Using MBatch Corrections: EBNPlus_TrainAndValidateFromVector_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_TrainAndValidateFromVector_Structures creates and returns a training and validation set and corrected data, based on the training and validation set and provided vectors of samples to use for training and validation, instead of using replicates and percentages.

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_TrainAndValidateFromVector_Structures(theDataMatrix1, theDataMatrix2, theBatchId1, theBatchId2, theEBNP_PsuedoReplicates1Train, theEBNP_PsuedoReplicates2Train, theEBNP_PsuedoReplicates1Validation, theEBNP_PsuedoReplicates2Validation, theEBNP_BatchWithZero, theEBNP_FixDataSet, theEBNP_CorrectForZero, theEBNP_ParametricPriorsFlag, theEBNP_TestRatio=0, theSeed=NULL, theTestSeed=NULL, theEBNP_PriorPlotsFile = NULL, theEBNP_MinSampleNum = 3)

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_PsuedoReplicates1Train

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

5.6 theEBNP_PsuedoReplicates2Train

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

5.7 theEBNP_PsuedoReplicates1Validation

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

5.8 theEBNP_PsuedoReplicates2Validation

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

5.9 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.

5.10 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.

5.11 theEBNP_CorrectForZero

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

5.12 theEBNP_ParametricPriorsFlag

Use parametric adjustments for corrections (TRUE or FALSE)

5.13   theSeed

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

5.14   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.

5.15   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.

5.16   theEBNP_MinSampleNum

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

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])
```

```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)
  #
```

```r
unlink(theOutputDir, recursive=TRUE)
dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)
# read the files in. This can be done however you want
print("read the files")
theDataMatrix1 <- readAsGenericMatrix(theDataFile1)
theDataMatrix2 <- readAsGenericMatrix(theDataFile2)
# this is the reduce genes to just gene symbols, handling those from standardized data
print("reduce to gene symbols")
rownames(theDataMatrix1) <- trimGenes(rownames(theDataMatrix1))
rownames(theDataMatrix2) <- trimGenes(rownames(theDataMatrix2))
# remove any duplicates (this is a requirement for EBNplus)
print("remove duplicates")
theDataMatrix1 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix1))
theDataMatrix2 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix2))
# pseudoreplicate vectors
  theEBNP_PsuedoReplicates1Train <- c("TCGA-A1-A0SH-01A-11R-A084-07", "TCGA-A1-A0SJ-01A-11R-A084-07"
theEBNP_PsuedoReplicates2Train <- c("TCGA-A1-A0SH-01A-11R-A084-07", "TCGA-A1-A0SJ-01A-11R-A084-07",
theEBNP_PsuedoReplicates1Validation <- c("TCGA-A1-A0SD-01A-11R-A115-07", "TCGA-A1-A0SK-01A-12R-A084-
theEBNP_PsuedoReplicates2Validation <- c("TCGA-A1-A0SD-01A-11R-A115-07", "TCGA-A1-A0SK-01A-12R-A084-
# call function
print("EBNPlus_TrainAndValidateReplicates_Structures")
resultsList <- EBNPlus_TrainAndValidateFromVector_Structures(theDataMatrix1, theDataMatrix2,
      theBatchId1, theBatchId2,
      theEBNP_PsuedoReplicates1Train,
      theEBNP_PsuedoReplicates2Train,
      theEBNP_PsuedoReplicates1Validation,
      theEBNP_PsuedoReplicates2Validation,
      theEBNP_BatchWithZero="1",
      theEBNP_FixDataSet=as.numeric(NA),
      theEBNP_CorrectForZero=TRUE,
      theEBNP_ParametricPriorsFlag=TRUE,
      theEBNP_TestRatio=0.3,
      theSeed=theRandomSeed,
      theTestSeed=theRandomSeed,
      theEBNP_PriorPlotsFile=file.path(theOutputDir, "priorplots.PNG"))
print("TestSet1")
printMatrix(resultsList$TestSet1)
print("TestSet2")
printMatrix(resultsList$TestSet2)
print("TrainingSet1")
printMatrix(resultsList$TrainingSet1)
print("TrainingSet2")
printMatrix(resultsList$TrainingSet2)
print("TrainingResults")
printMatrix(resultsList$TrainingResults)
print("ValidationSet1")
printMatrix(resultsList$ValidationSet1)
print("ValidationSet2")
printMatrix(resultsList$ValidationSet2)
print("ValidationResults")
printMatrix(resultsList$ValidationResults)
print("CorrectedResults")
printMatrix(resultsList$CorrectedResults)
```

