

Very Brief Normalization

Aedín Culhane

December 14, 2011

Contents

1	Sample workflows	1
2	normalization Affymetrix data	1
3	normalization Affymetrix data -custom cdf	1
4	ArrayQualityMetrics - checking for data quality	2
5	vsn normalization of data.. for information only	2

1 Sample workflows

Sample workflows for processing gene expression and other genomics data are available on the Bioconductor website. Gene Expression data are specifically discussed <http://www.bioconductor.org/help/workflows/oligo-arrays/>

2 normalization Affymetrix data

Download a .zip of 9 cel file cels.zip from the coures website. Clicking on this will extract the files to a folder cels.

```
> require(affy)
> require(made4)
> eset1<-justRMA(celfile.path="cels")

> overview(eset1)
> annt<-read.table("annt.txt", header=TRUE)
> overview(eset1, labels=annt$Donor, classvec=annt$Donor)
```

3 normalization Affymetrix data -custom cdf

Several groups, including the University of Michigan BrainArray project provide custom cdf files will allow one to summarize Affymetrix probes to the gene, transcript rather than probeset. <http://brainarray.mbni.med.umich.edu/brainarray/default.asp>

For R version 2.14 the custom cdf packages are available on <http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/14.1.0/ensg.asp>. Download the EnSEMBL gene summaries for hgu95av2.

Whilst the following should work

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite("HGU95Av2_Hs_ENSG", respos="http://brainarray.mbni.med.umich.edu/bioc")
```

It may not in which case you need to download 2 package and do a local install of the .zip files (or source, depending on your Operating system)

The are 2 files: the cdf file http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/14.1.0/ensg.download/hgu95av2hsensgcdf_14.1.0.zip and the probe file http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/14.1.0/ensg.download/hgu95av2hsensgprobe_14.1.0.zip

Download these, install and now run the normalization

```
> abatch<-ReadAffy(celfile.path="cels",cdfname = "HGU95Av2_Hs_ENSG" )
> eset2<-rma(abatch)
```

4 ArrayQualityMetrics - checking for data quality

5 fRMA

fRMA is frozen RMA and is useful when trying to collate many studies from U133a or U133plus2 Affymetrix chips

6 vsn normalization of data

For information only, please don't repeat in this today. *vsn* was used to normalize the Affymetrix data. To produce the normalized data, the cel files were downloaded, and then in R, use File -> Change directory (or `setwd` to select the directory containing the cel files. Then

```
getwd()
dir()

library(affy)
```

```

library(vsn)
cels <- list.celfiles()
data <- ReadAffy(filenamees= cels)
normalize.AffyBatch.methods <- c(normalize.AffyBatch.methods, "vs")

data.vsn <- es1 = expresso(data, bg.correct = FALSE,
normalize.method = "vs", pmcorrect.method = "pmonly",
summary.method    = "medianpolish")

exprs2excel(data.vsn, file="data.vsn.csv")

  A tab delimited text file could also be saved using

write.exprs(data.vsn, file="data.rma.txt")

```