XP-GWAS: a method for identifying trait-associated variants by sequencing pools of individuals

Schnable Lab

July 8, 2015

SUMMARY:

Although approaches for conducting genome-wide association studies (GWAS) are well developed, conventional GWAS requires the high-density genotyping of large numbers of individuals from a diversity panel. Here we report a method for conducting GWAS that does not require the genotyping of large numbers of individuals. Instead XP-GWAS (extreme phenotype GWAS) relies on genotyping pools of individuals from a diversity panel having extreme phenotypes. This analysis measures allele frequencies in the extreme pools, enabling the discovery of associations between genetic variants and traits of interest.

CITATION:

Yang et al., The Plant Journal, 2015, submitted

Statistical Procedure:

After SNP and small indel discovery, allele counts were estimated at each polymorphic site from each pool as the input data (see below example). Note the column names should match exactly as the below example data (or under directory of **data/**).

```
input <- read.table("../data/input_sample.txt", header=TRUE)
head(input)
```

```
##     snpid chr  pos high_ref high_alt low_ref low_alt random_ref random_alt
## 1 10_3005  10 3005        0        1       1       7          0          0
## 2 10_3219  10 3219        5        5       9       5          6          4
## 3 10_3452  10 3452       26       28      32      92         32         52
## 4 10_3523  10 3523        4        8       8      27          9         11
## 5 10_3658  10 3658        9       25      10      18          6         17
## 6 10_4099  10 4099       12        5      14       3         13         11
```

The following scripts should be ran in `R` console. An add-on package `GenABEL` should be installed first.

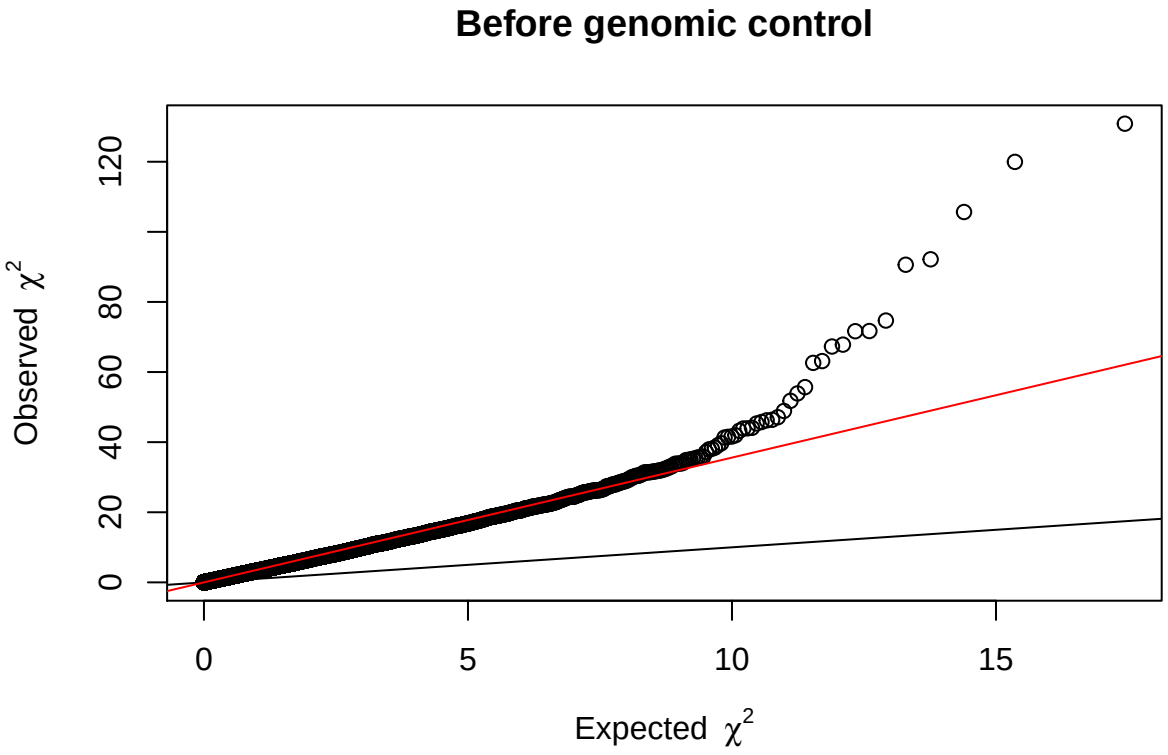
```
#install.packages("GenABEL")
```

```
setwd("../")
source("lib/xpgwas.R")
library("GenABEL")
```

```
## Loading required package: MASS
## Loading required package: GenABEL.data
```

```
### get FDR corrected p-values for sites passed your filtering criterion (DEFAULT filter = 50)
### plotlambda: indicate whether to plot the genomic control results, default = TRUE.
qval <- xpgwas(input, filter=200,  plotlambda=TRUE)
```