```
  }
}
```

```
## [1] "read the files"
## [1] "reduce to gene symbols"
## [1] "remove duplicates"
## [1] "EBNPlus_TrainAndValidateReplicates_Structures"
## 2018 06 21 10:38:43.833 DEBUG megazone23 EBNPlus_TrainAndValidateFromVector_Structures - start
## 2018 06 21 10:38:43.833 DEBUG megazone23 removeDuplicatesFromRows
## 2018 06 21 10:38:43.835 DEBUG megazone23 removeDuplicatesFromRows
## 2018 06 21 10:38:43.835 DEBUG megazone23 removeDuplicatesFromColumns
## 2018 06 21 10:38:43.836 DEBUG megazone23 removeDuplicatesFromColumns
## 2018 06 21 10:38:43.836 DEBUG megazone23 EBNPlus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_
## 2018 06 21 10:38:43.836 DEBUG megazone23 dim(theDataMatrix1)=20530 1215
## 2018 06 21 10:38:43.836 DEBUG megazone23 dim(theDataMatrix2)=17814 600
## 2018 06 21 10:38:43.837 DEBUG megazone23 remove unknown genes, that start with ?
## 2018 06 21 10:38:44.966 DEBUG megazone23 dim(theDataMatrix1)=20501 1215
## 2018 06 21 10:38:44.966 DEBUG megazone23 dim(theDataMatrix2)=17814 600
## 2018 06 21 10:38:44.967 DEBUG megazone23 make EBNplus
## 2018 06 21 10:38:44.967 DEBUG megazone23 before callNextMethod
## 2018 06 21 10:38:44.967 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:38:44.968 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:38:44.968 DEBUG megazone23 after callNextMethod
## 2018 06 21 10:38:44.968 DEBUG megazone23 dim(mData1)=20501 1215
## 2018 06 21 10:38:44.969 DEBUG megazone23 dim(mData2)=17814 600
## 2018 06 21 10:38:44.969 DEBUG megazone23 after EBNplus
## 2018 06 21 10:38:44.969 DEBUG megazone23 dim(ebObj@mData1)=20501 1215
## 2018 06 21 10:38:44.969 DEBUG megazone23 dim(ebObj@mData2)=17814 600
## 2018 06 21 10:38:44.973 DEBUG megazone23 getBiComOrder
## 2018 06 21 10:38:44.974 DEBUG megazone23 makeCommonRows
## 2018 06 21 10:38:45.935 DEBUG megazone23 dim(mData1)=16146 1215
## 2018 06 21 10:38:45.935 DEBUG megazone23 dim(mData2)=16146 600
## 2018 06 21 10:38:45.935 DEBUG megazone23 makeCommonCols
## 2018 06 21 10:38:46.140 DEBUG megazone23 dim(mData1)=16146 586
## 2018 06 21 10:38:46.140 DEBUG megazone23 dim(mData2)=16146 586
## 2018 06 21 10:38:46.140 DEBUG megazone23 asSameOrder
## 2018 06 21 10:38:46.141 DEBUG megazone23 asSameOrder before
## 2018 06 21 10:38:46.141 DEBUG megazone23 asSameOrder after if 1
## 2018 06 21 10:38:46.256 DEBUG megazone23 asSameOrder after if 2
## 2018 06 21 10:38:46.257 DEBUG megazone23 row.names(mat1)
## 2018 06 21 10:38:46.257 DEBUG megazone23 rownames(mat1)
## 2018 06 21 10:38:46.257 DEBUG megazone23 row.names(mat2)
## 2018 06 21 10:38:46.257 DEBUG megazone23 rownames(mat2)
## 2018 06 21 10:38:46.259 DEBUG megazone23 m.i
## 2018 06 21 10:38:46.370 DEBUG megazone23 asSameOrder after if 3
## 2018 06 21 10:38:46.370 DEBUG megazone23 dim(ebObj@mat1Com)=16146 586
## 2018 06 21 10:38:46.370 DEBUG megazone23 dim(ebObj@mat2Com)=16146 586
## 2018 06 21 10:38:46.371 DEBUG megazone23 get Validation and Training sets from vectors
## 2018 06 21 10:38:46.538 DEBUG megazone23 dim(ebObj@mat1Validation)=20501 175
## 2018 06 21 10:38:46.539 DEBUG megazone23 dim(ebObj@mat2Validation)=17814 175
## 2018 06 21 10:38:46.539 DEBUG megazone23 dim(ebObj@mat1Train)=20501 288
## 2018 06 21 10:38:46.539 DEBUG megazone23 dim(ebObj@mat2Train)=17814 288
## 2018 06 21 10:38:46.539 DEBUG megazone23 before Test Set
## 2018 06 21 10:38:46.627 DEBUG megazone23 after Test Set
```

