View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Kansas State University

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture 2004 - 16th Annual Conference Proceedings

GENETIC MAPPING OF GENE EXPRESSION LEVELS: EXPRESSION LEVEL POLYMORPHISM ANALYSIS FOR DISSECTING REGULATORY NETWORKS OF PLANT DISEASE RESISTANCE

Kyunga Kim

Marilyn A. L. West

Richard W. Michelmore

Dina A. St. Clair

R. W. Doerge

See next page for additional authors

Follow this and additional works at: https://newprairiepress.org/agstatconference

Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Kim, Kyunga; West, Marilyn A. L.; Michelmore, Richard W.; St. Clair, Dina A.; and Doerge, R. W. (2004). "GENETIC MAPPING OF GENE EXPRESSION LEVELS: EXPRESSION LEVEL POLYMORPHISM ANALYSIS FOR DISSECTING REGULATORY NETWORKS OF PLANT DISEASE RESISTANCE," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1153

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Author Information

Kyunga Kim, Marilyn A. L. West, Richard W. Michelmore, Dina A. St. Clair, and R. W. Doerge

This is available at New Prairie Press: https://newprairiepress.org/agstatconference/2004/proceedings/7

GENETIC MAPPING OF GENE EXPRESSION LEVELS: EXPRESSION LEVEL POLYMORPHISM ANALYSIS FOR DISSECTING REGULATORY NETWORKS OF PLANT DISEASE RESISTANCE

Kyunga Kim[1], Marilyn A. L. West[2], Richard W. Michelmore[2], Dina A. St. Clair[2], and R. W. Doerge[1,3]

[1] Purdue University, Department of Statistics, West Lafayette, IN 47907, USA
[2] University of California, Department of Vegetable Crops, Davis, CA 95616, USA
[3] Purdue University, Department of Agronomy, West Lafayette, IN 47907, USA

Abstract

The genetic basis of inherited traits has been studied through different approaches in many areas of science. Examples include quantitative trait locus (QTL) analysis and mutant analysis in genetics, genome sequencing and gene expression analysis in genomics. Each of these approaches is used for the investigation of complex traits, such as disease resistance, but also provides knowledge on components of complex biological systems. We introduce a novel functional genomics approach that integrates two areas, genetics and genomics, by applying QTL analysis to quantitative differences in the mRNA abundance of trait-related genes. This approach allows comprehensive dissection of regulatory networks for complex traits at a systems biology level. We also address statistical issues, and suggest guidelines for future experiments in this new framework. (*Key words:* QTL mapping, gene expression, microarrays, expression level polymorphism, gene regulatory networks, e-QTL.)

1. Introduction

The molecular basis of complex traits (e.g., gene regulatory networks of disease response) has been studied with increasing interest using various approaches, such as quantitative trait locus (QTL) analysis (Mauricio, 2001; Doerge, 2002) and mutant analysis (Glazebrook, 2001) in genetics; genome sequencing (The Arabidopsis Genome Initiative, 2000) and gene expression analysis (Duggan et al., 1999; Lipshutz et al., 1999) in genomics. While the underlying regulatory mechanism of complex traits has been investigated in each area of genome science, our understanding of the entire process as a whole remains a great challenge (Kitano, 2002). In particular, QTL analysis identifies genomic regions associated with disease resistance, and mutant analysis even isolates the resistance genes that encode disease resistance (Michelmore, 2000), but the regulation and function of these regions/genes remain poorly understood. Recent technological and methodological progress is providing ways to improve our understanding of the molecular dissection of complex traits at a systems biological level, by incorporating the methodologies of genetics and genomics (Jansen and Nap, 2001; Doerge, 2002) to pursue a more comprehensive dissection of complex traits and their genomic architecture.

QTL analysis is an approach that relies on a detailed genetic map and statistical theory

to reveal the association between the genotype of genetic markers with a quantitative trait (phenotype) for the purpose of identifying chromosomal regions associated with the trait (Lander and Botstein, 1989; Jansen, 1993; Zeng, 1993). These QTL (or genomic regions) may contain structural genes that encode for proteins that are responsible for the realization of the quantitative trait, but these QTL may also be regulatory genes that may be responsible for the control (or regulation) of structural genes. Differentiating between structural and regulatory genes that are associated with a quantitative trait of interest is the challenge at hand and can be addressed by assessing the transcript abundance of genes that are expressed during the control of a quantitative trait. This can be accomplished using microarray technology (Duggan et al., 1999; Lipshutz et al., 1999) which measures transcript abundance for thousands of genes under different conditions, treatments, or cell types, and hence enables gene expression analysis on a global scale.

