

Using MBatch Corrections: EBNPlus__TrainAndValidateReplicates__Structures

Tod Casasent

2018-06-21

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"
## 2018 06 21 10:37:35.135 DEBUG megazone23 starting BeaEBNplus
## 2018 06 21 10:37:35.136 DEBUG megazone23 MBatch Version: 2017-09-19-1530
## 2018 06 21 10:37:35.136 DEBUG megazone23 BeaEBNplus theEBNP_AddData1Rows= FALSE
## 2018 06 21 10:37:35.136 DEBUG megazone23 BeaEBNplus theEBNP_AddData2Rows= FALSE

```

```

## 2018 06 21 10:37:35.137 DEBUG megazone23 EBNPlus theEBNP_AddData1Rows= FALSE
## 2018 06 21 10:37:35.137 DEBUG megazone23 EBNPlus theEBNP_AddData2Rows= FALSE
## 2018 06 21 10:37:35.137 DEBUG megazone23 EBNPlus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2018 06 21 10:37:35.137 DEBUG megazone23 dim(theData1)=20530 1215
## 2018 06 21 10:37:35.138 DEBUG megazone23 dim(theData2)=17814 600
## 2018 06 21 10:37:35.138 DEBUG megazone23 remove unknown genes, that start with ?
## 2018 06 21 10:37:36.222 DEBUG megazone23 check on adding missing rows
## 2018 06 21 10:37:36.223 DEBUG megazone23 dim(theData1)=20501 1215
## 2018 06 21 10:37:36.223 DEBUG megazone23 dim(theData2)=17814 600
## 2018 06 21 10:37:36.223 DEBUG megazone23 make EBNplus
## 2018 06 21 10:37:36.224 DEBUG megazone23 before callNextMethod
## 2018 06 21 10:37:36.224 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:37:36.224 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:37:36.224 DEBUG megazone23 removing column duplicates
## 2018 06 21 10:37:36.225 DEBUG megazone23 removeDuplicatesFromColumns
## 2018 06 21 10:37:36.225 DEBUG megazone23 removeDuplicatesFromColumns
## 2018 06 21 10:37:36.225 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:37:36.226 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:37:36.226 DEBUG megazone23 removing row duplicates
## 2018 06 21 10:37:36.226 DEBUG megazone23 removeDuplicatesFromRows
## 2018 06 21 10:37:36.227 DEBUG megazone23 removeDuplicatesFromRows
## 2018 06 21 10:37:36.228 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:37:36.228 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:37:36.229 DEBUG megazone23 after callNextMethod
## 2018 06 21 10:37:36.229 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:37:36.229 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:37:36.229 DEBUG megazone23 after EBNplus
## 2018 06 21 10:37:36.229 DEBUG megazone23 dim(ebObj@mData1)=20501 1215
## 2018 06 21 10:37:36.230 DEBUG megazone23 dim(ebObj@mData2)=17814 600
## 2018 06 21 10:37:36.234 DEBUG megazone23 getBiComOrder
## 2018 06 21 10:37:36.234 DEBUG megazone23 makeCommonRows
## 2018 06 21 10:37:37.164 DEBUG megazone23 dim(mData1)=16146 1215
## 2018 06 21 10:37:37.164 DEBUG megazone23 dim(mData2)=16146 600
## 2018 06 21 10:37:37.165 DEBUG megazone23 makeCommonCols
## 2018 06 21 10:37:37.381 DEBUG megazone23 dim(mData1)=16146 586
## 2018 06 21 10:37:37.381 DEBUG megazone23 dim(mData2)=16146 586
## 2018 06 21 10:37:37.382 DEBUG megazone23 asSameOrder
## 2018 06 21 10:37:37.382 DEBUG megazone23 asSameOrder before
## 2018 06 21 10:37:37.382 DEBUG megazone23 asSameOrder after if 1
## 2018 06 21 10:37:37.477 DEBUG megazone23 asSameOrder after if 2
## 2018 06 21 10:37:37.477 DEBUG megazone23 row.names(mat1)
## 2018 06 21 10:37:37.478 DEBUG megazone23 rownames(mat1)
## 2018 06 21 10:37:37.478 DEBUG megazone23 row.names(mat2)
## 2018 06 21 10:37:37.478 DEBUG megazone23 rownames(mat2)
## 2018 06 21 10:37:37.479 DEBUG megazone23 m.i
## 2018 06 21 10:37:37.580 DEBUG megazone23 asSameOrder after if 3
## 2018 06 21 10:37:37.581 DEBUG megazone23 dim(ebObj@mat1Com)=16146 586
## 2018 06 21 10:37:37.581 DEBUG megazone23 dim(ebObj@mat2Com)=16146 586
## 2018 06 21 10:37:37.581 DEBUG megazone23 getValidationSet
## 2018 06 21 10:37:37.581 INFO megazone23 getValidationSet seed= 314
## 2018 06 21 10:37:37.582 DEBUG megazone23 validationRatio=0.3
## 2018 06 21 10:37:37.582 DEBUG megazone23 numSamples=586
## 2018 06 21 10:37:37.726 DEBUG megazone23 dim(ebObj@mat1Validation)=16146 175
## 2018 06 21 10:37:37.726 DEBUG megazone23 dim(ebObj@mat2Validation)=16146 175

```

