Scan variance QTL for GxE interaction

Xiaoran, Dmitri

Overview

1. Two types of variance loci (VLC)
2. identify variance loci
3. Simulations
4. Proof
5. Pit falls

GxE induced variance loci

Let $\boldsymbol{y}$ be affected by $\boldsymbol{g} \in \{0, 1, 2\}$, unobserved $\boldsymbol{u}$, and interaction $\boldsymbol{gu}$

$$\boldsymbol{y} = m_0 + a\boldsymbol{g} + b\boldsymbol{gu} + c\boldsymbol{u} + \boldsymbol{\epsilon}, \quad \epsilon \sim \mathcal{N}(0, v_\epsilon), \quad \boldsymbol{g} \perp \boldsymbol{u} \tag{1}$$

Fitting the mean of $\boldsymbol{y}$ given $\boldsymbol{g}$, but since $\boldsymbol{u}$ is unseen, we have

$$\boldsymbol{y} = m_0 + a\boldsymbol{g} + \boldsymbol{e} \quad \textbf{or} \quad \mathrm{E}(\boldsymbol{y}|\boldsymbol{g}) = m_0 + a\boldsymbol{g} \tag{2}$$

Unexplained $(c + b\boldsymbol{g})\boldsymbol{u} + \boldsymbol{\epsilon}$ flood into residual $\boldsymbol{e}$, as a result,

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \mathrm{E}(\boldsymbol{e}^2|\boldsymbol{g}) = (v_\epsilon + c^2 v_u) + 2bcv_u\boldsymbol{g} + b^2 v_u \boldsymbol{g}^2 \tag{3}$$

The phenotype variance depends on the genotype, thus $\boldsymbol{g}$ is a "variance loci"; when $2bcv_u \neq 0$ or $b^2 v_u > 0$, we know $b \neq 0$, and the presence of GxE, $b\boldsymbol{gu}$, even if we do not observe $\boldsymbol{u}$.

Direct variance loci

Suppose $\boldsymbol{g}$ directly determine the mean and variance of $\boldsymbol{y}$,

$$\boldsymbol{y} \sim \mathcal{N}(m_0 + a\boldsymbol{g},\ \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2)) \tag{4}$$

again, explain the mean of $\boldsymbol{y}$ with $\boldsymbol{g}$ gets

$$\boldsymbol{y} = m_0 + a\boldsymbol{g} + \boldsymbol{e}, \quad \textbf{or} \quad \mathrm{E}(\boldsymbol{y}|\boldsymbol{g}) = m_0 + a\boldsymbol{g} \tag{5}$$

The squared leftover $\boldsymbol{e}^2$ is Gamma distributed,

$$\boldsymbol{e}^2 \sim \mathrm{Gamma}(\mathrm{shape} = 2,\ \mathrm{scale} = \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2)) \tag{6}$$

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \mathrm{E}(\boldsymbol{e}^2|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2) \tag{7}$$

When $\alpha \neq 0$ or $\beta \neq 0$, $\boldsymbol{g}$ is a variance loci by definition, without an origin.

Connection between two types of variance loci

If $\boldsymbol{y} = m_0 + a\boldsymbol{g} + b\boldsymbol{g}\boldsymbol{u} + c\boldsymbol{u} + \boldsymbol{\epsilon}$, $\boldsymbol{g}$ is a GxE induced variance loci, we see

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = (v_\epsilon + c^2 v_u) + 2bcv_u\boldsymbol{g} + b^2 v_u \boldsymbol{g}^2$$

If we force $\boldsymbol{g}$ to be a direct variance loci such that

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\tilde{\mu} + \tilde{\alpha}\boldsymbol{g} + \tilde{\beta}\boldsymbol{g}^2).$$

As a result, the following approximation holds

$$\tilde{\mu} \approx \log(v_e), \quad \tilde{\alpha} \approx 2bcv_u/v_e, \quad \tilde{\beta} \approx b^2 v_u/v_e - 2b^2c^2v_u^2/{v_e}^2 \tag{8}$$

where $v_e = v_\epsilon + c^2 v_u$ is part of the variance unaffected by $\boldsymbol{g}$. When white noise $v_\epsilon = 0$, the above simplifies to

$$\tilde{\mu} \approx 2\log c + \log v_u, \quad \tilde{\alpha} \approx 2b/c, \quad \tilde{\beta} \approx -b^2/c^2 \tag{9}$$

