

RESULTS MSC

Jupiter

```
variants <- read.delim(file="/media/jupiter/marina/project-msbt/final.txt", header = TRUE, sep = "\t", s
head(variants)
```

```
##  GENE VariantTYPE DNAchange  AAchange      ANNOTATION Anno.Impact Sample_ID
## 1 atpE          snp   c.172G>A p.Gly58Ser missense_variant  MODERATE      NA
## 2 atpE          snp   c.17C>T  p.Ala6Val  missense_variant  MODERATE      NA
## 3 atpE          snp   c.221T>C p.Phe74Ser missense_variant  MODERATE      NA
## 4 atpE          snp   c.40G>T  p.Gly14*   stop_gained      HIGH          NA
## 5 atpE          snp   c.139G>T p.Gly47Trp missense_variant  MODERATE      NA
## 6 atpE          snp   c.159C>A p.Phe53Leu missense_variant  MODERATE      NA
```

```
id <- read.delim(file="/media/jupiter/marina/project-msbt/sample_id", header = FALSE, col.names = "Sample_ID")
head(id)
```

```
##  Sample_ID
## 1  TN106495
## 2  TN106507
## 3  TN106508
## 4  TN106566
## 5  TN106566
## 6  TN106566
```

```
nrow(variants)==nrow(id)
```

```
## [1] TRUE
```

```
variants$Sample_ID <- id$Sample_ID
head(variants)
```

```
##  GENE VariantTYPE DNAchange  AAchange      ANNOTATION Anno.Impact Sample_ID
## 1 atpE          snp   c.172G>A p.Gly58Ser missense_variant  MODERATE  TN106495
## 2 atpE          snp   c.17C>T  p.Ala6Val  missense_variant  MODERATE  TN106507
## 3 atpE          snp   c.221T>C p.Phe74Ser missense_variant  MODERATE  TN106508
## 4 atpE          snp   c.40G>T  p.Gly14*   stop_gained      HIGH      TN106566
## 5 atpE          snp   c.139G>T p.Gly47Trp missense_variant  MODERATE  TN106566
## 6 atpE          snp   c.159C>A p.Phe53Leu missense_variant  MODERATE  TN106566
```

```
unq <- levels(id$Sample_ID) # samples found with variant
length(unq) # total number of samples with variants
```

```
## [1] 0
```

```
syn <- variants[variants$ANNOTATION != "synonymous_variant", ] # filters out the synonymous mutations
head(syn)
```

```
##      GENE VariantTYPE DNAchange  AAchange      ANNOTATION Anno.Impact Sample_ID
## 1 atpE      snp    c.172G>A p.Gly58Ser missense_variant    MODERATE  TN106495
## 2 atpE      snp    c.17C>T  p.Ala6Val  missense_variant    MODERATE  TN106507
## 3 atpE      snp    c.221T>C p.Phe74Ser missense_variant    MODERATE  TN106508
## 4 atpE      snp    c.40G>T   p.Gly14*   stop_gained        HIGH    TN106566
## 5 atpE      snp    c.139G>T p.Gly47Trp missense_variant    MODERATE  TN106566
## 6 atpE      snp    c.159C>A p.Phe53Leu missense_variant    MODERATE  TN106566
```

```
variants <- syn[syn$ANNOTATION!="intragenic_variant", ] # filters out gene 223 ??
nrow(variants)
```

```
## [1] 995
```

```
# types of variants identified
```

A total of 0 have been found to have variants in one or more genes

A total of 995 have been identified in all genes. These are the raw results without filter

Types of variants identified

These variants
have been identified in the raw unfiltered data
complex,mnp will be filtered out leaving only snp,del,ins.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
levels(variants$VariantTYPE)
```

```
## NULL
```

```
target <- c("snp","del","ins")
filter <- (variants$VariantTYPE %in% target)
variants <- variants[filter, ]
```

