# Heatmap Trials

*Nathan Siemers* *Genetically Defined Diseases and Genomics*

*2016-06-28*

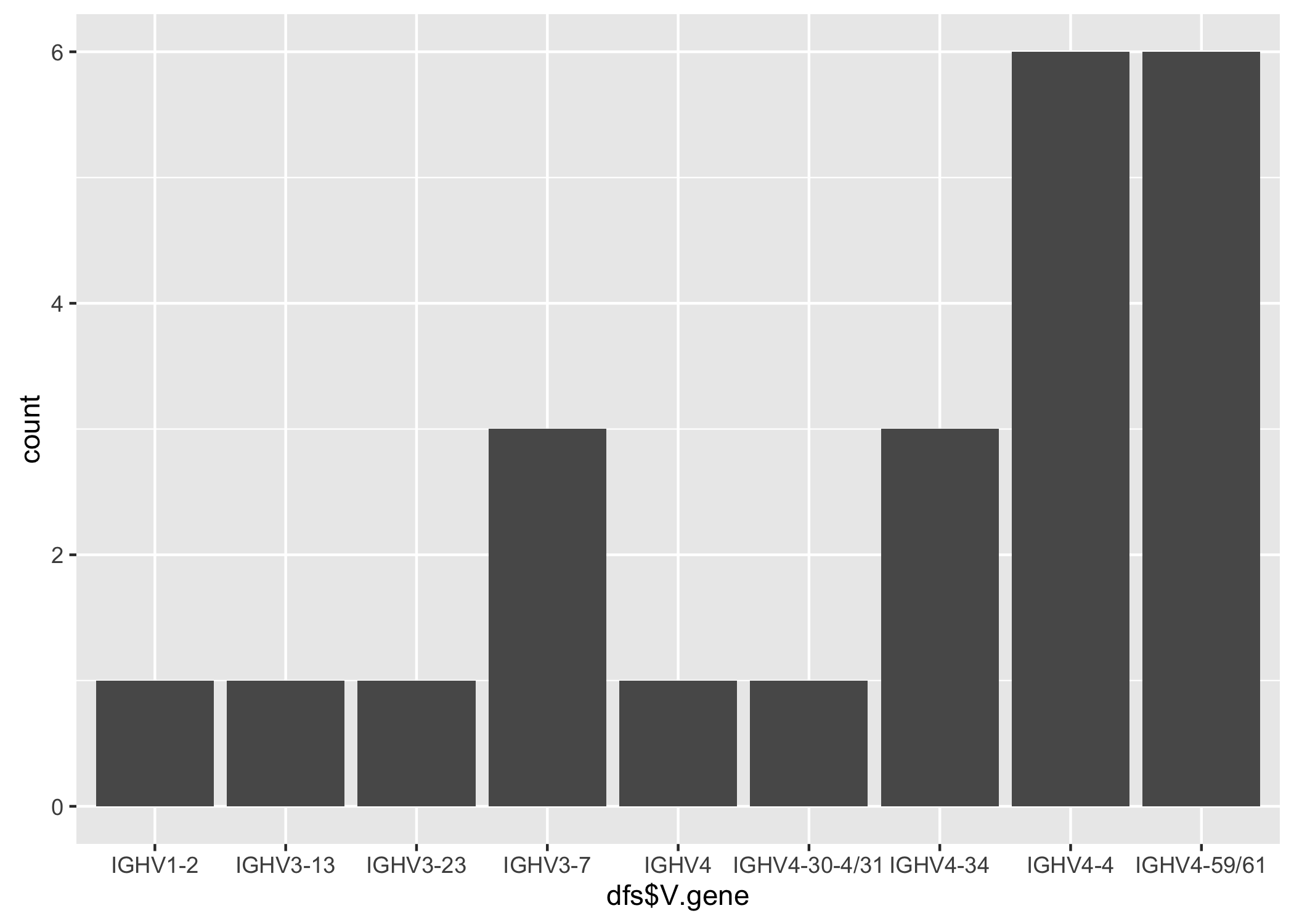
filename = 'heatmap-trials'  
library('plyr')  
library(gplots)  
library(knitr)  
library(knitit)  
library(pheatmap)  
library(Biostrings)  
library(ggplot2)  
  
