

# This is a note for Golden Gate Association analysis

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>. A quick review on the markdown cheatsheet here <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>.

This project is led by Ms. Jacqui Marzec. Our objectives are to find any significant association between SNPs (Golden Gate SNP candidate pool) in RSV positive patients.

## Association analysis on disease severity against GoldenGate SNP panels with several potential covariates.

In this analysis, we tried three different cases as following. Each setting will help to answer different hypothesis. The Analysis will be done with each case independently.

Table 1: Genotype case I

Disease-level	AA	AB	BB
Severe	SNPCnt	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt	SNPCnt

Table 2: Genotype case II

Disease-level	AA	AB/BB
Severe	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt

Table 3: Genotype case III

Disease-level	AA/AB	BB
Severe	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt

## Analysis set up and load necessary functions

```
suppressWarnings(library(HardyWeinberg))
suppressWarnings(library(XLConnect))
source("helperScripts/analysisFuncs.R")
```

## Load up preprocessed data

```
## Phenotype data
dt <- read.table("data/RSV_cohort_data_ASCII.txt",header = TRUE, sep= "\t")

## ONLY use for RSV positive candidates
row2use <- which (dt[,5] == 1)

## This is for reading Excel file
## We need to set memory at 24 GB
options(java.parameters = "-Xmx24g" )

snpsFiles <- "data/ASCII_genotype_data_GoldenGate_for_JYL_07122015.xlsx"

## There are five sheets of data
sheets <- c (seq (1,5))
```

## Analysis results on disease severity – logistic regression model

```
sig.asso.SAT <- c(list())
results.logit <- logisticAnalysis(sig.asso.SAT, snpsFiles, sheets,
  dt, 6, row2use)
results.logit
```

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##           SNP           pval
## 1 seq.rs3806496 0.03249843
## 2 seq.rs2066843 0.03161207
## 3 seq.rs17244587 0.04618051
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##           SNP           pval
## 1 seq.rs12987402 0.03525483
## 2 seq.rs469390 0.02409298
## 3 seq.rs3806265 0.03776518
## 4 seq.rs2066843 0.03371025
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##           SNP           pval
## 1 seq.rs17189298 0.04594427
## 2 seq.rs2770146 0.01637843
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##           SNP           pval
## 1 seq.rs2252930 0.0123745940
## 2 seq.rs2243592 0.0005072162
## 3 seq.rs2257167 0.0016167643
## 4 seq.rs2243599 0.0006175882
## 5 seq.rs2254180 0.0054136510
```

```

## 6 seq.rs2254315 0.0042648958
## 7 seq.rs1474347 0.0233376514
## 8 seq.rs8193036 0.0138688712
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##          SNP          pval
## 1 seq.rs2243592 0.003018021
## 2 seq.rs2257167 0.006062165
## 3 seq.rs2243599 0.006208847
## 4 seq.rs2254180 0.021865717
## 5 seq.rs2254315 0.015360578
## 6 seq.rs1800795 0.025821513
## 7 seq.rs2069832 0.012473056
## 8 seq.rs1474347 0.011935878
## 9 seq.rs1800893 0.048201983
## 10 seq.rs8193036 0.019022066
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
##          SNP          pval
## 1 seq.rs2280789 0.026207189
## 2 seq.rs2252930 0.010498389
## 3 seq.rs2243592 0.006836451
## 4 seq.rs2257167 0.011030534
## 5 seq.rs2243599 0.004429888
## 6 seq.rs2254180 0.019988072
## 7 seq.rs2254315 0.022174971
## 8 seq.rs9376268 0.042321508
## 9 seq.rs9376267 0.029287461
## 10 seq.rs369908 0.023580640
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##          SNP          pval
## 1 seq.rs1293749 0.041542545
## 2 seq.rs15895 0.034259581
## 3 seq.rs7966314 0.030823898
## 4 seq.rs2072133 0.007942976
## 5 seq.rs1041981 0.024407218
##
## $`OAS2 to TNFR`$`Case-II`
##          SNP          pval
## 1 seq.rs7966314 0.02788702
## 2 seq.rs2072133 0.01039301
##
## $`OAS2 to TNFR`$`Case-III`
##          SNP          pval
## 1 seq.rs1293755 0.04087340
## 2 seq.rs15895 0.04041382
## 3 seq.rs12693591 0.01505629
## 4 seq.rs3771300 0.03111144
## 5 seq.rs1800693 0.01013580
##
##

```

