Analyzing somatic mutations in RNA-seq data

Yang Hu

2017/5/3

Table of Contents

Abstract 1

A) Reading Annotated vcf, add gene names and annotations 2

B) Analyzing somatic mutations using FPKM 7

## Abstract

It is part of a workflow for analyzing RNA-seq data from Waldenstrom Macroglobulinemia patients, in a cohort called "zhunter," from Harvard University. After Alignment, samtools sort, VarScan and Annotation, an annotated zhunter\_Annotated.eff(data).vcf is generated like below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #CHROM | POS | ID | REF | ALT | QUAL | FILTER | INFO | FORMAT | ZH10\_NWM07\_CTTGTA\_L005\_R1\_001 | ... |
| chr11 | 47376915 | . | G | C | . | PASS | ..EFF=missense\_variant.. | GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:RDF:RDR:ADF:ADR | 0/0:22:11:11:11:0:0%:1E0:39:0:7:4:0:0 |  |

In part A, script will add gene name, label "MISSENSE/NONSENSE" , extract FREQ%, add counts, and filter out SNP like below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #CHROM | POS | gene\_POS | gene | ID | REF | ALT | QUAL | FILTER | INFO | counts |
| :-----: | :------: |  |
| chr11 | 47376915 | SPI1 47376915 | SPI1 | . | G | C | . | PASS | MISSENSE | 9 |

In part B, script will read FPKM results and cluster somatic mutations.

## A) Reading Annotated vcf, add gene names and annotations

#### A-1) Setup enviroment

setwd("/Users/yah2014/Dropbox/Public/Olivier/R/zhunter/Mutation");getwd();list.files() #set\_up\_enviroment

## [1] "/Users/yah2014/Dropbox/Public/Olivier/R/zhunter/Mutation"

## [1] "AnnotatedVcf.csv"   
## [2] "AnnotatedVcf\_COSM.csv"   
## [3] "AnnotatedVcf\_COSM.xls"   
## [4] "AnnotatedVcf\_new.csv"   
## [5] "AnnotatedVcf\_noval.csv"   
## [6] "AnnotatedVcf\_noval.xls"   
## [7] "AnnotatedVcf\_novel.csv"   
## [8] "AnnotatedVcf\_novel.xls"   
## [9] "AnnotatedVcf\_zhunter\_COSM.csv"   
## [10] "AnnotatedVcf\_zhunter\_new.csv"   
## [11] "AnnotatedVcf\_zhunter\_novel.csv"   
## [12] "Final"   
## [13] "Mut\_counts.name2.txt"   
## [14] "SampleList\_zhunter.txt"   
## [15] "SampleList\_zhunter\_name.txt"   
## [16] "WM.bed"   
## [17] "muta.csv"   
## [18] "mutatations.csv"   
## [19] "zhunter\_Annotated.eff.vcf"   
## [20] "zhunter\_Annotated.eff20170416.vcf"  
## [21] "zhunter\_Annotated.eff20170417.vcf"  
## [22] "zhunter\_snpEff\_summary.html"

SampleList\_zhunter <- read.csv("SampleList\_zhunter.txt", header=T)#Get\_SampleList\_zhunter\_Data  
SampleList\_zhunter\_name<- read.csv("SampleList\_zhunter\_name.txt", header=T)  
AnnotatedVcf\_zhunter <- read.csv("zhunter\_Annotated.eff20170416.vcf",sep="\t",header =T) #Get zhunter\_Annotated.eff.vcf  
dim(AnnotatedVcf\_zhunter)

## [1] 1415 84

Generate a new file AnnotatedVcf\_zhunter\_new.csv.

