

ADSP WES GWAS by GLMM

Xulong Wang

September 18, 2016

In an earlier GWAS of ADSP WES dataset, a missense variant in CR1 gene (1:207782889) showed significant association with AD status. This analysis addresses the following question: does this variant interact with APOE genotypes?

Genotype of the CR1 variant: 1:207782889

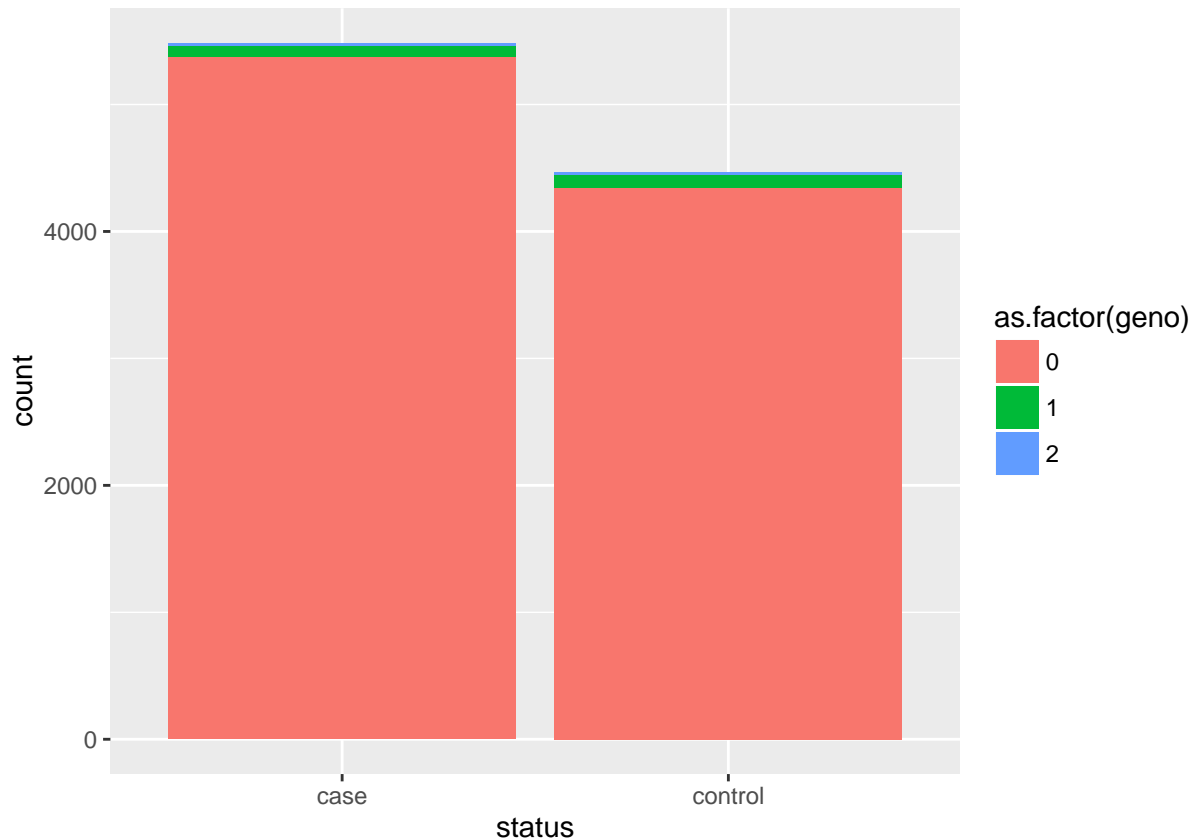
```
load("./mdata.rdt")
load("./genes/cr1.rdt")

geno = z[, mdata$ADSP_SM_ID]
rownames(geno)
```

