

ENIGMA2 | Protocol For Association Testing Using Unrelated Subjects in Cohorts with Patients

By Miguel E. Rentería, Derrek Hibar, Alejandro Arias Vasquez, Jason Stein and Sarah Medland

#####

Before we start, you need to download and install some required programs (which you may already have). The required programs are: R, ssh client, mach2qtl. Links to the download sites are available below. Please address any questions to: enigma2helpdesk@gmail.com.

#####

- R can be downloaded here: <http://cran.stat.ucla.edu/>
- 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.)

The following protocol can be split into three general categories based on cohort type. If you have a sample of unrelated, healthy subjects please follow the directions under Method A. If you have a sample of unrelated subjects with a mix of healthy controls and diagnosed patients please follow Method B. If you have a sample of related individuals, please follow Method C.

#####

Method B: Protocol for groups with patients

#####

You will need three files to run the association analysis (described below). We recommend you keep these files into your working directory. **Please, make sure to have exactly the same header labels and in the same order as shown below so that the commands used in this protocol need not to be changed:**

- [LandRvolumes.csv](#) Which contains your imaging phenotypes (after quality control) for the entire sample including patients. Make sure that the SubjectID’s in this file are in the proper format (i.e. that they match the format of the individual subject ID’s given in the IID column of the [SubCortCovs_havepatients.csv](#) file).

- Make sure that missing values and individual volume measures that were excluded from the analysis during QC in the LandRvolumes.csv are coded as “NA” without the quotes. Note that we originally suggested marking these values with an “x” in the imaging protocol. The following R scripts handle excluded values better if they are marked with NA. **Please do a “find and replace” in your favorite text editor for “x” and replace it with “NA” (again all without quotes).**
- **FSL FIRST Users:** The ICV values reported in your LandRvolumes.csv file are actually just ratios, in order to convert it to a volume measurement (and make it comparable to the ICV measure given in FreeSurfer) you need to multiply each value by the template volume. **If you used the default template in FSL FIRST (most likely this is true of everyone) then multiply each value in the ICV column by 1827243.** You can do this easily in a spreadsheet program like Excel or on the Linux command line using awk (remember to save it back as a CSV file).

NOTE (1): Missing values in both files: [SubCortCovs_havepatients.csv](#) and [LandRvolumes.csv](#) must be coded as “NA” (without the quotation marks -> “”). The spreadsheet should look like this:

SubjID	Lthal	Rthal	Lcaud	Rcaud	Lput	Rput	Lpal	Rpal	Lhippo	Rhippo	Lamyg	Ramyg	Laccumb	Raccumb	ICV
Subj1
Subj2

- [SubCortCovs_havepatients.csv](#) A spreadsheet generated using Excel or your favourite spreadsheet program, which contains the following columns: Family ID, Individual ID, age, sex, affection status (i.e. patient=1, control=0), and dummy covariates: i.e. a covariate to control for different MR acquisitions, if applicable, including patients. Save this spreadsheet as a comma delimited (.csv) text file called [SubCortCovs_havepatients.csv](#). The spreadsheet should look like this:

FID	IID	Age	Sex	AffectionStatus	Dummy1..	Dummy2...
Fam1	Subj1
Fam2	Subj1

NOTE (2): Sex must be specified as follows: (Males=1, Females=2), and “FID” and “IID” should be named exactly the same in all files; Likewise, for the protocol to run smoothly without requiring additional changes, please name the affection status column as **AffectionStatus**, and fill it in with values of 1 and 0 (patient=1 and control=0).

- The third file is [HM3mds2R.mds.csv](#) (a spreadsheet containing the following columns: individual ID (IID), 4 MDS components (C1, C2, C3 and C4), and PLINK's assigned solution code (SOL)).

FID	IID	SOL	C1	C2	C3	C4
Fam1	Subj1
Fam2	Subj2

NOTE (3): If you have no dummy covariates (or more than 1 dummy covariate) the commands below should still work (just add the extra dummy covariates to the end and modify the code where indicated in red).

These three files: [LandRvolumes.csv](#), [SubCortCovs_havepatients.csv](#) and [HM3mds2R.mds.csv](#) will be read into R to generate PED and DAT files that will be used for association with mach2qtl.