AnnotatedVcf\_zhunter\_new <-AnnotatedVcf\_zhunter[,1:2]

#### A-2) Add gene names

AnnotatedVcf\_zhunter\_new[,"gene\_POS"]<-NA #add a empty column to a dataframe   
colnames(AnnotatedVcf\_zhunter\_new)[1]<-"CHROM" # was #CHORM previously  
for(i in 1:(dim(AnnotatedVcf\_zhunter)[1]-1)){  
 if (AnnotatedVcf\_zhunter\_new[i,"POS"] >= 27022522 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=27108601 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr1")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("ARID1A ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "ARID1A"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 136871919 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=136873813 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr2")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("CXCR4 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "CXCR4"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 38179969 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=38184512 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr3")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("MYD88 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "MYD88"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 2743387 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=2757752 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr4")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TNIP2 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TNIP2"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 78432907 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=78532988 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr4")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("CXCL13 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "CXCL13"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 121613068 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=121844021 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr4")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("PRDM5 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "PRDM5"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 122052564 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=122137782 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr4")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TNIP3 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TNIP3"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 150409504 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=150460645 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr5")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TNIP1 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TNIP1"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 143072604 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=143266338 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr6")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("HIVEP2 ",AnnotatedVcf\_zhunter\_new[i,"POS"]);  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "HIVEP2"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 95947212 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=96081655 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr22")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("IGLL5 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "IGLL5"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 38177969 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=38186512 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr9")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("WNK2 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "WNK2"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 44953899 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=44971759 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr11")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TP53I11 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TP53I11"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 47376409 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=47400127 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr11")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("SPI1 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "SPI1"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 53773979 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=53810226 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr12")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("SP1 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "SP1 "}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 43699412 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=43785354 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr15")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TP53BP1 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TP53BP1"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 7571720 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=7590868 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr17")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("TP53 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "TP53"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 20715727 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=20840434 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr18")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("CABLES1 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "CABLES1"}  
 else if(AnnotatedVcf\_zhunter\_new[i,"POS"] >= 23229960 && AnnotatedVcf\_zhunter\_new[i,"POS"] <=23238013 && AnnotatedVcf\_zhunter\_new[i,"CHROM"]=="chr22")  
 {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0("IGLL5 ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- "IGLL5"}  
 else {AnnotatedVcf\_zhunter\_new[i,"gene\_POS"] <- paste0(i," ",AnnotatedVcf\_zhunter\_new[i,"POS"])  
 AnnotatedVcf\_zhunter\_new[i,"gene"] <- NA}  
}  
AnnotatedVcf\_zhunter<-AnnotatedVcf\_zhunter[!is.na(AnnotatedVcf\_zhunter\_new[,"gene"]),]  
AnnotatedVcf\_zhunter\_new<-AnnotatedVcf\_zhunter\_new[!is.na(AnnotatedVcf\_zhunter\_new[,"gene"]),]  
rownames(AnnotatedVcf\_zhunter\_new)<-AnnotatedVcf\_zhunter\_new[,"gene\_POS"]

#### A-3) Add MISSENSE & NONSSENSE annotations

AnnotatedVcf\_zhunter\_new<-cbind(AnnotatedVcf\_zhunter\_new,AnnotatedVcf\_zhunter[,3:7])  
  
AnnotatedVcf\_zhunter\_new[,"INFO"]<-NA #add a empty column to a dataframe   
AnnotatedVcf\_zhunter\_new[grep("MISSENSE",AnnotatedVcf\_zhunter[,"INFO"]),"INFO"]<-"MISSENSE" #Add MISSENSE INFO  
AnnotatedVcf\_zhunter\_new[grep("NONSENSE",AnnotatedVcf\_zhunter[,"INFO"]),"INFO"]<-"NONSENSE" #Add MISSENSE INFO  
intersect <- intersect(grep("NONSENSE",AnnotatedVcf\_zhunter[,"INFO"]),grep("MISSENSE",AnnotatedVcf\_zhunter[,"INFO"])) #find intersect  
AnnotatedVcf\_zhunter\_new[intersect,"INFO"]<-"MISSENSE & NONSENSE" #Add both