```

filter <- which(variants$ANNOTATION=="conservative_inframe_deletion")

filter<- c(filter, which(variants$ANNOTATION=="conservative_inframe_insertion"))

filter<- c(filter, which(variants$ANNOTATION=="splice_region_variant&stop_retained_variant"))
variants <- variants[-filter, ]

nrow(variants)

```

```
## [1] 974
```

```

ind <- seq(nrow(variants))
for (i in ind) {
  if (variants$ANNOTATION[i]=="stop_lost&splice_region_variant"){

    variants$ANNOTATION[i] <- "stop_lost"

  }
}

#length(which(filter))

```

After filtering out complex variants, i remained with 974 variants in all genes

```

atpe <- variants$GENE=="atpE"
atpe2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)

```

```
## [1] 21
```

```

atpe$AAchange <- as.character(atpe$AAchange)
atpe1 <- distinct(atpe,AAchange, .keep_all = TRUE)
table(atpe$AAchange)

```

```

##
## p.Ala21Gly p.Ala62Val p.Ala6Val p.Ala77Asp p.Gly13Arg p.Gly14* p.Gly14Arg
## 1 1 1 1 1 1 1
## p.Gly47Trp p.Gly58Ser p.Gly8Ser p.Phe50Ile p.Phe53Leu p.Phe70Ser p.Phe74Leu
## 1 1 1 1 2 1 1
## p.Phe74Ser p.Pro3His p.Pro79His p.Thr56Ala p.Thr56Ser p.Val75Leu
## 1 1 1 1 1 1

```

```

atpe <- levels(as.factor(atpe$AAchange))
length(atpe)

```

```
## [1] 20
```

There are 21 total variants in the atpE gene. Of these, 20 are unique variants, mutation p.Phe53Leu appeared twice

```
atpe <- variants$GENE=="Rv0678"
Rv06782 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 56
```

```
atpe$AAchange <- as.character(atpe$AAchange)
Rv0678 <- distinct(atpe, AAchange, .keep_all = TRUE)
nrow(atpe)
```

```
## [1] 56
```

```
table(atpe$AAchange)
```

```
##
## p.Ala101Thr p.Ala128Ser p.Ala57Val p.Ala59Glu p.Ala84Pro p.Arg105His
##          1          1          1          1          1          6
## p.Arg107Cys p.Arg107Leu p.Arg107Ser p.Arg34Leu p.Arg94fs p.Arg94Leu
##          1          1          1          1          1          1
## p.Asp116Tyr p.Asp152Tyr p.Asp15Val p.Asp88Tyr p.Asp8Tyr p.Gln115Lys
##          1          1          1          1          1          1
## p.Gln159His p.Glu104Asp p.Glu104Lys p.Glu106Gly p.Glu138Gly p.Glu21*
##          1          1          1          1          1          1
## p.Gly126Cys p.Gly41Val p.Gly65Arg p.Gly66fs p.Gly6Val p.Gly78Trp
##          1          3          1          1          1          1
## p.Gly87Val p.Ile108Phe p.Ile75fs p.Leu114Val p.Leu117Arg p.Leu122Pro
##          1          1          1          1          1          1
## p.Leu83Ile p.Met111Leu p.Met139Ile p.Met23Ile p.Phe100Leu p.Phe93Leu
##          1          1          1          1          1          1
## p.Pro129Gln p.Pro129Thr p.Ser52Tyr p.Ser63Gly p.Ser64Arg
##          3          1          1          1          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 47
```

```
write.table(variants, file="/media/jupiter/marina/project-msbt/final-results.csv", sep= "\t", quote = F)
```

There are 56 total variants in the Rv0678 gene. Of these, 47 are unique variants, mutation p.Arg105His appeared six times while mutations p.Pro129Gln and p.Gly41Val appeared three times

```
atpe <- variants$GENE=="pepQ"
pepQ2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 96
```

```
atpe$AAchange <- as.character(atpe$AAchange)
pepQ <- distinct(atpe, AAchange, .keep_all = TRUE)
nrow(atpe)
```

