 DEBUG megazone23 check column and row names
## 2019 10 10 11:19:42.092 DEBUG megazone23 in design.mat_plus
## 2019 10 10 11:19:42.092 DEBUG megazone23 in build.design_plus
## 2019 10 10 11:19:42.093 DEBUG megazone23 after build.design_plus loop
## 2019 10 10 11:19:42.093 DEBUG megazone23 design
## 2019 10 10 11:19:42.093 DEBUG megazone23 list.batch_plus(saminfo)
## 2019 10 10 11:19:42.094 DEBUG megazone23 in list.batch_plus
## 2019 10 10 11:19:42.094 DEBUG megazone23 list.batch_plus tmp1
## 2019 10 10 11:19:42.095 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2019 10 10 11:19:42.095 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2019 10 10 11:19:42.095 DEBUG megazone23 list.batch_plus batches
## 2019 10 10 11:19:42.095 DEBUG megazone23 after list.batch_plus(saminfo)
## 2019 10 10 11:19:42.095 DEBUG megazone23 n.batches
## 2019 10 10 11:19:42.096 DEBUG megazone23 n.array
## 2019 10 10 11:19:44.511 DEBUG megazone23 missbatch matrix
## 2019 10 10 11:19:44.531 DEBUG megazone23 There are 46 genes that were removed because of whole bat
## 2019 10 10 11:19:45.300 DEBUG megazone23 B.hat
## 2019 10 10 11:19:45.301 DEBUG megazone23 grand.mean
## 2019 10 10 11:19:45.563 DEBUG megazone23 var.pooled
## 2019 10 10 11:19:45.592 DEBUG megazone23 stand.mean 1
## 2019 10 10 11:19:45.655 DEBUG megazone23 stand.mean 2
## 2019 10 10 11:19:45.677 DEBUG megazone23 s.data
## 2019 10 10 11:19:46.294 DEBUG megazone23 gamma.hat
## 2019 10 10 11:19:46.640 DEBUG megazone23 plot priors

```

```

## 2019 10 10 11:19:46.641 DEBUG megazone23 plotPrior priorPlotsFile= /bea_testing/output/EBNPlus_Train
## 2019 10 10 11:19:47.098 DEBUG megazone23 parametric adjustments
## 2019 10 10 11:19:47.665 DEBUG megazone23 parametric adjustments
## 2019 10 10 11:19:47.954 DEBUG megazone23 in getBayesData
## 2019 10 10 11:19:47.954 DEBUG megazone23 dim(s.data)=7739 576
## 2019 10 10 11:19:47.955 DEBUG megazone23 dim(stand.mean)=7739 576
## 2019 10 10 11:19:48.077 DEBUG megazone23 dim(bayesdata)=7739 576
## 2019 10 10 11:19:48.156 DEBUG megazone23 Second dim(bayesdata)=7739 576
## 2019 10 10 11:19:48.156 DEBUG megazone23 keep the dim of Original.dat, keep the gene with all NA
## 2019 10 10 11:19:48.196 DEBUG megazone23 resultsDat
## 2019 10 10 11:19:48.196 DEBUG megazone23 after EB MBatch
## 2019 10 10 11:19:48.196 DEBUG megazone23 inside TEST branch MBatch
## 2019 10 10 11:19:48.199 DEBUG megazone23 test
## 2019 10 10 11:19:48.200 DEBUG megazone23 dim(mat1)=7785 123
## 2019 10 10 11:19:48.200 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:19:48.200 DEBUG megazone23 dim(mat2)=7785 123
## 2019 10 10 11:19:48.201 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:19:48.283 DEBUG megazone23 dim(dat)=7785 246
## 2019 10 10 11:19:48.284 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:19:48.293 DEBUG megazone23 after parlist
## 2019 10 10 11:19:48.294 DEBUG megazone23 dim(dat)=7785 246
## 2019 10 10 11:19:48.294 DEBUG megazone23 valid.genes
## 2019 10 10 11:19:48.295 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:19:48.300 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:19:48.303 DEBUG megazone23 after Non-valid
## 2019 10 10 11:19:48.304 DEBUG megazone23 dim(dat)=7739 246
## 2019 10 10 11:19:48.363 DEBUG megazone23 after getStandData
## 2019 10 10 11:19:48.364 DEBUG megazone23 dim(stand.data)=7739 246
## 2019 10 10 11:19:48.364 DEBUG megazone23 batches
## 2019 10 10 11:19:48.364 DEBUG megazone23 after batch.design
## 2019 10 10 11:19:48.365 DEBUG megazone23 dim(batch.design)=246 2
## 2019 10 10 11:19:48.365 DEBUG megazone23 n.batches
## 2019 10 10 11:19:48.365 DEBUG megazone23 in getBayesData
## 2019 10 10 11:19:48.365 DEBUG megazone23 dim(s.data)=7739 246
## 2019 10 10 11:19:48.366 DEBUG megazone23 dim(stand.mean)=7739 246
## 2019 10 10 11:19:48.396 DEBUG megazone23 dim(bayesdata)=7739 246
## 2019 10 10 11:19:48.405 DEBUG megazone23 Second dim(bayesdata)=7739 246
## 2019 10 10 11:19:48.405 DEBUG megazone23 after bayesData
## 2019 10 10 11:19:48.405 DEBUG megazone23 