ISSN 1022-7954, Russian Journal of Genetics, 2006, Vol. 42, No. 6, pp. 696–702. © Pleiades Publishing, Inc., 2006.
Original Russian Text © T.I. Axenovich, A.S. Zykovich, 2006, published in Genetika, 2006, Vol. 42, No. 6, pp. 850–857.

MATHEMATICAL MODELS AND METHODS

Power Estimation for in Silico Mapping

T. I. Axenovich[1] and A. S. Zykovich[2]

[1] *Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia; fax: (383) 333-12-78; e-mail: aks@bionet.nsc.ru*
[2] *Novosibirsk State University, Novosibirsk, 630090 Russia*

Received September 1, 2005

Abstract—A fundamentally new approach to gene mapping of complex traits was suggested recently. It consists in computer analysis of existing databases on the phenotypes and single nucleotide polymorphisms (SNPs) in inbred mouse strains and was termed in silico mapping. The power of this method has been studied by simulating quantitative traits controlled by one, two, or three genes. The results have demonstrated that the power of in silico mapping is high in the case of a monogenic trait. The probability of mapping all genes determining a digenic or, especially, trigenic trait is low. If two or three genes make equal phenotypic contributions to a trait, the proportions of experiments where none of them is localized are 17 and 25%, respectively. In the case of a major gene effect, when the phenotypic contribution of one gene considerably exceeds those of the other genes, the probability to map the major gene is 0.95 and 0.80 for the digenic and trigenic models, respectively. This shows that, in the case of polygenic control, the new method could localize only the genes with major effects, while most genes involved in the control of the trait would not be mapped.

DOI: 10.1134/S1022795406060160

INTRODUCTION

In 2001, a fundamentally new approach to gene mapping of complex traits was suggested. It consists in computer analysis of existing databases on the phenotypes and single nucleotide polymorphisms (SNPs) in inbred mouse strains [1]. The method was termed in silico mapping. It is based on the assumption that the larger the phenotypic difference between two inbred strains, the larger the differences in the nucleotide sequences of the loci controlling this phenotype. In silico mapping seems to be a convenient, rapid, and inexpensive method of gene mapping, because, in contrast to linkage analysis, it does not entail special hybridization experiments with genotyping the heterogeneous offspring. It allows the duration of analysis to be shortened to several minutes. In addition, linkage analysis deals with two contrasting mouse strains at a time, whereas in silico mapping simultaneously analyzes a large number of strains. Theoretically, this makes it possible to identify more genes involved in the control of a complex trait compared to linkage analysis.

The first results obtained by in silico mapping were promising. The localization of genes for some monogenic traits was confirmed [1, 2]. For example, Pletcher et al. [2] used the new method to determine the probable loci for the genes of different coat colors in mice and demonstrate that these loci were the same where the *Pde6b, Tyr*, and *Asip* genes were mapped. Analysis of complex polygenic traits yielded similar results. Some of the loci where the genes were mapped by the new method contained the genes controlling the studied trait that had been mapped earlier using linkage analysis [2, 3]. Hillebrandt et al. [3] studied the genetics of liver fibrosis in mice. They used for mapping 109 SNP markers in six inbred mouse strains. It was found that the loci characterized by significant correlation between genotypic and phenotypic interstrain differences (significant loci) contained two genes that had been identified earlier by means of linkage analysis: *Hfib1* (hepatic fibrogenic gene 1) in chromosome 15 and *C5* (complement factor 5 gene) in chromosome 2.

On the other hand, it was found that the in silico method applied to analysis of monogenic traits often yielded false positive results, detecting several significant loci only one of which (usually, the locus that had the strongest effect) contained the known gene. Therefore, in silico mapping was suggested to be used for preliminary analysis allowing researchers to exclude the loci that are certainly not involved in the control of the trait [4].

Analysis of polygenic traits yielded similar results: only some of the loci detected by the new method contained genes described earlier. In the aforementioned study [3], six significant loci were found besides the two loci containing the *Hfib1* and *C5* genes. These loci, unconfirmed by linkage analysis, may be false positive. However, at least some of these loci may be involved in the control of the trait, and their effect was detected because in silico mapping is based on the analysis of many strains, whereas linkage analysis uses only two strains. Therefore, it is suggested that all significant loci detected by the new method should be regarded as candidate and should be studied in detail with the use of other mapping methods [5].

