POLYGENIC RISK SCORE OF SPORADIC LATE ONSET ALZHEIMER DISEASE REVEALS A SHARED ARCHITECTURE WITH THE FAMILIAL AND EARLY ONSET FORMS

Supplementary Information

Supplementary Table 1: SNPs and Odd ratios employed to derive the PRS

Reported SNP	Tagging SNP	r^2	D'	Chr	Position	Gene	OR
rs6656401	rs2093761	0.9	0.97	1	207786542	CR1	1.25
rs6733839	rs7561528	0.4	0.70	2	127889637	BIN1	1.24
rs35349669				2	234068476	INPP5D	1.12
rs111418223				6	32578530	HLA-DRB5-HLA-DRB1	1.11
rs10948363				6	47487762	CD2AP	1.15
rs2718058				7	37841534	NME8	0.90
rs1476679				7	100004446	ZCWPW1	0.87
rs11771145				7	143110762	EPHA1	0.86
rs28834970				8	27195121	PTK2B	1.15
rs9331896	rs11136000	0.9	0.97	8	27464519	CLU	0.82
rs10838725				11	47557871	CELF1	1.12
rs983392				11	59923508	MS4A6A	0.86
rs10792832				11	85867875	PICALM	0.82
rs11218343				11	121435587	SORL1	0.69
rs17125944				14	53400629	FERMT2	1.21
rs4147929				19	1063443	ABCA7	1.22
rs3865444				19	51727962	CD33	0.91
rs7274581	rs927174	0.9	1.00	20	55020689	CASS4	0.84

Supplemental Table 2. Quantile regression for Polygenic Risk Score and the Age at Onset for the combined early and late onset sporadic cohorts. Panel A: Minimal model that corrects for sex and study that ascertained the participants. Panel B: Extended models that corrects for the effects of APOE alleles.

Minimal Model

Quantile	Effect*	SE	*p-value*	Variable
1	-11.83	4.10	3.94E-03	PRS
2	-12.27	3.56	5.88E-04	PRS
3	-9.53	3.16	2.58E-03	PRS
4	-9.37	3.80	1.38E-02	PRS
5	-8.73	4.04	3.06E-02	PRS

Extended Model

Quantile	Effect*	SE	*p-value*	Variable
1	-13.89	4.12	7.63E-04	PRS
1	-12.63	2.76	5.29E-06	APOE
2	-9.59	3.36	4.33E-03	PRS
2	-17.78	2.40	2.37E-13	APOE
3	-10.91	2.91	1.89E-04	PRS
3	-22.06	2.03	<1.00E-16	APOE
4	-4.89	3.46	1.58E-01	PRS
4	-25.68	2.57	<1.00E-16	APOE
5	-8.97	3.76	1.72E-02	PRS
5	-28.04	2.80	<1.00E-16	APOE

