Enigma Protocols for Association at each site using unrelated subjects – v1.1

By Sarah Medland, Derrek Hibar, Alejandro Arias Vasquez, and Jason Stein

##########

Before we start, you need to download and install some required programs (which you may already have). The required programs are: R, Mach, ssh client, mach2qtl. Links to the download sites are available below. Please address any questions to: enigma@lists.loni.ucla.edu.

##########

R can be downloaded here: http://cran.stat.ucla.edu/

Mach can be downloaded here: http://www.sph.umich.edu/csg/abecasis/MACH/download/

An ssh client can be downloaded here (though there are many to choose from): http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html.

Download mach2qtl here: http://www.sph.umich.edu/csg/abecasis/MACH/download/
(run tar -zxvf mach2qtl.tar.gz to decompress the files and then type "make all" in the same directory to build. You will then have an executable called mach2qtl that you should add to your path.)

##########

Compile a list of traits and covariates for all analyses to be conducted in the study without patients. This list should include hippocampal volume, ICV, brain volume, age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different MR acquisitions (if applicable). Then write out these files as Merlin compatible files and merge them with the genotype files.

##########

First, make a spreadsheet (using Excel or your favorite spreadsheet program) of familyID, individualID, hippocampal volume, intracranial volume, brain volume , age, sex, and dummy covariate(s) for different MR acquisitions (if applicable) excluding patients. The spreadsheet should look like this:

FID	IID	HippoVol	ICV	BrainVol	Age	Sex	Dummy
Family1	Subj1	3796.87	1780.30	1483.22	23	1	1
Family2	Subj1	4023.24	1801.19	1623.33	25	0	1
Family3	Subj1	4112.24	1792.84	1554.91	22	1	0
...	...	...	...	...	...	...	

Make sure to have exactly the same header labels in the same order so that the commands below do not need to be changed. If you have no dummy covariates (or more than 1 dummy covariate) these commands below should still work (just add the extra dummy covariates to the end). Save this file as a tab delimited text file called /enigma/genetics/HippoCovs_nopatients.txt.

Then read this file into R and add some required columns of age^2, sex*age, and sex*age^2, combine with the MDS components, and write out a .ped and .dat file to add the covariates to the current .ped and .dat files for each of the analyses to be conducted.

R

```
HippoCovs = read.table("HippoCovs_nopatients.txt", header=T);

HippoCovs$AgeSq = HippoCovs$Age*HippoCovs$Age; #add an age^2 term

HippoCovs$AgebySex = HippoCovs$Age*HippoCovs$Sex; #add an age*sex interaction term

HippoCovs$AgeSqbySex = HippoCovs$AgeSq*HippoCovs$Sex; #add an age^2*sex interaction term

mds.cluster = read.table("HM3mds2R.mds", header=T); #Read in the MDS components

mds.cluster$SOL = NULL; #Remove the "SOL" column in the MDS components since this is not a covariate to be included

merged = merge(HippoCovs, mds.cluster, by = c("FID","IID")); #Merge the MDS and other covariates

numcovs = length(colnames(merged))-4; #Number of covariates included for initial analyses

#Hippo | ICV, age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

write.table(cbind(c("T",rep("C",numcovs)),c("HippoVol","ICV",colnames(merged)[6:(numcovs+4)])),"HippoCov_ICV_nopatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_ICV_nopatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3:4],merged[6:(numcovs+4)])); #Create a HippoCov_ICV_nopatients.ped variable

write.table(pedfile,"HippoCov_ICV_nopatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_ICV_nopatients.ped file

#Hippo | BrainVol, age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

write.table(cbind(c("T",rep("C",numcovs)),c("HippoVol","BrainVol",colnames(merged)[6:(numcovs+4)])),"HippoCov_BrainVol_nopatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_BrainVol_nopatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3],merged[5],merged[6:(numcovs+4)])); #Create a HippoCov_BrainVol_nopatients.ped file

write.table(pedfile,"HippoCov_BrainVol_nopatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_BrainVol_nopatients.ped file
```

#Hippo | age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

```
write.table(cbind(c("T",rep("C",numcovs-1)),c("HippoVol",colnames(merged)[6:(numcovs+4)])),"HippoCov_nopatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_nopatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3],merged[6:(numcovs+4)])); #Create a HippoCov_nopatients.ped file

write.table(pedfile,"HippoCov_nopatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_nopatients.ped file
```