It remains unknown whether in silico mapping can be used for preliminary exclusion of some loci not involved in the control of the studied trait from further consideration [6]. To answer this question, it is necessary to estimate the power of the new method characterizing it ability to map genes involved in the control of the trait. Since complex traits are polygenic, and only a small proportion of them have been identified to date, the power of the method can be estimated only in a simulation experiment. Here, we estimated the power of in silico gene mapping of quantitative traits with various genetic determination modes.

MATERIALS AND METHODS

Mouse strains. We used the MPDsnps123 database of the distribution of SNP markers in inbred mouse strains (http://aretha.jax.org/pub-cgi/phenome/). We used information on 6550 SNP markers to analyze 16 strains: C57BL/6J, A/J, DBA/2J, 129X1/SvJ, C3H/HeJ, BALB/cByJ, AKR/J, NZB/B1NJ, NZW/LacJ, BALB/cJ, MRL/MpJ, A/HeJ, B10.D2-Hc<0>, LP/J, SM/J, and LG/J.

In silico mapping. A modified algorithm described earlier [1, 7] was used. It consists of the following main steps.

• Several (n) inbred strains with known SNP genotypes are selected from the database. In each strain, the value of the quantitative trait characterizing its phenotype is determined.

• All possible pairs of strains are formed; their number is $np = n(n-1)/2$.

• For each pair of strains (i and j), the phenotypic difference (x_{ij}) is determined.

• The genome is subdivided into overlapping regions 1000 nucleotides in length containing at least three SNP markers, and the proportion of differing alleles in the kth locus is determined for each pair of strains i and j). This proportion (g_{jik}) is the measure of genotypic differences between the ith and jth strains in the kth locus.

• For each kth locus, the coefficient of correlation between the phenotypic and genotypic differences is calculated:

$$C_k = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(x_{ij}-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{x_{ij}}{np}\right)\left(g_{jik}-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{g_{jik}}{np}\right)}{\left\{\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(x_{ij}-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{x_{ij}}{np}\right)^2\right]\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(g_{jik}-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{g_{jik}}{np}\right)^2\right]\right\}^{1/2}}.$$

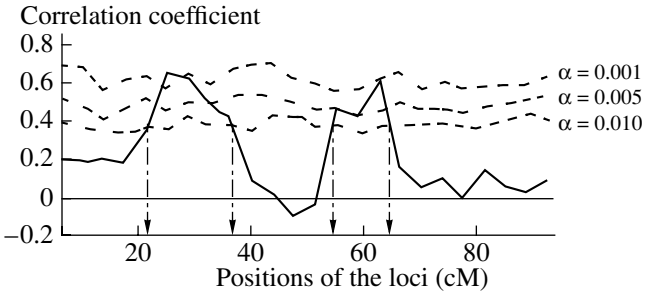
In this study, we formed 120 pairs out of 16 inbred strains. The genome was divided into 1635 loci.

To evaluate the contribution of each locus into the control of the trait we used an empirical threshold value. This approach, first suggested by Fisher (cited from [8]), is based on resampling. For this purpose, new samples were formed by means of permutations [9]. The essence of this technique is the following. It is assumed that the null hypothesis on the absence of linkage between the marker and the gene controlling the analyzed trait is true. Then, phenotypes and genotypes must be distributed between the strains independently. Therefore, random redistribution of quantitative traits between strains without changing their actual genotypes would suffice for resampling. We repeated this procedure 10 000 times, every time calculating the coefficients of correlation between the phenotypic and genotypic differences in each locus. This gave us 10 000 correlation coefficients for each locus. The boundary value of the correlation coefficient was chosen, separately for each locus, in such a way that it cut off the proportion of the highest correlation coefficients corresponding to the specified significance level. We used three significance levels corresponding to α values of 0.010, 0.005, and 0.001. If the correlation coefficient obtained for actual values of the trait exceeded the boundary value, then the null hypothesis was rejected for the given locus at the given significance level, and the locus was considered to be involved in the control of the trait studied. Figure illustrates the selection of the loci that make significant contribution in the control of a given trait.

Trait simulation. The in silico mapping method is based on the assumption that different loci have independent, additive effects on the formation of the trait. Since the differences between the values of the trait in pairs of strains, rather than the values themselves, are used for analysis, we assigned number 1 to strain C57BL/6J and assumed the value of the trait in this strain to be 0. The difference of this strain from all other strains ($i = 1$ and $j = 2, n$) with respect to the given trait was determined as

$$x_{1j} = \sum_{l=1}^{m}\mu_l g_{1jl},$$

where m is the number of loci controlling the trait and μ_l is the relative contribution of the lth locus to the control of the trait ($\sum\mu_l = 1$). The difference between strains i and j, where $i \neq 1$, was specified as $x_{ij} = |x_{1i} - x_{1j}|$.