#### A-4) Extract FREQ% data from column 10 to 84

library(stringr)  
AnnotatedVcf\_zhunter\_new[,"counts"]<-NA #add a empty column to a dataframe   
Sample<-AnnotatedVcf\_zhunter[,c(10,11)] #Generate a temp dataframe with correct dimension  
colnames(Sample)<-c("Sample1","Sample2")  
for(i in 1:75){  
 Sample[,"Sample1"]<-str\_split\_fixed(AnnotatedVcf\_zhunter[,i+9], ":", 14)[,7] #Split column and only extract FREQ%  
 Sample[,"Sample1"]<-as.numeric(sub("%","",Sample[,"Sample1"])) #convert character of percentage into numeric  
 Sample[,"Sample1"][is.na(Sample[,"Sample1"])]<- 0 #replace NA values with zeros  
 AnnotatedVcf\_zhunter\_new <- cbind(AnnotatedVcf\_zhunter\_new, Sample[,"Sample1"])   
 colnames(AnnotatedVcf\_zhunter\_new)[i+11] <-as.character(SampleList\_zhunter\_name[i,]) #rename columns  
}

#### A-5) Add counts

Counts are the total occurrence in 75 samples for one specific mutation.

counting <-function(x){#input is [1,85] vector  
 c = 0   
 for(i in 1:(ncol(x)-11)){ #"Counts" is at column 11  
 if(x[1,i+11]>0) # total occurrence of >0%  
 c<-c+1  
 }  
 return(c)  
 }  
for(i in 1:nrow(AnnotatedVcf\_zhunter\_new)){  
 AnnotatedVcf\_zhunter\_new[i,"counts"]<-counting(AnnotatedVcf\_zhunter\_new[i,]) #  
}

#### A-6) Filter out regular SNP

Generate file AnnotatedVcf\_zhunter\_COSM.csv contains COSM mutations only. Generate file AnnotatedVcf\_zhunter\_novel.csv contains "COSM || MISSENSE || NONSENSE" mutations only.

AnnotatedVcf\_zhunter\_COSM <- AnnotatedVcf\_zhunter\_new[grep("COSM",AnnotatedVcf\_zhunter\_new[,"ID"]),]  
AnnotatedVcf\_zhunter\_novel <- AnnotatedVcf\_zhunter\_new[AnnotatedVcf\_zhunter\_new[,"ID"] %in% ".",]  
AnnotatedVcf\_zhunter\_novel<- AnnotatedVcf\_zhunter\_novel[c(grep("MISSENSE",AnnotatedVcf\_zhunter\_novel[,"INFO"]),  
 grep("NONSENSE",AnnotatedVcf\_zhunter\_novel[,"INFO"])),]  
AnnotatedVcf\_zhunter\_novel<- rbind(AnnotatedVcf\_zhunter\_COSM,AnnotatedVcf\_zhunter\_novel)  
AnnotatedVcf\_zhunter\_novel<- AnnotatedVcf\_zhunter\_novel[order(rownames(AnnotatedVcf\_zhunter\_novel)),]  
AnnotatedVcf\_zhunter\_novel<- AnnotatedVcf\_zhunter\_novel[unique(AnnotatedVcf\_zhunter\_novel[,"gene\_POS"]),]

#### A-6) Export csv file (optional)

write.csv(AnnotatedVcf\_zhunter\_new,"AnnotatedVcf\_zhunter\_new.csv")  
write.csv(AnnotatedVcf\_zhunter\_COSM,"AnnotatedVcf\_zhunter\_COSM.csv")  
write.csv(AnnotatedVcf\_zhunter\_novel,"AnnotatedVcf\_zhunter\_novel.csv")

## B) Analyzing somatic mutations using FPKM

#### B-1) Setup enviroment and Get FPKM\_zhunter Data

Read each genes.FPKM\_tracking data into the file FPKM\_zhunter\_temp.csv. Generate file FPKM\_zhunter contains FPKM data from 1st sample only.