```
## [1] 96
```

```
table(atpe$AAchange)
```

```
##
## p.Ala141Val p.Ala153fs p.Ala196Val p.Ala198Asp p.Ala243Thr p.Ala254Val
##          1          1          1          1          1          1
## p.Ala268Ser p.Ala47Thr p.Ala65Ser p.Ala90Asp p.Arg123Leu p.Arg167*
##          1          1          1          1          1          1
## p.Arg206Leu p.Arg206Trp p.Arg235Ser p.Arg337Cys p.Arg61His p.Arg61Ser
##          1          1          1          1          1          2
## p.Asp216Asn p.Asp221Asn p.Asp48Tyr p.Asp8Gly p.Gln213His p.Gln246Lys
##          1          1          1          1          1          1
## p.Gln251Glu p.Gln257* p.Gln258His p.Gln262His p.Gln262Lys p.Gln5fs
##          1          1          2          1          1          1
## p.Gln5Lys p.Gln63Lys p.Gln67Lys p.Glu115Asp p.Glu139Val p.Glu191*
##          1          2          1          1          1          1
## p.Glu255fs p.Glu339Gly p.Glu49Asp p.Glu89* p.Gly162Cys p.Gly184Ser
##          1          1          1          1          1          1
## p.Gly197Val p.Gly215Cys p.Gly223Cys p.Gly223Val p.Gly228Trp p.Gly228Val
##          1          1          2          1          1          1
## p.Gly272Val p.Gly58Cys p.Gly93Glu p.His100Asn p.His100Gln p.His183Arg
##          1          1          1          1          1          1
## p.His183Tyr p.Leu107fs p.Leu163Met p.Leu176Pro p.Leu179Met p.Leu179Pro
##          1          1          1          1          1          1
## p.Leu225Met p.Leu252Val p.Leu54Phe p.Leu71Ile p.Lys117Glu p.Lys350Ile
##          1          1          1          1          1          1
## p.Lys94Arg p.Met180Lys p.Met351Leu p.Phe36Leu p.Phe46Leu p.Pro203Gln
##          1          1          1          1          1          1
## p.Pro203Leu p.Pro207Gln p.Pro266Ser p.Pro352Thr p.Pro359Thr p.Ser4Pro
##          2          1          1          1          1          1
## p.Thr168Ala p.Thr234Ser p.Thr236Ala p.Thr56Ala p.Trp245Arg p.Tyr286Cys
##          1          1          1          1          1          1
## p.Val194Met p.Val211Ala p.Val214Ile p.Val273Ala p.Val336Asp p.Val53Ala
##          1          1          1          1          1          1
## p.Val79fs
##          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 91
```

There are 96 total variants in the pepQ gene. Of these, 91 are unique variants

```

atpe <- variants$GENE=="Rv1979c"
Rv1979c2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)

```

```
## [1] 156
```

```

atpe$AAchange <- as.character(atpe$AAchange)
Rv1979c <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)

```