The procedure for the selection of loci that make significant contributions to the control of a given trait. The solid line shows the coefficients of correlation between the phenotypic and genotypic differences for each analyzed locus. The dashed lines show the experimental threshold values for the significance levels $\alpha = 0.001$, 0.005, and 0.010. The arrows indicate the two loci that are the most likely to contain the genes to be identified ($\alpha = 0.010$).

We studied the monogenic, digenic, and trigenic models. For the digenic model, we considered the variants

$$\mu_1 = 0.5, \quad \mu_2 = 0.5;$$

$$\mu_1 = 0.83, \quad \mu_2 = 0.17;$$

$$\mu_1 = 0.91, \quad \mu_2 = 0.09;$$

for the trigenic model,

$$\mu_1 = 0.33, \quad \mu_2 = 0.33, \quad \mu_3 = 0.33;$$

$$\mu_1 = 0.625, \quad \mu_2 = 0.25, \quad \mu_3 = 0.125.$$

We selected for simulation 24 loci in which the studied strains considerably differed from one another and which were additive, i.e., any three strains met the condition $g_{12} = g_{13} + g_{23}$, where the indices 1 and 2 designate the two most different of these three strains. Monogenic traits were simulated for each of these loci; 120 pairs of loci were used for simulating digenic traits; and 498 triads of loci, for simulating trigenic traits.

The power of in silico mapping was estimated by the proportion of experiments in which the contribution of the simulated locus was considerable. In addition, we estimated the proportion of experiments in which the simulated locus was the best among significant loci.

We used a set of software written in FORTRAN-90, intended for simulating the traits, calculating correlation coefficients, and determining the boundary value.

RESULTS

The monogenic model. Similar results were obtained for all the 24 traits studied:

• the simulated locus showed a significant effect at all significance levels used;

• the simulated locus was always the best among all significant loci; and

• in addition to the simulated locus, other loci showed significant effects, the number of these loci decreasing with an increase in significance level.

The digenic model. Table 1 shows the results obtained for the digenic model. As evident from the table, there were experiments where none of the simulated loci showed a significant effect. The proportion of these experiments increased with an increase in significance level and was as large as 17% at $\alpha = 0.001$. If the effect of one gene was prevailing, the proportion of these experiments was extremely low, i.e., at least one simulated locus showed a significant effect.

The frequency of experiments where both simulated loci showed a significant effect was low. It decreased with an increase in significance level and in the difference between the gene effects. For example, it varied from 0.617 to 0.325 at different significance levels and was as low as 0.013 if the contributions of the genes differed from each other by a factor of ten.

Conversely, the proportion of experiments in which only one simulated locus showed a significant effect increased with an increase in significance level and in the difference between gene effects. If the difference between gene contributions was maximum, this proportion was 0.987, the simulated locus being the best among significant loci in most cases.

The trigenic model. Table 2 shows the results obtained for the trigenic model. If the gene contributions were equal to one another, then the experiments in which none of the simulated loci showed a significant effect were even more frequent than in the case of the digenic model (25% at $\alpha = 0.001$). If the effect of one gene was prevailing, the proportion of these experiments was low.

The proportion of experiments where all simulated loci showed significant effects (M_{SSS}) was low. It decreased with an increase in significance level and in the difference between the phenotypic contributions of the genes. If the contributions were equal to one another, it varied from 0.184 to 0.032 at different significance levels; if the effect of one gene was prevailing, it was no higher than 0.066.

In most experiments, two or one of the three simulated loci showed significant effects. With an increase in significance level, the proportion of experiments where only one of the simulated loci was significant. Gene contributions being different, the proportions of experiments in which one locus or two loci had a significant effect were 0.761 and 0.199, respectively.

