

# Exclusions

*Rebekah Wrigley*

*26/01/2015*

I followed these steps using the data Cushla gave me which included both Europeans and Polynesians. Exclusions for the urate analysis are described elsewhere.

```
# Read in data
For.Rebekah2 <- read.delim("/Volumes/staff_groups/merrimanlab/Merriman Documents/Rebekah/ABCG2_files/Im

# Separate by ethnicity.
Cau_1<- For.Rebekah2[For.Rebekah2$ETHCLASS.1.cau..2.ep..3.wp..9.epwp==3, ] # Include only Europeans

Pol_1<- For.Rebekah2[is.na(For.Rebekah2$FULLEPGTAFF)!=TRUE | is.na(For.Rebekah2$FULLWPGTAFF)!=TRUE, ]

# Remove people with kidney disease, maybe gout, related people, or on urate lowering therapy.
Cau_2<- Cau_1[Cau_1$SUAGROUP.1.use!= 3 & Cau_1$SUAGROUP.1.use!= 5 & Cau_1$SUAGROUP.1.use!= 7 & Cau_1$SU

Pol_2<- Pol_1[Pol_1$SUAGROUP.1.use!= 3 & Pol_1$SUAGROUP.1.use!= 5 & Pol_1$SUAGROUP.1.use!= 7 & Pol_1$SU
```

I removed controls with AGE1ATK (age at first gout attack) data from Europeans but not Polynesians (only realised these were present late in the project). To the best of my knowledge controls with AGE1ATK info (i.e. not NA) were present in the file I got from Cushla but not in files I got from Mandy (that were downloaded from SNPmax).

```
AGE1ATK<- Cau_2[Cau_2$MYGOUTSUM==1 & is.na(Cau_2$AGE1ATK)!=TRUE,] # Check if there are any controls wi
Cau_3<- Cau_2[(Cau_2$MYGOUTSUM==1 & is.na(Cau_2$AGE1ATK)!=TRUE)!=TRUE,] # Remove controls with an AGE1

# Create .txt files
write.table(Cau_3, file="Cau.txt",row.names=FALSE,quote=FALSE,sep="\t",na="NA")

write.table(Pol_2, file="Pol.txt",row.names=FALSE,quote=FALSE,sep="\t",na="NA")
```

For NPH I removed people with kidney disease using the KIDNEY column but did not exclude those with no kidney disease info.

```
NPH_2<- NPH_1[NPH_1$KIDNEY==1 | is.na(NPH_1$KIDNEY)==TRUE, ]
```