Homework 3 review

There seems to be a big discrepancy

Directly use pre-normalized data

```
gds2938 <- getGEO("GDS2938")
eset <- GDS2eSet(gds2938)
x1 <- exprs(eset) # this is directly from GEO, pre-
     nomalized data
pdata=pData(eset)
TS=pdata[,3]
design <- model.matrix(~TS-1)
colnames(design)<-
     c('control','IFN.gamma','IL1.beta','both')
fit1 = lmFit(x1,design)
cont.matrix = makeContrasts(IFN=IFN.gamma-control,
                            IL1=IL1.beta-control,
                            BOTH=both-control,
                            levels=design)
fit1 = contrasts.fit(fit1, cont.matrix)
fit1 = eBayes(fit1)
colnames(x1)
 [1] "GSM113926" "GSM113927" "GSM113928" "GSM113929"
     "GSM113930" "GSM113931"
 [7] "GSM113932" "GSM113933" "GSM113934" "GSM113935"
     "GSM113936" "GSM113937"
 topTable(fit1,1,5)
            logFC AveExpr    t  P.Value adj.P.Val    B
202531_at   1.241    3.70 15.7 8.64e-09  0.000181 9.33
209545_s_at 0.706    3.44 14.5 1.99e-08  0.000181 8.78
217478_s_at 0.752    3.27 13.9 3.10e-08  0.000181 8.48
213537_at   0.779    2.79 13.8 3.24e-08  0.000181 8.45
212671_s_at 1.625    2.87 13.3 4.79e-08  0.000213 8.18
```

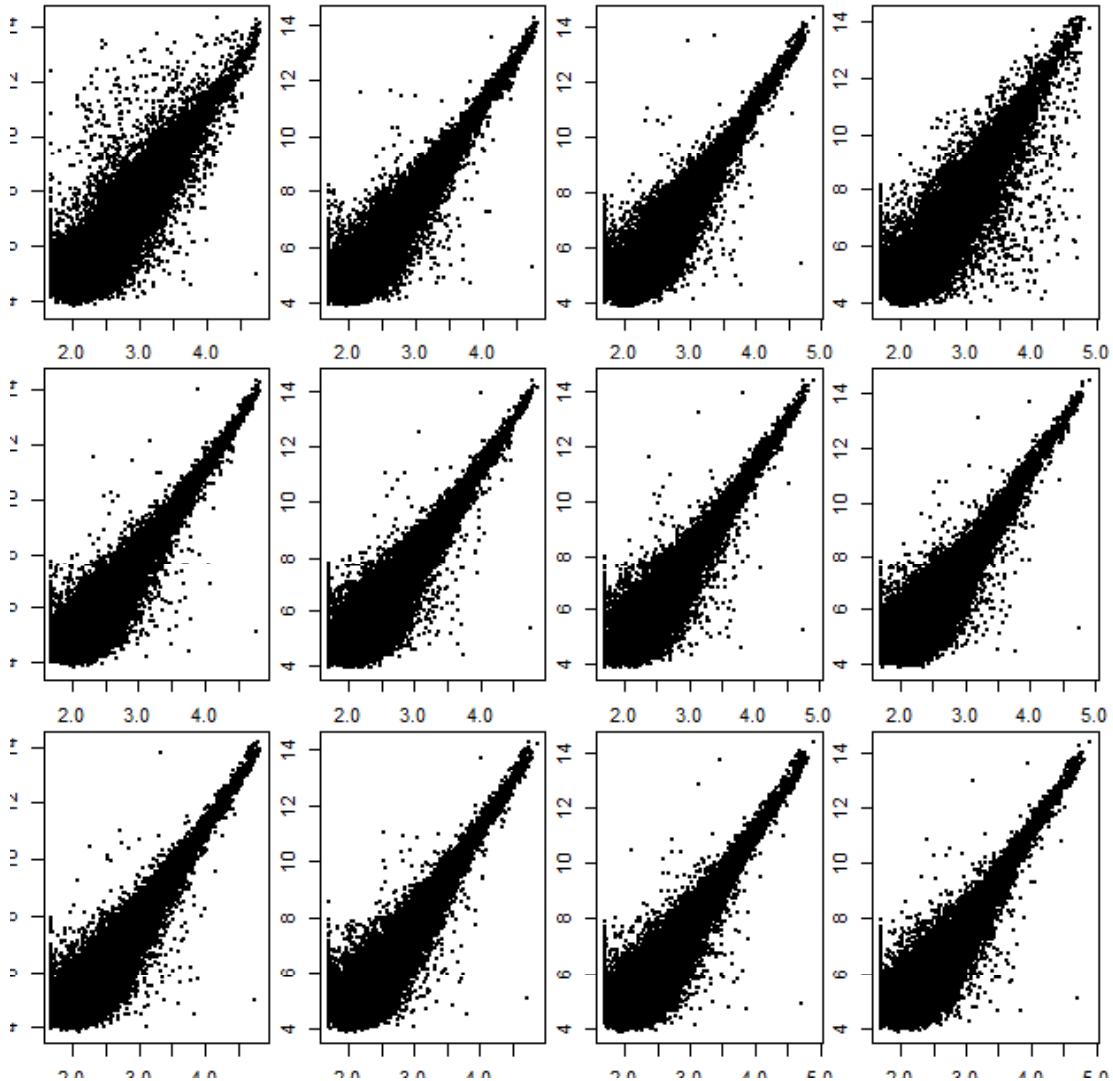
Download celfiles, preprocess by RMA

```
x0=ReadAffy(filenames=list.celfiles()[1:12])
x0.rma=rma(x0)
x0.exprs=exprs(x0.rma)
colnames(x0)
 [1] "GSM113926.CEL.gz" "GSM113927.CEL.gz"
     "GSM113928.CEL.gz" "GSM113929.CEL.gz"
 [5] "GSM113930.CEL.gz" "GSM113931.CEL.gz"
     "GSM113932.CEL.gz" "GSM113933.CEL.gz"
 [9] "GSM113934.CEL.gz" "GSM113935.CEL.gz"
     "GSM113936.CEL.gz" "GSM113937.CEL.gz“

topTable(fit0,1,5)
             logFC AveExpr     t  P.Value adj.P.Val     B
218573_at   -1.710    7.72 -6.65 6.03e-05         1 -2.78
210946_at    0.560    8.21  5.73 1.99e-04         1 -2.92
222258_s_at  0.737    9.26  5.57 2.48e-04         1 -2.95
218624_s_at -0.507    7.53 -5.18 4.27e-04         1 -3.04
203932_at    2.148    7.12  5.05 5.21e-04         1 -3.07
```

