Mutation\_Summary

C.J. Liu

12/5/2016

Table of Contents

suppressMessages(library(dplyr))  
suppressMessages(library(ggplot2))  
library(knitr)  
library(biomaRt)

## Somatic Mutation Calling

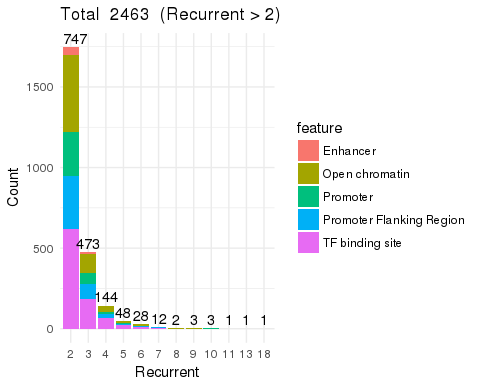
1. Download 20 cases of BRCA WXS bam files from GDC-portal.
2. Download Regulatory Build annotation from Ensembl.
3. Call Regulatory element mutations by GATK (DP > 10, AD > 3 )
4. Exclude dbSNP, avsnp147 (ANNOVAR)
5. Find recurrent tumor and nomal respectively.
6. Somatic tumor.recurrent - normal.recurrent.

## Result

resultHeaders <-c("chrom","pos","ref","alt","ensid","desc","feature","recurrent")  
chromOrder = c("chr1","chr2","chr3","chr4","chr5","chr6","chr7","chr8","chr9","chr10","chr11","chr12","chr13","chr14","chr15","chr16","chr17","chr18","chr19","chr20","chr21","chr22","chrX")  
somatic <- as\_data\_frame(read.table('/home/cliu18/liucj/projects/1.Mutation\_calling\_in\_non-condig\_region\_through\_EXOME/4.compare\_tumor\_normal/somatic.recurrency', header = F, sep = "\t"))  
colnames(somatic) <- resultHeaders  
somatic <- filter(somatic, ensid != "." & recurrent > 1) %>% arrange(-recurrent)  
somatic$chrom <- factor(somatic$chrom, levels = chromOrder)

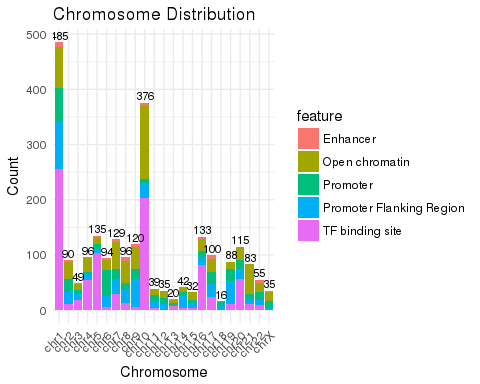
### Recurrent count and feature distribution

ggplot(somatic, aes(x = factor(recurrent))) + geom\_bar(aes(fill = feature)) + geom\_text(stat = 'count', aes(label = ..count..), vjust = -0.3) + theme\_minimal() + xlab("Recurrent") + ylab("Count") + ggtitle(paste("Total ",nrow(somatic)," (Recurrent > 2)"))



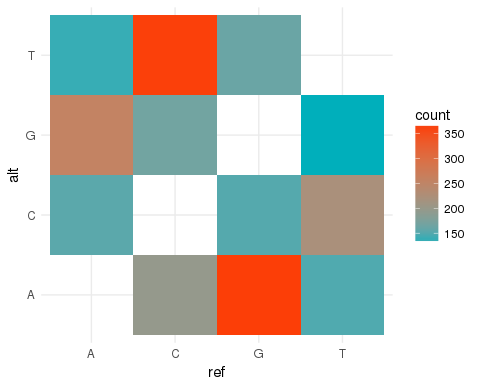
### Feature distribution on the chromosome

ggplot(somatic, aes(x = chrom)) + geom\_bar(aes(fill = feature)) + geom\_text(stat = 'count', aes(label = ..count..), vjust = -0.3, size = 3) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + xlab("Chromosome") + ylab("Count") + ggtitle("Chromosome Distribution")



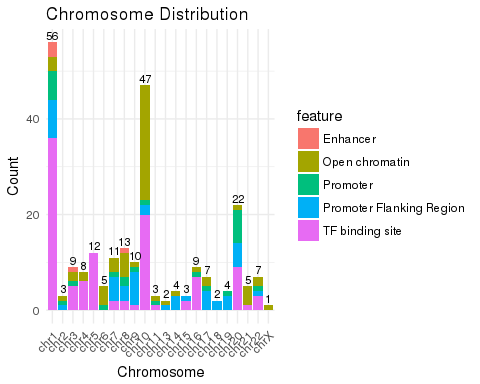
### Mutation type

ggplot(somatic, aes(ref,alt)) + geom\_bin2d() + theme\_minimal() + scale\_fill\_gradient(low = "#00AFBB", high = "#FC3E07")



### Filter recurrent larger than or equal to 5

somatic5 <- filter(somatic, recurrent >= 4)  
# ggplot(somatic5, aes(x = ensid)) + geom\_bar(aes(fill = feature)) + geom\_text(stat = 'count', aes(label = ..count..), vjust = -0.3) + theme\_minimal() + xlab("Recurrent") + ylab("Count") + ggtitle(paste("Total ",nrow(somatic5)," (Recurrent > 2)")) + theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
  
ggplot(somatic5, aes(x = chrom)) + geom\_bar(aes(fill = feature)) + geom\_text(stat = 'count', aes(label = ..count..), vjust = -0.3, size = 3) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + xlab("Chromosome") + ylab("Count") + ggtitle("Chromosome Distribution")



ensCount <- somatic5 %>% group\_by(ensid) %>% summarise(count = n()) %>% arrange(-count)  
#kable(ensCount)