

Reanalysis Towfic2014 - figures and code

Peter Hettegger

27 9 2020

For reproducibility, the `sessionInfo()` is provided at the end of the document.

```
rm(list = ls())

library(GGally)
## Loading required package: ggplot2
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
library(ggplot2)
library(limma)
library(sva)
## Loading required package: mgcv
## Loading required package: nlme
## This is mgcv 1.8-31. For overview type 'help("mgcv-package")'.
## Loading required package: genefilter
## Loading required package: BiocParallel
library(randRotation)
```

File from <https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE40566&format=file&file=GSE40566%5Fnon%5Fnormalized%2Etxt%2Egz>

```
edata <- read.table(file = "GSE40566_non_normalized.txt", header = TRUE,
                     sep = "\t", dec = ".", stringsAsFactors = FALSE,
                     row.names = 1)
```

File from <https://github.com/ous-uio-bioinfo-core/batch-adjust-warning-figures/blob/master/reanalysis/Towfic2014/sampleannotation.txt>. See also Nygaard, V., Rodland, E. A. & Hovig, E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* kxv027 (2015). doi:10.1093/biostatistics/kxv027.

```
pdata <- read.table(file = "sampleAnnotation.txt", header = TRUE,
                     sep = "\t", dec = ".", stringsAsFactors = TRUE,
                     row.names = 1)

all.equal(colnames(edata), as.character(pdata$title))
## [1] "2 string mismatches"
edata <- edata[,as.character(pdata$title)]

pdata$batch <- as.factor(pdata$batch)
```

Reanalysis Towfic2014 - figures and code

```
# for debugging
debug = FALSE
if(debug) edata <- edata[1:1000,]

# quantile normalisation
edata.quan <- normalizeBetweenArrays(edata, method = "quantile")

##### ComBat - "p ComBat" values
mod.com <- model.matrix(~covariate, pdata)
edata.com <- ComBat(edata.quan,
                     batch = pdata$batch,
                     mod = mod.com)
## Found 6 genes with uniform expression within a single batch (all zeros); these will not be adjusted for batch
## Found 18 batches
## Adjusting for 15 covariate(s) or covariate level(s)
## Standardizing Data across genes
## Fitting L/S model and finding priors
## Finding parametric adjustments
## Adjusting the Data

mod.fit <- model.matrix(~0+covariate, pdata)
fit1 <- lmFit(edata.com, design = mod.fit)

DP corresponds to Copaxone, N corresponds to Glatimer

cont.mat <- makeContrasts(c1="covariateDP-covariateN", levels=mod.fit)

fit2 <- contrasts.fit(fit1, cont.mat)
fit2 <- eBayes(fit2)

ps.com <- topTable(fit2, number = Inf, sort.by = "none")$P.Value
fdr.com <- topTable(fit2, number = Inf, sort.by = "none")$adj.P.Val

sum(ps.com<0.005)
## [1] 2954
sum(fdr.com<0.05)
## [1] 2011

##### limma batch as covariate - "p Limma (+batch)" values

mod.fit <- model.matrix(~0+covariate + batch, pdata)
fit1 <- lmFit(edata.quan, design = mod.fit)

cont.mat <- makeContrasts(c1="covariateDP-covariateN", levels=mod.fit)

fit2 <- contrasts.fit(fit1, cont.mat)
fit2 <- eBayes(fit2)
```

