Supplementary Material 1

Conditional genotype distribution of the checkerboard model

Penetrance $p_i = \dfrac{\exp(\alpha + x(g_i)\beta + z_i\gamma)}{1 + \exp(\alpha + x(g_i)\beta + z_i\gamma)}$ where $\alpha = -5.29$, $\beta = 2.5$, and $\gamma = 1$,

respectively. $x(g_i) = 1$ when $g_i \in$ {AABb, AaBB, Aabb, aaBb} for the high-risk

genotypes, and $x(g_i) = 0$ for the rest genotypes when $g_i \in$ {AABB, AAbb, AaBb, aaBB,

aabb}. Take MAF=0.5 and assume Hardy-Weinberg equilibrium, the frequencies of the

nine genotypes can be calculated. z has the form of the normal distribution $Z \sim N(0, 10)$,

and the probability density function $f(z) = \dfrac{1}{\sqrt{20\pi}}\exp\left(-\dfrac{z^2}{20}\right)$.

AABb and AABB are employed for high- and low-risk genotypes, respectively.

Using numerical solution to integrate the expressions of the penetrances,

$$P(\mathcal{A}|\text{AABb}) = \int_{-\infty}^{+\infty} \frac{\exp(-2.79 + z)}{1 + \exp(-2.79 + z)} f(z)\mathrm{d}z = 0.221$$

$$P(\mathcal{U}|\text{AABB}) = \int_{-\infty}^{+\infty} \frac{\exp(-5.29 + z)}{1 + \exp(-5.29 + z)} f(z)\mathrm{d}z = 0.073$$

Consider the case sample first, by repeated application of Bayes's theorem, for genotype

k, $P(k|\mathcal{A}) = \dfrac{P(\mathcal{A}|k)P(k)}{\sum_i P(\mathcal{A}|i)P(i)}$

$$P(\text{AABb}|\mathcal{A}) = 0.188$$

$$P(\text{AABB}|\mathcal{A}) = 0.031$$

Similarly, for the control sample,

$$P(\text{AABb}|\mathcal{U}) = 0.114$$

$$P(\text{AABB}|\mathcal{U}) = 0.068$$

The rest of genotypic frequencies in the case and control samples can be calculated, respectively. After multiplying these frequencies by sample size, the distribution illustrated in Figure 1A can be calculated.

Supplementary Material 2

The conditional score distribution

$$s_i = y_i - p_i \text{ is nonlinear as } p_i = \frac{\exp(\alpha + z_i\gamma)}{1 + \exp(\alpha + z_i\gamma)} \text{ where } \alpha \text{ and } \gamma \text{ are MLE under}$$

the null distribution, yielding $\hat{\alpha} = -1.891$ and $\hat{\gamma} = 0.884$ for the case demonstrated.

According to the definition of mathematical expectation, $E(s_i) = \int_z (y_i - p_i)\eta(z)\mathrm{d}z$,

where $\eta(z) = \dfrac{\exp(\alpha + z\gamma)}{1 + \exp(\alpha + z\gamma)} f(z)$, the probability a subject being affected. When

$\int_z \eta(z)\mathrm{d}z = 1$ is not guaranteed,

$$E(s_i) = \frac{\int_z (y_i - p_i)\eta(z)\mathrm{d}z}{\int_z \eta(z)\mathrm{d}z}$$

where $\int_z \eta(z)\mathrm{d}z$ is the normalization factor. Consider an affected subject first. The

penetrance function makes every subject bearing genotype k and covariate z has the

probability of $P(\mathcal{A}|k, z) = \dfrac{\exp(\alpha + x(k)\beta + z\gamma)}{1 + \exp(\alpha + x(k)\beta + z\gamma)}$ being affected.

Take AABb and AABB as representatives for high- and low-risk genotypes, the

estimated score for the subjects affected, respectively

$$E(s)_{\text{AABb}}^{\mathcal{A}} = \frac{\int_{-\infty}^{+\infty}\left(1 - \frac{\exp(\hat{\alpha} + z\hat{\beta})}{1 + \exp(\hat{\alpha} + z\hat{\beta})}\right)\eta(\text{AABb}, z)\mathrm{d}z}{\int_{-\infty}^{+\infty}\eta(\text{AABb}, z)\mathrm{d}z} = 0.263$$

$$E(s)^{\mathcal{A}}_{\text{AABB}} = \frac{\int_{-\infty}^{+\infty}\left(1 - \frac{\exp(\hat{\alpha} + z\hat{\beta})}{1 + \exp(\hat{\alpha} + z\hat{\beta})}\right)\eta(\text{AABB}, z)\mathrm{d}z}{\int_{-\infty}^{+\infty}\eta(\text{AABB}, z)\mathrm{d}z} = 0.124$$

and, for unaffected ones, $\eta(k, z) = \left(1 - \frac{\exp\left(\alpha + x(k)\beta + z\gamma\right)}{1 + \exp(\alpha + x(k)\beta + z\gamma)}\right)f(z)$, the probability

a subjected being unaffected

$$E(s)^{\mathcal{U}}_{\text{AABb}} = \frac{\int_{-\infty}^{+\infty}\left(0 - \frac{\exp(\hat{\alpha} + z\hat{\beta})}{1 + \exp(\hat{\alpha} + z\hat{\beta})}\right)\eta(\text{AABb}, z)\mathrm{d}z}{\int_{-\infty}^{+\infty}\eta(\text{AABb}, z)\mathrm{d}z} = -0.188$$

$$E(s)^{\mathcal{U}}_{\text{AABB}} = \frac{\int_{-\infty}^{+\infty}\left(0 - \frac{\exp(\hat{\alpha} + z\hat{\beta})}{1 + \exp(\hat{\alpha} + z\hat{\beta})}\right)\eta(\text{AABB}, z)\mathrm{d}z}{\int_{-\infty}^{+\infty}\eta(\text{AABB}, z)\mathrm{d}z} = -0.265$$

When the expectations are calculated above, it is easy to calculate the sum of score for

each genotypic cell. For example, the sum over the scores for the affected subjects in cell

AABb is

$$\sum_i E((s_i)^{\mathcal{A}}_{\text{AABb}}) = \mathcal{N}^{\mathcal{A}}_{\text{AABb}} \times E(s)^{\mathcal{A}}_{\text{AABb}} = 49.5$$

and, for subjects unaffected

$$\sum_i E((s_i)^{\mathcal{U}}_{\text{AABb}}) = \mathcal{N}^{\mathcal{U}}_{\text{AABb}} \times E(s)^{\mathcal{U}}_{\text{AABb}} = -21.5$$

$\mathcal{N}^{\mathcal{A}}_{\text{AABb}}$ is estimated from genotypic distribution that $\frac{\mathcal{N}}{2} \times P(\mathcal{A}|\text{AABb})$, and, similarly,

can apply to the rest genotypic cells, respectively.