#####

The following R script assumes your files are all kept in the same folder, which is also the working directory of R.

```
R
getwd() #Check that you are in the correct directory
SubCort <- read.table("LandRvolumes.csv", colClasses=c("character",
rep("numeric",15)), sep="," , header=T); #Read in the phenotypes file
Covs <- read.table("SubCortCovs_havepatients.csv",
colClasses=c(rep("character",2), rep("numeric",3)), sep="," , header=T); #Read
in the covariates file
SubCort$IID = SubCort$SubjID #This just renames a column for easier merging
SubCort$SubjID = NULL
SubCortCovs <- merge(SubCort, Covs, by="IID"); #Merge into a single dataframe

SubCortCovs$AgeSq <- SubCortCovs$Age*SubCortCovs$Age; #add an age^2 term
SubCortCovs$Mthal <- rowMeans(SubCortCovs[,c("Lthal","Rthal")]); #calculate
mean Thalamus
SubCortCovs$Mcaud <- rowMeans(SubCortCovs[,c("Lcaud","Rcaud")]); #calculate
mean Caudate
SubCortCovs$Mput <- rowMeans(SubCortCovs[,c("Lput","Rput")]); #calculate mean
Putamen
SubCortCovs$Mpal <- rowMeans(SubCortCovs[,c("Lpal","Rpal")]); #calculate mean
Pallidum
SubCortCovs$Mhippo <- rowMeans(SubCortCovs[,c("Lhippo","Rhippo")]);
#calculate mean Hippocampus
```

```

SubCortCovs$Mamyg <- rowMeans(SubCortCovs[,c("Lamyg","Ramyg")]); #calculate
mean Amygdala
SubCortCovs$Maccumb <- rowMeans(SubCortCovs[,c("Laccumb","Raccumb")]);
#calculate mean Accumbens

mds.cluster <- read.table("HM3mds2R.mds.csv",
colClasses=c(rep("character",2), rep("numeric",5)), sep=",", 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_temp <- merge(SubCortCovs, mds.cluster, by=c("FID","IID")); #Merge the
MDS and other covariates

merged_ordered <- merged_temp[,c("FID", "IID", "Sex", "Lthal", "Lcaud",
"Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "Rcaud", "Rput",
"Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal",
"Mhippo", "Mamyg", "Maccumb", "ICV", "Age", "AgeSq", "C1", "C2", "C3", "C4",
"AffectionStatus")] #Create an ordered data frame with left and hemisphere
volumes, as well as mean volumes and covariates. If you have additional dummy
covariates to accommodate different scanners you will need to modify this
command in order to work properly. For an example, see below:
#merged_ordered <- merged_temp[,c("FID", "IID", "Sex", "Lthal","Lcaud",
"Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal","Rcaud", "Rput",
"Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal","Mcaud", "Mput", "Mpal",
"Mhippo", "Mamyg", "Maccumb", "ICV","Age", "AgeSq", "C1", "C2", "C3", "C4",
"AffectionStatus", "Dummy1", "Dummy2"...)]

numcovs <- length(colnames(merged_ordered))-24; #Number of Covariates(ICV,
age, age2, population stratification (4 MDS components), dummy covariate for
different scanners/acquisitions).

merged_ordered[,1:(24+numcovs)][is.na(merged_ordered[,1:(24+numcovs)])] <-
"x" #recode "NAs" into "x", to comply with required association format

## * * * * *
* * ##
##Create three PED files containing 21 traits (7 x Left, 7 x Right and 7 x
Mean Hemispheric Volumes)for the different combined-sexes analyses:
Males+Females patients/controls combined, Males+Females patients only,
Males+Females controls only.
## * * * * *
* * ##

merged_MF_ordered_combined <- merged_ordered; #Create a Males+Females
variable

merged_MF_ordered_combined$Sex -> merged_MF_ordered_combined$SexPED; #Rename
Sex column as SexPED Variable

