Genome-Wide Association Study (GWAS)

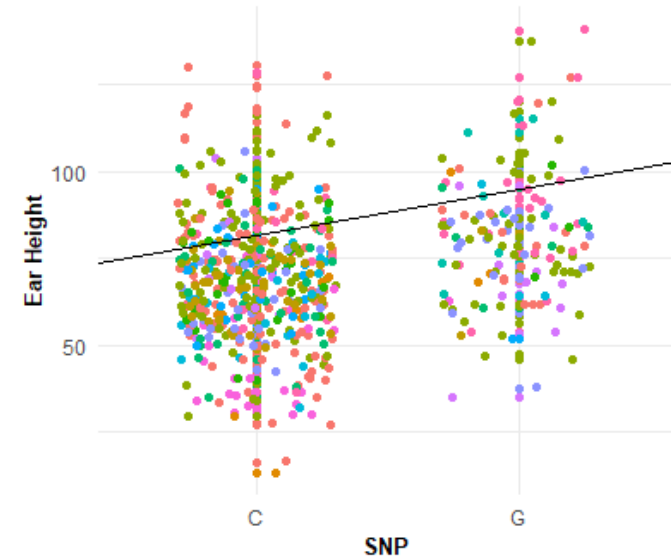An observational study of a genome-wide set of genetic variants (SNPs) in different individuals to see if any variant is associated with a trait.

Pros and Cons of GWAS

Bi-parental QTL mapping

Genome-wide association study

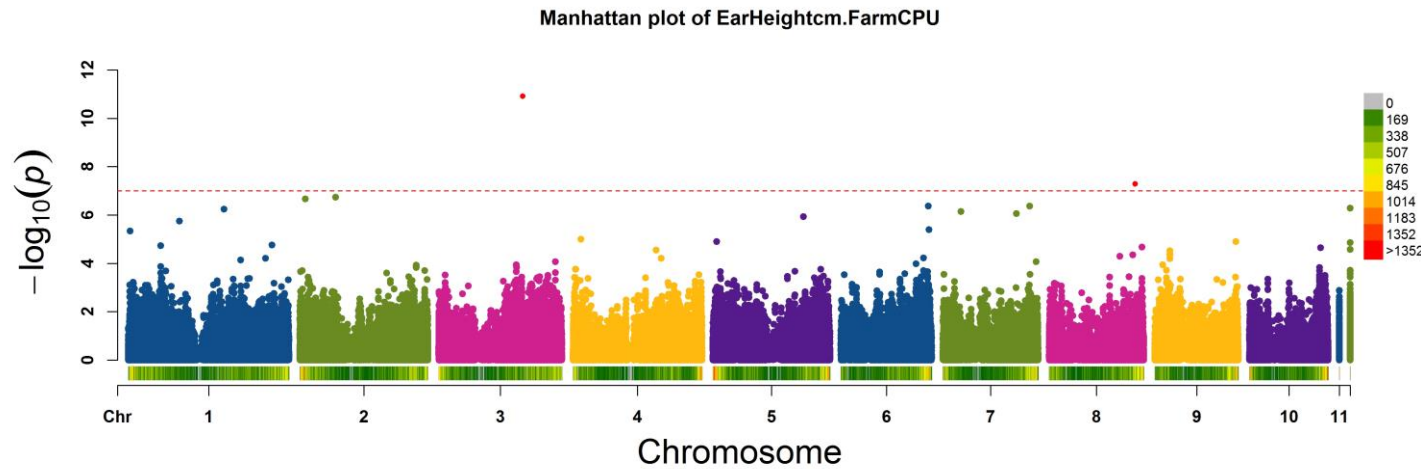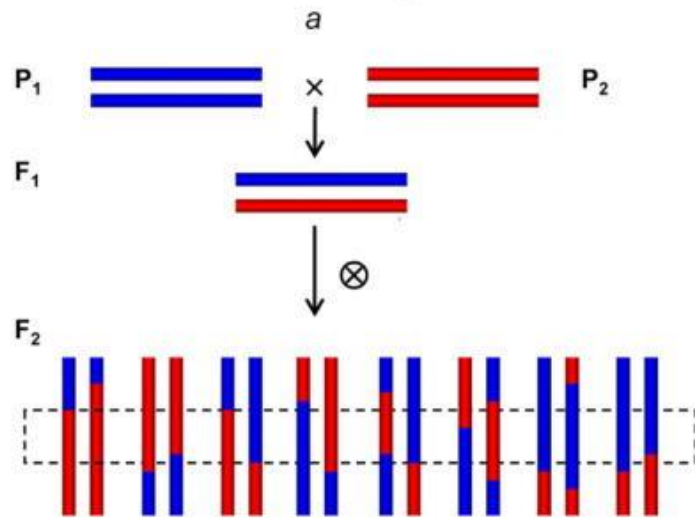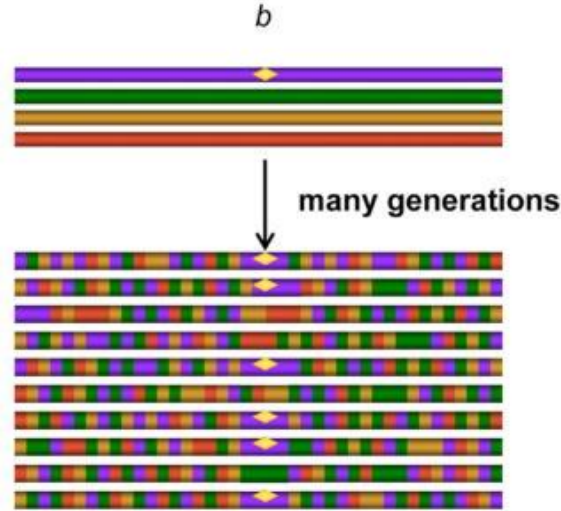Zhu et al., 2008. The Plant Genome

❖Pros
- No need to create population
- Take advantage of historical recombination
- Higher resolution
- Wider application

❖Cons
- Low power for rare alleles
- Cannot account for epistasis

GAPIT introduction

❖GAPIT: developed by Zhiwu Zhang lab at Washington State University

❖Operated in R environment, rely on several R libraries

❖Uses a minimal amount of code

❖Several statistical methods included
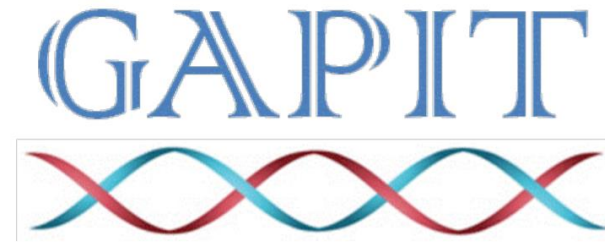
GAPIT

Genomic Association and Prediction Integrated Tool

(Version 3)