```

##
## p.Ala10Val p.Ala150Pro p.Ala173Ser p.Ala211Thr p.Ala283Ser p.Ala284Thr
##          6          1          1          1          1          1
## p.Ala294Ser p.Ala366Glu p.Ala378Glu p.Ala381Asp p.Ala461Glu p.Ala66Ser
##          1          1          1          1          12         1
## p.Ala68Thr p.Ala68Val p.Arg157Leu p.Arg254Ile p.Arg382Leu p.Arg408Trp
##          1          1          1          1          1          1
## p.Arg473Gln p.Arg473Leu p.Arg477Gln p.Arg70Gly p.Asn191fs p.Asn297Lys
##          1          1          1          1          1          1
## p.Asn445Ser p.Asp124Glu p.Asp129fs p.Asp286Gly p.Asp287Ala p.Asp287Tyr
##          1          1          1          39         1          1
## p.Gln132Arg p.Glu131fs p.Glu131Lys p.Glu38Asp p.Gly107Trp p.Gly155fs
##          1          1          1          2          1          1
## p.Gly15Ser p.Gly264Arg p.Gly29Ser p.Gly307* p.Gly334Val p.Gly41Trp
##          1          1          1          1          2          1
## p.Gly55Asp p.Gly87Trp p.Gly91Cys p.Ile238Phe p.Ile28fs p.Ile452Phe
##          1          1          1          1          1          1
## p.Ile56Val p.Leu117Pro p.Leu139Met p.Leu142Phe p.Leu149Ser p.Leu14Pro
##          1          1          1          1          1          1
## p.Leu153Phe p.Leu309fs p.Leu336Met p.Leu428Arg p.Lys215Asn p.Lys281Asn
##          1          1          1          1          1          1
## p.Lys415* p.Phe122Cys p.Phe33Leu p.Phe353Leu p.Phe363Leu p.Phe46Leu
##          1          1          1          1          1          1
## p.Phe57Ser p.Phe86Leu p.Pro204Gln p.Pro444Thr p.Ser175Trp p.Ser206*
##          1          1          1          1          1          1
## p.Ser262fs p.Ser262Leu p.Ser279* p.Ser333Gly p.Ser431Tyr p.Thr116Arg
##          1          1          1          1          1          1
## p.Thr288Ala p.Thr97Ile p.Thr97Pro p.Trp183Leu p.Tyr432* p.Tyr432Cys
##          1          1          1          1          1          2
## p.Tyr79fs p.Val108fs p.Val179Ile p.Val273fs p.Val389fs p.Val402Leu
##          1          1          1          3          1          1
## p.Val424Gly p.Val64Ala
##          6          1

```

```

atpe <- levels(as.factor(atpe$AAchange))
length(atpe)

```

```
## [1] 92
```

```
bdq <- rbind(atpe1,Rv0678,pepQ,Rv1979c)
write.table(bdq, file="/media/jupiter/marina/project-msbt/bdq.txt", quote = FALSE, sep = "\t")
```

```
unique(variants$ANNOTATION)
```

```
## [1] "missense_variant"      "stop_gained"
## [3] "frameshift_variant"    "stop_lost"
## [5] "gene_fusion"           "disruptive_inframe_deletion"
## [7] "initiator_codon_variant" "start_lost"
```

```
#levels(variants$ANNOTATION)
```

```
table(variants$ANNOTATION)
```

```
##
## disruptive_inframe_deletion      frameshift_variant
##                               1                55
##                gene_fusion      initiator_codon_variant
##                               46                4
##                missense_variant      start_lost
##                               819                1
##                stop_gained          stop_lost
##                               43                5
```

```
atpe <- variants$GENE=="ddn"
ddn2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 24
```

```
atpe$AAchange <- as.character(atpe$AAchange)
ddn <- distinct(atpe,AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
##      p.Ala120fs      p.Arg68Leu      p.Arg72Leu      p.Asn25fs
##           1           1           1           1
##      p.Gln101Arg      p.Gln134Arg      p.Glu121*      p.Glu35Asp
##           1           1           1           1
##      p.Gly34Val      p.Gly81Ser      p.Leu9Pro      p.Lys103Asn
##           2           1           1           1
##      p.Lys84Arg      p.Met21Ile      p.Phe41Tyr      p.Pro151His
##           1           1           1           1
##      p.Pro45Thr      p.Pro96His      p.Ter152Cysext*?      p.Trp27fs
##           1           1           1           1
##      p.Tyr29*      p.Tyr29Cys      p.Val147Ala
##           1           1           1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 23
```

```
atpe <- variants$GENE=="fbiA"
fbiA2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 83
```