#ICV | age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

```
write.table(cbind(c("T",rep("C",numcovs-1)),c("ICV",colnames(merged)[6:(numcovs+4)])),"ICVCov_nopatients.dat",col.names=F,row.names=F,quote=F); #Create a ICVCov_nopatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[4],merged[6:(numcovs+4)])); #Create a ICVCov_nopatients.ped file

write.table(pedfile,"ICVCov_nopatients.ped",quote=F,col.names=F,row.names=F); #Write a ICVCov_nopatients.ped file
```

#BrainVol | age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

```
write.table(cbind(c("T",rep("C",numcovs-1)),c("BrainVol",colnames(merged)[6:(numcovs+4)])),"BrainVolCov_nopatients.dat",col.names=F,row.names=F,quote=F); #Create a BrainVolCov_nopatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[5],merged[6:(numcovs+4)])); #Create a BrainVolCov_nopatients.ped file

write.table(pedfile,"BrainVolCov_nopatients.ped",quote=F,col.names=F,row.names=F); #Write a BrainVolCov_nopatients.ped file
```

##

Check the files you just produced to make sure they have the correct information. There was a lot of text manipulation we just did, so please make sure to look at the files you created to see if they have the correct number of subjects, correct columns, and correct .dat files.

##

Here is an example of the ICVCov_nopatients.dat file:

```
less ICVCov_nopatients.dat
```

```
T ICV
C Age
C Sex
C Dummy
C AgeSq
C AgebySex
C AgeSqbySex
C C1
C C2
C C3
C C4
ICVCov_nopatients.dat (END)
```

Here is an example of the ICVCov_nopatients.ped file (all the data is fake):

less ICVCov_nopatients.ped

```
002_S_0295 002_S_0295 0 0 0 2006.950742 85 0 0 7225 0 0 -0.0528871 0.0275052 -0.00781102 0.0011899
002_S_0413 002_S_0413 0 0 0 1184.641612 77 0 0 5929 0 0 -0.0509072 0.0319183 -0.00827646 0.00297379
002_S_0559 002_S_0559 0 0 0 1347.785842 79 0 0 6241 0 0 -0.0517766 0.0284207 -0.00688331 0.00339018
002_S_0619 002_S_0619 0 0 0 1869.845489 78 1 0 6084 78 6084 -0.0519674 0.0293724 -0.00772354 0.00546803
002_S_0685 002_S_0685 0 0 0 2242.068869 90 0 0 8100 0 0 -0.050936 0.0298941 -0.00571402 0.00132586
002_S_0729 002_S_0729 0 0 0 1294.901934 65 1 1 4225 65 4225 -0.0520809 0.0307134 -0.00664119 0.00279299
002_S_0782 002_S_0782 0 0 0 1911.919205 82 1 1 6724 82 6724 -0.0530871 0.02956 -0.00742942 0.00230579
ICVCov_nopatients.ped
```

Check that the file has the same number of rows as subjects:

wc ICVCov_nopatients.ped

```
jstein@cerebro-dev.data.cluster.loni.ucla.edu[57]% wc ICVCov_nopatients.ped
  743 11888 75357 ICVCov_nopatients.ped
jstein@cerebro-dev.data.cluster.loni.ucla.edu[58]%
```

In this case 743 subjects are contained in the .ped file.

Please check all the rest of the files to make sure they have the correct information.

##

Add the covariates to the .ped file and run the association on each chromosome. First create a .dat file for the population.ped files, then merge with the covariate files generated above, then run mach2qtl.

##

```
for x in `seq 1 22`
        do
        echo "S1 dummy_phenotype" > population${x}.dat
        awk '{print "M", $2}' population${x}.map >> population${x}.dat #Creates a .dat file for each population.ped chromosome file
        for reg in HippoCov_ICV_nopatients  HippoCov_BrainVol_nopatients  HippoCov_nopatients ICVCov_nopatients BrainVolCov_nopatients
                do
```

```
        pedmerge ${reg} population${x} population${x}_${reg}

        mach2qtl --datfile population${x}_${reg}.dat --pedfile population${x}_${reg}.ped --infofile enigmalmp_${x}.mlinfo --dosefile enigmalmp_${x}.mldose --probfile enigmalmp_${x}.mlprob --samplesize > ${reg}_chr${x}-mach2qtl.out #Perform the association on the imputed data

    done

done
```

##

Compile a list of traits and covariates for all analyses to be conducted in the study with all subjects including patients. This list should include hippocampal volume, ICV, brain volume, sex, age, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different MR acquisitions (if applicable). Then write out these files as Merlin compatible files and merge them with the genotype files.

