

Supplementary Text

1. Likelihood Approximation for Homozygous Individuals

Here, we provide details on the approximation used for homozygous individuals. All notations follow the main text. Recall that we denote y_i and r_i as the methylated read count and total read count for i th individual, for $i = 1, \dots, n$. We denote the corresponding read counts mapped to the haplotype of the two alleles of i th individual as y_{i1} and r_{i1} , for $l = 1$ or 2 ; thus we have $y_i = y_{i1} + y_{i2}$ and $r_i = r_{i1} + r_{i2}$. Recall that, for homozygous individuals, we derive a model on y_i and r_i by summing over all possible values of y_{i1} and r_{i1} :

$$P(y_i|r_i, \pi_{i1}, \pi_{i2}) = \sum_{y_{i1}=0}^{\min(r_{i1}, y_i)} \sum_{r_{i1}=0}^{r_i} P(y_{i1}|r_{i1}, \pi_{i1})P(y_i - y_{i1}|r_i - r_{i1}, \pi_{i2})P(r_{i1}|r_i), \quad (1)$$

where $P(y_{i1}|r_{i1}, \pi_{i1})$ follows a binomial distribution $y_{i1} \sim \text{Bin}(r_{i1}, \pi_{i1})$; $P(y_i - y_{i1}|r_i - r_{i1}, \pi_{i2})$ follows a binomial distribution $y_i - y_{i1} \sim \text{Bin}(r_i - r_{i1}, \pi_{i2})$; and $P(r_{i1}|r_i)$ follows a binomial distribution $r_{i1} \sim \text{Bin}(r_i, 0.5)$. Here, y_{i1} and y_{i2} are conditionally independent given π_{i1} and π_{i2} , though π_{i1} and π_{i2} are correlated with each other. To ensure scalable computation, we approximate the above $P(y_i|r_i, \pi_{i1}, \pi_{i2})$ using a binomial mixed model in the following simple form:

$$y_i \sim \text{Bin}(r_i, \tilde{\pi}_i), \quad (2)$$

where $\tilde{\pi}_i$ is a mean that depends on π_{i1} and π_{i2} , and thus is a function of $(\mu, \beta, x_{i1}, x_{i2}, g_i, u_i, e_{i1}, e_{i2})$. We obtain an estimate of $\tilde{\pi}_i$ by matching the first moment in the likelihood function defined by equation (2) to the first moment in the likelihood function defined by equation (1). Specifically, the expectation of y_i under equation (1) can be expressed as:

$$E(y_i) = E(y_{i1} + y_{i2}) = E(E(y_{i1} + y_{i2}|r_{i1}, r_{i2})) = r_i \frac{\pi_{i1} + \pi_{i2}}{2}.$$

Therefore, by matching moments, we have

$$\tilde{\pi}_i = (\pi_{i1} + \pi_{i2})/2 \\ = (g^{-1}(\mu + x_{i1}\beta + g_i + u_i + e_{i1}) + g^{-1}(\mu + x_{i2}\beta + g_i + u_i + e_{i2}))/2, \quad (3)$$

with

$$\begin{pmatrix} u_i + e_{i1} \\ u_i + e_{i2} \end{pmatrix} \sim N(0, (\sigma_u^2 + \sigma_e^2) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}), \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2},$$

where μ is the intercept; x_{il} is the l th allele type for i th individual for the SNP of interest ($x_{il} = 0$ or 1 , corresponding to the reference allele and alternative allele, respectively); β is the corresponding allele/genotype effect size; g_i represents the genetic background/polygenic effect of i th individual; u_i represents the common individual-level environmental effect that varies across individuals but is shared between the two alleles within the same individual; e_{il} represents the residual error and is used to account for independent environmental noise that varies across both individuals and alleles; $g(\cdot)$ denotes the logit link function.

Because we focus on homozygous individuals where $x_{i1} = x_{i2} = x_i$, we denote $\lambda_0 = \mu + x_i\beta + g_i + u_i + (e_{i1} + e_{i2})/2$, which is the average of $\mu + x_{i1}\beta + g_i + u_i + e_{i1}$ and $\mu + x_{i2}\beta + g_i + u_i + e_{i2}$. Due to the g^{-1} function, equation (3) is not in an easy-to-evaluate form. Instead, we further approximate each term in equation (3) through a first order Taylor expansion at λ_0 :

$$g^{-1}(\mu + x_{i1}\beta + g_i + u_i + e_{i1}) \approx \frac{e^{\lambda_0}}{e^{\lambda_0} + 1} \left(1 + \frac{\frac{e_{i1} - e_{i2}}{2}}{e^{\lambda_0} + 1} \right),$$

$$g^{-1}(\mu + x_{i1}\beta + g_i + u_i + e_{i2}) \approx \frac{e^{\lambda_0}}{e^{\lambda_0} + 1} \left(1 + \frac{\frac{e_{i2} - e_{i1}}{2}}{e^{\lambda_0} + 1} \right).$$