```

## 2018 06 21 10:37:37.727 DEBUG megazone23 dim(ebObj@mat1Train)=16146 411
## 2018 06 21 10:37:37.727 DEBUG megazone23 dim(ebObj@mat2Train)=16146 411
## 2018 06 21 10:37:37.727 DEBUG megazone23 before Test Set
## 2018 06 21 10:37:37.971 DEBUG megazone23 after Test Set
## 2018 06 21 10:37:38.187 DEBUG megazone23 train
## 2018 06 21 10:37:38.188 DEBUG megazone23 inside train
## 2018 06 21 10:37:38.188 DEBUG megazone23 train theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_T
## 2018 06 21 10:37:38.211 DEBUG megazone23 colnames(Object@mat1Train)=TCGA-A1-AOSH-01A-11R-A084-07.RNA
## 2018 06 21 10:37:38.212 DEBUG megazone23 colnames(Object@mat2Train)=TCGA-A1-AOSH-01A-11R-A084-07.Agi
## 2018 06 21 10:37:38.212 DEBUG megazone23 inside train, call getData4EB
## 2018 06 21 10:37:38.579 DEBUG megazone23 data4EB <- as.matrix
## 2018 06 21 10:37:38.580 DEBUG megazone23 dim(data4EB)=16146 576
## 2018 06 21 10:37:38.580 DEBUG megazone23 cbinds and rbinds
## 2018 06 21 10:37:38.580 DEBUG megazone23 Object@DF1batch= RNASeqV2
## 2018 06 21 10:37:38.580 DEBUG megazone23 Object@DF2batch= Agilent4502
## 2018 06 21 10:37:38.582 DEBUG megazone23 row.names(sampBatch)
## 2018 06 21 10:37:38.583 DEBUG megazone23 TCGA-A1-AOSH-01A-11R-A084-07.RNASeqV2 TCGA-A1-AOSJ-01A-11R-
## 2018 06 21 10:37:38.583 DEBUG megazone23 colnames(sampBatch)
## 2018 06 21 10:37:38.583 DEBUG megazone23 sample batch
## 2018 06 21 10:37:38.584 DEBUG megazone23 inside train, after getData4EB
## 2018 06 21 10:37:38.584 DEBUG megazone23 call EB MBatch
## 2018 06 21 10:37:38.584 DEBUG megazone23 TDC HERE trainEB mbatch
## 2018 06 21 10:37:38.584 DEBUG megazone23 EBNplus args Obj
## 2018 06 21 10:37:38.585 DEBUG megazone23 EBNplus args par.prior
## 2018 06 21 10:37:38.585 DEBUG megazone23 EBNplus args minSampleNum
## 2018 06 21 10:37:38.585 DEBUG megazone23 EBNplus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2018 06 21 10:37:38.585 DEBUG megazone23 dat
## [1] 16146 576
## 2018 06 21 10:37:38.586 DEBUG megazone23 saminfo
## [1] 576 2
## 2018 06 21 10:37:38.587 DEBUG megazone23 check column and row names
## 2018 06 21 10:37:38.747 DEBUG megazone23 in design.mat_plus
## 2018 06 21 10:37:38.748 DEBUG megazone23 in build.design_plus
## 2018 06 21 10:37:38.748 DEBUG megazone23 after build.design_plus loop
## 2018 06 21 10:37:38.749 DEBUG megazone23 design
## 2018 06 21 10:37:38.749 DEBUG megazone23 list.batch_plus(saminfo)
## 2018 06 21 10:37:38.749 DEBUG megazone23 in list.batch_plus
## 2018 06 21 10:37:38.750 DEBUG megazone23 list.batch_plus tmp1
## 2018 06 21 10:37:38.750 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2018 06 21 10:37:38.750 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2018 06 21 10:37:38.751 DEBUG megazone23 list.batch_plus batches
## 2018 06 21 10:37:38.751 DEBUG megazone23 after list.batch_plus(saminfo)
## 2018 06 21 10:37:38.751 DEBUG megazone23 n.batches
## 2018 06 21 10:37:38.751 DEBUG megazone23 n.array
## 2018 06 21 10:37:44.079 DEBUG megazone23 missbatch matrix
## 2018 06 21 10:37:44.178 DEBUG megazone23 There are 144 genes that were removed because of whole ba
## 2018 06 21 10:37:45.579 DEBUG megazone23 B.hat
## 2018 06 21 10:37:45.579 DEBUG megazone23 grand.mean
## 2018 06 21 10:37:46.425 DEBUG megazone23 var.pooled
## 2018 06 21 10:37:46.440 DEBUG megazone23 stand.mean 1
## 2018 06 21 10:37:46.610 DEBUG megazone23 stand.mean 2
## 2018 06 21 10:37:46.667 DEBUG megazone23 s.data
## 2018 06 21 10:37:48.132 DEBUG megazone23 gamma.hat
## 2018 06 21 10:37:49.583 DEBUG megazone23 plot priors

```

