MuGenT Statistical Overview

1. FORM A GENE-SPECIFIC SNP SET

−50Kb   +50Kb

Gene body

DNA
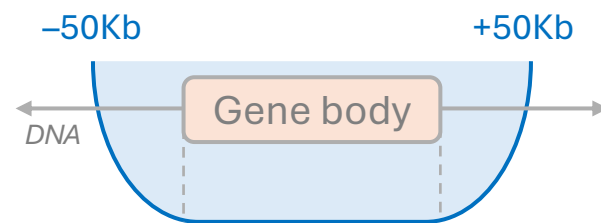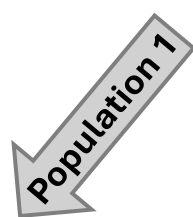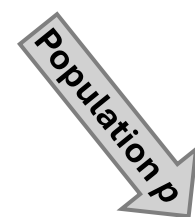
$\{ \text{SNP}_1, \text{SNP}_2, \ldots, \text{SNP}_M \}$

In-sample or population-matched LD reference panel

Population 1   ...   Population p
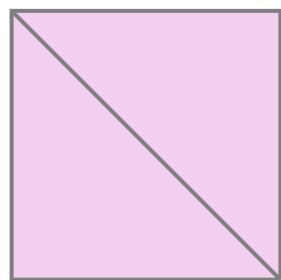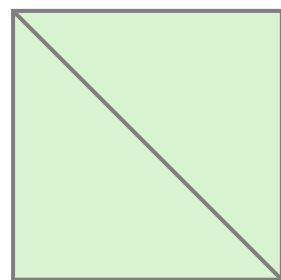
LD matrix ($\mathbf{R}_1$)   ...   LD matrix ($\mathbf{R}_p$)

2. ESTIMATE MARGINAL SNP EFECT SIZES

GWAS marginal models
(population k)

$$Y_k = \text{SNP}_1 \times \beta_{k1} + e_{k1}$$
$$Y_k = \text{SNP}_2 \times \beta_{k2} + e_{k2}$$
$$\vdots$$
$$Y_k = \text{SNP}_M \times \beta_{kM} + e_{kM}$$

Estimated effect sizes and SEs

$$\{ \hat{\beta}_{k1}, \hat{\beta}_{k2}, \ldots, \hat{\beta}_{kM} \}$$
$$\{ \text{SE}(\hat{\beta}_{k1}), \text{SE}(\hat{\beta}_{k2}), \ldots, \text{SE}(\hat{\beta}_{kM}) \}$$

SNP Z-statistics

$$Z_{kj} = \hat{\beta}_{kj} \times \text{SE}(\hat{\beta}_{kj})^{-1}$$
$$\mathbf{z}_1 = (Z_{11}, Z_{12}, \ldots, Z_{1M})^\top$$
$$\vdots$$
$$\mathbf{z}_p = (Z_{p1}, Z_{p2}, \ldots, Z_{pM})^\top$$

3. PERFORM MULTI-POPULATION GENE-BASED ASSOCIATION TEST (MuGenT)

MuGenT hypotheses

$$H_0: \cap_{j=1}^{M} \cap_{k=1}^{p} \beta_{kj} = 0, \qquad H_1: \cup_{j=1}^{M} \cup_{k=1}^{p} \beta_{kj} \neq 0$$
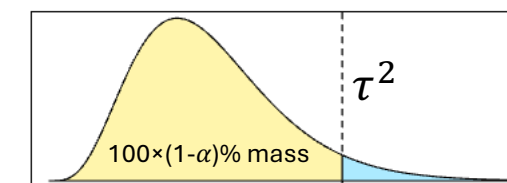
Test statistic T

$$\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_p), \qquad \mathbf{1} = (1, \ldots, 1)^\top$$
$$T = \mathbf{1}^\top \mathbf{Z}^\top \mathbf{Z} \mathbf{1} M^{-1}$$

Null distribution

$$T \mid H_0 \sim \text{Gamma}(\alpha_0, \xi_0)$$
$$\xi_0 = M \upsilon^{-1}, \qquad \alpha_0 = M \xi_0, \qquad \upsilon = \text{f}(\mathbf{R}_1, \ldots, \mathbf{R}_p)$$

100×(1−α)% mass τ^2

Test at level α

$$T > \tau^2 \rightarrow \text{Reject } H_0$$
$$T \leq \tau^2 \rightarrow \text{Do not reject } H_0$$