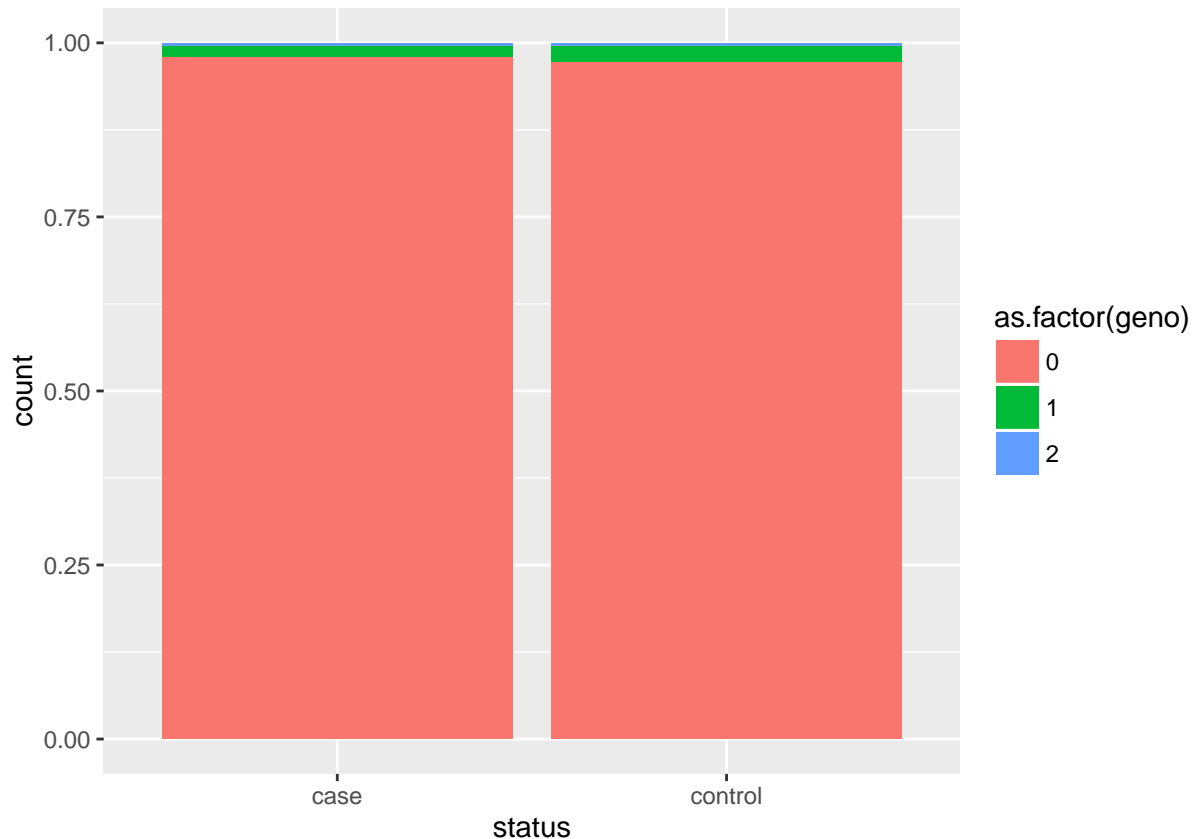
```
## [1] "1:207782889" "1:207782916"
```

```
data = within(mdata, { geno = geno[1, ] })
data$y = as.numeric(data$status == "case")

ggplot(data, aes(x = status, fill = as.factor(geno))) + geom_bar()
```



```
ggplot(data, aes(x = status, fill = as.factor(geno))) + geom_bar(position = "fill")
```



```
(tbl <- table(data$geno, data$status))
```

```
##
##      case control
##  0 5377    4344
##  1   87     103
##  2   20      18
```

Tested allele of the variant showed significant negative association with AD status

```
tmp = data
glm(y ~ Age + Sex + geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0725  -0.7000   0.1300   0.5573   3.1793
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 20.197371   0.415171  48.648 < 2e-16 ***
## Age        -0.248456   0.005007 -49.622 < 2e-16 ***
## Sex         0.413146   0.055473   7.448 9.5e-14 ***
## geno       -1.441906   0.152987  -9.425 < 2e-16 ***
## lsscBroad   0.225440   0.066806   3.375 0.000739 ***
## lsscWashU  -0.027790   0.071602  -0.388 0.697925
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 13687.7 on 9948 degrees of freedom
## Residual deviance: 8595.5 on 9943 degrees of freedom
## AIC: 8607.5
##
## Number of Fisher Scoring iterations: 5
```