The union of quantitative genetics with genomics (i.e., gene expression) in this manner has been referred to by Jansen and Nap (2001) as 'genetical genomics', and has been accomplished in various ways (Jin et al., 2001; Brem et al., 2002; Wayne and McIntyre, 2002; Yvert et al., 2003). All of these investigations, however, employed rather simplistic statistical analyses (e.g., the Wilcoxon-Mann-Whitney test, χ^2 test, ANOVA) with limited population sizes (6-40 individuals). A more advanced investigation by Schadt et al. (2003) employed an F_2 population of 111 mice from two standard inbred strains that were known to have significant differential expression among the 23,574 genes in the mouse genome. Each gene that was differentially expressed between the parents was treated as a quantitative trait and evaluated for transcript abundance in each of the 111 F_2 individuals. Interval mapping (Lander and Botstein, 1989) was employed to identify genomic regions, called expression quantitative trait locus (e-QTL), that accounted for variation of gene expression for each gene (i.e., quantitative trait). While these investigations have produced results, the statistical methods that have been used in current genetical-genomic approaches have not been examined as to their statistical power in these applications. We expect the power to detect e-QTL to be low due to the simplicity of the statistical models relative to the complexity of the biological systems, small sample size, and the technological issues that are associated with any microarray experiment.

In our simulation study we use *Arabidopsis* which is the model organism for plants and spotted cDNA microarrays to assess gene expression levels induced by two experimental treatment conditions. The conditions of interest are a control and the response to an application of salicylic acid (SA). The role of SA as a signaling intermediate in disease resistance pathways has been studied through mutant analysis (Glazebrook, 2001). We are particularly interested in the genes that respond to the SA treatment, especially the regulatory genes that are involved in the control of the structural genes that are providing the response. We define the variation involved in the genes that are providing the response to be an 'expression level polymorphism' (ELP), and employ an existing QTL methodology, namely multiple trait analysis (Jiang and Zeng, 1995) to decompose the 'variation in the expression levels' (ELP) across individuals in a recombinant inbred segregating population into genetic and non-genetic factors, and their interactions. In addition to the identification of genetic deter-

minants of ELP, this application of multiple trait analysis also provides statistical tests for ELP-by-treatment interactions, and may give us insight into the gene regulatory system of complex traits. Through simulation studies, we demonstrate the consequences of the particular sources of variation (e.g., treatment, dye, array, etc.) on the application of multiple trait QTL analysis to these data, as well as address other statistical issues, such as statistical power, sample size, and effective number of technical (spot) replicates. We also put forth a global view of ELP analysis, including experimental design, data collection, ELP mapping, bioinformatics on mapping results, and reconstruction of gene regulatory networks (Figure 1).

2. Methods

Recently, genetic/allelic variation in gene expression levels have been recognized in several studies (Cowles et al., 2002; Oleksiak et al., 2002; Yan et al., 2002). As mentioned previously, natural variation that is seen in induced gene expression levels across individuals in a population is defined as expression level polymorphism, or ELP (Doerge, 2002). By decomposing sources of ELP into genetic and nongenetic factors, and their interactions, one can identify genetic determinants of ELPs, which contribute to gene expression changes, and map them to specific genomic regions. This approach will help to further characterize a complex trait by using the information gained from ELP analysis to infer the genes involved in the pathway of the trait. In cooperation with physical and genetic map information, the putative determinants of expression level polymorphism can be categorized into a 'structural locus' that is a gene whose expression levels are under investigation and 'regulatory loci' that are other genes which control/regulate the expression of the structural gene.

