TNFa GSE64233

The file length\_bias\_analysis.R , the counts file and the Tx.len.file should be located in Counts\_data folder within the working directory. The output file is cnts\_data\_Pgenes.txt

## Step 1 : Filter genes

Create filtered counts file with genes that have at least 1.0 cpm in all replicates of at least one biological condition in the dataset, and have a length and GC content data. The output is cnts\_data\_Pgenes file, this file is being use as a counts file input for the next steps.

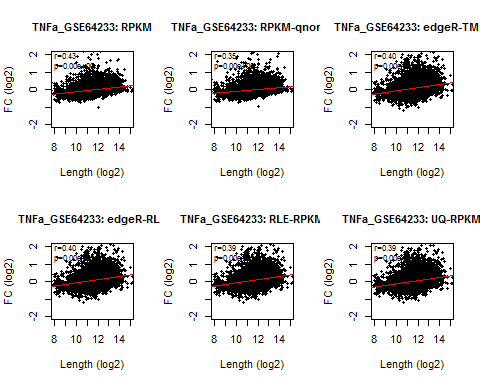
N.replicates <- 3  
Conds <- list("control"=1:N.replicates, "treatment"=(N.replicates+1):(N.replicates\*2))  
Tx.LenFile <- "Counts\_data/Hs\_Ens\_v96\_APLen\_forPCgenes\_MaxTxLenForRest\_plus\_GCcontent.txt";  
countsFile <- "Counts\_data/GSE64233\_Counts.txt"  
DatasetTitle <- "TNFa\_GSE64233"  
identify\_genes\_with\_leninfo\_and\_cpm\_above\_thresh (countsFile, Conds,N.samples.cutoff=N.replicates,  
 CPM.CUTOFF=1.0, Tx.LenFile,  
 DatasetTitle=DatasetTitle)

##   
## Num of genes passing filter: 12135   
##   
##   
## Writing Pgenes count file to TNFa\_GSE64233\_cnts\_data\_Pgenes.txt  
##   
##

## Step 2 : compare biological conditions

The count file is the filtered counts file from the previous step (cnts\_data\_Pgenes file) and it need to be located in the wd. The output includes 2 files: treatment\_vs\_control\_analysis\_results.txt biological\_conditions\_analysis\_plots.pdf

countsFile.Pgenes <- ("TNFa\_GSE64233\_cnts\_data\_Pgenes.txt")  
length\_bias\_analysis\_compare\_biological\_conditions(countsFile.Pgenes, Tx.LenFile,  
 DatasetTitle=DatasetTitle,  
 X.lim=c(8, 15), Y.lim=c(-2,2))



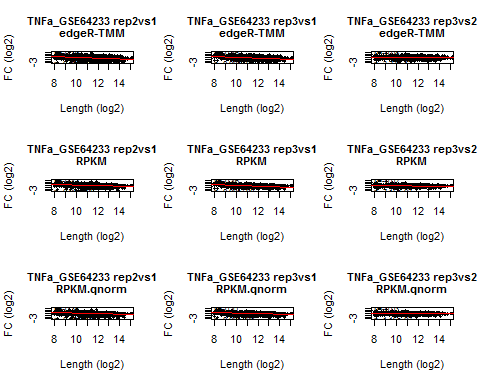
##   
## Results recorded in TNFa\_GSE64233\_treatment\_vs\_control\_analysis\_results.txt  
##

## RPKM RPKM.qnorm edgeR.TMM edgeR.RLE RLE.RPKM UQ.RPKM  
## Spearman.r 0.4255115 0.3532551 0.3958778 0.397017 0.394832 0.3920928  
## pval 0.0000000 0.0000000 0.0000000 0.000000 0.000000 0.0000000

## Step 3 : compare replicate samples

The count file is the filtered counts file from the step 1 (cnts\_data\_Pgenes file) and it need to be located in the wd. The analysis is on the treatment replicates. For control replicates set Rep.first.col to 2. The output includes 2 files: replicate\_samples\_analysis\_results.txt replicate\_samples\_analysis\_plots.pdf