dim(bayesData)=7739 246
## 2019 10 10 11:19:48.406 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:19:48.414 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:19:48.415 DEBUG megazone23 Check the adjusted test set
## 2019 10 10 11:19:48.415 DEBUG megazone23 after train
## 2019 10 10 11:19:48.415 DEBUG megazone23 EBadj
## 2019 10 10 11:19:48.416 DEBUG megazone23 EBadj validation for validation results
## 2019 10 10 11:19:48.489 DEBUG megazone23 validation
## 2019 10 10 11:19:48.490 DEBUG megazone23 dim(mat1)=7785 175
## 2019 10 10 11:19:48.490 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:19:48.490 DEBUG megazone23 dim(mat2)=7785 175
## 2019 10 10 11:19:48.490 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:19:48.597 DEBUG megazone23 dim(dat)=7785 350
## 2019 10 10 11:19:48.598 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:19:48.608 DEBUG megazone23 after parlist

```



```

## 2019 10 10 11:19:48.608 DEBUG megazone23 dim(dat)=7785 350
## 2019 10 10 11:19:48.609 DEBUG megazone23 valid.genes
## 2019 10 10 11:19:48.610 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:19:48.617 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:19:48.617 DEBUG megazone23 after Non-valid
## 2019 10 10 11:19:48.617 DEBUG megazone23 dim(dat)=7739 350
## 2019 10 10 11:19:48.853 DEBUG megazone23 after getStandData
## 2019 10 10 11:19:48.853 DEBUG megazone23 dim(stand.data)=7739 350
## 2019 10 10 11:19:48.854 DEBUG megazone23 batches
## 2019 10 10 11:19:48.854 DEBUG megazone23 after batch.design
## 2019 10 10 11:19:48.854 DEBUG megazone23 dim(batch.design)=350 2
## 2019 10 10 11:19:48.854 DEBUG megazone23 n.batches
## 2019 10 10 11:19:48.855 DEBUG megazone23 in getBayesData
## 2019 10 10 11:19:48.855 DEBUG megazone23 dim(s.data)=7739 350
## 2019 10 10 11:19:48.855 DEBUG megazone23 dim(stand.mean)=7739 350
## 2019 10 10 11:19:48.901 DEBUG megazone23 dim(bayesdata)=7739 350
## 2019 10 10 11:19:48.914 DEBUG megazone23 Second dim(bayesdata)=7739 350
## 2019 10 10 11:19:48.914 DEBUG megazone23 after bayesData
## 2019 10 10 11:19:48.915 DEBUG megazone23 dim(bayesData)=7739 350
## 2019 10 10 11:19:48.915 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:19:48.927 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:19:48.927 DEBUG megazone23 objafterEB@validationEB
## 2019 10 10 11:19:48.928 DEBUG megazone23 EBadj train for training results
## 2019 10 10 11:19:48.928 DEBUG megazone23 train
## 2019 10 10 11:19:48.928 DEBUG megazone23 dim(mat1)=7785 288
## 2019 10 10 11:19:48.928 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:19:48.928 DEBUG megazone23 dim(mat2)=7785 288
## 2019 10 10 11:19:48.929 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:19:49.060 DEBUG megazone23 dim(dat)=7785 576
## 2019 10 10 11:19:49.061 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:19:49.086 DEBUG megazone23 after parlist
## 2019 10 10 11:19:49.087 DEBUG megazone23 dim(dat)=7785 576
## 2019 10 10 11:19:49.087 DEBUG megazone23 valid.genes
## 2019 10 10 11:19:49.088 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:19:49.115 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:19:49.115 DEBUG megazone23 after Non-valid
## 2019 10 10 11:19:49.115 DEBUG megazone23 dim(dat)=7739 576
## 2019 10 10 11:19:49.229 DEBUG megazone23 after getStandData
## 2019 10 10 11:19:49.229 DEBUG megazone23 dim(stand.data)=7739 576
## 2019 10 10 11:19:49.229 DEBUG megazone23 batches
## 2019 10 10 11:19:49.230 DEBUG megazone23 after batch.design
## 2019 10 10 11:19:49.230 DEBUG megazone23 dim(batch.design)=576 2
## 2019 10 10 11:19:49.230 DEBUG megazone23 n.batches
## 2019 10 10 11:19:49.230 DEBUG megazone23 in getBayesData
## 2019 10 10 11:19:49.231 DEBUG megazone23 dim(s.data)=7739 576
## 2019 10 10 11:19:49.231 DEBUG megazone23 dim(stand.mean)=7739 576
## 2019 10 10 11:19:49.427 DEBUG megazone23 dim(bayesdata)=7739 576
## 2019 10 10 11:19:49.504 DEBUG megazone23 Second dim(bayesdata)=7739 576
## 2019 10 10 11:19:49.504 DEBUG megazone23 after bayesData
## 2019 10 10 11:19:49.504 DEBUG megazone23 dim(bayesData)=7739 576
## 2019 10 10 11:19:49.505 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:19:49.555 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:19:49.556 DEBUG megazone23 trainingResults@trainEB
## 2019 10 10 11:19:49.556 DEBUG megazone23 EBadj train for test results

```

