DNA Methylation
Practical 2 - Normalisation

Allan McRae
a.mcrae@uq.edu.au

Sonia Shah
sonia.shah@uq.edu.au

Overview

This practical will cover

- Performing a functional normalisation of methylation data

- Visualising quality control metrics

- Estimating blood cell counts from methylation data

Read Saved Data

This practical continues the analysis of the QCed methylation data. Read the data in by:

```
library(meffil)

load("qc.objects.clean.Robj")
length(qc.objects)

load("qcsummary.clean.Robj")
```

Estimating PCs on Control Probes

We estimate the number of PCs that explain variation on the control probes. These are included in subsequent functional normalisation

```
y <- meffil.plot.pc.fit(qc.objects)
ggsave(y$plot,filename="pc.fit.pdf",height=6,width=6)

pc <- ??
```

Functional Normalisation

Jean-Philippe Fortin et al., **Functional normalization of 450k methylation array data improves replication in large cancer studies** *Genome Biology* 2014, 15:503

- Quantile normalisation can remove global differences in methylation data

- Particularly important in case-control studies that have expected widespread methylation changes (e.g. cancer)

- Control probes have no expected differences

- Extend quantile normalisation of control probes to all probes

Functional Normalisation

Perform functional normalisation:

```
norm.objects = meffil.normalize.quantiles(qc.objects,
                                          number.pcs=pc)

save(norm.objects, file=”norm.obj.Robj")


norm.beta = meffil.normalize.samples(norm.objects,
                cpglist.remove=qc.summary$bad.cpgs$name)

save(norm.beta, file=paste("norm.beta.Robj"))
```

Create Normalisation Report

Set parameters for report

```
batch_var<-c("Slide",
             "sentrix_row",
             "sentrix_col",
             "Sex")

norm.parameters <- meffil.normalization.parameters(
    norm.objects,
    variables=batch_var,
    control.pcs=1:10,
    batch.pcs=1:10,
    batch.threshold=0.01
)
```

Create Normalisation Report

Estimate PCs from 20,000 most variable probes and show relationship to batch variables

```
pcs= meffil.methylation.pcs(norm.beta,probe.range=20000)

save(pcs,file="pcs.norm.beta.Robj")


norm.summary= meffil.normalization.summary(norm.objects,
                    pcs=pcs,parameters=norm.parameters)

meffil.normalization.report(norm.summary,
                  output.file="normalization-report.html")
```

Extract Blood Cell Counts

Blood cell counts were estimated during QC. Extract these for use in the next practical

```
cc = t(sapply(qc.objects, function(obj)
                                obj$cell.counts$counts))

cc = data.frame(IID=row.names(cc),cc)

write.table(cc,"cellcounts.txt",sep="\t",
                          row.names=F,col.names=T,quote=F)
```