```

```

merged_MF_ordered_combined$SexPED -> merged_MF_ordered_combined$Sex; #Create
a SexCOV Variable

merged_MF_ordered_combined$Sex[merged_MF_ordered_combined$Sex==1] <- 0;
#recode males from "1" into "0", in the sex covariate.
merged_MF_ordered_combined$Sex[merged_MF_ordered_combined$Sex==2] <- 1;
#recode females from "2" into "1", in the sex covariate.

merged_MF_ordered_combined <- merged_MF_ordered_combined[,c("FID", "IID",
"SexPED", "Lthal", "Lcaud", "Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb",
"Rthal", "Rcaud", "Rput", "Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal",
"Mcaud", "Mput", "Mpal", "Mhippo", "Mamyg", "Maccumb", "ICV", "Age", "Sex",
"AgeSq", "C1", "C2", "C3", "C4", "AffectionStatus")] #Create an ordered data
frame with left and hemisphere volumes, as well as mean volumes and
covariates. If you have additional dummy covariates to accommodate different
scanners you will need to modify this command in order to work properly. For
an example, see below:
#merged_ordered <- merged_temp[,c("FID", "IID", "Sex", "Lthal", "Lcaud",
"Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "Rcaud", "Rput",
"Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal",
"Mhippo", "Mamyg", "Maccumb", "ICV", "Age", "AgeSq", "C1", "C2", "C3", "C4",
"AffectionStatus", "Dummy1", "Dummy2"...)]

pedfile=as.data.frame(c(merged_MF_ordered_combined[1:2],rep(0,length(merged_M
F_ordered_combined[1])),rep(0,length(merged_MF_ordered_combined[1])),merged_M
F_ordered_combined[3:24],merged_MF_ordered_combined[25:(numcovs+25)]));
#Create a pedfile variable containing all individuals in the sample.
write.table(pedfile,"MalesFemales_subcortCov_combined.ped",quote=F,col.names=
F,row.names=F); #Write out MalesFemales_subcortCov_patientsonly.ped file

merged_MF_ordered_havepatients <- subset(merged_MF_ordered_combined,
AffectionStatus==1); #Create a PATIENTS ONLY Subset
merged_MF_ordered_havepatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_MF_ordered_havepatients[1:2],rep(0,length(merg
ed_MF_ordered_havepatients[1])),rep(0,length(merged_MF_ordered_havepatients[1
])),merged_MF_ordered_havepatients[3:24],merged_MF_ordered_havepatients[25:(n
umcovs+24)])); #Create a pedfile variable containing patients only.
write.table(pedfile,"MalesFemales_subcortCov_patientsonly.ped",quote=F,col.na
mes=F,row.names=F); #Write out MalesFemales_subcortCov_patientsonly.ped file

merged_MF_ordered_nopatients <- subset(merged_MF_ordered_combined,
AffectionStatus==0); #Create a CONTROLS ONLY Subset
merged_MF_ordered_nopatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_MF_ordered_nopatients[1:2],rep(0,length(merged
_MF_ordered_nopatients[1])),rep(0,length(merged_MF_ordered_nopatients[1])),me
rged_MF_ordered_nopatients[3:24],merged_MF_ordered_nopatients[25:(numcovs+24)
])); #Create a pedfile variable containing Controls only.