```
## 2018 06 21 10:38:46.883 DEBUG megazone23 train
## 2018 06 21 10:38:46.883 DEBUG megazone23 inside train
## 2018 06 21 10:38:46.883 DEBUG megazone23 train theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_T:
## 2018 06 21 10:38:46.915 DEBUG megazone23 colnames(Object@mat1Train)=TCGA-A1-A0SH-01A-11R-A084-07.RNA$
## 2018 06 21 10:38:46.916 DEBUG megazone23 colnames(Object@mat2Train)=TCGA-A1-A0SH-01A-11R-A084-07.Agi:
## 2018 06 21 10:38:46.916 DEBUG megazone23 inside train, call getData4EB
## 2018 06 21 10:38:47.455 DEBUG megazone23 data4EB <- as.matrix
## 2018 06 21 10:38:47.455 DEBUG megazone23 dim(data4EB)=16146 404
## 2018 06 21 10:38:47.456 DEBUG megazone23 cbinds and rbinds
## 2018 06 21 10:38:47.456 DEBUG megazone23 Object@DF1batch= RNASeqV2
## 2018 06 21 10:38:47.456 DEBUG megazone23 Object@DF2batch= Agilent4502
## 2018 06 21 10:38:47.458 DEBUG megazone23 row.names(sampBatch)
## 2018 06 21 10:38:47.458 DEBUG megazone23 TCGA-A1-A0SH-01A-11R-A084-07.RNASeqV2 TCGA-A1-A0SJ-01A-11R-/
## 2018 06 21 10:38:47.458 DEBUG megazone23 colnames(sampBatch)
## 2018 06 21 10:38:47.458 DEBUG megazone23 sample batch
## 2018 06 21 10:38:47.459 DEBUG megazone23 inside train, after getData4EB
## 2018 06 21 10:38:47.459 DEBUG megazone23 call EB MBatch
## 2018 06 21 10:38:47.459 DEBUG megazone23 TDC HERE trainEB mbatch
## 2018 06 21 10:38:47.459 DEBUG megazone23 EBNplus args Obj
## 2018 06 21 10:38:47.460 DEBUG megazone23 EBNplus args par.prior
## 2018 06 21 10:38:47.460 DEBUG megazone23 EBNplus args minSampleNum
## 2018 06 21 10:38:47.460 DEBUG megazone23 EBNplus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_
## 2018 06 21 10:38:47.460 DEBUG megazone23 dat
## [1] 16146   404
## 2018 06 21 10:38:47.461 DEBUG megazone23 saminfo
## [1] 404   2
## 2018 06 21 10:38:47.461 DEBUG megazone23 check column and row names
## 2018 06 21 10:38:47.568 DEBUG megazone23 in design.mat_plus
## 2018 06 21 10:38:47.569 DEBUG megazone23 in build.design_plus
## 2018 06 21 10:38:47.570 DEBUG megazone23 after build.design_plus loop
## 2018 06 21 10:38:47.571 DEBUG megazone23 design
## 2018 06 21 10:38:47.571 DEBUG megazone23 list.batch_plus(saminfo)
## 2018 06 21 10:38:47.572 DEBUG megazone23 in list.batch_plus
## 2018 06 21 10:38:47.573 DEBUG megazone23 list.batch_plus tmp1
## 2018 06 21 10:38:47.573 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2018 06 21 10:38:47.574 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2018 06 21 10:38:47.574 DEBUG megazone23 list.batch_plus batches
## 2018 06 21 10:38:47.574 DEBUG megazone23 after list.batch_plus(saminfo)
## 2018 06 21 10:38:47.575 DEBUG megazone23 n.batches
## 2018 06 21 10:38:47.575 DEBUG megazone23 n.array
## 2018 06 21 10:38:52.489 DEBUG megazone23 missbatch matrix
## 2018 06 21 10:38:52.547 DEBUG megazone23 There are  208  genes that were removed because of whole ba
## 2018 06 21 10:38:53.672 DEBUG megazone23 B.hat
## 2018 06 21 10:38:53.672 DEBUG megazone23 grand.mean
## 2018 06 21 10:38:54.307 DEBUG megazone23 var.pooled
## 2018 06 21 10:38:54.317 DEBUG megazone23 stand.mean 1
## 2018 06 21 10:38:54.393 DEBUG megazone23 stand.mean 2
## 2018 06 21 10:38:54.424 DEBUG megazone23 s.data
## 2018 06 21 10:38:55.588 DEBUG megazone23 gamma.hat
## 2018 06 21 10:38:56.671 DEBUG megazone23 plot priors
## 2018 06 21 10:38:56.671 DEBUG megazone23 plotPrior priorPlotsFile= /bea_testing/output/EBNPlus_Train