Therefore, we have

$$\tilde{\pi}_i \approx \frac{e^{\lambda_0}}{e^{\lambda_0} + 1},$$

or equivalently

$$g(\tilde{\pi}_i) \approx \lambda_0 = \mu + x\beta + g_i + u_i + \frac{e_{i1} + e_{i2}}{2} \quad (5)$$

for homozygous individuals. Note the above equation (5) holds exactly for heterozygous individuals, since we have $g(\pi_{il}) = \mu + x_{il}\beta + g_i + u_i + e_{il}$ for each allele of the heterozygous individuals (main text equation (2)). Subsequently, we can define an approximate background heritability as

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_u^2 + \sigma_e^2/2}. \quad (6)$$

2. IMAGE Inference Algorithm

For homozygous individual i , if we denote the random effects for individual i as $b_i = g_i + u_i + \frac{e_{i1} + e_{i2}}{2}$, we have:

$$y_i \sim \text{Bin}(r_i, \tilde{\pi}_i), \text{logit}(\tilde{\pi}_i) = \mu + x_i\beta + b_i. \quad (7)$$

For heterozygous individual i , the corresponding model is as follows:

$$y_{il} \sim \text{Bin}(r_{il}, \pi_{il}), \text{logit}(\pi_{il}) = \mu + x_{il}\beta + b_{il}, l = 1, 2. \quad (8)$$

We treat the data for a homozygous individual, (y_i, r_i) , as one observation. We treat the data for a heterozygous individual $(y_{i1}, r_{i1}, y_{i2}, r_{i2})$ as two observations $((y_{i1}, r_{i1}), (y_{i2}, r_{i2}))$. Given n_1 homozygous individuals (both types of homozygotes) and n_2 heterozygous individuals, we have $n' = n_1 + 2n_2$ observations in total. To facilitate algebra derivation, we ordered observations by listing homozygous observations first, followed by heterozygous observations. The first and second moments in the likelihood functions defined in equations (7) and (8) can be written in vector and matrix forms as:

$$E(g(\boldsymbol{\pi})) = \mu + \boldsymbol{x}\boldsymbol{\beta},$$

$$V(g(\boldsymbol{\pi})) = \sigma_g^2 \begin{pmatrix} \mathbf{K}_{n_1 n_1} & \mathbf{K}_{n_1 n_2} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ \mathbf{K}_{n_2 n_1} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \mathbf{K}_{n_2 n_2} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{pmatrix} + \sigma_u^2 \begin{pmatrix} \mathbf{I}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{I}_{n_2 \times n_2} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{pmatrix} + \sigma_e^2 \begin{pmatrix} \frac{1}{2} \mathbf{I}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{I}_{n_2 \times n_2} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{pmatrix}, \quad (9)$$

where both $g(\boldsymbol{\pi})$ and \boldsymbol{x} are n' dimensional vectors by stacking the corresponding values from homozygous observations and heterozygous observations; $\mathbf{K}_{n_1 n_1}$, $\mathbf{K}_{n_2 n_2}$, and $\mathbf{K}_{n_1 n_2}$ are the corresponding kinship matrices computed among homozygous observations, among heterozygous

observations, and between homozygous and heterozygous observations, respectively; and \otimes denotes a Kronecker product.

We denote $\mathbf{b} = \mathbf{g} + \mathbf{u} + \mathbf{e}$, as the n' -vector of random effects. Our likelihood for all observations requires integrating out the random effects and thus is in a form of an n -dimensional integration

$$L = \int \prod_{i=1}^{n'} f(y_i | \mu, \beta, \mathbf{b}) f(\mathbf{b} | \sigma_g^2, \sigma_e^2, \rho) d\mathbf{b}.$$