```

## 2018 06 21 10:37:49.584 DEBUG megazone23 plotPrior priorPlotsFile= /bea_testing/output/EBNPlus_Train
## 2018 06 21 10:37:50.452 DEBUG megazone23 parametric adjustments
## 2018 06 21 10:37:51.679 DEBUG megazone23 parametric adjustments
## 2018 06 21 10:37:52.676 DEBUG megazone23 in getBayesData
## 2018 06 21 10:37:52.676 DEBUG megazone23 dim(s.data)=16002 576
## 2018 06 21 10:37:52.677 DEBUG megazone23 dim(stand.mean)=16002 576
## 2018 06 21 10:37:52.920 DEBUG megazone23 dim(bayesdata)=16002 576
## 2018 06 21 10:37:52.965 DEBUG megazone23 Second dim(bayesdata)=16002 576
## 2018 06 21 10:37:52.966 DEBUG megazone23 keep the dim of Original.dat, keep the gene with all NA
## 2018 06 21 10:37:52.998 DEBUG megazone23 resultsDat
## 2018 06 21 10:37:52.999 DEBUG megazone23 after EB MBatch
## 2018 06 21 10:37:52.999 DEBUG megazone23 inside TEST branch MBatch
## 2018 06 21 10:37:53.016 DEBUG megazone23 test
## 2018 06 21 10:37:53.016 DEBUG megazone23 dim(mat1)=16146 123
## 2018 06 21 10:37:53.017 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:37:53.017 DEBUG megazone23 dim(mat2)=16146 123
## 2018 06 21 10:37:53.017 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:37:53.203 DEBUG megazone23 dim(dat)=16146 246
## 2018 06 21 10:37:53.203 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:37:53.237 DEBUG megazone23 after parlist
## 2018 06 21 10:37:53.238 DEBUG megazone23 dim(dat)=16146 246
## 2018 06 21 10:37:53.238 DEBUG megazone23 valid.genes
## 2018 06 21 10:37:53.240 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:37:53.274 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:37:53.275 DEBUG megazone23 after Non-valid
## 2018 06 21 10:37:53.275 DEBUG megazone23 dim(dat)=16002 246
## 2018 06 21 10:37:53.327 DEBUG megazone23 after getStandData
## 2018 06 21 10:37:53.327 DEBUG megazone23 dim(stand.data)=16002 246
## 2018 06 21 10:37:53.327 DEBUG megazone23 batches
## 2018 06 21 10:37:53.328 DEBUG megazone23 after batch.design
## 2018 06 21 10:37:53.328 DEBUG megazone23 dim(batch.design)=246 2
## 2018 06 21 10:37:53.328 DEBUG megazone23 n.batches
## 2018 06 21 10:37:53.329 DEBUG megazone23 in getBayesData
## 2018 06 21 10:37:53.329 DEBUG megazone23 dim(s.data)=16002 246
## 2018 06 21 10:37:53.329 DEBUG megazone23 dim(stand.mean)=16002 246
## 2018 06 21 10:37:53.429 DEBUG megazone23 dim(bayesdata)=16002 246
## 2018 06 21 10:37:53.457 DEBUG megazone23 Second dim(bayesdata)=16002 246
## 2018 06 21 10:37:53.458 DEBUG megazone23 after bayesData
## 2018 06 21 10:37:53.458 DEBUG megazone23 dim(bayesData)=16002 246
## 2018 06 21 10:37:53.458 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:37:53.490 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:37:53.490 DEBUG megazone23 Check the adjusted test set
## 2018 06 21 10:37:53.491 DEBUG megazone23 after train
## 2018 06 21 10:37:53.491 DEBUG megazone23 EBadj
## 2018 06 21 10:37:53.491 DEBUG megazone23 EBadj validation for validation results
## 2018 06 21 10:37:53.492 DEBUG megazone23 validation
## 2018 06 21 10:37:53.508 DEBUG megazone23 dim(mat1)=16146 175
## 2018 06 21 10:37:53.509 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:37:53.509 DEBUG megazone23 dim(mat2)=16146 175
## 2018 06 21 10:37:53.509 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:37:53.731 DEBUG megazone23 dim(dat)=16146 350
## 2018 06 21 10:37:53.731 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:37:53.778 DEBUG megazone23 after parlist

```