  
  
  
  
## example code to get antibody data from abgenesis.rdcloud.bms.com - you would need account  
## don't need this as I have sample data in the repo  
do.download = FALSE  
vh.unix = '\*VH-clone-table.xls'  
if ( do.download ) {  
 my.project = 25  
 my.experiment = 189  
 abgenesis.home = 'siemersn@abgenesis.rdcloud.bms.com:/data/abgenesis/data/2'  
 # download all the files  
 vh.files = paste(abgenesis.home, my.project, my.experiment, vh.unix, sep='/')  
 vh.call = paste0('rsync -v --progress ', '\'', vh.files, '\'', ' . ')  
 system(vh.call)  
}  
  
  
  
## remove mouse V regions from tables - given column names this returns non mouse columns  
only.human.v = function(x) {  
 x = x[! grepl("m$", names(x))]  
 x = x[! grepl("^m", names(x))]  
 order(x)  
 }  
  
  
## read test data  
df = read.csv('trial.data.xls.gz', sep="\t", stringsAsFactors = FALSE)  
## for testing, remove duplicates  
  
dim(df)

[1] 499999 49

dfs = df [ ! duplicated(paste(  
 df$V.gene,  
 df$J.gene,  
 df$CDR3  
 ) ) , ]  
dim(dfs)

[1] 23 49

## tiny now  
  
  
  
  
## simple qc  
qplot(dfs$V.gene)



which(is.na(dfs$V.gene)) # integer(0) is good

integer(0)

length(unique(dfs$V.gene)) # number of distinct v genes in library

[1] 9

colnames(dfs)

[1] "Clone" "Lineage" "Freq"   
 [4] "X." "V.gene" "D.gene"   
 [7] "J.gene" "germ" "CDR1"   
[10] "CDR2" "CDR3" "ACC"   
[13] "ORIG.HEADER" "Source" "VH\_ALLELES"   
[16] "VH\_BASE" "JH\_SEGMENT" "VL\_ALLELES"   
[19] "VL\_VBASE" "JL\_SEGMENT" "liabilities"   
[22] "DNA.liabilities" "VHAA" "VLAA"   
[25] "JHAA" "JLAA" "H1\_CANONICAL"   
[28] "H2\_CANONICAL" "H3\_CANONICAL" "L1\_CANONICAL"   
[31] "L2\_CANONICAL" "L3\_CANONICAL" "VH\_Kabat\_FW1"   
[34] "VH\_Kabat\_CDR1" "VH\_Kabat\_FW2" "VH\_Kabat\_CDR2"   
[37] "VH\_Kabat\_FW3" "VH\_Kabat\_CDR3" "VH\_Kabat\_FW4"   
[40] "VL\_Kabat\_FW1" "VL\_Kabat\_CDR1" "VL\_Kabat\_FW2"   
[43] "VL\_Kabat\_CDR2" "VL\_Kabat\_FW3" "VL\_Kabat\_CDR3"   
[46] "VL\_Kabat\_FW4" "AA" "DNA"   
[49] "VH"

# Heatmap of clones

abheatmap = function(abslice) {  
 annot.col = data.frame(  
 V = abslice$V.gene,  
 D = abslice$D.gene,  
 J = abslice$J.gene,  
 L = abslice$liabilities,  
 Freq = abslice$Freq  
 )  
 rownames(annot.col) = 1:dim(abslice)[[1]]  
 mydist = stringDist(abslice$CDR3)  
 pheatmap(mydist, fontsize = 6, annotation\_row = annot.col)  
}  
  
abheatmap(dfs)