We develop an algorithm based on the penalized quasiliquelihood (PQL) [1] for parameter inference for IMAGE. Briefly, PQL employs an iterative numerical optimization procedure. In each iteration, we introduce a set of pseudo-data $\tilde{\mathbf{y}}$ to replace the originally observed count data \mathbf{y} . The pseudo-data $\tilde{\mathbf{y}}$ is obtained based on a second order Taylor expansion using the conditional distribution $P(y_i | \mu, \beta, \mathbf{b})$ using the first and second order moments $E(y_i | \mu, \beta, \mathbf{b})$ and $V(y_i | \mu, \beta, \mathbf{b})$, both evaluated at the current estimates of the fixed coefficients as well as the random effects \mathbf{b} . With the pseudo-data, the complex likelihood function for the original data \mathbf{y} is replaced by a much simpler linear mixed model (LMM) likelihood function for the pseudo-data $\tilde{\mathbf{y}}$, thereby alleviating much of the computational burden associated with our count model. With pseudo-data $\tilde{\mathbf{y}}$, we can perform inference and update parameters using the standard average information (AI) algorithm for LMMs [2-4]. By iterating between the approximation step of obtaining the pseudo-data $\tilde{\mathbf{y}}$ and the inference step of updating the parameter estimates via the AI algorithm, PQL allows us to perform inference in a computationally efficient fashion. Below, we describe the estimation and inference procedure in detail.

The observations (y_i, r_i) ($i = 1, 2, \dots, n'$; $n' = n_1 + 2n_2$; note our slight abuse of notation here) are independent conditional on the unobserved random effects \mathbf{b} and the fixed effects $\mu + \mathbf{x}\beta$, with conditional mean $E(y_i | \mu, \beta, \mathbf{b}) = \pi_i r_i = g^{-1}(\mu + x_i \beta + b_i)$ and conditional variance $V(y_i | \mu, \beta, \mathbf{b}) = v(\pi_i) = \pi_i(1 - \pi_i)r_i$. We use these two conditional moments to obtain the quasi-likelihood for i th observation as $ql_i(\mu, \beta | \mathbf{b}) = \int_{y_i}^{\pi_i r_i} \frac{y_i - t}{v(t)} dt$, which serve as an approximation for the conditional likelihood. The joint likelihood function can be approximated by the joint quasi-likelihood function

$$ql(\mu, \beta, \sigma_g^2, \sigma_u^2, \sigma_e^2) = \log \int \left(\prod_{i=1}^n ql_i(\mathbf{b}, \mu, \beta) \right) P(\mathbf{b} | \sigma_g^2, \sigma_u^2, \sigma_e^2) d\mathbf{b}.$$

We use Laplace approximation to further approximate the above function and obtain:

$$\tilde{ql}(\mu, \beta, \sigma_g^2, \sigma_u^2, \sigma_e^2) = \frac{1}{2} \log |\mathbf{V}\mathbf{D} + \mathbf{I}| + \sum_{i=1}^n ql_i(\tilde{\mathbf{b}}, \mu, \beta) - \frac{1}{2} \tilde{\mathbf{b}}^T \mathbf{V}^{-1} \tilde{\mathbf{b}}, \quad (10)$$

where $\mathbf{V} = \sigma_g^2 \mathbf{V}_1 + \sigma_u^2 \mathbf{V}_2 + \sigma_e^2 \mathbf{V}_3$, $\tilde{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmax}} \left(\sum_{i=1}^n ql_i(\mu, \beta | \mathbf{b}) - \frac{1}{2} \mathbf{b}^T \mathbf{V}^{-1} \mathbf{b} \right)$ and $\mathbf{D} = \operatorname{diag}(1/g'(\lambda_i))$ is a diagonal weight matrix. $\mathbf{V}_1 = \begin{pmatrix} K_{n_1 n_1} & K_{n_1 n_2} \otimes [1 \ 1] \\ K_{n_2 n_1} \otimes [1] & K_{n_2 n_2} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{pmatrix}$, $\mathbf{V}_2 = \begin{pmatrix} \mathbf{I}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{I}_{n_2 \times n_2} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{pmatrix}$, $\mathbf{V}_3 = \begin{pmatrix} \frac{1}{2} \mathbf{I}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{I}_{n_2 \times n_2} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{pmatrix}$.

We treat the approximated quasi-likelihood function \tilde{ql} in equation (10) as the target function. And we obtain estimates for (\mathbf{b}, μ, β) and $(\sigma_g^2, \sigma_u^2, \sigma_e^2)$ alternately from equation (10).