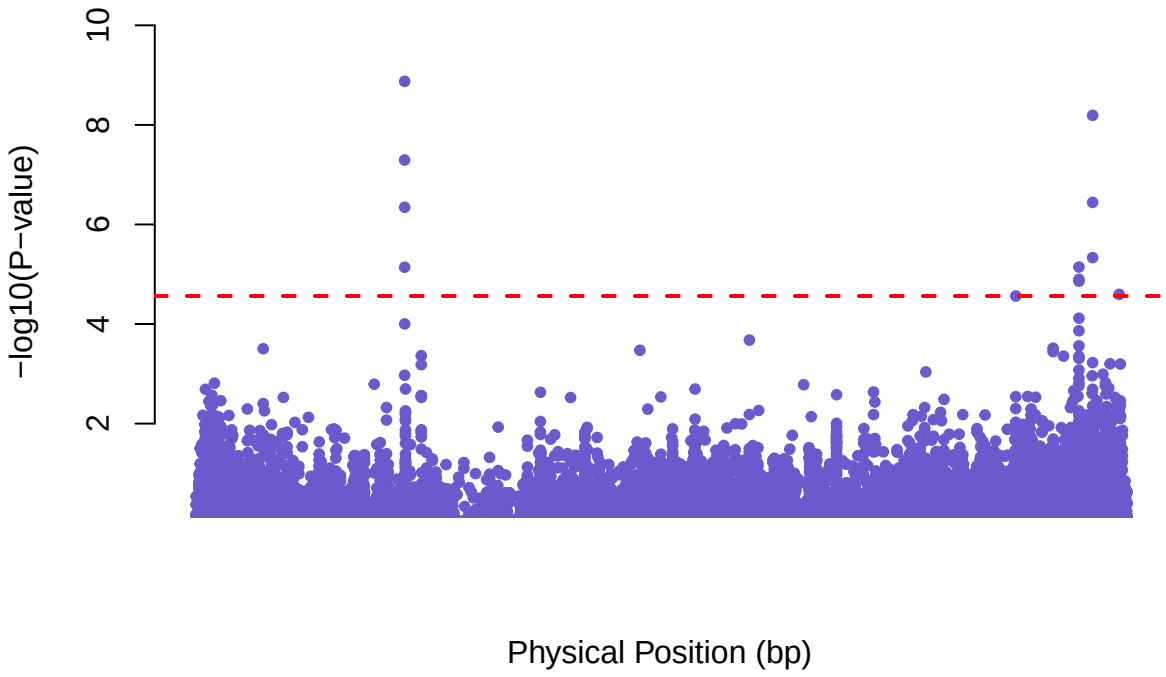
```
## ###>>> input [ 347628 ] variants, remaining [ 16879 ] after depth filtering [ > 200 ]
## ###>>> DONE!
## ###>>> [ 12 ] significant sites using FDR < 0.05
```




```
### save results
#save(file="../cache/qval.RData", list=c("input", "qval"))
```

Plot your results

```
source("../lib/xpplot.R")
xpplot(qval)
```


−log10(P−value)
2
4
6
8
10
Physical Position (bp)