Reanalysis Towfic2014 - figures and code

```
ps.lim <- topTable(fit2, number = Inf, sort.by = "none")$P.Value
fdr.lim <- topTable(fit2, number = Inf, sort.by = "none")$adj.P.Val

sum(ps.lim<0.005)
## [1] 566
sum(fdr.lim<0.05)
## [1] 11

##### ComBat with random rotations - "p ComBat - 2000 rot." values

library(randRotation)

mod.fit <- model.matrix(~0+covariate, pdata)
cont.mat <- makeContrasts(c1="covariateDP-covariateN", levels=mod.fit)

X.s <- contrastModel(mod.fit, cont.mat)

rr1 <- initBatchRandrot(edata.quan, X.s, 16, pdata$batch)

statistic <- function(Y, batch, mod.com, mod.fit, cont.mat){
  edata.com <- sva::ComBat(Y,
                            batch = batch,
                            mod = mod.com, mean.only = FALSE)

  fit1 <- limma::lmFit(edata.com, design = mod.fit)
  fit2 <- limma::contrasts.fit(fit1, cont.mat)
  fit2 <- limma::eBayes(fit2)

  abs(limma::topTable(fit2, number = Inf, sort.by = "none")$t)
}

rs1 <- rotateStat(rr1, R = 2000, statistic = statistic, pdata$batch, mod.com,
                   mod.fit, cont.mat, parallel = TRUE)

ps.rot <- pFdr(rs1)

sum(ps.rot < 0.005)
## [1] 498

fdr.rot <- p.adjust(ps.rot, "BH")
sum(fdr.rot<0.05)
## [1] 8

ps <- cbind(ps.com, ps.lim, ps.rot = ps.rot[,1])
colnames(ps) <- c("p ComBat", "p Limma (+batch)", "p ComBat - 2000 rot.")

##### p-vals scatterplot

df1 <- data.frame(ps)
colnames(df1) <- colnames(ps)
```

Reanalysis Towfic2014 - figures and code

```

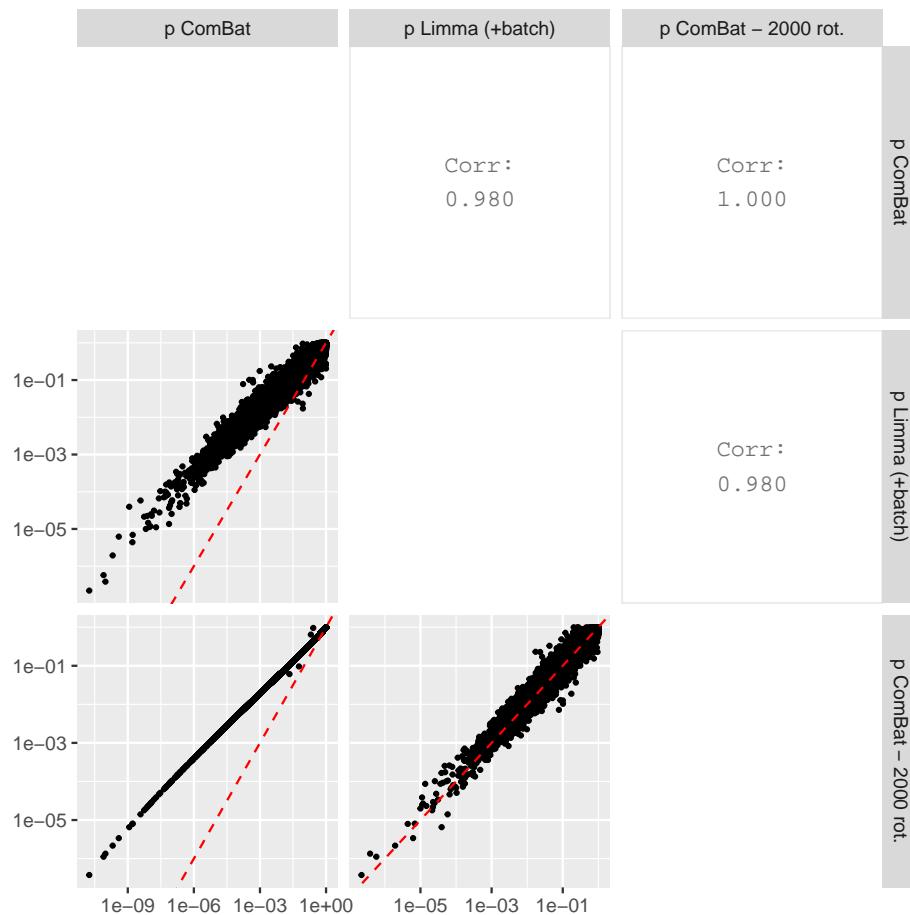
F1 <- function(...){
  ggally_points(..., size = 0.7) +
    scale_y_log10() +
    scale_x_log10() +
    geom_abline(slope = 1, intercept = 0, lty = 2, lwd = 0.5, col = "red")
}
lower.pan <- list(continuous = F1, combo = "facethist", discrete = "facetbar",
                  na = "na")

my.cor <- function(...)
  ggally_statistic(
    text_fn = function(x,y)
      formatC(cor(log(x),log(y)),digits = 3, format = "f"), title = "Corr",
      sep = ":\n",...)

upper.pan <- list(continuous = my.cor, combo = "box_no_facet",
                  discrete = "count", na = "na")

ggpairs(df1, lower = lower.pan, upper = upper.pan, diag = NULL)

