

GS01 0163

Analysis of Microarray Data

Keith Baggerly and Kevin Coombes
Section of Bioinformatics
Department of Biostatistics and Applied Mathematics
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org
kcoombes@mdanderson.org

2 November 2006

dChip is Tricky?

Reported answers:

444, 594, 595*, 604, 604, 605, 609, 609, 615, 896*.

Answers we got:

610, using dChip2004

604, using dChip2006

900 or so, using PM-MM.

We didn't check the intersections. Reporting version as well as filters may be important.

justRMA() is tricky

I grouped all 44 CEL files into a single folder and ran justRMA(). The resulting exprSet had 44 columns, even though 2 of the CEL files were U95Av2. The annotation indicated that the arrays were hgu95a.

It trusts you.

adding phenoData is tricky

After putting the 2 U95Av2 files in a different folder, I ran justRMA() again. I also loaded in the sample info file that we'd used for dChip in order to use it as phenoData here.

```
> zed <- justRMA();  
> pD1 <- read.phenoData("krc-sample-info.xls",  
                        sep="\t", header=TRUE);
```

what do we have?

```
> pData(zed)[1:4,]
```

```
              sample
CL2001011101AA.CEL      1
CL2001011102AA.CEL      2
CL2001011104AA.CEL      3
CL2001011105AA.CEL      4
```

```
> pData(pD1)[1:4,]
```

```
      Scan.name  Sample.Name  type  Split
1 CL2001011101AA      ALL01  ALL  Training
2 CL2001011104AA      ALL02  ALL  Training
3 CL2001011105AA      ALL03  ALL  Training
4 CL2001011108AA      ALL04  ALL  Training
```

Note that the order is not the same...

ok, let's assign something

```
> phenoData(zed) <- pD1;
> pData(zed)[1:4,]
      Scan.name Sample.Name type      Split
1 CL2001011101AA      ALL01  ALL Training
2 CL2001011104AA      ALL02  ALL Training
3 CL2001011105AA      ALL03  ALL Training
4 CL2001011108AA      ALL04  ALL Training
> colnames(zed@exprs)[1:4]
[1] "CL2001011101AA.CEL" "CL2001011102AA.CEL"
[3] "CL2001011104AA.CEL" "CL2001011105AA.CEL"
```

oops...

The columns weren't reordered when the phenoData changed.

phenoData assignment is trusting

There is very little validity checking involved. You can assign a phenoData object that has the wrong number of rows, or that has an order different from the ordering of the CEL files in the exprs matrix.

You need to check to make sure that the order matches.

If you supply the filenames option to `just.rma()` or `justRMA()`, the files will be loaded in the order specified.

how can we reorder things?

One option:

```
zed2 <- as.character(pData(pD1)$Scan.name);
sampleOrder <- order(charmatch(zed2,
                               colnames(zed@exprs)));
pD2 <- pD1;
pData(pD2) <- pData(pD2)[sampleOrder,];
phenoData(zed) <- pD2;
```

Other checks

```
if (length(colnames(zed@exprs)) !=  
    dim(pData(zed))[1]) {  
  print("Houston, we have a problem");  
}
```

Now, given that we have associated a phenoData object with the exprSet in the right way, we want to invoke contrasts using variables from here.

```
myTs <- MultiTtest(zed, "type");
```