Connection between two types of variance loci

If $\boldsymbol{y} \sim \mathcal{N}(m_0 + a\boldsymbol{g}, \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2))$, $\boldsymbol{g}$ is a direct variance locus,

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2)$$

Perceive $\boldsymbol{g}$ as a GxE induced variance loci with latent environment $\tilde{\boldsymbol{u}}$,

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = (v_{\tilde{\epsilon}} + \tilde{c}^2 v_{\tilde{u}}) + 2\tilde{b}\tilde{c}v_{\tilde{u}}\boldsymbol{g} + \tilde{b}^2 v_{\tilde{u}}\boldsymbol{g}^2,$$

As a result, the following approximation holds

$$(v_{\tilde{\epsilon}} + \tilde{c}^2 v_{\tilde{u}}) \approx e^{\mu}, \quad 2\tilde{b}\tilde{c}v_{\tilde{u}} \approx \alpha e^{\mu}, \quad \tilde{b}^2 v_{\tilde{u}} \approx (\alpha^2 + 2\beta)e^{\mu}/2 \tag{10}$$

If we let $v_{\tilde{\epsilon}} = 0$, naturally, $\tilde{c}^2 v_{\tilde{u}} \approx e^{\mu}$, and consequently

$$\mu \approx 2\log\tilde{c} + \log v_{\tilde{u}}, \quad \alpha \approx 2\tilde{b}/\tilde{c}, \quad \beta \approx -\tilde{b}^2/\tilde{c}^2, \tag{11}$$

Are they equivalent?

Direct variance loci are too generic

- given v_u, b, c uniquely determine $\tilde{\mu}$, $\tilde{\alpha}$, and $\tilde{\beta}$;
- given v_u, μ, α, and β can result in unreal $\tilde{b}$, since $\tilde{b}^2 \propto (\alpha^2 + 2\beta)e^{\mu}$, and $\alpha^2 + 2\beta \in (-\infty, \infty)$

Direct variance loci do not necessarily indicate GxE.

Direct variance loci are too rigid, if restricting $\beta = 0$

- restricting $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g})$ is a mainstream practice.
- no risk of $\tilde{b}$ being unreal, but imposes $\tilde{b}/\tilde{c} = \alpha$.

Direct variance loci with $\beta = 0$ do not capture all GxE.

Are they equivalent?

How about restricted GxE vQTL ($b = 0$)

- restricted $V(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_2\boldsymbol{g}$ is popular;
  - $a_2\boldsymbol{g} \approx 2bcv_u\boldsymbol{g}$ also capture GxE effect b;
  - when MAF is low, absorb $a_2\boldsymbol{g}$ part of $b^2v_u\boldsymbol{g}^2$, since $\boldsymbol{g} \not\perp \boldsymbol{g}^2$,
- lower power when environmental main effect is low ($c \approx 0$), or when MAF is moderate, causing $\boldsymbol{g}$ and $\boldsymbol{g}^2$ nearly independent.
- less specific when environmental main effect is big ($|c| > 0$), since a significant $|a_2|$ does not necessarily imply GxE effect $|b|$ is large.
- deviated from the nature result of GxE, since $b^2v_u\boldsymbol{g}^2$ is deduced from the premises that $b\boldsymbol{gu}$ exists.

Are they equivalent?

The essence of GxE caused vQTL

Fitting the squared residual $\boldsymbol{e}^2$ with squared SNP $\boldsymbol{g}^2$

$$\boldsymbol{e}^2 = (v_\epsilon + c^2 v_u) + 2bcv_u\boldsymbol{g} + b^2 v_u \boldsymbol{g}^2 \tag{12}$$

implies

$$\frac{\partial \boldsymbol{e}^2}{\partial \boldsymbol{g}} = 2bcv_u + 2b^2 v_u \boldsymbol{g} \tag{13}$$

as a result, rejecting $h_0 : b^2 v_u \leq 0$ concludes that the derivative of e^2 wrt. $\boldsymbol{g}$ is positive, or, the curve of e^2 over $\boldsymbol{g}$ must bend upwards.