Specifically, we first obtain estimates for (\mathbf{b}, μ, β) conditional on the current estimates of $(\sigma_g^2, \sigma_u^2, \sigma_e^2)$. To do so, following [1], we assume that the iterative weights vary slowly with respect to the conditional mean; that is

$$\frac{\partial \mathbf{D}}{\partial \lambda_i} \approx 0.$$

We then obtain the first order derivatives with respect to either (μ, β) or \mathbf{b} , and set the two first order derivatives to zero; that is

$$(\mathbf{1}, \mathbf{x})^T \mathbf{D} \Delta (\mathbf{y} - \lambda \mathbf{r}) = \mathbf{0}, \quad (11)$$

$$\mathbf{D} \Delta (\mathbf{y} - \lambda \mathbf{r}) - \mathbf{V}^{-1} \mathbf{b} = \mathbf{0}, \quad (12)$$

where $\lambda = (\lambda_1, \dots, \lambda_{n_1+2n_2})$ and $\Delta = \text{diag}\{g'(\lambda_i)\}$.

We now define the pseudo-data

$$\tilde{y}_i = \eta_i + g'(\lambda_i)(y_i - \lambda_i r_i), \quad (13)$$

and our equation (11) becomes

$$\mathbf{b} = \mathbf{V} \mathbf{H}^{-1} \left[\tilde{\mathbf{y}} - (\mathbf{1}, \mathbf{x}) \begin{pmatrix} \mu \\ \beta \end{pmatrix} \right], \quad (14)$$

where $\mathbf{H} = \mathbf{D}^{-1} + \mathbf{V}$. Substituting equation (14) into equation (12), we can obtain the estimates

$$\begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} = [(\mathbf{1}, \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{1}, \mathbf{x})]^{-1} (\mathbf{1}, \mathbf{x})^T \mathbf{H}^{-1} \tilde{\mathbf{y}}, \quad (15)$$

and

$$\hat{\mathbf{b}} = \mathbf{V} \mathbf{H}^{-1} \left[\tilde{\mathbf{y}} - (\mathbf{1}, \mathbf{x}) \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} \right], \quad (16)$$

Both are conditional on the variance component estimates $(\sigma_g^2, \sigma_u^2, \sigma_e^2)$.

Next, we obtain estimates for the variance components $(\sigma_g^2, \sigma_u^2, \sigma_e^2)$ conditional on the current estimates of (\mathbf{b}, μ, β) . Here we first define the variance components $\tau_1 = \sigma_g^2$, $\tau_2 = \sigma_u^2$ and $\tau_3 = \sigma_e^2$. We then integrate out the fixed effects μ and β in equation (10) to obtain the restricted likelihood function as

$$\tilde{q}l_R(\tau_1, \tau_2, \tau_3) = c_R - \frac{1}{2} \log |\mathbf{H}| - \frac{1}{2} \log |(\mathbf{1}, \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{1}, \mathbf{x})| - \frac{1}{2} \tilde{\mathbf{y}}^T \mathbf{P} \tilde{\mathbf{y}},$$

where $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1} (\mathbf{1}, \mathbf{x})^T ((\mathbf{1}, \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{1}, \mathbf{x}))^{-1} (\mathbf{1}, \mathbf{x}) \mathbf{H}^{-1}$, and c_R is a constant. We use the AI algorithm to obtain variance component estimates. In particular, we obtain the first derivatives as

$$\frac{\partial \tilde{q}l_R(\sigma_g^2, \sigma_u^2, \sigma_e^2)}{\partial \tau_i} = \frac{1}{2} \{ \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_i \mathbf{P} \tilde{\mathbf{y}} - \text{tr}(\mathbf{P} \mathbf{V}_i) \}, i \in \{1, 2, 3\},$$

and the second derivatives as

$$\frac{\partial^2 \tilde{q}l_R(\sigma_g^2, \sigma_u^2, \sigma_e^2)}{\partial \tau_i \partial \tau_j} = \frac{1}{2} \text{tr}(\mathbf{V}_i \mathbf{P} \mathbf{V}_j \mathbf{P}) - \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_i \mathbf{P} \mathbf{V}_j \mathbf{P} \tilde{\mathbf{y}}; i, j \in \{1, 2, 3\}$$

