

# Log transformation of data

2023-10-25

```
#####log transformation of data

library(MASS)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##      select

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

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

library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## vforcats    1.0.0    vreadr     2.1.4
## vggplot2    3.4.4    vstringr   1.5.0
## vlubridate  1.9.3    vtibble    3.2.1
## vpurrr      1.0.2    vtidyr    1.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Root_necrosis <- read.csv("~/Desktop/QPCR_PAPER/Rootnecrosis_paper/Data/git/Root-necrosis-images/root_necrosis")
dim(Root_necrosis)

## [1] 51374    47
```

```

#####RENAME COLUMNS AND USING ONLY THOSE THAT WERE NORMALIZED

Data1 <- Root_necrosis %>%
  rename(SN = solidity_necrosis, CHVN = convex_hull_vertices_necrosis, EAN= ellipse_angle_necrosis, EEN= ellipse_eccentricity_necrosis, NECRO= ellipse_neck_ratio_necrosis, NAF= ellipse_axis_ratio_necrosis, NWF= ellipse_width_factor_necrosis)

# Select specific columns using dplyr
selected_data <- select(Data1, SN, CHVN, EAN, EEN, NECRO, NAF, NWF)

#####NA does not mean that the image is missing it means no root necrosis was identified
# Replace NA values with 0.0001

selected_data [is.na(selected_data )] <- 0.0001

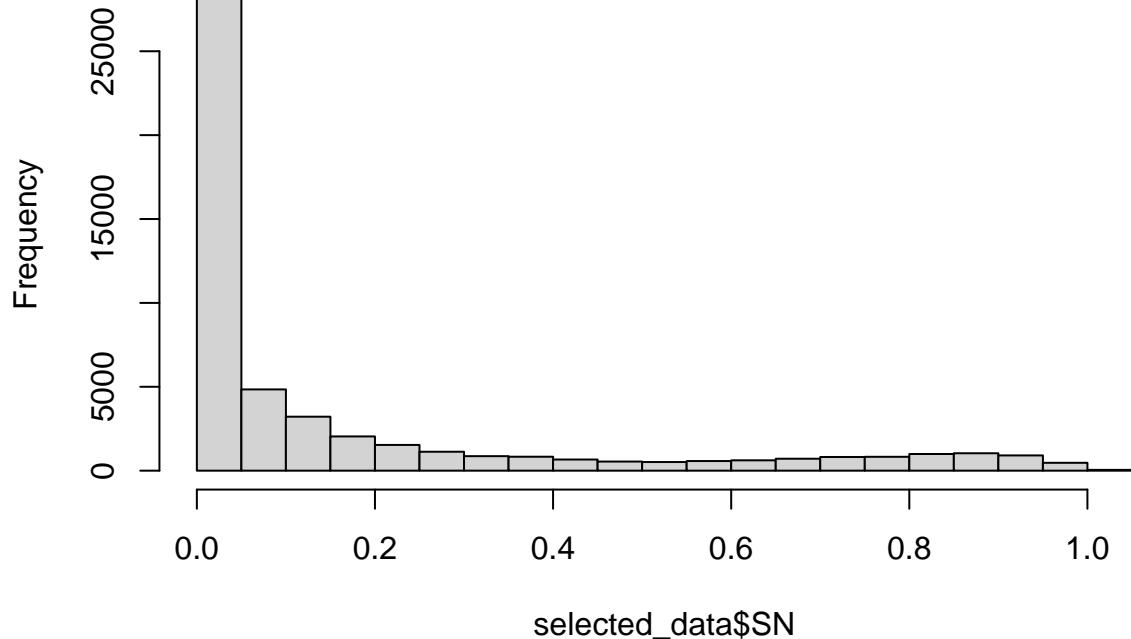
```

#CHECK DISTRIBUTION OF RAW DATA

####HISTOGRAM

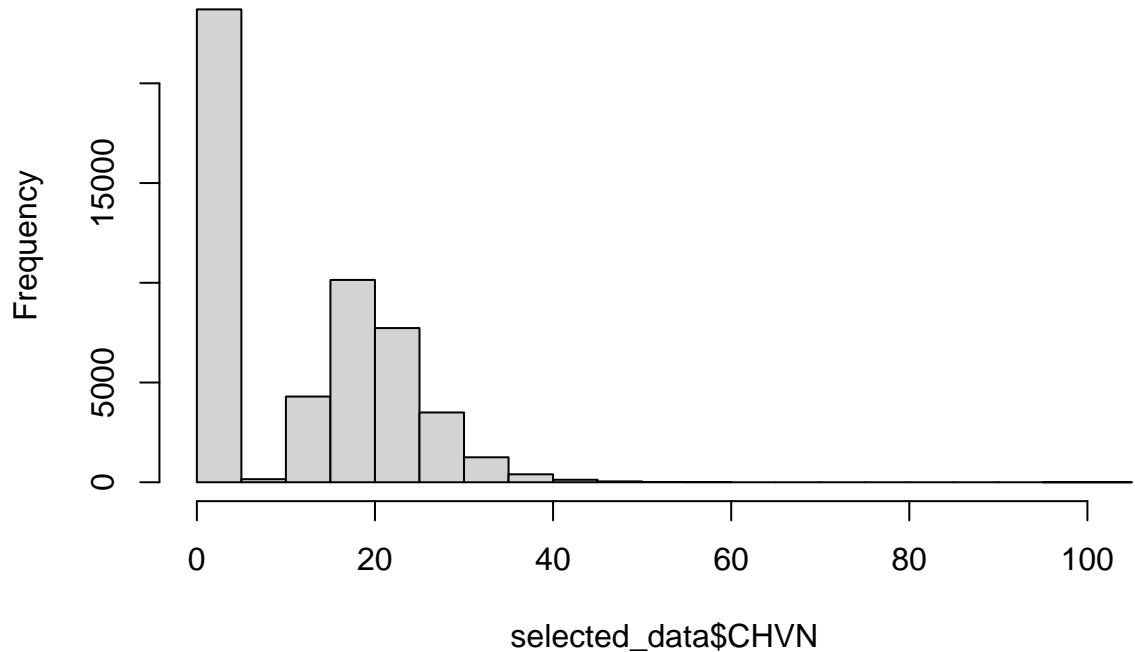
```
hist(selected_data$SN)
```

## Histogram of selected\_data\$SN



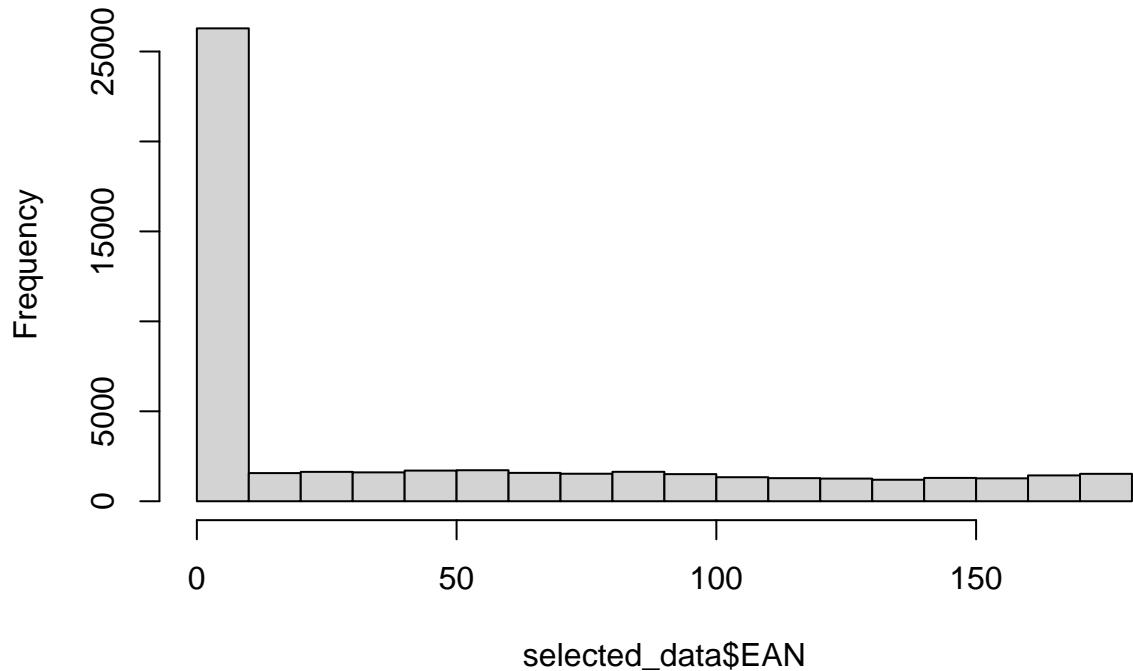
```
hist(selected_data$CHVN)
```

**Histogram of selected\_data\$CHVN**



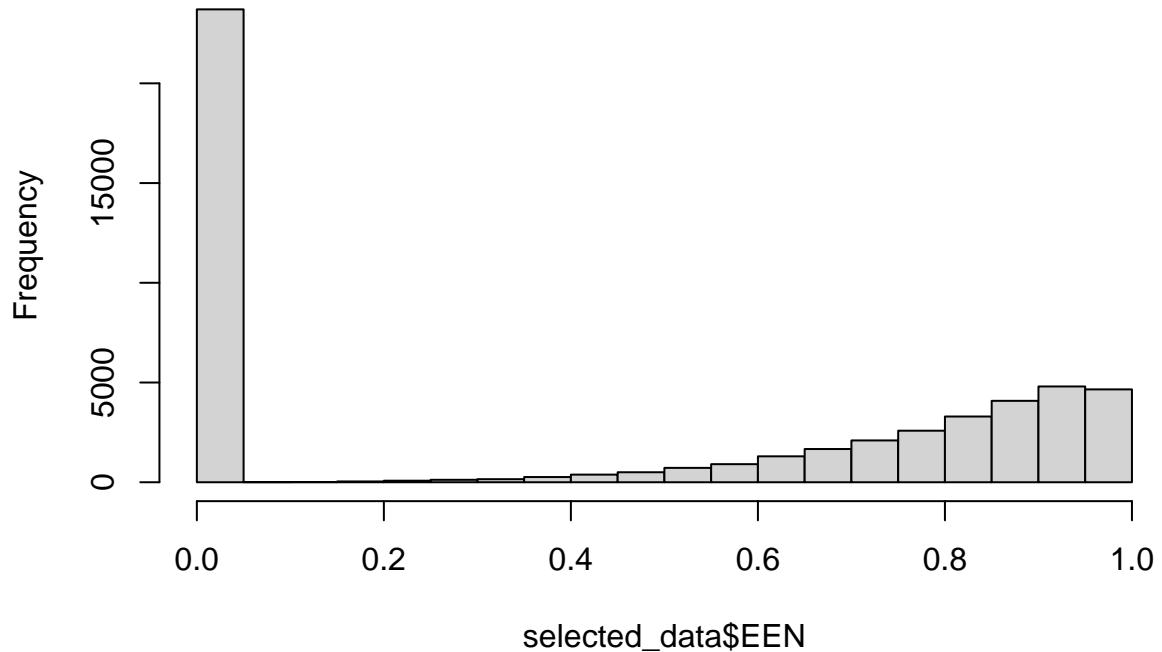
```
hist(selected_data$EAN)
```

### Histogram of selected\_data\$EAN



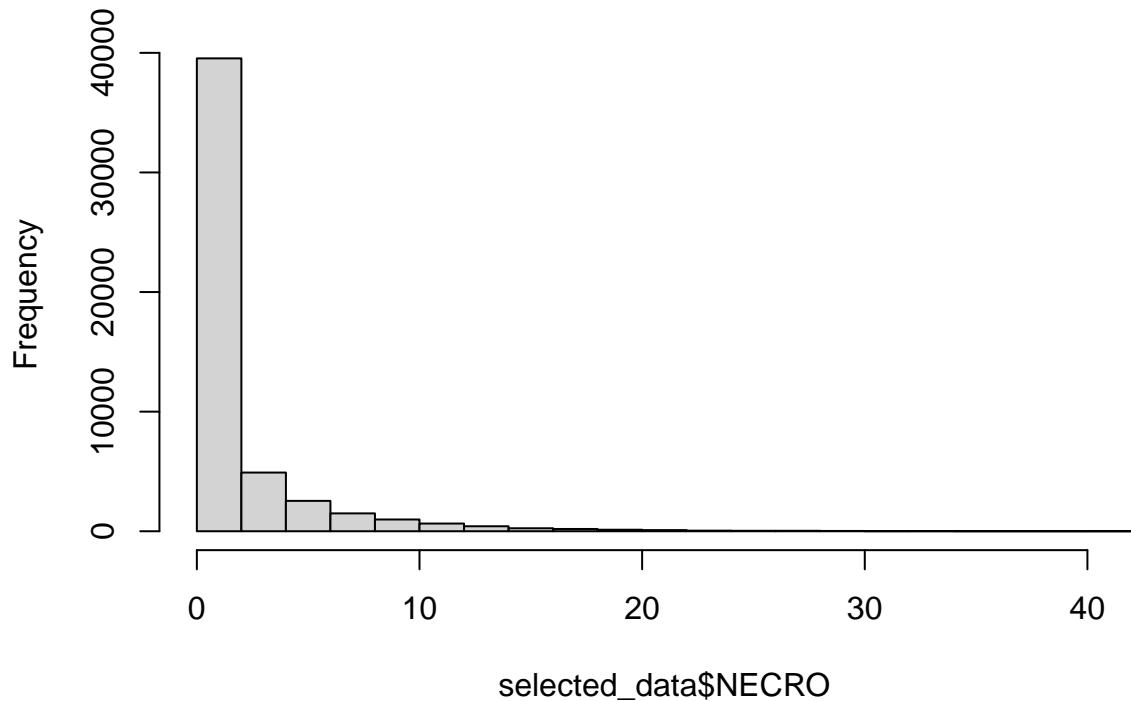
```
hist(selected_data$EAN)
```

**Histogram of selected\_data\$EEN**



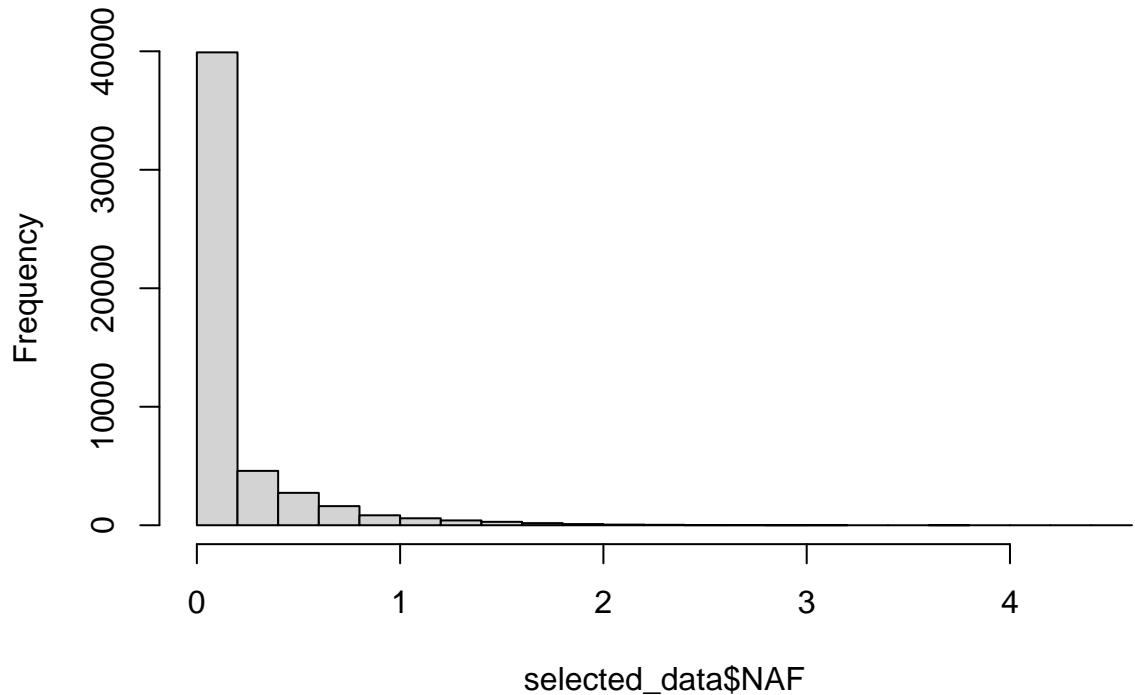
```
hist(selected_data$NECRO)
```

**Histogram of selected\_data\$NECRO**

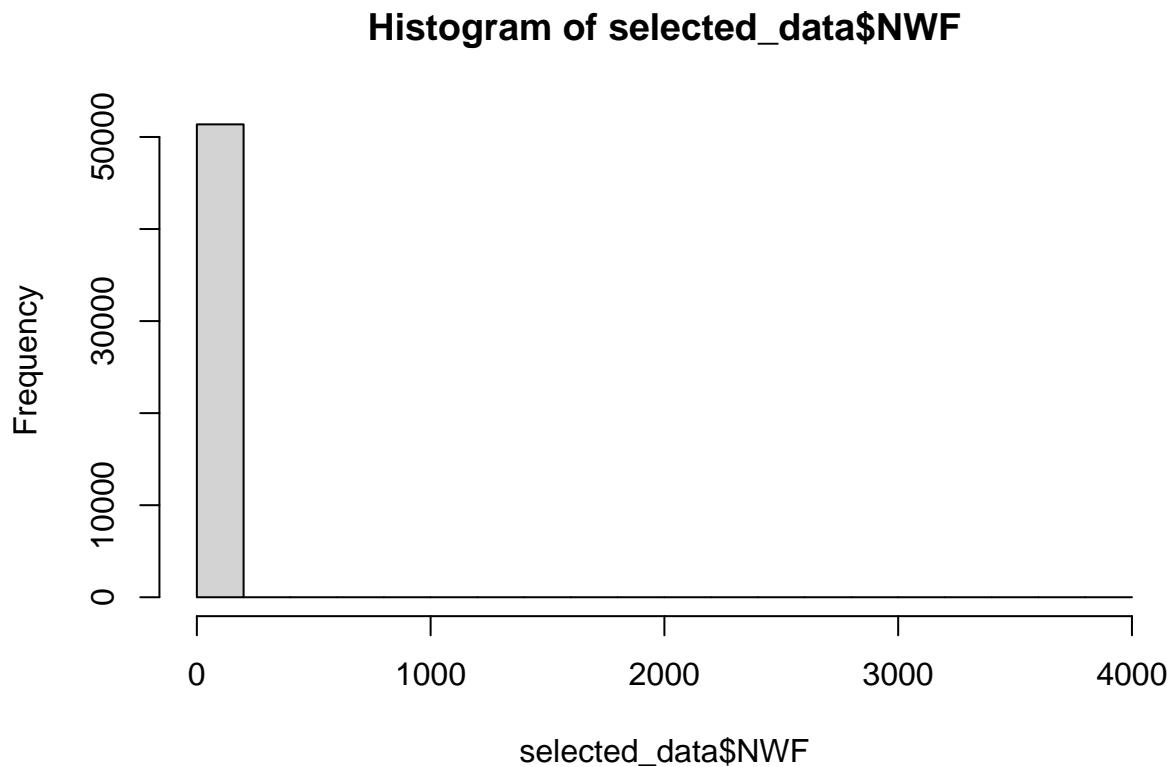


```
hist(selected_data$NAF)
```

**Histogram of selected\_data\$NAF**



```
hist(selected_data$NWF)
```

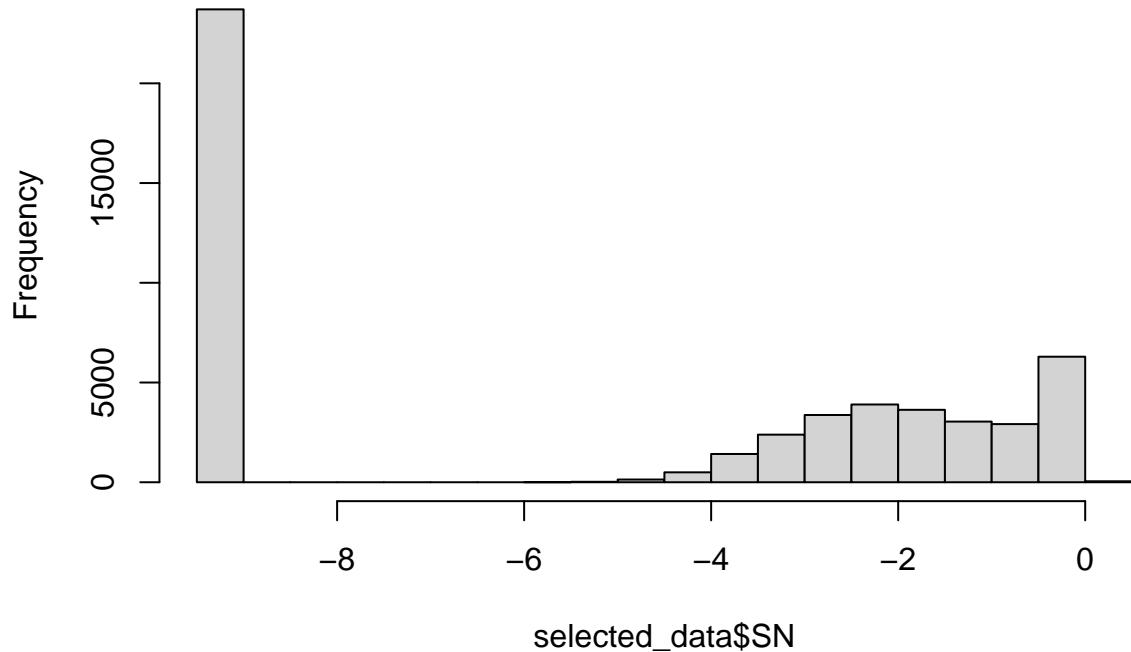


```
# Apply natural log transformation to selected columns
columns_to_transform <- c("SN", "CHVN", "EAN", "EEN", "NECRO", "NAF", "NWF")
selected_data[columns_to_transform] <- lapply(selected_data[columns_to_transform], function(x) log(x))

##### PLOTTING LOG TRANSFORMED DATA

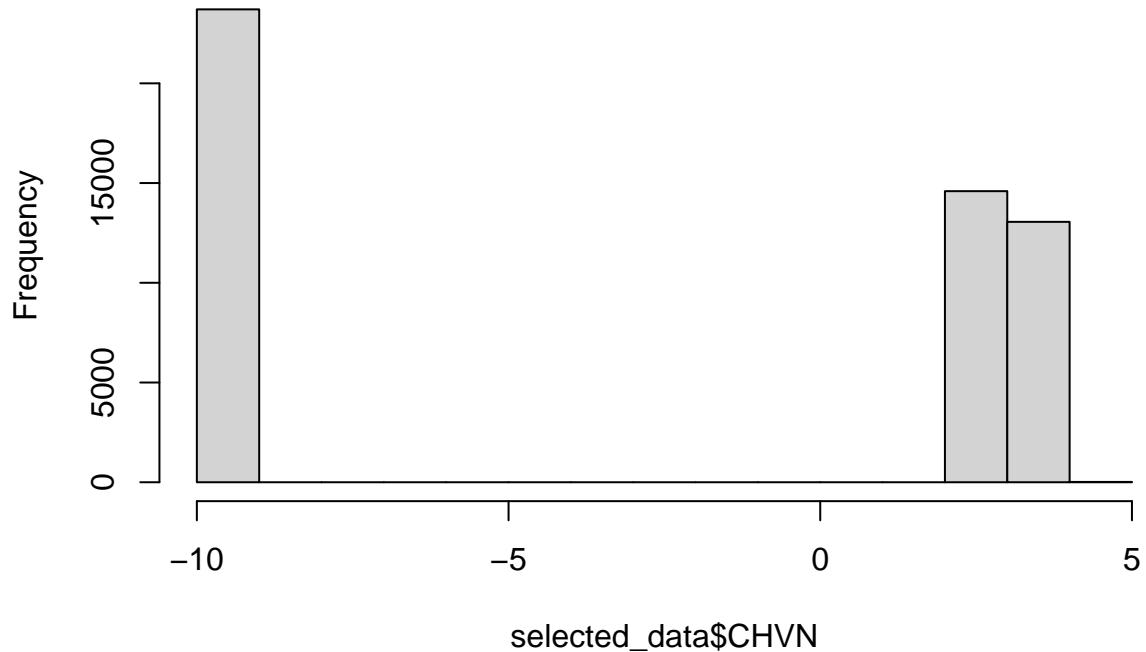
hist(selected_data$SN)
```