## 2018 06 21 10:38:57.493 DEBUG megazone23 parametric adjustments
## 2018 06 21 10:38:58.340 DEBUG megazone23 parametric adjustments
```

```
## 2018 06 21 10:38:58.828 DEBUG megazone23 in getBayesData
## 2018 06 21 10:38:58.829 DEBUG megazone23 dim(s.data)=15938 404
## 2018 06 21 10:38:58.829 DEBUG megazone23 dim(stand.mean)=15938 404
## 2018 06 21 10:38:59.018 DEBUG megazone23 dim(bayesdata)=15938 404
## 2018 06 21 10:38:59.073 DEBUG megazone23 Second dim(bayesdata)=15938 404
## 2018 06 21 10:38:59.074 DEBUG megazone23 keep the dim of Original.dat, keep the gene with all NA
## 2018 06 21 10:38:59.108 DEBUG megazone23 resultsDat
## 2018 06 21 10:38:59.108 DEBUG megazone23 after EB MBatch
## 2018 06 21 10:38:59.108 DEBUG megazone23 inside TEST branch MBatch
## 2018 06 21 10:38:59.156 DEBUG megazone23 test
## 2018 06 21 10:38:59.156 DEBUG megazone23 dim(mat1)=20501 86
## 2018 06 21 10:38:59.157 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:38:59.157 DEBUG megazone23 dim(mat2)=17814 86
## 2018 06 21 10:38:59.157 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:38:59.324 DEBUG megazone23 dim(dat)=16146 172
## 2018 06 21 10:38:59.324 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:38:59.347 DEBUG megazone23 after parlist
## 2018 06 21 10:38:59.348 DEBUG megazone23 dim(dat)=16146 172
## 2018 06 21 10:38:59.348 DEBUG megazone23 valid.genes
## 2018 06 21 10:38:59.350 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:38:59.374 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:38:59.374 DEBUG megazone23 after Non-valid
## 2018 06 21 10:38:59.375 DEBUG megazone23 dim(dat)=15938 172
## 2018 06 21 10:38:59.406 DEBUG megazone23 after getStandData
## 2018 06 21 10:38:59.407 DEBUG megazone23 dim(stand.data)=15938 172
## 2018 06 21 10:38:59.407 DEBUG megazone23 batches
## 2018 06 21 10:38:59.407 DEBUG megazone23 after batch.design
## 2018 06 21 10:38:59.408 DEBUG megazone23 dim(batch.design)=172 2
## 2018 06 21 10:38:59.408 DEBUG megazone23 n.batches
## 2018 06 21 10:38:59.408 DEBUG megazone23 in getBayesData
## 2018 06 21 10:38:59.408 DEBUG megazone23 dim(s.data)=15938 172
## 2018 06 21 10:38:59.409 DEBUG megazone23 dim(stand.mean)=15938 172
## 2018 06 21 10:38:59.472 DEBUG megazone23 dim(bayesdata)=15938 172
## 2018 06 21 10:38:59.486 DEBUG megazone23 Second dim(bayesdata)=15938 172
## 2018 06 21 10:38:59.487 DEBUG megazone23 after bayesData
## 2018 06 21 10:38:59.487 DEBUG megazone23 dim(bayesData)=15938 172
## 2018 06 21 10:38:59.488 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:38:59.501 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:38:59.502 DEBUG megazone23 Check the adjusted test set
## 2018 06 21 10:38:59.502 DEBUG megazone23 after train
## 2018 06 21 10:38:59.503 DEBUG megazone23 EBadj
## 2018 06 21 10:38:59.503 DEBUG megazone23 EBadj train
## 2018 06 21 10:38:59.504 DEBUG megazone23 train
## 2018 06 21 10:38:59.504 DEBUG megazone23 dim(mat1)=20501 202
## 2018 06 21 10:38:59.504 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:38:59.505 DEBUG megazone23 dim(mat2)=17814 202
## 2018 06 21 10:38:59.505 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:38:59.779 DEBUG megazone23 dim(dat)=16146 404
## 2018 06 21 10:38:59.780 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:38:59.832 DEBUG megazone23 after parlist
## 2018 06 21 10:38:59.832 DEBUG megazone23 dim(dat)=16146 404
## 2018 06 21 10:38:59.833 DEBUG megazone23 valid.genes
## 2018 06 21 10:38:59.834 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:38:59.895 DEBUG megazone23 Non-valid genes were removed before adjustment
```

