Insights into fine-mapping causal variants of complex traits from diverse populations

Masahiro Kanai

PhD student working with Mark Daly and Hilary Finucane

HARVARD MEDICAL SCHOOL

BROAD INSTITUTE

MASSACHUSETTS GENERAL HOSPITAL

大阪大学 OSAKA UNIVERSITY

SuSiE on meta-analysis

- Original model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$$

$$\boldsymbol{e} \sim N_n(0, \sigma^2 I_n)$$

$$\boldsymbol{b} = \sum_{l=1}^{L} \boldsymbol{b}_l$$

$$\boldsymbol{b}_l = \boldsymbol{\gamma}_l b_l$$

$$\boldsymbol{\gamma}_l \sim \text{Mult}(1, \boldsymbol{\pi})$$

$$b_l \sim N_1(0, \sigma_{0l}^2).$$

- Sufficient summary statistics for SuSiE:

$$\hat{\beta},\ \text{SE}(\hat{\beta}),\ R = X^T X/n,\ \text{Var}(y),\ \text{and}\ n$$

OR

$$X^T X, X^T y, y^T y,\ \text{and}\ n$$

Inverse-variance weighted fixed-effect meta-analysis

For jth SNP:

$$\langle \widehat{\beta_j} \rangle = \frac{\sum_i \widehat{\beta_{ij}}/\mathrm{Var}(\widehat{\beta_{ij}})}{\sum_i 1/\mathrm{Var}(\widehat{\beta_{ij}})}, \langle \mathrm{SE}(\widehat{\beta_j}) \rangle = \sqrt{\frac{1}{\sum_i 1/\mathrm{Var}(\widehat{\beta_{ij}})}}$$

Setup:

SNP $m_1 = 1$

$$X = \begin{matrix} X_1 \\ X_2 \end{matrix} \begin{matrix} \text{Pop1 } n_1 \\ \text{Pop2 } n_2 \end{matrix}$$

If SNP is Pop1-specific, then $X_2 = 0$

Then:

$$\hat{\beta} = (X^T X)^{-1} X^T y = \left(\left({X_1}^T {X_2}^T \right) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right)^{-1} \left({X_1}^T {X_2}^T \right) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{{X_1}^T y_1 + {X_2}^T y_2}{{X_1}^T X_1 + {X_2}^T X_2}$$

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} = \sigma^2 \left({X_1}^T X_1 + {X_2}^T X_2 \right)^{-1}$$

Weighted average of LD matrix?

Setup:

SNP1 m_1 SNP2 m_2

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

Pop1 n_1

Pop2 n_2

If SNP2 is Pop1-specific, then $D = 0$

Given,

$$X_1{}^T X_1 = \begin{pmatrix} A^T \\ B^T \end{pmatrix} (A \quad B) = \begin{pmatrix} A^T A & A^T B \\ B^T A & B^T B \end{pmatrix}, \; R_1 = \frac{X_1{}^T X_1}{n_1}$$

$$X_2{}^T X_2 = \begin{pmatrix} C^T \\ D^T \end{pmatrix} (C \quad D) = \begin{pmatrix} C^T C & 0 \\ 0 & 0 \end{pmatrix}, \; R_2 = \frac{X_2{}^T X_2}{n_2}$$