##

This is basically a repeat of the previous section (just including patients now). First, make a spreadsheet (using Excel or your favorite spreadsheet program) of familyID, individualID, hippocampal volume, intracranial volume, brain volume , age, sex, and dummy covariate(s) for different MR acquisitions including patients. The spreadsheet should look like this:

FID	IID	HippoVol	ICV	BrainVol	Age	Sex	Dummy
Family1	Subj1	3796.87	1780.30	1483.22	23	1	1
Family2	Subj1	4023.24	1801.19	1623.33	25	0	1
Family3	Subj1	4112.24	1792.84	1554.91	22	1	0
...	...	...	...	...	...	...	

Make sure to have exactly the same header labels in the same order so that the commands below do not need to be changed. If you have no dummy covariates (or more than 1 dummy covariate) these commands below should still work (just add the extra dummy covariates to the end). Save this file as a tab delimited text file called /enigma/genetics/HippoCovs_wpatients.txt.

Then read this file into R and add some required columns of age^2, sex*age, and sex*age^2, combine with the MDS components, and write out a .ped and .dat file to add the covariates to the current .ped and .dat files.

```
R

HippoCovs = read.table("HippoCovs_wpatients.txt", header=T);

HippoCovs$AgeSq = HippoCovs$Age*HippoCovs$Age; #add an age^2 term

HippoCovs$AgebySex = HippoCovs$Age*HippoCovs$Sex; #add an age*sex interaction term

HippoCovs$AgeSqbySex = HippoCovs$AgeSq*HippoCovs$Sex; #add an age^2*sex interaction term
```

```
mds.cluster = read.table(“HM3mds2R.mds”, header=T); #Read in the MDS components

mds.cluster$SOL = NULL; #Remove the “SOL” column in the MDS components since this is not a covariate to be included

merged = merge(HippoCovs, mds.cluster, by = c("FID","IID")); #Merge the MDS and other covariates

numcovs = length(colnames(merged))-4; #Number of covariates included for initial analyses

#Hippo | ICV, age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

write.table(cbind(c("T",rep("C",numcovs)),c(“HippoVol”,”ICV”,colnames(merged)[6:(numcovs+4)])),"HippoCov_ICV_wpatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_ICV_wpatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3:4],merged[6:(numcovs+4)])); #Create a HippoCov_ICV_wpatients.ped variable

write.table(pedfile,"HippoCov_ICV_wpatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_ICV_wpatients.ped file

#Hippo | BrainVol, age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

write.table(cbind(c("T",rep("C",numcovs)),c(“HippoVol”,”BrainVol”,colnames(merged)[6:(numcovs+4)])),"HippoCov_BrainVol_wpatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_BrainVol_wpatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3],merged[5],merged[6:(numcovs+4)])); #Create a HippoCov_BrainVol_wpatients.ped file

write.table(pedfile,"HippoCov_BrainVol_wpatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_BrainVol_wpatients.ped file

#Hippo | age, sex, age^2, sex*age, sex*age^2, population stratification (4 MDS components), dummy covariate for different scanners/acquisitions

write.table(cbind(c("T",rep("C",numcovs-1)),c(“HippoVol”,colnames(merged)[6:(numcovs+4)])),"HippoCov_wpatients.dat",col.names=F,row.names=F,quote=F); #Create a HippoCov_wpatients.dat file

pedfile=as.data.frame(c(merged[1:2],rep(0,length(merged[1])),rep(0,length(merged[1])),rep(0,length(merged[1])),merged[3],merged[6:(numcovs+4)])); #Create a HippoCov_wpatients.ped file

write.table(pedfile,"HippoCov_wpatients.ped",quote=F,col.names=F,row.names=F); #Write a HippoCov_wpatients.ped file
```

##

Add the covariates to the .ped file and run the association on each chromosome

##

First create a .dat file for the population.ped files, then merge with the covariate files generated above, then run mach2qtl.

```
for x in `seq 1 22`
        do
        for reg in HippoCov_ICV_wpatients  HippoCov_BrainVol_wpatients  HippoCov_wpatients
                do
                pedmerge ${reg} population${x} population${x}_${reg}
                mach2qtl --datfile population${x}_${reg}.dat --pedfile population${x}_${reg}.ped --infofile enigmalmp_${x}.mlinfo --dosefile enigmalmp_${x}.mldose --probfile enigmalmp_${x}.mlprob --samplesize > ${reg}_chr${x}-mach2qtl.out  #Perform the association on the imputed data
        done
done
```

##

Visually check the results files. If the program errors, you will be able to tell by looking at the results files.