```

## 2018 06 21 10:37:53.778 DEBUG megazone23 dim(dat)=16146 350
## 2018 06 21 10:37:53.779 DEBUG megazone23 valid.genes
## 2018 06 21 10:37:53.780 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:37:53.827 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:37:53.828 DEBUG megazone23 after Non-valid
## 2018 06 21 10:37:53.828 DEBUG megazone23 dim(dat)=16002 350
## 2018 06 21 10:37:53.909 DEBUG megazone23 after getStandData
## 2018 06 21 10:37:53.909 DEBUG megazone23 dim(stand.data)=16002 350
## 2018 06 21 10:37:53.910 DEBUG megazone23 batches
## 2018 06 21 10:37:53.910 DEBUG megazone23 after batch.design
## 2018 06 21 10:37:53.910 DEBUG megazone23 dim(batch.design)=350 2
## 2018 06 21 10:37:53.911 DEBUG megazone23 n.batches
## 2018 06 21 10:37:53.911 DEBUG megazone23 in getBayesData
## 2018 06 21 10:37:53.911 DEBUG megazone23 dim(s.data)=16002 350
## 2018 06 21 10:37:53.911 DEBUG megazone23 dim(stand.mean)=16002 350
## 2018 06 21 10:37:54.058 DEBUG megazone23 dim(bayesdata)=16002 350
## 2018 06 21 10:37:54.082 DEBUG megazone23 Second dim(bayesdata)=16002 350
## 2018 06 21 10:37:54.082 DEBUG megazone23 after bayesData
## 2018 06 21 10:37:54.082 DEBUG megazone23 dim(bayesData)=16002 350
## 2018 06 21 10:37:54.083 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:37:54.110 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:37:54.110 DEBUG megazone23 objafterEB@validationEB
## 2018 06 21 10:37:54.111 DEBUG megazone23 EBadj train for training results
## 2018 06 21 10:37:54.111 DEBUG megazone23 train
## 2018 06 21 10:37:54.111 DEBUG megazone23 dim(mat1)=16146 288
## 2018 06 21 10:37:54.111 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:37:54.112 DEBUG megazone23 dim(mat2)=16146 288
## 2018 06 21 10:37:54.112 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:37:54.612 DEBUG megazone23 dim(dat)=16146 576
## 2018 06 21 10:37:54.613 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:37:54.711 DEBUG megazone23 after parlist
## 2018 06 21 10:37:54.711 DEBUG megazone23 dim(dat)=16146 576
## 2018 06 21 10:37:54.712 DEBUG megazone23 valid.genes
## 2018 06 21 10:37:54.713 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:37:54.814 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:37:54.814 DEBUG megazone23 after Non-valid
## 2018 06 21 10:37:54.814 DEBUG megazone23 dim(dat)=16002 576
## 2018 06 21 10:37:54.972 DEBUG megazone23 after getStandData
## 2018 06 21 10:37:54.972 DEBUG megazone23 dim(stand.data)=16002 576
## 2018 06 21 10:37:54.973 DEBUG megazone23 batches
## 2018 06 21 10:37:54.973 DEBUG megazone23 after batch.design
## 2018 06 21 10:37:54.973 DEBUG megazone23 dim(batch.design)=576 2
## 2018 06 21 10:37:54.973 DEBUG megazone23 n.batches
## 2018 06 21 10:37:54.974 DEBUG megazone23 in getBayesData
## 2018 06 21 10:37:54.974 DEBUG megazone23 dim(s.data)=16002 576
## 2018 06 21 10:37:54.974 DEBUG megazone23 dim(stand.mean)=16002 576
## 2018 06 21 10:37:55.213 DEBUG megazone23 dim(bayesdata)=16002 576
## 2018 06 21 10:37:55.255 DEBUG megazone23 Second dim(bayesdata)=16002 576
## 2018 06 21 10:37:55.256 DEBUG megazone23 after bayesData
## 2018 06 21 10:37:55.256 DEBUG megazone23 dim(bayesData)=16002 576
## 2018 06 21 10:37:55.256 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:37:55.389 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:37:55.390 DEBUG megazone23 trainingResults@trainEB
## 2018 06 21 10:37:55.391 DEBUG megazone23 EBadj train for test results

```