Interaction was not detected by including all APOE genotypes

```
tmp = data
glm(y ~ Age + Sex + APOE * geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + APOE * geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8933  -0.7174   0.1298   0.5490   3.3128
##
## Coefficients: (1 not defined because of singularities)
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 19.932350   0.439419  45.361 < 2e-16 ***
## Age        -0.244017   0.005241 -46.562 < 2e-16 ***
## Sex         0.414875   0.056031   7.404 1.32e-13 ***
## APOE22      -0.805763   0.336519  -2.394 0.01665 *
## APOE23      -0.828169   0.084538  -9.796 < 2e-16 ***
## APOE24      -0.312836   0.234993  -1.331 0.18310
## APOE34       0.224388   0.074898   2.996 0.00274 **
## APOE44      -0.235808   0.531043  -0.444 0.65701
## geno       -1.373398   0.211072  -6.507 7.68e-11 ***
## lsscBroad   0.204726   0.067395   3.038 0.00238 **
## lsscWashU  -0.074461   0.072364  -1.029 0.30349
## APOE22:geno      NA         NA      NA      NA
## APOE23:geno   0.060501   0.624221   0.097 0.92279
## APOE24:geno  -0.344633   0.840450  -0.410 0.68176
## APOE34:geno  -0.256436   0.314049  -0.817 0.41419
## APOE44:geno   8.445095 196.968511   0.043 0.96580
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13687.7  on 9948  degrees of freedom
## Residual deviance:  8460.2  on 9934  degrees of freedom
## AIC: 8490.2
##
## Number of Fisher Scoring iterations: 10
```

Interaction was not detected by including each APOE genotype separately

```
tmp = data
glm(y ~ Age + Sex + Apoe2 * geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + Apoe2 * geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9076  -0.6839   0.1234   0.5449   3.3691
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  20.326062   0.422109  48.154 < 2e-16 ***
## Age         -0.248328   0.005076 -48.924 < 2e-16 ***
## Sex          0.409369   0.055899   7.323 2.42e-13 ***
## Apoe2       -0.765117   0.073273 -10.442 < 2e-16 ***
## geno        -1.478291   0.160721  -9.198 < 2e-16 ***
## lsscBroad    0.218155   0.067267   3.243 0.00118 **
## lsscWashU   -0.055367   0.072065  -0.768 0.44232
## Apoe2:geno   0.036553   0.500952   0.073 0.94183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13687.7  on 9948  degrees of freedom
## Residual deviance:  8478.3  on 9941  degrees of freedom
## AIC: 8494.3
##
## Number of Fisher Scoring iterations: 5
```

```
glm(y ~ Age + Sex + Apoe3 * geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + Apoe3 * geno + lssc, family = "binomial",
```

```

##      data = tmp)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -4.0185   -0.6773    0.1329    0.5808    3.2488
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 20.218315   0.414592  48.767 < 2e-16 ***
## Age         -0.253592   0.005153 -49.212 < 2e-16 ***
## Sex          0.406651   0.055556   7.320 2.49e-13 ***
## Apoe3        0.247684   0.053165   4.659 3.18e-06 ***
## geno        -1.474993   0.415030  -3.554 0.000380 ***
## lsscBroad    0.237099   0.066996   3.539 0.000402 ***
## lsscWashU   -0.018459   0.071707  -0.257 0.796853
## Apoe3:geno   0.017537   0.263383   0.067 0.946913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13687.7  on 9948  degrees of freedom
## Residual deviance:  8572.8  on 9941  degrees of freedom
## AIC: 8588.8
##
## Number of Fisher Scoring iterations: 5

```

```

glm(y ~ Age + Sex + Apoe4 * geno + lssc, data = tmp, family = "binomial") %>% summary

```

```

##
## Call:
## glm(formula = y ~ Age + Sex + Apoe4 * geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -4.0271   -0.7098    0.1261    0.5682    3.1063
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 19.650605   0.431177  45.574 < 2e-16 ***
## Age         -0.242528   0.005164 -46.962 < 2e-16 ***
## Sex          0.420281   0.055584   7.561 3.99e-14 ***
## Apoe4        0.315394   0.068741   4.588 4.47e-06 ***
## geno        -1.319730   0.198733  -6.641 3.12e-11 ***
## lsscBroad    0.207994   0.066901   3.109 0.00188 **
## lsscWashU   -0.051335   0.071846  -0.715 0.47491
## Apoe4:geno  -0.300394   0.297110  -1.011 0.31199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13687.7  on 9948  degrees of freedom

```

```
## Residual deviance: 8574.6 on 9941 degrees of freedom
## AIC: 8590.6
##
## Number of Fisher Scoring iterations: 5
```