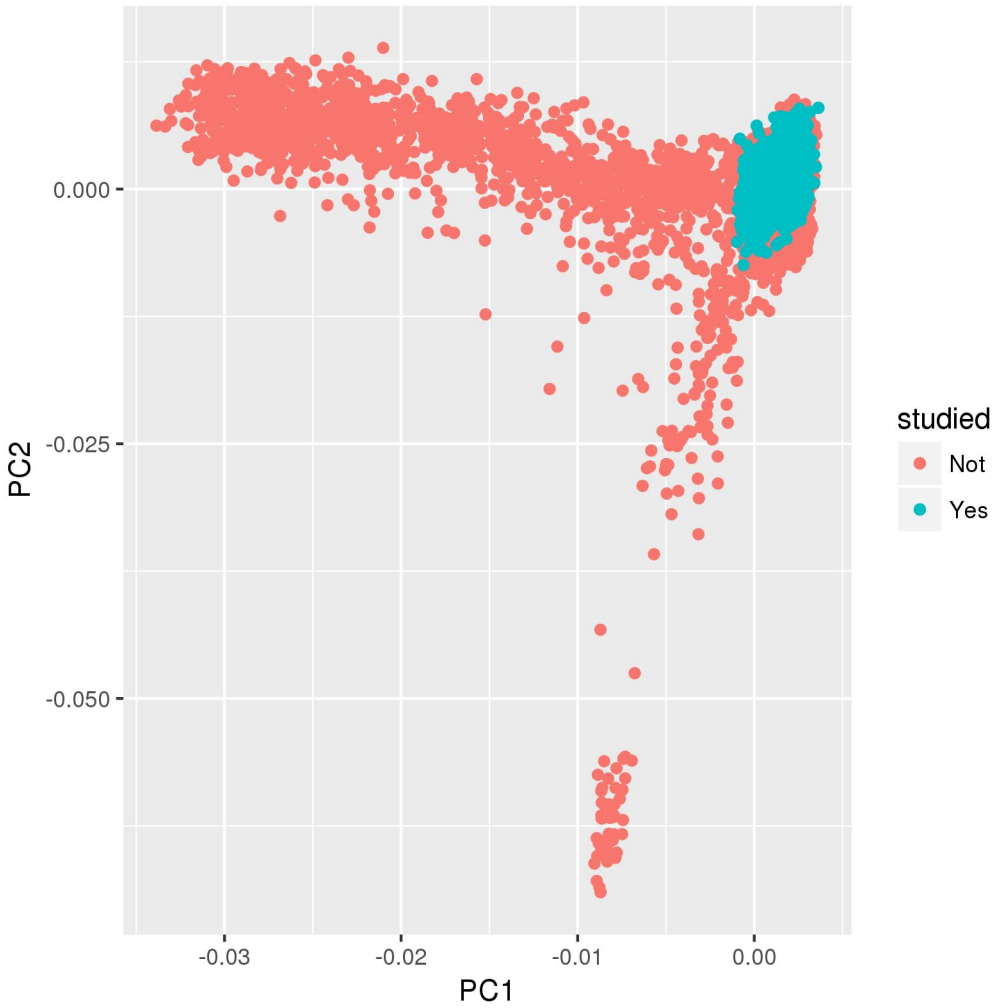
The effect display is the months associated per unit of SD

Supplemental Table 3: Association results of the logistic regression models for Polygenic Risk Scores for the entire range of PRS derived for each of the cohorts and compared to non-demented participants