```
## 2018 06 21 10:38:59.896 DEBUG megazone23 after Non-valid
## 2018 06 21 10:38:59.896 DEBUG megazone23 dim(dat)=15938 404
## 2018 06 21 10:39:00.015 DEBUG megazone23 after getStandData
## 2018 06 21 10:39:00.015 DEBUG megazone23 dim(stand.data)=15938 404
## 2018 06 21 10:39:00.016 DEBUG megazone23 batches
## 2018 06 21 10:39:00.016 DEBUG megazone23 after batch.design
## 2018 06 21 10:39:00.016 DEBUG megazone23 dim(batch.design)=404 2
## 2018 06 21 10:39:00.017 DEBUG megazone23 n.batches
## 2018 06 21 10:39:00.017 DEBUG megazone23 in getBayesData
## 2018 06 21 10:39:00.017 DEBUG megazone23 dim(s.data)=15938 404
## 2018 06 21 10:39:00.018 DEBUG megazone23 dim(stand.mean)=15938 404
## 2018 06 21 10:39:00.253 DEBUG megazone23 dim(bayesdata)=15938 404
## 2018 06 21 10:39:00.283 DEBUG megazone23 Second dim(bayesdata)=15938 404
## 2018 06 21 10:39:00.284 DEBUG megazone23 after bayesData
## 2018 06 21 10:39:00.285 DEBUG megazone23 dim(bayesData)=15938 404
## 2018 06 21 10:39:00.285 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:39:00.314 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:39:00.315 DEBUG megazone23 objafterEB@trainEB
## 2018 06 21 10:39:00.315 DEBUG megazone23 EBadj validation
## 2018 06 21 10:39:00.315 DEBUG megazone23 validation
## 2018 06 21 10:39:00.333 DEBUG megazone23 dim(mat1)=20501 175
## 2018 06 21 10:39:00.334 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:39:00.334 DEBUG megazone23 dim(mat2)=17814 175
## 2018 06 21 10:39:00.335 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:39:00.646 DEBUG megazone23 dim(dat)=16146 350
## 2018 06 21 10:39:00.647 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:39:00.716 DEBUG megazone23 after parlist
## 2018 06 21 10:39:00.717 DEBUG megazone23 dim(dat)=16146 350
## 2018 06 21 10:39:00.717 DEBUG megazone23 valid.genes
## 2018 06 21 10:39:00.723 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:39:00.779 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:39:00.780 DEBUG megazone23 after Non-valid
## 2018 06 21 10:39:00.780 DEBUG megazone23 dim(dat)=15938 350
## 2018 06 21 10:39:00.851 DEBUG megazone23 after getStandData
## 2018 06 21 10:39:00.852 DEBUG megazone23 dim(stand.data)=15938 350
## 2018 06 21 10:39:00.852 DEBUG megazone23 batches
## 2018 06 21 10:39:00.854 DEBUG megazone23 after batch.design
## 2018 06 21 10:39:00.854 DEBUG megazone23 dim(batch.design)=350 2
## 2018 06 21 10:39:00.855 DEBUG megazone23 n.batches
## 2018 06 21 10:39:00.856 DEBUG megazone23 in getBayesData
## 2018 06 21 10:39:00.857 DEBUG megazone23 dim(s.data)=15938 350
## 2018 06 21 10:39:00.858 DEBUG megazone23 dim(stand.mean)=15938 350
## 2018 06 21 10:39:00.999 DEBUG megazone23 dim(bayesdata)=15938 350
## 2018 06 21 10:39:01.035 DEBUG megazone23 Second dim(bayesdata)=15938 350
## 2018 06 21 10:39:01.036 DEBUG megazone23 after bayesData
## 2018 06 21 10:39:01.036 DEBUG megazone23 dim(bayesData)=15938 350
## 2018 06 21 10:39:01.037 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:39:01.248 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:39:01.249 DEBUG megazone23 objafterEB@validationEB
## 2018 06 21 10:39:01.250 DEBUG megazone23 EBadj train for test results
## 2018 06 21 10:39:01.250 DEBUG megazone23 test
## 2018 06 21 10:39:01.250 DEBUG megazone23 dim(mat1)=20501 86
## 2018 06 21 10:39:01.250 DEBUG megazone23 dim(mat1)=
## 2018 06 21 10:39:01.251 DEBUG megazone23 dim(mat2)=17814 86
```

