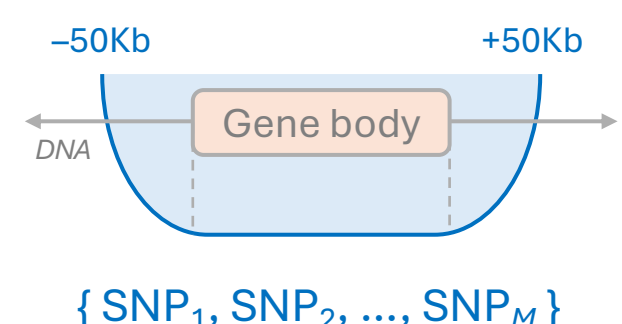


GenT Statistical Overview

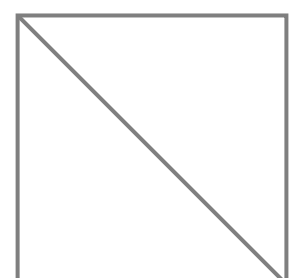
1. FORM A GENE-SPECIFIC SNP SET



$\{ \text{SNP}_1, \text{SNP}_2, \dots, \text{SNP}_M \}$

In-sample or population-matched LD reference panel

LD matrix (**R**)



2. ESTIMATE MARGINAL SNP EFFECT SIZES

GWAS marginal models

$$\begin{aligned} Y &= \text{SNP}_1 \times \beta_1 + e_1 \\ Y &= \text{SNP}_2 \times \beta_2 + e_2 \\ &\vdots \\ Y &= \text{SNP}_M \times \beta_M + e_M \end{aligned}$$

Estimated effect sizes and SEs

$$\begin{aligned} &\{ \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M \} \\ &\{ \text{SE}(\hat{\beta}_1), \text{SE}(\hat{\beta}_2), \dots, \text{SE}(\hat{\beta}_M) \} \end{aligned}$$

SNP Z-statistics

$$\begin{aligned} Z_j &= \hat{\beta}_j \times \text{SE}(\hat{\beta}_j)^{-1} \\ \mathbf{z} &= (Z_1, Z_2, \dots, Z_M)^\top \end{aligned}$$

3. PERFORM GENE-BASED ASSOCIATION TEST (GenT)

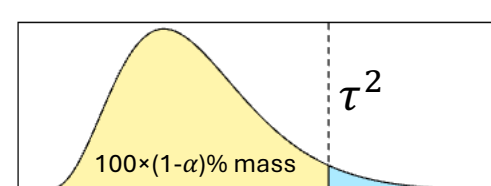
GenT hypotheses

$$H_0: \cap_{j=1}^M \beta_j = 0, \quad H_1: \cup_{j=1}^M \beta_j \neq 0$$

Test statistic T

$$T = \mathbf{z}^\top \mathbf{z} = \sum_{j=1}^M Z_j^2$$

Null distribution

$$T \mid H_0 \sim \text{Gamma}(\alpha_0, \xi_0)$$
$$\xi_0 = M / \text{trace}(\mathbf{2RR}), \quad \alpha_0 = M \xi_0$$


Test at level α

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

GenT Computational Overview

Input

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
⋮	⋮	⋮	⋮	⋮

Allele-matched LD reference

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	...
⋮	⋮	⋮	⋮	⋮

Data preparation

Gene A

rs1	rs2	rs3
Z1	Z2	Z3

Za

Gene B

rs5	rs6
Z5	Z6

Zb

Gene C

rs6	rs7	rs8	rs9
Z6	Z7	Z8	Z9

Zc

Computation

```
# Gene A
test_a <- gent(Za, Ra)

# Gene B
test_b <- gent(Zb, Rb)

# Gene C
test_c <- gent(Zc, Rc)
```

Output

```
# Example for gene A
> str(test_a)
List of 5
 $ pval      : num 1.96e-06
 $ shape     : num 12.7
 $ rate      : num 0.0158
 $ mu_h0     : int 805
 $ sigma2_h0: num 51026
```

← P-value testing GenT H_0
← α_0 (null Gamma rate)
← ξ_0 (null Gamma shape)
← M (number of tested SNPs)
← $2\text{trace}(\mathbf{RR})$ (null variance)