DISCUSSION

To estimate the characteristics of the in silico method, we used a simulation experiment assuming that a trait was strictly determined by the genotypes of the analyzed strains. The method of the trait determination entirely corresponded to the basic assumptions of in silico mapping: (1) the change in a trait in a series of

Table 1. The results of mapping digenic traits

Relative contributions of the genes	Power	Significance level		
		0.01	0.005	0.001
$\mu_1 = 0.5$ $\mu_2 = 0.5$	P_{11}	0.042	0.042	0.042
	P_{1s}	0.417	0.358	0.225
	P_{1n}	0.175	0.233	0.367
	P_{ss}	0.158	0.108	0.058
	P_{sn}	0.2	0.175	0.142
	P_{nn}	0.008	0.083	0.167
	P_{SS}	0.617	0.508	0.325
	P_{SN}	0.375	0.408	0.509
$\mu_1 = 0.83$ $\mu_2 = 0.17$	P_{11}	0.021	0.021	0.021
	P_{1s}	0.154	0.113	0.033
	P_{1n}	0.788	0.829	0.908
	P_{ss}	0.008	0.008	0
	P_{sn}	0.029	0.029	0.033
	P_{nn}	0	0	0.004
	P_{SS}	0.183	0.142	0.054
	P_{SN}	0.817	0.858	0.946
$\mu_1 = 0.9$ $\mu_2 = 0.1$	P_{11}	0.013	0.013	0.013
	P_{1s}	0.088	0.071	0.017
	P_{1n}	0.896	0.913	0.967
	P_{ss}	0	0	0
	P_{sn}	0.004	0.004	0.004
	P_{nn}	0	0	0
	P_{SS}	0.013	0.013	0.013
	P_{SN}	0.987	0.987	0.987

P_{11}, the proportion of experiments in which the effects of both simulated loci were significant and maximum.
P_{1s}, the proportion of experiments in which the effects of both simulated loci were significant and the effect of one locus was maximum.
P_{1n}, the proportion of experiments in which the effect of one simulated locus was significant and maximum and the effect of the other locus was nonsignificant.
P_{ss}, the proportion of experiments in which the effects of both simulated loci were significant but not maximum.
P_{sn}, the proportion of experiments in which the effect of one simulated locus were significant but not maximum and the effect of the other locus was nonsignificant.
P_{nn}, the proportion of experiments in which the effects of both simulated loci were nonsignificant.
P_{SS}, the proportion of experiments in which the effects of both simulated loci were significant.
P_{SN}, the proportion of experiments in which the effect of one simulated locus was significant and the effect of the other locus was nonsignificant.
The subscript index 1 designates loci with significant and maximum effects; the index s, loci with significant but not maximum effects; and index n, loci with nonsignificant effects.

strains was proportional to the relative number of different nucleotides in the loci controlling this trait and (2) there was no interaction between genes. In addition, we did not take into consideration environmental modifications. This model allows only the upper limit of the power of the method to be estimated, because any deviation from the basic assumption must deteriorate the estimated power.

We analyzed three models of inheritance of a trait: monogenic, digenic, and trigenic. In the case of the monogenic model, the obtained results agreed with the results of the in silico mapping of factual data on traits that had proven to be monogenic. According to these data, a locus controlling a monogenic trait always shows a significant effect. Often, this locus is not the only significant locus, but its effect is the strongest. The identification of additional loci (false positive results) by in silico mapping may be explained by the so-called mirror loci characterized by the same distribution of genotypic differences (g_{ij}) for all pairs of strains. Three

Table 2. The results of mapping trigenic traits

Relative contributions of the genes	Power	Significance level		
		0.01	0.005	0.001
$\mu_1 = 0.33$ $\mu_2 = 0.33$ $\mu_3 = 0.33$	P_{111}	0	0	0
	P_{11s}	0.024	0.020	0.012
	P_{11n}	0.014	0.018	0.026
	P_{1ss}	0.090	0.060	0.012
	P_{1sn}	0.241	0.221	0.151
	P_{1nn}	0.104	0.155	0.273
	P_{sss}	0.070	0.050	0.008
	P_{ssn}	0.225	0.169	0.074
	P_{snn}	0.187	0.223	0.193
	P_{nnn}	0.044	0.084	0.251
	P_{SSS}	0.184	0.130	0.032
	P_{SSN}	0.480	0.408	0.251
	P_{SNN}	0.291	0.378	0.466
$\mu_1 = 0.625$ $\mu_2 = 0.250$ $\mu_3 = 0.125$	P_{111}	0	0	0
	P_{11s}	0.004	0.004	0
	P_{11n}	0.026	0.026	0.030
	P_{1ss}	0.052	0.034	0.002
	P_{1sn}	0.325	0.277	0.147
	P_{1nn}	0.396	0.462	0.624
	P_{sss}	0.010	0.002	0
	P_{ssn}	0.074	0.056	0.022
	P_{snn}	0.108	0.131	0.137
	P_{nnn}	0.005	0.008	0.038
	P_{SSS}	0.066	0.040	0.002
	P_{SSN}	0.425	0.359	0.199
	P_{SNN}	0.504	0.593	0.761