```
## 2018 06 21 10:39:01.251 DEBUG megazone23 dim(mat2)=
## 2018 06 21 10:39:01.403 DEBUG megazone23 dim(dat)=16146 172
## 2018 06 21 10:39:01.404 DEBUG megazone23 dim(dat)=
## 2018 06 21 10:39:01.428 DEBUG megazone23 after parlist
## 2018 06 21 10:39:01.428 DEBUG megazone23 dim(dat)=16146 172
## 2018 06 21 10:39:01.428 DEBUG megazone23 valid.genes
## 2018 06 21 10:39:01.430 DEBUG megazone23 nonValidGenes
## 2018 06 21 10:39:01.455 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2018 06 21 10:39:01.455 DEBUG megazone23 after Non-valid
## 2018 06 21 10:39:01.455 DEBUG megazone23 dim(dat)=15938 172
## 2018 06 21 10:39:01.491 DEBUG megazone23 after getStandData
## 2018 06 21 10:39:01.491 DEBUG megazone23 dim(stand.data)=15938 172
## 2018 06 21 10:39:01.492 DEBUG megazone23 batches
## 2018 06 21 10:39:01.492 DEBUG megazone23 after batch.design
## 2018 06 21 10:39:01.493 DEBUG megazone23 dim(batch.design)=172 2
## 2018 06 21 10:39:01.493 DEBUG megazone23 n.batches
## 2018 06 21 10:39:01.494 DEBUG megazone23 in getBayesData
## 2018 06 21 10:39:01.494 DEBUG megazone23 dim(s.data)=15938 172
## 2018 06 21 10:39:01.494 DEBUG megazone23 dim(stand.mean)=15938 172
## 2018 06 21 10:39:01.552 DEBUG megazone23 dim(bayesdata)=15938 172
## 2018 06 21 10:39:01.564 DEBUG megazone23 Second dim(bayesdata)=15938 172
## 2018 06 21 10:39:01.565 DEBUG megazone23 after bayesData
## 2018 06 21 10:39:01.565 DEBUG megazone23 dim(bayesData)=15938 172
## 2018 06 21 10:39:01.565 DEBUG megazone23 EBadj datNonValid
## 2018 06 21 10:39:01.577 DEBUG megazone23 EBadj EBadj after non valid
## 2018 06 21 10:39:01.578 DEBUG megazone23 testResults@testEB
## [1] "TestSet1"
## [1] TRUE
## [1] 20501     86
##         TCGA-A2-A04N-01A-11R-A115-07 TCGA-A2-A04T-01A-21R-A034-07
## A1BG                     7.521504                   6.03884438
## A1CF                    -2.016249                  -2.01624935
## A2BP1                   -2.016249                  -0.08512937
## A2LD1                    6.995044                   6.25703325
##         TCGA-A2-A04X-01A-21R-A034-07 TCGA-A2-A0CT-01A-31R-A056-07
## A1BG                     5.431493                    8.0208855
## A1CF                    -2.016249                   -2.0162494
## A2BP1                   -2.016249                   -0.5879231
## A2LD1                    5.904990                    6.2884564
## [1] "TestSet2"
## [1] TRUE
## [1] 17814     86
##         TCGA-A2-A04N-01A-11R-A115-07 TCGA-A2-A04T-01A-21R-A034-07
## 15E1.2                  -0.94900000                  -1.2670000
## 2'-PDE                  -0.11818750                   0.2290625
## 7A5                      2.64550000                   1.2170000
## A1BG                     0.08566667                  -0.0885000
##         TCGA-A2-A04X-01A-21R-A034-07 TCGA-A2-A0CT-01A-31R-A056-07
## 15E1.2                  -0.6882500                   -0.664000
## 2'-PDE                   0.3431250                    0.466875
## 7A5                      2.1730000                   -0.791000
## A1BG                    -0.6243333                    1.232000
## [1] "TrainingSet1"
## [1] TRUE
```