```
atpe$AAchange <- as.character(atpe$AAchange)
fbiA <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
##      p.Ala204Ser      p.Ala212Ser      p.Ala222Ser      p.Ala231Glu
##           1           2           1           1
##      p.Ala231Thr      p.Ala269Ser      p.Ala281Thr      p.Ala315Glu
##           1           1           1           1
##      p.Ala322Ser      p.Arg234Leu      p.Arg274Trp      p.Arg321Ser
##           1           1           1           1
##      p.Asp257Gly      p.Asp257Tyr      p.Asp286Glu      p.Asp312Tyr
##           1           1           1           1
##      p.Asp36Gly       p.Asp63Gly       p.Cys287*         p.Cys59*
##           1           1           1           1
##      p.Gln117fs       p.Gln120His       p.Gln21*         p.Gln76His
##           1           1           1           1
##      p.Glu191Lys       p.Glu198*         p.Glu297*         p.Glu84Gln
##           1           1           1           1
##      p.Gly105Trp       p.Gly18Trp        p.Gly18Val        p.Gly242Arg
##           1           2           1           1
##      p.Gly24Cys        p.Gly264Trp       p.Gly283Glu       p.Gly293Cys
##           1           2           1           1
##      p.Gly300Trp       p.Gly328*         p.Gly48Ser        p.Gly71Asp
##           2           1           1           1
##      p.Gly71Cys        p.His112Asn       p.His183Arg       p.His291Tyr
##           1           1           1           1
##      p.His295Asn       p.Ile229Phe       p.Ile247Met       p.Leu125Met
##           1           1           1           1
##      p.Leu132Pro       p.Leu224Val       p.Met89Ile        p.Phe28Val
##           1           1           1           2
##      p.Pro124His       p.Pro159Thr       p.Pro239Thr       p.Pro251Thr
##           1           1           1           1
##      p.Pro307Gln       p.Pro313Gln       p.Pro313Leu       p.Ser163fs
##           1           1           1           1
##      p.Ser267Tyr       p.Ser32Pro        p.Ser35* p.Ter332Leuext*?
##           1           1           1           2
##      p.Thr182Met       p.Thr64Ala        p.Trp172Leu       p.Trp79Leu
```



```
##          1          1          1          1
##      p.Trp86Leu      p.Val114fs      p.Val226Ala      p.Val290Met
##          1          1          1          1
##      p.Val301Ala      p.Val306Glu      p.Val306Leu      p.Val58Phe
##          1          1          1          1
##      p.Val5Phe
##          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 77
```

```
atpe <- variants$GENE=="fbiB"
fbiB2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 112
```

```
atpe$AAchange <- as.character(atpe$AAchange)
fbiB <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
## p.Ala106Thr p.Ala127Asp p.Ala171Thr p.Ala199Thr p.Ala201Thr p.Ala202Thr
##          1          1          1          1          1          1
## p.Ala202Val p.Ala237Ser p.Ala279Glu p.Ala288Val p.Ala336Thr p.Ala344Thr
##          1          1          1          1          1          1
## p.Ala404Val p.Ala405Thr p.Ala423Ser p.Ala423Thr p.Ala429Val p.Arg115Gln
##          1          1          1          1          1          1
## p.Arg115Trp p.Arg154Cys p.Arg180Cys p.Arg218Leu p.Arg24His p.Arg296*
##          1          1          1          1          2          1
## p.Arg308Leu p.Arg312Trp p.Arg334Leu p.Arg365His p.Asn112Lys p.Asn238fs
##          1          1          1          1          1          1
## p.Asn238Lys p.Asp124Glu p.Asp124His p.Asp162fs p.Asp203Asn p.Asp203Tyr
##          1          1          1          1          1          1
## p.Asp274Tyr p.Asp28fs p.Asp324Gly p.Asp329Tyr p.Asp355Tyr p.Asp410Tyr
##          1          1          1          1          1          1
## p.Asp413Tyr p.Asp417Gly p.Asp42Glu p.Asp437Tyr p.Asp71Gly p.Gln254His
##          1          1          1          1          1          1
## p.Gln301His p.Glu197Asp p.Glu22fs p.Glu369* p.Glu369Gly p.Glu411*
##          1          1          1          1          1          1
## p.Glu5* p.Glu65Asp p.Glu68Asp p.Glu81* p.Gly100Arg p.Gly100Trp
##          1          1          1          1          1          1
## p.Gly107Cys p.Gly107Ser p.Gly169fs p.Gly169Val p.Gly244Trp p.Gly252Cys
##          1          1          1          1          1          1
## p.Gly356Val p.Gly394Val p.Gly427Val p.Gly434Trp p.Gly58Val p.His290Leu
##          1          1          1          1          1          1
## p.His358Leu p.His370Tyr p.His6Arg p.Ile331Leu p.Ile331Thr p.Ile78Phe
##          1          1          1          1          1          1
```

