Conducting eQTL analysis

Files required
All files should be created using the same order of samples

Gene expression
- Samples in rows
- Genes in columns
- log2(cpm1+1)

Principal components of gene expression
- Samples in rows
- Value for PCs in columns

Genotyping
- Samples in rows
- SNPs in columns
- Coded as 0, 1 and 2 (number of copies of the minor allele)

Principal components of genotyping
- Samples in rows
- Value for PCs in columns

Environmental factors
- Samples in rows
- Value for each environmental factor in columns

SNP gene pairs to test
- Gene name in first column
- SNP name in second column

Example

```{r}
load("files_for_pfizer_eqtl.rda")

```

This contains the following files

`cpm1.379.t` gene expression

`pcas.cpm1` principal components of expression

`g.pcs.379` genotying

`geno.379` principal components of genotyping

`int.terms` environmental factors to test for interaction (e.g. IFN and drug)

`pairs.0.05.tss` SNP gene pairs to test for eQTL

`pairs.int.tss` top eQTL SNP gene pairs to test for interaction

`subject` subject ID

eQTL script
This tests the first 10 SNP gene pairs, change `irange` to test more

```{r}
library(lme4)

irange<-1:10
results<-as.data.frame(matrix(0, ncol=4, nrow=length(irange)))

results<-do.call(rbind, lapply(irange, function(i){
  model.null <- lmer(cpm1.379.t[,pairs.0.05.tss[i,1]] ~ pcas.cpm1[,1:25]
    + g.pcs.379[,1:5]
    + (1|subject),
    REML=FALSE)

    model.test <- lmer(cpm1.379.t[,pairs.0.05.tss[i,1]] ~ pcas.cpm1[,1:25]
      + g.pcs.379[,1:5]
      + geno.379[,pairs.0.05.tss[i,2]]
      + (1|subject),
      REML=FALSE)

#Check there aren't any individuals with a missing genotype otherwise update the model
    if (all(complete.cases(geno.379[,pairs.0.05.tss[i,2]]))){
        results[i, ]<-c(summary(model.test)$coefficients[32,],
            anova(model.null, model.test)$'Pr(>Chisq)'[2])


    } else {
```

```
      model.null.subset<-update(model.null,
subset=complete.cases(geno.379[,pairs.0.05.tss[i,2]]))
      model.test.subset<-update(model.test,
subset=complete.cases(geno.379[,pairs.0.05.tss[i,2]]))
      results[i, ]<-c(summary(model.test)$coefficients[32,],
          anova(model.null.subset, model.test.subset)$'Pr(>Chisq)'[2])
   }
}))

colnames(results)<-c("eQTL_beta", "eQTL_SE", "eQTL_t", "eQTL_pval")
-----

results <- data.frame(
   Gene = pairs.0.05.tss[irange, 1],
   SNP = pairs.0.05.tss[irange, 2],
   results
)
saveRDS(results, "eQTL.rds")
```

Interaction script for drug
This tests the first 10 eQTL SNP gene pairs for interactions, change `irange` to test more

```{r}
irange<-1:10

results.int<-as.data.frame(matrix(0, ncol=10, nrow=length(irange)))

results.int<-do.call(rbind, lapply(irange, function(i){
  model.null<-lmer(cpm1.379.t[,pairs.int.tss[i,1]] ~ pcas.cpm1[,1:25] +
   g.pcs.379[,1:5]+
   int.terms$Drug +
   geno.379[,pairs.int.tss[i,2]] +
   (1|subject),
   REML=FALSE)

  model.test<-lmer(cpm1.379.t[,pairs.int.tss[i,1]] ~ pcas.cpm1[,1:25] +
   g.pcs.379[,1:5] +
   int.terms$Drug +
   geno.379[,pairs.int.tss[i,2]] +
   int.terms$Drug*geno.379[,pairs.int.tss[i,2]] +
   (1|subject),
   REML=FALSE)
```

```
#Include a filter so eQTL are only tested for an interaction if there is more than
#one minor homozygous individual in each of the environmental factor groups
  if (length(unique(subject[which(geno.379[,pairs.int.tss[i,2]]==2
    & int.terms$Drug==0)]))>1 &
    length(unique(subject[which(geno.379[,pairs.int.tss[i,2]]==2
    & int.terms$Drug==1)]))>1){

    results.int[i, ]<-c(summary(model.test)$coefficients[32,],
      summary(model.test)$coefficients[33,],
      summary(model.test)$coefficients[34,],
      anova(model.null, model.test)$'Pr(>Chisq)'[2])


  } else {
    results.int[i,]<-rep(NA, 10)

  }

}))

colnames(results.int)<-c("Drug_estimate", "Drug_SE", "Drug_t", "Geno_estimate",
  "Geno_SE", "Geno_t", "Int_estimate", "Int_SE", "Int_t", "pval")

results.int <- data.frame(
    Gene = pairs.int.tss[irange, 1],
    SNP = pairs.int.tss[irange, 2],
    results.int
)

saveRDS(results.int, "drug.interaction.rds")

```