GAPIT input file - Genotype

rs	alleles	chrom	pos	strand	assembly	center	protLSID	assayLSID	panel	QCcode	33-16	38-11	4226	4722	A188	A214N	A239	A272	A441-5
PZB00859.1	A/C	1	157104	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC	CC	AA	CC	AA	AA	CC
PZA01271.1	C/G	1	1947984	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	GG	CC	GG	CC	CC	CC	CC	CC
PZA03613.2	G/T	1	2914066	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	GG	GG	GG	GG	TT	TT	TT	GG
PZA03613.1	A/T	1	2914171	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT	TT	TT	AA	TT	TT	TT
PZA03614.2	A/G	1	2915078	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	GG	GG	GG	GG	GG	AA	AA	GG
PZA03614.1	A/T	1	2915242	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT	TT	TT	AA	AA	AA	TT
PZA00258.3	C/G	1	2973508	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	CC	CC	CG	CC	CC	CC	GG	CC
PZA02962.13	A/T	1	3205252	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT	TT	TT	TT	TT	TT	TT
PZA02962.14	C/G	1	3205262	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC	CC	CC	CC	CC	CC	CC

required   SNP attributes

Genotype	AA	CC	GG	TT	AG	CT	CG	AT	GT	AC
Code	A	C	G	T	R	Y	S	W	K	M

GAPIT input file - Phenotype

Taxa	EarHT	dpoll	EarDia
811	59.50	NaN	NaN
4226	65.50	59.5	32.21933
4722	81.13	71.5	32.42100
33-16	64.75	64.5	NaN
38-11	92.25	68.5	37.89700
A188	27.50	62.0	31.41900
A214N	65.00	69.0	32.00600
A239	47.88	61.0	36.06400
A272	35.63	70.0	NaN
A441-5	53.50	67.5	35.00800
A554	38.50	66.0	33.41775
A556	28.00	65.0	31.92900
A6	109.50	80.5	31.51750
A619	36.00	61.0	40.63000
A632	60.00	61.0	35.95300

GAPIT – Statistical methods

- **GLM (General Linear Model)**

test marker-trait association individually

- **MLM (Mixed Linear Model)**

include individual as random effects, population structure and kinship as fixed effects.

- **MLMM (Multiple Locus Mixed linear Model)**
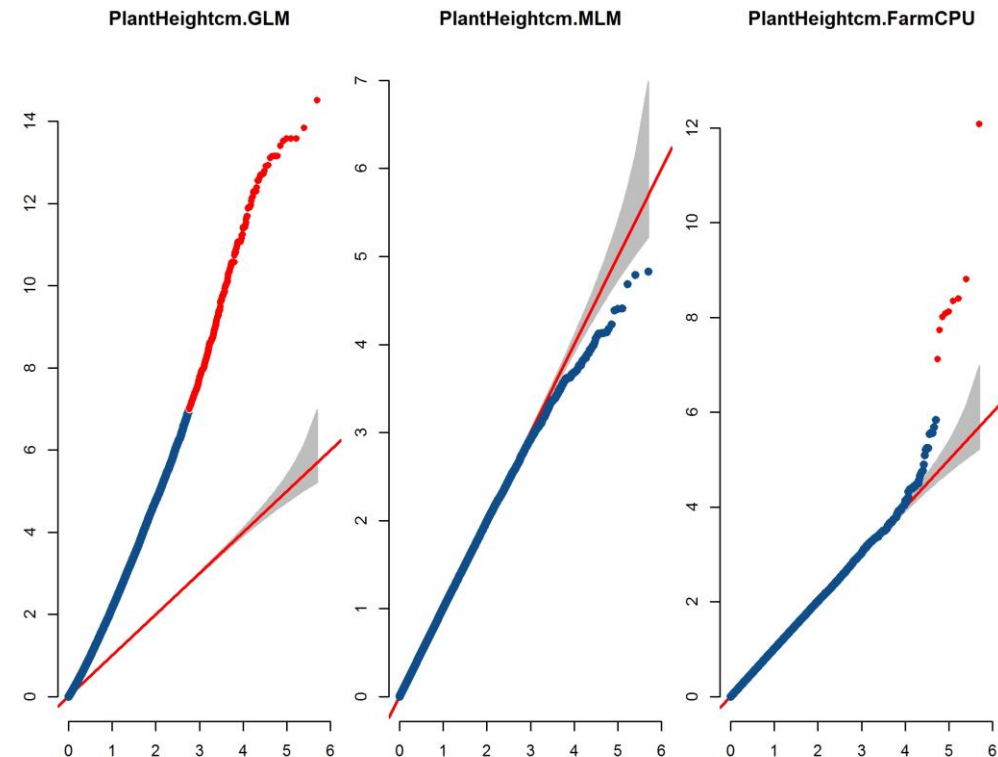
forward-backward stepwise linear mixed-model regression

significant SNP is added into model as co-factor to detect other S

- **FarmCPU (Fixed and random model circulating probability uni**

designed for GWAS on large data

address confounding problem between covariates and test marke

can control both false positive and false negative.

Get started

1. Create a new folder named "R-GAPIT" on your desktop

2. Save "mdp_genotype_test.hmp.txt" and "mdp_traits.txt" to "R-GAPIT" folder

3. Open R studio – go to session – set working directory –choose directory – navigate to "R-GAPIT"

4. Run R code (see next page)

Install and load required libraries

```r
if (!requireNamespace("BiocManager", quietly = TRUE))

  install.packages("BiocManager")

BiocManager::install("multtest")

install.packages("gplots")

install.packages("LDheatmap")

install.packages("genetics")

install.packages("ape")

install.packages("EMMREML")

install.packages("scatterplot3d")

library(multtest)

library(gplots)

library(LDheatmap)

library(genetics)

library(ape)

library(EMMREML)

library(compiler) #this library is already installed in R

library("scatterplot3d")

source("http://zzlab.net/GAPIT/gapit_functions.txt")

source("http://zzlab.net/GAPIT/emma.txt")
```

Load in data and run analysis

```r
geno_demo <- read.delim("mdp_genotype_test.hmp.txt",header = FALSE)
dim(geno_demo)
View(geno_demo[1:10,1:20])
pheno_demo <- read.delim("mdp_traits.txt",header = TRUE)
dim(pheno_demo)
View(pheno_demo[1:15,])

#run GAPIT
myGAPIT <- GAPIT(
  Y=pheno_demo,
  G=geno_demo,
  PCA.total = 3,
  model = c(""GLM,"MLM",”MLMM”,"FarmCPU")
)
```

GAPIT – Outputs

**GAPIT.MLM.EarDia.GWAS.Results.csv**

SNP	Chromosome	Position	P.value	maf	nobs	Rsquare.of.Model.without.SNP	Rsquare.of.Model.with.SNP	FDR_Adjusted_P-values	effect
PZA00453.7	4	166281276	0.000185	0.104418	249	0.249775477	0.293916878	0.537324968	1.636501
PZB01915.1	1	9029842	0.000347	0.212851	249	0.249775477	0.290087127	0.537324968	1.184324
PZB00237.2	1	240501065	0.000832	0.080321	249	0.249775477	0.284839401	0.545540086	1.681858
PZA03379.2	4	167650309	0.000935	0.028112	249	0.249775477	0.284147776	0.545540086	3.377588
PZA03186.1	5	5476347	0.001156	0.086345	249	0.249775477	0.282890765	0.545540086	-1.47772
PZA00396.10	2	4679610	0.001672	0.375502	249	0.249775477	0.280709757	0.545540086	1.031925
PZA00280.14	2	62870552	0.001675	0.024096	249	0.249775477	0.280700103	0.545540086	3.04718
PZB00063.2	1	268373269	0.001922	0.13253	249	0.249775477	0.279892492	0.545540086	1.29007
PZA00186.2	3	165800369	0.002123	0.353414	249	0.249775477	0.27931134	0.545540086	0.923693
PZA02808.12	2	44606596	0.002267	0.064257	249	0.249775477	0.278925792	0.545540086	-1.52639
PZA00210.6	3	29693448	0.002309	0.022088	249	0.249775477	0.278818933	0.545540086	2.79886
PZD00098.2	1	23267835	0.002311	0.11245	249	0.249775477	0.27881422	0.545540086	-1.31213
PZB02516.2	3	193566873	0.002464	0.184739	249	0.249775477	0.278440517	0.545540086	1.106617