Stratify population by APOE genotypes suggested the protective effect of the allele was most profound in APOE/E4 carriers. Phil De Jager's paper reported this as well.

```
tmp = data[data$Apoe3 == 2, ]
glm(y ~ Age + Sex + geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4149  -0.7960  -0.4974   0.7751   2.9623
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 18.400123   0.555387  33.130 < 2e-16 ***
## Age         -0.226970   0.006645 -34.158 < 2e-16 ***
## Sex          0.579313   0.068112   8.505 < 2e-16 ***
## geno        -1.242933   0.208003  -5.976 2.29e-09 ***
## lsscBroad    0.203638   0.079715   2.555  0.0106 *
## lsscWashU    -0.088964   0.086045  -1.034  0.3012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 7840.0 on 5662 degrees of freedom
## Residual deviance: 5720.1 on 5657 degrees of freedom
## AIC: 5732.1
##
## Number of Fisher Scoring iterations: 5
```

```
tmp = data[data$Apoe2 > 0, ]
glm(y ~ Age + Sex + geno + lssc, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Age + Sex + geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1735  -0.8035  -0.5736   0.7341   2.7879
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.18354    0.91384  14.426 <2e-16 ***
## Age         -0.17030    0.01115 -15.270 <2e-16 ***
## Sex          0.33997    0.13425   2.532  0.0113 *
## geno        -0.86082    0.43677  -1.971  0.0487 *
## lsscBroad    0.08224    0.15559   0.529  0.5971
## lsscWashU    -0.14096    0.17662  -0.798  0.4248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 1769.0  on 1379  degrees of freedom
## Residual deviance: 1438.1  on 1374  degrees of freedom
## AIC: 1450.1
##
## Number of Fisher Scoring iterations: 4

tmp = data[data$Apoe4 > 0, ]
glm(y ~ Age + Sex + geno + lssc, data = tmp, family = "binomial") %>% summary

##
## Call:
## glm(formula = y ~ Age + Sex + geno + lssc, family = "binomial",
##      data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.9224   0.0292   0.1176   0.2778   3.0334
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 28.8636    1.0605  27.217 < 2e-16 ***
## Age         -0.3471    0.0128 -27.126 < 2e-16 ***
## Sex         -0.3899    0.1502  -2.596  0.00942 **
## geno        -2.2716    0.2651  -8.570 < 2e-16 ***
## lsscBroad    0.1877    0.2054   0.914  0.36087
## lsscWashU    0.1025    0.2063   0.497  0.61916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 3222.6  on 3087  degrees of freedom
## Residual deviance: 1241.4  on 3082  degrees of freedom
## AIC: 1253.4
##
## Number of Fisher Scoring iterations: 7
```

APOE genotype stratification versus interaction model

```
tmp = data[c("y", "geno", "Apoe3")]
tmp$geno = as.numeric(tmp$geno > 0)
tmp$Apoe3 = as.numeric(tmp$Apoe3 > 0)

with(tmp, table(geno, Apoe3))
```

```
##      Apoe3
## geno    0    1
##      0  411 9310
##      1    9  219
```

```
glm(y ~ Apoe3 * geno, data = tmp, family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ Apoe3 * geno, family = "binomial", data = tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.585  -1.257   1.100   1.100   1.236
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9214     0.1093   8.429 < 2e-16 ***
## Apoe3         -0.7370     0.1113  -6.623 3.51e-11 ***
## geno          -0.6983     0.6797  -1.027  0.304
## Apoe3:geno     0.3767     0.6933   0.543  0.587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13688  on 9948  degrees of freedom
## Residual deviance: 13634  on 9945  degrees of freedom
## AIC: 13642
##
## Number of Fisher Scoring iterations: 4
```

```
glm(y ~ geno, data = tmp[tmp$Apoe3 == 0, ], family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ geno, family = "binomial", data = tmp[tmp$Apoe3 ==
##      0, ])
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5852  -1.5852   0.8185   0.8185   1.0842
##
```



```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.9214     0.1093   8.429  <2e-16 ***
## geno        -0.6983     0.6797  -1.027   0.304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 504.37  on 419  degrees of freedom
## Residual deviance: 503.36  on 418  degrees of freedom
## AIC: 507.36
##
## Number of Fisher Scoring iterations: 4
```

```
glm(y ~ geno, data = tmp[tmp$Apoe3 == 1, ], family = "binomial") %>% summary
```

```
##
## Call:
## glm(formula = y ~ geno, family = "binomial", data = tmp[tmp$Apoe3 ==
##    1, ])
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -1.257  -1.257   1.100   1.100   1.236
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.18441     0.02082   8.859  <2e-16 ***
## geno        -0.32161     0.13706  -2.347   0.0189 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 13136  on 9528  degrees of freedom
## Residual deviance: 13130  on 9527  degrees of freedom
## AIC: 13134
##
## Number of Fisher Scoring iterations: 3
```