

(A)

Conventional model:

$$M = \sum_{1Mb} b_j SNP_j + \epsilon$$

Incorporate **local ancestry** (LA) by dissecting the original genotype SNP_j by AFR and EUR ancestry

$$\begin{aligned} SNP_j &= SNP_{j,AFR} + SNP_{j,EUR} \\ &= SNP_{j1}I(LA_{j1} = AFR) + SNP_{j2}I(LA_{j2} = AFR) \\ &\quad + SNP_{j1}I(LA_{j1} = EUR) + SNP_{j2}I(LA_{j2} = EUR) \end{aligned}$$

Basic LA model:

$$M = \sum_{1Mb} (b_{j,AFR}SNP_{j,AFR} + b_{j,EUR}SNP_{j,EUR}) + \epsilon$$

Preselection:
determine if the j th SNP has LA-specific effects and belongs to the *LA set* by comparing the ancestral model with the null model

	CpG _m	
	Methylated count	Unmethylated count
id ₁	9	1
id ₂	900	100
id ₃	70	30
...		
id _N	80	20

Input

LA Methylation Predictor with Preselection step (LAMPP) :

$$\begin{aligned} M &= \sum_{j \notin LA \text{ set}} b_j SNP_j + \sum_{j \in LA \text{ set}} (b_{j,AFR}SNP_{j,AFR} + b_{j,EUR}SNP_{j,EUR}) + \epsilon \\ &= \sum_{j \notin LA \text{ set}} b_j SNP_j + \sum_{j \in LA \text{ set}} (b_{j,average} \cdot SNP_j + b_{j,diff} \frac{SNP_{j,AFR} - SNP_{j,EUR}}{2}) + \epsilon \end{aligned}$$

Regularized logistic regression

Beta coefficients to predict DNA methylation levels for CpG_m

(B) Preselection

Null model

$$M \sim b_j SNP_j$$

Ancestry model

$$\begin{aligned} M &\sim b_{j,AFR}SNP_{j,AFR} + b_{j,EUR}SNP_{j,EUR} \\ &\sim \frac{b_{j,AFR} + b_{j,EUR}}{2} (SNP_{j,AFR} + SNP_{j,EUR}) + \frac{b_{j,AFR} - b_{j,EUR}}{2} (SNP_{j,AFR} - SNP_{j,EUR}) \\ &\sim b_{j,average} \cdot SNP_j + b_{j,diff} \frac{(SNP_{j,AFR} - SNP_{j,EUR})}{2} \end{aligned}$$