**Histogram of selected\_data\$SN**



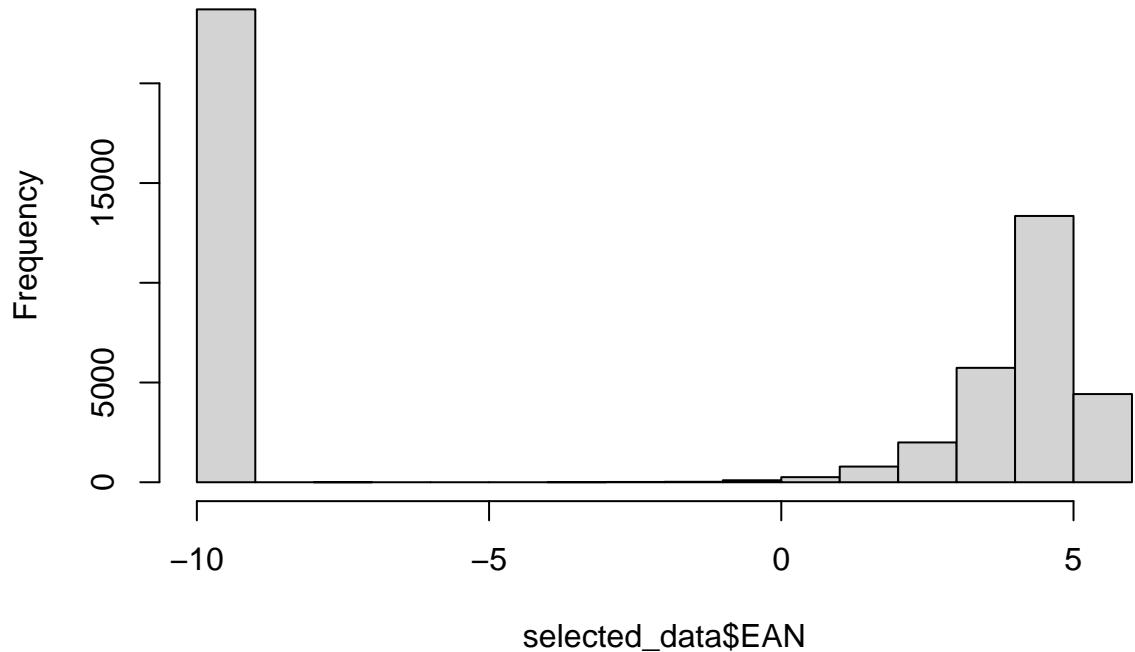
```
hist(selected_data$CHVN)
```

### Histogram of selected\_data\$CHVN



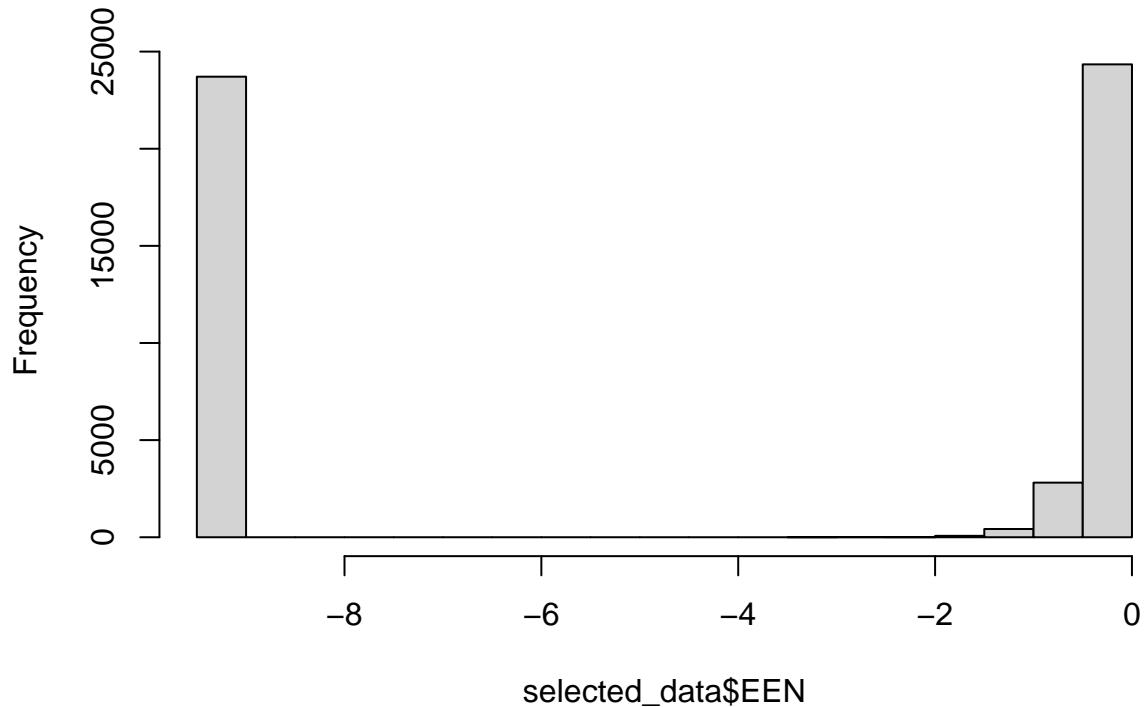
```
hist(selected_data$EAN)
```

**Histogram of selected\_data\$EAN**



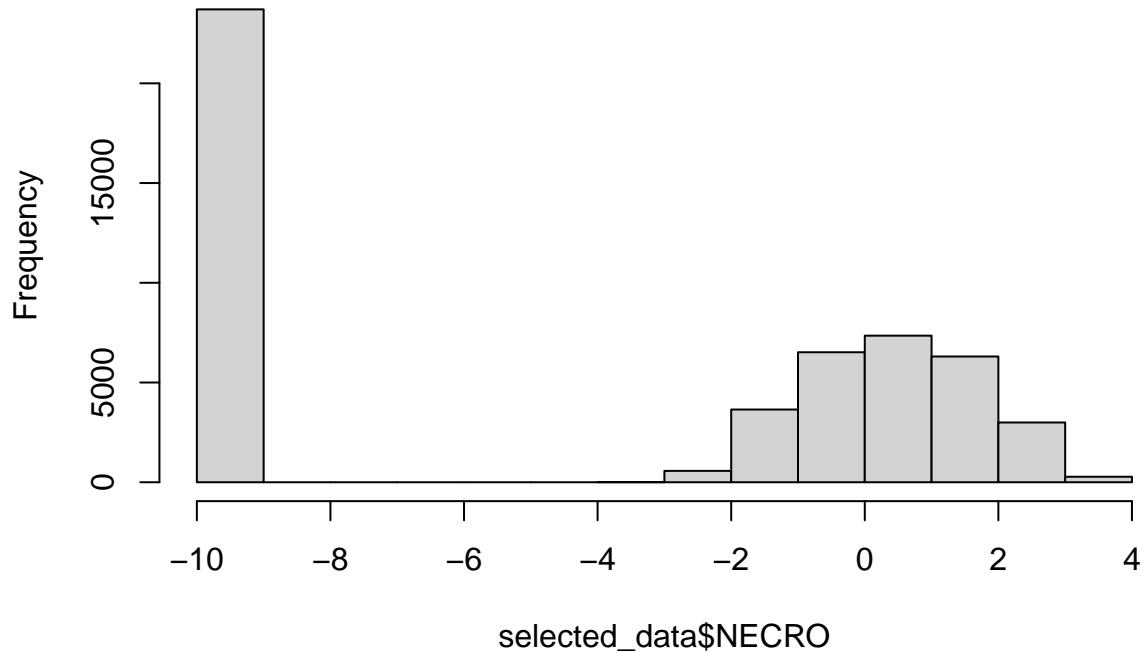
```
hist(selected_data$EAN)
```

**Histogram of selected\_data\$EEN**



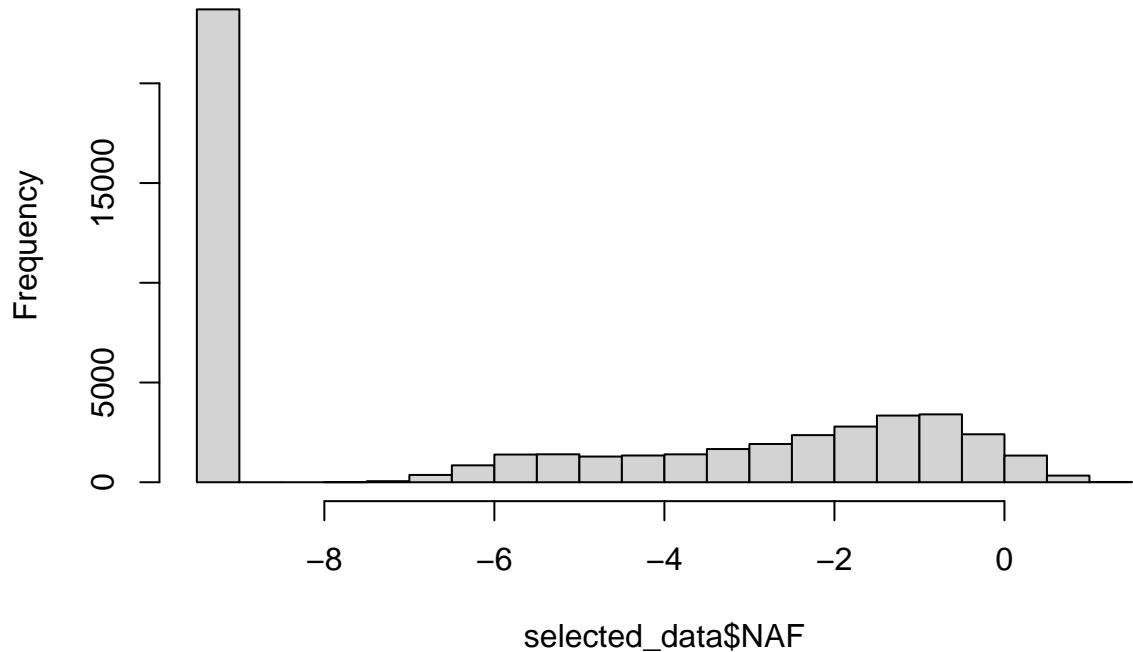
```
hist(selected_data$NECRO)
```

**Histogram of selected\_data\$NECRO**



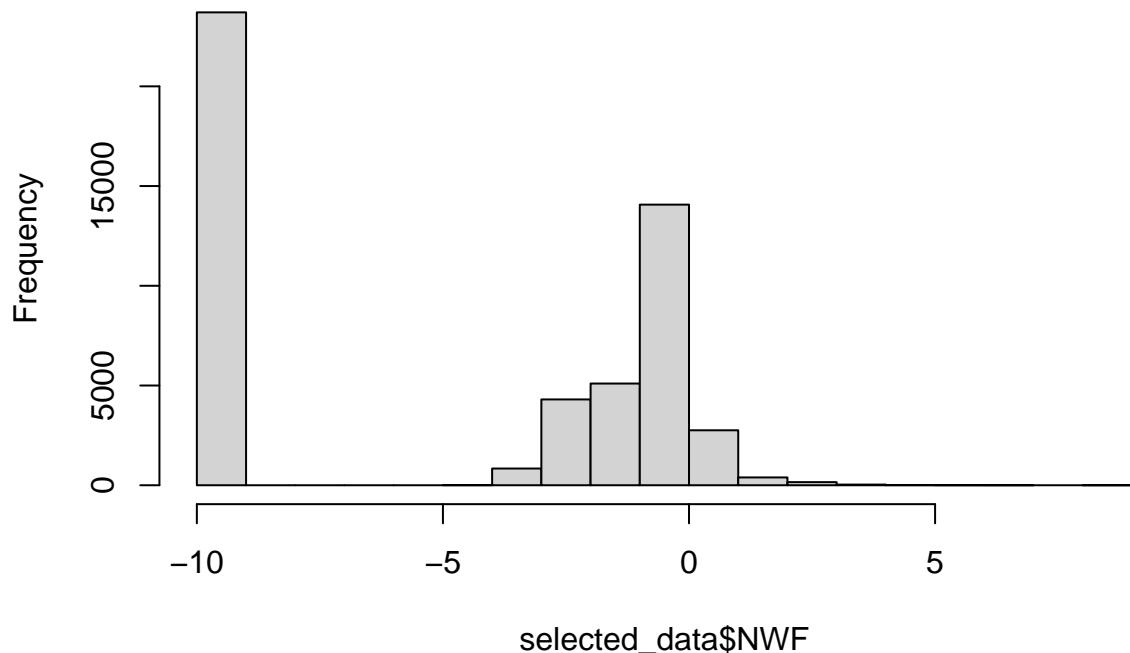
```
hist(selected_data$NAF)
```

**Histogram of selected\_data\$NAF**



```
hist(selected_data$NWF)
```

## Histogram of selected\_data\$NWF



```
##### PHENOTYPIC CORRELATIONS OF THE SEVEN NORMALIZED TRAITS
```

```
##### phenotypic correlations
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##     %+%, alpha
```

```
# Remove rows with NA values using complete.cases()
data_clean <- selected_data[selected_data$EAN != -Inf, ]
```

```
# Create scatter plot matrix with histograms and correlation values
pairs.panels(data_clean,
             method = "pearson", # Correlation method (you can change it to "spearman" for Spearman cor
             hist.col = "#75AADB", # Histogram color
             density = TRUE, # Show density plots on histograms
             ellipses = TRUE, # Show correlation ellipses
             stars = TRUE, # Show significance stars for correlations
             pch = 16, # Point type for scatter plots
```

```

    bg = "#FFA07A", # Point color
    # main = "Phenotypic Correlations", # Main title
    cex.cor = 1.2, # Font size of correlation values
    tl.cex = 0.6, # Font size of variable names
    col = "#2E86C1" # Scatter plot color
)

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
  
## Warning in title(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
  
## Warning in title(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
  
## Warning in title(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
  
## Warning in title(...): "tl.cex" is not a graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter
```

```
## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter
```

```

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter
```

```
## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter
```

```
## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "tl.cex" is not a  
## graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a  
## graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter  
## Warning in title(...): "tl.cex" is not a graphical parameter  
## Warning in plot.window(...): "tl.cex" is not a graphical parameter  
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter
```

```

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in plot.window(...): "tl.cex" is not a graphical parameter

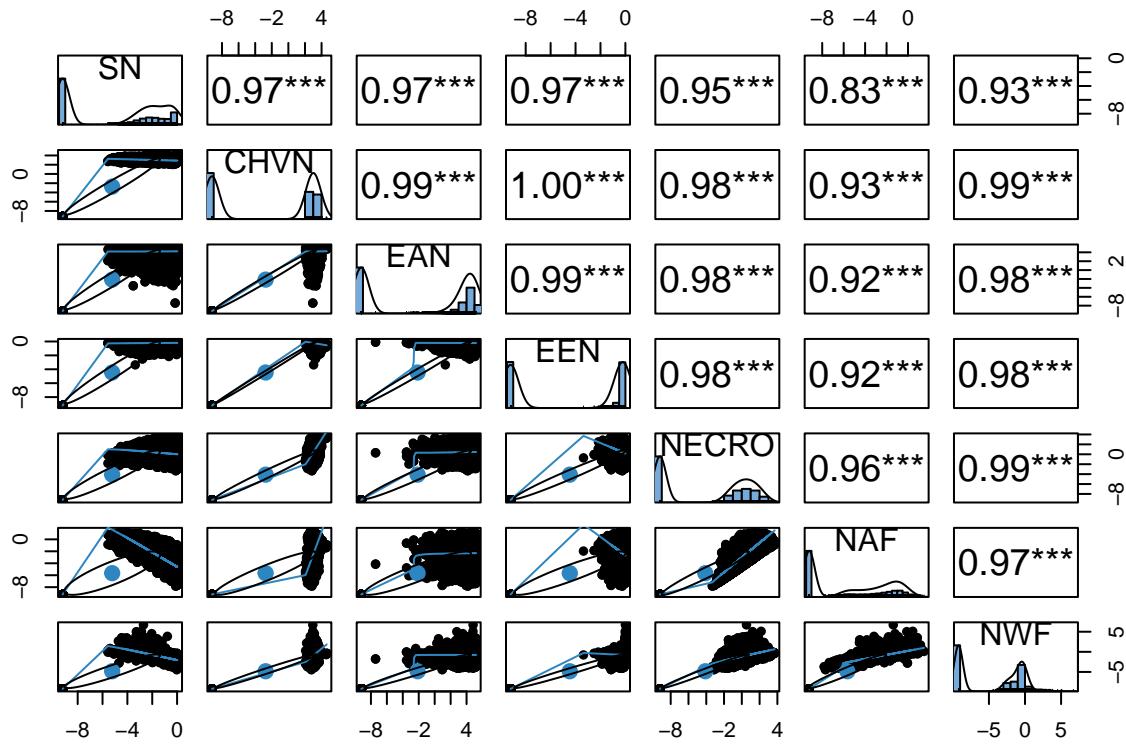
## Warning in plot.xy(xy, type, ...): "tl.cex" is not a graphical parameter

## Warning in title(...): "tl.cex" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "tl.cex" is not a
## graphical parameter

```



##### Biplot analysis

```

library(dplyr)
library(tidyverse)
library(corrplot)

## corrplot 0.92 loaded

library(Hmisc)

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:psych':
##      describe

## The following objects are masked from 'package:dplyr':
##      src, summarize

## The following objects are masked from 'package:base':
##      format.pval, units

library(MASS)
library(psych)
library(metan)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

## | ====== |
## | Multi-Environment Trial Analysis (metan) v1.18.0          |
## | Author: Tiago Olivoto                                     |
## | Type 'citation('metan')' to know how to cite metan       |
## | Type 'vignette('metan_start')' for a short tutorial        |
## | Visit 'https://bit.ly/pkgmetan' for a complete tutorial   |
## | ====== |

##
## Attaching package: 'metan'

```

```

## The following object is masked from 'package:psych':
##
##      skew

## The following object is masked from 'package:forcats':
##
##      as_factor

## The following object is masked from 'package:tidyverse':
##
##      replace_na

## The following objects are masked from 'package:tibble':
##
##      column_to_rownames, remove_rownames, rownames_to_column

## The following object is masked from 'package:dplyr':
##
##      recode_factor

## The following object is masked from 'package:MASS':
##
##      select

library(ggpubr)
library(lme4)

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyverse':
##
##      expand, pack, unpack

#####data with all root necrosis traits and the 1-5 scoring scale

FULL_ROOTNECROSIS_IMAGEDATA_2023 <- read.csv("FULL_ROOTNECROSIS_IMAGEDATA_2023.csv")

selected_data <- FULL_ROOTNECROSIS_IMAGEDATA_2023 [c("SN", "CHVN", "EAN", "EEN", "NECRO", "NAF", "NWF",

#####adding the enviromental variable by pasting location and year effects to do the biplot analysis

FULL_ROOTNECROSIS_IMAGEDATA_2023$ENV <- paste(FULL_ROOTNECROSIS_IMAGEDATA_2023$Location, FULL_ROOTNECROSIS_IMAGEDATA_2023$Year)

Data <- FULL_ROOTNECROSIS_IMAGEDATA_2023 %>% select(germplasmName, ENV)

```

```

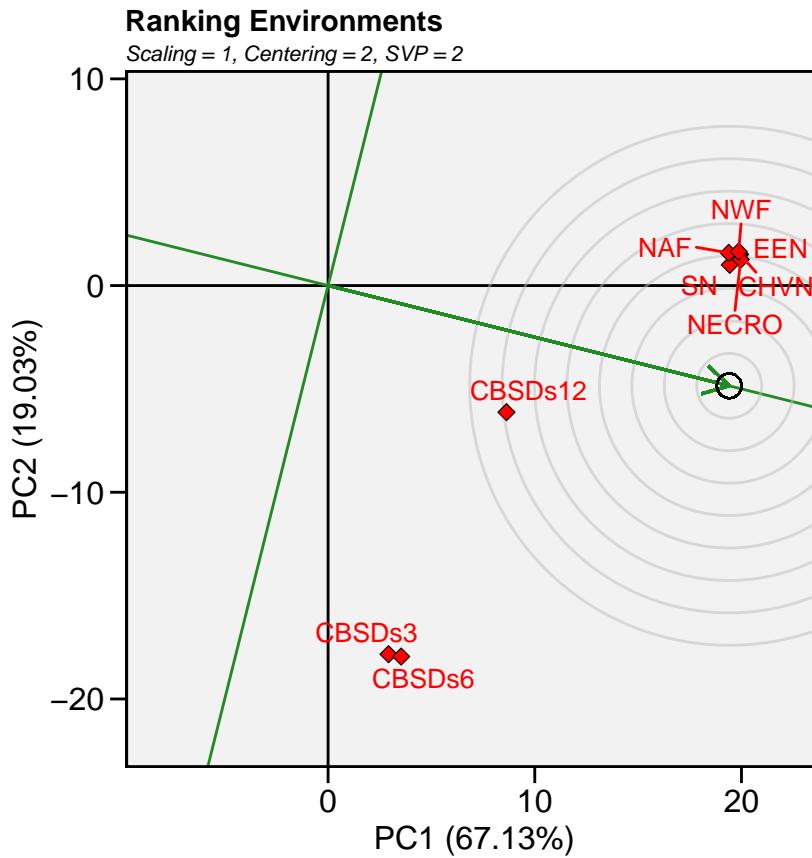
Data6 <- cbind(Data,selected_data)

#Ranking environments
Model1 <- gtb(Data6 , germplasmName, resp = c( SN,CHVN,EAN,EEN,NECRO,NAF,NWF,CBSDs3,CBSDs6,CBSDs12))
summary(Model1)

##      Length Class Mode
## mod 17     gge   list

plot(Model1,
      type = 6,
      col.gen = "blue",
      col.env = "red",
      size.text.gen = 4)

```

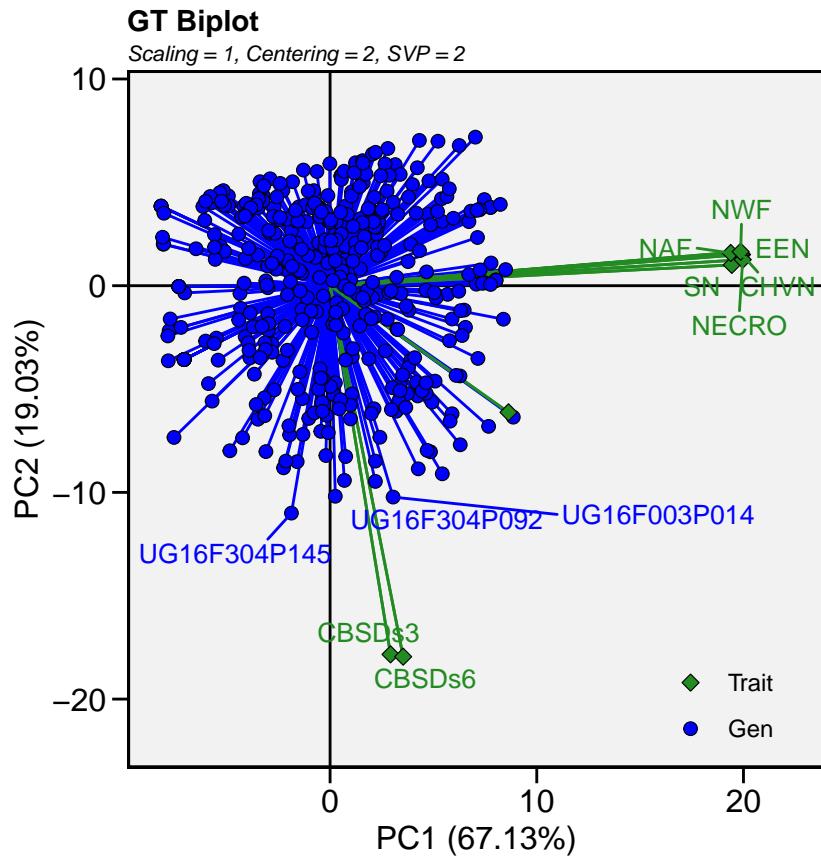


```

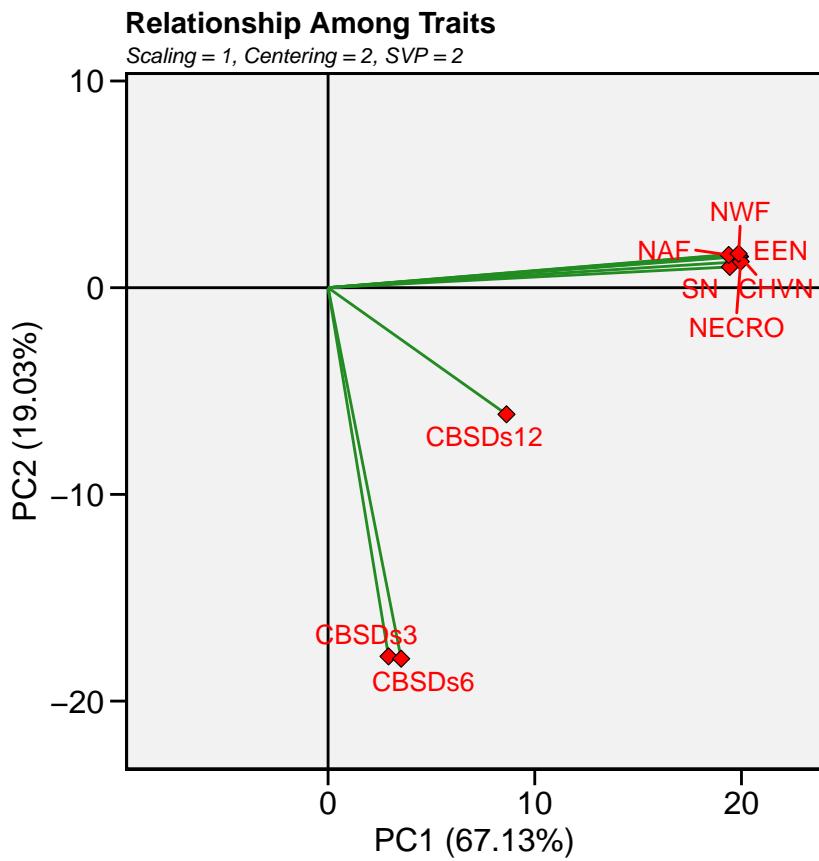
#####genotype by trait rankings
plot(Model1, sel_env = "ENV")

## Warning: ggrepel: 402 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```

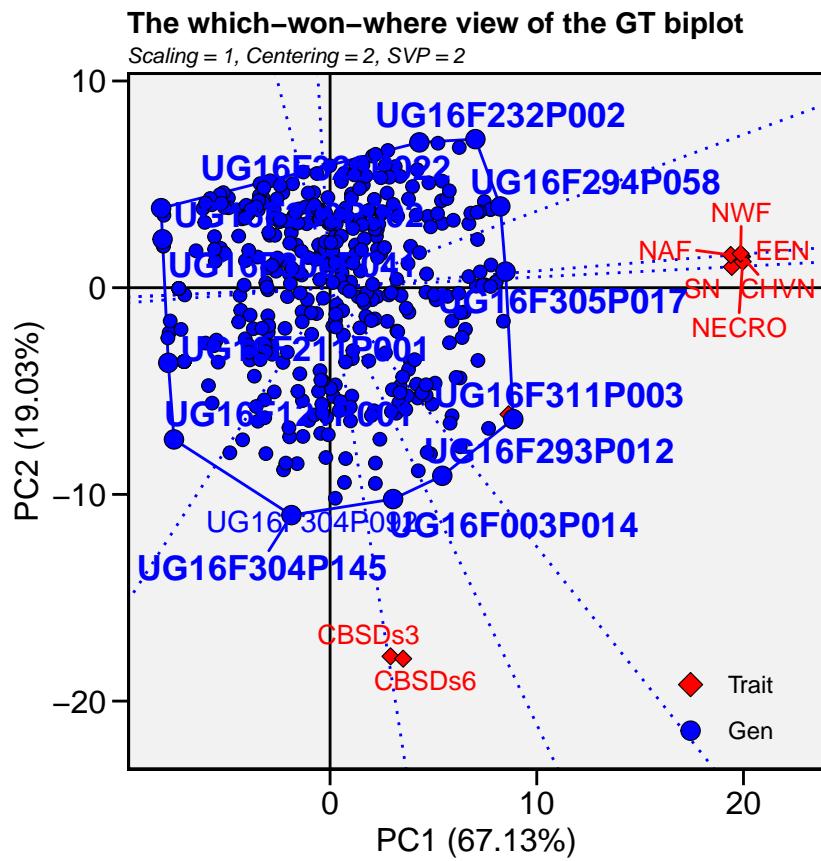