Suggestions

The goal is to screen for GxE candidate

- favor the nature model deduced from GxE, over Gaussian vQTL.
  - $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = (v_\epsilon + c^2 v_u) + 2bcv_u\boldsymbol{g} + b^2 v_u \boldsymbol{g}^2$
- consider the quadratic term $\boldsymbol{g}^2$,
  - model $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_2\boldsymbol{g} + a_3\boldsymbol{g}^2$, test $h_0 : a_2 = a_3 = 0$.
- exploit the mutual absorption between $\boldsymbol{g}$ and $\boldsymbol{g}^2$
  - model use $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_3\boldsymbol{g}^2$, test $h_0 : a_3 \leq 0$
  - $a_3\boldsymbol{g}^2 \approx b^2 v_u \boldsymbol{g}^2$ is GxE specific;
  - absorb $a_2 \approx 2bcv_u\boldsymbol{g}$ when $\boldsymbol{g}$ and $\boldsymbol{g}^2$ is near parallel;
  - more powerful one tail t-test, while expecting $a_3 \geq 0$.

Gaussian Environment, Gaussian Genotype

Simulated Truth:

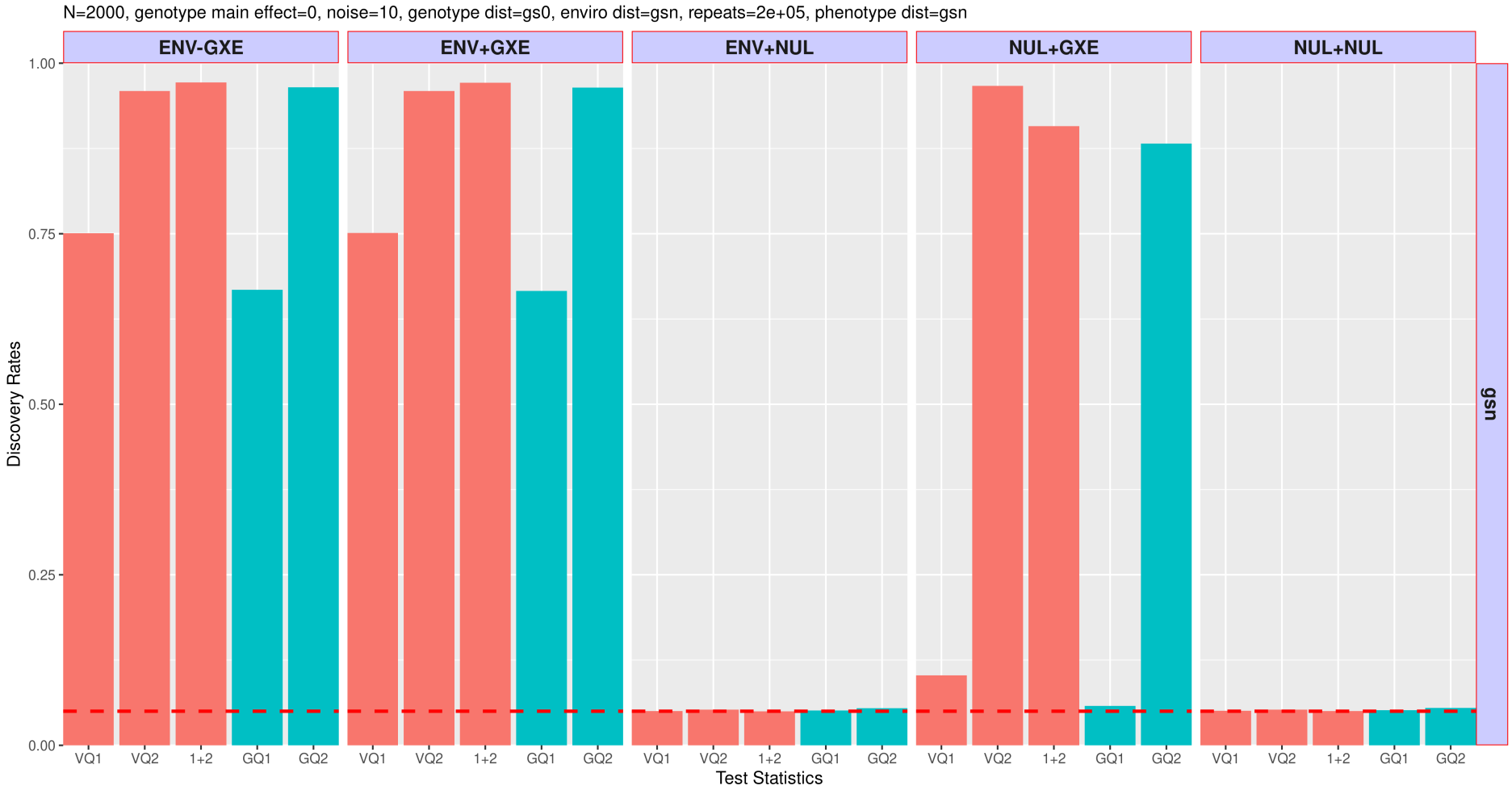
- for power check:
  - ENV $\pm$ GXE: $\boldsymbol{y} = c\boldsymbol{u} \pm b\boldsymbol{gu}$, environmental main and GxE;
  - NUL + GXE: $\boldsymbol{y} = 0\boldsymbol{u} + b\boldsymbol{gu}$, GxE only;
- for type 1 error check
  - ENV + NUL: $\boldsymbol{y} = c\boldsymbol{u} + 0\boldsymbol{gu}$, environmental main only;
  - NUL + NUL: $\boldsymbol{y} = 0\boldsymbol{u} + 0\boldsymbol{gu}$, total null;