GAPIT.FarmCPU.EarDia.GWAS.Results.csv

| SNP	Chromosome	Position	P.value	maf	nobs	Rsquare.of.Model.without.SNP	Rsquare.of.Model.with.SNP	FDR_Adjusted_P-values	effect
PZB00237.2	1	240501065	8.89E-07	0.078853	279	NA	NA	0.002749854	1.828343
PZB01230.5	2	39040780	1.05E-05	0.1541219	279	NA	NA	0.01403432	1.025081
PZA00714.6	1	74580051	1.77E-05	0.1666667	279	NA	NA	0.01403432	1.064631
PZA00942.2	6	102566000	1.81E-05	0.2867384	279	NA	NA	0.01403432	0.952351
PZA00793.2	8	64421988	2.65E-05	0.437276	279	NA	NA	0.014538246	0.729054
PZA00442.3	2	63547199	2.82E-05	0.3602151	279	NA	NA	0.014538246	-0.75982
PZB02035.3	2	27403594	3.32E-05	0.2598566	279	NA	NA	0.014684549	-0.79974
PZB00093.2	4	122796082	4.56E-05	0.2401434	279	NA	NA	0.017612683	0.863066
PZB02516.2	3	193566873	0.00015972	0.1756272	279	NA	NA	0.05453162	0.83783
PZA00339.3	1	260297200	0.00017631	0.3476703	279	NA	NA	0.05453162	0.709649