##

```
less *mach2qtl.out
```

```
Mach2Qtl V1.0.8 (2009-12-16) -- QTL Association Mapping with Imputed Allele Counts
(c) 2007 Goncalo Abecasis, Yun Li


The following parameters are in effect:

Available Options
        Phenotypic Data : --datfile [population22_BrainVolCov_nopatients.dat],
                          --pedfile [population22_BrainVolCov_nopatients.ped]
   Imputed Allele Counts : --infofile [enigmalmp_22.mlinfo],
                           --dosefile [enigmalmp_22.mldose],
                           --probfile [enigmalmp_22.mlprob]
       Analysis Options : --useCovariates [ON], --quantileNormalization,
                          --dominant, --recessive, --additive [ON]
                 Output : --samplesize [ON]

FITTED MODELS (for covariate adjusted residuals)
================================================
         Trait     Raw Mean Raw Variance         Mean     Variance
      BrainVol   1531.24338  89851.96841     -0.03466  89412.48894

Loading marker information ...
   20085 markers will be analyzed

Processing prob file ...

Executing first [ass analysis of imputed genotypes ...
Executing second pass analysis of imputed genotypes ...

                     INFORMATION FROM .info FILE             QUTL ASSOCIATION ADDITIVE
                     ===================================     ==============================
TRAIT                MARKER          ALLELES  FREQ1  RSQR    EFFECT2  STDERR  CHISQ     PVALUE N
BrainVol             rs4819391       A,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs11089128      A,G      .9299  .0393   -202.00  152.998 1.7432    0.1867    743
BrainVol             rs11912265      A,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs4321465       C,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs8138488       T,C      .6634  .0585    14.415  67.970  0.0450    0.8321    743
BrainVol             rs2027649       A,G      .8335  .0471   -25.828  95.707  0.0728    0.7873    743
BrainVol             rs4911642       T,C      .8665  .1092    45.873  68.880  0.4435    0.5054    743
BrainVol             rs9617162       G,A      .9650  .1232    60.876  119.927 0.2577    0.6117    743
BrainVol             rs6010348       A,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs4568049       C,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs131538        G,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs140378        C,G      .9224  .2611    25.118  56.833  0.1953    0.6585    743
BrainVol             rs7287144       A,G      .7531  .1698   -67.030  43.548  2.3692    0.1238    743
BrainVol             rs5994019       C,-      1.00   .0000   -0.000   1000000000000000 0.0000         1    743
BrainVol             rs5748662       G,A      .7309  .2136   -44.074  37.730  1.3646    0.2427    743
BrainVol             rs4276106       C,A      .9958  .1216    8.429   345.160 0.0006    0.9805    743
BrainVolCov_nopatients_chr22-mach2qtl.out
```

The results files should look something like this. (Disregard the actual numbers because this is done on fake data). If there is an error and you need help, please send an email to enigma@lists.loni.ucla.edu with the message.

##

Compiling the results for upload to the LONI servers

##

#Compress the .mlinfo files and prepare to send to LONI servers.

tar czvf enigmalmp_allchr.mlinfo.tar.gz enigmalmp_*.mlinfo #This will create a file called enigmalmp_allchr.mlinfo.tar.gz that should be uploaded to the server.

#Compress the association results files and prepare to send to LONI servers.

for reg in HippoCov_ICV_nopatients HippoCov_BrainVol_nopatients HippoCov_nopatients ICVCov_nopatients BrainVolCov_nopatients HippoCov_ICV_wpatients HippoCov_BrainVol_wpatients HippoCov_wpatients

do

tar czvf ${reg}_allchr.mach2qtl.tar.gz ${reg}_chr*-mach2qtl.out #This will create a file of all the GWAS results for each analysis that will be sent to LONI servers

done

##

Upload results to LONI servers

##

Please upload everything here: http://enigma.loni.ucla.edu/meta-analysis/. You will need to upload the 9 files that you just created here:

1. enigmalmp_allchr.mlinfo.tar.gz
2. HippoCov_ICV_nopatients _allchr.mach2qtl.tar.gz
3. HippoCov_BrainVol_nopatients_allchr.mach2qtl.tar.gz
4. HippoCov_nopatients_allchr.mach2qtl.tar.gz
5. ICVCov_nopatients_allchr.mach2qtl.tar.gz
6. BrainVolCov_nopatients_allchr.mach2qtl.tar.gz
7. HippoCov_ICV_wpatients_allchr.mach2qtl.tar.gz
8. HippoCov_BrainVol_wpatients_allchr.mach2qtl.tar.gz
9. HippoCov_wpatients_allchr.mach2qtl.tar.gz

In addition there will be other files to upload on the site.