Competing Test Statistics

- VQ1: $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_2\boldsymbol{g}$, mainstream linear vQTL 2-tail test
- VQ2: $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_3\boldsymbol{g}^2$, proposed 1-tail test ($a_3 \geq 0$)
- 1+2: $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = a_1 + a_2\boldsymbol{g} + a_3\boldsymbol{g}^2$, full model
- GQ1: $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g})$, mainstream Gaussian vQTL 2-tail test
- GQ2: $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2)$, full Gaussian vQTL 2-tail test

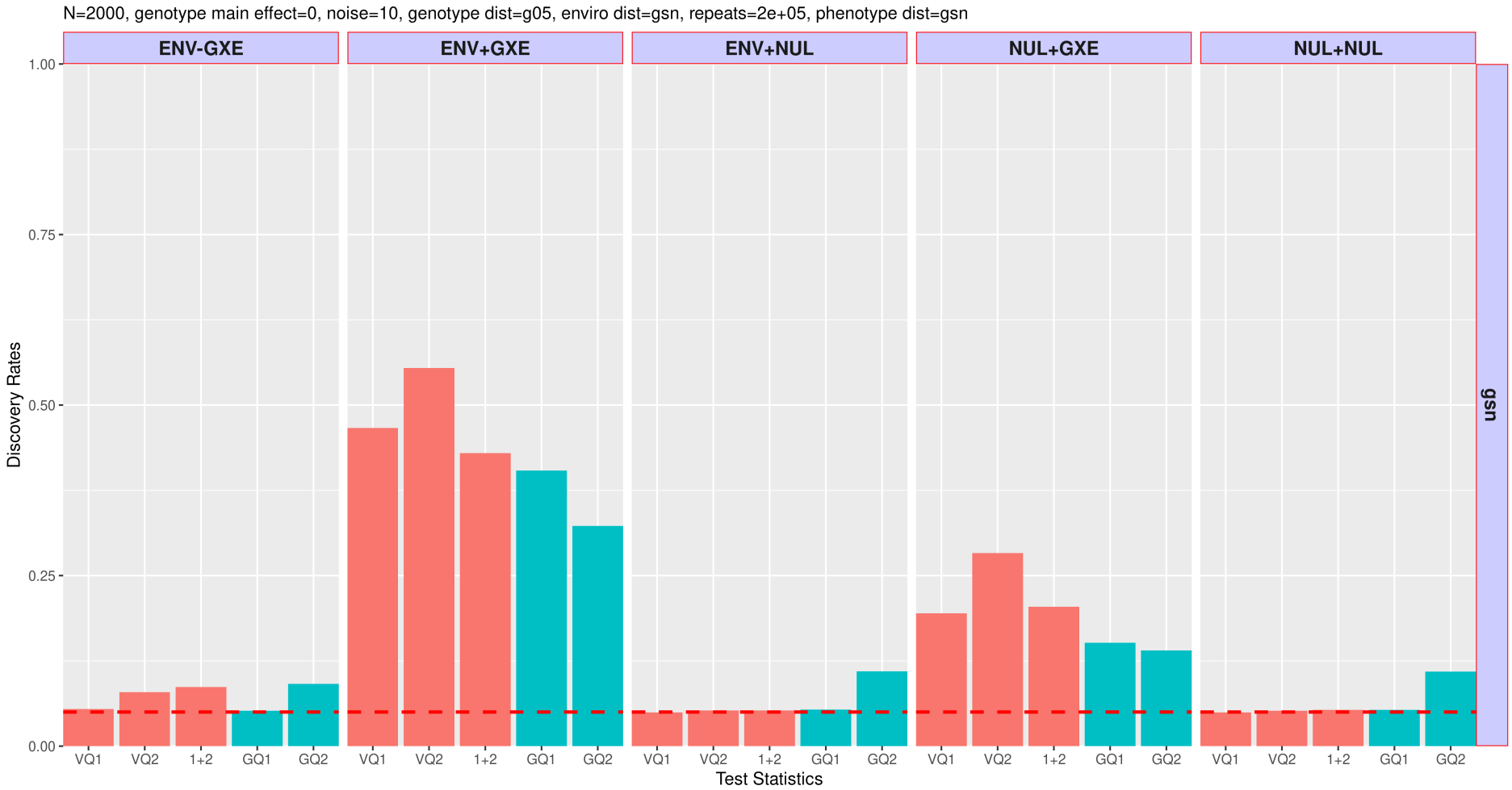
Gaussian Environment, Gaussian Genotype

Genotype normal, unrealistically eliminate absorption between $\boldsymbol{g}$ and $\boldsymbol{g}^2$



