

Concordance

Molly Garrett

2023-06-13

#Load packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.2      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr    1.5.0
```

```
## v ggplot2    3.4.3      v tibble     3.2.1
```

```
## v lubridate  1.9.2      v tidyr      1.3.0
```

```
## v purrr      1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Prep

```
#set rad and gt
```

```
rad <- read.csv('NIDGS_RADvsGT_RAD_04272023.csv')
```

```
gt <- read.csv('NIDGS_RADvsGT_GT_04272023.csv')
```

```
# remove first column
```

```
rad <- rad[, -1]
```

```
head(rad)
```

```
##   LZ49 LZ89 LZA9 LZC9 LZN9 LZS9 LZX9 LZZ9 MZ69 MZ79 MZ89 MZA9 MZC9 MZD9 MZE9
## 1   CC   NN   CC   GC   NN   CC   NN   CC   CC   NN   NN   NN   GG   CC   GC
## 2   GG   NN   GG   GG   NN   GG   GG   GA   GG   GG   GA   GG   GA   GA   AA
## 3   CC   NN   CC   CC   NN   CC   NN   CC   CC   TT   CC   NN   CC   CC   CC
## 4   CC   NN   CC   CC   NN   CC   CC   NN   CC   NN   TT   NN   CT   NN   NN
## 5   GG   NN   GG   GG   GA   GG   GG   GA   GA   GA   AA   NN   GG   GA   GG
## 6   NN   NN   CT   CC   NN   NN   NN   CC   CC   CC   CC   NN   NN   CC   CC
##   MZG9 MZI9 MZT9 MZU9 MZV9 MZW9 NZ49 NZ59 NZ69 NZ79 NZA9 NZB9 NZC9 NZD9 NZG9
## 1   CC   GG   GG   GC   CC   GC   GC   GC   GC   GC   GC   GC   NN   NN   GC
## 2   GG   AA   AA   GA   AA   GG   AA   GA   GA   GA   GG   GG   GA   GA   GA
## 3   CT   CC   CT   CC   CC   CT   CT   CC   NN   CC   CC   CC   CC   NN   CC
## 4   CC   CC   CC   CC   TT   NN   CT   CT   NN   CC   CC   CC   CC   CC   NN
## 5   GG   NN   GA   GG   GA   GA   GA   AA   GG   GA   AA   AA   AA   AA   AA
## 6   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   NN   CC   CC
##   NZK9 NZX9 OZ19 PZX9 QZ79 TZCJ TZZJ UZ2J UZ3J UZ7J UZ8J UZDJ UZEJ UZFJ UZGJ
## 1   NN   NN   CC   CC   GC   CC   NN   NN   GG   GC   GG   GC   NN   GG   GG
## 2   GA   NN   GA   AA   NN   GA   GA   GG   GA   AA   GG   GA   GA   AA   GA
## 3   CC   NN   CC   CC   NN   CC   CT   CC   TT   CT   NN   CT   CC   CT   CC
## 4   NN   NN   CC   NN   NN   NN   NN   CC   CC   CC   NN   CC   NN   CC   CC
```

```
## 5   GA   GG   GG   GG   GG   GG   NN   GG   GG   GA   GA   GG   GA   GA   GG
## 6   NN   CT   TT   CT   NN   CC   CC   NN   NN   CT   CT   CC   CC   NN   CC
##    UZHJ UZIJ UZJJ UZVJ UZZJ WZ1J ZZS8 ZZY8
## 1   NN   GG   NN   NN   CC   NN   CC   GC
## 2   AA   AA   NN   GA   GA   GA   NN   AA
## 3   NN   NN   CC   NN   NN   CC   CC   CC
## 4   NN   NN   NN   CC   CT   NN   NN   CC
## 5   GA   GA   GG   NN   GG   GA   NN   GG
## 6   NN   CT   NN   NN   CC   NN   NN   NN
```

```
gt <- gt[, -1]
head(gt)
```

```
##    LZ49 LZ89 LZA9 LZC9 LZN9 LZS9 LZX9 LZZ9 MZ69 MZ79 MZ89 MZA9 MZC9 MZD9 MZE9
## 1   CC   CC   CC   GC   CC   CC   CC   CC   CC   GG   GC   CC   GG   CC   GC
## 2   GG   GA   GG   GG   AA   GG   GG   GA   GG   GG   GA   GG   GA   GA   AA
## 3   CC   CC   CC   CC   CC   CC   CC   CC   CC   TT   CC   CC   CC   CC   CC
## 4   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   TT   TT   CT   CC   CC
## 5   GA   GA   GG   GG   GA   GG   GG   GA   GA   GA   AA   GA   GG   GA   GG
## 6   CC   CC   CT   CC   NN   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC
##    MZG9 MZI9 MZT9 MZU9 MZV9 MZW9 NZ49 NZ59 NZ69 NZ79 NZA9 NZB9 NXC9 NZD9 NZG9
## 1   CC   GG   GG   GC   GC   GC   GC   GC   GC   GC   GC   GC   GC   GC   GC
## 2   GG   AA   AA   GA   AA   GG   AA   GA   GA   GA   GG   GG   GA   GA   GA
## 3   CT   CC   CT   CC   CC   CT   CT   CC   CC   CC   CC   CC   CC   CC   CC
## 4   CC   CC   CC   CT   CT   CT   CT   CT   CT   CC   CC   CC   CC   CC   CC
## 5   GG   GG   GA   GG   GA   GA   GA   AA   GG   GA   AA   AA   AA   AA   AA
## 6   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CT   CC
##    NZK9 NZX9 OZ19 PZX9 QZ79 TZCJ TZZJ UZ2J UZ3J UZ7J UZ8J UZDJ UZEJ UZFJ UZGJ
## 1   CC   GC   CC   CC   GC   CC   GC   GC   GG   GC   GG   GC   GC   GG   GG
## 2   GA   GA   GA   AA   GA   GA   GA   GG   GA   AA   GG   GA   GA   AA   GA
## 3   CC   CC   CC   CC   CC   CC   CT   CT   TT   CT   CC   CT   CT   CT   CC
## 4   CC   CC   CC   CC   CC   CC   CC   CC   CC   CC   CT   CC   CT   CC   CC
## 5   GA   GG   GG   GG   GG   GG   GA   GA   GG   GA   GA   GG   GA   GA   GG
## 6   CC   TT   TT   CT   CC   CC   CC   CC   CC   CT   CT   CC   CC   CC   CC
##    UZHJ UZIJ UZJJ UZVJ UZZJ WZ1J ZZS8 ZZY8
## 1   GG   GC   CC   CC   CC   GC   CC   GC
## 2   AA   AA   GA   NN   NN   GA   AA   AA
## 3   CT   CC   CC   CC   CC   CC   CC   CC
## 4   TT   TT   CC   CC   CT   CC   CC   CC
## 5   GA   GA   GG   GG   GG   GA   GG   GG
## 6   CC   CT   TT   CC   CC   CC   CC   CC
```

```
# set rcol
rcol <- ncol(rad)
```

```
# make dataframe
```

```
concord<- data.frame("sample" = c(rep("NA", rcol)), "rad_only" = c(rep("NA", rcol)), "gt_only" = c(rep(
  "both" = c(rep("NA", rcol)), "n_concord" = c(rep("NA", rcol)), "p_con
  "mismatch" = c(rep("NA", rcol)), "mismatch_rate" = c(rep("NA", rcol))
  "both_hetero" = c(rep("NA", rcol)), "rad_homo_gt_het" = c(rep("NA", r
```

