Statistics of BRCA/KIRC test data result

C.J. Liu

11/16/2016

Table of Contents

# Statistics of BRCA/KIRC 20 test data result

## Load data

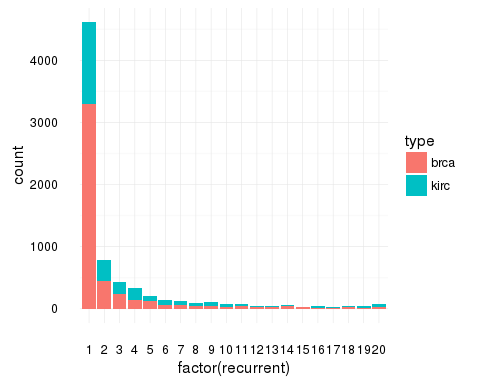
suppressMessages(library(dplyr))  
suppressMessages(library(ggplot2))  
#project/3.calling/BRCA\_WXS\_20/recurrency/SNP  
#project/3.calling/KIRC\_WXS\_20/recurrency/SNP  
#read recurrency.sortByChrom to brca and kirc  
brca <- as\_data\_frame(read.table('brca.recurrency.sortByChrom', header = F, sep = "\t"))  
kirc <- as\_data\_frame(read.table('kirc.recurrency.sortByChrom', header = F, sep = "\t"))  
brca <- mutate(brca, type = 'brca')  
kirc <- mutate(kirc, type = 'kirc')  
tcga <- bind\_rows(brca, kirc, .id = NULL)

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

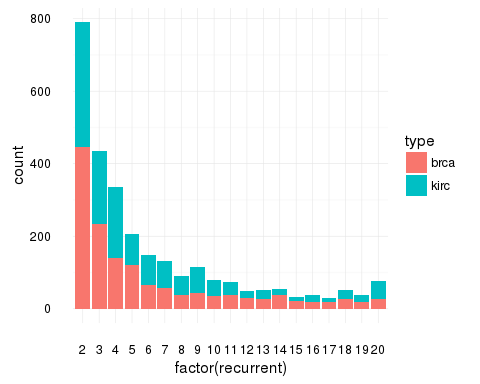
header <-c("chrom","pos","ref","alt","ensid","desc","feature","recurrent", "type")  
colnames(tcga) <- header  
  
tcga.filterEns <- filter(tcga, ensid != ".")  
  
tcga.filterEns.recurrent2 <- filter(tcga.filterEns, recurrent > 1)  
chromOrder = c("chr1","chr2","chr3","chr4","chr5","chr6","chr7","chr8","chr9","chr10","chr11","chr12","chr13","chr14","chr15","chr16","chr17","chr18","chr19","chr20","chr21","chr22","chrX")  
# reorder the chrom levels  
tcga.filterEns.recurrent2$chrom <- factor(tcga.filterEns.recurrent2$chrom, levels = chromOrder)

## Basic statistics

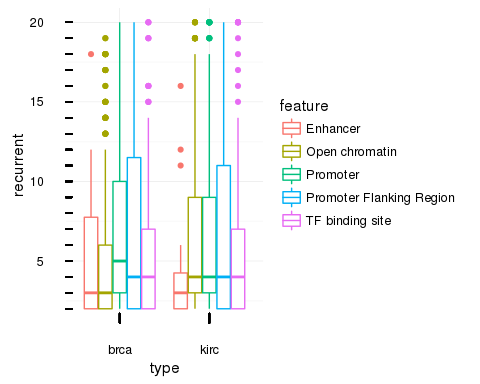
# all result  
ggplot(tcga.filterEns, aes(x = factor(recurrent))) + geom\_bar(aes(fill = type)) + theme\_minimal()



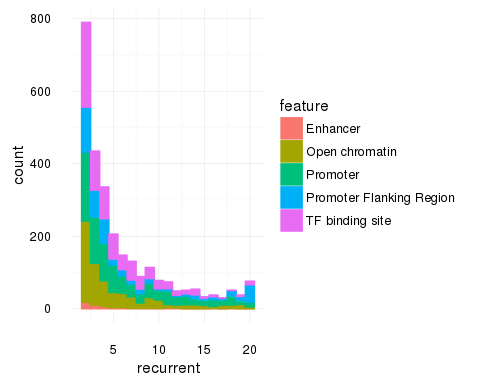
# following analyis based on the recurrent > 2  
  