We know preprocessing matters. But I did not expect it to be so drastically different. Are we even analyzing the same data??

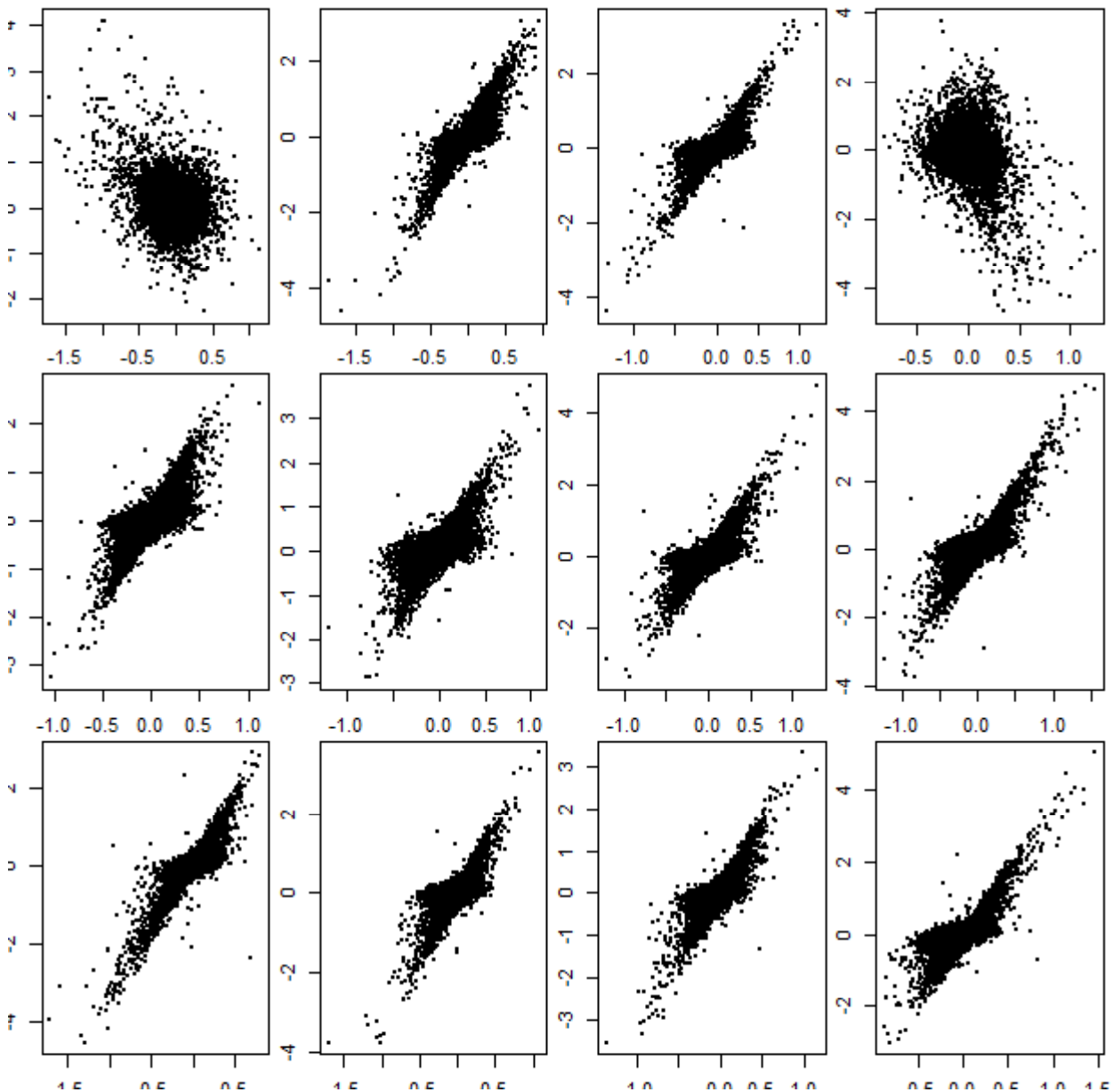
It appears reasonably reproduced



```
par(mfrow=c(3,4))
par(mai=c(.2,.2,.1,.1))
for( i in 1:12) plot(x1[,i],x0.exprs[,i],pch=16, cex=.3)
### but we know we'd expect good correlation here, even if we are comparing unrelate
    samples. The strong probe effect, and the wide dynamic range of gene expression
    (some genes are always high, some are always low) would make two samples
    correlated.


### However, if we are comparing two samples that are really measuring the same thing,
    they should be highly correlated even when we remove the gene's baseline effect.
### that means, compared to each gene's own reference, it should be higher in the same
    samples and lower in the same other samples.


X0=x0.exprs
X1=x1
X0=sweep(X0,1,rowMeans(X0))
X1=sweep(X1,1,rowMeans(X1))
for( i in 1:12) plot(X1[,i],X0[,i],pch=16, cex=.3)
```