```
## p.Ile94Val p.Leu275Pro p.Leu310fs p.Leu407Arg p.Leu421Phe p.Leu77Met
##          1          1          1          1          3          1
## p.Phe220Ile p.Phe23Ser p.Pro122Gln p.Pro122Thr p.Pro270Gln p.Pro270Ser
##          1          1          1          1          1          1
## p.Pro289Gln p.Pro289Thr p.Pro294Gln p.Pro345Thr p.Pro350Gln p.Pro415Leu
##          1          1          1          1          1          1
## p.Pro420Gln p.Pro440Ser p.Pro62Ser p.Ser128Arg p.Thr245Ser p.Thr366Ser
##          1          1          1          1          1          1
## p.Trp317Leu p.Val102Phe p.Val113Phe p.Val214Leu p.Val222Ala p.Val271Leu
##          1          1          1          1          1          1
## p.Val335Met
##          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 109
```

```
atpe <- variants$GENE=="fbiC"
fbiC2 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 161
```

```
atpe$AAchange <- as.character(atpe$AAchange)
fbiC <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
## p.Ala11Thr p.Ala176Val p.Ala239Glu p.Ala271Thr p.Ala297Asp p.Ala297Ser
##          1          1          2          1          1          1
## p.Ala304Asp p.Ala333Glu p.Ala345Gly p.Ala379Thr p.Ala421Glu p.Ala444Glu
##          1          1          2          1          1          1
## p.Ala497Asp p.Ala498Thr p.Ala507Ser p.Ala518Ser p.Ala520Asp p.Ala529Asp
##          1          1          1          1          1          1
## p.Ala530Thr p.Ala56fs p.Ala596fs p.Ala615Ser p.Ala68Val p.Ala73Thr
##          1          1          1          1          1          1
## p.Ala844Gly p.Ala852Pro p.Arg156Leu p.Arg165Leu p.Arg175Cys p.Arg250Leu
##          1          1          1          1          1          1
## p.Arg322Leu p.Arg330Cys p.Arg344Gly p.Arg411Trp p.Arg458Pro p.Arg463Gly
##          1          1          1          1          1          1
## p.Arg486Ser p.Arg504Gly p.Arg504His p.Arg550Cys p.Arg758Ser p.Arg75Gln
##          1          1          1          1          1          1
## p.Arg75Leu p.Arg818Gln p.Arg842His p.Arg845His p.Asn40Lys p.Asn429Asp
##          1          1          1          1          1          1
## p.Asn724Asp p.Asp375fs p.Asp387Tyr p.Asp492Tyr p.Asp538Tyr p.Asp586Tyr
##          1          1          1          1          1          1
## p.Asp674Glu p.Asp728Asn p.Cys105Phe p.Gln293His p.Gln332Leu p.Gln332Lys
##          1          2          1          2          1          1
## p.Glu200* p.Glu299* p.Glu312Asp p.Glu43Asp p.Glu54* p.Glu543Asp
```