2.1 Genetic models for regulatory pathways of ELP

Gene regulatory networks are typically very complex and include a large number of components (e.g., stimuli, structural and regulatory determinants). For simplicity we illustrate three simple pathway models based on one structural gene with one or two regulatory determinants (Figure 2; taken from Kim et al. (2005)). Our investigation will be based upon the simplest model (Model 1 in Figure 2) that consists of a single structural gene and one regulatory locus that controls the expression changes of the structural gene that is in turn providing the response to the stimulus. According to the proposed pathway models, the regulatory loci/genes react to an outside stimulus, and their products physically bind to the promoter site of the structural gene to promote or repress expression of the structural gene (Lewin, 2000). Each pathway model can be characterized by a genetic model that reflects components of both the pathway and the experiment. Assuming there is allelic variation in gene expression, one can decompose the variation in gene expression into several components: genetic components such as sequence polymorphism (allele difference) of each ELP determinant, non-genetic component such as treatment effect (the nature of the stimulus), interactions between genetic and non-genetic component (genotype and the stimulus), and also systematic or technological components such as dye effect and array effect.

A decomposition of the sources of variation of this type often relies on linear models, which are well established, and have been applied extensively in both QTL mapping (Zeng, 1993; Jiang and Zeng, 1995) and microarray analysis (Kerr et al., 2000). As such, we propose a linear genetic model for the dissection of a regulatory pathway of a single structural gene with a regulatory locus. It is straightforward to generalize this model with multiple regulatory loci (e.g., Models 2 and 3 in Figure 2). Under the assumption of additive genetic effects (no epistasis), and a segregating recombinant inbred line (RIL) population, let i and j denote the genotypes of regulatory and structural loci, respectively, for each individual in the RIL population, and let y_{ijklmr} denote the measurement of transcript abundance for the structural gene as gained from spotted cDNA microarrays. This expression measurement is in the original or log scale under treatment k with dye l on array m in replication r:

$$y_{ijklmr} = \mu + \alpha_i + \beta_j + \tau_k + (\alpha\tau)_{ik} + \delta_l + A_m + \epsilon_{ijklmr} \tag{1}$$

where $i = 1, 2$, $j = 1, 2$, $k = 1, 2$, $l = 1, 2$, $m = 1, 2$, and $r = 1, \cdots, R$; μ is overall mean; α, β, τ and δ are the additive effects of the regulatory and structural loci, treatment effect, and dye effect; $\alpha\tau$ is interaction between the regulatory ELP and the treatment; A is the array effect and assumed to be distributed as a random normal with mean 0 and variance σ_A^2; ϵ_{ijklmr} is the measurement error distributed as a random normal with mean 0 and variance σ_ϵ^2. The array effect and measurement error are assumed to be independent of each other and between individuals.

2.2 ELP experimental design

As in any QTL analysis use of the appropriate segregating population is crucial to the success of the mapping. Typically, such populations are developed via experimental mating systems from two homozygous and inbred parental lines that are genetically different for the phenotype of interest. Recombinant inbred line (RIL) populations are commonly used in QTL experiments, and especially recommended for complex trait research due to the advantage of being able to study low-heritability traits, and/or QTL by environmental interaction effects (Mackay, 2001; Jansen, 2003). In order to determine which parental inbred lines will produce reasonable differences in the expression levels of a gene (i.e., ELP; Figure 3), a preliminary experiment using oligonucleotide microarrays (Affymetrix chips) was conducted with a set of natural accessions of *Arabidopsis thaliana*, and differential changes in expression between each parental pair were examined based on a mixed linear model and t-tests with a multiple comparisons adjustment, such as Holm's procedure or false discovery rate (West et al., 2004).

Microsatellites provide information on the *Arabidopsis* genome via marker genotypes that are observed for each of N RILs as derived from the parental pair selected in the preliminary analysis (Figure 4). Because we are using spotted cDNA arrays for transcript assessment, a collection of cDNAs that represent genes (including those genes detected with potential ELPs in the preliminary analysis) are derived and spotted on glass slides to provide the arrays that will be hybridized. Once the arrays are made, then each RIL individual provides mRNA samples collected under two treatment conditions (i.e., no SA or the control, and SA). A dye-swap experimental design (Black, 2002) is used, and four measurements of

expression levels for each gene are collected from two arrays for each individual RIL. The benefits of the dye-swap design are to control the dye effect which is often the result of an unequal incorporation of one dye over the other, and to provide technological replicates (i.e., two measurement for each treatment condition, with different dyes) which contribute to improving the power of statistical analysis.