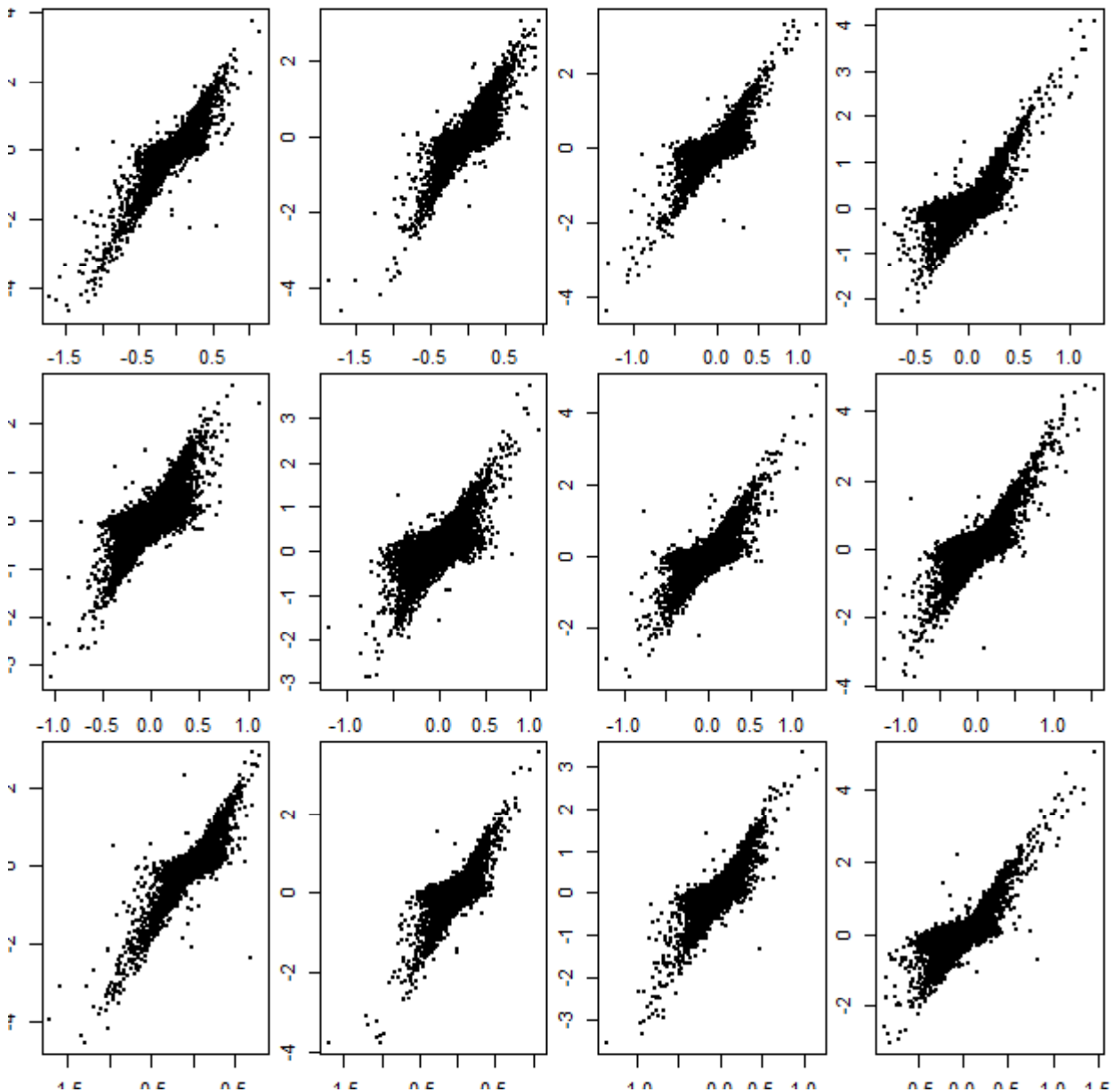


After we switch the two samples

```
x2=x0.exprs
x2=x2[,c(4,2,3,1,5:12)]
colnames(x2)=colnames(x0.exprs)
X2=sweep(x2,1,rowMeans(x2))
for( i in 1:12) plot(X1[,i],X2[,i],pch=16,
    cex=.3)

fit2 = lmFit(x2,design)
fit2 = contrasts.fit(fit2, cont.matrix)
fit2 = eBayes(fit2)
```

```
topTable(fit2,1,20)
             logFC AveExpr      t  P.Value adj.P.Val     B
212671_s_at  4.889    7.85  20.93 1.11e-09  2.48e-05 10.60
202531_at    4.012    9.98  16.23 1.37e-08  1.53e-04  9.14
200628_s_at  3.400   11.54  14.10 5.39e-08  2.89e-04  8.21
222288_at   -1.826    5.71 -13.97 5.92e-08  2.89e-04  8.14
209545_s_at  2.301    9.34  13.65 7.39e-08  2.89e-04  7.98
217478_s_at  2.749    8.65  13.58 7.78e-08  2.89e-04  7.94
210645_s_at -1.543    9.19 -12.39 1.87e-07  5.63e-04  7.28
213537_at    2.535    6.63  12.28 2.05e-07  5.63e-04  7.21
210029_at    6.643   10.32  12.09 2.37e-07  5.63e-04  7.09
218322_s_at  1.583    7.67  12.01 2.53e-07  5.63e-04  7.04
208306_x_at  3.001   10.64  11.53 3.73e-07  7.36e-04  6.73
208664_s_at -1.222    7.76 -11.46 3.97e-07  7.36e-04  6.68
203932_at    2.474    7.12  11.20 4.92e-07  8.44e-04  6.51
204670_x_at  2.969   10.68  10.93 6.18e-07  9.83e-04  6.32
209312_x_at  3.401   10.71  10.36 1.02e-06  1.51e-03  5.91
200075_s_at  1.090   10.10  10.27 1.11e-06  1.55e-03  5.83
203911_at   -0.893    7.75  -9.98 1.45e-06  1.83e-03  5.61
211991 s at  3.960    8.44   9.96 1.48e-06  1.83e-03  5.59
204770_at    2.177    7.65   9.67 1.94e-06  2.11e-03  5.36
215193_x_at  3.770    9.54   9.66 1.95e-06  2.11e-03  5.35
```



Conclusion?

- It appears there is some labeling error.
- I don’t know which way it is.
- Anyone interested in further investigation?