```

## 2018 06 21 10:37:55.391 DEBUG megazone23 test
## 2018 06 21 10:37:55.391 DEBUG megazone23 dim(mat1)=16146 123
## 2018 06 21 10:37:55.391 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:37:55.392 DEBUG megazone23 dim(mat2)=16146 123
## 2018 06 21 10:37:55.392 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:37:55.573 DEBUG megazone23 dim(dat)=16146 246
## 2018 06 21 10:37:55.573 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:37:55.608 DEBUG megazone23 after parlist
## 2018 06 21 10:37:55.608 DEBUG megazone23 dim(dat)=16146 246
## 2018 06 21 10:37:55.609 DEBUG megazone23 valid.genes
## 2018 06 21 10:37:55.610 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:37:55.644 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:37:55.645 DEBUG megazone23 after Non-valid
## 2018 06 21 10:37:55.645 DEBUG megazone23 dim(dat)=16002 246
## 2018 06 21 10:37:55.694 DEBUG megazone23 after getStandData
## 2018 06 21 10:37:55.695 DEBUG megazone23 dim(stand.data)=16002 246
## 2018 06 21 10:37:55.695 DEBUG megazone23 batches
## 2018 06 21 10:37:55.695 DEBUG megazone23 after batch.design
## 2018 06 21 10:37:55.695 DEBUG megazone23 dim(batch.design)=246 2
## 2018 06 21 10:37:55.696 DEBUG megazone23 n.batches
## 2018 06 21 10:37:55.696 DEBUG megazone23 in getBayesData
## 2018 06 21 10:37:55.696 DEBUG megazone23 dim(s.data)=16002 246
## 2018 06 21 10:37:55.697 DEBUG megazone23 dim(stand.mean)=16002 246
## 2018 06 21 10:37:55.791 DEBUG megazone23 dim(bayesdata)=16002 246
## 2018 06 21 10:37:55.814 DEBUG megazone23 Second dim(bayesdata)=16002 246
## 2018 06 21 10:37:55.814 DEBUG megazone23 after bayesData
## 2018 06 21 10:37:55.815 DEBUG megazone23 dim(bayesData)=16002 246
## 2018 06 21 10:37:55.815 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:37:55.879 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:37:55.880 DEBUG megazone23 testResults@testEB
## 2018 06 21 10:37:56.133 DEBUG megazone23 finishing BeaEBNPlus
## [1] "TestSet1"
## [1] TRUE
## [1] 16146 123
## TCGA-A1-AOSE-01A-11R-A084-07 TCGA-A1-AOSO-01A-22R-A084-07
## A1BG 7.592955 8.196403
## A2BP1 -2.016249 -2.016249
## A2M 13.941115 11.999331
## A2ML1 0.303810 4.520856
## TCGA-A1-AOSP-01A-11R-A084-07 TCGA-A2-A04Q-01A-21R-A034-07
## A1BG 5.210315 6.838973
## A2BP1 -2.016249 -2.016249
## A2M 12.357505 13.641990
## A2ML1 9.738008 11.163480
## [1] "TestSet2"
## [1] TRUE
## [1] 16146 123
## TCGA-A1-AOSE-01A-11R-A084-07 TCGA-A1-AOSO-01A-22R-A084-07
## A1BG 0.2798333 1.542333
## A2BP1 -1.5346667 -1.574833
## A2M 1.0385000 -0.829500
## A2ML1 0.3715000 0.473500
## TCGA-A1-AOSP-01A-11R-A084-07 TCGA-A2-A04Q-01A-21R-A034-07
## A1BG -1.291167 0.2260

```