2.3 ELP mapping using multiple trait mapping

ELP mapping in this context uses multiple measurements of gene expression over different treatment and dye conditions to identify and map genetic determinants of a single ELP. These multiple expression measurements can be regarded as multiple phenotype evaluations over different environments as in multiple trait analysis (Jiang and Zeng, 1995). Multiple trait analysis is an advanced QTL methodology extended from single trait QTL analysis to multiple correlated traits. For each individual in a RIL population, w_p denotes the phenotypic value of the pth trait among P quantitative traits. The statistical model for multiple trait composite interval mapping (CIM) is:

$$w_p = \beta_{0p} + \beta_p^* x^* + \sum_{l=1}^{t} \beta_{lp} x_l + \epsilon_p \tag{2}$$

where $p = 1, \cdots, P$; β_{0p} is overall mean for trait p; β_p^* is the additive effect of the putative QTL on trait p; x^* is the number of one allele of the putative QTL for the individual; β_{lp} is partial regression coefficient of w_p on x_l; x_l is the number of the allele of marker l (among t markers used for controlling residual genetic variation) on the individual; ϵ_p is the residual effect on trait p for the individual, and the residual effects are correlated among traits within individuals while they are independent among individuals. By accounting for the correlation structure among the traits, multiple trait CIM provides a statistical test for the gene-by-treatment interactions, as well as for genomic regions that are associated with the multiple traits. Four transcript measurements of a structural gene from the ELP experiment (see equation (1) and section 2.2) can be viewed as four traits in the context of multiple trait CIM, as described in Kim et al. (2005):

$$\begin{aligned}
w_1 &= \frac{1}{R}\sum_{r=1}^{R} y_{ij111r} = \mu + \alpha_i + \beta_j + \tau_1 + (\alpha\tau)_{i1} + \delta_1 + A_1 + \frac{1}{R}\sum_{r=1}^{R} \epsilon_{ij111r} \\
w_2 &= \frac{1}{R}\sum_{r=1}^{R} y_{ij221r} = \mu + \alpha_i + \beta_j + \tau_2 + (\alpha\tau)_{i2} + \delta_2 + A_1 + \frac{1}{R}\sum_{r=1}^{R} \epsilon_{ij221r} \\
w_3 &= \frac{1}{R}\sum_{r=1}^{R} y_{ij122r} = \mu + \alpha_i + \beta_j + \tau_1 + (\alpha\tau)_{i1} + \delta_2 + A_2 + \frac{1}{R}\sum_{r=1}^{R} \epsilon_{ij122r} \\
w_4 &= \frac{1}{R}\sum_{r=1}^{R} y_{ij212r} = \mu + \alpha_i + \beta_j + \tau_2 + (\alpha\tau)_{i2} + \delta_1 + A_2 + \frac{1}{R}\sum_{r=1}^{R} \epsilon_{ij212r}
\end{aligned} \tag{3}$$

Based on the representation in equation (3), we can apply multiple trait analysis to our ELP data to identify chromosomal regions or genetic determinants of expression level polymorphism that are associated to the treatment response (e.g., salicylic acid).

2.4 Bioinformatics and regulatory networks

ELP mapping yields genomic locations of ELP determinants, as well as information about their interactions with treatment. However, it is hard to assess whether these determinants are structural genes, or *cis*- or *trans*-regulatory regions. The *cis*-acting regulatory elements reside in the vicinity of the structural gene (e.g., promoter regions), whereas the *trans*-acting regulatory genes are usually located away from the structural gene, even on different chromosomes. Incorporation of the statistical interaction information with the physical map information enables further characterization of putative *cis*- and/or *trans*-acting regulatory regions, and allows one to construct a regulatory pathway for each gene. By integrating pathways for each of many genes, a putative regulatory network of related genes may be estimated.

3. Simulation study

We conducted simulation studies based on the simple genetic model (Model 1 in Figure 2; equation (1)) as the first step in validating the application of multiple trait QTL analysis to the ELP framework, and to examine the effect of sample size, as well as spot number (technical) replication. We simulated segregating populations consisting of 100, 200, 300, 400, 500, 700, and 1000 RILs. For the details about simulation settings, see Kim et al. (2005). The *JZmapqtl* procedure in QTL Cartographer (Basten et al., 2002) was employed to conduct multiple trait composite interval mapping on the simulated data.