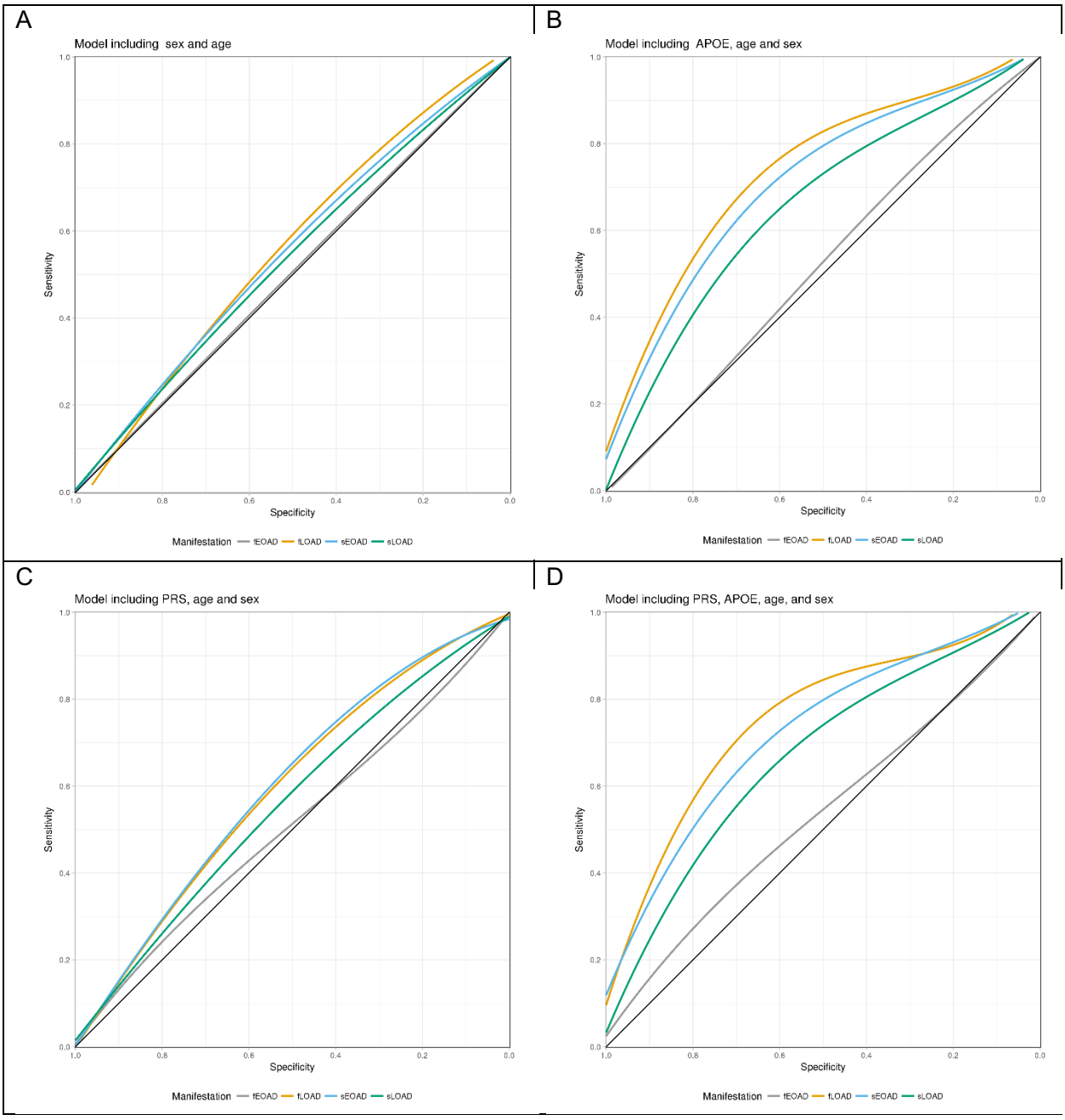
Cohort	PRS (model 1)		PRS w/APOE effects (model 2)		PRS APOE corrected (model 3)	
	OR	*p-value*	OR	*p-value*	OR	*p-value*
Autosomal Dominant Early Onset AD	1.1	5.52×10^{-1}	0.99	9.03×10^{-1}	1.25	1.94×10^{-1}
Sporadic Early Onset AD	2.03	1.85×10^{-6}	1.92	3.65×10^{-36}	2.01	1.38×10^{-5}
Familial Late Onset AD	1.77	1.27×10^{-8}	2.08	2.19×10^{-48}	2.08	1.16×10^{-7}
Sporadic Late Onset AD	1.48	6.05×10^{-5}	1.68	2.21×10^{-39}	1.51	6.98×10^{-5}

Supplemental Table 4: Association results of the logistic regression models for extreme tertiles of the Polygenic Risk Scores derived for the subjects not included in the IGAP study compared to non-demented participants

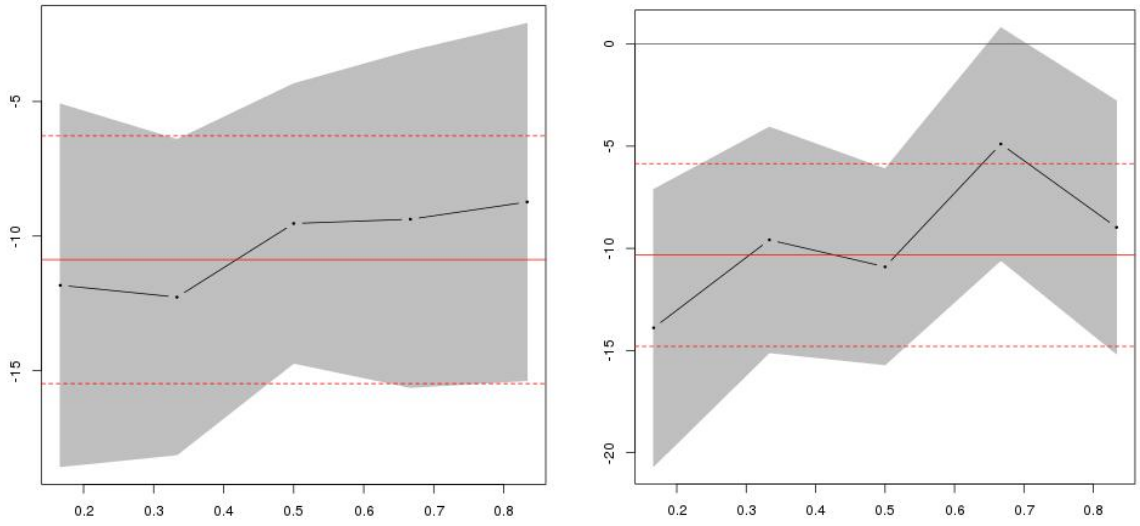
	PRS (model 1)		**PRS APOE corrected (model 2)**		**PRS w/APOE effects (model 3)**	
Cohort	OR	*p-value*	OR	*p-value*	OR	*p-value*
Familial Late Onset AD	1.7	5.53×10^{-3}	1.81	3.60×10^{-3}	6.49	3.49×10^{-4}
Sporadic Late Onset AD	1.22	3.27×10^{-2}	1.23	3.58×10^{-2}	3.79	7.10×10^{-3}