```

## A2BP1 -1.133833 -1.1005
## A2M -0.549500 0.9565
## A2ML1 1.332000 2.3620
## [1] "TrainingSet1"
## [1] TRUE
## [1] 16146 288
## TCGA-A1-AOSH-01A-11R-A084-07 TCGA-A1-AOSJ-01A-11R-A084-07
## A1BG 7.4964200 7.82672198
## A2BP1 -0.7169079 -2.01624935
## A2M 13.7729442 14.43268308
## A2ML1 0.7585590 -0.03091399
## TCGA-A2-A04N-01A-11R-A115-07 TCGA-A2-A04R-01A-41R-A109-07
## A1BG 7.5215039 7.7133865
## A2BP1 -2.0162494 -2.0162494
## A2M 13.2630050 12.8595157
## A2ML1 -0.6362122 -0.7145386
## [1] "TrainingSet2"
## [1] TRUE
## [1] 16146 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-A2-A04N-01A-11R-A115-07 TCGA-A2-A04R-01A-41R-A109-07
## A1BG 0.0856667 1.2780
## A2BP1 -1.2888333 -1.5805
## A2M 0.3242500 -0.1500
## A2ML1 0.5495000 0.6305
## [1] "TrainingResults"
## [1] TRUE
## [1] 16146 576
## TCGA-A1-AOSH-01A-11R-A084-07.Agilent4502
## A1BG 4.3272798
## A2BP1 -0.2723037
## A2M 7.5700011
## A2ML1 1.3918137
## TCGA-A1-AOSH-01A-11R-A084-07.RNASeqV2
## A1BG 4.01423955
## A2BP1 -1.27715076
## A2M 7.36665554
## A2ML1 0.09205367
## TCGA-A1-AOSJ-01A-11R-A084-07.Agilent4502
## A1BG 4.3290105
## A2BP1 -0.3523965
## A2M 8.0929148
## A2ML1 1.8502446
## TCGA-A1-AOSJ-01A-11R-A084-07.RNASeqV2
## A1BG 4.2921431
## A2BP1 -2.1452478
## A2M 8.0226132
## A2ML1 -0.4700872
## [1] "ValidationSet1"
## [1] TRUE

```