MuGenT + MuGenT-PH Computational Overview

Input

population 1, ..., **population p**

GWAS summary statistics

rsID	Z	A1	CHR	POS
rs1	-1.2	A	1	1223
rs2	1.4	G	3	985
rs3	0.6	C	8	6583
⋮	⋮	⋮	⋮	⋮

rsID	Z	A1	CHR	POS
rs1	-0.1	A	1	1223
rs2	0.8	G	3	985
rs3	2.6	C	8	6583
⋮	⋮	⋮	⋮	⋮

Allele-matched LD references

Subject	rs1	rs2	rs3	...
ID1	A/A	G/T	C/A	...
ID2	A/G	G/T	C/C	...
ID3	G/G	T/T	A/A	...
⋮	⋮	⋮	⋮	⋱

Subject	rs1	rs2	rs3	...
IDA	A/G	T/T	A/A	...
IDB	G/G	G/T	C/C	...
IDC	A/A	G/G	A/A	...
⋮	⋮	⋮	⋮	⋱

Data preparation

Gene A
rs1 rs2 rs3: Z_{11} Z_{12} Z_{13} Za1
Z_{p1} Z_{p2} Z_{p3} Zap
+ Ra1, Rap

Gene B
rs5 rs6: Z_{15} Z_{16} Zb1
Z_{p5} Z_{p6} Zbp
+ Rb1, Rbp

Gene C
rs6 rs7 rs8 rs9: Z_{16} Z_{17} Z_{18} Z_{19} Zc1
Z_{p6} Z_{p7} Z_{p8} Z_{p9} Zcp
+ Rc1, Rcp
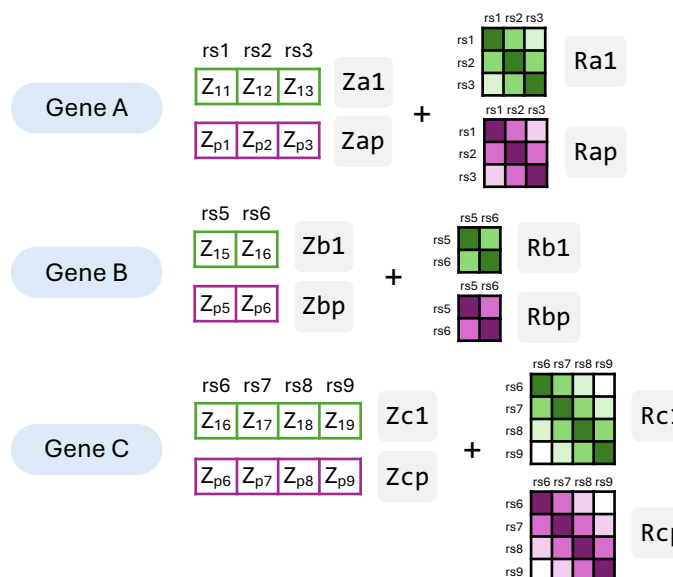
Computation

```
# Gene A
Za <- cbind(Za1, Zap)
Ra <- list(Ra1, Rap)
test_a <- mugent(Za, Ra)


# Gene B
Zb <- cbind(Zb1, Zbp)
Rb <- list(Rb1, Rbp)
test_b <- mugent(Zb, Rb)


# Gene C
Zc <- cbind(Zc1, Zcp)
Rc <- list(Rc1, Rcp)
test_c <- mugent(Zc, Rc)
```

Output

```
# MuGenT Example for gene A
> str(test_a)
List of 5
 $ pval     : num 4.61e-47
 $ shape    : num 8.44
 $ rate     : num 1.69
 $ mu_h0    : num 5
 $ sigma2_h0: num 2.96
```
← P-value testing MuGenT H_0
← α_0 (null Gamma shape)
← ξ_0 (null Gamma rate)
← number of tested populations
← υ (null variance)

```
# MuGenT-PH Example for gene A
> mugent_ph(Za, Ra) %>% str()
List of 5
 $ pval     : num 0.00788
 $ shape    : num 20.2
 $ rate     : num 0.0269
 $ mu_h0    : int 750
 $ sigma2_h0: num 27856
```
← P-value testing MuGenT-PH H_0
← null Gamma shape
← null Gamma rate
← number of tested SNPs
← null variance