```
####relationships among traits
plot(Model1,
      type = 10,
      col.gen = "blue",
      col.env = "red",
      size.text.gen = 4)
```

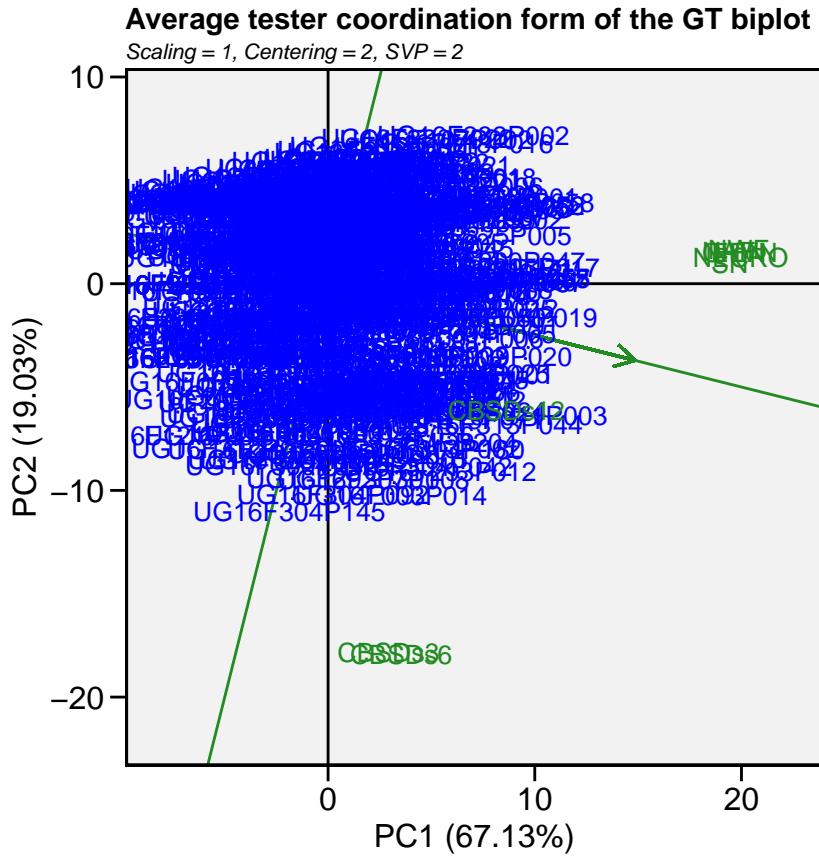


```
plot(Model1,
  type = 3,
  col.gen = "blue",
  col.env = "red",
  size.text.gen = 4)
```

```
## Warning: ggrepel: 392 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
plot(Model1,
      type = 2,
      second= "PC3")
```



```
ge_details(Data6 ,
  env = ENV,
  gen = germplasmName,
  resp = everything())
```

```
## # A tibble: 10 x 10
##   Parameters SN          CHVN EEN    NECRO NAF     NWF    CBSDs3 CBSDs6 CBSDs12
##   <chr>     <chr>      <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Mean      "-5.26"    "-2.~" "-4.~" "-4.~" "-5.~" "-4.~" "0.27" "0.41" "0.51"
## 2 SE        "0.02"     "0.0~" "0.0~" "0.0~" "0.0~" "0.0~" "0"     "0"     "0"
## 3 SD        "3.83"     "6.1"  "4.4~" "4.8~" "3.6~" "4.2~" "0.3"  "0.38" "0.42"
## 4 CV        "-72.74"    "-22~" "-10~" "-11~" "-64~" "-87~" "110.~" "91.8~" "82.54"
## 5 Min      "-9.21 (UG16F~" "-9.~" "-9.~" "-9.~" "-9.~" "-9.~" "0 (U~" "0 (M~" "0 (UG~"
## 6 Max      "0.03 (UG16F2~" "4.6~" "0 (~" "3.7~" "1.5~" "8.2~" "1.1 ~" "1.31~" "1.61 ~
## 7 MinENV   "Serere_2020 ~" "Ser~" "Ser~" "Ser~" "Ser~" "Ser~" "Sere~" "Sere~" "Serer~"
## 8 MaxENV   "Serere_2019 ~" "Ser~" "Ser~" "Ser~" "Ser~" "Ser~" "Sere~" "Sere~" "Namul~"
## 9 MinGEN   "UG16F118P001~" "UG1~" "UG1~" "UG1~" "UG1~" "UG1~" "UG16~" "UG16~" "UG16F~"
## 10 MaxGEN  "UG16F294P069~" "UG1~" "UG1~" "UG1~" "UG1~" "UG1~" "UG16~" "UG16~" "UG16F~"
```

```
GGEMODEL11 <- gge(Data6,ENV, germplasmName,resp = c(SN,CHVN,EAN,EEN,NECRO,NAF,NWF,CBSDs3,CBSDs6,CBSDs12)
```

```
## Warning: Data imputation used to fill the GxE matrix
```

```
## Warning: Data imputation used to fill the GxE matrix
```

```

## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix
## Warning: Data imputation used to fill the GxE matrix

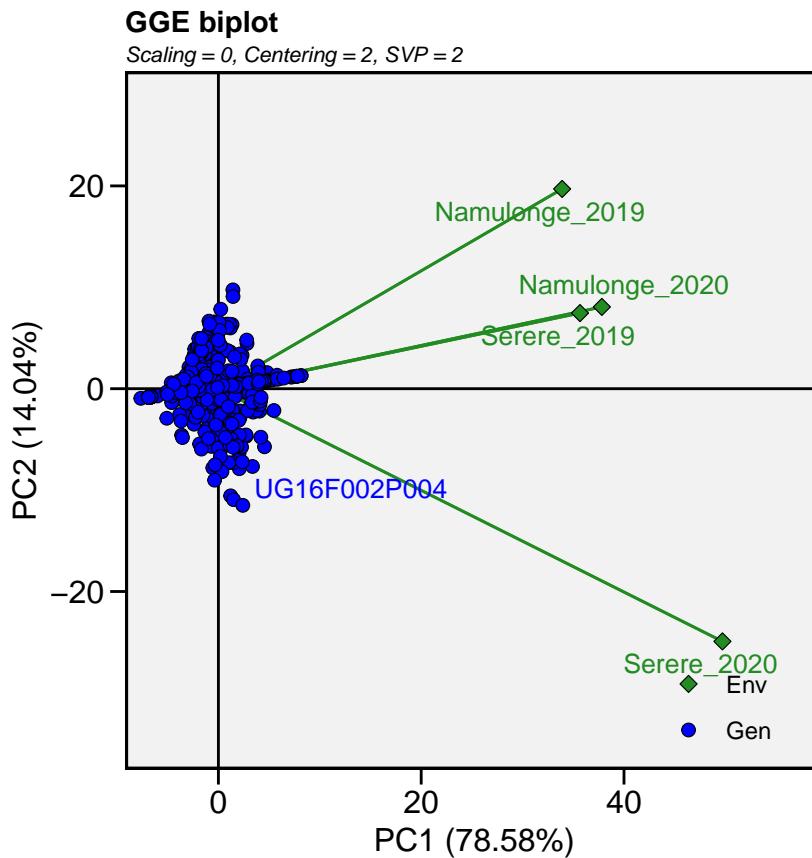
plot(GGEMODEL11)

```

```

## Warning: ggrepel: 403 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



```

#####broadsense heritability

```

```

#the mixed models were run on the server

```

```

#this is the code used

```

```

Explore R Script

```

```
parallel::detectCores() # get number of cores
n.cores <- parallel::detectCores() - 28 # select # required n.cores
```

## create cluster object

```
my.cluster <- parallel::makeCluster( n.cores, type = "PSOCK" )
doParallel::registerDoParallel(cl = my.cluster) # register cluster
foreach::getDoParRegistered() # check if running
foreach::getDoParWorkers() # verify worker

#run code or script

#automatic install of packages if they are not installed already
list.of.packages <- c( "foreach", "doParallel",
"ranger", "palmerpenguins", "tidyverse", "kableExtra", "data.table", "dplyr", "sommer" )

new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages) > 0){ install.packages(new.packages, dep=TRUE) }

#loading packages for(package.i in list.of.packages){ suppressPackageStartupMessages( library( package.i,
character.only = TRUE ) ) }

#this dataset has been log transformed to ensure normality
Full_rootnecrosis_data2023 <- read_csv("~/Desktop/Leah_images/Full_rootnecrosis_data2023.csv")
dim(Full_rootnecrosis_data2023)

#accounting for the augmented design

#closeAllConnections() data <- within(Full_rootnecrosis_data2023, { new <- ifelse(Full_rootnecrosis_data2023$Entry_type == 1, 999, data$germplasmName)}) str(data)
tail(data) #####create a unique Id Column that will be used to create a unique number
data1 %>% mutate(unique_ID = 1:n())

#####
#####create unique variable for plot number
data1Plot_number <- paste(data1$Location, data1$Year, data1$plot_number, sep = "_")
#####plant number
data1Plant_number <- paste(data1$plotNumber,data1$Plant_Num, sep = "_")
#####plant root number
data1Root_number <- paste(data1$Plant_Num,data1$Root_number, sep = "_")

#Factors are location, year, block plot, plant number, root number, image position
data1$Location <- as.factor(data1$Location) data1$Year <- as.factor(data1$Year) data1$block_number <- as.factor(data1$block_number)
data1$Image_number <- as.factor(data1$Image_number)

#random variables
data1$Plot_number <- as.factor(data1$plot_number) data1$Plant_number <- as.factor(data1$Plant_number)
data1$Root_number <- as.factor(data1$Root_number)
```

## list of traits

```

train_names <- c( "SN", "CHVN", "EAN", "EEN", "NECRO", "NAF", "NWF", "CBSDs3", "CBSDs6",
"CBSDs12")

foreach (i = train_names) %dopar% {

# load libraries library(nlme) library(dplyr) library(lme4) library(sommer)

Model1 <- lmer(get(i) ~ Location + Year + entryc + Image_number + (1|germplasmName:new) +
(1|Plot_number) + (1|block_number) + (1|plantnumber) + (1|rootnumber), data=data1)

saveRDS(Model1, paste(i,sep = "_.", "Model1.rds"))

}

parallel::stopCluster(cl = my.cluster) # stop cluster when done

##### loading mixed models from the server and determining broad sense heritability

#accounting for the augmented design

#closeAllConnections()
data <- within(FULL_ROOTNECROSIS_IMAGEDATA_2023 , {
  new <- ifelse(FULL_ROOTNECROSIS_IMAGEDATA_2023$Entry_type!="check",0,1)
});
tail(data)

```

```

##          X             Plot_id      SN      CHVN       EAN
## 46153 46153 C2_Serere_GWAS_CET_2019_99 -2.6222149  2.890372 5.170463968
## 46154 46154 C2_Serere_GWAS_CET_2019_99 -0.1447462  2.772589 4.973984965
## 46155 46155 C2_Serere_GWAS_CET_2019_99 -9.2103404 -9.210340 -9.210340372
## 46156 46156 C2_Serere_GWAS_CET_2019_99 -9.2103404 -9.210340 -9.210340372
## 46157 46157 C2_Serere_GWAS_CET_2019_99 -9.2103404 -9.210340 -9.210340372
## 46158 46158 C2_Serere_GWAS_CET_2019_99 -9.2103404 -9.210340 -9.210340372
##          EEN      NECRO      NAF      NWF Root_number Image_number
## 46153 -0.2335222 -0.6657755 -2.648731 -1.020406      Root1    image3
## 46154 -0.2033160 -1.5685173 -6.028941 -2.892200      Root2    image2
## 46155 -9.2103404 -9.2103404 -9.210340 -9.210340      Root1    image1
## 46156 -9.2103404 -9.2103404 -9.210340 -9.210340      Root1    image3
## 46157 -9.2103404 -9.2103404 -9.210340 -9.210340      Root1    image3
## 46158 -9.2103404 -9.2103404 -9.210340 -9.210340      Root1    image2
##          Plant_Num Location Year block_number plot_number germplasmName Entry_type
## 46153        7   Serere 2019         5        99 UG16F301P011      test
## 46154        1   Serere 2019         5        99 UG16F301P011      test
## 46155        4   Serere 2019         5        99 UG16F301P011      test
## 46156        4   Serere 2019         5        99 UG16F301P011      test
## 46157        1   Serere 2019         5        99 UG16F301P011      test
## 46158        1   Serere 2019         5        99 UG16F301P011      test
##          CBSDs3     CBSDs6 CBSDs12      ENV new
## 46153        0 0.1335314 1.55228 Serere_2019     0
## 46154        0 0.1335314 1.55228 Serere_2019     0
## 46155        0 0.1335314 1.55228 Serere_2019     0
## 46156        0 0.1335314 1.55228 Serere_2019     0
## 46157        0 0.1335314 1.55228 Serere_2019     0
## 46158        0 0.1335314 1.55228 Serere_2019     0

```

```

#checks are labeled 1 while test clones are zero
data1 <- within(data, {
  entryc <- ifelse(data$new == 1, 999,data$germplasmName)})
str(data1)

## 'data.frame':   46158 obs. of  24 variables:
##   $ X          : int  1 2 3 4 5 6 7 8 9 10 ...
##   $ Plot_id    : chr "2020_C2_CET_GWAS_Namulonge_323" "2020_C2_CET_GWAS_Namulonge_323" "2020_C2_CET...
##   $ SN         : num -9.21 -0.487 -9.21 -9.21 -9.21 ...
##   $ CHVN       : num -9.21 3.09 -9.21 -9.21 -9.21 ...
##   $ EAN         : chr "-9.210340372" "5.026214998" "-9.210340372" "-9.210340372" ...
##   $ EEN         : num -9.2103 -0.0485 -9.2103 -9.2103 -9.2103 ...
##   $ NECRO      : num -9.21 -0.623 -9.21 -9.21 -9.21 ...
##   $ NAF         : num -9.21 -4.74 -9.21 -9.21 -9.21 ...
##   $ NWF         : num -9.21 -1.86 -9.21 -9.21 -9.21 ...
##   $ Root_number: chr "Root2" "Root1" "Root2" "Root1" ...
##   $ Image_number: chr "image3" "image3" "image2" "image3" ...
##   $ Plant_Num  : int 5 5 5 2 5 5 2 2 5 10 ...
##   $ Location    : chr "Namulonge" "Namulonge" "Namulonge" "Namulonge" ...
##   $ Year        : int 2020 2020 2020 2020 2020 2020 2020 2020 ...
##   $ block_number: int 12 12 12 12 12 12 12 12 12 ...
##   $ plot_number : int 323 323 323 323 323 323 323 323 324 ...
##   $ germplasmName: chr "UG16F305P003" "UG16F305P003" "UG16F305P003" "UG16F305P003" ...
##   $ Entry_type  : chr "test" "test" "test" "test" ...
##   $ CBSDs3     : num 0.182 0.182 0.182 0.182 0.182 ...
##   $ CBSDs6     : num 0.531 0.531 0.531 0.531 0.531 ...
##   $ CBSDs12    : num 0.329 0.329 0.329 0.329 0.329 ...
##   $ ENV         : chr "Namulonge_2020" "Namulonge_2020" "Namulonge_2020" "Namulonge_2020" ...
##   $ new         : num 0 0 0 0 0 0 0 0 0 ...
##   $ entryc     : chr "UG16F305P003" "UG16F305P003" "UG16F305P003" "UG16F305P003" ...

####create unique variables

data1$Plot_number <- paste(data1$Location, data1$Year, data1$plot_number, sep = "_")
#####plant number
data1$plantnumber <- paste(data1$Plot_number, data1$Plant_Num, sep = "_")
#####plant root number
data1$rootnumber <- paste(data1$plantnumber, data1$Root_number, sep = "_")

####determining harmonic means

##plot harmonic mean
unique_varieties <- unique(data1$germplasmName)

# Function to figure out how many plots each variety has been evaluated in
plotsPerVariety <- function(varName){
  dataVar <- data1 %>% dplyr::filter(germplasmName == varName)
  unique_plots <- unique(dataVar$Plot_number)
  return(length(unique_plots))
}
plotsPerVar <- sapply(unique_varieties, plotsPerVariety)

```

```

# Calculate the harmonic mean of plotsPerVar for var of plots
plantsPerVariety <- function(varName){
  dataVar <- data1 %>% dplyr::filter(germplasmName == varName)
  unique_plants <- unique(dataVar$plantnumber)
  return(length(unique_plants))
}

plantsPerVar <- sapply(unique_varieties, plantsPerVariety)

# Calculate the harmonic mean of plantsPerVar for var of plants within plots
rootsPerVariety <- function(varName){
  dataVar <- data1 %>% dplyr::filter(germplasmName == varName)
  unique_roots <- unique(dataVar$rootnumber)
  return(length(unique_roots))
}
rootsPerVar <- sapply(unique_varieties, rootsPerVariety)

# Calculate the harmonic mean of rootsPerVar for var of roots within plants
imagesPerVar <- table(data1$germplasmName)

# Calculate the harmonic mean of imagesPerVar for residual
imagesPerVar <- table(data1$germplasmName)

##plot harmonic mean

plot_harmonic_mean <- 1 / mean(1 / plotsPerVar)

plot_harmonic_mean

## [1] 1.762977

##plant harmonic mean

plant_harmonic_mean <- 1 / mean(1 / plantsPerVar)

plant_harmonic_mean

## [1] 3.708938

##Root harmonic mean

Root_harmonic_mean <- 1 / mean(1 / rootsPerVar)

Root_harmonic_mean

## [1] 5.485583

```

```

##image harmonic mean / residual variance

image_harmonic_mean <- 1/ mean(1/imagesPerVar)

image_harmonic_mean

## [1] 7.775424

#####mixed models and broadsense heritability

#SN
one <- readRDS("~/Downloads/oct23_v2/SN_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r

Broad_sense <- 2.44523 / (2.44523 + 2.576121/plot_harmonic_mean + 0.00648/plant_harmonic_mean + 0.01417
Broad_sense #0.4498814

## [1] 0.4498814

#NWF
two <- readRDS("~/Downloads/oct23_v2/NWF_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r

Broad_sense2 <- 3.010467 / (3.010467 + 4.221826/plot_harmonic_mean + 0.006726 /plant_harmonic_mean + 0
Broad_sense2 # 0.4260881

## [1] 0.4260881

#NECRO
three <- readRDS("~/Downloads/oct23_v2/NECRO_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r

Broad_sense3 <- 5.47955 / (5.47955 + 5.64765/plot_harmonic_mean + 0.01160 /plant_harmonic_mean + 0
Broad_sense3 # 0.5022727

## [1] 0.5022727

#NAF
four <- readRDS("~/Downloads/oct23_v2/NAF_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r

Broad_sense4 <- 2.766034 / (2.766034 + 3.532798/plot_harmonic_mean + 0.005521 /plant_harmonic_mean +
Broad_sense4 # 0.4652036

```

```

## [1] 0.4652036

#EEN
five <- readRDS("~/Downloads/oct23_v2/EEN_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r
Broad_sense5 <- 2.91513 / (2.91513 + 4.33485/plot_harmonic_mean + 0.00999 /plant_harmonic_mean + 0.
Broad_sense5 # 0.3965869

## [1] 0.3965869

#EAN
six <- readRDS("~/Downloads/oct23_v2/EAN_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r
Broad_sense6 <- 6.45458 / (6.45458 + 9.68386/plot_harmonic_mean + 0.02437 /plant_harmonic_mean + 0.03
Broad_sense6 # 0.3943223

## [1] 0.3943223

#CHVN
seven <- readRDS("~/Downloads/oct23_v2/CHVN_Model1.rds")

#Broad_sense <- geno_variance / (geno_variance + plot_variance + plant_variance/harmonicmean_plants + r
Broad_sense7 <- 5.72242/ (5.72242 + 8.24094 /plot_harmonic_mean + 0.01664 /plant_harmonic_mean + 0.0334
Broad_sense7 # 0.4079832

## [1] 0.4079832

#####deregressed BLUPs for GEMMA GWAS analysis

BLUP <- ranef(one , condVar=TRUE)$`germplasmName:new`
PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(one),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blups
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_SN.csv")

```

```

BLUP <- ranef(two , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(two),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blups
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_NWF.csv")

```

```

BLUP <- ranef(three , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(three),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blups
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_NECKO.csv")

```

```

BLUP <- ranef(four , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(four),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blups
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_NECKO.csv")

```

```

BLUP <- ranef(four , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(four),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blups
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_NECKO.csv")

```

```

BLUP <- ranef(five , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(five),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blupS
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_NAF.csv")

```

```

BLUP <- ranef(six , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(six),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blupS
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_EAN.csv")

```

```

BLUP <- ranef(seven , condVar=TRUE)$`germplasmName:new`  

PEV <- c(attr(BLUP, "postVar"))
VARCOMPS <- as.data.frame(VarCorr(seven),comp="Variance")
CLONE <- VARCOMPS[2,4]
ResidVar <- (VARCOMPS[4,4])
#deregressing OF blupS
out <- BLUP/(1-(PEV/CLONE))

write.csv(out, file = "Deregressed_Blups_CHVN.csv")

```

```

#####GWAS USING GEMMA
#####DEREGRESSED blupS WERE USED GWAS
scp ln242@login.sgn.cornell.edu:/home/ln242/*_march2023.* .
March 2023
##-- Run GEMMA for Image phenotypes ----##

```

**keep only samples with phenotype and genotype**

**GEMMA\_Rootnecrosis\_pheno.csv to GEMMA\_Rootnecrosis\_pheno\_sub298.csv**

**map image phenotype to the .fam pheno input file from previous GEMMA analysis .fam file**

combinedC2\_final\_PLINK\_Mar2023.fam

**copy phenotype file to the server**

```
scp ~/Desktop/Leah_image_analysis/GEMMA_Mar2023/combinedC2_final_PLINK_Mar2023.fam.txt  
ln242@login.sgn.cornell.edu:/home/ln242/  
scp ln242@solanine:/home/ln242/combinedC2_final_PLINK_Mar2023.fam.txt .  
mv combinedC2_final_PLINK_Mar2023.fam.txt combinedC2_final_2_PLINK.fam
```

**SWITCH directory to “/home/ln242/gemma2/genetic-data-analysis-2” to carry out GEMMA analysis**

```
cd /home/ln242/gemma2/genetic-data-analysis-2
```

**run GEMMA**

```
#generate centered relationship matrix# ./gemma -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK  
-gk 1 -o combinedC2_final_2_PLINK_image_Mar2023
```

```
Reading Files ... ## number of total individuals = 3194 ## number of analyzed individuals  
= 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs =  
51860 ## number of analyzed SNPs = 30750 Calculating Relatedness Matrix ... Reading SNPs  
=====100.00% #####
```

```
#run GEMMA # ./gemma -bfile ../../GEMMA-0.98.4/combinedC2_final_PLINK -k output/combinedC2_final_2_PLINK  
-lmm 2 -o combinedC2_final_PLINK_GEMMA #
```

univariate GWAS

```
###UNIVARIATE GWAS
```

#for CHVN

```
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k out-  
put/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 3 -o combinedC2_final_2_PLINK_GEMMA_mar2
```

```
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##  
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of co-  
variates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of  
analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =3.23627e-06 se(pve) =0.0675396  
=====100% **** INFO:  
Done.
```

#for EAN

```

./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 4 -o combinedC2_final_2_PLINK_GEMMA_mar2
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.0498661 se(pve) =0.0818999
=====
100% **** INFO: Done.

#for EEN
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 5 -o combinedC2_final_2_PLINK_GEMMA_mar2
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.0327276 se(pve) =0.0501022
=====
100% **** INFO: Done.

#for NAF
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 10 -o combinedC2_final_2_PLINK_GEMMA_mar
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =3.23627e-06 se(pve) =0.0711944
=====
100% **** INFO: Done.

#for NECRO
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 11 -o combinedC2_final_2_PLINK_GEMMA_mar
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.0136163 se(pve) =0.0424404
=====
100% **** INFO: Done.

#for NWF
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 12 -o combinedC2_final_2_PLINK_GEMMA_mar
#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.00310162 se(pve) =0.040102
=====
100% **** INFO: Done.

#for SN
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 14 -o combinedC2_final_2_PLINK_GEMMA_mar

```

```

#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.0419753 se(pve) =0.105102
=====
===== 100% **** INFO: Done.

#for CBSDs3
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 16 -o combinedC2_final_2_PLINK_GEMMA_mar

#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =0.0756482 se(pve) =0.0858772
=====
===== 100% **** INFO: Done.

#for CBSDs6
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 17 -o combinedC2_final_2_PLINK_GEMMA_mar

#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =3.23627e-06 se(pve) =0.046063
=====
===== 100% **** INFO: Done.

#for CBSDs12
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_Mar2023.cXX.txt -lmm 2 -n 18 -o combinedC2_final_2_PLINK_GEMMA_mar

#log GEMMA 0.98.4 (2021-01-29) by Xiang Zhou and team (C) 2012-2021 Reading Files ... ##
number of total individuals = 3194 ## number of analyzed individuals = 320 ## number of covariates = 1 ## number of phenotypes = 1 ## number of total SNPs/var = 51860 ## number of analyzed SNPs = 30750 Start Eigen-Decomposition... pve estimate =3.23627e-06 se(pve) =0.075352
=====
===== 100% **** INFO: Done.