```

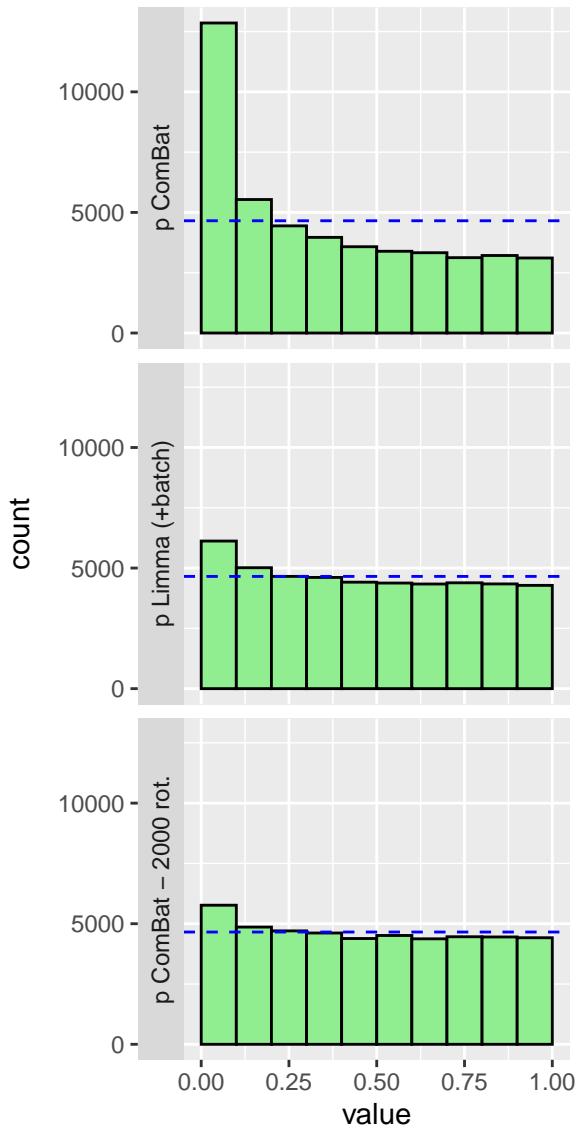


Reanalysis Towfic2014 - figures and code

```
#### Histograms

df2 <- reshape2::melt(ps)

ggplot(df2, aes(x=value))+
  geom_histogram(colour="black", fill="lightgreen", binwidth = 0.1, boundary=0)+
  facet_grid(Var2 ~ ., switch = "y")+
  geom_abline(slope = 0, intercept = nrow(ps)/10, lty = 2, col = "blue")+
  theme(axis.title.y = element_text(vjust=+3.3))
```



```
## Histograms 2

ind <- 1:15
h.com <- hist(ps.com, breaks = 100, plot = FALSE)
h.lim <- hist(ps.lim, breaks = 100, plot = FALSE)
```