FPKM\_zhunter\_files\_path <-paste0("/Users/yah2014/Documents/Programs/R/FPKM/", #Create FPKM\_zhunter\_files\_path file contains folder names  
 SampleList\_zhunter$Sample\_ID,"\_CuffLinks/genes.FPKM\_tracking")   
FPKM\_zhunter <-data.frame(x= str(0), y= integer(0)) #Generate empty FPKM\_zhunter dataframe

## num 0

FPKM\_zhunter\_temp <- read.csv(FPKM\_zhunter\_files\_path[1], header=T, sep="\t") #Read data from 1st sample  
FPKM\_zhunter\_temp <- FPKM\_zhunter\_temp[order(FPKM\_zhunter\_temp$tracking\_id),] #Reorder the gene name  
FPKM\_zhunter<-FPKM\_zhunter\_temp$FPKM #Fill up FPKM\_zhunter dataframe with tracking id

Generate file FPKM\_zhunter contains FPKM data from all samples. This might take less than one minute.

for(i in 2:nrow(SampleList\_zhunter)){ #skip the first one, which is added already  
 FPKM\_zhunter\_temp <- read.csv(FPKM\_zhunter\_files\_path[i], header=T, sep="\t")  
 FPKM\_zhunter\_temp <- FPKM\_zhunter\_temp[order(FPKM\_zhunter\_temp$tracking\_id),] #Reorder the gene name  
 FPKM\_zhunter<-cbind(FPKM\_zhunter,FPKM\_zhunter\_temp$FPKM)   
}

Rename rows and columns ofFPKM\_zhunter

dim(FPKM\_zhunter)

## [1] 25286 75

colnames(FPKM\_zhunter) <-SampleList\_zhunter\_name$Sample\_ID # or annotation$SimpleLabel  
rownames(FPKM\_zhunter) <-FPKM\_zhunter\_temp$tracking\_id  
  
FPKM\_name<-data.frame(x= str(0), y= integer(0))

## num 0

FPKM\_name<-as.character(FPKM\_zhunter\_temp$tracking\_id)  
FPKM\_name<-cbind(FPKM\_name,as.character(FPKM\_zhunter\_temp$tracking\_id))

#### B-2) Charaterize somatic mutations

Setup envirment

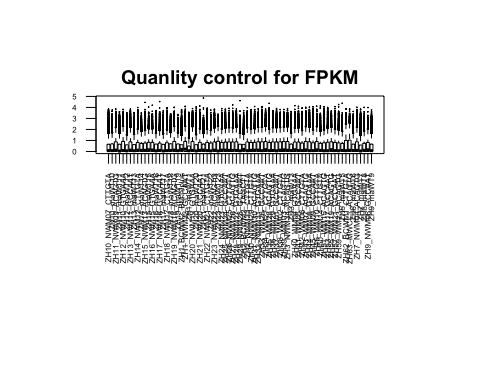
setwd("/Users/yah2014/Dropbox/Public/Olivier/R/zhunter/Mutation");getwd();list.files()

## [1] "/Users/yah2014/Dropbox/Public/Olivier/R/zhunter/Mutation"

## [1] "AnnotatedVcf.csv"   
## [2] "AnnotatedVcf\_COSM.csv"   
## [3] "AnnotatedVcf\_COSM.xls"   
## [4] "AnnotatedVcf\_new.csv"   
## [5] "AnnotatedVcf\_noval.csv"   
## [6] "AnnotatedVcf\_noval.xls"   
## [7] "AnnotatedVcf\_novel.csv"   
## [8] "AnnotatedVcf\_novel.xls"   
## [9] "AnnotatedVcf\_zhunter\_COSM.csv"   
## [10] "AnnotatedVcf\_zhunter\_new.csv"   
## [11] "AnnotatedVcf\_zhunter\_novel.csv"   
## [12] "Final"   
## [13] "Mut\_counts.name2.txt"   
## [14] "SampleList\_zhunter.txt"   
## [15] "SampleList\_zhunter\_name.txt"   
## [16] "WM.bed"   
## [17] "muta.csv"   
## [18] "mutatations.csv"   
## [19] "zhunter\_Annotated.eff.vcf"   
## [20] "zhunter\_Annotated.eff20170416.vcf"  
## [21] "zhunter\_Annotated.eff20170417.vcf"  
## [22] "zhunter\_snpEff\_summary.html"