```

## copy file

```

scp output/mar23 ln242@solanine:/home/ln242
scp ln242@login.sgn.cornell.edu:/home/ln242/mar23 .
Multivariate gwas
#— for multi-traits —#

```

## CHVN\_NWF\_NAF

```

./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_May2023.cXX.txt -lmm 2 -n 1 2 14 9 3 7 5 11 -o combinedC2_final_2_PLINK_GEMMA_CHVN_NWF_NAF_may2023

```

## EAN\_NECRO

```
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_May2023.cXX.txt -lmm 2 -n 14 3 7 12 -o combinedC2_final_2_PLINK_GEMMA
```

## NWF\_NAF

```
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_May2023.cXX.txt -lmm 2 -n 5 13 -o combinedC2_final_2_PLINK_GEMMA_NW
```

## NECRO\_CBSDs3\_CBSDs6

```
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_May2023.cXX.txt -lmm 2 -n 15 8 6 4 11 -o combinedC2_final_2_PLINK_GEMMA_N
```

## NECRO\_CBSDs12

```
./gemma-0.98.4-linux-static-AMD64 -bfile ../../GEMMA-0.98.4/combinedC2_final_2_PLINK -k output/combinedC2_final_2_PLINK_image_May2023.cXX.txt -lmm 2 -n 6 4 11 -o combinedC2_final_2_PLINK_GEMMA_N  
##-- copy results --# zip -r output_may2023.zip output  
scp output_may2023.zip ln242@solanine:/home/ln242  
#copy result from local machine scp ln242@login.sgn.cornell.edu:/home/ln242/output_may2023.zip .
```

## Plotting Manhattan plots -----

```
#Previous calculation of MEFF effective number of markers  
Meff = 6477  
p_threshold = (1 - (1 - 0.05))1/6477  
p_threshold #7.719623e-06  
  
## [1] 7.719623e-06  
  
# Plotting Manhattan plots -----  
  
colR <- c("blue4", "orange3", "blue4", "orange3", "blue4", "orange3", "blue4", "orange3", "blue4", "orange3")  
  
# QQPLOTS -----  
  
gg_qqplot <- function(P, ci = 0.95) {  
  n <- length(P)  
  df <- data.frame(  
    observed = -log10(sort(P)),  
    expected = -log10(ppoints(n)),  
    clower = -log10(qbeta(p = (1 - ci) / 2, shape1 = 1:n, shape2 = n:1)),
```

```

    cupper = -log10(qbeta(p = (1 + ci) / 2, shape1 = 1:n, shape2 = n:1))
)
log10Pe <- expression(paste("Expected -log"[10], plain(P)))
log10Po <- expression(paste("Observed -log"[10], plain(P)))
ggplot(df) +
  geom_point(aes(expected, observed), shape = 1, size = 3) +
  geom_abline(intercept = 0, slope = 1, alpha = 0.5) +
  xlab(log10Pe) +
  ylab(log10Po)
}

```

```

gwscan1 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_CBSDs3.assoc.txt", header = TRUE)
summary(gwscan1)

```

```

##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708   Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844 1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006 3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0         af      logl_H1
##  Length:29708  Length:29708   Min.   :0.0100  Min.   :8197
##  Class  :character  Class  :character  1st Qu.:0.0760 1st Qu.:8197
##  Mode   :character  Mode   :character  Median :0.1810  Median :8197
##                      Mean   :0.2101  Mean   :8197
##                      3rd Qu.:0.3260 3rd Qu.:8197
##                      Max.   :0.7370  Max.   :8210
##      l_mle        p_lrt
##  Min.   :0.000010  Min.   :0.0000002
##  1st Qu.:0.000010  1st Qu.:0.2813864
##  Median :0.000010  Median :0.5341052
##  Mean   :0.002203  Mean   :0.5208212
##  3rd Qu.:0.000010  3rd Qu.:0.7635283
##  Max.   :0.410631  Max.   :0.9999985

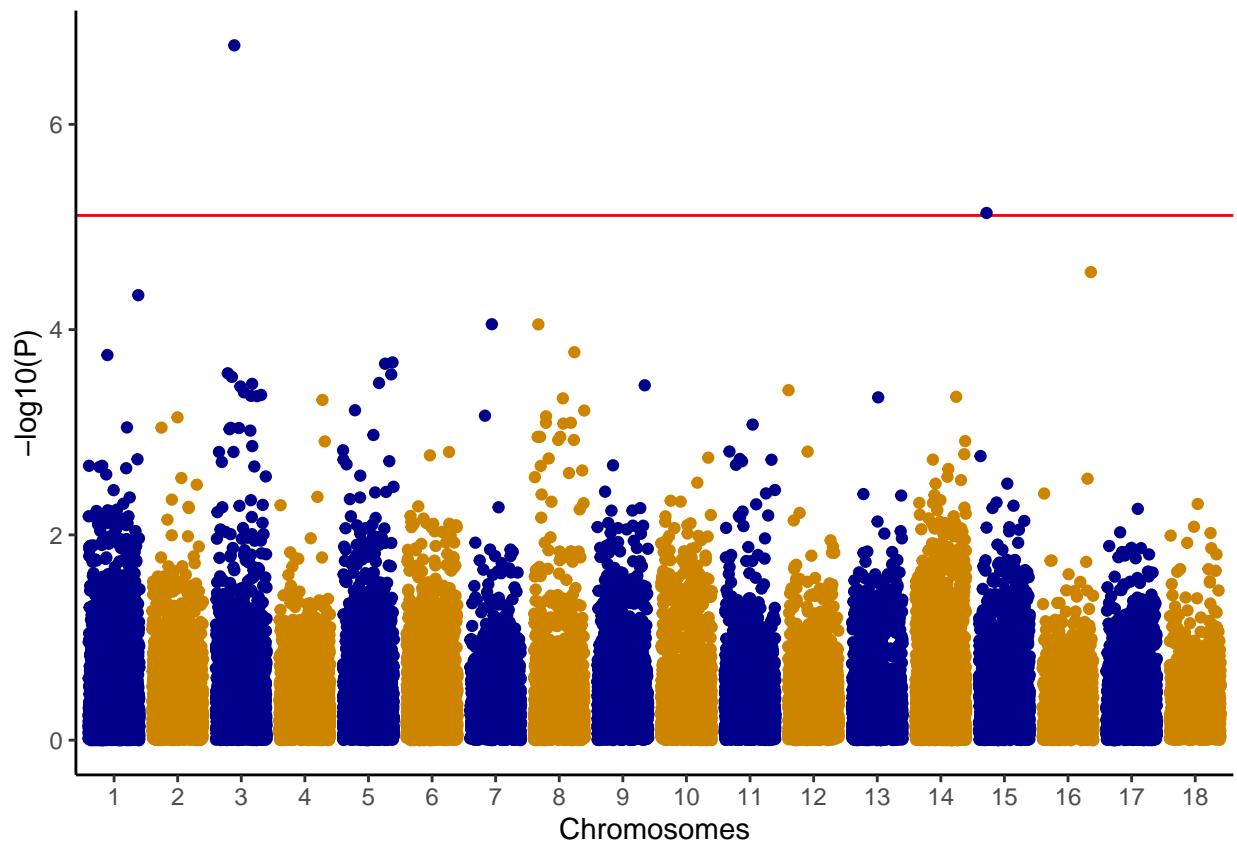
```

```

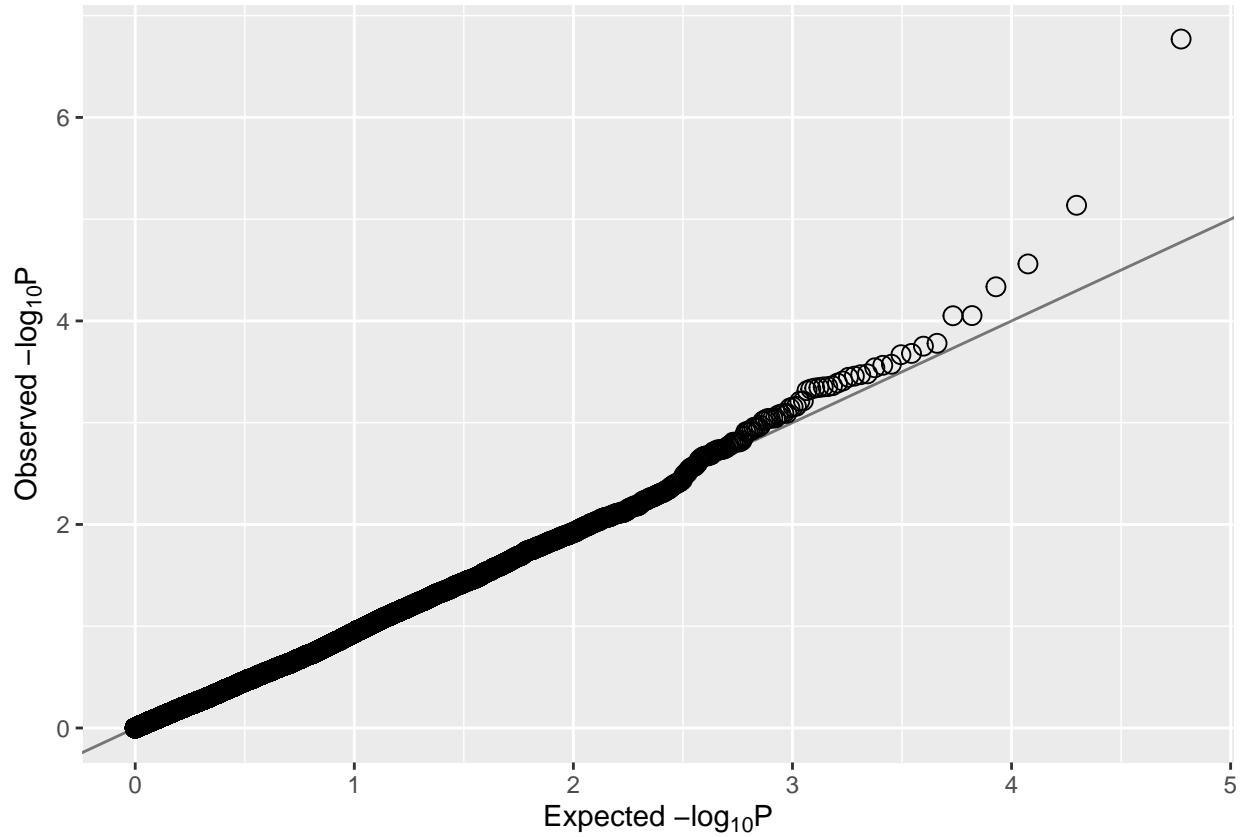
df1 <- gwscan1[,c(1,3,10,2)]
colnames(df1) <- c("CHR", "BP", "P", "SNP")
df1$BP <- as.numeric(df1$BP)
df1$CHR <- as.numeric(df1$CHR)

ggplot(data = df1,
       aes(x = as.factor(CHR),
            y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color = colR)
geom_jitter() +
  labs(x = "Chromosomes") +
  theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df1$P)
```



```
gwscan2 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_CBSDs6.assoc.txt", header=TRUE)
summary(gwscan2)
```

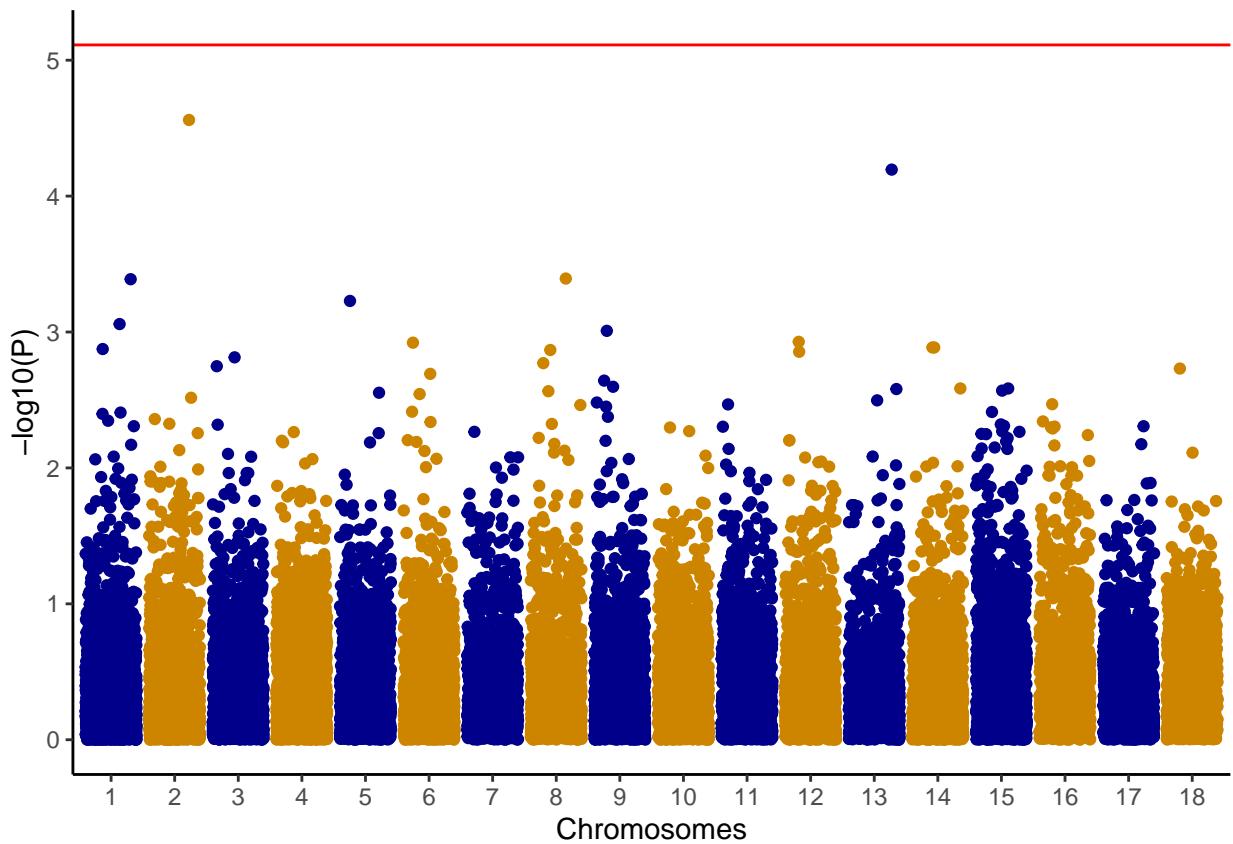
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                  3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                  Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :8102
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:8102
##  Mode   :character  Mode   :character  Median :0.1810  Median :8102
##                      Mean   :0.2101  Mean   :8102
##                      3rd Qu.:0.3260  3rd Qu.:8102
##                      Max.   :0.7370  Max.   :8111
##      l_mle          p_lrt
##  Min.   :1.000e-05  Min.   :0.0000275
##  1st Qu.:1.000e-05  1st Qu.:0.2971398
##  Median :1.000e-05  Median :0.5460620
##  Mean   :4.098e-05  Mean   :0.5339637
##  3rd Qu.:1.000e-05  3rd Qu.:0.7749813
##  Max.   :6.626e-02  Max.   :0.9999957
```

```

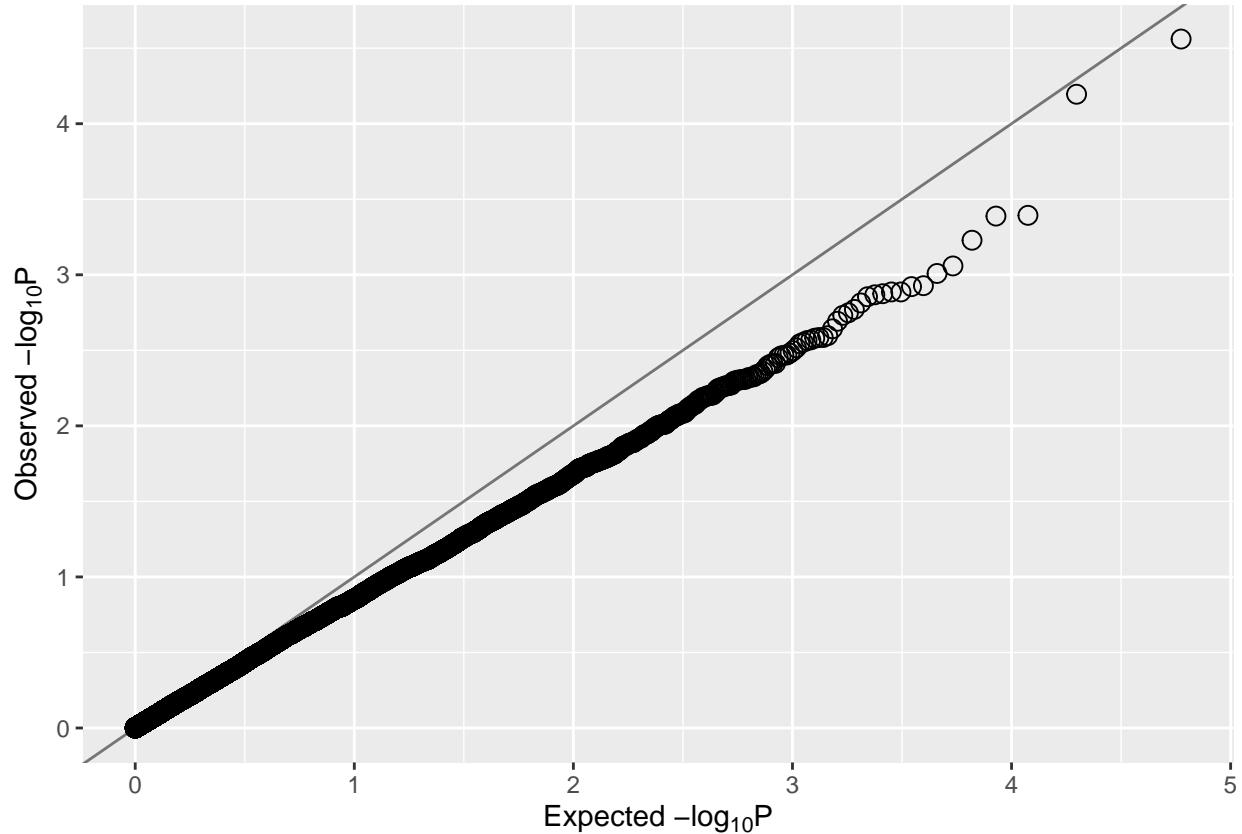
df2 <- gwscan2[,c(1,3,10,2)]
colnames(df2) <- c("CHR", "BP", "P", "SNP")
df2$BP <- as.numeric(df2$BP)
df2$CHR <- as.numeric(df2$CHR)

ggplot(data = df2,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df2$P)
```



```
gwscan3 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_CBSDs12.assoc.txt", head=TRUE)
summary(gwscan3)
```

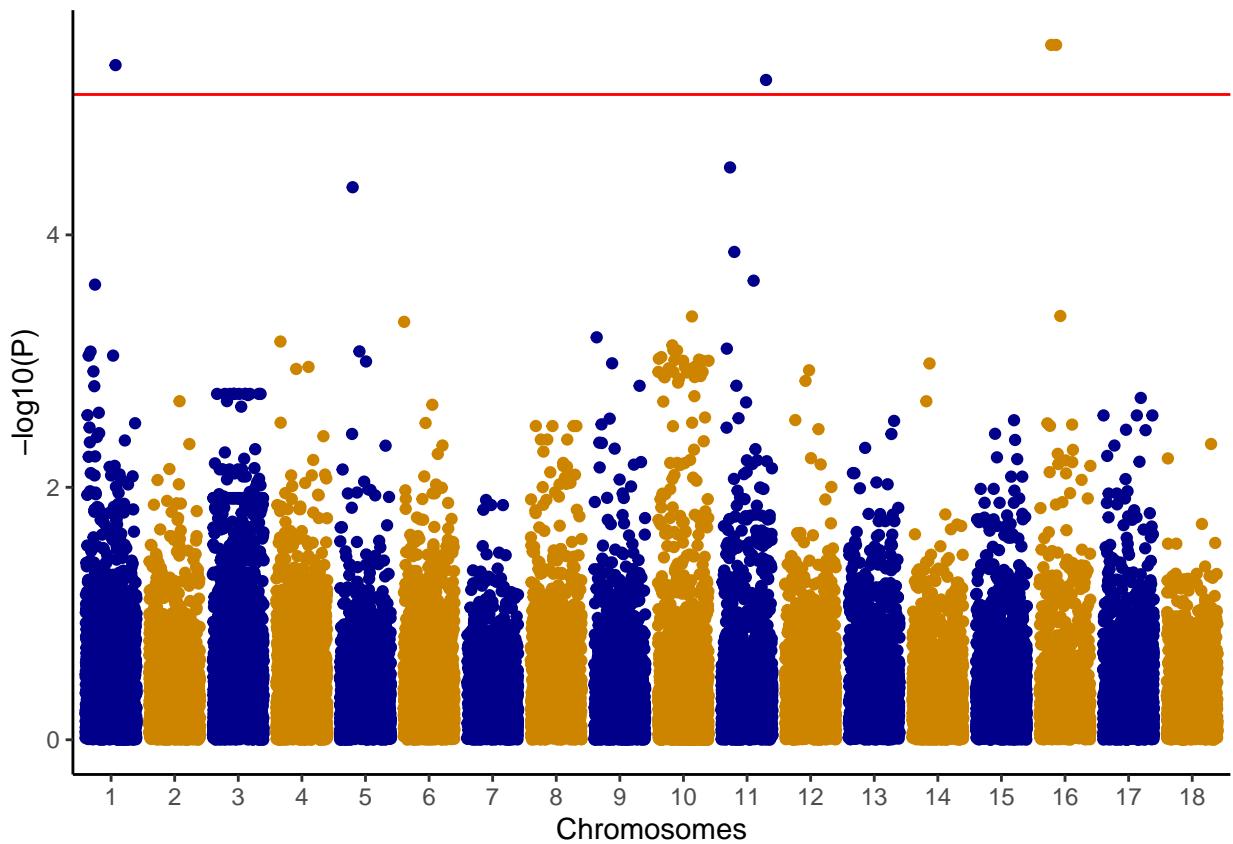
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :6747
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:6747
##  Mode   :character  Mode   :character  Median :0.1810  Median :6747
##                           Mean   :0.2101  Mean   :6747
##                           3rd Qu.:0.3260  3rd Qu.:6747
##                           Max.   :0.7370  Max.   :6758
##      l_mle          p_lrt
##  Min.   :0.0000100  Min.   :0.0000031
##  1st Qu.:0.0000100  1st Qu.:0.2816577
##  Median :0.0000100  Median :0.5171842
##  Mean   :0.0002305  Mean   :0.5150665
##  3rd Qu.:0.0000100  3rd Qu.:0.7544500
##  Max.   :0.1413707  Max.   :0.9997442
```

```

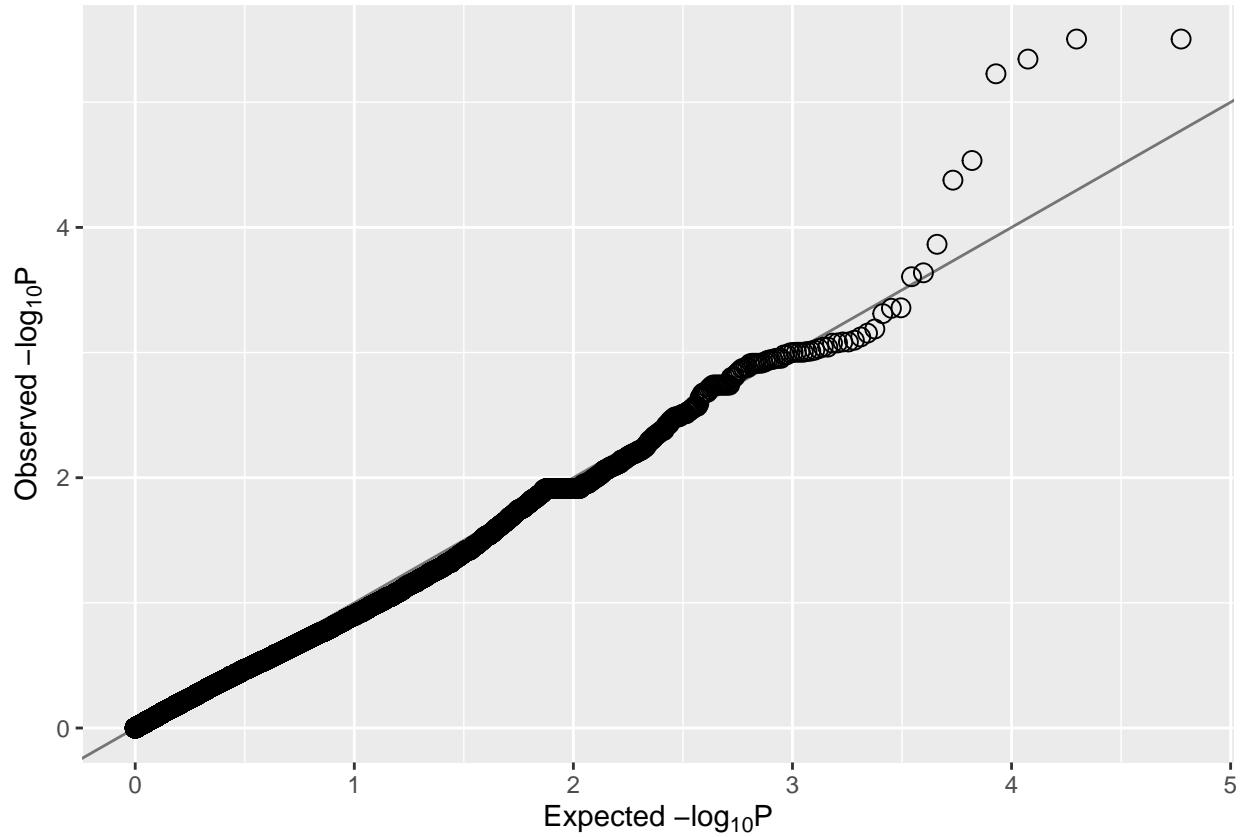
df3 <- gwscan3[,c(1,3,10,2)]
colnames(df3) <- c("CHR", "BP", "P", "SNP")
df3$BP <- as.numeric(df3$BP)
df3$CHR <- as.numeric(df3$CHR)

ggplot(data = df3,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df3$P)
```



```
gwscan4 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_CHVN.assoc.txt", header = TRUE)
summary(gwscan4)
```