The second derivatives constitute the observed information matrix. Because the elements in the expected information matrix are

$$E \left[\frac{\partial^2 \tilde{q}l_R(\sigma_g^2, \sigma_u^2, \sigma_e^2)}{\partial \tau_i \partial \tau_j} \right] = -\frac{1}{2} \text{tr}(\mathbf{V}_i \mathbf{P} \mathbf{V}_j \mathbf{P}); i, j \in \{1, 2, 3\},$$

we can obtain the average information (AI) matrix as an average of the above two matrices; that is

$$\mathbf{AI} = \frac{1}{2} \begin{bmatrix} \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_1 \mathbf{P} \mathbf{V}_1 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_1 \mathbf{P} \mathbf{V}_2 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_1 \mathbf{P} \mathbf{V}_3 \mathbf{P} \tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_2 \mathbf{P} \mathbf{V}_1 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_2 \mathbf{P} \mathbf{V}_2 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_2 \mathbf{P} \mathbf{V}_3 \mathbf{P} \tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_3 \mathbf{P} \mathbf{V}_1 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_3 \mathbf{P} \mathbf{V}_2 \mathbf{P} \tilde{\mathbf{y}} & \tilde{\mathbf{y}}^T \mathbf{P} \mathbf{V}_3 \mathbf{P} \mathbf{V}_3 \mathbf{P} \tilde{\mathbf{y}} \end{bmatrix}.$$

With the first and second order derivatives, we can perform Newton-Raphson update with the AI algorithm and obtain estimates for $(\sigma_g^2, \sigma_u^2, \sigma_e^2)$.

As a summary, we implement the PQL algorithm that consists of the following steps:

1. Initialize the parameters, $\mu^{(0)}, \beta^{(0)}, \boldsymbol{\tau}^{(0)} = (\tau_1^{(0)}, \tau_2^{(0)}, \tau_3^{(0)})^T$, and obtain the pseudo-data $\tilde{\mathbf{y}}^{(0)}$ as in equation (12). Set $t = 1$.
2. Update $\boldsymbol{\tau}^{(t)} = \boldsymbol{\tau}^{(t-1)} + \mathbf{AI}^{-1} \left(\frac{\partial q_{LR}(\sigma_g^2, \sigma_u^2, \sigma_e^2)}{\partial \boldsymbol{\tau}} \right)$;
3. Update $\mu^{(t)}, \beta^{(t)}$ and $\mathbf{b}^{(t)}$ with $\boldsymbol{\tau}^{(t)}$ and $\tilde{\mathbf{y}}^{(t-1)}$ as in equations (15) and (16);
4. Update $\tilde{\mathbf{y}}^{(t)}$ using the $\mu^{(t)}, \beta^{(t)}$ and $\mathbf{b}^{(t)}$ as in equation (13);
5. Set $t = t + 1$, and repeat steps 2-4 until convergence.

Once we obtain parameter estimates $(\mu, \beta, \sigma_g^2, \sigma_u^2, \sigma_e^2)$, we can construct the Wald test based on

$$\hat{\beta} = (\mathbf{x}^T \mathbf{P}_c \mathbf{x})^{-1} \mathbf{x}^T \mathbf{P}_c \tilde{\mathbf{y}} \text{ and } \text{Var}(\hat{\beta}) = (\mathbf{x}^T \mathbf{P}_c \mathbf{x})^{-1},$$

where $\mathbf{P}_c = \mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{W}^T (\mathbf{W}^T \mathbf{H}^{-1} \mathbf{W}^T)^{-1} \mathbf{W}^T \mathbf{H}^{-1}$.

Because of the approximation using PQL, when sample size is small (<150), we observe an inflation of p-values in the simulations and real data applications (e.g. the genomic control factor equals 1.21 and 1.12 in the two real data sets). We therefore perform an additional step of genomic control to adjust for p-values in a *post hoc* fashion. Specifically, we normalize χ^2 statistics (across all SNP-CpG pairs) by dividing $\max(1, \lambda_{gc})$ and then use the normalized χ^2 statistics to obtain the final p-values.

3. M-values Calculation

For methods that use normalized data, we performed “M” value transformation. To do so, we first divided the number of methylated reads by the number of unmethylated reads. Afterwards, we performed a log2-transformation to obtain the final normalized data. The final normalized data is in the form of $\log_2 \left(\frac{\text{methylated reads} + a}{\text{unmethylated reads} + a} \right)$, where a small value $a = 0.01$ is added to avoid zero values inside the log function.

REFERENCES

1. Breslow, N.E. and D.G. Clayton, *Approximate Inference in Generalized Linear Mixed Models*. Journal of the American Statistical Association, 1993. **88**(421): p. 9-25.
2. Chen, H., et al., *Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models*. American Journal of Human Genetics, 2016. **98**(4): p. 653-666.
3. Gilmour, A.R., R. Thompson, and B.R. Cullis, *Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models*. Biometrics, 1995. **51**(4): p. 1440-1450.
4. Yang, J.A., et al., *GCTA: A Tool for Genome-wide Complex Trait Analysis*. American Journal of Human Genetics, 2011. **88**(1): p. 76-82.