```
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##          SNP          pval
## 1 seq.rs3769823 0.0402697
##
## $`Additional_CASP8 to VDR`$`Case-II`
## NULL
##
## $`Additional_CASP8 to VDR`$`Case-III`
##          SNP          pval
## 1 seq.rs3769823 0.01005097
## 2 seq.rs4844590 0.03257378
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`
##          SNP          pval
## 1 rs3761624 0.02619375
##
## $`Genotyped prior`$`Case-II`
##          SNP          pval
## 1 rs3761624 0.02858216
##
## $`Genotyped prior`$`Case-III`
## NULL
```

## Analysis results on disease severity – logistic with gender as covariate

```
sig.asso.SAT <- c(list())
results.logit <- logisticAnalysisWCorrV(sig.asso.SAT, snpFiles,
  sheets, dt, 6, row2use, params$genStatus)
results.logit
```

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##          SNP          pval
## 1 seq.rs3738447 0.03095243
## 2 seq.rs2806117 0.02231901
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##          SNP          pval
## 1 seq.rs2806117 0.02550568
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##          SNP          pval
## 1 seq.rs2066842 0.02627682
## 2 seq.rs2917666 0.04877192
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##          SNP          pval
## 1 seq.rs2107538 0.01438484
```

```

## 2 seq.rs2280789 0.03260841
## 3 seq.rs3024492 0.02204524
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##          SNP          pval
## 1 seq.rs2107538 0.01714037
## 2 seq.rs4845623 0.01818976
## 3 seq.rs4129267 0.02053515
## 4 seq.rs3024492 0.01755576
## 5 seq.rs3848722 0.02404389
## 6 seq.rs443099 0.04551007
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
##          SNP          pval
## 1 seq.rs2280789 0.04637065
## 2 seq.rs6694817 0.01024136
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##          SNP          pval
## 1 seq.rs1293749 0.04924504
## 2 seq.rs12367468 0.04644929
## 3 seq.rs6581568 0.01465038
##
## $`OAS2 to TNFR`$`Case-II`
##          SNP          pval
## 1 seq.rs1293749 0.0165843840
## 2 seq.rs15895 0.0482663324
## 3 seq.rs12367468 0.0022164286
## 4 seq.rs6581568 0.0009284449
##
## $`OAS2 to TNFR`$`Case-III`
## NULL
##
##
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##          SNP          pval
## 1 seq.rs2229095 0.04752476
##
## $`Additional_CASP8 to VDR`$`Case-II`
## NULL
##
## $`Additional_CASP8 to VDR`$`Case-III`
## NULL
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`
##          SNP          pval
## 1 rs6721961 0.04200231
##
## $`Genotyped prior`$`Case-II`

```

```
##          SNP          pval
## 1 rs2569190 0.04132563
##
## $`Genotyped prior`$`Case-III`
## NULL
```

## Analysis results on disease severity – logistic with region as covariate

```
sig.asso.SAT <- c(list())
results.logit <- logisticAnalysisWCorrV(sig.asso.SAT, snpFiles,
    sheets, dt, 6, row2use, params$regStatus)
results.logit
```