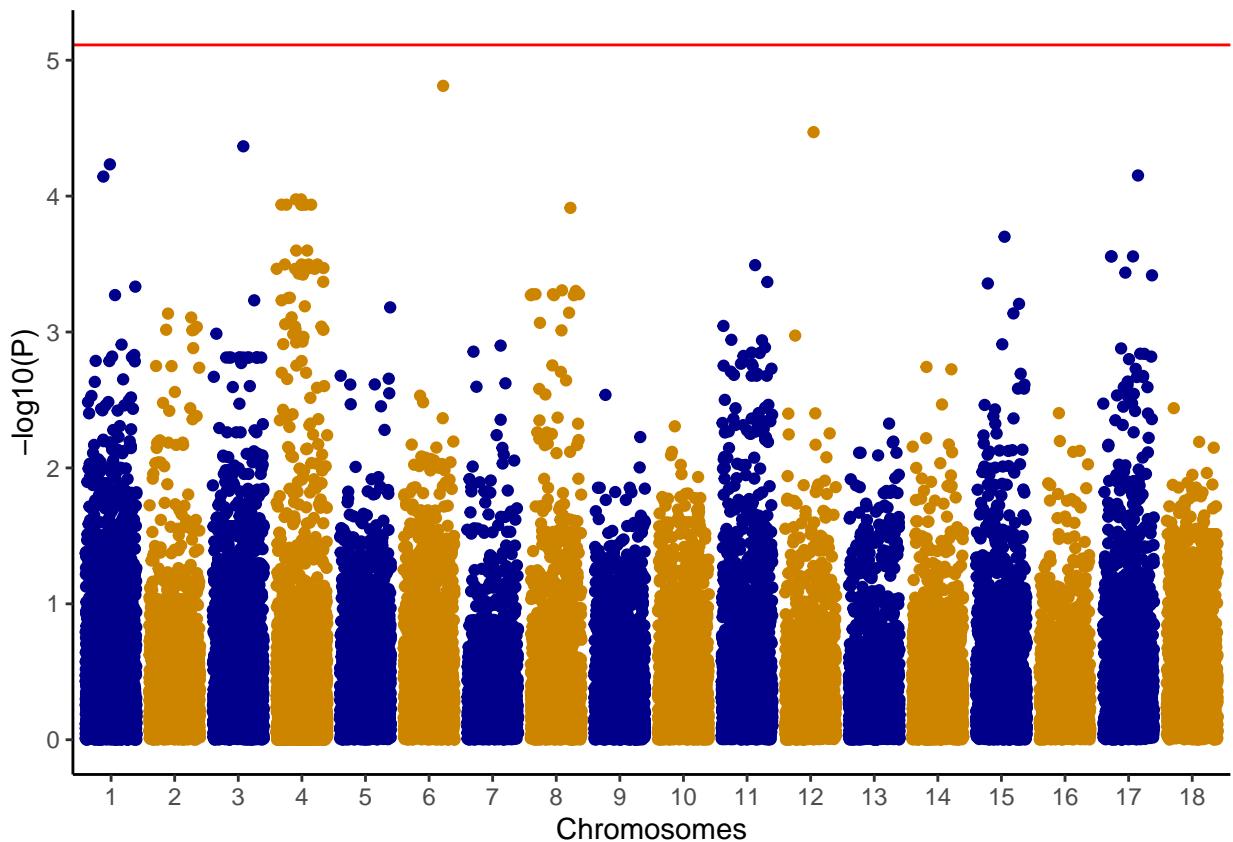
	chr	rs	ps	n_miss
##	Min. : 1.000	Length:29708	Min. : 7103	Min. :0
##	1st Qu.: 4.000	Class :character	1st Qu.: 5291844	1st Qu.:0
##	Median : 9.000	Mode :character	Median :16448302	Median :0
##	Mean : 8.721		Mean :16909538	Mean :0
##	3rd Qu.:13.000		3rd Qu.:28000006	3rd Qu.:0
##	Max. :18.000		Max. :40505410	Max. :0
##	allele1	allele0	af	logl_H1
##	Length:29708	Length:29708	Min. :0.0100	Min. :7840
##	Class :character	Class :character	1st Qu.:0.0760	1st Qu.:7840
##	Mode :character	Mode :character	Median :0.1810	Median :7840
##			Mean :0.2101	Mean :7840
##			3rd Qu.:0.3260	3rd Qu.:7841
##			Max. :0.7370	Max. :7849
##	l_mle	p_lrt		
##	Min. :0.00001	Min. :0.0000154		
##	1st Qu.:0.03267	1st Qu.:0.2368962		
##	Median :0.04895	Median :0.4926905		
##	Mean :0.04268	Mean :0.4934787		
##	3rd Qu.:0.05331	3rd Qu.:0.7513334		
##	Max. :0.25037	Max. :0.9999925		

```

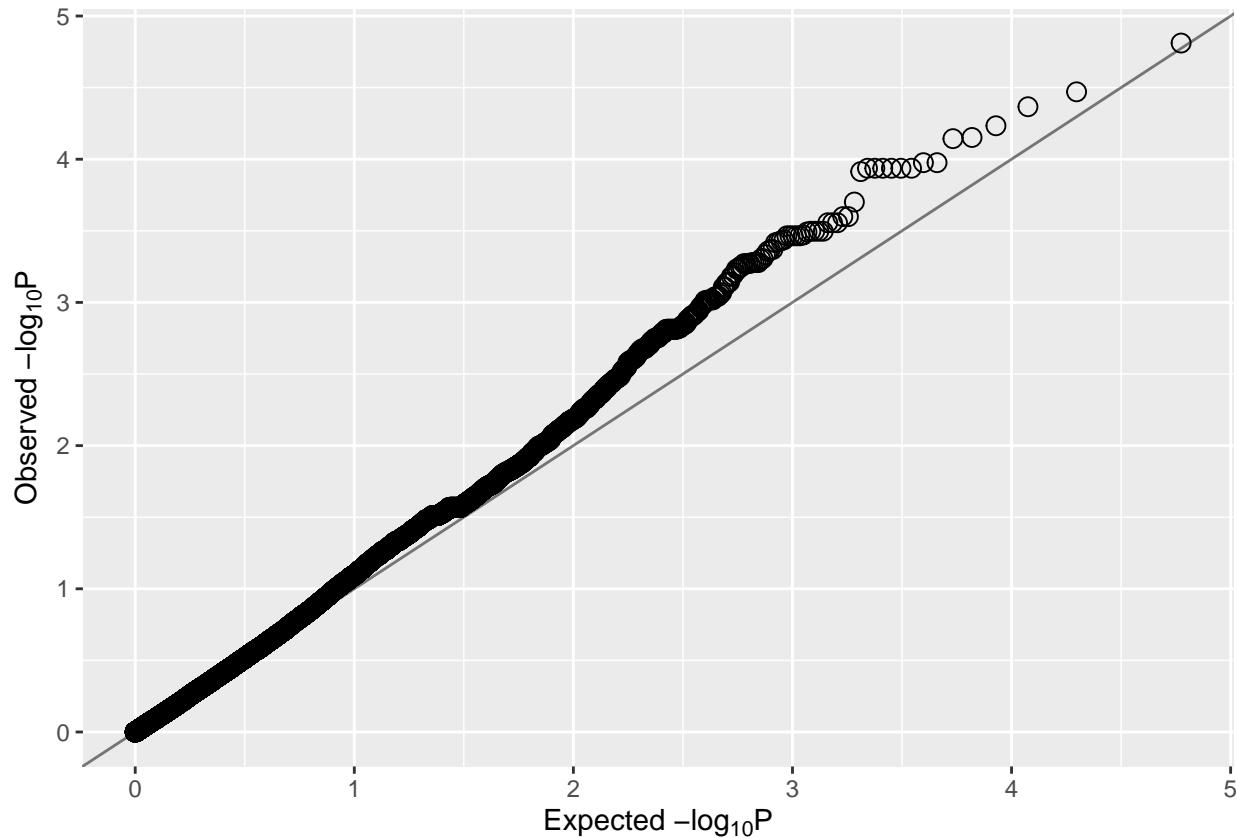
df4 <- gwscan4[,c(1,3,10,2)]
colnames(df4) <- c("CHR", "BP", "P", "SNP")
df4$BP <- as.numeric(df4$BP)
df4$CHR <- as.numeric(df4$CHR)

ggplot(data = df4,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df4$P)
```



```
gwscan5 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_EEN.assoc.txt", header = TRUE)
summary(gwscan5)
```

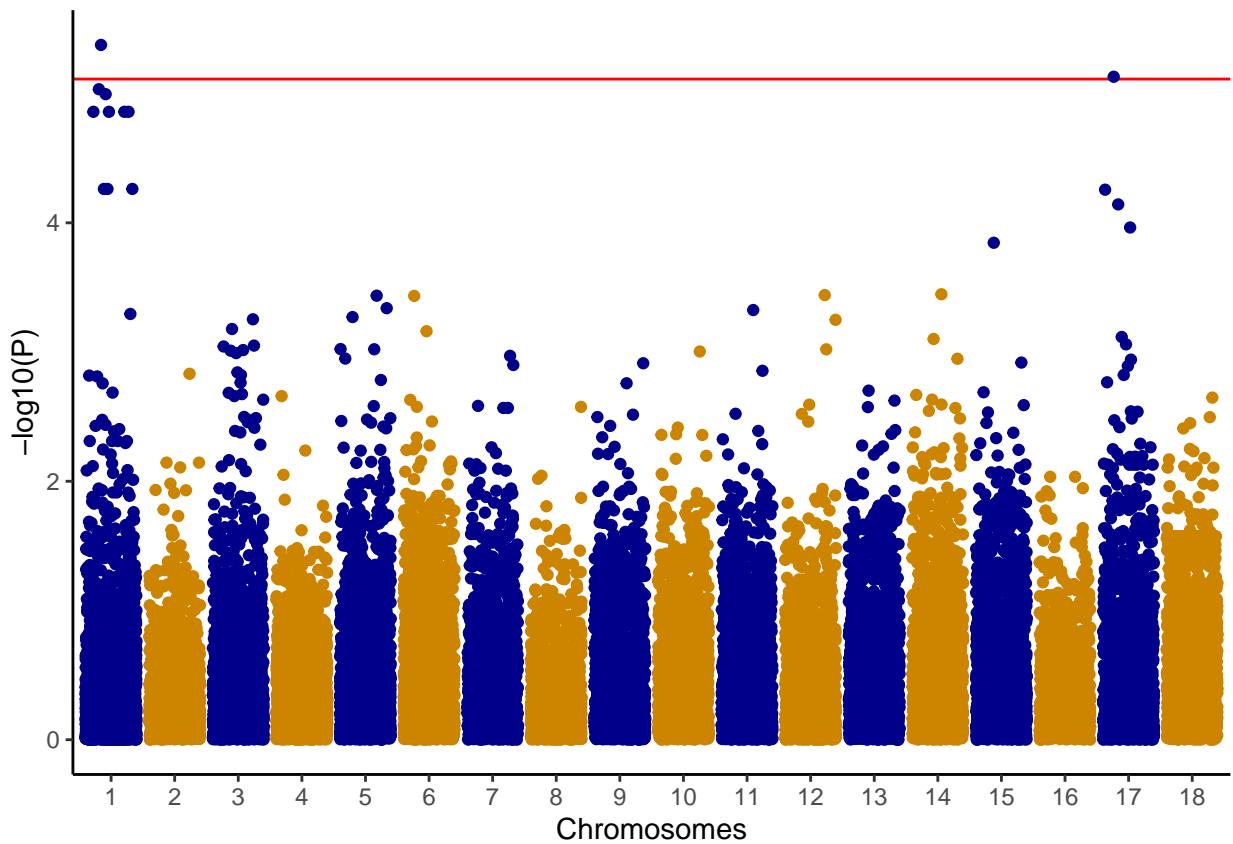
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :7714
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:7714
##  Mode   :character  Mode   :character  Median :0.1810  Median :7714
##                           Mean   :0.2101  Mean   :7714
##                           3rd Qu.:0.3260  3rd Qu.:7714
##                           Max.   :0.7370  Max.   :7724
##      l_mle          p_lrt
##  Min.   :0.000001  Min.   :0.0000042
##  1st Qu.:0.01036   1st Qu.:0.2254679
##  Median :0.02970   Median :0.4775745
##  Mean   :0.02733   Mean   :0.4847515
##  3rd Qu.:0.03638   3rd Qu.:0.7416849
##  Max.   :0.22253   Max.   :0.9999908
```

```

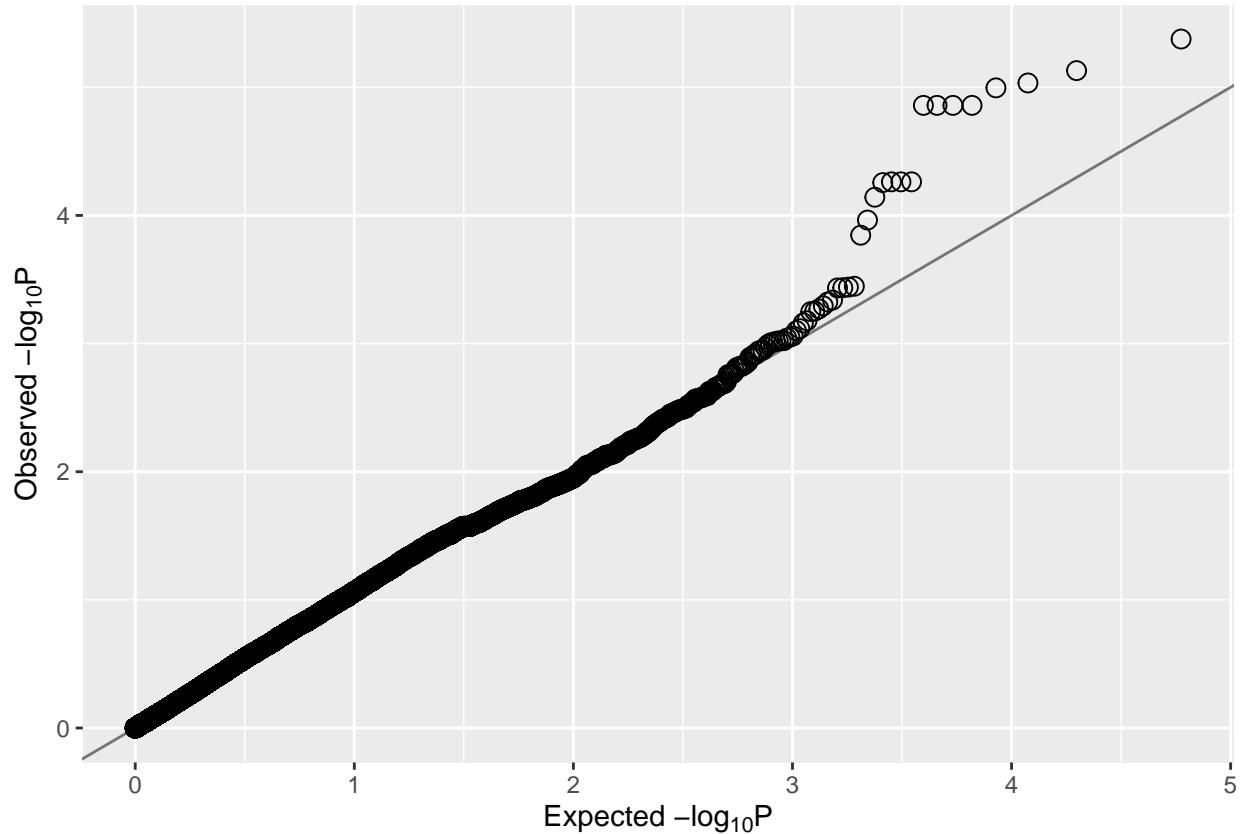
df5 <- gwscan5[,c(1,3,10,2)]
colnames(df5) <- c("CHR", "BP", "P", "SNP")
df5$BP <- as.numeric(df5$BP)
df5$CHR <- as.numeric(df5$CHR)

ggplot(data = df5,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df5$P)
```



```
gwscan6 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_EAN.assoc.txt", header = TRUE)
summary(gwscan6)
```

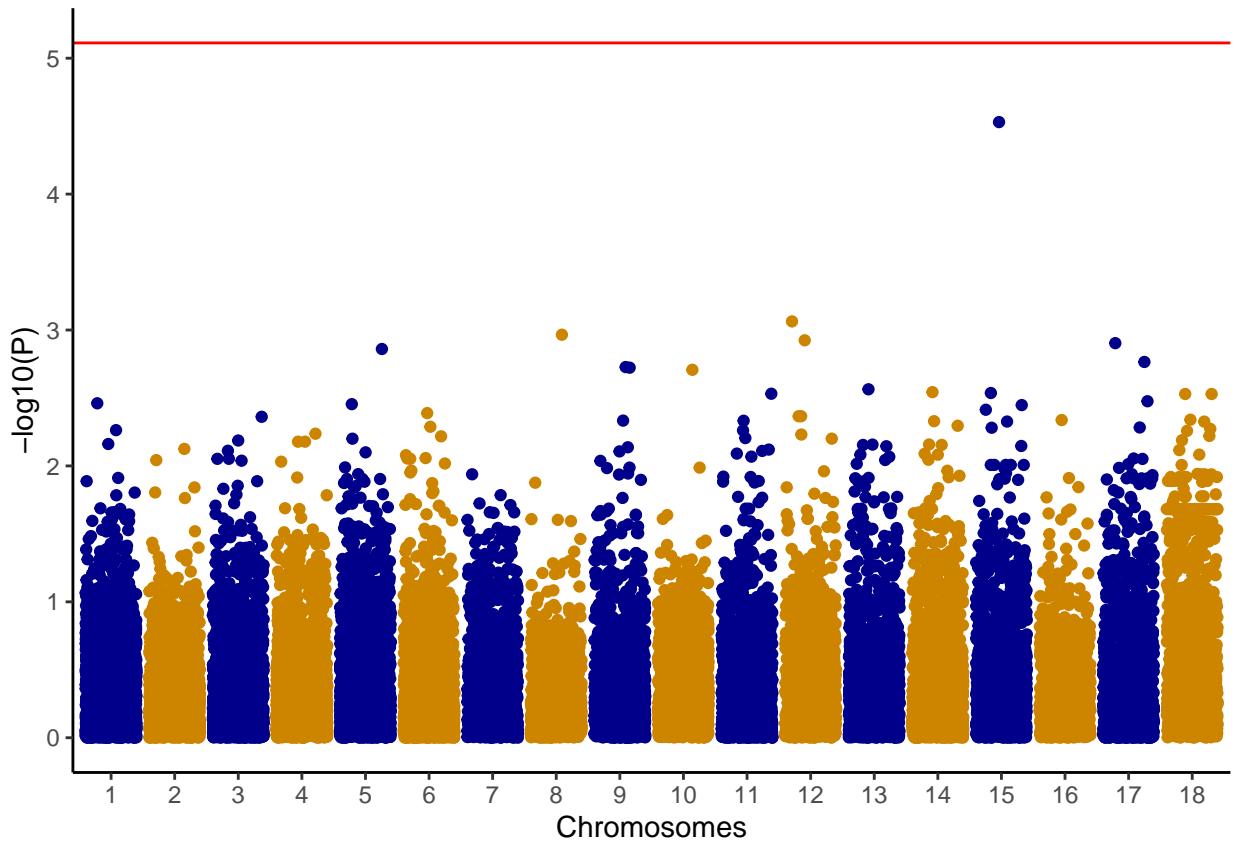
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0         af      logl_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :7549
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:7549
##  Mode   :character  Mode   :character  Median :0.1810  Median :7550
##                               Mean   :0.2101  Mean   :7550
##                               3rd Qu.:0.3260  3rd Qu.:7550
##                               Max.   :0.7370  Max.   :7558
##      l_mle          p_lrt
##  Min.   :1.000e-05  Min.   :0.0000295
##  1st Qu.:1.000e-05  1st Qu.:0.2906205
##  Median :1.000e-05  Median :0.5333822
##  Mean   :2.449e-05  Mean   :0.5283553
##  3rd Qu.:1.000e-05  3rd Qu.:0.7730396
##  Max.   :8.500e-02  Max.   :0.9999202
```

```

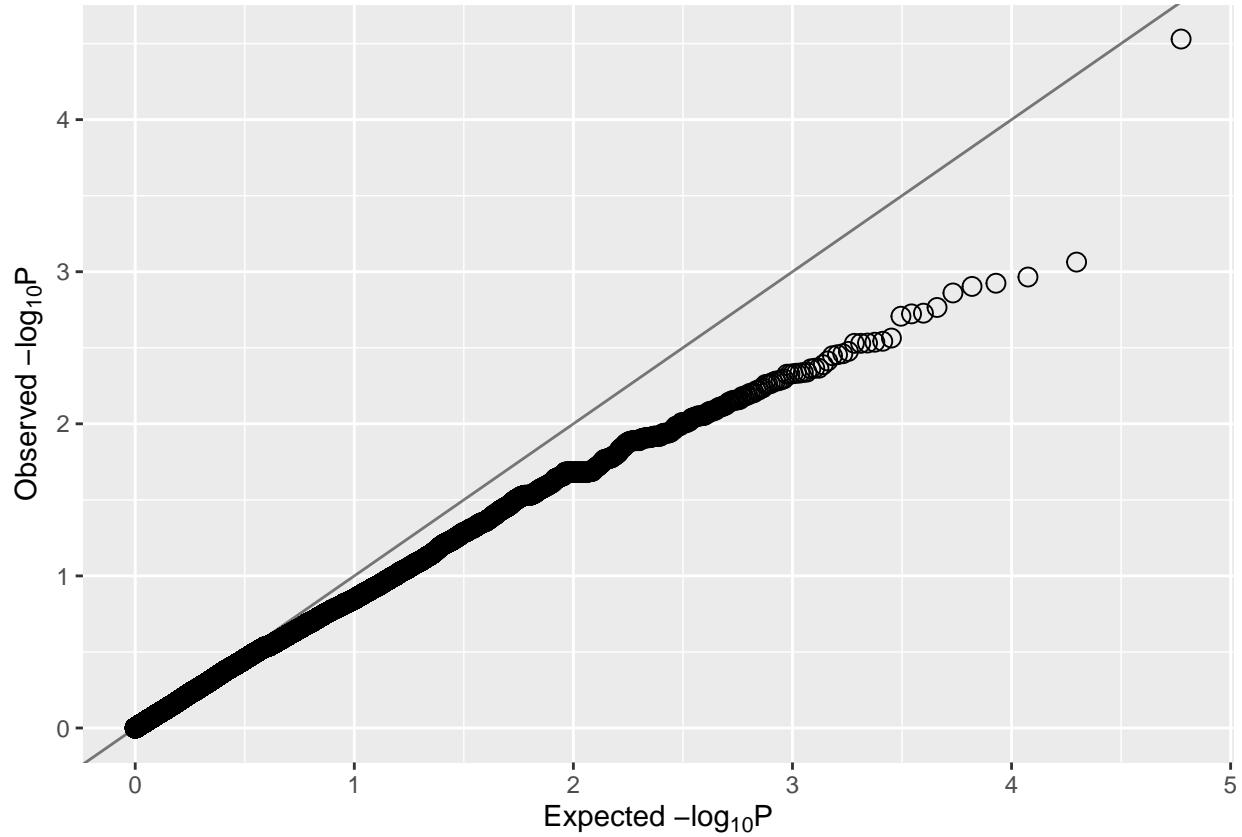
df6 <- gwscan6[,c(1,3,10,2)]
colnames(df6) <- c("CHR", "BP", "P", "SNP")
df6$BP <- as.numeric(df6$BP)
df6$CHR <- as.numeric(df6$CHR)

ggplot(data = df6,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df6$P)
```



```
gwscan7 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_NAF.assoc.txt", header = TRUE)
summary(gwscan7)
```