PZA03559.1	2	15810363	0.00025675	0.4265233	279	NA	NA	0.072193836	0.615314
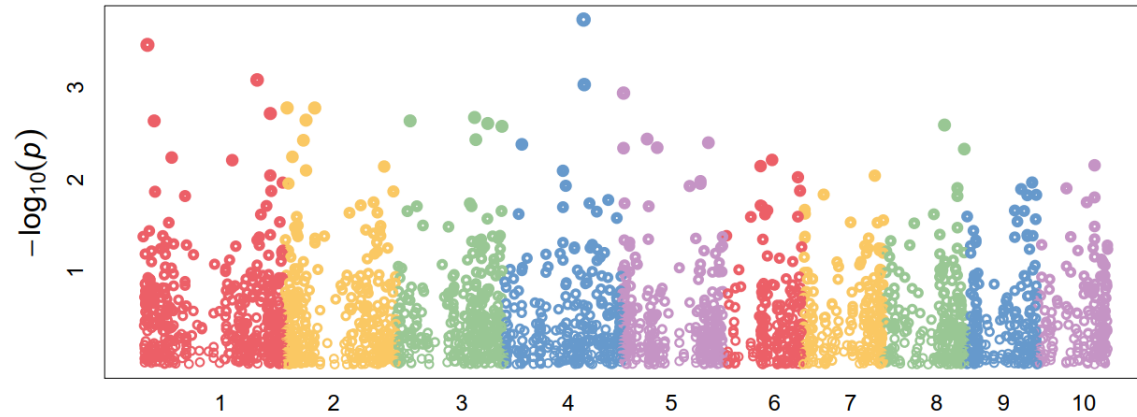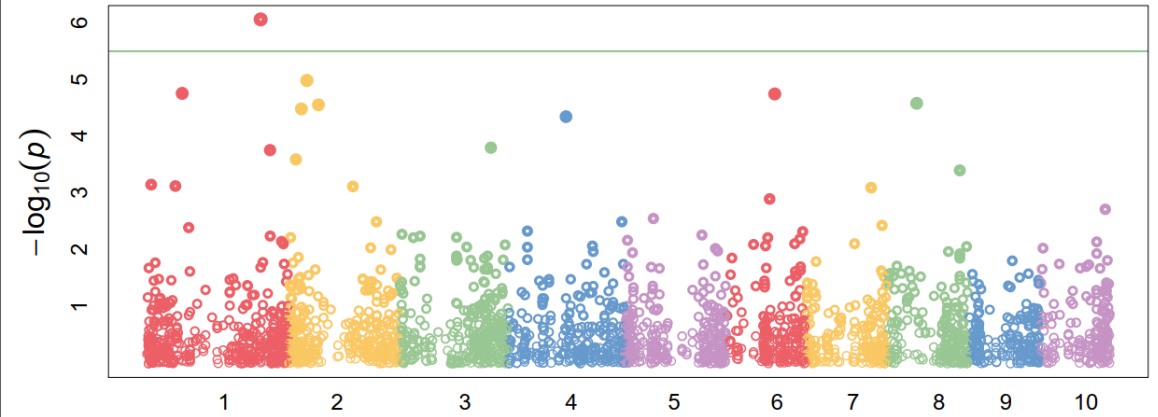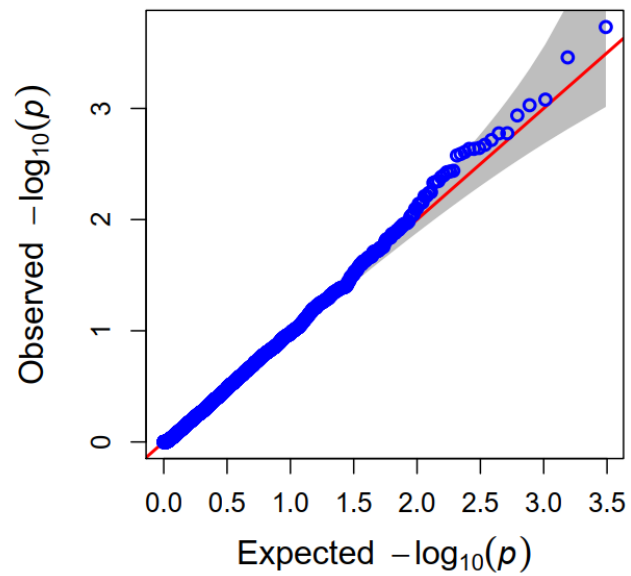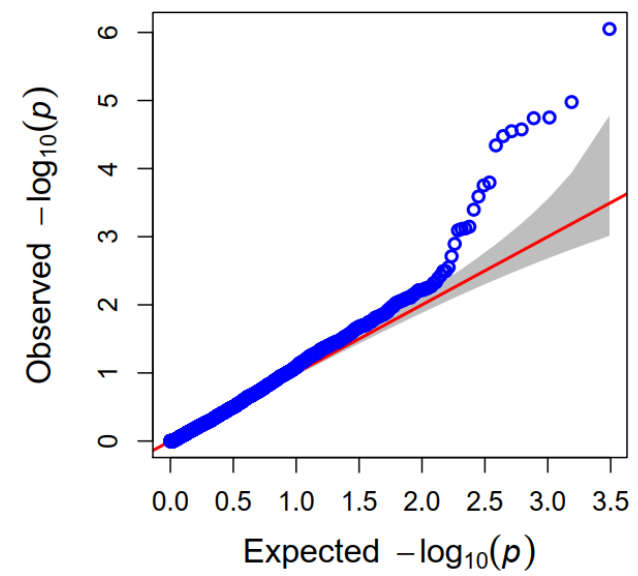
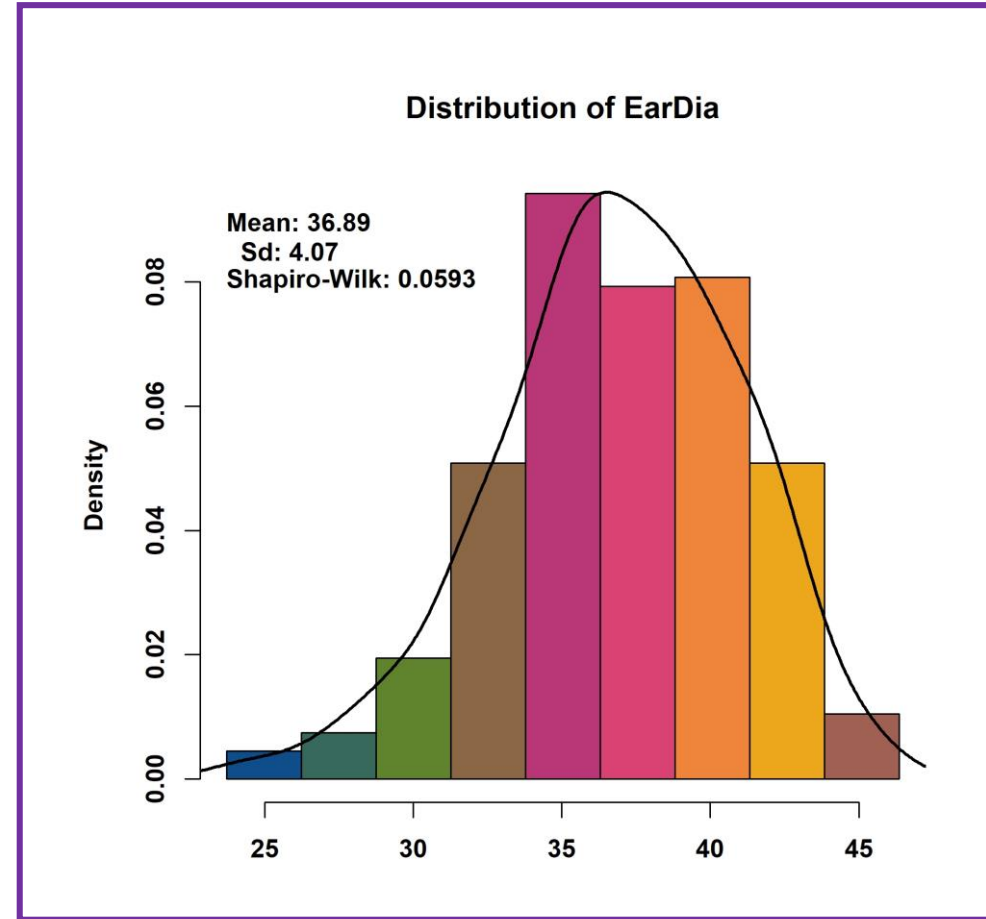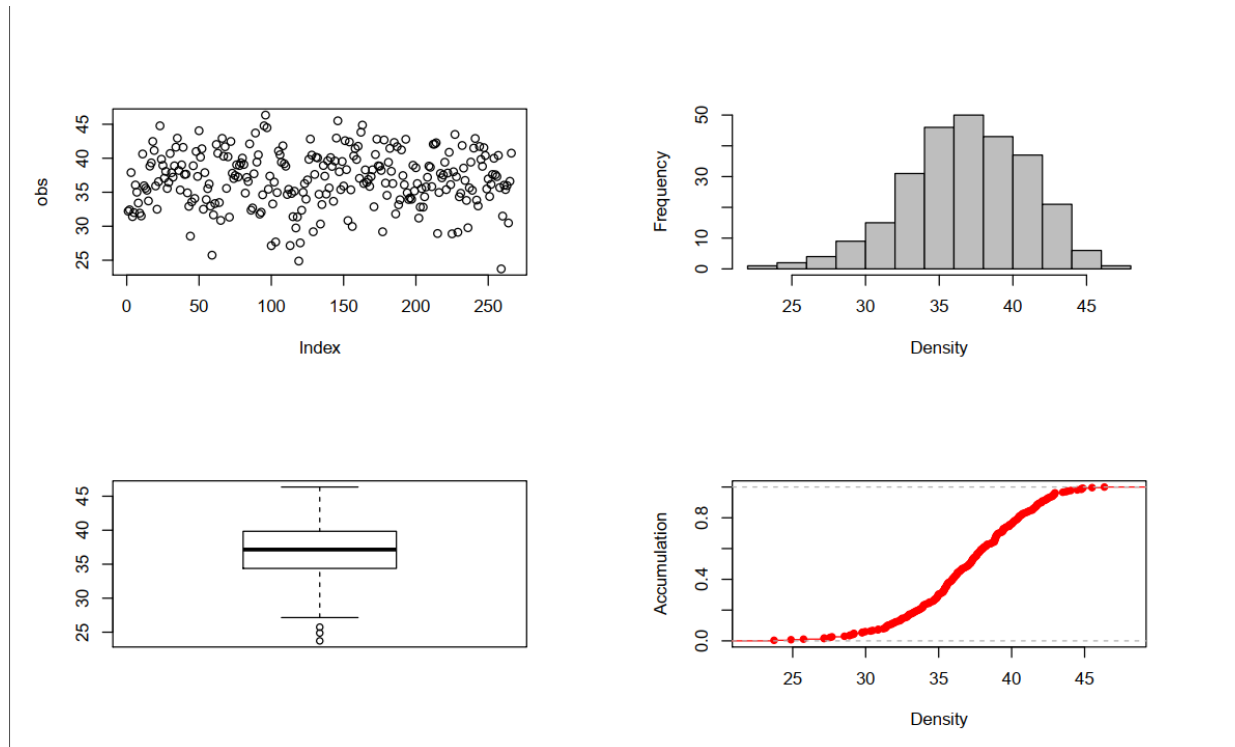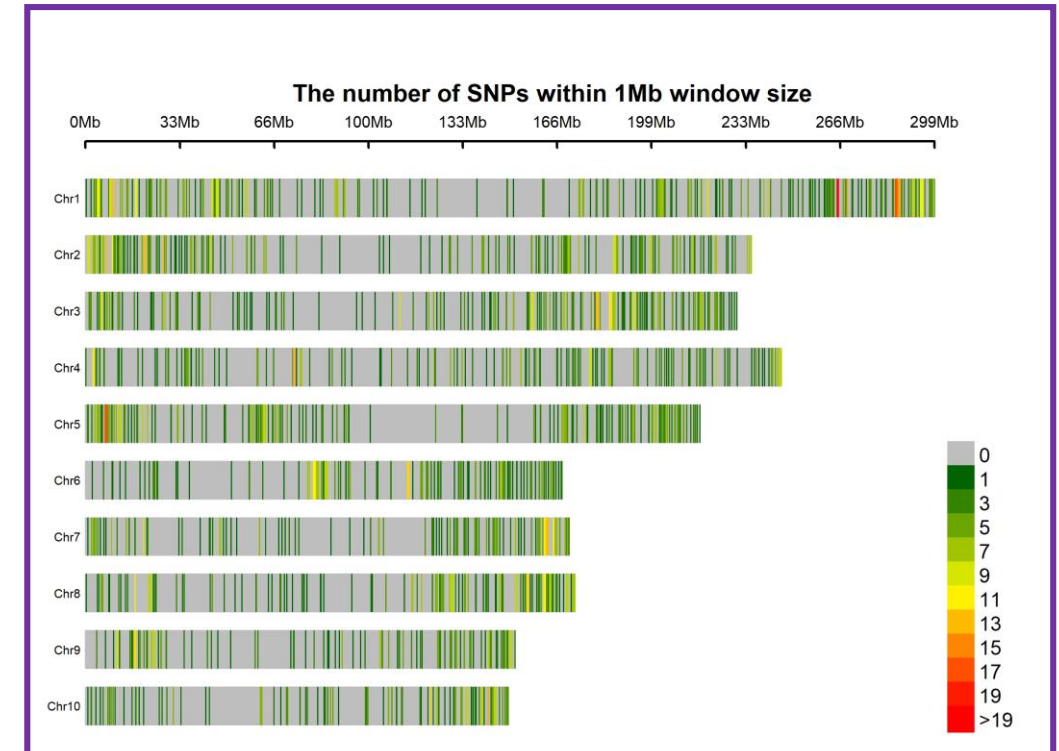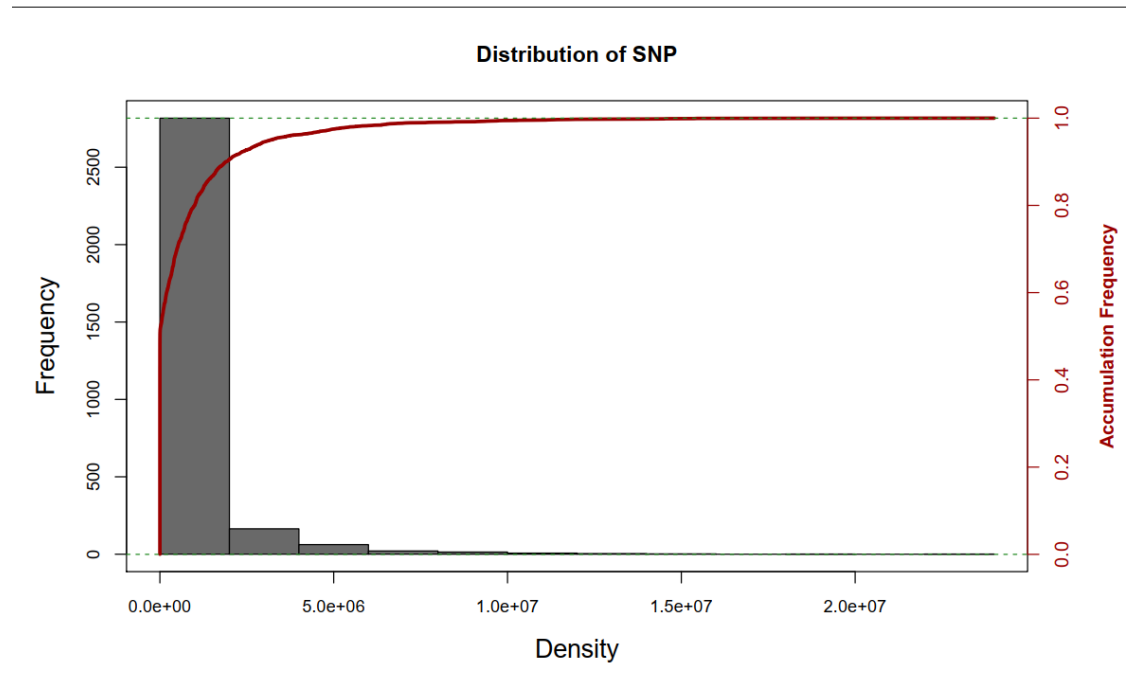# GAPIT – Outputs