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##          SNP          pval
## 1 seq.rs5743291 0.007589368
## 2 seq.rs17244587 0.029565288
## 3 seq.rs4986790 0.037814172
## 4 seq.rs10759930 0.024301793
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##          SNP          pval
## 1 seq.rs6722472 0.033356496
## 2 seq.rs3731612 0.023233536
## 3 seq.rs5743291 0.007166732
## 4 seq.rs17244587 0.024779377
## 5 seq.rs75059843 0.013924264
## 6 seq.rs4986790 0.037448234
## 7 seq.rs4986791 0.025379831
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##          SNP          pval
## 1 seq.rs2917666 0.01499783
## 2 seq.rs3936292 0.01410505
## 3 seq.rs7022323 0.02726623
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##          SNP          pval
## 1 seq.rs7296163 0.046356411
## 2 seq.rs9272346 0.033613178
## 3 seq.rs2252930 0.007116561
## 4 seq.rs2243592 0.018099573
## 5 seq.rs2243599 0.049535474
## 6 seq.rs6517154 0.029281023
## 7 seq.rs1554606 0.039698733
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##          SNP          pval
## 1 seq.rs13069802 0.048974384
## 2 seq.rs2252930 0.006260596
```

```

## 3 seq.rs2243592 0.006443205
## 4 seq.rs2243599 0.042282970
## 5 seq.rs6517154 0.021185432
## 6 seq.rs1554606 0.010554988
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
##          SNP          pval
## 1 seq.rs1046161 0.02513435
## 2 seq.rs9272346 0.03415691
## 3 seq.rs2250226 0.04812289
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##          SNP          pval
## 1 seq.rs1293745 0.03445199
## 2 seq.rs1800630 0.04190544
## 3 seq.rs1041981 0.03180766
##
## $`OAS2 to TNFR`$`Case-II`
##          SNP          pval
## 1 seq.rs2072137 0.007765022
## 2 seq.rs7953402 0.046495039
## 3 seq.rs1799964 0.044793060
## 4 seq.rs1800630 0.047761580
## 5 seq.rs1041981 0.045486371
##
## $`OAS2 to TNFR`$`Case-III`
##          SNP          pval
## 1 seq.rs1293745 0.04176694
## 2 seq.rs2280232 0.03598239
##
##
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##          SNP          pval
## 1 seq.rs3890733 0.007209354
##
## $`Additional_CASP8 to VDR`$`Case-II`
##          SNP          pval
## 1 seq.rs11208545 0.03722931
## 2 seq.rs6580642 0.04844294
##
## $`Additional_CASP8 to VDR`$`Case-III`
##          SNP          pval
## 1 seq.rs3890733 0.003656025
## 2 seq.rs11168293 0.004391315
## 3 seq.rs4516035 0.011011403
## 4 seq.rs7139166 0.004782714
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`
##          SNP          pval

```

```
## 1 rs2569190 0.03129927
## 2 rs6721961 0.02408053
##
## $`Genotyped prior`$`Case-II`
##      SNP      pval
## 1 rs2569190 0.01943644
## 2 rs6721961 0.01451079
##
## $`Genotyped prior`$`Case-III`
## NULL
```

## Analysis results on disease severity – logistic with social as covariate

```
sig.asso.SAT <- c(list())
results.logit <- logisticAnalysisWCorrV(sig.asso.SAT, snpFiles,
      sheets, dt, 6, row2use, params$sesStatus)
results.logit
```

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##      SNP      pval
## 1 seq.rs2193035 0.034177766
## 2 seq.rs5743291 0.004372155
## 3 seq.rs4986790 0.005361488
## 4 seq.rs4986791 0.027518773
## 5 seq.rs10759930 0.015191125
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##      SNP      pval
## 1 seq.rs2193035 0.036090953
## 2 seq.rs5743291 0.003114019
## 3 seq.rs75059843 0.049610011
## 4 seq.rs1927911 0.017150741
## 5 seq.rs4986790 0.003999410
## 6 seq.rs4986791 0.006687538
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##      SNP      pval
## 1 seq.rs3936292 0.02484239
## 2 seq.rs7022323 0.02953915
## 3 seq.rs10759930 0.02168704
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##      SNP      pval
## 1 seq.rs7296163 0.016802503
## 2 seq.rs2252930 0.009035721
## 3 seq.rs2243592 0.023568224
## 4 seq.rs2243599 0.038480553
## 5 seq.rs6517154 0.037488503
## 6 seq.rs3024496 0.045334709
##
```



```

## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##          SNP          pval
## 1 seq.rs7296163 0.021309670
## 2 seq.rs2252930 0.003443911
## 3 seq.rs2243592 0.019517942
## 4 seq.rs6517154 0.020560818
## 5 seq.rs10211925 0.043408749
## 6 seq.rs9376268 0.046019117
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
##          SNP          pval
## 1 seq.rs1046161 0.02900484
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##          SNP          pval
## 1 seq.rs1800630 0.03458350
## 2 seq.rs1041981 0.01775615
##
## $`OAS2 to TNFR`$`Case-II`
##          SNP          pval
## 1 seq.rs2072137 0.03815757
## 2 seq.rs7953402 0.03677857
## 3 seq.rs1800630 0.03404981
## 4 seq.rs1041981 0.02103300
##
## $`OAS2 to TNFR`$`Case-III`
## NULL
##
##
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##          SNP          pval
## 1 seq.rs11574663 0.04743326
## 2 seq.rs11208545 0.04210803
## 3 seq.rs3890733 0.00734571
##
## $`Additional_CASP8 to VDR`$`Case-II`
##          SNP          pval
## 1 seq.rs3769823 0.04331380
## 2 seq.rs11208545 0.01539316
##
## $`Additional_CASP8 to VDR`$`Case-III`
##          SNP          pval
## 1 seq.rs2734648 0.046999929
## 2 seq.rs3890733 0.008839898
## 3 seq.rs11168293 0.009520885
## 4 seq.rs4516035 0.012611197
## 5 seq.rs7139166 0.006271190
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`

