Simulation models of gene with time-specific expression

We first modeled the age-dependent expression of a given gene. Its mean expression level, $f(x)$, as a function of age x, is described by a Gaussian function:

$$f(x) = A\exp\left(-\frac{(x-c)^2}{d}\right)$$

where A is the maximum expression amplitude, c is the age of peak expression, and d is a scale parameter controlling the width of the expression profile. This function approximates biological scenarios where gene activity is highest at a specific developmental stage.

We then assume that part of the genes with age-dependent expression are disease susceptibility genes. Each gene has at most k independent (LD r^2<0.01) disease susceptibility loci (sampling without replacement). These loci (m in total) collectively explain h_g^2 liability to a disease. An individual's liability for a disease, L, is a linear combination of the m loci weighted by w, plus an additional error term ϵ:

$$L = \sum_{t=1}^{m} w_t\, g_t + \epsilon$$

where $g_i \in \{0,1,2\}$ is the encoded genotype and the term ϵ represents environmental and unmodeled genetic factors, following a normal distribution $\epsilon \sim N(0, \sigma_\epsilon^2)$. Its variance, σ_ϵ^2, is set to ensure that the total genetic component explains a target heritability of liability (h_g^2), such that $\sigma_\epsilon^2 = \mathrm{Var}(\sum_{t=1}^{m} w_t\, g_t)\left(\frac{1}{h_g^2} - 1\right)$. For simplicity, all the weights w_t were set to 1.

Disease status was determined using the liability-threshold model[48]. Given a population prevalence K, individuals with a standardized liability L_x exceeding the corresponding threshold z (where $z = \Phi^{-1}(1-K)$ and Φ^{-1} is the inverse normal cumulative distribution function) were classified as cases.