Rep.first.col <- N.replicates + 2  
length\_bias\_analysis\_compare\_replicate\_samples(countsFile.Pgenes, Tx.LenFile,N.replicates=N.replicates,  
 DatasetTitle=DatasetTitle, Rep.first.col=Rep.first.col,  
 X.lim=c(8, 15), Y.lim=c(-3,3))



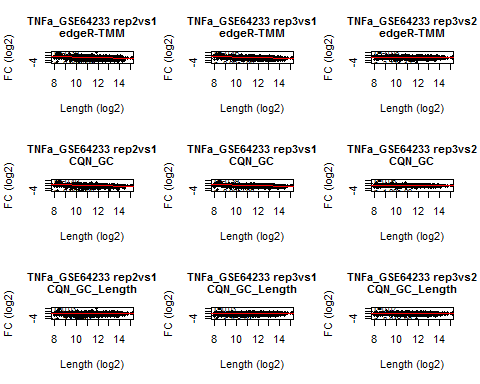
##   
## Results recorded in TNFa\_GSE64233\_replicate\_samples\_analysis\_results.txt  
##

## r.rep2vs1 r.rep3vs1 r.rep3vs2 p.rep2vs1 p.rep3vs1  
## edgeR.TMM -0.3925942 -0.4612455 -0.07632911 0 0  
## RPKM -0.4172103 -0.4614282 -0.02441696 0 0  
## RPKM.qnorm -0.3544737 -0.4050252 -0.02947935 0 0  
## RLE.RPKM -0.4010907 -0.4694072 -0.09334458 0 0  
## UQ.RPKM -0.4083818 -0.4714776 -0.08207404 0 0  
## p.rep3vs2  
## edgeR.TMM 3.772723e-17  
## RPKM 7.147957e-03  
## RPKM.qnorm 1.163065e-03  
## RLE.RPKM 6.745987e-25  
## UQ.RPKM 1.358926e-19

## Step 4 : compare biological and replicate samples after cqn normalization

The count file is the filtered counts file from the step 1 (cnts\_data\_Pgenes file) and it need to be located in the wd. The analysis is on the treatment replicates. For control replicates set Rep.first.col to 2. The output includes 4 files: cqn\_replicate\_samples\_analysis\_results.txt CQN\_replicate\_samples\_analysis\_plots.pdf cqn\_treatment\_vs\_control\_analysis\_results.txt CQN\_biological\_conditions\_analysis\_plots.pdf

length\_bias\_analysis\_cqn\_normalization (countsFile.Pgenes,Tx.LenFile,N.replicates=N.replicates,  
 DatasetTitle=DatasetTitle, Rep.first.col=Rep.first.col,  
 X.lim=c(8, 15), Y.lim=c(-4,4))



##   
## Results recorded in TNFa\_GSE64233\_cqn\_replicate\_samples\_analysis\_results.txt  
##

##   
## Results recorded in TNFa\_GSE64233\_cqn\_treatment\_vs\_control\_analysis\_results.txt  
##   
## cqn treatment vs control analysis results:  
## edgeR.TMM CQN\_GC CQN\_GC\_Length  
## Spearman.r 0.3958778 0.3595877 0.01506222  
## pval 0.0000000 0.0000000 0.09708315  
##   
## cqn replicates results:  
## r.rep2vs1 r.rep3vs1 r.rep3vs2 p.rep2vs1 p.rep3vs1  
## edgeR.TMM -0.39259422 -0.4612455 -0.07632911 0.0000000000 0.000000000  
## CQN\_GC -0.40241559 -0.3892724 0.05541850 0.0000000000 0.000000000  
## CQN\_GC\_Length 0.03060401 0.0340400 0.01587176 0.0007469242 0.000176474  
## p.rep3vs2  
## edgeR.TMM 3.772723e-17  
## CQN\_GC 1.003132e-09  
## CQN\_GC\_Length 8.040399e-02

