

SUGEN 8.6 GxE

Misa Graff, July 2017

Running a GxE analysis

- For a GxE the environmental variable has to be numeric. If we want to use sex or group as the environmental variable, then we should use the binary dummy variables in the data set "sex_f" or "group_uw".

```
$ head phenotypes_sugen.txt | awk '{printf("%5s %5s %5s %5s %5s\n", $2, $5, $9, $11, $12)}'
```

```
user1@ubuntu-sugen:~/data$ head phenotypes_sugen.txt | awk '{printf("%5s %5s %5s %5s %5s\n", $2, $5, $9, $11, $12)}'
```

scanID	sex	group	sex_f	group_uw
p207	F	unc	1	0
p36	M	uw	0	1
p202	F	unc	1	0
p370	F	unc	1	0
p290	M	unc	0	0
p402	F	unc	1	0
p421	M	unc	0	0
p321	F	unc	1	0
p379	M	unc	0	0

Script for a GxE analysis

```
$ ../cd/results
```

```
$ ~/SUGEN-master/SUGEN --pheno
```

```
../data/phenotypes_sugen.txt --id-col scanID --family-  
col fid --vcf ../data/SISG_genotype.vcf.gz --formula  
"trait=sex_f+age+EV1+EV2+group" --unweighted --model  
linear --ge sex_f --out-prefix quan_trait_gxe
```

- Note flag for `--ge sex_f`

- Output:

- *quan_trait_gxe.wald.out*

- Association results

- *quan_trait_gxe.log*

****Copy *quan_trait_gxe.wald.out* to "C:\easystrata\"**

*wald.out description

Results columns in a GxE association analyses

```
$ head quan_trait_gxe.wald.out
```

Standard columns:

- CHROM, POS, VCF_ID, REF, ALT, ALT_AF, ALT_AC, N_INFORMATIVE, N_REF, N_HET, N_ALT, N_DOSE

Additional columns:

- PVALUE_G main genetic variable p-value.
- PVALUE_INTER interaction p-value (SNP by environmental variable).
- PVALUE_BOTH p-value of both the main genetic variable and the interaction terms.
- BETA_G Effect estimate of the genetic variable.
- BETA_sex_f Effect estimate of the environmental variable (sex).
- BETA_G:sex_f Effect estimate of the interaction term between the genetic variable and the environment variable.
- COV_G_G Variance estimate of Beta G.
- COV_sex_f_sex_f Variance estimate of Beta sex_f.
- COV_G:sex_f_G:sex_f Variance estimate of Beta G:sex_f.
- COV_G_sex_f Covariance estimate between Beta G and Beta sex_f.
- COV_G_G:sex_f Covariance estimate between Beta G and Beta G:sex_f.
- COV_sex_f_G:sex_f Covariance estimate between Beta sex_f and Beta G:sex_f.

Clean and plot results in EasyStrata

- Open EasyStrata script (we copied this to you laptop C drive yesterday):
“C:\easystрата\Easystрата_quant_trait_gex.ecf”
 - *Make any necessary changes to paths or file name*
 - *We will go through the script to QC and plot results for the quantitative trait (other scripts are similar)*
-

##PLEASE add path to where results, plots, etc will be output

DEFINE --pathOut C:\easystрата

##define column names and column classes; don't need to change these

--acolIn CHROM;POS;VCF_ID;REF;ALT;ALT_AF;N_INFORMATIVE;PVALUE_G;PVALUE_INTER;PVALUE_BOTH;BETA_G;BETA_sex_f;BETA_G:sex_f;COV_G_G;COV_sex_f_sex_f;COV_G:sex_f_G:sex_f

--acolInClasses character;numeric;character;character;character;numeric;numeric;numeric;numeric;numeric;numeric;numeric;numeric;numeric;numeric

--strMissing NA

--strSeparator TAB

##PLEASE add Define path to the results:

#optional: tag name to be used in any merging where 2 columns have the same name; ShortName will be used in plots, reports, etc

EASYIN --fileIn C:\easystрата\quan_trait_gxe.wald.out

--fileInTag quan_trait_gxe

--fileInShortName quan_trait_gxe

Clean and plot results in EasyStrata

```
#####  
## QQplot: all pvalues  
#####
```

QQPLOT

```
--acolQQPlot PVALUE_G;PVALUE_INTER;PVALUE_BOTH  
--astrColour black;red;blue  
--blnLegend 1  
--astrLegendText c("PVALUE_G","PVALUE_INTER","PVALUE_BOTH"), col=c("black","red","blue")
```

```
#calculate lambda for all pvalues; suppress GC correction
```

```
GC      --colPval PVALUE_G  
        --blnSuppressCorrection 1
```

```
#calculate lambda for interaction; suppress GC correction
```

```
GC      --colPval PVALUE_INTER  
        --blnSuppressCorrection 1
```

```
#calculate lambda for 2df pvalue (main + interaction effects); suppress GC correction
```

```
GC      --colPval PVALUE_BOTH  
        --blnSuppressCorrection 1
```

Clean and plot results in EasyStrata

```
#####  
## Miami plot of interaction pvalue and 2df pvalue (main + interaction effects)  
#####  
MHPLOT --colMHPlot PVALUE  
        --colInChr CHROM  
        --colInPos POS  
        --numWidth 1200  
        --numHeight 800  
        --anumAddPvalLine 7e-4;7e-6  
        --anumAddPvalLineLty 6;6  
        --astrAddPvalLineCol blue;orangered  
        --numDefaultSymbol 20  
        --arcdSymbolCrit PVALUE<7e-4;PVALUE<7e-6  
        --anumSymbol 8;8  
        --arcdColourCrit PVALUE<7e-4;PVALUE<7e-6  
        --astrColour blue;orangered  
        --arcdCexCrit PVALUE<7e-4;PVALUE<7e-6  
        --anumCex 1.2;1.3  
        --astrDefaultColourChr gray51;gray66  
        --blnYAxisBreak 1  
        ##Path to list of known loci for your trait if available; includes 3 columns "Chr" "Pos" "Colour",  
        #--fileAnnot /path2/known_loci.txt  
        #--numAnnotPosLim 500000
```

Clean and plot results in EasyStrata

```
#####  
## Clump results by 'suggestive' significance (PVALUE_INTER<7e-4) using physical distance criterion distance<500kB to either side  
#####  
INDEP      --rcdCriterion PVALUE_INTER<7e-4  
            --colIndep PVALUE_INTER  
            --colInChr CHROM  
            --colInPos POS  
            --numPosLim 500000  
## Function INDEP creates two files labeled _indep.txt and _indepX.txt  
## _indep.txt contains ALL SNPs that passed P<7e-4  
##      Column aLociTag can be used to distinguish between independent loci (each number represents an indep locus)  
##      Column aTopHit can be used to identify the TopHit for the respective locus (all other SNPs are set to NA)  
## _indepX.txt only contains the Top Hits
```


Clean and plot results in EasyStrata

```
#####
```

```
## this will output cleaned results after dropping based on QC
```

```
WRITE --strMode txt  
      --strPrefix CLEANED.  
      --strSep TAB  
      --strMissing .
```

```
STOP EASYSTRATA
```

```
#####
```