Using the *Arabidopsis* genome as our model organism, we considered 5 chromosomes whose lengths are 135, 100, 100, 125, and 140 cM. Over the genome, 125 markers were equally distributed. Two unlinked ELP determinants (a regulatory and a structural) were generated with equal additive effects (0.1 and 0.5) and placed at 98 cM from the left end of chromosome 1 and 27cM from the left end of chromosome 2, respectively. The other chromosomes contain no ELP determinants. Each ELP determinant, as well as each marker was assumed to be biallelic, and the genotypes of each ELP determinant and each unlinked marker were randomly assigned according to their allele frequencies. Assuming a recombination process with moderate interference, we used Kosambi's map function. The genotype of each linked marker on chromosome 1 and 2 was assigned based on both the recombination fraction and genotype of the corresponding ELP determinant.

One structural gene, whose expression levels are assumed to be significantly different over treatment conditions between parental lines, was considered. Transcript abundance of the structural gene was simulated according to a dye-swap experimental design using spotted cDNA arrays. Gene expression for each RIL individual based on the linear genetic model (1), with various number of spot (technical) replicates (R = 1, 2, 3, 4, 5, 7, 10) resulted. Corresponding to various examples of potential ELP (Figure 3), a range of parameter values

(Table 1 in Kim et al. (2005)) was considered.

The simulated data were then analyzed using the *JZmapqtl* procedure (QTL Cartographer version 1.61v; Basten et al. (2002)). Empirical thresholds based on 1000 permutations (Churchill and Doerge, 1994) were estimated for both tests, namely detecting a significant determinant for ELP and its significant interaction with the treatment. When estimating permutation thresholds of multiple traits mapping, it is essential that the randomizations maintain the correlation structure between traits within each individual. To validate our approach and to investigate the effective sample size, we conducted multiple trait CIM with 200 data sets which were repeatedly simulated under each parameter configuration and with each sample size and no replicate spots. The statistical power to detect an ELP determinant was estimated by calculating the proportion of genes that multiple trait CIM significantly identified as well as located in a 1 cM neighborhood of its true location. The power to detect the ELP-by-treatment interaction was also examined based on how many times the interaction was significantly detected out of 200 simulation runs. The effective number of spot replication was also investigated via simulations under each parameter configuration and with sample size of 500.

4. Results

Using representative examples of the simulation results (Figures 5 and 6), several findings based on simulations with various parameter configurations can be summarized. First, ELP mapping via multiple trait CIM detected the regulatory determinant relatively well with good identification of the treatmental interaction, whereas it showed poor power to detect the structural gene, especially with a low additive effect. The estimated detection power for the regulatory locus was observed to be higher than for the structural locus over all parameter settings, although their additive effects are same. Accounting for gene-by-treatment interaction boosts the power of our approach to detect a genetic determinant which responds to treatment condition. By examining the increasing pattern of the power over sample size (Figure 5), we found that the continued gain in power becomes less after the sample size approaches 500. Furthermore, the benefit of technological (spot) replicates in ELP mapping via multiple trait CIM is largest when replicates are introduced, even for a small number of technical replicates (e.g., from no replicate to two replicates; Figure 6).

For more complicated genetic models with multiple regulatory genes (e.g., Models 2 and 3 in Figure 2) simulations were conducted, but are not presented here. However, it was demonstrated that ELP mapping using multiple trait CIM has much less power to detect ELP determinants under complex genetic models, and hence requires even larger sample sizes. We also performed further exploration on the properties of the estimates when applying multiple trait CIM by examining the sample means and standard deviations (over all 200 runs) of positions and additive effects for ELP determinants. In addition to higher detection power, the position and additive effect were more precisely estimated for the regulatory determinant than for the structural determinant. However, the estimates of the additive effect show relatively large bias for the regulatory locus (Figure 5), and this phenomenon may result from the existing non-genetic effects in statistical model, such as

treatment effect and ELP-by-treatment interaction. Interestingly, the employment of spot replicates reduces this bias (Figure 6), which is not adjusted by an increment of sample size.