```

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

```
##Create six DAT files: With and without ICV as a Covariate including ALL Volumes, Left, Right and Mean (and two more versions of the same files that include affection status)##
```

```
## * * * * *
* * ##
```

```
##Without ICV and no affection status
```

```
write.table(cbind(c(rep("T",21),"S",rep("C", (numcovs-1))),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_MF_ordered_havepatients)[25:(numcovs+24)])), "subcort_nICV_SexCov_noaffect.dat", col.names=F, row.names=F, quote=F);
```

```
##Without ICV with affection status
```

```
write.table(cbind(c(rep("T",21),"S",rep("C", (numcovs))),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_MF_ordered_combined)[25:(numcovs+25)])), "subcort_nICV_SexCov_withaffect.dat", col.names=F, row.names=F, quote=F);
```

```
##Without ICV and without affection status in the ped file
```

```
write.table(cbind(c(rep("T",21),"S",rep("C", (numcovs-2))),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_MF_ordered_havepatients)[25:(numcovs+23)])), "subcort_nICV_SexCov_Sep.dat", col.names=F, row.names=F, quote=F);
```

```
##With ICV and no affection status
```

```
write.table(cbind(c(rep("T",21),rep("C", (numcovs))),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_MF_ordered_havepatients)[25:(numcovs+24)])), "subcort_wICV_SexCov_noaffect.dat", col.names=F, row.names=F, quote=F);
```

```
##With ICV with affection status
```

```
write.table(cbind(c(rep("T",21),rep("C", numcovs+1)),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_MF_ordered_combined)[25:(numcovs+25)])), "subcort_wICV_SexCov_withaffect.dat", col.names=F, row.names=F, quote=F);
```

```
##With ICV withoutout affection status in the ped file
```

```
write.table(cbind(c(rep("T",21),rep("C", numcovs)),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames
```

```
(merged_MF_ordered_combined)[25:(numcovs+24)]), "subcort_wICV_SexCov_Sep.dat"
, col.names=F, row.names=F, quote=F);
```

```
## * * * * *
* * ##
```

```
##Create six PED files containing 21 traits (7 x Left, 7 x Right and 7 x Mean Hemispheric Volumes) for the different sex-specific analyses: Males-Only patients/controls combined, Males-Only patients only, Males-Only controls only; Females-Only patients/controls combined, Females-Only patients only, Females-Only controls only.
```

```
## * * * * *
* * ##
```

```
merged_M_Ordered_combined <- subset(merged_ordered, Sex==1); #Create a MALES ONLY subset
pedfile=as.data.frame(c(merged_M_Ordered_combined[1:2], rep(0, length(merged_M_Ordered_combined[1])), rep(0, length(merged_M_Ordered_combined[1])), merged_M_Ordered_combined[3:24], merged_M_Ordered_combined[25:(numcovs+24)])); #Create a pedfile variable containing Males-only.
write.table(pedfile, "Males_subcortCov_combined.ped", quote=F, col.names=F, row.names=F); #Write out Males_subcortCov_combined.ped file
```

```
merged_M_Ordered_havepatients <- subset(merged_M_Ordered_combined, AffectionStatus==1); #Create a MALE PATIENTS ONLY Subset
merged_M_Ordered_havepatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_M_Ordered_havepatients[1:2], rep(0, length(merged_M_Ordered_havepatients[1])), rep(0, length(merged_M_Ordered_havepatients[1])), merged_M_Ordered_havepatients[3:24], merged_M_Ordered_havepatients[25:(numcovs+23)])); #Create a pedfile variable containing Male patients only.
merged_M_Ordered_havepatients$AffectionStatus <- NULL
write.table(pedfile, "Males_subcortCov_patientonly.ped", quote=F, col.names=F, row.names=F); #Write out Males_subcortCov_patientonly.ped file
```

```
merged_M_Ordered_nopatients <- subset(merged_M_Ordered_combined, AffectionStatus==0); #Create a MALE CONTROLS ONLY Subset
merged_M_Ordered_nopatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_M_Ordered_nopatients[1:2], rep(0, length(merged_M_Ordered_nopatients[1])), rep(0, length(merged_M_Ordered_nopatients[1])), merged_M_Ordered_nopatients[3:24], merged_M_Ordered_nopatients[25:(numcovs+23)])); #Create a pedfile variable containing Male Controls only.
merged_M_Ordered_nopatients$AffectionStatus <- NULL
write.table(pedfile, "Males_subcortCov_nopatients.ped", quote=F, col.names=F, row.names=F); #Write out Males_subcortCov_nopatients.ped file
```

```
merged_F_Ordered_combined <- subset(merged_ordered, Sex==2); #Create a FEMALES ONLY subset
```

```

pedfile=as.data.frame(c(merged_F_Ordered_combined[1:2],rep(0,length(merged_F_Ordered_combined[1])),rep(0,length(merged_F_Ordered_combined[1])),merged_F_Ordered_combined[3:24],merged_F_Ordered_combined[25:(numcovs+24)])); #Create a pedfile variable containing Females-only.
write.table(pedfile,"Females_subcortCov_combined.ped",quote=F,col.names=F,row.names=F); #Write out Females_subcortCov_combined.ped file

merged_F_Ordered_havepatients <- subset(merged_F_Ordered_combined, AffectionStatus==1); #Create a FEMALE PATIENTS ONLY subset
merged_F_Ordered_havepatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_F_Ordered_havepatients[1:2],rep(0,length(merged_F_Ordered_havepatients[1])),rep(0,length(merged_F_Ordered_havepatients[1])),merged_F_Ordered_havepatients[3:24],merged_F_Ordered_havepatients[25:(numcovs+23)])); #Create a pedfile variable containing Female Patients only.
write.table(pedfile,"Females_subcortCov_patientonly.ped",quote=F,col.names=F,row.names=F); #Write out Females_subcortCov_havepatients.ped file

merged_F_Ordered_nopatients <- subset(merged_F_Ordered_combined, AffectionStatus==0); #Create a FEMALE CONTROLS ONLY subset
merged_F_Ordered_nopatients$AffectionStatus <- NULL
pedfile=as.data.frame(c(merged_F_Ordered_nopatients[1:2],rep(0,length(merged_F_Ordered_nopatients[1])),rep(0,length(merged_F_Ordered_nopatients[1])),merged_F_Ordered_nopatients[3:24],merged_F_Ordered_nopatients[25:(numcovs+23)])); #Create a pedfile variable containing Female Controls only.
write.table(pedfile,"Females_subcortCov_nopatients.ped",quote=F,col.names=F,row.names=F); #Write out Females_subcortCov_nopatients.ped file

## * * * * *
##
##Create four DAT files: With and without ICV as a Covariate including ALL Volumes, Left, Right and Mean (and two more versions of the same files that include affection status)##
## * * * * *
##

##Without ICV and no affection status
write.table(cbind(c(rep("T",21),"S",rep("C", (numcovs-2)), "S"),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_ordered)[25:(numcovs+24)]),"subcort_nICV_noSexCov_noaffect.dat",col.names=F,row.names=F,quote=F);

##Without ICV with affection status
write.table(cbind(c(rep("T",21),"S",rep("C", (numcovs-1))),c("Lthal","Lcaud","Lput","Lpal","Lhippo","Lamyg","Laccumb","Rthal","Rcaud","Rput","Rpal","Rhippo","Ramyg","Raccumb","Mthal","Mcaud","Mput","Mpal","Mhippo","Mamyg","Maccumb",colnames(merged_ordered)[25:(numcovs+24)]),"subcort_nICV_noSexCov_withaffect.dat",col.names=F,row.names=F,quote=F);