Note: Designations are the same as in Table 1.

hundred and seventy-six out of 1635 loci used in this study had mirror variants located at least 5 Mb away from them. The number of mirror loci depends on the density of the genetic map and the method of the genome division into separate regions. However, the existence of mirror regions is not the only cause of false positive results of mapping. For example, Pletcher et al. [2] using a less dense genetic map did not find mirror loci. However, more than one significant region was always found on mapping monogenic traits. Probably, incomplete coincidence of the genotypic differences (g_{ij}) between strains in pairs is sufficient for obtaining significant correlations. Taking into account the high power and the existence of false positive results, in silico mapping can be recommended only as a method for preliminary rejection of the loci that are certainly not involved in the genetic control of a given monogenic trait.

In the case of the digenic model, we obtained the following results. Experiments where the effects of all simulated loci were significant and exceeded those of other loci were extremely rare. In most cases, only one of simulated loci had a significant effect. The proportion of these experiments increased with an increase in the relative phenotypic contribution of one of the loci and in the significance level. At $\alpha = 0.001$ and a difference between the contributions of individual loci by a factor of five, the proportion of experiments where one of the simulated loci showed the maximum effect (and the effect was significant) was larger than 90%; if the contributions differed by a factor of ten, this proportion was 96%.

Similar results were obtained in the case of the trigenic model. Experiments in which all the three loci showed significant effects were very rare, none of these loci ever being the best of significant loci. In most

cases, one or two of the three simulated loci were significant. If the simulated effect of one of the loci was considerably stronger than those of others, this locus was often detected during mapping, whereas the other two loci were nonsignificant.

Thus, the results of our study indicate that the statistical power of in silico mapping is high if the studied trait is monogenic. The proportion of experiments where none of the genes was mapped was 0.17 and 0.25 in the cases of the digenic and trigenic models. Equal contributions of genes into the control of a trait are assumed in all models. If the phenotypic contribution of one gene was considerably larger than those of other genes involved in the control of the studied trait, this major gene was mapped in 95 and 80% of experiments in the digenic and trigenic models, respectively. Thus, we can expect that, in the case of polygenic control, the new method makes it possible to map only the major gene, whereas most genes with smaller phenotypic contributions are unlikely to be mapped.

In contrast to linkage analysis, in silico mapping requires the use of many inbred strains. Therefore, we expected that the new method would identify more genes involved in the control of the studied traits than linkage analysis. In fact, however, only one locus was reliably detected, and even this occurred only if it was the only locus controlling a trait (i.e., a trait was monogenic) or its effect was considerably stronger than those of other genes (major gene control). It should be borne in mind that our results estimate the upper limit of the statistical power, because we did not take into account gene interactions and modificational variation and used for simulation of traits only the loci with the greatest diversity of genotypic differences. In reality, estimates of the statistical power are likely to be even lower.

One explanation of such a low power of the method is the incorrect assumption that the difference between two strains with respect to analyzed traits is proportional to the number of different nucleotides in the genomic region responsible for their control. Another variant of in silico mapping based on the segmental structure of the genomes of inbred mouse strains [10] was proposed recently [2]. Each locus in different strains may be represented by a limited number of haplotypes regarded as fixed factors affecting the value of a quantitative trait. The involvement of a given locus in the control of a given trait is inferred from a significant contribution of a factor (haplotype) into the diversity of the trait determined by analysis of variance, rather than the correlation of the difference between two strains in the mean values of the trait with their difference in the genotypes of the given locus. The new variant of the method was used to analyze several quantitative traits. Many of the loci studied contained the genes that were earlier demonstrated to be involved in the control of the trait with the use of other methods. This allowed the authors to evaluate the new variant of in silico method as very promising.

To estimate the negative effect of the assumption on a quantitative dependence between genotypic and phenotypic differences, we repeated the analysis of two traits, the concentration of high-density lipoprotein cholesterol and the number of biliary calculi (http://aretha.jax.org), using the variant of mapping that we previously applied to estimating the power. We found about as many loci as the authors of the study [2] did. About three quarters of significant loci revealed by either variant of the method contained the genes that had earlier been demonstrated to control the analyzed traits. However, only one quarter of these loci were identified by both variants. This indicates that either variant could identify only some of the genes involved in the control of the traits. Therefore, the powers of both methods were low and close to each other.

