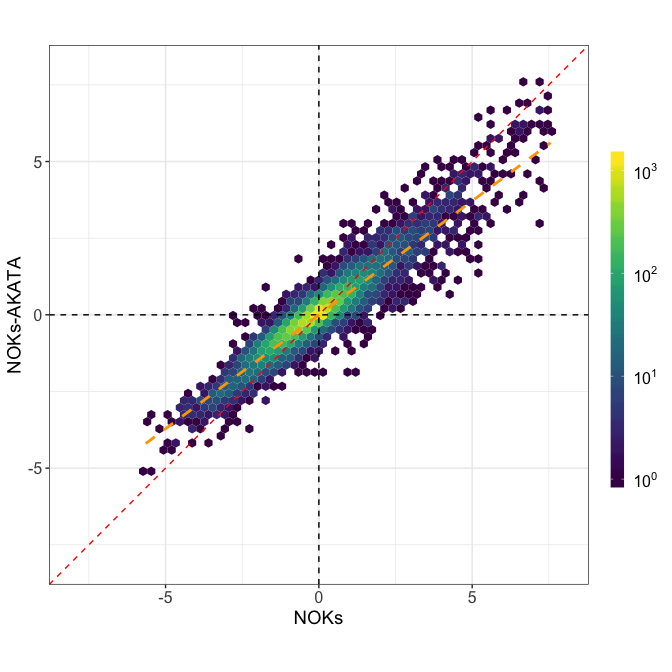
Fold change analysis

**Scott Fold change:** The idea of this figure is two fold. a) Compare the MC over NT fold change between NOKS and EBV cell lines. b) Quantify the ratio between cell fold changes.

For b), we considered the model, where and are the for the NOKs, and NOKs-AKATA samples respectively. Below, we calculated the test, representing that



update\_geom\_defaults("smooth", list(size = .5))  
  
fc\_model = lm( EBV ~ 0 + NOKS , data = scott\_fc)  
  
multcomp::glht(model = fc\_model, linfct = c("NOKS = 1")) %>%   
 summary()

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Fit: lm(formula = EBV ~ 0 + NOKS, data = scott\_fc)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## NOKS == 1 0.750509 0.002312 -107.9 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

**Refereee’s comment**

“- Plotting a histogram of the residuals from Figure 6 as a new panel would show if there is anyone gene that is genuinely affected by EBV. If not, then the result that”EBV infection attenuates the differentiation response globally " (line 287) is not very exciting."

**Mark’s idea**

I suspect that the reviewer is not happy that we cannot provide a short list of genes that are the most significantly different between NOKs and NOKs-Akata. My thought is that I can create a histogram of the ratios of gene changes, and then from that create a list of the “top 100” genes, or genes with a ratio above a specific threshold.

**Response:**

I agree with Mark that the referee is not happy that we didn’t provided a list of the genes that are differentially expressed. However, I feel that both the referee’s and Mark’s ideas are a bit lackcluster because those are not proper method to rank genes that are differentially expressed.

When we were working in figuring out to which MC differentiation cell line, the drdz differentiation look like, I provided an analysis that tested for interaction effects using DESeq2, and that test was equivalent to determining the genes such that the fold change look the most different between cell lines.

ratio\_of\_ratios\_deseq <- function(rsem\_data,thr = 20)  
{  
 count\_matrix = rsem\_data %>%   
 dplyr::select(file,rsem) %>%   
 unnest() %>%   
 dplyr::select(file,gene\_id,expected\_count) %>%   
 mutate(  
 expected\_count = floor(expected\_count)  
 ) %>%   
 spread(file,expected\_count) %>%   
 as\_matrix()  
  
 coldata = rsem\_data %>%   
 dplyr::select(file,cell,treatment) %>%   
 mutate(interac = paste(cell,treatment, sep = ".")) %>%   
 as.data.frame() %>%   
 tibble::remove\_rownames() %>%   
 tibble::column\_to\_rownames("file")  
  
 deseq = DESeqDataSetFromMatrix(  
 count\_matrix,colData = coldata,  
 design = ~ cell + treatment + cell:treatment)  
  
 deseq = deseq[ rowSums(assay(deseq) ) > thr,]  
 deseq = DESeq(deseq, test = "LRT", reduced = ~ cell + treatment)  
  
 deseq  
   
}  
  
fc\_diff\_genes = ratio\_of\_ratios\_deseq(scott\_data$rsem\_data,20)  
gene\_list = results(fc\_diff\_genes,cooksCutoff = FALSE,tidy = TRUE) %>%   
 as\_tibble() %>% rename(gene\_id = row) %>%   
 select(gene\_id,pvalue,padj) %>%   
 mutate( padj = if\_else(is.na(padj),1,padj)) %>%   
 inner\_join(scott\_fc,by = "gene\_id") %>%   
 separate(gene\_id,into = c("ensembl","symbol"),remove = FALSE) %>%   
 mutate(  
 reg = case\_when(EBV > 0 & NOKS > 0 ~ "Upregulated",  
 EBV < 0 & NOKS < 0 ~ "Downregulated",  
 TRUE ~ "Mixed")  
 )

