File S1

Supporting Information

MATÉRN COVARIANCE FUNCTION

The Matérn covariance function is a family of covariance functions widely used for simulating and studying Gaussian processes (BANERJEE ET AL., 2004). The covariance between two points x and y is defined as $\mathrm{var}(x, y) = C(\|x - y\|)$, where $\|\|$ denotes a distance function (usually Euclidean norm), and

$$C(t) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}t\phi)^{\nu}K_{\nu}(2\sqrt{\nu}t\phi), \;\; t > 0,$$

where $K_{\nu}(\cdot)$ is the modified Bessel function. The σ^2 parameter is the amplitude parameter that controls the variance, ϕ is the scale parameter that controls the span of dependence (in space or time), and ν is a smoothness parameter that controls how rough the resulting error process is. See BANERJEE ET AL. (2004) for details, and PATIL (2010) for examples and interpretation.

SIMULATION STUDIES: COMPARISON WITH NON-FUNCTIONAL APPROACHES

We present the simulation studies of Type-I error and power of our Wald statistic under a Gaussian process noise. For power, we assessed the effect of both sample size and the number of time points.

Model We used a functional linear model $y(t) = z\beta(t) + \epsilon(t)$, where the design matrix z was random genotypes encoded as 0's and 1's and $\beta(t)$ is a genetic effect function; we only simulated dominant effects (one allele out of two is dominant).

Type-I error In order to assess Type-I error, we simulated data from a linear model under the null hypothesis. Since under the null hypothesis there is no genetic effect, $\beta(t)$ is identically zero. We assumed one genetic locus of three genotypes with probabilities 0.25, 0.5, 0.25. The random processes were sampled at 20 evenly spaced time points and there are 5000 runs for each of four sample sizes, 300, 400, 500, and 600. The results are in Table 1. We can see that proportions at cutoffs agree with the theoretical values. This confirms the use of χ^2 distribution and degrees of freedom as a valid reference distribution for the Wald statistic.

Power To evaluate the performance of the functional linear models for identifying QTLs, we compared their power with that of the traditional cross-sectional models for QTLs. We considered a single trait locus, and the frequency of two genotypes at the trait locus were assumed to be equal. The genetic model used the functional linear model mentioned above. The power is the number of times the p-values are over the significance level of 0.05. We used three functions as genetic effect functions and the random process was generated with zero mean and Matérn covariance functions (as in the Type-I error simulations above). The three functions were

1. Quadratic function: $\beta(t) = 2.5 + \frac{t}{10} + \frac{t^2}{1000}$
2. Exponential function: $\beta(t) = 1 - \frac{1}{10}\exp(-\frac{5t}{1000})$
3. Logistic function: $\beta(t) = \frac{1}{1+\exp(-t)}$

A total of 1,000 simulations were conducted. The cross-sectional method averaged the trait over all time points: $\frac{1}{m}\sum_{j=1}^{m} y(t_j)$. Our functional method used B-spline basis functions of order 4 with 2 knots, and used the Wald test statistic. We computed power either as a function of the number of time points, where 400 subjects were sampled, or as a function of sample sizes where 5, 6 and 7 time points were assumed for exponential, logistic and quadratic effect functions, respectively. The functions were simulated over intervals $[-50, -38]$, $[-460, -316]$, and $[-6, 2]$, respectively, for the three functions. The powers curves are in Figure 1. Several features emerge: First, power increased with the number of time points. Second, in general, the functional linear models had higher power to detect a QTL than the cross-sectional approach, sometimes dramatically so. Third, difference in power between the functional approach and cross-sectional approach depends on the types of genetic effect functions. We observed the largest difference in power between the functional linear models and cross-sectional models for the logistic genetic effect function.

UNSTRUCTURED COVARIANCE MATRIX

The "unstructured" covariance matrix, Σ_3 used by YAP ET AL. (2009) and used in our simulations is given below.

$$\Sigma_3 = \begin{bmatrix} 0.72 & 0.39 & 0.45 & 0.48 & 0.50 & 0.53 & 0.60 & 0.64 & 0.68 & 0.68 \\ 0.39 & 1.06 & 1.61 & 1.60 & 1.50 & 1.48 & 1.55 & 1.47 & 1.35 & 1.29 \\ 0.45 & 1.61 & 3.29 & 3.29 & 3.17 & 3.09 & 3.19 & 3.04 & 2.78 & 2.53 \\ 0.48 & 1.60 & 3.29 & 3.98 & 4.07 & 4.01 & 4.17 & 4.18 & 4.00 & 3.69 \\ 0.50 & 1.50 & 3.17 & 4.07 & 4.70 & 4.68 & 4.66 & 4.78 & 4.70 & 4.36 \\ 0.53 & 1.48 & 3.09 & 4.07 & 4.68 & 5.56 & 6.23 & 6.87 & 7.11 & 6.92 \\ 0.60 & 1.55 & 3.19 & 4.17 & 4.66 & 6.23 & 8.59 & 10.16 & 10.80 & 10.70 \\ 0.64 & 1.47 & 3.04 & 4.18 & 4.78 & 6.87 & 10.16 & 12.74 & 13.80 & 13.80 \\ 0.68 & 1.35 & 2.78 & 4.00 & 4.70 & 7.11 & 10.80 & 13.80 & 15.33 & 15.35 \\ 0.68 & 1.29 & 2.53 & 3.69 & 4.36 & 6.92 & 10.70 & 13.80 & 15.35 & 15.77 \end{bmatrix}.$$

LITERATURE CITED

BANERJEE, S., B. CARLIN, AND A. GELFAND (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall.

PATIL, A. (2010) *PyMC Gaussian process module Users guide*.

YAP, J. S., J. FAN, AND R. WU (2009) Nonparametric Modeling of Longitudinal Covariance Structure in Functional Mapping of Quantitative Trait Loci. *Biometrics*, **65**(4):1068–1077.