```
## [1] 20501   202
##        TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A1-A0SJ-01A-11R-A084-07
## A1BG                      7.4964200                     7.826722
## A1CF                     -2.0162494                    -2.016249
## A2BP1                    -0.7169079                    -2.016249
## A2LD1                     7.0294683                     7.035494
##        TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04W-01A-31R-A115-07
## A1BG                      7.7133865                     7.185537
## A1CF                     -0.7145386                    -2.016249
## A2BP1                    -2.0162494                    -2.016249
## A2LD1                     6.0826238                     6.549330
## [1] "TrainingSet2"
## [1] TRUE
## [1] 17814   202
##        TCGA-A1-A0SH-01A-11R-A084-07 TCGA-A1-A0SJ-01A-11R-A084-07
## 15E1.2                   -1.9440000                    -2.0980000
## 2'-PDE                   -0.0295625                     0.7919375
## 7A5                       1.8550000                     1.4080000
## A1BG                      0.7486667                     0.7500000
##        TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04W-01A-31R-A115-07
## 15E1.2                    -1.652250                    -1.6490000
## 2'-PDE                     0.635625                     0.4400625
## 7A5                       -0.392000                     3.4470000
## A1BG                       1.278000                    -0.7923333
## [1] "TrainingResults"
## [1] TRUE
## [1] 16146   404
##        TCGA-A1-A0SH-01A-11R-A084-07.Agilent4502
## A1BG                                    4.357751
## A2BP1                                  -0.292937
## A2M                                     7.568959
## A2ML1                                   1.486490
##        TCGA-A1-A0SH-01A-11R-A084-07.RNASeqV2
## A1BG                                3.99436879
## A2BP1                              -1.26534147
## A2M                                 7.36653343
## A2ML1                               0.07972427
##        TCGA-A1-A0SJ-01A-11R-A084-07.Agilent4502
## A1BG                                   4.3594847
## A2BP1                                 -0.3726113
## A2M                                    8.0890437
## A2ML1                                  1.9105608
##        TCGA-A1-A0SJ-01A-11R-A084-07.RNASeqV2
## A1BG                                 4.2718734
## A2BP1                               -2.1246373
## A2M                                  8.0269204
## A2ML1                               -0.4831787
## [1] "ValidationSet1"
## [1] TRUE
## [1] 20501   175
##        TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A1-A0SK-01A-12R-A084-07
## A1BG                      7.1552716                     6.441132
## A1CF                     -2.0162494                    -2.016249
## A2BP1                    -0.7908586                    -2.016249
```

11