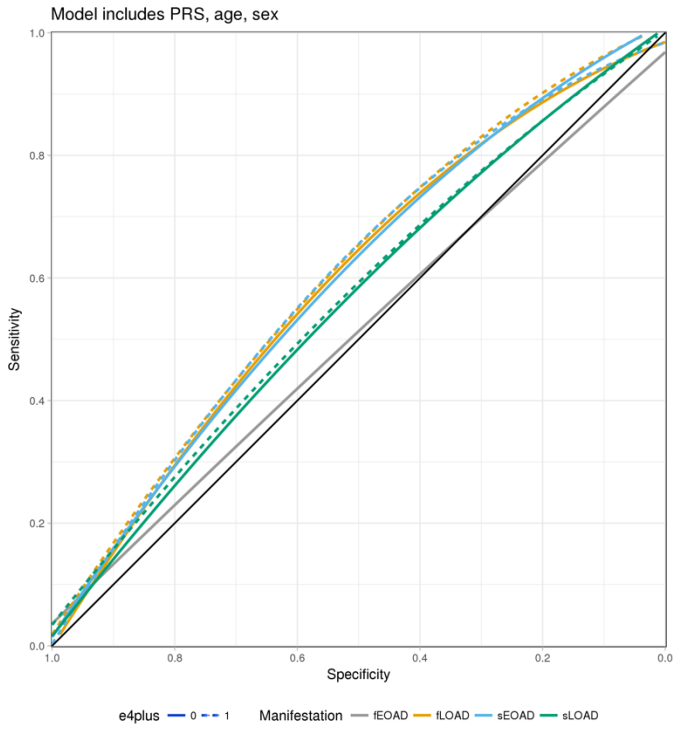
Supplementary Figure 1: Population stratification. We calculated the principal component factors for each sample to confirm the ethnicity of the samples (blue: samples included in the analysis)



Supplementary Figure 2. ROC curves for the distinct datasets using alternative models. A: Model includes gender, and age (for the late-onset dataset). B) Extends previous model including the effect conferred by *APOE* genotypes. C) Extends the model in (A) including the PRS that includes the GWAS loci but not APOE. D) Extends the model in (A) including both *APOE* and PRS effects. Please refer to Supplementary Figure 4 for the ROC curves of model 2 stratified by APOE e4 alleles.



Supplementary Figure 3. Quintile (5-quantile) regression estimates for the Polygenic Risk Score and the Age at Onset for the combined early- and late-onset sporadic cohorts. The solid red line represents the estimates of calculated by the Ordinary Least Square method, and the dashed red lines the 95% confidence interval. Left Panel: Minimal model that corrects for the sex and study that ascertained the participant. Right Panel: Extended model that also correct for the risk conferred by *APOE* genotypes



Supplemental Figure 4: ROC curves representing the sensitivity and specificity of the models including PRS, gender and age (for the late-onset datasets) stratified APOE e4 carrier status.

Supplementary Information.