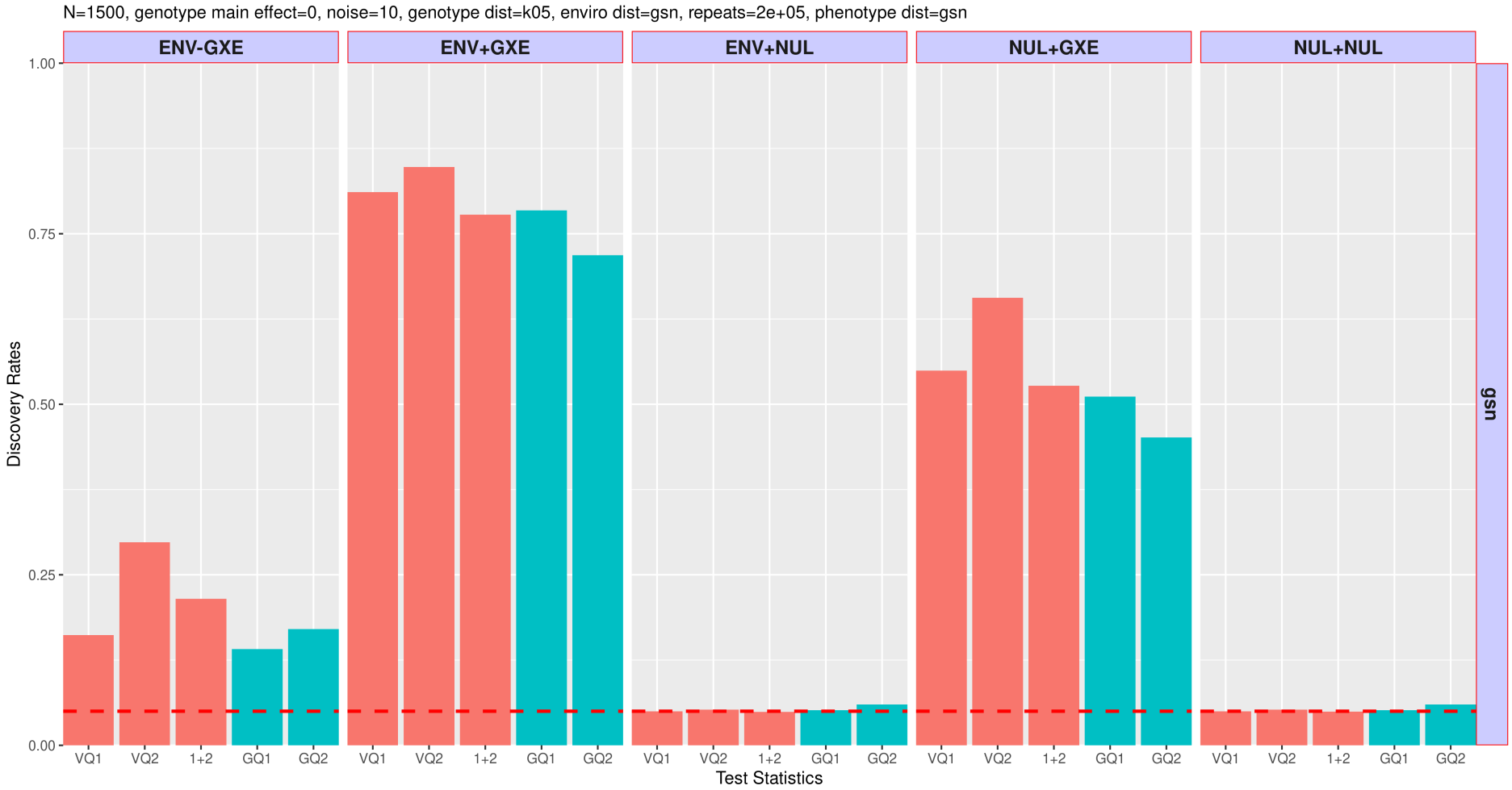
Gaussian Environment, Binomial Genotype

Genotype is Binomial(0.05, 2), $\boldsymbol{g} \not\perp \boldsymbol{g}^2$, heavy absorption.



Gaussian Environment, 1000G project data

Genotype is drawn from 1000G, MAF>0.05, realistic absorption.



Pseudo Real Data analysis for BMI

Pseudo ground truth of GxE

- `ENV: env = alc + smk + edu + inc`
- `GWA: bmi = snp + age * sex + env`
- `GXE: bmi = snp + age * sex + env + snp : env`
- `test GWA = GXE`, rejection counts snp as a Pseudo true GxE SNP

Q: how many pseudo true GxE SNP passes various vQTL screening tests – the positive predictive value (PPV), relative to no screening at all?

Address issues with binary outcome

I1: T2D case, I2: T2D non-case

Pseudo Real Data analysis for BMI

Proof: polynomial Gaussian vQTL $\rightarrow$ interaction

If $\boldsymbol{g}$ is a polynmial Guassian vQTL that

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2) \tag{14}$$

$$\rightarrow \frac{\mathrm{V}(\boldsymbol{y}|\boldsymbol{g})}{e^{\mu}} = 1 + \alpha\boldsymbol{g} + \frac{(\alpha^2 + 2\beta)}{2}\boldsymbol{g}^2 + \sum_{n=3}^{\infty} \frac{s^{(n)}}{n!}\boldsymbol{g}^n \tag{15}$$