# filter 1 recurrent  
ggplot(tcga.filterEns.recurrent2, aes(x = factor(recurrent))) + geom\_bar(aes(fill = type)) + theme\_minimal()



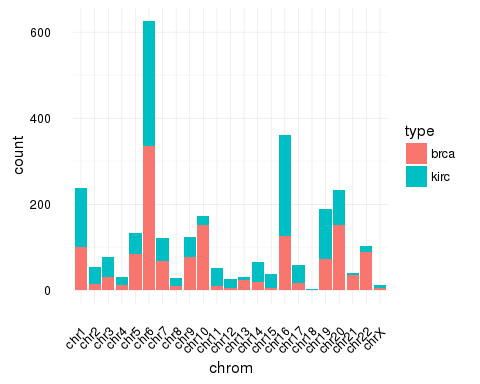
# feature distribution  
ggplot(tcga.filterEns.recurrent2, aes(type, recurrent)) + geom\_boxplot(aes(color = feature), position = position\_dodge(0.8)) + theme\_minimal() + geom\_rug() #+ geom\_jitter(aes(color = feature),position = position\_dodge(0.8))



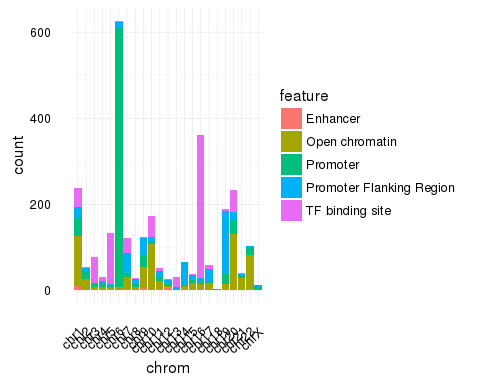
# recurrent distribution  
ggplot(tcga.filterEns.recurrent2, aes(x = recurrent)) + geom\_histogram(aes(color = feature, fill = feature), bins = 19) + theme\_minimal()



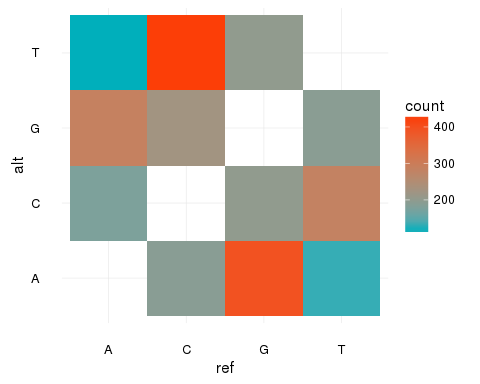
# chrom distribution  
ggplot(tcga.filterEns.recurrent2, aes(x = chrom)) + geom\_bar(aes(fill = type)) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# chrom feature distribution  
ggplot(tcga.filterEns.recurrent2, aes(x = chrom)) + geom\_bar(aes(fill = feature)) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# mutation  
# ggplot(tcga.filterEns.recurrent2, aes(ref, alt)) + geom\_jitter(aes(color = ref), size = 0.5) + theme\_minimal()  
  
ggplot(tcga.filterEns.recurrent2, aes(ref,alt)) + geom\_bin2d() + theme\_minimal() + scale\_fill\_gradient(low = "#00AFBB", high = "#FC3E07")



# ggplot(tcga.filterEns.recurrent2, aes(x = feature)) + geom\_bar(aes(color = alt, fill = alt)) + theme\_minimal()  
  
# mutations in ens region  
# ggplot(tcga.filterEns.recurrent2, aes(x = ensid)) + geom\_bar()  
tcga.filterEns.recurrent2.ensid.top10 <- head(sort(table(tcga.filterEns.recurrent2$ensid), decreasing = T),10)  
library(knitr)  
kable(as.data.frame(tcga.filterEns.recurrent2.ensid.top10))

|  |  |
| --- | --- |
|  | tcga.filterEns.recurrent2.ensid.top10 |
| ENSR00000195346 | 240 |
| ENSR00000195701 | 182 |
| ENSR00000195522 | 170 |
| ENSR00000085420 | 135 |
| ENSR00000180289 | 117 |
| ENSR00000085424 | 70 |
| ENSR00000085419 | 67 |
| ENSR00000154776 | 59 |
| ENSR00000143419 | 51 |
| ENSR00000111531 | 44 |