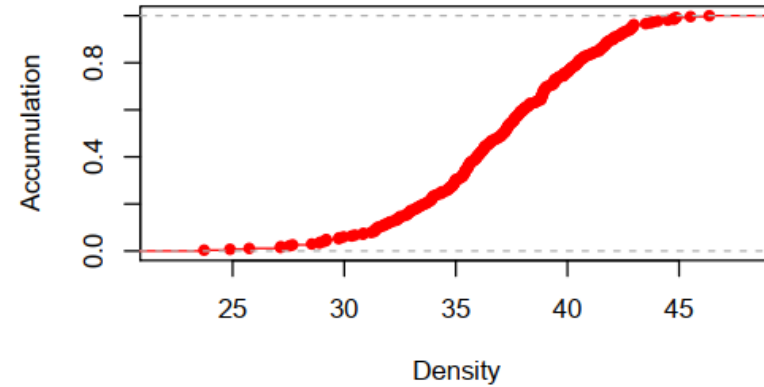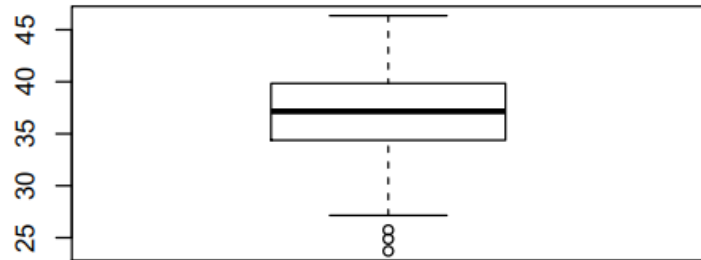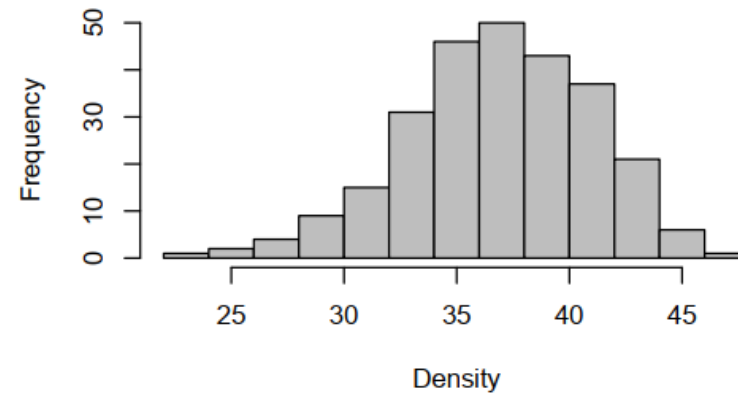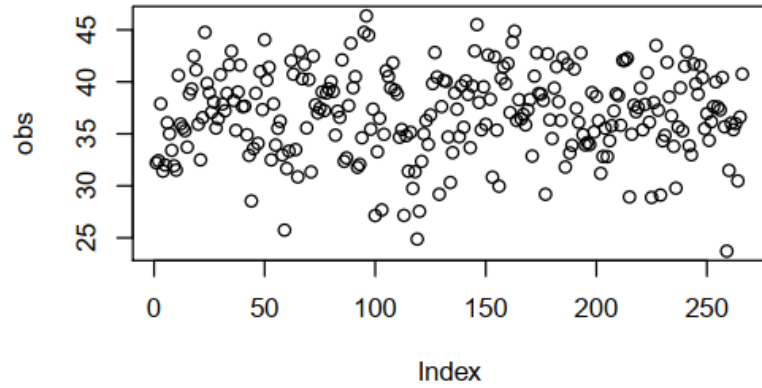
GAPIT – Outputs
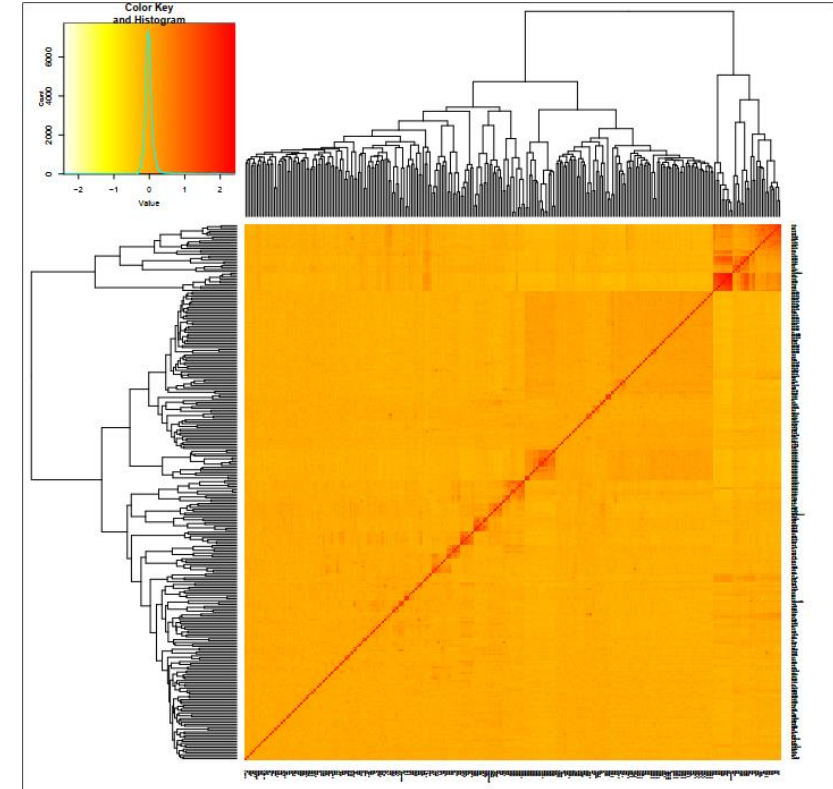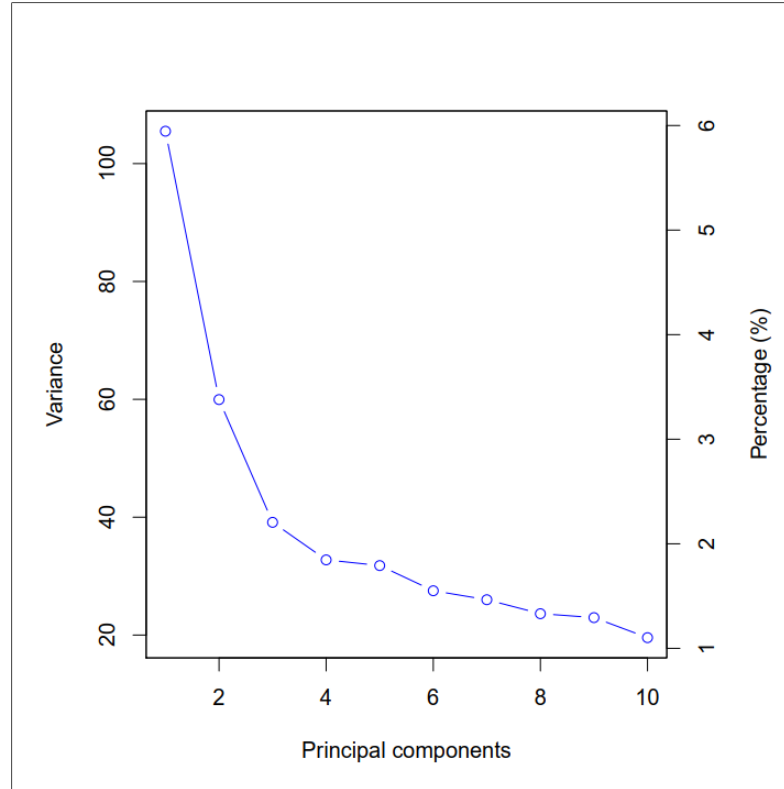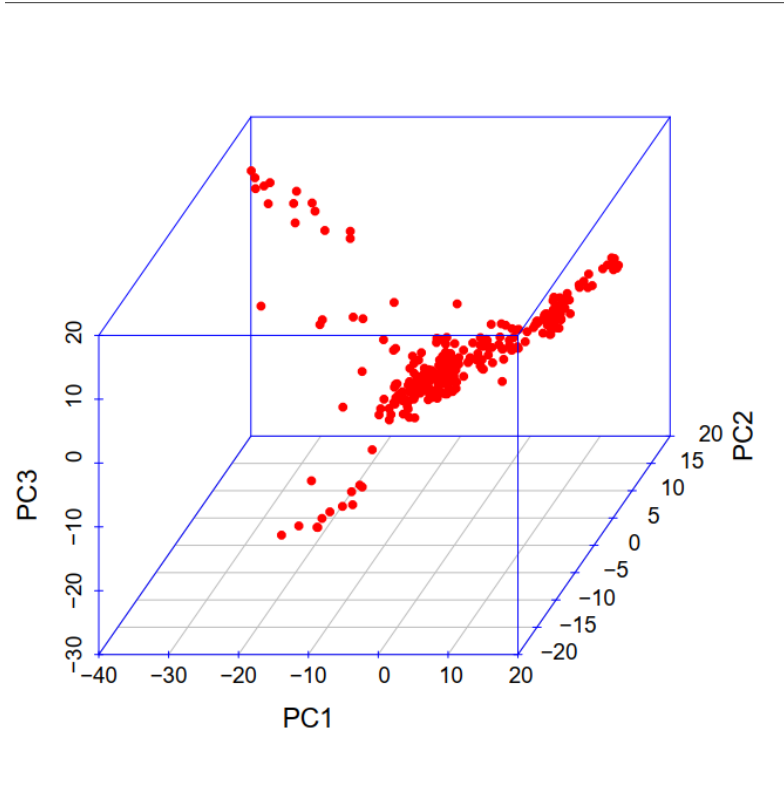
Demo – Redraw GWAS figure with rMVP


```
#phenotype histogram
MVP.Hist(phe=pheno_demo,file.type = "jpg",breakNum = 10,dpi = 300)
```