$$s^{(n)} = \begin{cases} s^{(n-1)}\alpha + 2s^{(n-2)}\beta, & n > 0 \\ 1, & n = 0 \\ 0, & n < 0 \end{cases} \tag{16}$$

Here we Taylor expand $f(\boldsymbol{g}) = e^{\alpha\boldsymbol{g}+\beta\boldsymbol{g}^2}$ at $\mathbf{0}$.

Proof: polynomial Gaussian vQTL $\rightarrow$ interaction

When $\boldsymbol{g} \sim \mathrm{N}(0, v_g)$, $v_g < 1$, $|\alpha| < 1$ and $|\beta| < 1$, approximation

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) \approx e^{\mu} + \alpha e^{\mu}\boldsymbol{g} + \frac{(\alpha^2 + 2\beta)e^{\mu}}{2}\boldsymbol{g}^2 \tag{17}$$

is accurate. If we force the vQTL to be interaction induced, that is,

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = c^2 v_u + 2bcv_u\boldsymbol{g} + b^2 v_u\boldsymbol{g}^2,$$

Naturally, $e^{\mu} = c^2 v_u$, and

$$\mu = 2\log c + \log v_u, \quad \alpha = \frac{2b}{c}, \quad \beta = -\frac{b^2}{c^2}, \tag{18}$$

Proof: interaction $\rightarrow$ polynomial Gussian vQTL

If $\boldsymbol{g}$ is interaction induced vQTL, $\boldsymbol{y} = a\boldsymbol{g} + b\boldsymbol{g}\boldsymbol{u} + c\boldsymbol{u} + \boldsymbol{\epsilon}$, we know

$$\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = v_\epsilon + c^2 v_u + 2bcv_u\boldsymbol{g} + b^2 v_u \boldsymbol{g}^2$$