Apparently, the low power of in silico mapping is accounted for by the fact that loci are analyzed independently from one another. In fact, various genes contributing to most quantitative traits have effects that depend on one another and/or are directed differently. All mapping methods face the problem of multilocus control of complex traits. Some authors believe that sequential analysis of individual loci, rather than simultaneous study of their complex sets, is the main cause of the slow progress in identifying the genes responsible for the most prevalent human diseases [11–13]. At present, the multilocus approach is being intensely developed in the framework of mapping human genes [14]. Probably, the introduction of multilocus models into in silico mapping will enhance the power of this method and make it useful for practical applications.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 04-04-48074) and the Program of the Russian Academy of Sciences "Plant, Animal, and Human Gene Pool Dynamics."

REFERENCES

1. Grupe, A., Germer, S., Usuka, J., et al., In Silico Mapping of Complex Disease-Related Traits in Mice, *Science*, 2001, vol. 292, no. 5523, pp. 1915–1958.
2. Pletcher, M.T., McClurg, P., Batalov, S., et al., Use of a Dense Single Nucleotide Polymorphism Map for in Silico Mapping in the Mouse, *PLoS Biol.*, 2004, vol. 2, no. 12, p. e393.
3. Hillebrandt, S., Goos, C., Matern, S., and Lammert, F., Genome-Wide Analysis of Hepatic Fibrosis in Inbred Mice Identifies the Susceptibility Locus *Hfib1* on Chromosome 15, *Gastroenterology*, 2002, vol. 123, no. 6, pp. 2041–2051.
4. Darvasi, A., In Silico Mapping of Mouse Quantitative Trait Loci, *Science*, 2001, vol. 294, no. 5551, pp. 2424–2426.
5. Savov, J.D., Whitehead, G.S., Wang, J., et al., Ozone-Induced Acute Pulmonary Injury in Inbred Mouse

Strains, *Am. J. Respir. Cell Mol. Biol.*, 2004, vol. 31, no. 1, pp. 69–77.

6. Chesler, E.J., Rodriguez-Zas, S.L., and Mogil, J.S., In Silico Mapping of Mouse Quantitative Trait Loci, *Science*, 2001, vol. 294, no. 5551, pp. 2423–2424.
7. Smith, J.D., James, D., Dansky, H.M., et al., In Silico Quantitative Trait Locus Map for Atherosclerosis Susceptibility in Apolipoprotein E-Deficient Mice, *Arterioscler. Thromb. Vasc. Biol.*, 2003, vol. 23, no. 1, pp. 117–122.
8. Lynch, M. and Walsh, B., *Genetics and Analysis of Quantitative Traits*, Sunderland: Sinauer, 1998.
9. Churchill, G.A. and Doerge, R.W., Empirical Threshold Values for Quantitative Trait Mapping, *Genetics*, 1994, vol. 138, no. 3, pp. 963–971.
10. Frazer, K.A., Wade, C.M., Hinds, D.A., et al., Segmental Phylogenetic Relationships of Inbred Mouse Strains Revealed by Fine-Scale Analysis of Sequence Variation across 4.6 Mb of Mouse Genome, *Genome Res.*, 2004, vol. 14, pp. 1493–1500.
11. Hirschhorn, J.N. and Daly, M.J., Genome-Wide Association Studies for Common Diseases and Complex Traits, *Nat. Rev. Genet.*, 2005, vol. 6, no. 2, pp. 95–108.
12. Nelson, M.R., Kardia, S.L., Ferrell, R.E., and Sing, C.F., A Combinatorial Partitioning Method to Identify Multilocus Genotypic Partitions that Predict Quantitative Trait Variation, *Genome Res.*, 2001, vol. 11, no. 3, pp. 458–470.
13. Wang, W.Y., Barratt, B.J., Clayton, D.G., and Todd, J.A., Genome-Wide Association Studies: Theoretical and Practical Concerns, *Nat. Rev. Genet.*, 2005, vol. 6, no. 2, pp. 109–118.
14. Hoh, J. and Ott, J., Mathematical Multi-locus Approaches to Localizing Complex Human Trait Genes, *Nat. Rev. Genet.*, 2003, vol. 4, no. 9, pp. 701–709.