```



```
##With ICV and no affection status
write.table(cbind(c(rep("T",21),rep("C", (numcovs-
1)), "S")),c("Lthal", "Lcaud", "Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "
Rcaud", "Rput", "Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal"
, "Mhippo", "Mamyg", "Maccumb", colnames(merged_ordered) [25:(numcovs+24)]), "subc
ort_wICV_noSexCov_noaffect.dat", col.names=F, row.names=F, quote=F);
```

```
##With ICV with affection status
write.table(cbind(c(rep("T",21),rep("C", numcovs)),c("Lthal", "Lcaud", "Lput", "L
pal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "Rcaud", "Rput", "Rpal", "Rhippo", "Ramyg
", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal", "Mhippo", "Mamyg", "Maccumb", colnames
(merged_ordered) [25:(numcovs+24)]), "subcort_wICV_noSexCov_withaffect.dat", co
l.names=F, row.names=F, quote=F);
```

```
## * * * * *
* * ##
```

```
##Create four DAT files: With and without ICV as a Covariate including ALL
Volumes, Left, Right and Mean (without the affection status column)##
```

```
## * * * * *
* * ##
```

```
##Without ICV without affection status
write.table(cbind(c(rep("T",21), "S", rep("C", (numcovs-
2))),c("Lthal", "Lcaud", "Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "Rcau
d", "Rput", "Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal", "Mh
ippo", "Mamyg", "Maccumb", colnames(merged_ordered) [25:(numcovs+23)]), "subcort_
nICV_noSexCov_Sep.dat", col.names=F, row.names=F, quote=F);
```

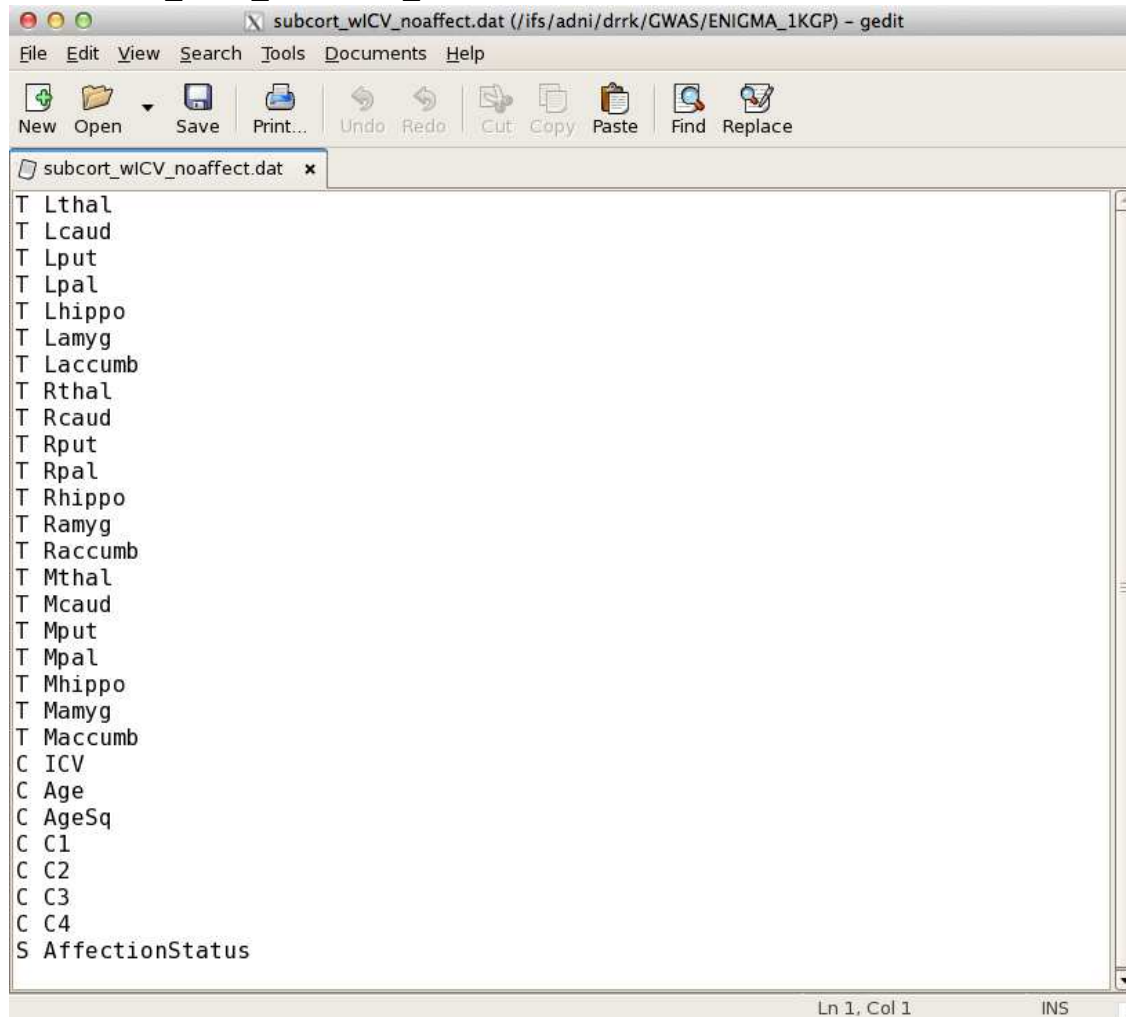
```
##With ICV with affection status
write.table(cbind(c(rep("T",21),rep("C", numcovs-
1)),c("Lthal", "Lcaud", "Lput", "Lpal", "Lhippo", "Lamyg", "Laccumb", "Rthal", "Rcaud
", "Rput", "Rpal", "Rhippo", "Ramyg", "Raccumb", "Mthal", "Mcaud", "Mput", "Mpal", "Mhi
ppo", "Mamyg", "Maccumb", colnames(merged_ordered) [25:(numcovs+23)]), "subcort_w
ICV_noSexCov_Sep.dat", col.names=F, row.names=F, quote=F);
```

```
#####
```