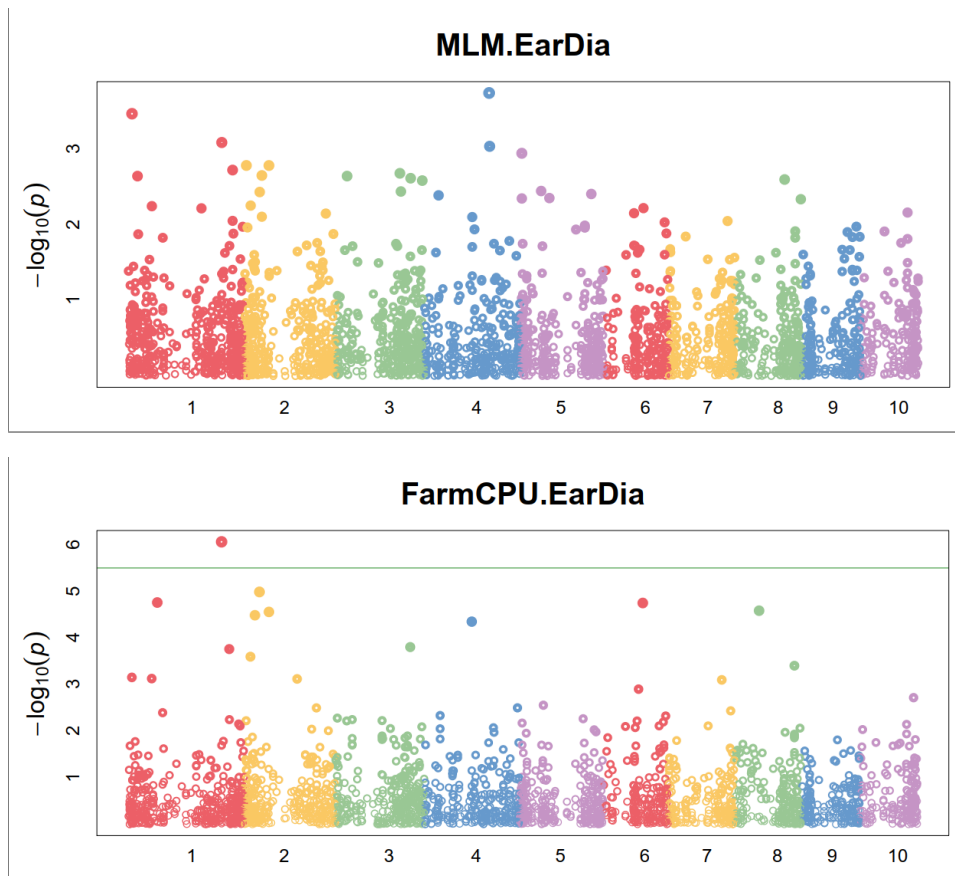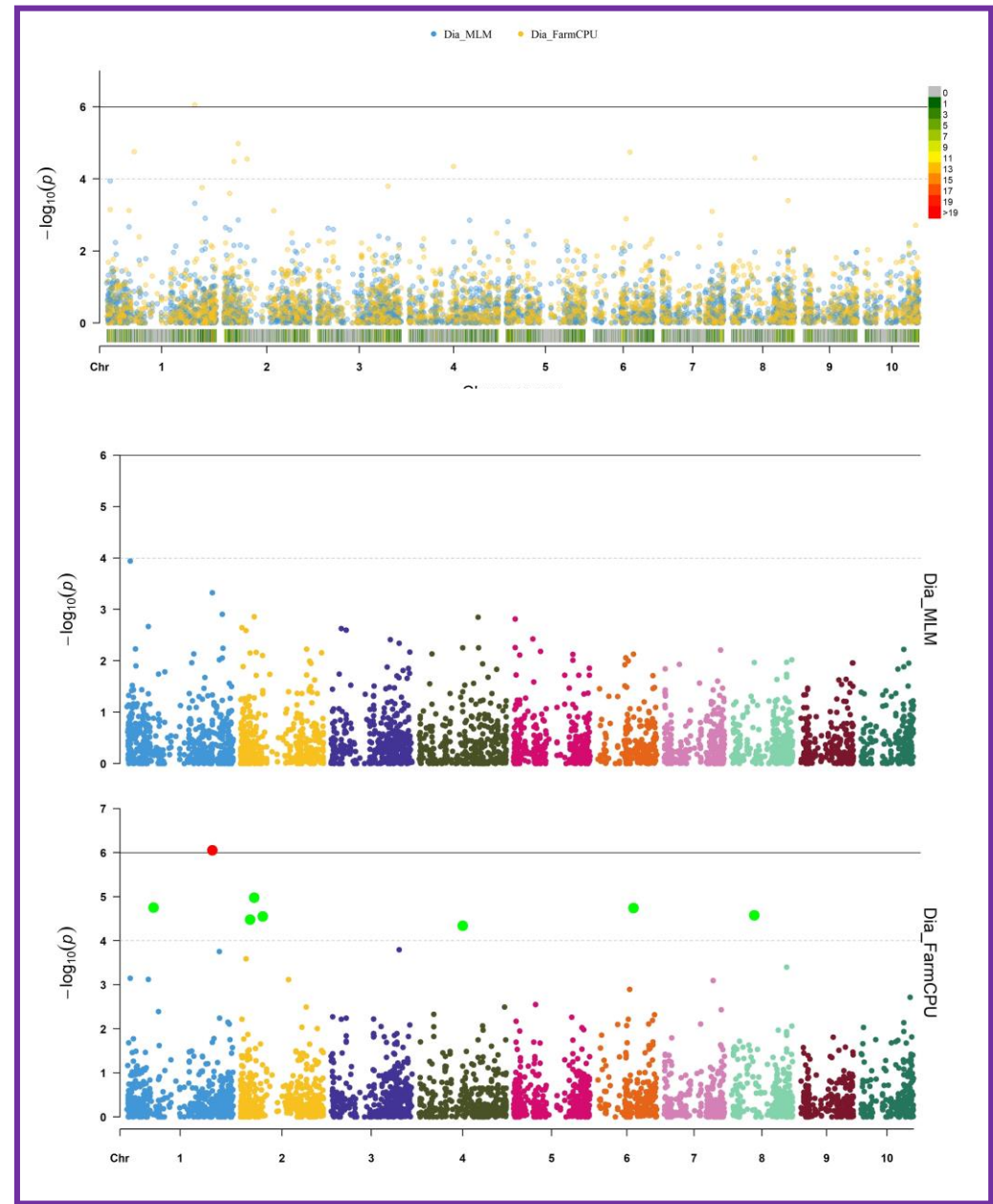# Demo – Redraw GWAS figure with rMVP


```
#SNP density
geno_mvp <- read.delim("mdp_genotype_test.hmp.txt",header = TRUE)
MVP.Report(geno_mvp[, c(1,3,4)], plot.type="d", col=c("darkgreen", "yellow", "red"), file.type="jpg", dpi=300)
```