```

## 2019 10 10 11:19:49.559 DEBUG megazone23 test
## 2019 10 10 11:19:49.559 DEBUG megazone23 dim(mat1)=7785 123
## 2019 10 10 11:19:49.559 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:19:49.560 DEBUG megazone23 dim(mat2)=7785 123
## 2019 10 10 11:19:49.560 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:19:49.626 DEBUG megazone23 dim(dat)=7785 246
## 2019 10 10 11:19:49.626 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:19:49.634 DEBUG megazone23 after parlist
## 2019 10 10 11:19:49.634 DEBUG megazone23 dim(dat)=7785 246
## 2019 10 10 11:19:49.635 DEBUG megazone23 valid.genes
## 2019 10 10 11:19:49.635 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:19:49.640 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:19:49.641 DEBUG megazone23 after Non-valid
## 2019 10 10 11:19:49.641 DEBUG megazone23 dim(dat)=7739 246
## 2019 10 10 11:19:49.663 DEBUG megazone23 after getStandData
## 2019 10 10 11:19:49.664 DEBUG megazone23 dim(stand.data)=7739 246
## 2019 10 10 11:19:49.664 DEBUG megazone23 batches
## 2019 10 10 11:19:49.664 DEBUG megazone23 after batch.design
## 2019 10 10 11:19:49.664 DEBUG megazone23 dim(batch.design)=246 2
## 2019 10 10 11:19:49.665 DEBUG megazone23 n.batches
## 2019 10 10 11:19:49.665 DEBUG megazone23 in getBayesData
## 2019 10 10 11:19:49.665 DEBUG megazone23 dim(s.data)=7739 246
## 2019 10 10 11:19:49.665 DEBUG megazone23 dim(stand.mean)=7739 246
## 2019 10 10 11:19:49.705 DEBUG megazone23 dim(bayesdata)=7739 246
## 2019 10 10 11:19:49.722 DEBUG megazone23 Second dim(bayesdata)=7739 246
## 2019 10 10 11:19:49.723 DEBUG megazone23 after bayesData
## 2019 10 10 11:19:49.723 DEBUG megazone23 dim(bayesData)=7739 246
## 2019 10 10 11:19:49.724 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:19:49.742 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:19:49.742 DEBUG megazone23 testResults@testEB
## 2019 10 10 11:19:49.811 DEBUG megazone23 finishing BeaEBNPlus
## [1] "TestSet1"
## [1] TRUE
## [1] 7785 123
##          TCGA-A1-AOSE-01A-11R-A084-07 TCGA-A1-AOSP-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"
## [1] TRUE
## [1] 7785 123
##          TCGA-A1-AOSE-01A-11R-A084-07 TCGA-A1-AOSP-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-01A-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-AOSH-01A-11R-A084-07 TCGA-A1-AOSJ-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-AOSK-01A-12R-A084-07 TCGA-A1-AOSM-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-AOSH-01A-11R-A084-07 TCGA-A1-AOSJ-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-AOSK-01A-12R-A084-07 TCGA-A1-AOSM-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-AOSH-01A-11R-A084-07.Agilent4502
## A1BG              4.2922928
## A2BP1             -0.3358343
## A2M               7.5542896
## A2ML1             1.4633094
##          TCGA-A1-AOSH-01A-11R-A084-07.RNASeqV2
## A1BG              4.02642841
## A2BP1             -1.30645343
## A2M               7.37907814
## A2ML1             0.06540906
##          TCGA-A1-AOSJ-01A-11R-A084-07.Agilent4502
## A1BG              4.2939693
## A2BP1             -0.4097362
## A2M               8.0769652
## A2ML1             1.9279318
##          TCGA-A1-AOSJ-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-AOSD-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-AOSD-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-AOSD-01A-11R-A115-07.Agilent4502
## A1BG                                4.544614
## A2BP1                              2.235393
## A2M                                6.626155
## A2ML1                              1.653776
##      TCGA-A1-AOSD-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

```

7 Example Output

A list of matrices with the following names().