```
##          2          1          1          1          1          1
## p.Glu543fs p.Glu590Asp p.Gly112Ala p.Gly166Cys p.Gly194Asp p.Gly356Cys
##          1          1          1          1          1          1
## p.Gly52Cys p.Gly666Ser p.Gly711Val p.Gly74Ala p.Gly821Ser p.His481Asn
##          1          1          1          4          1          1
## p.His698Asn p.His746Gln p.Ile287fs p.Ile654Phe p.Ile693Val p.Leu323Met
##          2          1          1          1          1          1
## p.Leu325Met p.Leu337Met p.Leu374Phe p.Leu531Pro p.Leu680Arg p.Leu772fs
##          1          1          1          1          1          1
## p.Lys684Asn p.Lys828Arg p.Met123Ile p.Met214Ile p.Phe567fs p.Pro209Gln
##          1          1          1          1          1          1
## p.Pro238Gln p.Pro307His p.Pro327fs p.Pro367His p.Pro372His p.Pro420Gln
##          1          1          1          2          1          1
## p.Pro434Gln p.Pro607His p.Pro665Gln p.Pro743Thr p.Pro761His p.Pro78Ser
##          1          1          1          1          1          2
## p.Pro843Gln p.Pro93fs p.Pro93Leu p.Ser280fs p.Ser534* p.Ser634fs
##          1          1          1          1          1          1
## p.Ser705* p.Ser87Thr p.Thr109Ser p.Thr10Ala p.Thr273Ala p.Thr301Ile
##          1          1          1          1          2          1
## p.Thr361Ser p.Thr519Ala p.Thr696Asn p.Thr810Ala p.Trp435Leu p.Trp589Cys
##          1          1          1          1          1          1
## p.Trp652Cys p.Trp690Leu p.Trp788Leu p.Tyr851Cys p.Val110Met p.Val320Ala
##          1          1          1          1          1          1
## p.Val380Ala p.Val41Glu p.Val41Met p.Val462Leu p.Val539Gly p.Val611Asp
##          1          1          9          1          1          1
## p.Val611Ile p.Val740Ile p.Val830Ala
##          1          1          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 141
```

```
atpe <- variants$GENE=="Rv2983"
Rv29832 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 73
```