Now, 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.

```
#####
```

Below is an example of the contents of subcort_wICV_noaffect.dat file:
less subcort_wICV_noSexCov_noaffect.dat



```
subcort_wICV_noaffect.dat x
T Lthal
T Lcaud
T Lput
T Lpal
T Lhippo
T Lamyg
T Laccumb
T Rthal
T Rcaud
T Rput
T Rpal
T Rhippo
T Ramyg
T Raccumb
T Mthal
T Mcaud
T Mput
T Mpal
T Mhippo
T Mamyg
T Maccumb
C ICV
C Age
C AgeSq
C C1
C C2
C C3
C C4
S AffectionStatus
Ln 1, Col 1 INS
```

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

```

Males_subcortCov_patientsonly.ped (/ifs/adni/drrk/GWAS/ENIGMA_1KGP) - gedit
File Edit View Search Tools Documents Help
New Open Save Print... Undo Redo Cut Copy Paste Find Replace
Males_subcortCov_patientsonly.ped x
002_S_0619 002_S_0619 0 0 8007.3 3140.32 4060.45 1355.59 3146.63 1205.39 177.968 7217.17 3104.97 4594
002_S_0816 002_S_0816 0 0 8119.15 3201.03 5254.45 1998.63 3738.29 1656.93 696.296 7616.27 2920.57 517
003_S_1257 003_S_1257 0 0 7818.52 2894.71 4520.99 1961.66 2581.82 1538.22 355.983 8095.81 2864.73 489
005_S_0221 005_S_0221 0 0 7716.31 3270.67 4531.09 1878.87 2791.07 1387.53 535.143 7639.4 3457.6 4410.
005_S_0929 005_S_0929 0 0 7417.03 3324.8 4073.54 1633.81 2615.64 1362.24 317.747 7320.27 3152.18 4690
006_S_0547 006_S_0547 0 0 11308.5 3574.91 5842.3 1944.88 6296 1708.74 498.52 9791.1 3682.05 5249.76 2
007_S_0316 007_S_0316 0 0 7370.79 3349.49 4576.18 1864.49 2739.33 1293.66 551.693 6674.53 3149.65 488
009_S_1334 009_S_1334 0 0 8578.61 3597.24 5411.97 1991.91 3762.61 1786.81 545.492 8156.61 3455.5 5256
010_S_0786 010_S_0786 0 0 7109.02 2927.94 4721.82 1470.94 3186.5 1603.98 391.607 6649.82 3073.85 4643
011_S_0003 011_S_0003 0 0 8586.43 3396.21 5421.52 2277.3 4575.29 1816.58 432.518 8029.79 3962.24 5007
011_S_0053 011_S_0053 0 0 7419.75 3994.1 5080.51 2099.89 3199.39 1596.89 476.824 7586.17 3702.39 4620
012_S_0689 012_S_0689 0 0 8001.09 3547.23 4422.6 2708.71 2895.18 1367.76 684.436 7440.59 3185.48 4513
012_S_0712 012_S_0712 0 0 6278.88 2363.29 4019.51 1623.02 2212.23 560.289 256.041 6303.52 2463.99 416
013_S_0592 013_S_0592 0 0 7550.17 3160.71 5079.58 1939.44 2848.38 1404.55 521.797 7163.03 3072.81 492
013_S_0699 013_S_0699 0 0 7661.36 3700.46 3979.63 1744.37 3874.71 957.435 526.496 8669.38 2965.99 486
013_S_1161 013_S_1161 0 0 7782.67 3451.48 4602.59 1901.68 3840.8 1587.23 527.829 7722.78 3254.95 4538
013_S_1205 013_S_1205 0 0 7384.29 3228.06 5350.15 1689.82 2881.86 1306.19 578.243 6690.03 2900.57 476
014_S_0328 014_S_0328 0 0 7879.22 2854.61 4553.29 1742.41 4081.97 1726.41 486.253 7203.16 3031.62 507
014_S_0356 014_S_0356 0 0 7422.59 2833.59 3968.31 1577.3 2963.7 1563.44 354.066 6560.89 2510.46 3512.
018_S_0286 018_S_0286 0 0 8970.89 3849.23 4609.06 2064.22 3564.7 2063.15 467.833 8073.59 3643.56 5226
018_S_0633 018_S_0633 0 0 6604.11 2731.96 4053.64 1447.65 2551.54 1253.35 406.752 6406.61 2762.92 441
018_S_0682 018_S_0682 0 0 7199.69 2825.64 4323 1741.96 2846.91 1297.43 499.831 6857.25 2997.92 4451.6
020_S_0213 020_S_0213 0 0 8140.83 3991.19 5836.71 2140.07 4093.72 1711.31 816.506 7300.09 3666.82 529
021_S_0343 021_S_0343 0 0 7199.26 2816.59 5124.26 1973 3184.2 1414.14 295.365 7045.2 2897.34 5180.57
021_S_0642 021_S_0642 0 0 7745.91 2978.62 4588.63 1687.64 2235.3 2696.82 340.293 7296.09 3210.45 4713
021_S_0753 021_S_0753 0 0 7199.52 2848.21 4620.33 1734.76 2962.44 1424.1 403.532 6722.33 2830.06 4356
023_S_0083 023_S_0083 0 0 7396.62 2887.26 4913.02 1779.08 2528.46 1381.03 332.642 7103.22 2846.15 499
023_S_0916 023_S_0916 0 0 7914.1 3796.45 4550.89 1508.87 3443.51 1365.08 343.604 7344.54 3643.32 5273