$$\begin{aligned}
\rightarrow \log \mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) &= \log(v_\epsilon + c^2 v_u) + \frac{2bcv_u}{v_\epsilon + c^2 v_u}\boldsymbol{g} + \left[\frac{b^2 v_u}{v_\epsilon + c^2 v_u} - \frac{2b^2 c^2 v_u^2}{(v_\epsilon + c^2 v_u)^2}\right]\boldsymbol{g}^2 \\
&\quad + \sum_{n=3}^{\infty}\left[\frac{(-1)^n (2b^2 v_u)^{n-2}}{n(v_\epsilon + c^2 v_u)^{n-1}} + \frac{(-1)^{n-1}(2bcv_u)^n}{n(v_\epsilon + c^2 v_u)^n}\right]\boldsymbol{g}^n \\
&\approx \log(v_\epsilon + c^2 v_u) + \frac{2bcv_u}{v_\epsilon + c^2 v_u}\boldsymbol{g} + \left[\frac{b^2 v_u}{v_\epsilon + c^2 v_u} - \frac{2b^2 c^2 v_u^2}{(v_\epsilon + c^2 v_u)^2}\right]\boldsymbol{g}^2
\end{aligned}$$

Here we Taylor expand $f(\boldsymbol{g}) = \log \mathrm{V}(\boldsymbol{y}|\boldsymbol{g})$ at $\boldsymbol{0}$.

Interaction induced vQTL fit by log-Gamma

The approximation up to the second order

$$\log \phi \approx \log v_u + 2 \log c + \frac{2b}{c}\boldsymbol{g} - \frac{b^2}{c^2}\boldsymbol{g}^2 = \tilde{\mu} + \tilde{\alpha}\boldsymbol{g} + \tilde{\beta}\boldsymbol{g}^2 \qquad (19)$$

is accurate when $\boldsymbol{g} \sim \mathrm{N}(0, v_g)$, $v_g < 1$, and $|c| > |b|$. Thus, $\boldsymbol{y} = a\boldsymbol{g} + b\boldsymbol{g}\boldsymbol{u} + c\boldsymbol{u}$, give rise to $\mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = e^{\tilde{\mu}+\tilde{\alpha}\boldsymbol{g}+\tilde{\beta}\boldsymbol{g}^2}$; $\tilde{\mu}$ is the effect of $\boldsymbol{u}$ on $\boldsymbol{y}$; $\tilde{\alpha}$ and $\tilde{\beta}$ relate to the effect size of interaction $\boldsymbol{g}\boldsymbol{u}$ on $\boldsymbol{y}$ relative to the effect size of $\boldsymbol{u}$ on $\boldsymbol{y}$.

A polynomial Gaussian vQTL induce interaction between genotype $\boldsymbol{g}$ and latent variable $\tilde{\boldsymbol{u}}$, when $\beta = -0.25\alpha^2$.

Non-centered interaction

When $\boldsymbol{u}$ is the difference from non-zero mean u_0

$$\boldsymbol{y} = a\boldsymbol{g} + b\boldsymbol{g}(\boldsymbol{u} + u_0) + c(\boldsymbol{u} + u_0), \quad \boldsymbol{g} \perp \boldsymbol{u} \tag{20}$$

Explain the mean of $\boldsymbol{y}$ with $\boldsymbol{u} + u_0$ unseen,

$$\begin{aligned} \boldsymbol{y} &= cu_0 + (a + bu_0)\boldsymbol{g} + \boldsymbol{e}, \quad \textbf{or} \\ \mathrm{E}(\boldsymbol{y}|\boldsymbol{g}) &= cu_0 + (a + bu_0)\boldsymbol{g} \end{aligned} \tag{21}$$

Notice spurious genetic effect bu_0. Still, $(c + b\boldsymbol{g})\boldsymbol{u}$ flow into $\boldsymbol{e}$, therefore vQTL is unaffected.

Exposome Challenge 2021

Suppose any one of the phenotypes follows a multivariate normal. Under the basic null, we say

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}, \sigma_0^2 \boldsymbol{I} + \boldsymbol{\Sigma}) \tag{22}$$

$$\boldsymbol{\mu} = \boldsymbol{X}_c \boldsymbol{\alpha}, \tag{23}$$

$$\boldsymbol{\Sigma} = \sigma_t^2 \boldsymbol{K}_t + \sigma_m^2 \boldsymbol{K}_m + \sigma_s^2 \boldsymbol{K}_s + \sigma_u^2 \boldsymbol{K}_u \tag{24}$$

where $\boldsymbol{X}_c$ is the matrix of confounders; and $\boldsymbol{K}_*$ are kernels of transcriptom, methylation, expression, seurm metabolom, and urine metabolome.
This null model left out exposome.