R code employed to evaluate the statistical significance of the association of the Polygenic Risk Score. We also include R code to generate toy datasets to allow the initial evaluation of the test. The code is also available in plain text format (file PRS.r)

```
# test.GRS: Evaluates the Odd ratios utilizing logistic regression and quantiles for
the Polygenic risk scores
# Parameters
#     data: dataframe containing the phenotypic and polygenic risk score calculated
for each individual
#           (e.g cohort, score, gender, age.test (either  Age at Onset for cases or
Age of Last Assessment for controls)
#     variable:  Indicates which column of the dataset "data" describes to which
cohort the samples are asigned (e.g. "cohort" )
#     cohort1:   Indicates  the first cohort which is being compared (e.g. a value
"Sporadic EO"  )
#     cohort2:   Indicates  the second cohort which is being compared (e.g. a value
"Control"  )
#     score_var: Indicates which column of the dataframe "data" contains the PRS
calculated for each subject
#     ntiles:     Refers to the number of quantiles (either 3 or 4) to use
# Returns:
# Eaxmple test.GRS(database,"cohort","Sporadic Early
Onset","Control","Log_OR_Score",ntiles=3)
test.GRS = function ( data, variable, cohort1, cohort2, score_var, ntiles=3,
verbose=T) {

  sel = data[,variable] %in% c( cohort1, cohort2)
  data.test = data[sel,]
  data.test$test =  data.test[,variable]
  data.test$riskScore =  data.test[,score_var]

  if (verbose)  {
    cat(  sprintf( "Cohorts %s vs. %s\n", cohort1, cohort2) )
    print( table( data.test$test) )
  }
  data.test <- data.test[!is.na(data.test$riskScore),]
  data.test$quant = NA
  if ( ntiles == 3 ) {
    q = quantile(data.test$riskScore, probs=c( 0, 33, 66, 100)/100, na.rm = TRUE)
    data.test[data.test$riskScore <= q[2],]$quant = 0
    data.test[data.test$riskScore >= q[3],]$quant = 1
  }
  else {
    q = quantile(data.test$riskScore, probs=c( 0, 25, 50, 75, 100)/100, na.rm = TRUE)
    data.test[data.test$riskScore <= q[2],]$quant = 0
    data.test[data.test$riskScore >= q[4],]$quant = 1
  }
  cat( "\n")
  cat(  sprintf( "Test the basic PRS model %s vs. %s. Quantiles=%i\n", cohort1,
cohort2, ntiles ) )
  m = glm(test == cohort1   ~ quant , family = binomial(link='logit'), data=data.test)
  dump.model( m)
  if ( verbose ) {
    print(summary(m))
  }
  cat(  sprintf( "Test the PRS, corrected by gender\n") )
  m = glm(test == cohort1   ~ quant + gender, family = binomial(link='logit'),
data=data.test)
```

```
  dump.model( m)
  if ( verbose ) {
    print(summary(m))
  }
  cat(  sprintf( "Test the PRS, corrected by age\n") )
  m = glm(test == cohort1   ~ quant + age.test, family = binomial(link='logit'),
data=data.test)
  res = try( dump.model( m), silent = T )
  if(inherits(res, "try-error")) {
    print( "Could not get CI when Age is included into the model")
  }
  if ( verbose ) {
    print(summary(m))
  }
  cat(  sprintf( "Test the PRS corrected by  gender and age model\n") )
  m = glm(test == cohort1   ~ quant + gender + age.test, family =
binomial(link='logit'), data=data.test)
  res = try( dump.model( m), silent = T )
  if(inherits(res, "try-error")) {
    print( "Could not get CI when sex and Age is included into the model")
  }
  if ( verbose ) {
    print(summary(m))
  }
  #Uncomment these if interested in looking at APOE score effect separated from the
main PRS
  #cat(  sprintf( "Test the GRS +  gender + APOE Score model %s vs. %s\n", cohort1,
cohort2) )
  #m = glm(test == cohort1   ~ quant + gender + APOE_Score, family =