```

## [1] 16146 175
## TCGA-A1-AOSD-01A-11R-A115-07 TCGA-A1-AOSK-01A-12R-A084-07
## A1BG 7.1552716 6.441132
## A2BP1 -0.7908586 -2.016249
## A2M 14.5382944 12.032942
## A2ML1 2.9738317 4.652188
## TCGA-A1-AOSM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG 8.2655724 6.768147
## A2BP1 -0.8290494 -2.016249
## A2M 13.1294046 13.182368
## A2ML1 -0.8290494 10.684743
## [1] "ValidationSet2"
## [1] TRUE
## [1] 16146 175
## TCGA-A1-AOSD-01A-11R-A115-07 TCGA-A1-AOSK-01A-12R-A084-07
## A1BG 0.9493333 -0.5471667
## A2BP1 0.5391667 -2.1845000
## A2M 0.2420000 -0.8597500
## A2ML1 0.4235000 0.3060000
## TCGA-A1-AOSM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG 0.5976667 -0.1568333
## A2BP1 -1.1756667 -0.8194000
## A2M 0.5745000 0.2440000
## A2ML1 0.2635000 1.9145000
## [1] "ValidationResults"
## [1] TRUE
## [1] 16146 350
## TCGA-A1-AOSD-01A-11R-A115-07.Agilent4502
## A1BG 4.587749
## A2BP1 2.514321
## A2M 6.641444
## A2ML1 1.579742
## TCGA-A1-AOSD-01A-11R-A115-07.RNASeqV2
## A1BG 3.727210
## A2BP1 -1.326558
## A2M 8.127619
## A2ML1 1.669429
## TCGA-A1-AOSK-01A-12R-A084-07.Agilent4502
## A1BG 2.645263
## A2BP1 -1.601643
## A2M 5.527630
## A2ML1 1.245173
## TCGA-A1-AOSK-01A-12R-A084-07.RNASeqV2
## A1BG 3.126360
## A2BP1 -2.145248
## A2M 5.636626
## A2ML1 2.864496
## [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.

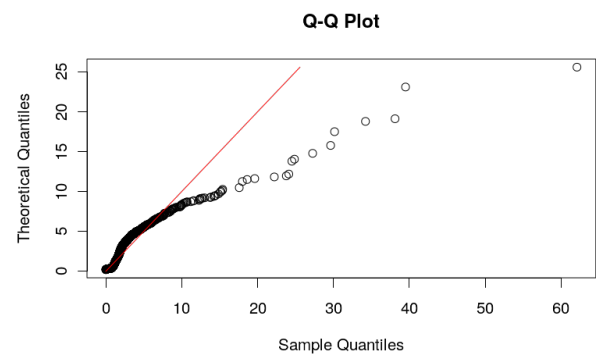
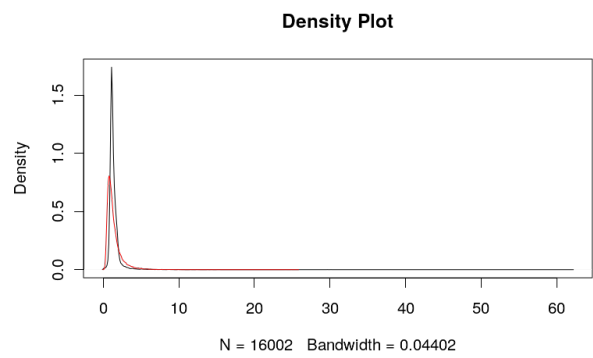
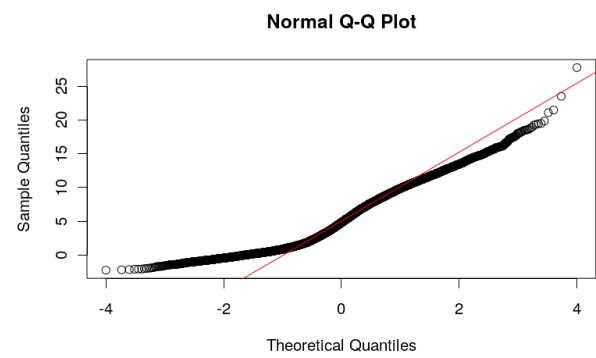
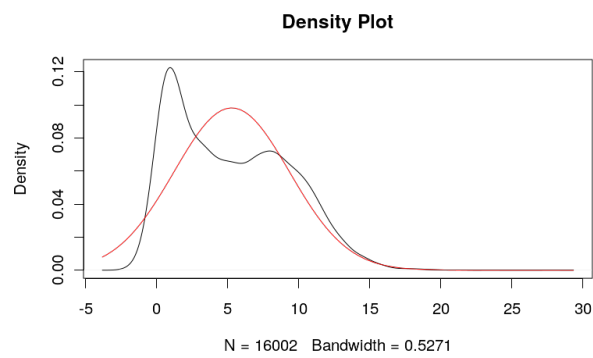


Figure 1: Prior Plots File