```

```
## NULL
##
## $`Genotyped prior`$`Case-II`
##      SNP      pval
## 1 rs2569190 0.02675012
##
## $`Genotyped prior`$`Case-III`
## NULL
```

## Analysis results on disease severity – logistic with breastfeeding as covariate

```
sig.asso.SAT <- c(list())
results.logit <- logisticAnalysisWCorrV(sig.asso.SAT, snpFiles,
    sheets, dt, 6, row2use, params$brfStatus)
results.logit
```

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##      SNP      pval
## 1 seq.rs4925659 0.04857201
## 2 seq.rs7502875 0.02966858
## 3 seq.rs5030725 0.03166908
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##      SNP      pval
## 1 seq.rs11795343 0.03356908
## 2 seq.rs10813831 0.03897348
## 3 seq.rs7502875 0.04425052
## 4 seq.rs5030725 0.03166908
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##      SNP      pval
## 1 seq.rs469012 0.02343009
## 2 seq.rs469390 0.01958672
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##      SNP      pval
## 1 seq.rs11063084 0.045638507
## 2 seq.rs2107538 0.009151977
## 3 seq.rs2257167 0.017594107
## 4 seq.rs2275913 0.013270682
## 5 seq.rs8193038 0.014041868
## 6 seq.rs7744 0.026364135
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##      SNP      pval
## 1 seq.rs2107538 0.018246758
## 2 seq.rs2275913 0.008774505
## 3 seq.rs8193038 0.006582946
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
```

```

##          SNP          pval
## 1 seq.rs2257167 0.02022729
## 2 seq.rs2250889 0.04664627
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##          SNP          pval
## 1 seq.rs16824035 0.005067530
## 2 seq.rs1800692 0.003280719
##
## $`OAS2 to TNFR`$`Case-II`
##          SNP          pval
## 1 seq.rs10199181 0.042982420
## 2 seq.rs16824035 0.006314110
## 3 seq.rs1800692 0.009707555
##
## $`OAS2 to TNFR`$`Case-III`
## NULL
##
##
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##          SNP          pval
## 1 seq.rs11208545 0.03037701
##
## $`Additional_CASP8 to VDR`$`Case-II`
##          SNP          pval
## 1 seq.rs310202 0.01642526
## 2 seq.rs11208545 0.02557307
##
## $`Additional_CASP8 to VDR`$`Case-III`
##          SNP          pval
## 1 seq.rs310247 0.03893488
## 2 seq.rs12743599 0.01310665
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`
## NULL
##
## $`Genotyped prior`$`Case-II`
## NULL
##
## $`Genotyped prior`$`Case-III`
## NULL

```

## Analysis results on disease severity – chi-square test based model