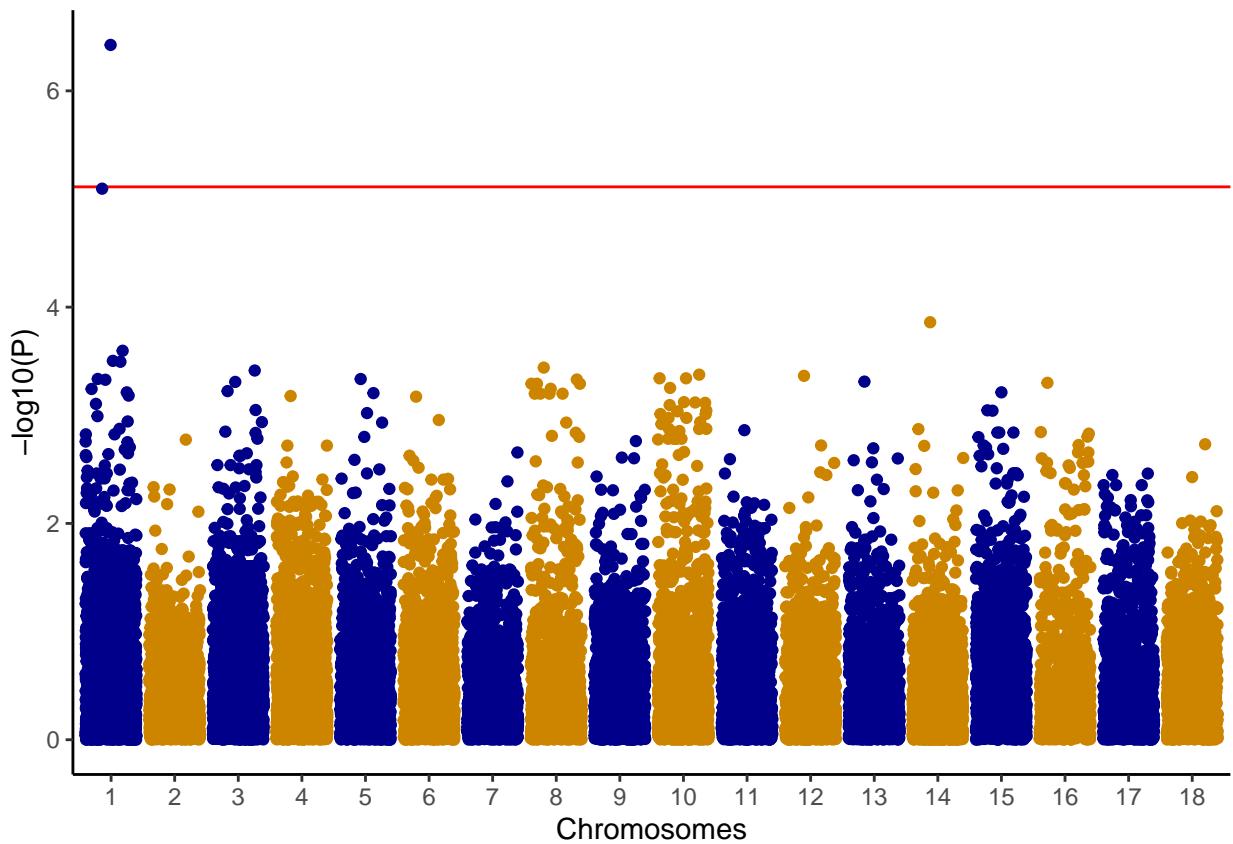
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :7757
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:7757
##  Mode   :character  Mode   :character  Median :0.1810  Median :7757
##                           Mean   :0.2101  Mean   :7757
##                           3rd Qu.:0.3260  3rd Qu.:7757
##                           Max.   :0.7370  Max.   :7770
##      l_mle          p_lrt
##  Min.   :0.000001  Min.   :0.0000004
##  1st Qu.:0.05477  1st Qu.:0.2214045
##  Median :0.07509  Median :0.4836919
##  Mean   :0.06669  Mean   :0.4868037
##  3rd Qu.:0.08052  3rd Qu.:0.7457590
##  Max.   :0.41036  Max.   :0.9999852
```

```

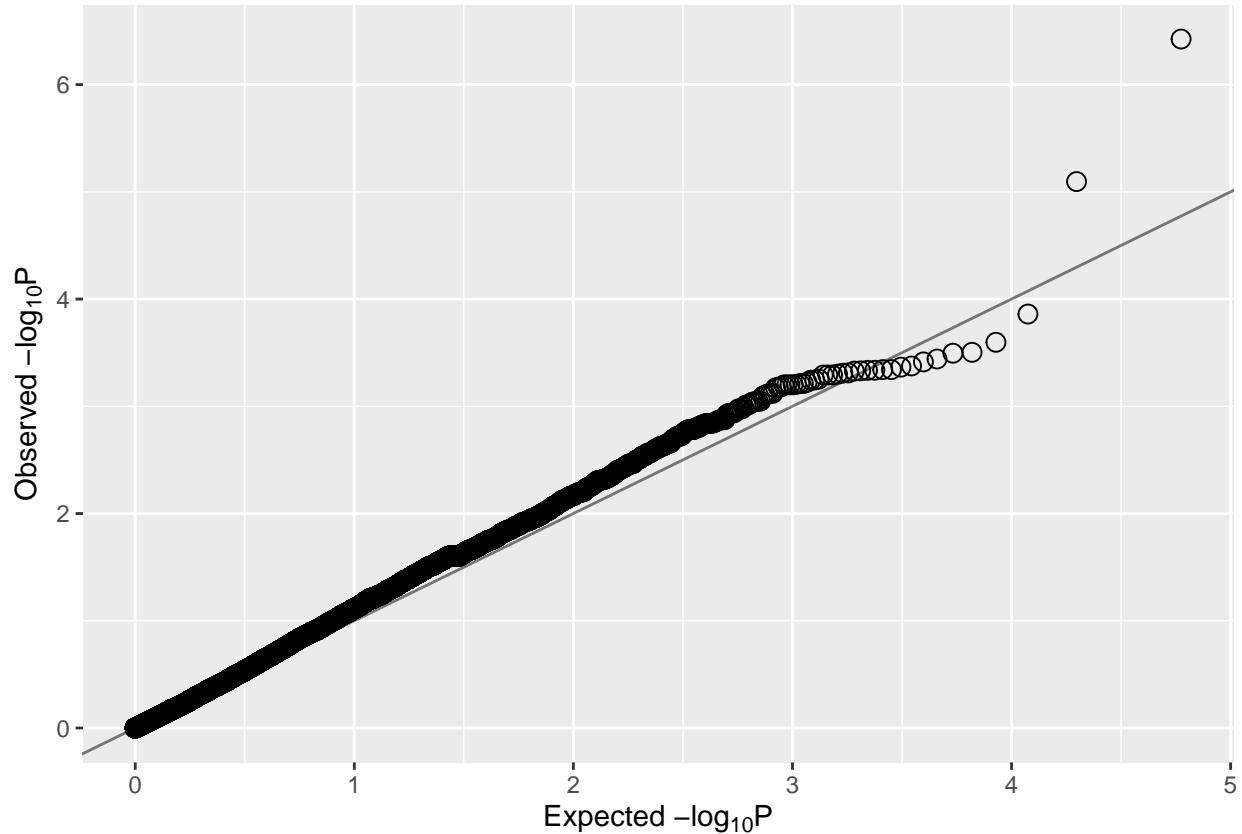
df7 <- gwscan7[,c(1,3,10,2)]
colnames(df7) <- c("CHR", "BP", "P", "SNP")
df7$BP <- as.numeric(df7$BP)
df7$CHR <- as.numeric(df7$CHR)

ggplot(data = df7,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color
geom_jitter() +
  labs(x = "Chromosomes")+
  theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df7$P)
```



```
gwscan8 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_NWF.assoc.txt", header = TRUE)
summary(gwscan8)
```

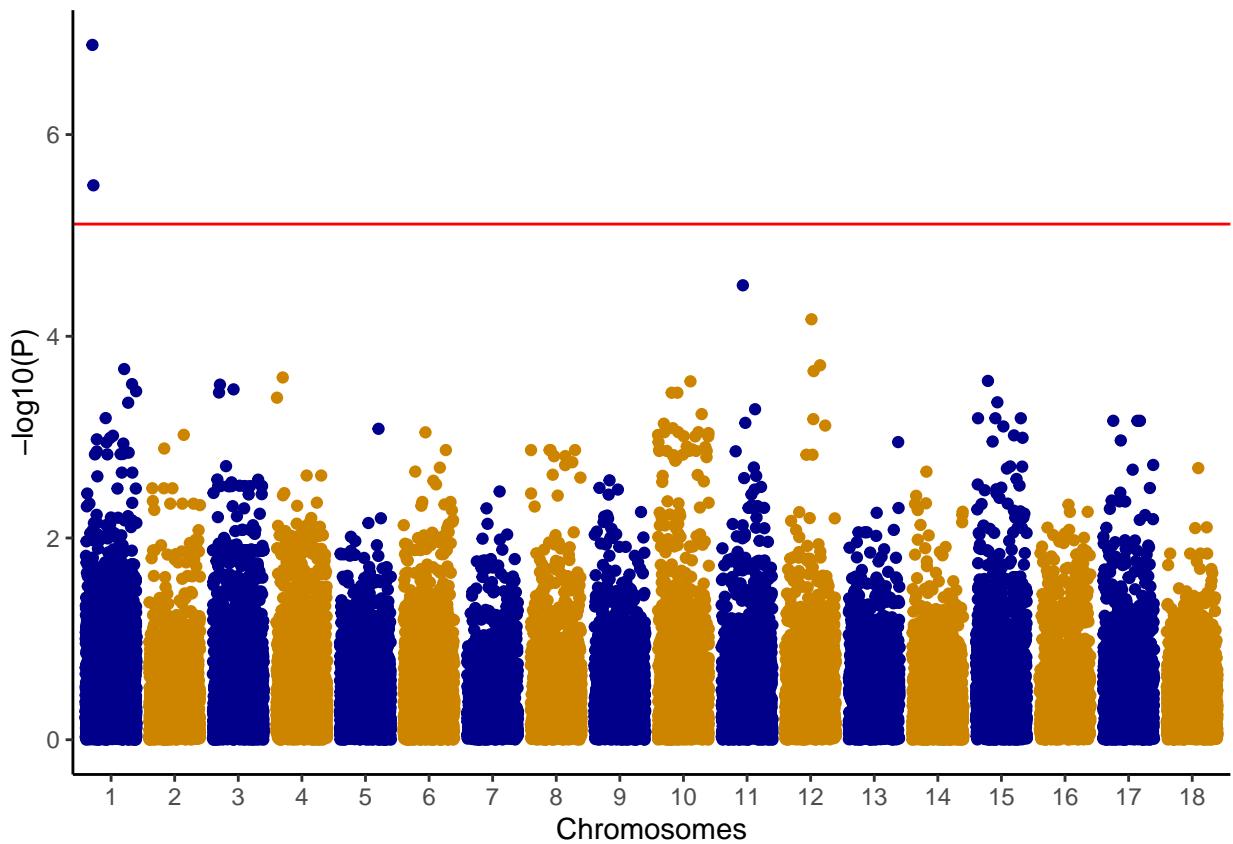
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :7740
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:7740
##  Mode   :character  Mode   :character  Median :0.1810  Median :7740
##                      Mean   :0.2101  Mean   :7741
##                      3rd Qu.:0.3260  3rd Qu.:7741
##                      Max.   :0.7370  Max.   :7754
##      l_mle          p_lrt
##  Min.   :0.0000010  Min.   :0.0000001
##  1st Qu.:0.004601  1st Qu.:0.2265866
##  Median :0.021724  Median :0.4705061
##  Mean   :0.019899  Mean   :0.4817659
##  3rd Qu.:0.027185  3rd Qu.:0.7353123
##  Max.   :0.284266  Max.   :0.9999660
```

```

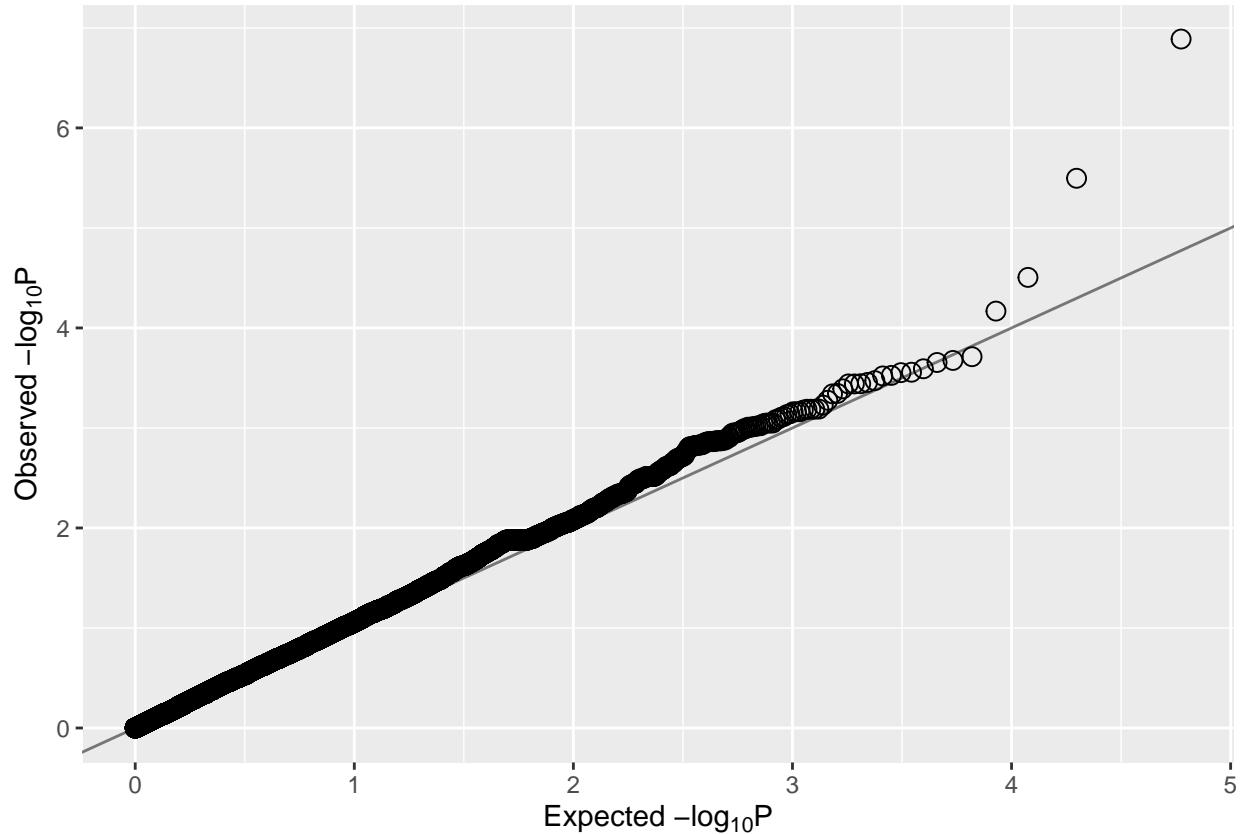
df8 <- gwscan8[,c(1,3,10,2)]
colnames(df8) <- c("CHR", "BP", "P", "SNP")
df8$BP <- as.numeric(df8$BP)
df8$CHR <- as.numeric(df8$CHR)

ggplot(data = df8,
        aes(x = as.factor(CHR),
            y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color
geom_jitter() +
labs(x = "Chromosomes")+
theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df8$P)
```



```
gwscan9 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_NECKO.assoc.txt", header=TRUE)
```

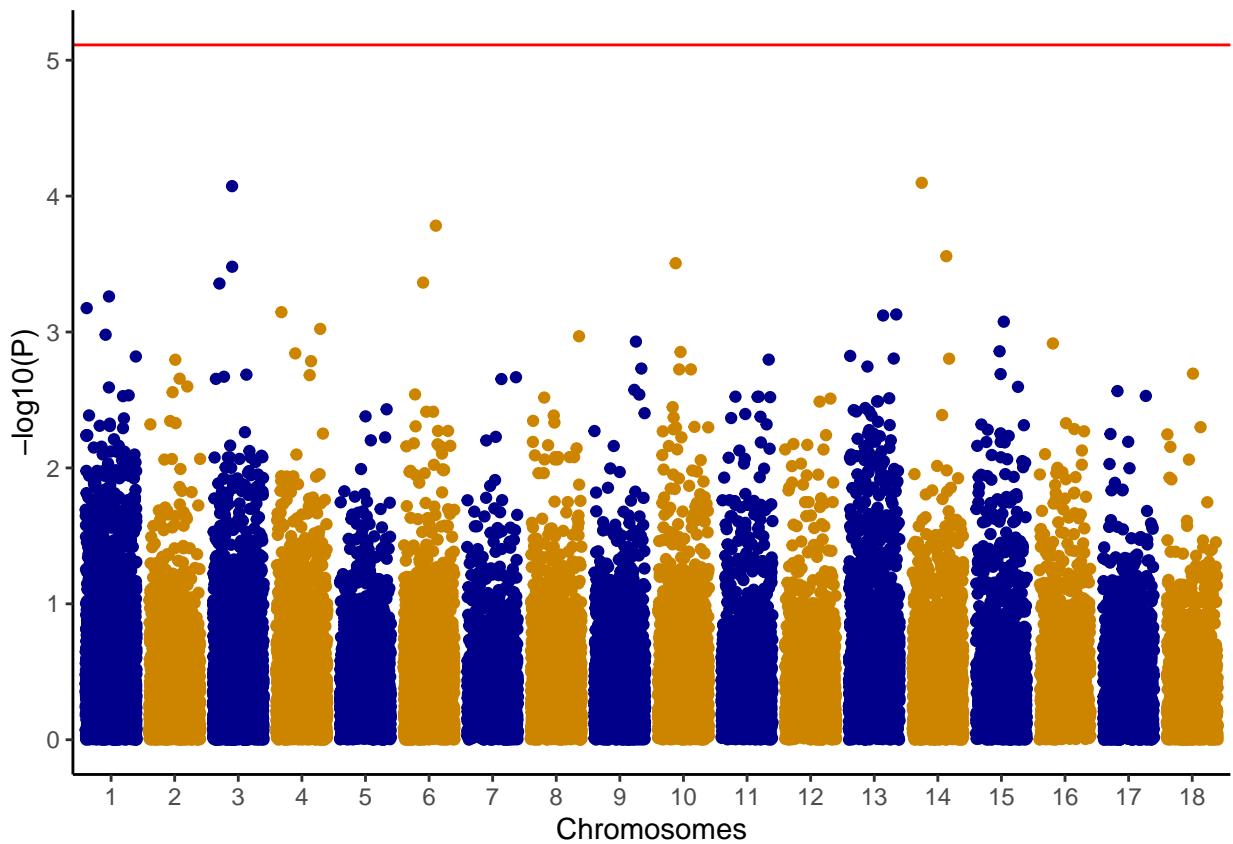
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.:5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0       af      log1_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :7639
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:7639
##  Mode   :character  Mode   :character  Median :0.1810  Median :7639
##                           Mean   :0.2101  Mean   :7639
##                           3rd Qu.:0.3260  3rd Qu.:7639
##                           Max.   :0.7370  Max.   :7646
##      l_mle          p_lrt
##  Min.   :0.000010  Min.   :0.0000798
##  1st Qu.:0.000010  1st Qu.:0.2391258
##  Median :0.008039  Median :0.4768199
##  Mean   :0.007940  Mean   :0.4874172
##  3rd Qu.:0.011042  3rd Qu.:0.7309164
##  Max.   :0.164405  Max.   :0.9999572
```

```

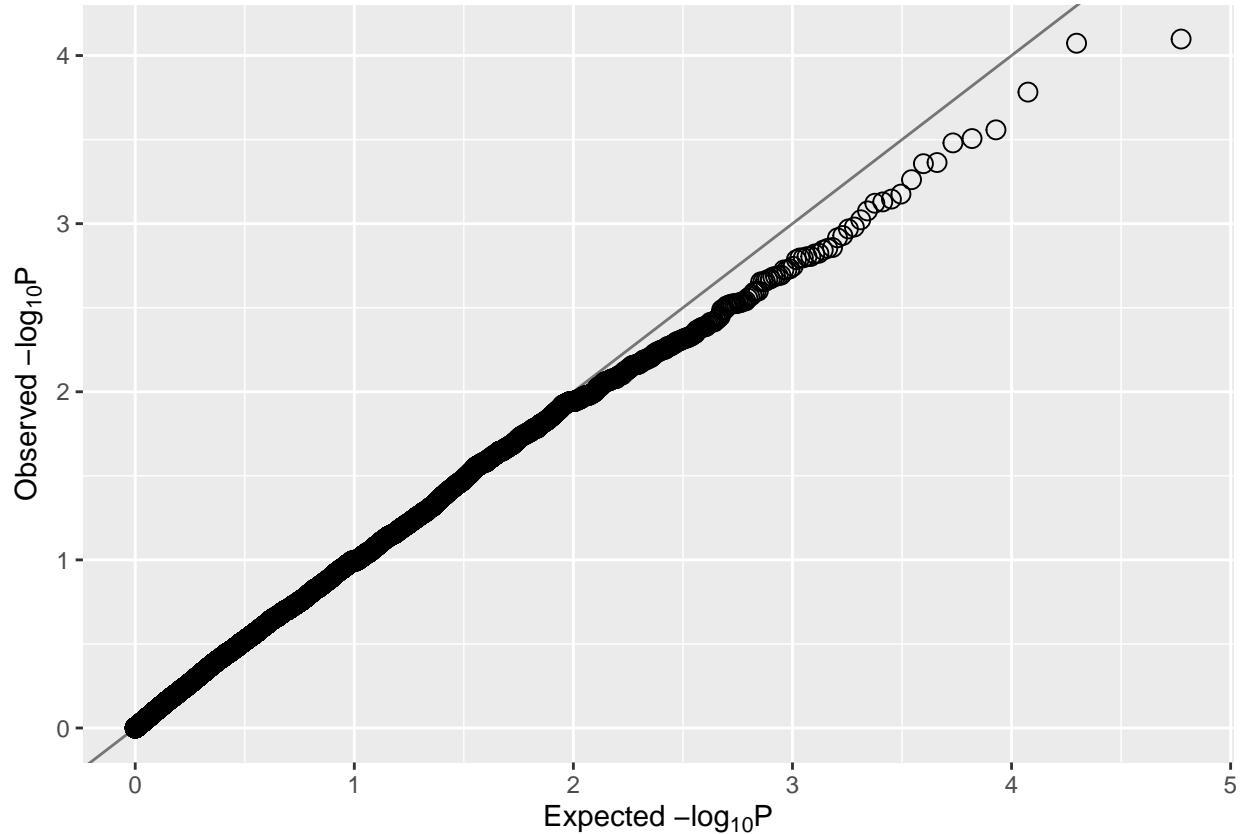
df9 <- gwscan9[,c(1,3,10,2)]
colnames(df9) <- c("CHR", "BP", "P", "SNP")
df9$BP <- as.numeric(df9$BP)
df9$CHR <- as.numeric(df9$CHR)

ggplot(data = df9,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes")+
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df9$P)
```



```
gwscan10<- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_may23_SN.assoc.txt", header = TRUE)
summary(gwscan10)
```

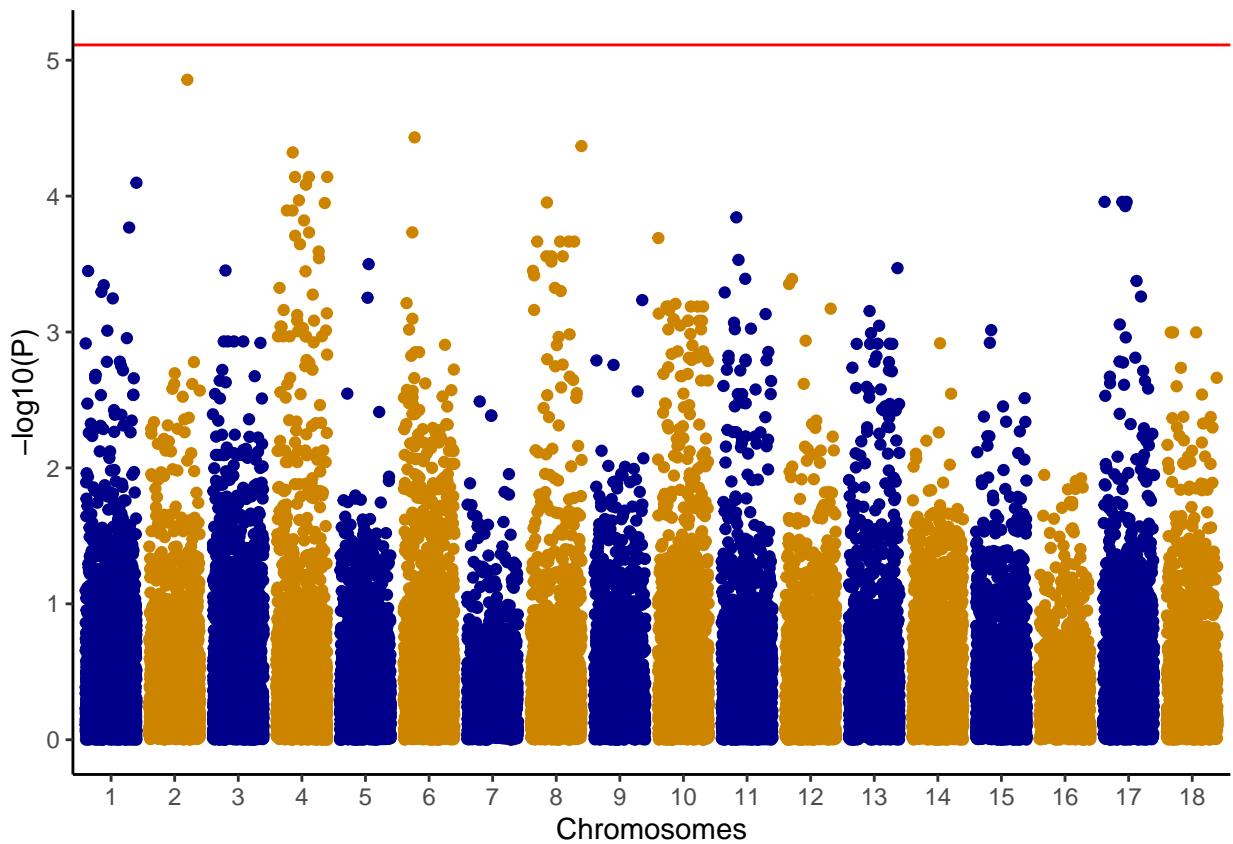
```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0         af      logl_H1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :8381
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:8382
##  Mode   :character  Mode   :character  Median :0.1810  Median :8382
##                               Mean   :0.2101  Mean   :8382
##                               3rd Qu.:0.3260  3rd Qu.:8382
##                               Max.   :0.7370  Max.   :8391
##      l_mle          p_lrt
##  Min.   :0.00001  Min.   :0.0000139
##  1st Qu.:0.05138  1st Qu.:0.2472250
##  Median :0.06377  Median :0.4886814
##  Mean   :0.05668  Mean   :0.4917499
##  3rd Qu.:0.06780  3rd Qu.:0.7347741
##  Max.   :0.21655  Max.   :0.9999598
```

```

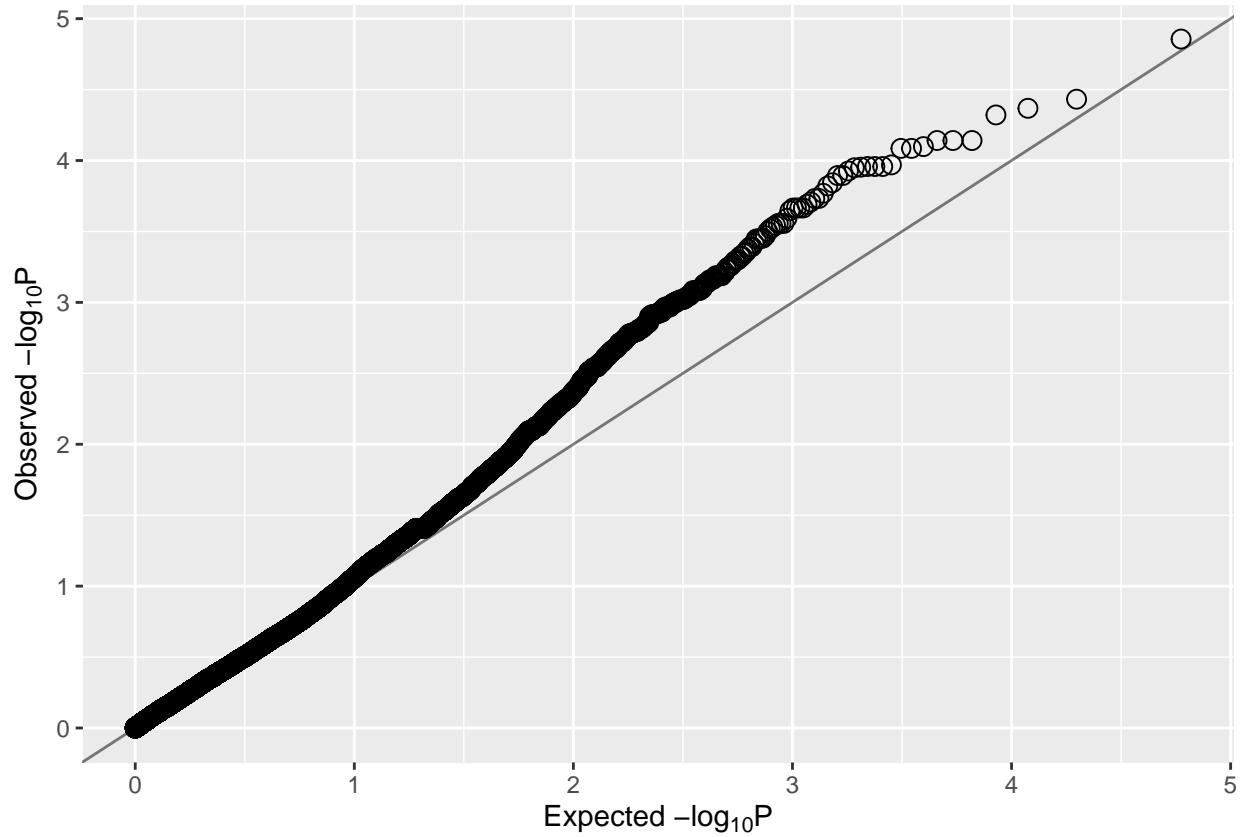
df10 <- gwscan10[,c(1,3,10,2)]
colnames(df10) <- c("CHR", "BP", "P", "SNP")
df10$BP <- as.numeric(df10$BP)
df10$CHR <- as.numeric(df10$CHR)

ggplot(data = df10,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df10$P)
```



```
##### Multivariate GWAS
```

```
Candidate Gene Analysis
```

```
Candidate Gene Analysis
```

```
#step1: #copy vcf file to the dir '/home/ln242/snpEff' cp /home/ln242/GEMMA-0.98.4/chr18.gbs_dart_phased_imputed.vcf . cp /home/ln242/GEMMA-0.98.4/chr13.gbs_dart_phased_imputed.vcf.vcf.gz . cp /home/ln242/GEMMA-0.98.4/chr1.gbs_dart_phased_imputed.vcf.vcf.gz . cp /home/ln242/GEMMA-0.98.4/chr2.gbs_dart_phased_imputed.vcf . cp /home/ln242/GEMMA-0.98.4/chr5.gbs_dart_phased_imputed.vcf.vcf.gz . cp /home/gjb99/genomes/manihot_ref6/M
```

.