binomial(link='logit'), data=data.test)
  #dump.model( m)
  #cat(  sprintf( "Test the GRS +  gender + age + APOE Score model %s vs. %s\n",
cohort1, cohort2) )
  #m = glm(test == cohort1   ~ quant + gender + age.test + APOE_Score, family =
binomial(link='logit'), data=data.test)
  #dump.model( m)
  #print(summary(m))
}


#Auxiliary function to dump the results of the logistic regressions in the test.GRS
function
dump.model = function ( m ) {
  s = summary( m)
  coeff = s$coefficients
  ci = suppressMessages( exp( confint(m) ) )
  cat( sprintf( "Odd\t2.5%%\t97.5%%\tp-value\n"))
  for ( i in 2:dim( coeff)[1] ) {
    var.name=rownames(coeff )[i]
    cat( sprintf( "%f\t%f\t%f\t%e\t%s\n", exp( coeff[i,1]), ci[i,1], ci[i,2],
coeff[i,4], var.name) )
  }
  cat("")
}


#ORDdiff: determine the significance of differences between the OR of alternative
regression models.
#   Parameteres:
#       OR1 odds ratios (not log odds) of the fisrt model
#       OR2 aodds ratios (not log odds) of the second model
#       SE1 Standard Errors from the model 1
#       SE2 Standard Errors from the model 2
```

```
#   Returns:  Evaluation of the statistical significance of the OR (p-value)

ORdiff <- function(OR1,OR2,SE1,SE2){
  odds1 <- OR1
  odds2 <- OR2
  sterr1 <- SE1
  sterr2 <- SE2
  absdiff <- abs(odds1 - odds2)
  delta <- sqrt(sterr1^2 + sterr2^2)
  Z <- absdiff/delta
  diffpval <- 2*(1-pnorm(Z))
  return(diffpval)
}


#Build toy dataset
toy.db = data.frame(
  cohort = c ( rep( "Control", 1000), rep( "sEOAD", 500), rep( "sLOAD", 1500) ),
  gender = sample( c(0,1), size = 3000, replace = T),
  age.test = c( sample( 65:85, size=1000, replace = T), sample( 55:64, size=500,
replace = T), sample( 65:85, size=1500, replace = T) ),
  PRS =  c( runif( 1000, 1, 5 ), runif( 500, 1.2, 5.5 ), runif( 1500, 1, 5.5) ),
  PRS_APOE =  c( runif( 1000, 0, 7 ), runif( 500, 1.2, 10 ), runif( 1500, 0, 9) ),
  stringsAsFactors = F
)

#Examples:
#Compares sEOAD vs Contols, using the PRS (model1)
#Should not return any meaningful results when age (colinearity) is added
test.GRS(toy.db, "cohort", "sEOAD", "Control", "PRS", verbose = F)
#Compares sLOAD vs Contols, using the PRS (model1)
test.GRS(toy.db, "cohort", "sLOAD", "Control", "PRS", verbose = F)

#Compares sEOAD vs Contols, using the PRS with APOE effects (model2)
#Should not return any meaningful results when age (colinearity) is added
test.GRS(toy.db, "cohort", "sEOAD", "Control", "PRS_APOE", verbose = F)
#Compares sLOAD vs Contols, using the PRS with APOE effects (model2)
test.GRS(toy.db, "cohort", "sLOAD", "Control", "PRS_APOE", verbose = F)

#The eADAD are analyzed differently:
toy.db.eadad = data.frame(
  cohort = c ( rep( "Control", 1000), rep( "fADAD", 150)),
  gender = sample( c(0,1), size = 1150, replace = T),
  age.test = c( sample( 65:85, size=1000, replace = T), sample( 42:55, size=150,
replace = T)),
  FID =  c(rep( 0, 1000), sample( 1:45, 150, replace = T)),
  PRS =  c( runif( 1000, 1, 5 ), runif( 150, 1, 5)),
  PRS_APOE =  c( runif( 1000, 1.5, 7 ), runif( 150, 1.9, 7.9)),
  stringsAsFactors = F
)
library( lme4)
library( test.lme )
#Test the association for the PRS (model 1), only correcting for gender
fm2 <- glmer(cohort =="fADAD" ~ PRS + gender + (1 | FID), toy.db.eadad,
family=binomial() )
summary(fm2)
#confint( fm2) #can take a while
#Test the association for the PRS and APOE effects (model 2), only correcting for
gender
fm2 <- glmer(cohort =="fADAD" ~ PRS_APOE  + (1 | FID), toy.db.eadad,
family=binomial() )
summary(fm2)
#confint( fm2) #can take a while
```