7.1 TestSet1

a matrix of the test set from data set 1

7.2 TestSet2

a matrix of the test set from data set 2

7.3 TrainingSet1

a matrix of the training set from data set 1

7.4 TrainingSet2

a matrix of the training set from data set 2

7.5 TrainingResults

a matrix of the corrected training data

7.6 ValidationSet1

a matrix of the validation set from data set 1

7.7 ValidationSet2

a matrix of the validation set from data set 2

7.8 ValidationResults

a matrix of the corrected validation data

7.9 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.

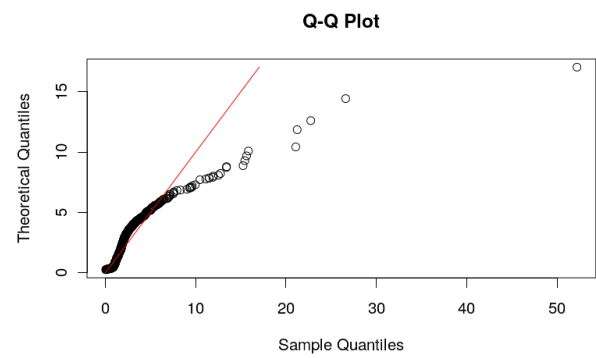
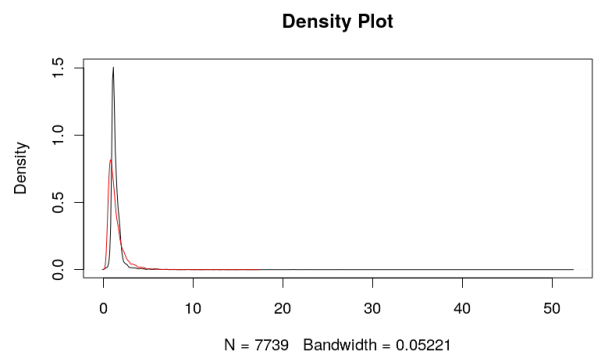
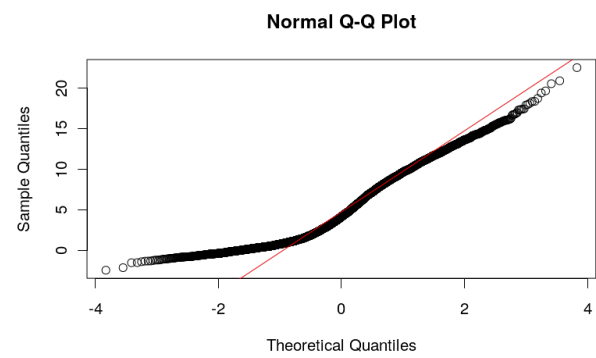
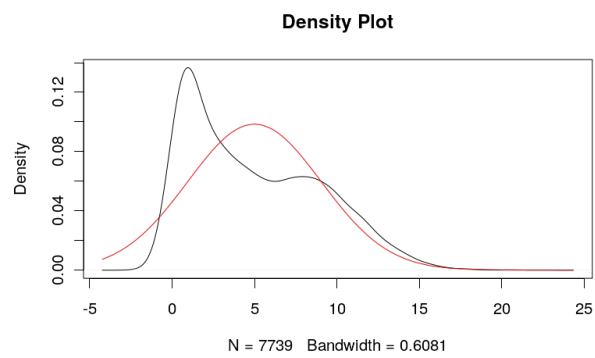


Figure 1: Prior Plots File