```

sig.asso.SAT <- c(list())
results.chisquare <- AssoAnalysis(sig.asso.SAT, snpFiles, sheets,
  dt, 6, row2use)

```

# results.chisquare

```
## $`Top Tier IRF7 to TLR4`
## $`Top Tier IRF7 to TLR4`$`Case-I`
##           SNP           pval
## 1  seq.rs469390 0.03553412
## 2  seq.rs2770146 0.03031926
##
## $`Top Tier IRF7 to TLR4`$`Case-II`
##           SNP           pval
## 1  seq.rs12987402 0.04484957
## 2  seq.rs469390 0.03136834
## 3  seq.rs3806265 0.04864552
## 4  seq.rs2066843 0.04292077
##
## $`Top Tier IRF7 to TLR4`$`Case-III`
##           SNP           pval
## 1  seq.rs2770146 0.02269727
##
##
## $`BPIIFA1(PLUNC) to MYD88`
## $`BPIIFA1(PLUNC) to MYD88`$`Case-I`
##           SNP           pval
## 1  seq.rs2252930 0.032963115
## 2  seq.rs2243592 0.002065785
## 3  seq.rs2257167 0.004097110
## 4  seq.rs2243599 0.002391277
## 5  seq.rs2254180 0.016497081
## 6  seq.rs2254315 0.014203726
## 7  seq.rs2069832 0.026046287
## 8  seq.rs1474347 0.040922019
## 9  seq.rs8193036 0.046061333
## 10 seq.rs8193038 0.049507445
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-II`
##           SNP           pval
## 1  seq.rs2243592 0.003973553
## 2  seq.rs2257167 0.007940527
## 3  seq.rs2243599 0.008163886
## 4  seq.rs2254180 0.028029801
## 5  seq.rs2254315 0.019873260
## 6  seq.rs1800795 0.033217808
## 7  seq.rs2069832 0.016223152
## 8  seq.rs1474347 0.015568224
## 9  seq.rs8193036 0.024515805
##
## $`BPIIFA1(PLUNC) to MYD88`$`Case-III`
##           SNP           pval
## 1  seq.rs2280789 0.037358791
## 2  seq.rs2252930 0.013686847
## 3  seq.rs2243592 0.009214950
## 4  seq.rs2257167 0.014382366
## 5  seq.rs2243599 0.005967165
## 6  seq.rs2254180 0.027873558
```

```

## 7  seq.rs2254315 0.030957836
## 8  seq.rs9376268 0.033497757
## 9  seq.rs9376267 0.035090018
## 10 seq.rs369908 0.033457931
##
##
## $`OAS2 to TNFR`
## $`OAS2 to TNFR`$`Case-I`
##      SNP      pval
## 1  seq.rs2072133 0.02734773
## 2  seq.rs12693591 0.01358428
## 3  seq.rs361525 0.02179245
## 4  seq.rs1800693 0.01481453
##
## $`OAS2 to TNFR`$`Case-II`
##      SNP      pval
## 1  seq.rs7966314 0.03569512
## 2  seq.rs2072133 0.01355917
##
## $`OAS2 to TNFR`$`Case-III`
##      SNP      pval
## 1  seq.rs12693591 0.02004173
## 2  seq.rs3771300 0.03990646
## 3  seq.rs361525 0.02179672
## 4  seq.rs1800693 0.01382160
##
##
## $`Additional_CASP8 to VDR`
## $`Additional_CASP8 to VDR`$`Case-I`
##      SNP      pval
## 1  seq.rs3769823 0.03369201
##
## $`Additional_CASP8 to VDR`$`Case-II`
## NULL
##
## $`Additional_CASP8 to VDR`$`Case-III`
##      SNP      pval
## 1  seq.rs3769823 0.01319883
## 2  seq.rs4844590 0.04305651
##
##
## $`Genotyped prior`
## $`Genotyped prior`$`Case-I`
## NULL
##
## $`Genotyped prior`$`Case-II`
##      SNP      pval
## 1  rs3761624 0.03637326
##
## $`Genotyped prior`$`Case-III`
## NULL

```