Mapping Quantitative Traits

Hsiu-Khuern Tang
Hewlett Packard
1501 Page Mill Road, MS 1U-2
Palo Alto, California, USA
hktang@stanford.alumni.org

David Siegmund
Stanford University, Department of Statistics
390 Serra Mall
Stanford, California, USA
dos@stat.stanford.edu

Our simplest model for the phenotype Y influenced by a quantitative trait locus (QTL) at τ is

$$Y = \mu + \alpha_x + \alpha_y + \delta_{x,y} + e,$$

where α_a denotes the additive effect of allele a at τ, $\delta_{a,b}$ the dominance effect of alleles a and b, and e incorporates both environmental and residual genetic effects. In more general models we also allow gene-gene and gene-environment interactions, but we assume e is uncorrelated with the genetic and environmental effects that are explicitly modeled. Both Hardy-Weinberg and linkage equilibrium are assumed. Components of variance are $\sigma_A^2 = 2E\alpha_x^2$, $\sigma_D^2 = E\delta_{x,y}^2$, $\sigma_e^2 = Ee^2$ and $\sigma_Y^2 = \sigma_A^2 + \sigma_D^2 + \sigma_e^2$. Given two sibs with phenotypes Y_1 and Y_2, we also let $\rho = \mathrm{Corr}(Y_1, Y_2) = [\sigma_A^2/2 + \sigma_D^2/4 + r\sigma_e^2]/\sigma_Y^2$, where r is the residual genetic- environmental correlation of siblings. Important "segregation" parameters, which can be estimated without recourse to genetic data, are μ, σ_Y^2 and ρ.

Let $\nu = \nu(\tau)$ denote the number of alleles shared identical by descent by two siblings at the trait locus τ, and let

$$\rho_\nu = \mathrm{Corr}(Y_1, Y_2|\nu) = \rho + (\nu - 1)\alpha_0/\sigma_Y^2 + (1/2 - 1_{\{\nu=1\}})\delta_0/\sigma_Y^2,$$

where $\alpha_0 = \sigma_A^2/2 + \sigma_D^2/4$ and $\delta_0 = \sigma_D^2/4$. The parameters α_0 and δ_0 are the "linkage" parameters of interest, while the segregation parameters are nuisance parameters.

We consider N sibships of size s each and let $\mathbf{Y}$ denote the vector of phenotypes in the nth sibship. For simplicity we take $s = 2$. Under a normality assumption the components of the efficient score at τ are

$$\ell_\alpha(\tau) = \sum_{n=1}^{N} (\nu - 1)C, \quad \ell_\delta(\tau) = \sum (1/2 - 1_{\{\nu=1\}})C,$$

where $C = C_n(\alpha_0, \delta_0, \rho, \sigma_Y^2, \nu, \mathbf{Y})$. Under the hypothesis $\alpha_0 = 0$, the linkage parameters α_0, δ_0 are orthogonal to each other and to the segregation parameters μ, σ_Y, ρ, so the score statistic for α_0 at the putative QTL t is $Z_1(t) = \ell_\alpha(t)/I_{\alpha\alpha}^{1/2}$. Here $I_{\alpha\alpha}$ is the Fisher information for α_0 at $\alpha_0 = 0$, and unknown parameters in C have been replaced by their maximum likelihood estimates under $\alpha_0 = 0$. Since the true value of τ is unknown, we use $\max_t Z_1(t)$, where the maximum extends over all marker loci t.

Remarks. (i) To obtain a statistic that is nonparametric (under the hypothesis $\alpha_0 = 0$) and is fully efficient under the normality assumption, we can evaluate the Type I error probability conditional on the phenotypes. For large values of N this is equivalent to using the statistic $\max_t \ell_\alpha(t)/[\sum C^2/2]^{1/2}$.

(ii) One can also obtain a two dimensional statistic $(Z_1(t), Z_2(t))$ based on ℓ_α and ℓ_δ if it is thought that dominance may be important. Since $0 \leq \delta_0 \leq \alpha_0$, the second coordinate Z_2 often fails to compensate for the higher significance threshold required for the two degree of freedom statistic. For models incorporating gene-gene and gene-environment interaction there are additional variance components, and the corresponding additional degrees of freedom play an essential role.

(iii) For a sibship of size s the square of the asymptotic expectation of $Z_1(\tau)$ is

$$N \frac{\alpha_0^2}{2\sigma_Y^4} \binom{s}{2} \frac{\{[1 + (s-2)\rho]^2 + \rho^2\}}{\{(1-\rho)[1 + (s-1)\rho]\}^2}.$$

As this expression suggests, the power to detect linkage increases very rapidly with s and can be as large as the power of $\binom{s}{2}$ sib pairs. For traits involving gene-gene interactions, we find that the parameter α_0, which depends on σ_D^2 in the case of a single gene, also depends on interaction variance components. A statistic that ignores the interaction can be almost as powerful as a statistic that properly accounts for the interaction. However, for traits involving gene-environment interactions, we find that correctly accounting for the interaction can result in a substantial increase in power over a procedure that ignores the interaction.

REFERENCES

Tang, H.-K. and Siegmund, D. (2001). Mapping quantitative trait loci in oligogenic models, *Biostatistics*, **2**

Tang, H.-K. (2000). Using variance components to map quantitative trait loci in humans, Ph. D. thesis, Stanford University.

RESUMÉ

Nous développons une théorie asymptotique pour la localisation de traits quantitatifs sur le génome humain et étudions les interactions entre génes et entre les génes et l'environnement.