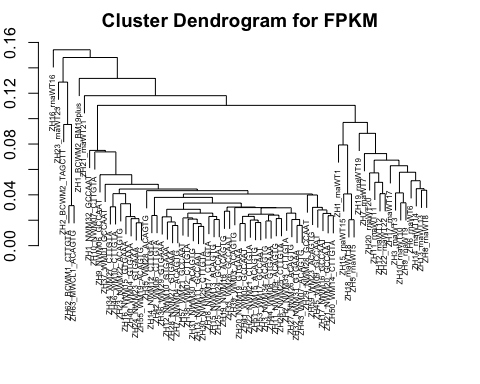
Run quanlity control with boxplot

# set a margin  
par(oma=c(10,3,3,3)) # all sides have 3 lines of space   
par(mar=c(2,2,2,2))  
FPKM\_zhunter\_clean <- FPKM\_zhunter[,!(colnames(FPKM\_zhunter) %in% c("ZH2\_rnaWT2","ZH4\_rnaWT4"))] #"ZH2\_rnaWT2","ZH4\_rnaWT4" have bad quanlity  
boxplot(log10(FPKM\_zhunter\_clean + 1),xlab="all test samples",cex=0.05, cex.axis=0.5, ylab = "log (base 10) RPKM + 1",main="Quanlity control for FPKM",las=2)



Group sample with hclust

par(mar=c(2,2,2,2))   
lm <-log(FPKM\_zhunter\_clean+1)  
dm<-as.dist(1-cor(lm))  
hm<-hclust(dm,method = "average")  
p<-plot(hm,col = "black",cex = 0.5,main="Cluster Dendrogram for FPKM")



Generate heatmap Prepare palette

my\_palette1 <- colorRampPalette(c("antiquewhite2", "green", "blue"))(n = 100)  
my\_palette2 <- colorRampPalette(c("red", "green"))(n = 2)  
my\_palette3 <- colorRampPalette(c("yellow", "orange", "red","green","blue","purple"))(n = 6)

Clean up data

mut\_gene <- AnnotatedVcf\_zhunter\_novel[,!(colnames(AnnotatedVcf\_zhunter\_novel) %in% c("CHROM","POS","gene\_POS","gene","ID","REF","ALT","QUAL","FILTER","INFO","counts"))]   
mut\_gene\_clean <- mut\_gene[,!(colnames(mut\_gene) %in% c("ZH2\_rnaWT2","ZH4\_rnaWT4"))]

Generate heatmap

library(gplots)

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(ComplexHeatmap)

## Loading required package: grid

par(oma=c(3,3,3,3))  
par(mar=c(2,2,2,2)+1)  
  
Heatmap(mut\_gene\_clean,   
 column\_title = "Cluster of RNA expression vs DNA mutation genes in WM",  
 show\_row\_names = TRUE,  
 col=my\_palette1,   
 row\_title\_gp = gpar(fontsize = 14),  
 row\_names\_gp = gpar(fontsize = 2),  
 column\_names\_gp = gpar(fontsize = 6),  
 clustering\_distance\_columns = "euclidean",  
 cluster\_rows = FALSE,  
 cluster\_columns = hm,  
 row\_dend\_width = unit(5, "cm"),  
 column\_dend\_height = unit(30, "mm"),  
 heatmap\_legend\_param = list(title = "mutation rate %")  
)