Exposome Challenge 2021

We include exposome, one at a time, say $\boldsymbol{z}$

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_0 + \boldsymbol{\mu}, \sigma_0^2 \boldsymbol{I} + \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{v_i}) \tag{25}$$

$$\boldsymbol{\mu} = \boldsymbol{X}_c \boldsymbol{\alpha} + \boldsymbol{z}\beta, \tag{26}$$

$$\boldsymbol{\Sigma} = \sigma_t^2 \boldsymbol{K}_t + \sigma_m^2 \boldsymbol{K}_m + \sigma_s^2 \boldsymbol{K}_s + \sigma_u^2 \boldsymbol{K}_u \tag{27}$$

$$+ \tau_t^2 \boldsymbol{V}\boldsymbol{K}_t + \tau_m^2 \boldsymbol{V}\boldsymbol{K}_m + \tau_s^2 \boldsymbol{V}\boldsymbol{K}_s + \tau_u^2 \boldsymbol{V}\boldsymbol{K}_u \tag{28}$$

$$\boldsymbol{V} = \boldsymbol{z}'\boldsymbol{z} \tag{29}$$

where $\boldsymbol{z}$ is an exposome variant; $\boldsymbol{V}$ is the kernel of that variants. This model the alternative, with exposome.
Testing H0: $\tau_*^2 = 0$ amounts to test, exposome is not interacting with the omics.

General vQTL

Let $\boldsymbol{g}$ be a vQTL by directly determine the variance of $\boldsymbol{y}$,

$$\boldsymbol{\mu}_y = h^{-1}(m_0 + a\boldsymbol{g}), \quad \boldsymbol{v}_y = \theta(\boldsymbol{\mu}_y)\phi^{-1}(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2), \tag{30}$$

If $\boldsymbol{y}$ is Gaussian, then $h^{-1}(\boldsymbol{x}) = \boldsymbol{x}$ and $\theta(\mu_y) = 1$, and

$$\boldsymbol{y} = m_0 + a\boldsymbol{g} + \boldsymbol{e}, \quad \textbf{or} \quad \mathrm{E}(\boldsymbol{y}|\boldsymbol{g}) = \boldsymbol{\mu}_y = m_0 + a\boldsymbol{g} \tag{31}$$

The squared residual $\boldsymbol{e}^2$ is Gamma distributed

$$\boldsymbol{e}^2 \sim \mathrm{Gamma}(1, \phi^{-1}(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2)), \tag{32}$$

$$\textbf{or} \quad \mathrm{V}(\boldsymbol{y}|\boldsymbol{g}) = \mathrm{E}(\boldsymbol{e}^2|\boldsymbol{g}) = \phi^{-1}(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2) \tag{33}$$

When $\alpha \neq 0$ or $\beta \neq 0$, $\boldsymbol{g}$ is a vQTL with no specif origin (i.e., due to GxE).

Connection between two types vQTL

If $\boldsymbol{y} \sim \mathcal{N}(a\boldsymbol{g}, \exp(\mu + \alpha\boldsymbol{g} + \beta\boldsymbol{g}^2))$ and $\beta = -0.25\alpha^2$, $\boldsymbol{g}$ is a constrained polynomial Gaussian vQTL, then $\boldsymbol{g}$ interacts with latent $\tilde{\boldsymbol{u}}$ such that

$$\boldsymbol{y} = a\boldsymbol{g} + \tilde{b}\boldsymbol{g}\tilde{\boldsymbol{u}} + \tilde{c}\tilde{\boldsymbol{u}} \tag{34}$$

$$\tilde{b} \approx \frac{\tilde{c}}{2}\alpha, \quad \tilde{c} \approx \sqrt{\frac{e^{\mu}}{v_{\tilde{u}}}}, \quad \tilde{\boldsymbol{u}} \perp \boldsymbol{g}, \quad v_{\tilde{u}} = \mathrm{Var}(\boldsymbol{u}) \tag{35}$$

Suggestions

The goal is to detect Gene by Environment interaction

- with measured environment (i.e., $\boldsymbol{u}$ become visible),
- with variants pre-selected by vQTL scan across genome.

We prefer interaction induced vQTL,

- prefer linear model over log-Gamma regression, the later may detect vQTL not specifically due to interaction.
- include quadratic term $\boldsymbol{g}^2$ to enhance the detection.

Our analysis of Type 2 diabetes:

- scan vQTL on UK Biobank blood assays for GxE SNPs;
- pairing the SNPs with measured environment in PEGS;
- analyze the pairs for GxE interaction affecting T2D.