024_S_1171 024_S_1171 0 0 7970.89 3846.51 4975.97 1950.3 3358.67 2027.72 594.953 7603.94 3816.82 5013
027_S_0850 027_S_0850 0 0 5837.16 2166.43 3086.79 890.36 1410 351.43 292.501 5882.16 1792.5 1637.15 5
027_S_1081 027_S_1081 0 0 7177 3055.58 3726.66 1799.85 2835.8 1226.93 365.945 6800.38 3059.85 3298.84
027_S_1254 027_S_1254 0 0 7768.55 2942.68 4215.57 1401.99 2912.81 1318.77 449.191 7554.09 2957.62 458
029_S_0836 029_S_0836 0 0 8174.62 3801.98 4597.29 1913.23 3236.63 1897.26 603.674 7674.22 3400.59 494
029_S_0999 029_S_0999 0 0 6888.29 2906.35 4215.81 1688.59 3256.64 1209.52 221.505 6109.42 2439.65 409
031_S_0321 031_S_0321 0 0 7413.27 2705.96 3676.12 1378.14 3785.34 1257.14 523.63 7159.49 2861.23 4042
032_S_0400 032_S_0400 0 0 8856.64 3320.04 5117.46 2020.89 3885.67 1989.88 677.908 8295.28 3609.81 507
032_S_1037 032_S_1037 0 0 7912.73 3347.78 4422.13 2342.7 3361.63 1665.89 348.527 7751.79 3012.04 4572
Ln 96, Col 1 INS

```

Check that the file has the same number of rows as subjects compared to your subject info:
 wc Males_subcortCov_patientsonly.ped

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

#####

Association with Mach2QTL

You should now have six PED files and four DAT files. This is all you will need to run the association on each chunk of chromosome you produced in the imputation section of these protocols. Use the shell script below to that end.

#####

Replace **highlighted** portions below to customise for your data. This code will generate a script called `mach2qtl_association.sh` that you need to tailor to your server/queuing system. The aim is to run association commands in as many chromosome chunks in parallel as possible. The files being generated will be zipped as they are produced to help preserve space.

```
#!/bin/bash

echo "#Mach2qtl association" > mach2qtl_association.sh
machdir=/home/1KGPref/Mach #give the directory to the imputed output from
Mach/minimac
peddatdir=/home/1KGPref #give the dir to the ped and dat files just created
samplename=QTIM #give abbreviated name of your sample, no spaces in the name
(i.e. ADNI)
mach2qtlout=/home/1KGPref/mach2qtl_out #make a folder for the output from
mach2qtl

### For Patients and Controls seperately
#loop over dat file options
for aff in Sep; do
#Males-only, Females-only
for group in Males Females; do
#with and without ICV as covariate
for cov in w n; do
#different subsets
for subset in patientonly nopatients; do
#loop over chromosomes
for ((i=1; i<=23; i++)); do
# loop over 'chunks'
for ((j=1; j<=15; j++)); do

if test -f ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz
then

#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".out" >> mach2qtl_association.sh

#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".out" >> gzip_results.sh
fi
```

```

if [ -f ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.dose.gz ] && [
${group} == "Females" ]
then
#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".female.out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".female.out" >> gzip_results.sh
fi
done
if [ -f ${machdir}/ready4mach."$i".male.imputed.dose.gz ] && [ ${group} ==
"Males" ]
then
#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/ready4mach."$i".male.imputed.info.gz \
--dosefile ${machdir}/ready4mach."$i".male.imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".male.out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".male.out" >> gzip_results.sh
fi
done
done
done
done

#### For Patients+Controls combined
#loop over dat file options
for aff in noaffect withaffect; do
#Males-only, Females-only
for group in Males Females; do