```
gwscan11 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_CHVN_NAF_NWF_may2023.assoc.txt")
```

```
summary (gwscan11)
```

##	chr	rs	ps	n_miss
##	Min. : 1.000	Length:29708	Min. : 7103	Min. :0
##	1st Qu.: 4.000	Class :character	1st Qu.: 5291844	1st Qu.:0
##	Median : 9.000	Mode :character	Median :16448302	Median :0
##	Mean : 8.721		Mean :16909538	Mean :0
##	3rd Qu.:13.000		3rd Qu.:28000006	3rd Qu.:0
##	Max. :18.000		Max. :40505410	Max. :0
##	allele1	allele0	af	beta_1
##	Length:29708	Length:29708	Min. :0.0100	Min. :-1.300e-12
##	Class :character	Class :character	1st Qu.:0.0760	1st Qu.:-6.200e-14
##	Mode :character	Mode :character	Median :0.1810	Median : 5.697e-15

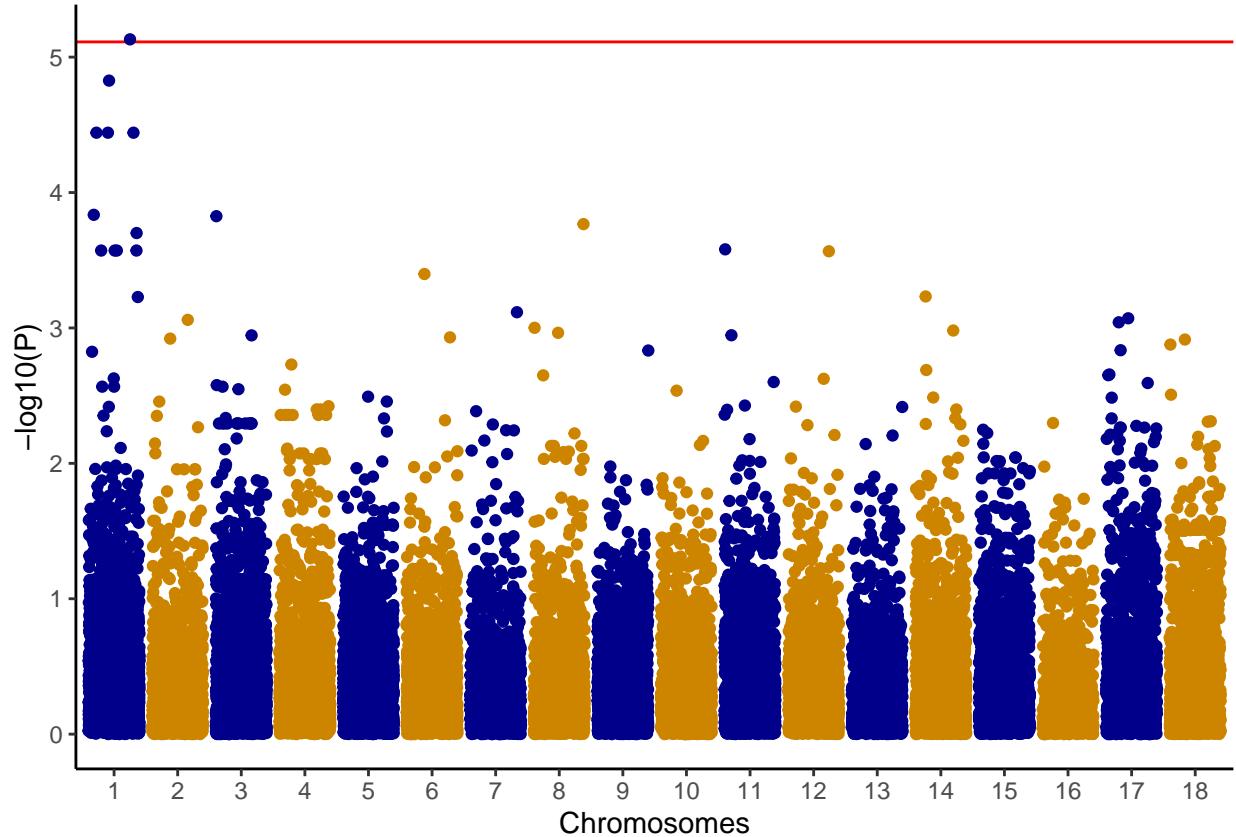
```

##                                     Mean    : 0.2101   Mean    : 2.855e-15
##                                     3rd Qu.: 0.3260   3rd Qu.: 7.486e-14
##                                     Max.    : 0.7370   Max.    : 7.415e-13
##      beta_2          beta_3          Vbeta_1_1
##  Min.  :-1.330e-12  Min.  :-1.420e-12  Min.  :3.548e-27
##  1st Qu.:-8.440e-14 1st Qu.:-7.890e-14 1st Qu.:5.877e-27
##  Median : 1.011e-14  Median : 1.549e-14  Median :9.608e-27
##  Mean   : 1.061e-14  Mean   : 1.014e-14  Mean   :1.912e-26
##  3rd Qu.: 1.043e-13 3rd Qu.: 1.081e-13 3rd Qu.:1.828e-26
##  Max.   : 1.370e-12  Max.   : 1.088e-12  Max.   :1.462e-25
##      Vbeta_1_2          Vbeta_1_3          Vbeta_2_2
##  Min.  :3.459e-27  Min.  :3.723e-27  Min.  :6.457e-27
##  1st Qu.:5.795e-27  1st Qu.:6.163e-27  1st Qu.:1.074e-26
##  Median :9.377e-27  Median :1.008e-26  Median :1.744e-26
##  Mean   :1.818e-26  Mean   :2.001e-26  Mean   :3.408e-26
##  3rd Qu.:1.752e-26  3rd Qu.:1.912e-26  3rd Qu.:3.269e-26
##  Max.   :1.388e-25  Max.   :1.530e-25  Max.   :2.602e-25
##      Vbeta_2_3          Vbeta_3_3          p_lrt
##  Min.  :5.951e-27  Min.  :6.697e-27  Min.  :0.0000074
##  1st Qu.:9.837e-27  1st Qu.:1.112e-26  1st Qu.:0.2907517
##  Median :1.608e-26  Median :1.822e-26  Median :0.5440907
##  Mean   :3.190e-26  Mean   :3.684e-26  Mean   :0.5301735
##  3rd Qu.:3.049e-26  3rd Qu.:3.515e-26  3rd Qu.:0.7738687
##  Max.   :2.438e-25  Max.   :2.819e-25  Max.   :0.9999516

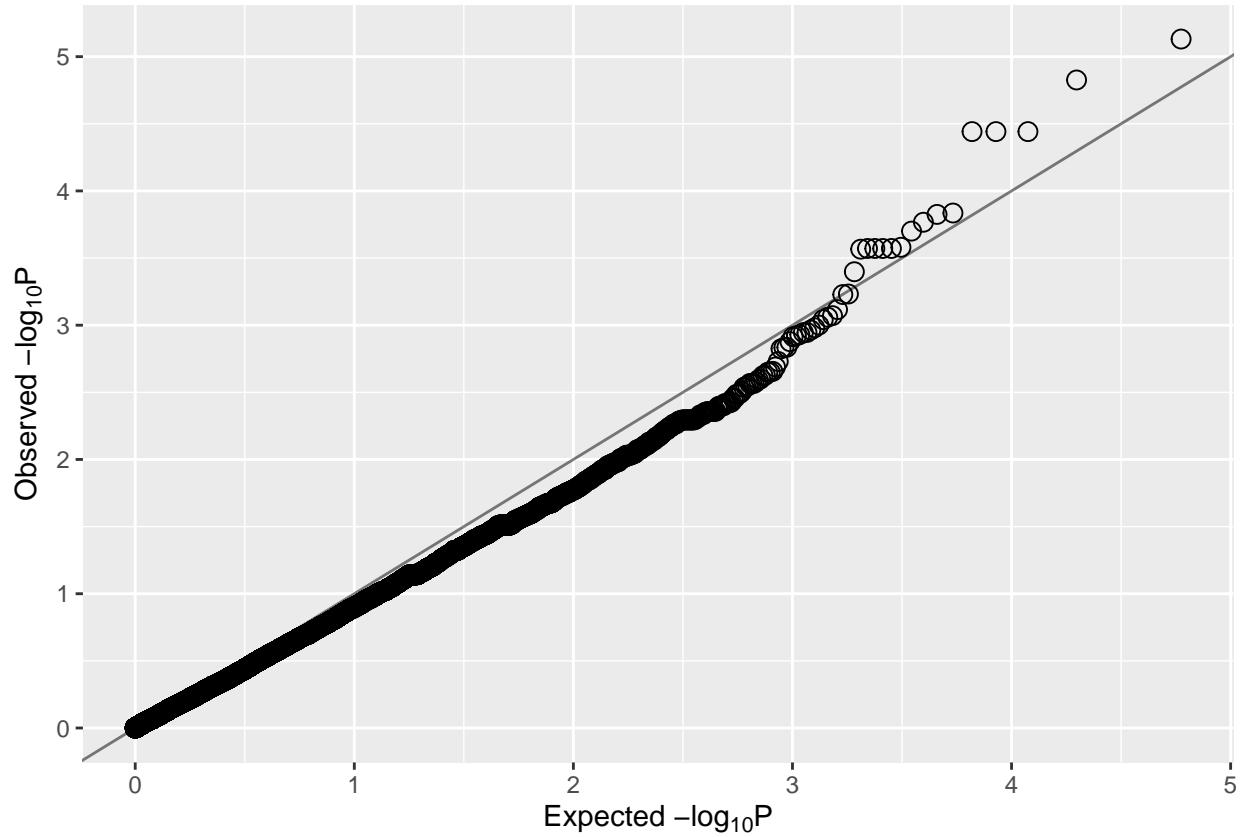
df11 <- gwscan11[,c(1,3,17,2)]
colnames(df11) <- c("CHR", "BP", "P", "SNP")
df11$BP <- as.numeric(df11$BP)
df11$CHR <- as.numeric(df11$CHR)

ggplot(data = df11,
        aes(x = as.factor(CHR),
            y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="black", size=1) +
        geom_jitter() +
        labs(x = "Chromosomes")+
        theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df11$P)
```



```
gwscan12 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_EAN_NECKRO_may2023.assoc.txt",
summary(gwscan12)
```

```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708   Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844 1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302 Median :0
##  Mean   : 8.721                   Mean   :16909538 Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006 3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410 Max.   :0
##      allele1        allele0        af      beta_1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :-2.060e-12
##  Class  :character  Class  :character  1st Qu.:0.0760 1st Qu.:-1.420e-13
##  Mode   :character  Mode   :character  Median :0.1810 Median : 1.309e-14
##                                Mean   :0.2101 Mean   : 2.420e-14
##                                3rd Qu.:0.3260 3rd Qu.: 1.739e-13
##                                Max.   :0.7370 Max.   : 3.277e-12
##      beta_2        Vbeta_1_1       Vbeta_1_2
##  Min.   :-2.640e-12  Min.   :1.171e-26  Min.   :2.138e-26
##  1st Qu.:-1.280e-13  1st Qu.:2.000e-26  1st Qu.:3.688e-26
##  Median : 1.866e-16  Median :3.271e-26  Median :6.028e-26
##  Mean   : 1.408e-15  Mean    :6.966e-26  Mean   :1.284e-25
##  3rd Qu.: 1.209e-13  3rd Qu.:6.614e-26  3rd Qu.:1.219e-25
##  Max.   : 1.792e-12  Max.   :5.370e-25  Max.   :9.859e-25
```

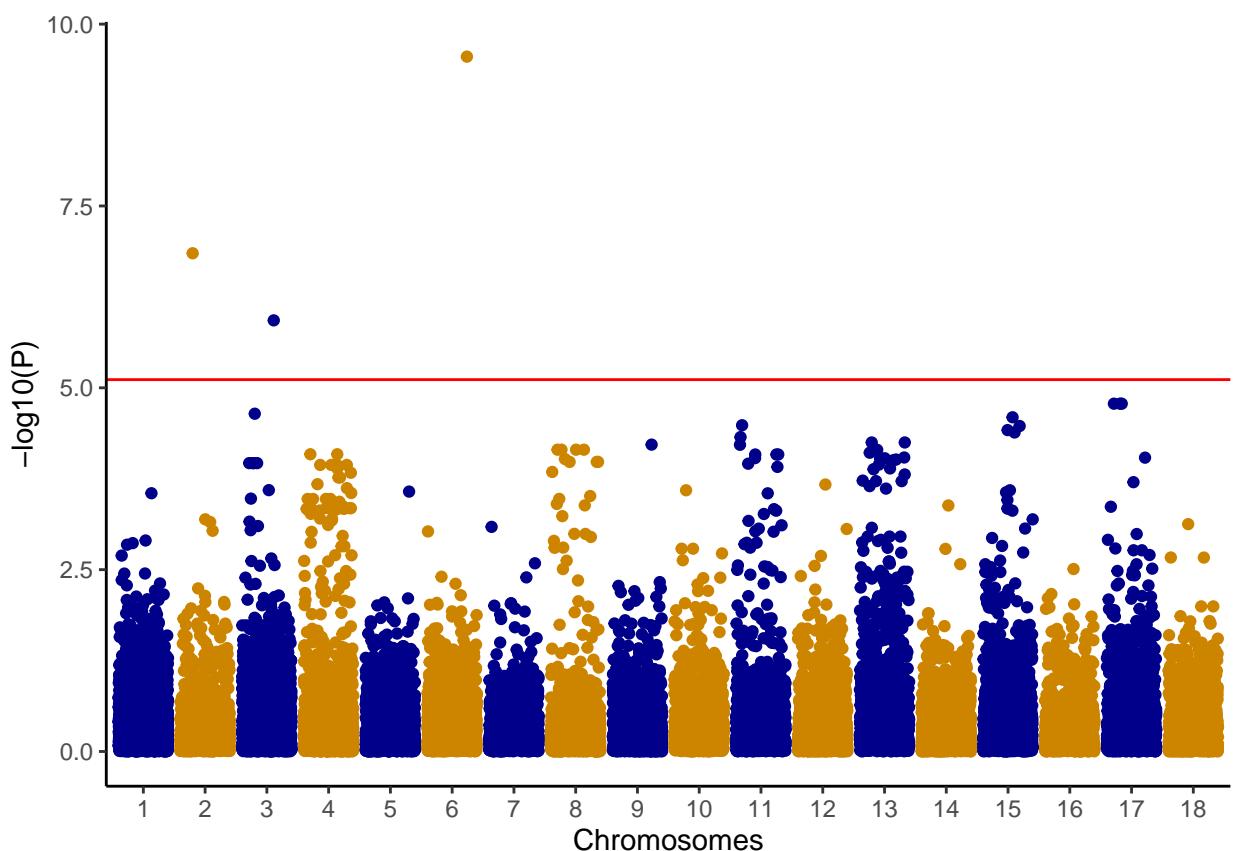
```

##      Vbeta_2_2                  p_lrt
##  Min.   :1.288e-26   Min.   :0.0000
##  1st Qu.:2.149e-26  1st Qu.:0.3082
##  Median :3.525e-26  Median :0.5559
##  Mean   :7.220e-26  Mean   :0.5384
##  3rd Qu.:6.878e-26  3rd Qu.:0.7844
##  Max.   :5.530e-25   Max.   :0.9995

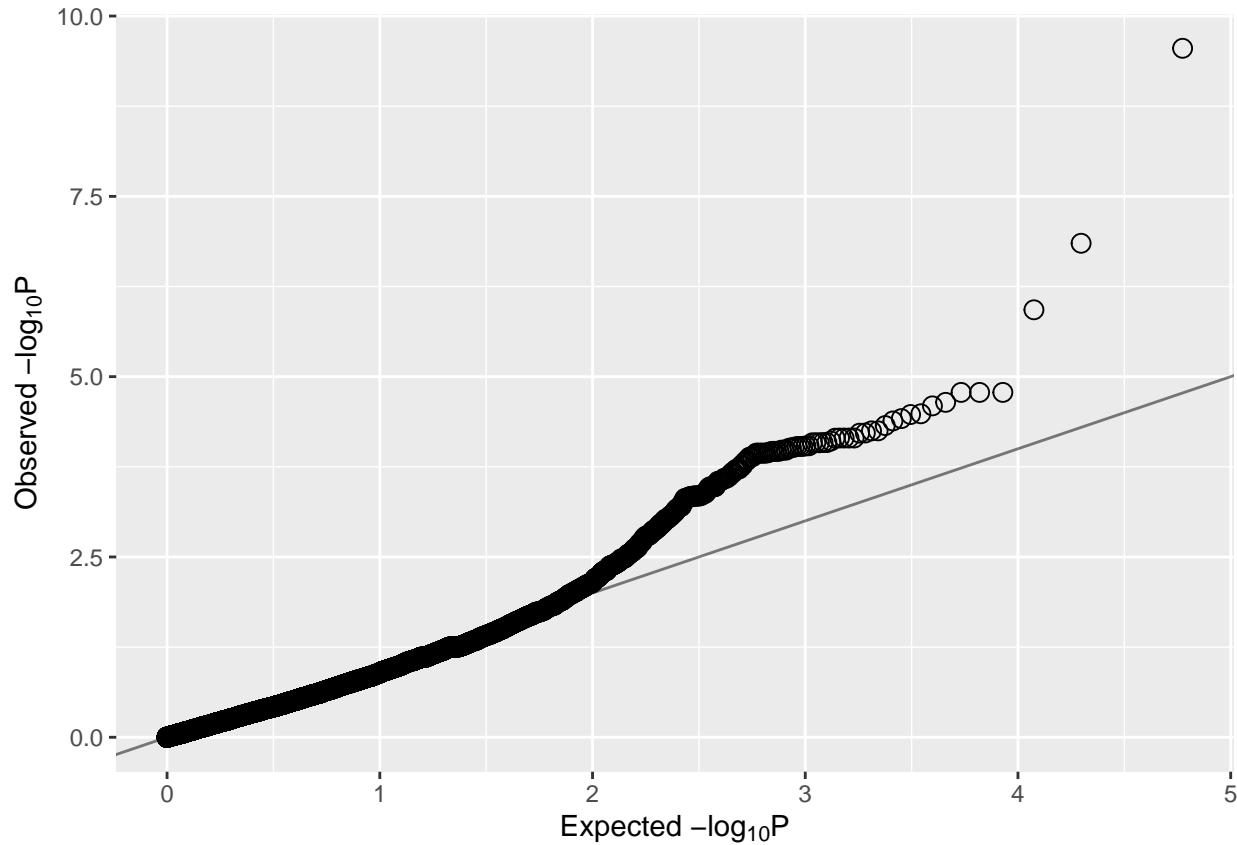
df12 <- gwscan12[,c(1,3,13,2)]
colnames(df12) <- c("CHR", "BP", "P", "SNP")
df12$BP <- as.numeric(df12$BP)
df12$CHR <- as.numeric(df12$CHR)

ggplot(data = df12,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes")+
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df12$P)
```



```
gwscan13 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_NWF_NAF_may2023.assoc.txt", h=TRUE)
summary(gwscan13)
```

```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0         af        beta_1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :-1.422e-12
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:-7.945e-14
##  Mode   :character  Mode   :character  Median :0.1810  Median : 1.538e-14
##                           Mean   :0.2101  Mean   : 9.791e-15
##                           3rd Qu.:0.3260  3rd Qu.: 1.086e-13
##                           Max.   :0.7370  Max.   : 1.085e-12
##      beta_2        Vbeta_1_1       Vbeta_1_2
##  Min.   :-1.335e-12  Min.   :6.856e-27  Min.   :6.091e-27
##  1st Qu.:-8.492e-14  1st Qu.:1.127e-26  1st Qu.:9.980e-27
##  Median : 1.025e-14  Median :1.845e-26  Median :1.626e-26
##  Mean   : 1.040e-14  Mean   :3.706e-26  Mean   :3.209e-26
```

```

## 3rd Qu.: 1.052e-13   3rd Qu.:3.534e-26   3rd Qu.:3.070e-26
## Max.    : 1.368e-12  Max.    :2.834e-25   Max.    :2.451e-25
## Vbeta_2_2           p_lrt
## Min.    :6.562e-27   Min.    :0.0000011
## 1st Qu.:1.085e-26   1st Qu.:0.2564861
## Median  :1.759e-26   Median  :0.5041278
## Mean    :3.421e-26   Mean    :0.4993478
## 3rd Qu.:3.284e-26   3rd Qu.:0.7437277
## Max.    :2.610e-25   Max.    :0.9853671