Demo – Redraw GWAS figure with rMVP


```
Dia_MLMM <- read.csv("GAPIT.MLMM.EarDia.GWAS.Results.csv",header = TRUE)
colnames(Dia_MLMM)[4] <- c("Dia_MLM")
Dia_FarmCPU <- read.csv("GAPIT.FarmCPU.EarDia.GWAS.Results.csv",header = TRUE)
colnames(Dia_FarmCPU)[4] <- c("Dia_FarmCPU")
Dia <- merge(Dia_MLMM[,1:4],Dia_FarmCPU[,c(1,4)],by="SNP")%>%arrange(Chromosome)
MVP.Report(Dia, plot.type="m", multracks=TRUE, threshold=c(1e-6,1e-4),threshold.lty=c(1,2),
        threshold.lwd=c(1,1), threshold.col=c("black","grey"), amplify=TRUE,bin.size=1e6,
        chr.den.col=c("darkgreen", "yellow", "red"), signal.col=c("red","green"),signal.cex=c(1,1),
        file.type="jpg",memo="",dpi=300)
```

Demo – Redraw GWAS figure with rMVP
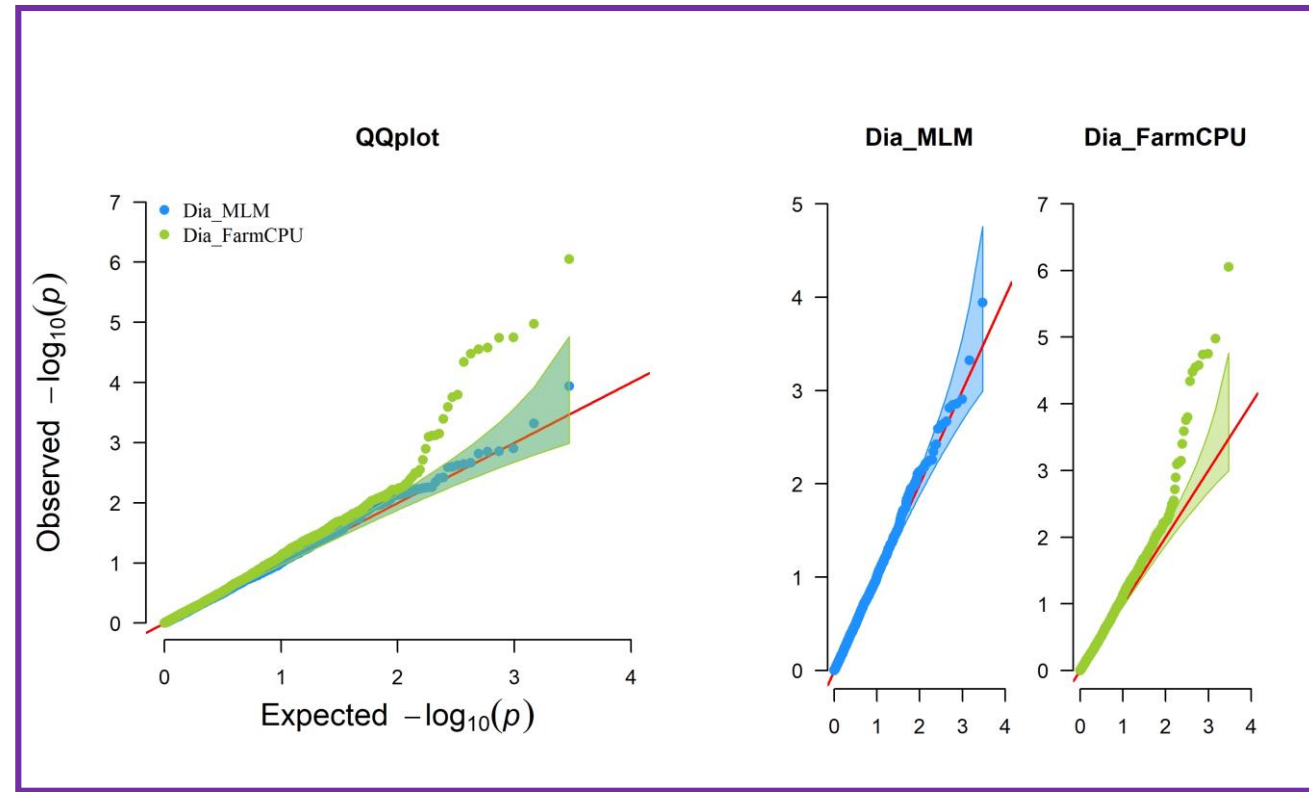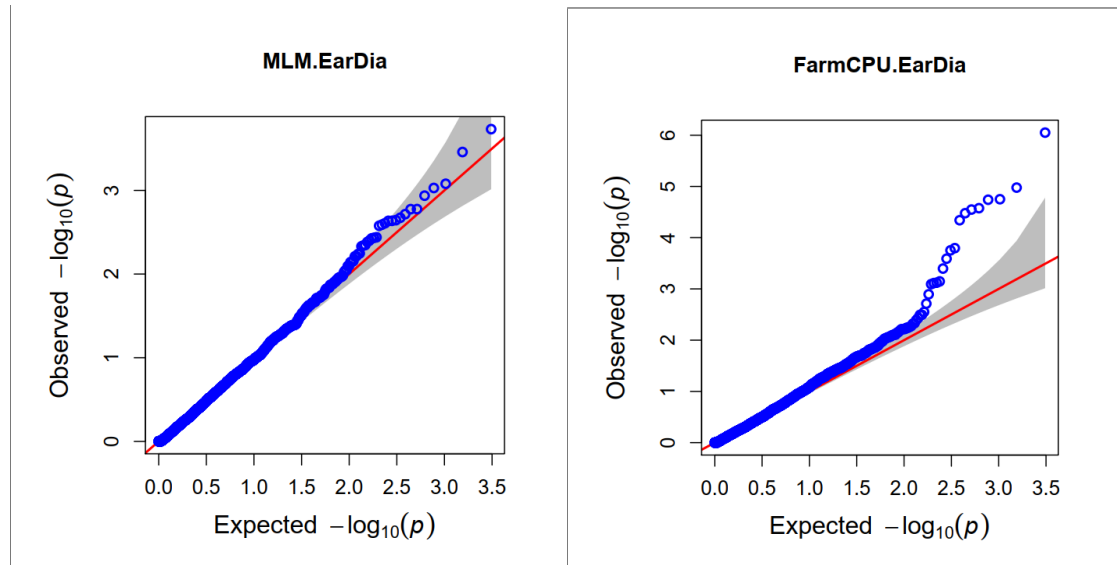
MVP.Report(Dia,plot.type="q",col=c("dodgerblue1", "olivedrab3", "darkgoldenrod1"),threshold=1e6,
signal.pch=19,signal.cex=1.5,signal.col="red",box=FALSE,multracks=
TRUE,file.type="jpg",memo="",dpi=300)