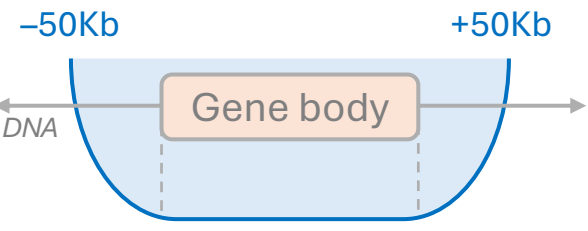


xGenT Statistical Overview

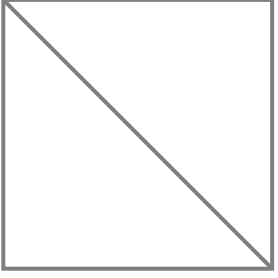
1. FORM A GENE-SPECIFIC SNP SET



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

In-sample or population-matched LD reference panel

LD matrix (**R**)



2. ESTIMATE MARGINAL SNP EFFECT SIZES

GWAS marginal models

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

Estimated effect sizes and SEs

$$\{ \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M \}$$
$$\{ \text{SE}(\hat{\beta}_1), \text{SE}(\hat{\beta}_2), \dots, \text{SE}(\hat{\beta}_M) \}$$

SNP Z-statistics

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

2. ESTIMATE MARGINAL xQTL EFFECT SIZES

xQTL marginal models
(xQTL trait L_k ; e.g., gene expression in tissue k)

$$L_k = \text{SNP}_1 \times \mu_{k1} + \epsilon_{k1}$$
$$L_k = \text{SNP}_2 \times \mu_{k2} + \epsilon_{k2}$$
$$\vdots$$
$$L_k = \text{SNP}_M \times \mu_{kM} + \epsilon_{kM}$$

Estimated effect sizes and SEs

$$\{ \hat{\mu}_{k1}, \hat{\mu}_{k2}, \dots, \hat{\mu}_{kM} \}$$
$$\{ \text{SE}(\hat{\mu}_{k1}), \text{SE}(\hat{\mu}_{k2}), \dots, \text{SE}(\hat{\mu}_{kM}) \}$$

xQTL Z-statistics

$$V_{kj} = \hat{\mu}_{kj} \times \text{SE}(\hat{\mu}_{kj})^{-1}$$
$$\mathbf{v}_1 = (V_{11}, V_{12}, \dots, V_{1M})^\top$$
$$\vdots$$
$$\mathbf{v}_K = (V_{K1}, V_{K2}, \dots, V_{KM})^\top$$

4. PERFORM xQTL-INTEGRATED GENE-BASED ASSOCIATION TEST (xGenT)

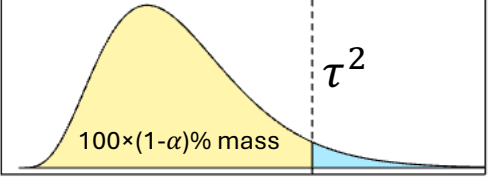
xGenT hypotheses

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

Test statistic T

$$T = \mathbf{z}^\top \mathbf{L} \mathbf{z}, \quad \mathbf{L} = \left(\sum_{k=1}^K \mathbf{v}_k \mathbf{v}_k^\top \right)$$

Null distribution

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


Test at level α

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

xGenT 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	...
⋮	⋮	⋮	⋮	⋮

xQTL summary statistics

rsID	Z	A1	CHR	POS
rs1	-0.1	A	1	1223
rs2	0.8	G	3	985
rs3	1.0	C	8	6583
⋮	⋮	⋮	⋮	⋮

...

rsID	Z	A1	CHR	POS
rs1	-5.8	A	1	1223
rs2	4.4	G	3	985
rs3	3.2	C	8	6583
⋮	⋮	⋮	⋮	⋮

Data preparation

Gene A

	rs1	rs2	rs3
Za	Z1	Z2	Z3

Tissue 1

	rs1	rs2	rs3
Z11	Z11	Z12	Z13

Tissue K

	rs1	rs2	rs3
ZK1	ZK1	ZK2	ZK3

xZa1 + Ra

Gene B

	rs6	rs7	rs8	rs9
Zb	Z6	Z7	Z8	Z9

Tissue 1

	rs6	rs7	rs8	rs9
Z16	Z16	Z17	Z18	Z19

Tissue K

	rs6	rs7	rs8	rs9
ZK6	ZK6	ZK7	ZK8	ZK9

xZb1 + Rb

Computation

```
# Gene A
eqtl_Za <- cbind(xZa1, xZaK)
test_a <- gent(
  Za,
  LD = Ra,
  xqtl_Z = eqtl_Za)

# Gene B
eqtl_Zb <- cbind(xZb1, xZbK)
test_b <- gent(
  Zb,
  LD = Rb,
  xqtl_Z = eqtl_Zb)
```

Output

```
# Example for gene A
> str(test_a)
List of 5
 $ pval      : num 0.0948
 $ shape     : num 0.523
 $ rate      : num 0.00388
 $ mu_h0     : num 135
 $ sigma2_h0: num 34818
```

← P-value testing xGenT H_0

← α_0 (null Gamma rate)

← ξ_0 (null Gamma shape)

← M (number of tested SNPs)

← $2\text{trace}(\mathbf{LRLR})$ (null variance)