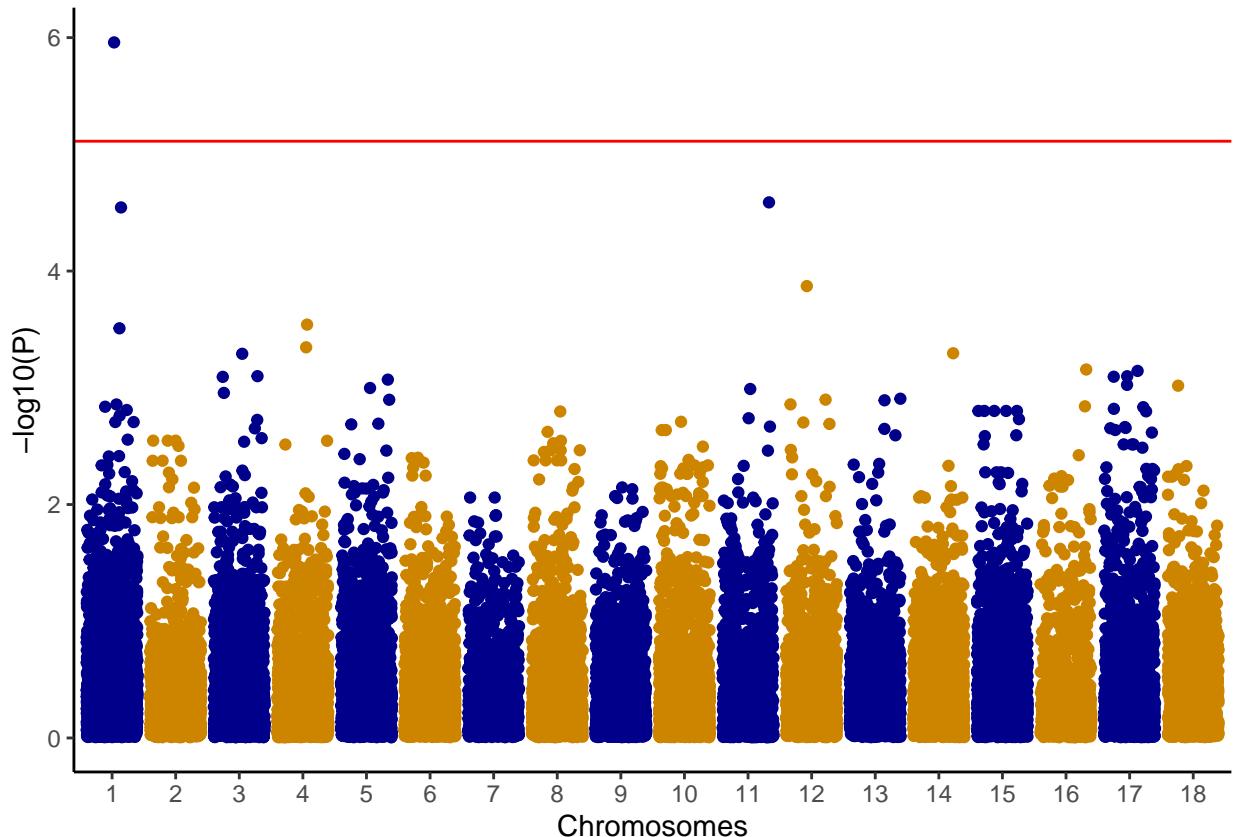
```

```

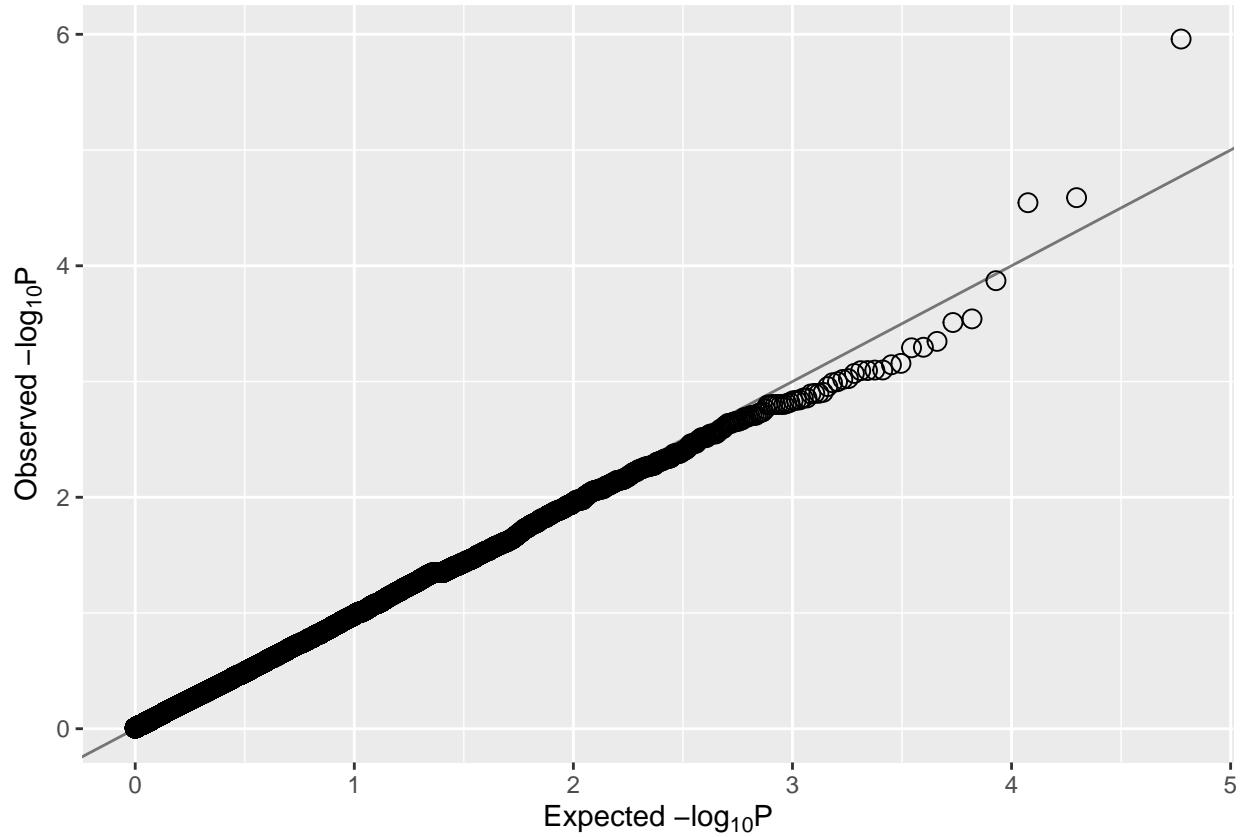
df13 <- gwscan13[,c(1,3,13,2)]
colnames(df13) <- c("CHR", "BP", "P", "SNP")
df13$BP <- as.numeric(df13$BP)
df13$CHR <- as.numeric(df13$CHR)

ggplot(data = df13,
        aes(x = as.factor(CHR),
            y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
        geom_jitter() +
        labs(x = "Chromosomes")+
        theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df13$P)
```



```
gwscan14 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_NEcro_CBSDs3_CBSDs6_may2023.a
summary (gwscan14)
```

```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708   Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844 1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302 Median :0
##  Mean   : 8.721                           Mean   :16909538 Mean   :0
##  3rd Qu.:13.000                           3rd Qu.:28000006 3rd Qu.:0
##  Max.   :18.000                           Max.   :40505410 Max.   :0
##      allele1        allele0         af      beta_1
##  Length:29708  Length:29708   Min.   :0.0100  Min.   :-2.648e-12
##  Class  :character  Class  :character  1st Qu.:0.0760 1st Qu.:-1.375e-13
##  Mode   :character  Mode   :character  Median :0.1810 Median : 3.027e-15
##                               Mean   :0.2101 Mean   : 3.889e-15
##                               3rd Qu.:0.3260 3rd Qu.: 1.312e-13
##                               Max.   :0.7370 Max.   : 1.781e-12
##      beta_2        beta_3        Vbeta_1_1
##  Min.   :-4.738e-13  Min.   :-4.223e-13  Min.   :1.235e-26
##  1st Qu.:-1.631e-14  1st Qu.:-2.553e-14  1st Qu.:2.095e-26
##  Median : 3.143e-15  Median :-1.220e-15  Median :3.423e-26
##  Mean   : 1.739e-15  Mean   :-2.232e-15  Mean   :7.150e-26
##  3rd Qu.: 2.071e-14  3rd Qu.: 2.292e-14  3rd Qu.:6.778e-26
##  Max.   : 2.162e-13  Max.   : 6.425e-13  Max.   :5.474e-25
```

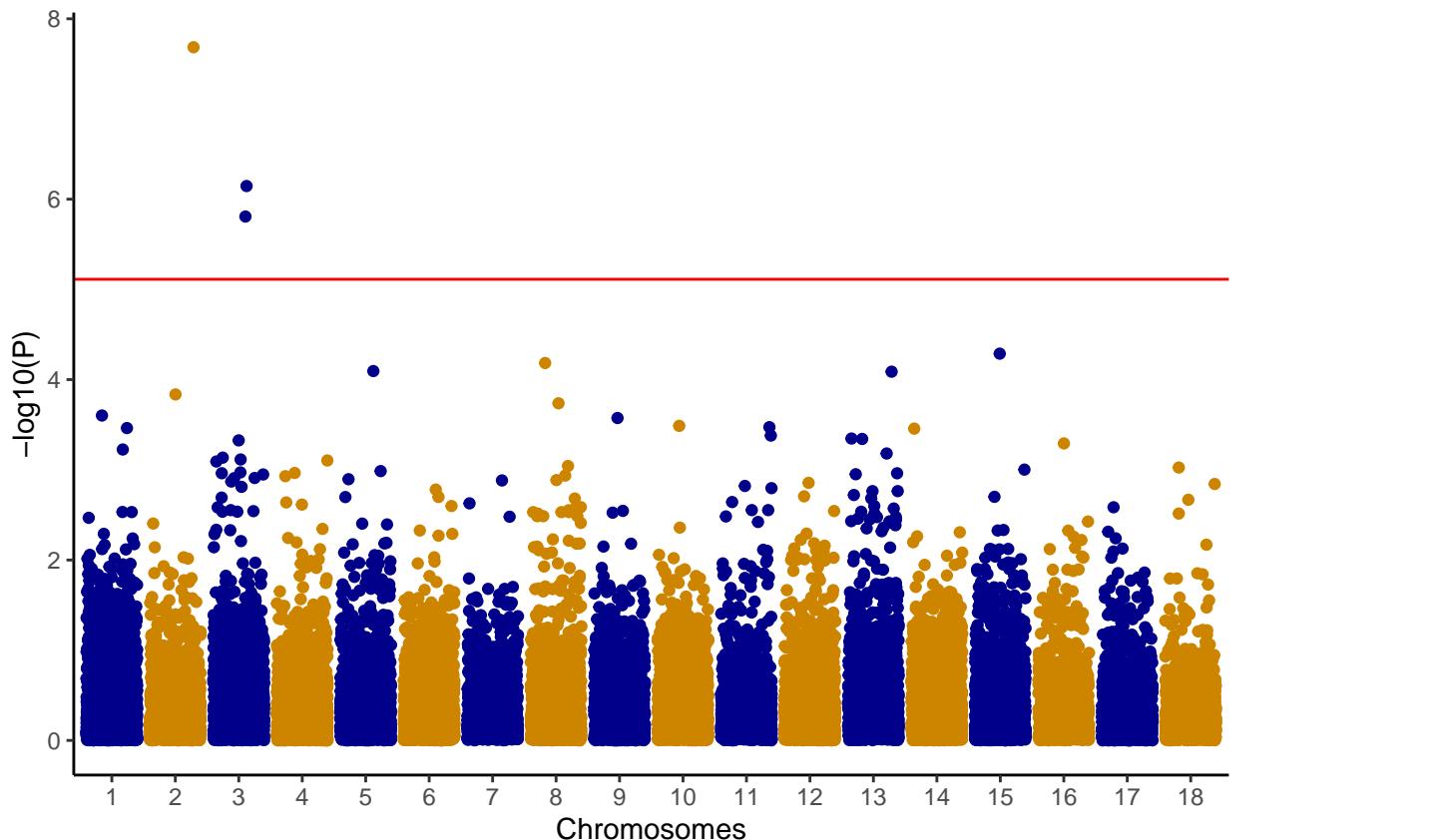
```

##      Vbeta_1_2          Vbeta_1_3          Vbeta_2_2
##  Min. :-2.813e-27  Min. :-2.888e-26  Min. :2.805e-28
##  1st Qu.:-2.771e-28 1st Qu.:-3.405e-27 1st Qu.:4.787e-28
##  Median :-1.381e-28 Median :-1.687e-27 Median :7.832e-28
##  Mean   :-2.906e-28 Mean  :-3.589e-27 Mean  :1.669e-27
##  3rd Qu.:-8.592e-29 3rd Qu.:-1.030e-27 3rd Qu.:1.582e-27
##  Max.  :-3.067e-29  Max. :-6.058e-28 Max.  :1.281e-26
##      Vbeta_2_3          Vbeta_3_3          p_lrt
##  Min. :8.157e-29  Min. :5.324e-28  Min. :0.0000
##  1st Qu.:1.399e-28 1st Qu.:9.046e-28 1st Qu.:0.2950
##  Median :2.286e-28 Median :1.478e-27 Median :0.5581
##  Mean   :4.859e-28 Mean  :3.149e-27 Mean  :0.5380
##  3rd Qu.:4.605e-28 3rd Qu.:2.987e-27 3rd Qu.:0.7863
##  Max.  :4.529e-27  Max. :2.418e-26 Max.  :0.9993

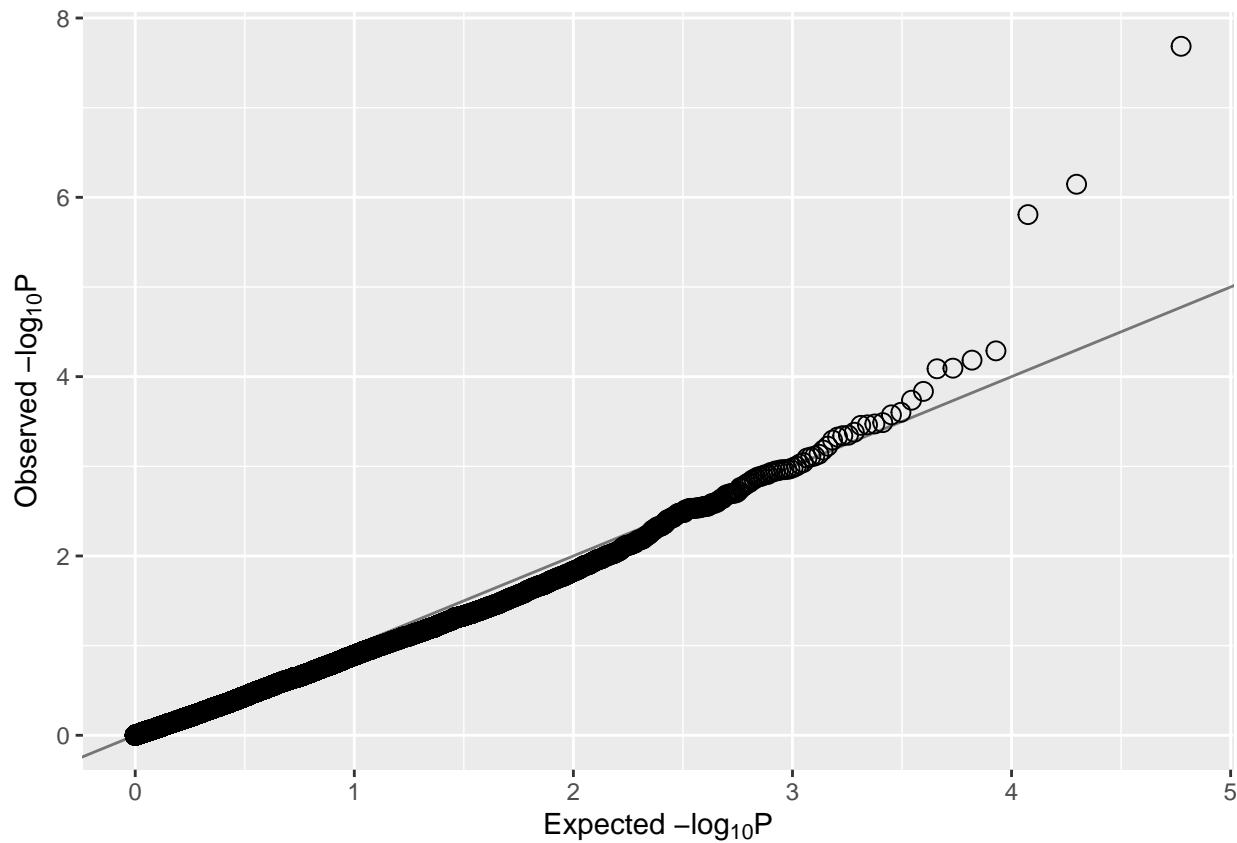
df14 <- gwscan14[,c(1,3,17,2)]
colnames(df14) <- c("CHR", "BP", "P", "SNP")
df14$BP <- as.numeric(df14$BP)
df14$CHR <- as.numeric(df14$CHR)

ggplot(data = df14,
       aes(x = as.factor(CHR),
           y = -log10(P), color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
       geom_jitter() +
       labs(x = "Chromosomes") +
       theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df14$P)
```



```
gwscan15 <- read.table("~/Downloads/output/combinedC2_final_2_PLINK_GEMMA_NECKRO_CBSDs12_may2023.assoc.txt")
summary(gwscan15)
```

```
##      chr          rs          ps      n_miss
##  Min.   : 1.000  Length:29708  Min.   : 7103  Min.   :0
##  1st Qu.: 4.000  Class  :character  1st Qu.: 5291844  1st Qu.:0
##  Median : 9.000  Mode   :character  Median :16448302  Median :0
##  Mean   : 8.721                   Mean   :16909538  Mean   :0
##  3rd Qu.:13.000                   3rd Qu.:28000006  3rd Qu.:0
##  Max.   :18.000                   Max.   :40505410  Max.   :0
##      allele1        allele0         af        beta_1
##  Length:29708  Length:29708  Min.   :0.0100  Min.   :-2.642e-12
##  Class  :character  Class  :character  1st Qu.:0.0760  1st Qu.:-1.291e-13
##  Mode   :character  Mode   :character  Median :0.1810  Median : 1.406e-15
##                           Mean   :0.2101  Mean   : 4.269e-15
##                           3rd Qu.:0.3260  3rd Qu.: 1.261e-13
##                           Max.   :0.7370  Max.   : 1.902e-12
##      beta_2        Vbeta_1_1       Vbeta_1_2
##  Min.   :-2.646e-11  Min.   :1.354e-26  Min.   :8.290e-26
##  1st Qu.:-2.883e-12  1st Qu.:2.226e-26  1st Qu.:1.434e-25
##  Median :-4.750e-13  Median :3.650e-26  Median :2.340e-25
##  Mean   :-1.371e-13  Mean   :7.333e-26  Mean   :4.978e-25
```

```

## 3rd Qu.: 2.061e-12   3rd Qu.:7.002e-26   3rd Qu.:4.718e-25
## Max.    : 5.800e-11   Max.    :5.593e-25   Max.    :4.049e-24
## Vbeta_2_2           p_lrt
## Min.    :4.728e-24   Min.    :0.0000
## 1st Qu.:8.047e-24   1st Qu.:0.2904
## Median  :1.316e-23   Median  :0.5348
## Mean    :2.802e-23   Mean    :0.5192
## 3rd Qu.:2.657e-23   3rd Qu.:0.7540
## Max.    :2.152e-22   Max.    :0.9948

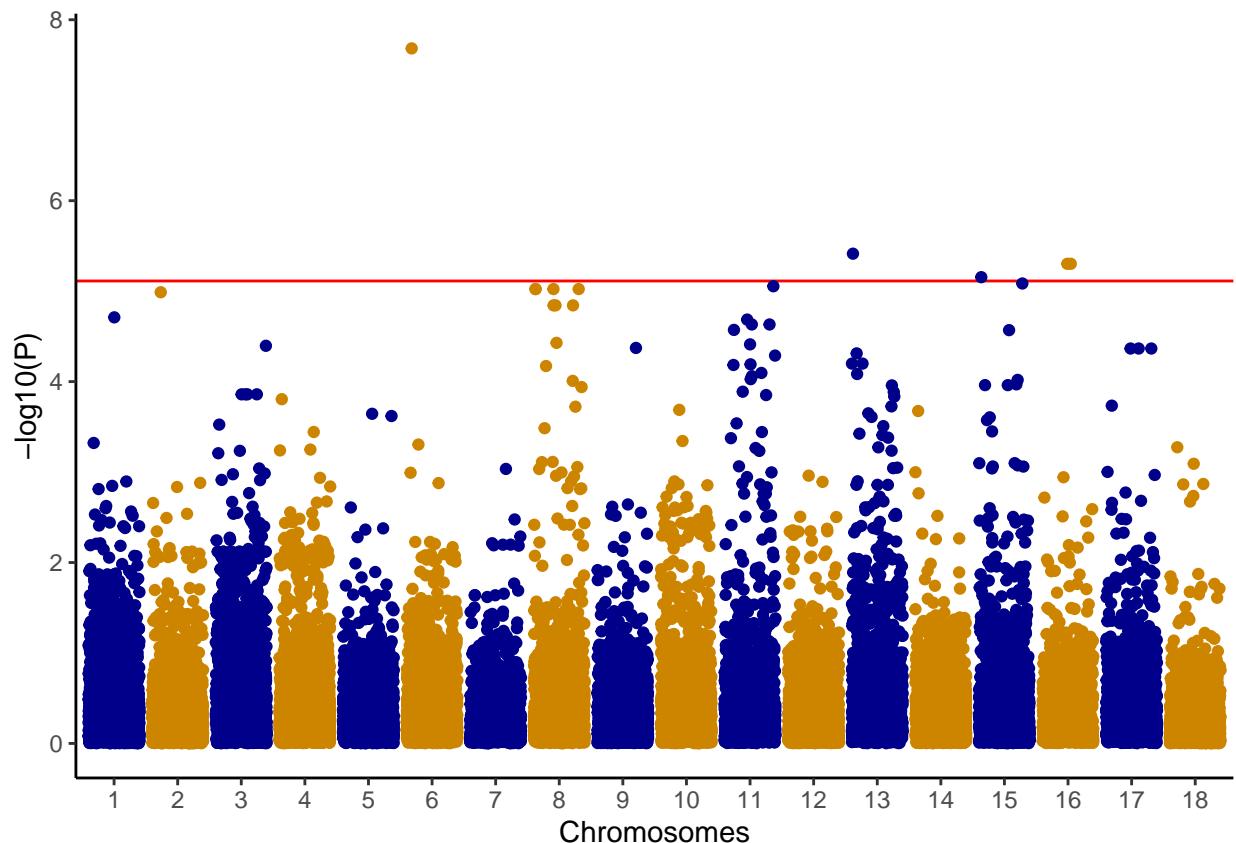
```

```

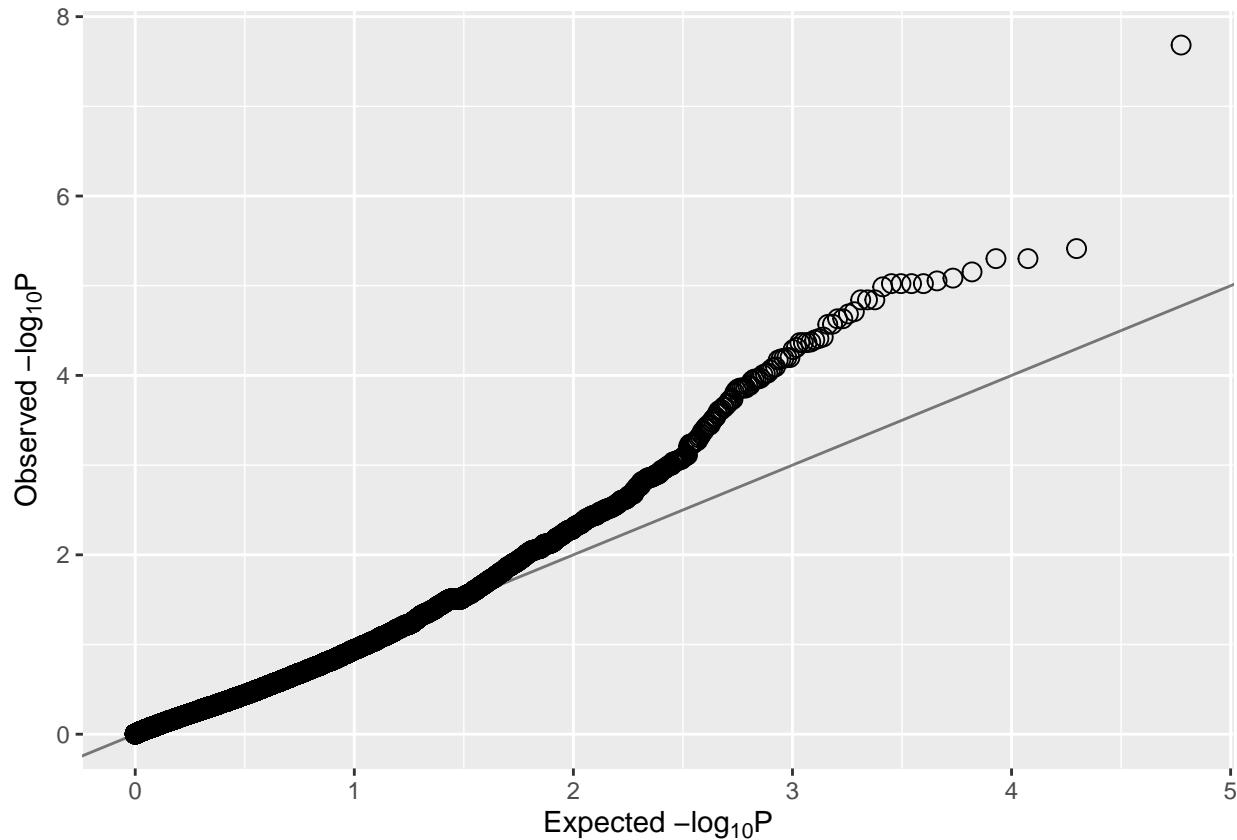
df15 <- gwscan15[,c(1,3,13,2)]
colnames(df15) <- c("CHR","BP","P","SNP")
df15$BP <- as.numeric(df15$BP)
df15$CHR <- as.numeric(df15$CHR)

ggplot(data = df15,
        aes(x = as.factor(CHR),
            y = -log10(P),color=as.factor(CHR))) + geom_hline(yintercept = -log10(p_threshold), color="red") +
        geom_jitter() +
        labs(x = "Chromosomes")+
        theme_classic() + theme(legend.position = 'none') + scale_color_manual(values=colR)

```



```
gg_qqplot(df15$P)
```



### **running snpEff (Candidate annotation)**

```
java -Xmx4g -jar snpEff.jar manihotv6_1 chr18.gbs_dart_phased_imputted.vcf.gz > c2_UGA.chr18.ann.vcf
java -Xmx4g -jar snpEff.jar manihotv6_1 chr13.gbs_dart_phased_imputted.vcf.gz > c2_UGA.chr13.ann.vcf
#Got error and did not proceed further
```

### **transfer cassava.gff3 file**

```
scp Mesculenta_305_v6.1.gene_exons.gff3 ln242@solanine:/home/ln242 scp ln242@login.sgn.cornell.edu:/home/ln242/Mesculenta_305_v6.1.gene_exons.gff3 .
```

**subset gff3 file based on significant region of interest**

**visit '<https://www.biostars.org/p/141968/#142129>' for more info**

**transfer file from “~/snpEff”**

```
scp *_annotation.gtf ln242@solanine:/home/ln242
scp ln242@login.sgn.cornell.edu:/home/ln242/*_annotation.gtf .
```