```
## A2LD1                         6.0684602                        8.659733
##       TCGA-A1-A0SM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG                         8.2655724                        6.768147
## A1CF                        -2.0162494                       -2.016249
## A2BP1                       -0.8290494                       -2.016249
## A2LD1                        6.6860985                        6.474090
## [1] "ValidationSet2"
## [1] TRUE
## [1] 17814   175
##        TCGA-A1-A0SD-01A-11R-A115-07 TCGA-A1-A0SK-01A-12R-A084-07
## 15E1.2                    -1.4477500                      -1.2850000
## 2'-PDE                     0.0153750                      -0.1911250
## 7A5                        2.4105000                      -1.3005000
## A1BG                       0.9493333                      -0.5471667
##        TCGA-A1-A0SM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## 15E1.2                    -2.1540000                      -1.5990000
## 2'-PDE                     0.3263125                      -0.3471250
## 7A5                        2.1600000                       1.7275000
## A1BG                       0.5976667                      -0.1568333
## [1] "ValidationResults"
## [1] TRUE
## [1] 16146   350
##        TCGA-A1-A0SD-01A-11R-A115-07.Agilent4502
## A1BG                                   4.618661
## A2BP1                                  2.479128
## A2M                                    6.645426
## A2ML1                                  1.660333
##        TCGA-A1-A0SD-01A-11R-A115-07.RNASeqV2
## A1BG                                3.707752
## A2BP1                              -1.314247
## A2M                                8.132635
## A2ML1                              1.659238
##        TCGA-A1-A0SK-01A-12R-A084-07.Agilent4502
## A1BG                                   2.672890
## A2BP1                                 -1.615330
## A2M                                    5.537639
## A2ML1                                  1.350840
##        TCGA-A1-A0SK-01A-12R-A084-07.RNASeqV2
## A1BG                                3.107764
## A2BP1                              -2.124637
## A2M                                5.624822
## A2ML1                              2.855925
## [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.
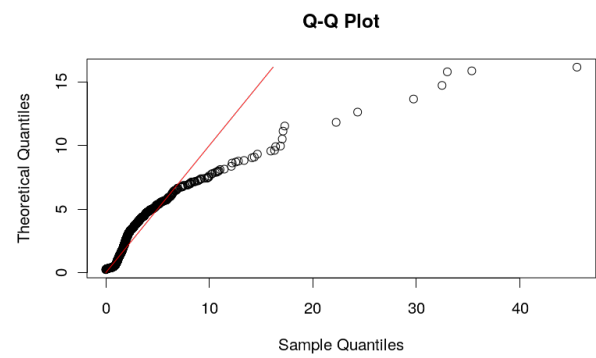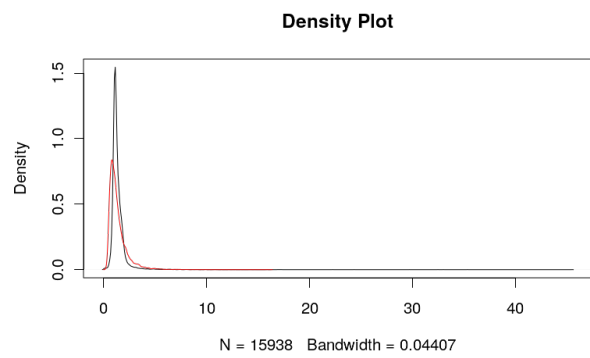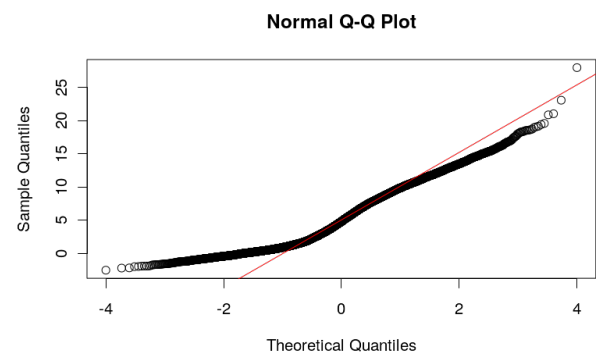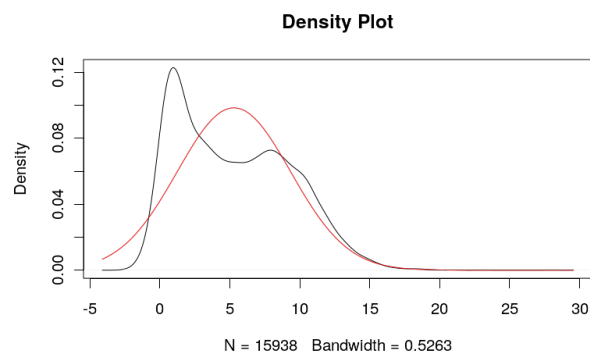
This also generates the prior plots output file like this one.

Figure 1: Prior Plots File