

R Code to Pre-process Data from Docetaxel Clinical Trial.

Paul Geeleher

September 3, 2013

This file contains the full instructions to download and pre-process the Docetaxel clinical trial data so that it can be used to reproduce Figure 2 (and associated results) in the paper.

The raw CEL files should be downloaded from GEO (accession numbers “*GSE349*” and “*GSE350*”). You must specify where your files have been downloaded. It is assumed that the sensitive samples are in a directory named “*GSE350_RAW/*” and the resistant “*GSE349_RAW/*” and that these are contained in the same directory, which is specified by “*downloadDir*”.

NOTE: These directories will need to be altered based on where you have downloaded and unzipped the .CEL files.

```
> # setwd("/path/to/this/file/")
> downloadDir <- "/home/pgeeher/postdoc_stuff_local/test_ridge/data/docetaxel_vivo/sensit
> sensDirectory <- paste(downloadDir, "GSE350_RAW/", sep="")
> resDirectory <- paste(downloadDir, "GSE349_RAW/", sep="")
```

Load the “affy” library. NOTE: If this hasn’t been installed you will need to install it.

```
> library(affy)
```

Read and RMA normalize the expression data. The array platform is Affymetrix Human Genome U95 Version 2 Array and the CDF file can be downloaded from BrainArray website:

<http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/17.0.0/entrezg.asp>.

Once the CDF name is specified @*cdfName*, it is automatically downloaded, but note that to reproduce our results exactly, you will need to use version 17 of the CDF.

```
> sensCELS <- paste(sensDirectory, dir(sensDirectory), sep="")
> resCELS <- paste(resDirectory, dir(resDirectory), sep="")
> doceVivoRaw <- ReadAffy(filename=c(sensCELS, resCELS))
> doceVivoRaw@cdfName <- "HG95Av2_Hs_ENTREZG"
> doceVivoNorm <- rma(doceVivoRaw)
```

Background correcting

Normalizing

Calculating Expression

Next, map the gene identifiers to gene symbols. To do this we need to download the annotation package from BrainArray. The file can be downloaded from the link below. On Linux/MAC this can be done by issuing the following commands from the command prompt:

```
$ wget http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/17.0.0/entrezg.download/
$ R CMD INSTALL hgu95av2hsentrezg.db_17.0.0.tar.gz
```

Then, in R, we need to load the annotation file and map the gene identifiers to gene symbols.

```
> library("hgu95av2hsentrezg.db")
> x <- hgu95av2hsentrezgSYMBOL
> mapped_probes_brain <- mappedkeys(x)
> names(mapped_probes_brain) <- as.character(x[mapped_probes_brain])
> entrez2sym_brain <- as.character(x[mapped_probes_brain])
> symBrain2entrez <- entrez2sym_brain
> names(symBrain2entrez) <- names(entrez2sym_brain)
> doceVivoNorm_syms <- exprs(doceVivoNorm)
> rownames(doceVivoNorm_syms) <- symBrain2entrez[rownames(exprs(doceVivoNorm))]
```

Finally, save the expression matrix and a factor to indicate which samples are sensitive/resistant.

```
> groupMembership <- c(rep("sens", 10), rep("res", 14))
> save(doceVivoNorm_syms, groupMembership,
+ file="../Data/docetaxelData/doce_rma_syms_brainArray.RData")

> sessionInfo()
```

```
R version 2.15.2 (2012-10-26)
Platform: x86_64-pc-linux-gnu (64-bit)
```

locale:

```
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
[5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=C               LC_NAME=C
[9] LC_ADDRESS=C             LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

other attached packages:

```
[1] hgu95av2hsentrezg.db_17.0.0 org.Hs.eg.db_2.8.0
[3] RSQLite_0.11.4              DBI_0.2-7
[5] hgu95av2hsentrezgcdf_17.1.0 AnnotationDbi_1.20.7
[7] affy_1.36.1                 Biobase_2.18.0
[9] BiocGenerics_0.4.0
```

loaded via a namespace (and not attached):

```
[1] affyio_1.26.0              BiocInstaller_1.8.3        IRanges_1.16.6
```

```
[4] parallel_2.15.2      preprocessCore_1.20.0 stats4_2.15.2
[7] tools_2.15.2          zlibbioc_1.4.0
```