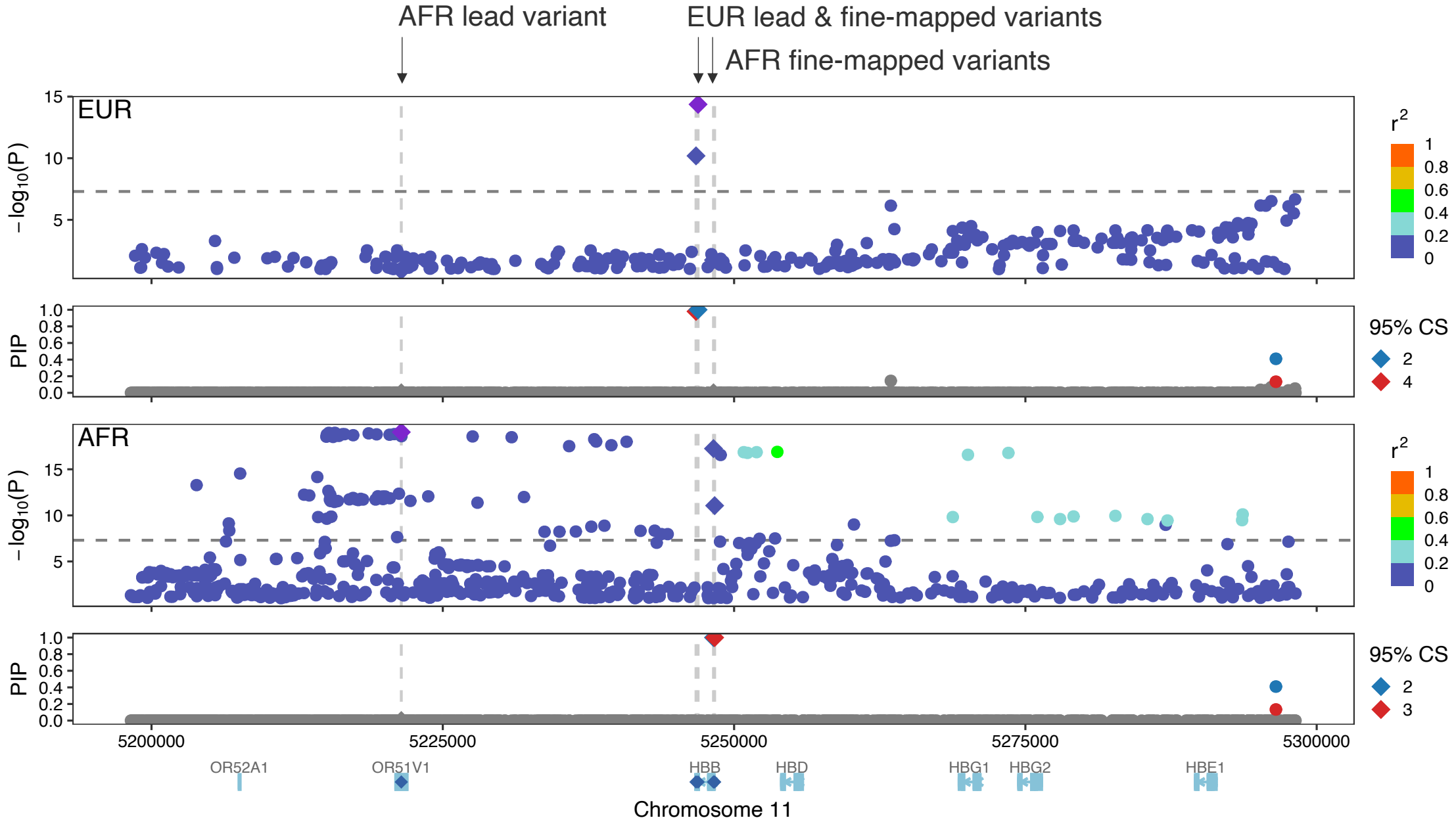
Then

$$X^T X = \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A^T A + C^T C & A^T B + C^T D \\ B^T A + D^T C & B^T B + D^T D \end{pmatrix} = \begin{pmatrix} A^T A + C^T C & A^T B \\ B^T A & B^T B \end{pmatrix},$$