```

```

#with and without ICV as covariate
for cov in w n; do
#different subsets
for subset in combined ; do
#loop over chromosomes
for ((i=1; i<=23; i++)); do
# loop over 'chunks'
for ((j=1; j<=15; j++)); do

if test -f ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz
then

#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".out" >> mach2qtl_association.sh

#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".out" >> gzip_results.sh
fi

if [ -f ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.dose.gz ] && [
${group} == "Females" ]
then
#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".female.imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".female.out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".female.out" >> gzip_results.sh
fi
done
if [ -f ${machdir}/ready4mach."$i".male.imputed.dose.gz ] && [ ${group} ==
"Males" ]
then

```

```

#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_noSexCov_"$aff".dat \
--pedfile ${peddatdir}/"$group"_subcortCov_"$subset".ped \
--infofile ${machdir}/ready4mach."$i".male.imputed.info.gz \
--dosefile ${machdir}/ready4mach."$i".male.imputed.dose.gz \
--samplesize > \
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".male.out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_"$group"_"$cov"_ICV_subcort_"$subset"_"$aff"_chr
"$i"_"$j".male.out" >> gzip_results.sh
fi
done

done
done
done
done

##### Males+Females Patients and Controls seperate
#loop over dat file options
for aff in Sep ; do
#with and without ICV as covariate
for cov in w n; do
#different subsets
for subset in patientonly nopatients; do
#loop over chromosomes
for ((i=1; i<=22; i++)); do
# loop over 'chunks'
for ((j=1; j<=15; j++)); do
if test -f ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz
then
#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_SexCov_"$aff".dat \
--pedfile ${peddatdir}/MalesFemales_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz \
--samplesize >
${mach2qtlout}/${samplename}_MalesFemales_"$cov"_ICV_subcort_"$subset"_"$aff"
_chr"$i"_"$j".out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_MalesFemales_"$cov"_ICV_subcort_"$subset"_"$aff"
_chr"$i"_"$j".out" >> gzip_results.sh
fi
done

```

```
done
done
done
done
```

```
##### Males+Females Patients and Controls combined
#loop over dat file options
for aff in noaffect withaffect; do
#with and without ICV as covariate
for cov in w n; do
#different subsets
for subset in combined ; do
#loop over chromosomes
for ((i=1; i<=22; i++)); do
# loop over 'chunks'
for ((j=1; j<=15; j++)); do
if test -f ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz
then
#Specify the commands, parameters and data files required for association
echo "mach2qtl --datfile ${peddatdir}/subcort_"$cov"ICV_SexCov_"$aff".dat \
--pedfile ${peddatdir}/MalesFemales_subcortCov_"$subset".ped \
--infofile ${machdir}/chunk"$j"-ready4mach."$i".imputed.info.gz \
--dosefile ${machdir}/chunk"$j"-ready4mach."$i".imputed.dose.gz \
--samplesize >
${mach2qtlout}/${samplename}_MalesFemales_"$cov"_ICV_subcort_"$subset"_"$aff"
_chr"$i"_"$j".out" >> mach2qtl_association.sh
#Generate a shell script to zip association results files to be uploaded to
the ENIGMA server
echo "gzip
${mach2qtlout}/${samplename}_MalesFemales_"$cov"_ICV_subcort_"$subset"_"$aff"
_chr"$i"_"$j".out" >> gzip_results.sh
fi
done
done
done
done
done
done
```

The code above will generate two shell script files: “mach2qtl_association.sh” and “gzip_results.sh”. Change the permission to make them executable and run “mach2qtl_association.sh”:

```
chmod +x mach2qtl_association.sh
chmod +x gzip_results.sh
```


You can run the association script directly (`./mach2qt1_association.sh`), but in the interest of time try to split the commands up to run in parallel in the format appropriate for your computing cluster.

When association has finished running for all chunks, run the [gzip_results.sh](#) script to compress the results files and save space (this will make it a lot easier and faster to upload them to the ENIGMA server):

```
./gzip_results.sh
```

Each group has a secure space on the ENIGMA upload server to upload the `.info.gz` and gzipped association result files. Please contact enigma2helpdesk@gmail.com to obtain upload information for your group's data.