Reanalysis Towfic2014 - figures and code

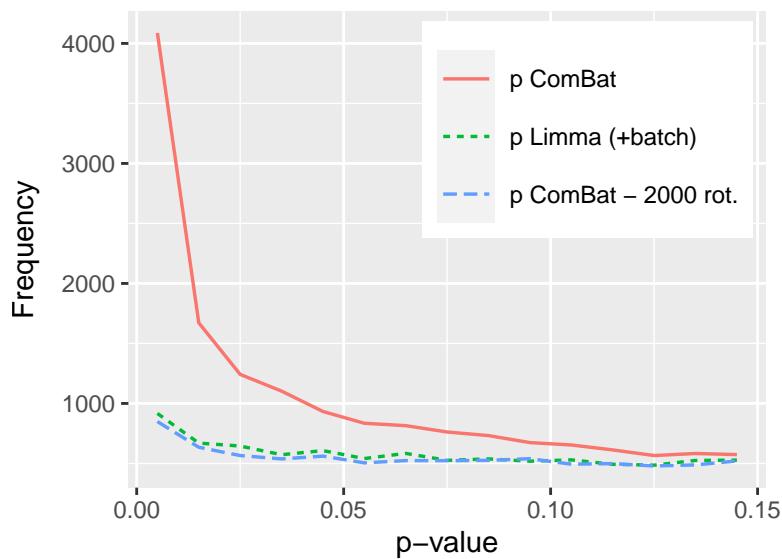
```
h.rot <- hist(ps.rot, breaks = 100, plot = FALSE)

maxcount <- max(h.com$counts, h.lim$counts, h.rot$counts)

lab <- factor(rep(colnames(ps), rep(length(ind),3)), levels = colnames(ps))

df1 <- data.frame(mids = c(h.com$mids[ind], h.lim$mids[ind], h.rot$mids[ind]),
                   counts =
                     c(h.com$counts[ind], h.lim$counts[ind], h.rot$counts[ind]),
                   lab = lab)

ggplot(df1, aes(x = mids, y = counts, colour = lab, lty = lab))+
  geom_line(lwd = 0.6) +
  xlab("p-value") + ylab("Frequency") +
  theme(axis.title.y = element_text(vjust=+3.3)) +
  theme(legend.justification=c(1,1), legend.position=c(0.98, 0.98),
        legend.title = element_blank())+
  theme(legend.key.size = unit(1.5,"line"))
```



1 Session Info

```
sessionInfo()
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Austria.1252  LC_CTYPE=German_Austria.1252
```

Reanalysis Towfic2014 - figures and code

```
## [3] LC_MONETARY=German_Austria.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Austria.1252
##
## attached base packages:
## [1] stats      graphics   grDevices  utils      datasets   methods    base
##
## other attached packages:
## [1] randRotation_1.1.4    sva_3.37.0        BiocParallel_1.23.2
## [4] genefilter_1.71.0     mgcv_1.8-31       nlme_3.1-148
## [7] limma_3.45.13       GGally_2.0.0       ggplot2_3.3.2
## [10] BiocStyle_2.17.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.5          locfit_1.5-9.4    lattice_0.20-41
## [4] snow_0.4-3          digest_0.6.25     R6_2.4.1
## [7] plyr_1.8.6          stats4_4.0.2      RSQLite_2.2.0
## [10] evaluate_0.14       pillar_1.4.6      Rdpack_1.0.0
## [13] rlang_0.4.7         annotate_1.67.1   blob_1.2.1
## [16] S4Vectors_0.27.12   Matrix_1.2-18     rmarkdown_2.3
## [19] labeling_0.3         splines_4.0.2     stringr_1.4.0
## [22] RCurl_1.98-1.2      bit_4.0.4        munsell_0.5.0
## [25] compiler_4.0.2      xfun_0.16        pkgconfig_2.0.3
## [28] BiocGenerics_0.35.4 htmltools_0.5.0   tidyselect_1.1.0
## [31] tibble_3.0.3        bookdown_0.20    edgeR_3.31.4
## [34] IRanges_2.23.10     matrixStats_0.56.0 XML_3.99-0.5
## [37] reshape_0.8.8       crayon_1.3.4     dplyr_1.0.2
## [40] withr_2.3.0        bitops_1.0-6     grid_4.0.2
## [43] xtable_1.8-4       gtable_0.3.0     lifecycle_0.2.0
## [46] DBI_1.1.0          magrittr_1.5     scales_1.1.1
## [49] bibtex_0.4.2.3     stringi_1.4.6   reshape2_1.4.4
## [52] farver_2.0.3       ellipsis_0.3.1  generics_0.0.2
## [55] vctrs_0.3.4        RColorBrewer_1.1-2 tools_4.0.2
## [58] bit64_4.0.5        Biobase_2.49.1   glue_1.4.2
## [61] purrrr_0.3.4       parallel_4.0.2  survival_3.1-12
## [64] yaml_2.2.1         AnnotationDbi_1.51.3 colorspace_1.4-1
## [67] BiocManager_1.30.10 gRd_0.4-11      memoise_1.1.0
## [70] knitr_1.30
```