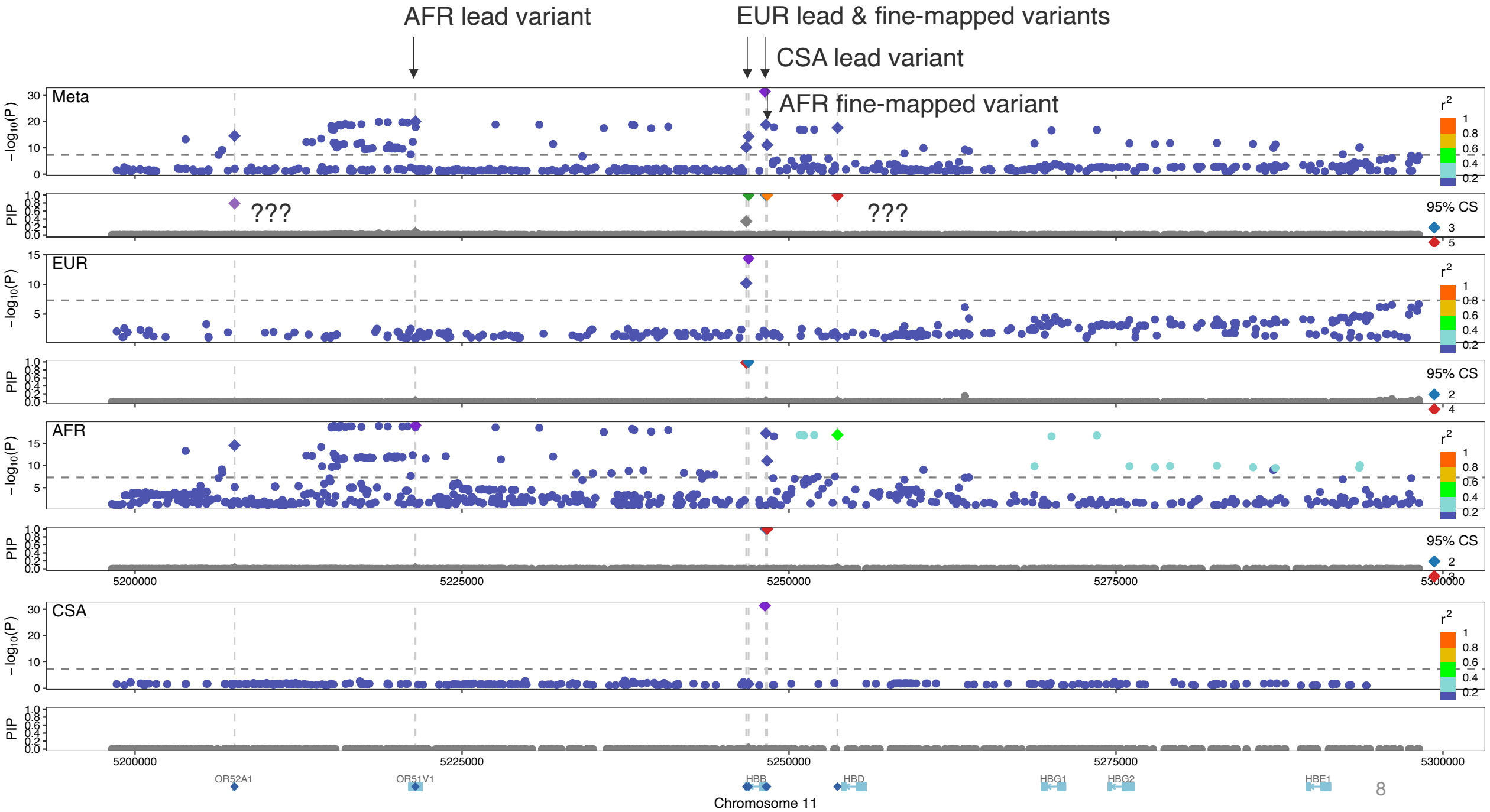
Thus:

$$R = \frac{X^T X}{n} = \frac{X_1{}^T X_1 + X_2{}^T X_2}{n_1 + n_2} = \frac{n_1 R_1 + n_2 R_2}{n_1 + n_2}$$

HBB locus of mean corpuscular volume (MCV) GWAS



HBB, a known pathogenic gene for sickle cell disease


AFR lead variant
EUR lead & fine-mapped variants
CSA lead variant
AFR fine-mapped variant
Meta
EUR
AFR
CSA
−log10(P)
PIP
???
???
95% CS
r2
5200000
5225000
5250000
5275000
5300000
OR52A1
OR51V1
HBB
HBD
HBG1
HBG2
HBE1
Chromosome 11

What breaks meta-analysis fine-mapping?

- Heterogeneity (assumption: true β is same across cohorts)
  - True heterogeneity
    - Environment (model: $y_k = X_k\beta + \varepsilon_k$ where $\varepsilon_k \sim N(0, \sigma^2 I)$)
  - Technical artifacts
    - E.g., phenotyping/genotyping/imputation error leads to systematic heterogeneity

- Missing causal variants
  - I think population-specific variants are fine… but…?