```
atpe$AAchange <- as.character(atpe$AAchange)
Rv2983 <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
## p.Ala104Thr p.Ala132Val p.Ala141Thr p.Ala15Ser
##          1          1          1          1
## p.Ala198Thr p.Ala22Ser p.Ala41Val p.Ala50Ser
##          1          1          1          1
## p.Ala51Thr p.Ala71Ser p.Ala73Gly p.Ala95Thr
##          1          1          1          1
```

##	p.Ala98Gly	p.Ala98Ser	p.Arg137Leu	p.Arg143Leu
##	1	1	1	1
##	p.Arg186Gly	p.Arg35His	p.Arg35Ser	p.Asn93Ser
##	1	1	1	1
##	p.Asp194Val	p.Asp7Tyr	p.Asp88Tyr	p.Asp90Asn
##	1	1	1	1
##	p.Gln114Arg	p.Gln114Leu	p.Gln33His	p.Glu100*
##	1	1	1	1
##	p.Glu177Gly	p.Glu36Asp	p.Gly106Trp	p.Gly145Val
##	1	1	1	1
##	p.Gly202Trp	p.Gly202Val	p.Gly54Val	p.Gly76Cys
##	1	1	1	1
##	p.His134Asn	p.His135del	p.His213Tyr	p.Ile10fs
##	1	1	1	2
##	p.Ile14Phe	p.Leu117Met	p.Leu151Pro	p.Leu158Val
##	1	1	1	1
##	p.Leu43Val	p.Leu47Ser	p.Met1?	p.Phe30Leu
##	1	1	2	1
##	p.Pro160Gln	p.Pro205Leu	p.Pro28Thr	p.Pro64His
##	1	1	1	1
##	p.Pro83Ala	p.Pro83Gln	p.Pro85fs	p.Ser173Leu
##	2	1	1	1
##	p.Ser58*	p.Ter215Leuext*?	p.Thr122Ile	p.Thr146Ala
##	1	1	1	1
##	p.Thr148Ala	p.Thr191Asn	p.Thr46Pro	p.Thr84Lys
##	1	1	1	1
##	p.Thr97Asn	p.Val102Glu	p.Val102Leu	p.Val16Ala
##	1	1	1	1
##	p.Val203fs	p.Val38Ala		
##	1	1		

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 70
```

```
atpe <- variants$GENE=="fgd1"
fgd12 <- variants[atpe, ]
atpe <- variants[atpe, ]
total <- nrow(atpe)
print(total)
```

```
## [1] 99
```

```
atpe$AAchange <- as.character(atpe$AAchange)
fgd1 <- distinct(atpe, AAchange, .keep_all = TRUE)
table(atpe$AAchange)
```

```
##
## p.Ala110Gly p.Ala271Asp p.Ala29Asp p.Arg100Gly p.Arg18Leu p.Arg220Gln
## 1 1 1 1 1 1
## p.Arg247Leu p.Arg334Leu p.Arg45Cys p.Asp146Glu p.Asp146Tyr p.Asp149Asn
```

```
##          1          1          1          1          1          1
## p.Asp164Glu p.Asp219Gly p.Asp219Tyr p.Asp226Asn p.Asp226Tyr p.Asp291Glu
##          1          1          1          1          3          1
## p.Asp327Tyr p.Asp39Val p.Cys194Tyr p.Cys95* p.Gln14Glu p.Gln299Lys
##          1          1          1          1          1          1
## p.Gln317fs p.Gln42Lys p.Glu19Lys p.Glu201fs p.Glu205* p.Glu240*
##          1          1          1          1          1          1
## p.Glu3Lys p.Gly106Ala p.Gly137Trp p.Gly159Ser p.Gly199Cys p.Gly314Glu
##          1          1          1          1          1          1
## p.Gly31Ser p.His315Tyr p.His46Tyr p.Ile162Phe p.Ile174Leu p.Ile225Phe
##          1          1          1          1          2          2
## p.Ile231Val p.Leu142Met p.Leu157Ile p.Leu308Pro p.Leu69Gln p.Lys158Asn
##          1          1          1          1          1          1
## p.Lys198Asn p.Lys198Glu p.Lys270Met p.Lys296Asn p.Lys296Glu p.Met228Ile
##          3          1          9          2          1          1
## p.Met58Ile p.Phe102fs p.Phe129Leu p.Phe53Leu p.Phe79fs p.Pro171Leu
##          1          1          1          1          1          1
## p.Pro171Thr p.Pro209Leu p.Pro239Gln p.Pro265Leu p.Pro276Ala p.Pro52Thr
##          1          2          1          1          1          1
## p.Ser11* p.Ser161* p.Ser221Tyr p.Ser253* p.Ser54* p.Thr255Ser
##          1          1          1          2          1          1
## p.Thr36Ser p.Thr59Ala p.Trp249* p.Trp303Cys p.Trp303Leu p.Tyr173*
##          1          1          1          1          1          1
## p.Val1? p.Val136Ala p.Val181fs
##          2          1          1
```

```
atpe <- levels(as.factor(atpe$AAchange))
length(atpe)
```

```
## [1] 81
```

```
d1m <- rbind(fbiA,fbiB,fbiC,Rv2983,fgd1,ddn)
write.table(d1m, file="/media/jupiter/marina/project-msbt/d1m.txt", quote = FALSE, sep = "\t")
```

```
fresults <- read.csv2(file="/media/jupiter/marina/project-msbt/final-results.csv", header = TRUE, sep="
```

```
resistant <- c("TN106506","TN106458","TN106541","TN106437","TN106558","TN106497","TN106838","TN106655",
```

```
list <- character()
for (i in seq(1:nrow(fresults))){
  if (fresults[i,7] %in% resistant){
    list[i] <- "resistant"
  }
  else {
    list[i] <- "susceptible"
  }
}
}
```

```
fresults$Resistance <- list
```