Run

```
for (i in 1:rcol) {
  temp_ind <- data.frame('rad'=rad[,i], 'gt'=gt[,i])
  gt_only <- nrow(filter(temp_ind, rad == "NN" & gt != "NN")) # RAD got NA, GT genotyped
  rad_only <- nrow(filter(temp_ind, rad != "NN" & gt == "NN")) # GT got NA, RAD genotyped
  n_concord <- nrow(filter(temp_ind, rad == "NN" & gt == "NN")) # both got NA
  both <- nrow(filter(temp_ind, rad != "NN" & gt != "NN")) # both genotyped

  temp_both <- filter(temp_ind, rad != "NN" & gt != "NN")
  p_concord <- nrow(filter(temp_both, rad==gt)) # when both got genotypes, they are the same
  mismatch <- nrow(filter(temp_both, rad!=gt)) # when both got genotypes, they are different calls
  temp_mismatch <- (filter(temp_both, rad!=gt))
  radhomo_gthet=0 # of mismatches, rad is homo, gt is hetero
  radhet_gthomo=0 # of mismatches, rad is hetero, gt is homo
  homo_homo_mm=0 # of mismatches, both are homo
  het_het_mm=0 # of mismatches, both are hetero

  if (mismatch >0){

    for (h in 1:mismatch){
      rada1 <- substr(temp_mismatch$rad[h],1,1)
      rada2 <- substr(temp_mismatch$rad[h],2,2)
      gta1 <- substr(temp_mismatch$gt[h],1,1)
      gta2 <- substr(temp_mismatch$gt[h],2,2)
      if (rada1 == rada2){
        radhomo = 1
      } else {
        radhomo = 0
      }
      if (gta1 == gta2){
        gthomo = 1
      } else {
        gthomo = 0
      }
      if (radhomo == 1 & gthomo == 0){
        radhomo_gthet <- radhomo_gthet +1
      } else if (radhomo == 0 & gthomo == 1){
        radhet_gthomo <- radhet_gthomo +1
      } else if (radhomo == 1 & gthomo == 1){
        homo_homo_mm <- homo_homo_mm +1
      } else if (radhomo == 0 & gthomo == 0) {
        het_het_mm <- het_het_mm +1
      }
    }
  }

  concord$sample[i] <- colnames(rad[i])
  concord$rad_only[i] <- rad_only
  concord$gt_only[i] <- gt_only
  concord$both[i] <- both
  concord$n_concord[i] <- n_concord
  concord$p_concord[i] <- p_concord
}
```

```
concord$mismatch[i] <- mismatch
concord$mismatch_rate[i] <- mismatch / both
concord$both_homo[i] <- homo_homo_mm
concord$both_hetero[i] <- het_het_mm
concord$rad_homo_gt_het[i] <- radhomo_gthet
concord$rad_het_gt_homo[i] <- radhet_gthomo

}

# Save out
write.csv(concord, './SNP_Panel/Read counts/Concordance_csvs_06132023/concord.csv', row.names = FALSE)
```