5. Summary

We have introduced expression level polymorphism (ELP) mapping as an approach that combines an advanced QTL analysis in genetics with gene expression experiments in genomics. This approach provides some insight into comprehensive dissection of regulatory pathways for complex traits. While other studies based on genetical genomics (Jin et al., 2001; Brem et al., 2002; Steinmetz et al., 2002; Wayne and McIntyre, 2002; Schadt et al., 2003; Yvert et al., 2003) have used gene expression levels for only one parental line or one treatment condition at a time as related to a complex trait, our ELP analysis uses all gene expression levels quantified under different treatment conditions at the same time so that one can assess proper surrogates of the complex trait and identify genetic determinants of ELPs. The approach proposed here relies on the application of multiple trait interval mapping to the ELP data, and provides a way to account for possible interactions between regulatory loci and the treatment for the purpose of effective detection of ELP determinants. The statistical consideration of the ELP-by-treatment interaction in multiple trait platform allows our approach to potentially outperform other approaches based on the standard interval mapping (Schadt et al., 2003). The results from the simulations also revealed that statistical power for detecting regulatory loci is increased when there are interactions with the treatment, and that the statistical models employed to dissection of complex traits should be able to accommodate these interactions.

Our limited simulation studies show that the proposed approach based on multiple trait QTL mapping has great potential. The statistical power of multiple trait CIM for the detection of treatment-interacting ELPs is fairly good when the additive effects of the ELPs are not too small, and the underlying genetic model is simple (Model 1 in Figure 2), especially when gene replicates are employed. However, as also revealed by these simulations, multiple trait CIM application has its limitations. The detection power decreases, and the benefit of spot replicates degrades as the biological system becomes complex (Models 2 and 3 in Figure 2). In order to overcome the limitations of the multiple trait CIM application, we are developing a new methodology that is tailored for the ELP analysis and dissects the sources of ELPs with better power by completely accounting for non-genetical factors from microarrays.

In this work, we not only proposed the ELP mapping method, but we also demonstrated an evaluation process for ELP analysis that initiates from experimental design and concludes with the reconstruction of regulatory networks involved in complex traits (Figure 1). Our simulation studies suggest some practical guidelines for experimental issues, such as the proper sample size and the effective number of technical (spot) replicates.

6. Acknowledgments

This work is supported by an NSF grant (No. NSF01-13MCB-0115109) to D. A. St. Clair, R. W. Doerge, and R. W. Michelmore.

7. References

Basten, C. J., Weir, B. S., and Zeng, Z.-B. (2002). *QTL Cartographer*. Department of Statistics, North Carolina University, Raleigh, NC. Version 1.16.

Black, M. A. (2002). *Statistical issues in the design and analysis of spotted microarray experiments*. PhD thesis, Purdue University, West Lafayeete, Indiana.

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in Budding Yeast. *Science*, 296:752–755.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138:963–971.

Cowles, C. R., Hirschhorn, J. N., Altshuler, D., and Lander, E. S. (2002). Detection of regulatory variation in mouse genes. *Nature Genetics*, 32:432–437.

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, 3:43–52.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nature Genetics*, 21(Suppliment):10–14.

Glazebrook, J. (2001). Genes controlling expression of defense responses in arabidopsis-2001 status. *Current Opinion in Plant Biology*, 4:301–308.

Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics*, 135:205–211.

Jansen, R. C. (2003). Studying complex biological systems using multifactorial perturbation. *Nature Reviews Genetics*, 4:145–151.

Jansen, R. C. and Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends in genetics*, 17:388–391.

Jiang, C. and Zeng, Z.-B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, 140:1111–1127.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in *drosophila melanogaster*. *Nature Genetics*, 29:389–395.

Kerr, M. K., Martin, M., and Churchill, G. A. (2000). Analysis of variance for gene expression microarray data. *Journal of computational biology*, 7:819–837.

Kim, K., West, M. A. L., Michelmore, R. W., St. Clair, D. A., and Doerge, R. W. (2005). *The 23rd Stadler Genetics Symposium-Genome Exploitation: Data Mining*, chapter "Old methods for new ideas: Genetic dissection of the determinants of gene expression levels". *In press*.

Kitano, H. (2002). Sytems biology: a brief overview. *Science*, 295:1662–1664.

Lander, E. S. and Botstein, D. (1989). Mapping mandelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121:185–199.

Lewin, B. (2000). *Genes VII*. Oxford University Press Inc., New York, NY.

Lipshutz, R. J., Fodor, S. P. A., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nature Genetics*, 21(Suppliment):20–24.

Mackay, T. F. (2001). Quantitative trait loci in *drosophila*. *Nature Reviews Genetics*, 2:11–20.

Mauricio, R. (2001). Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics*, 2:370–381.

Michelmore, R. (2000). Genomic approaches to plant disease resistance. *Current Opinion in Plant Biology*, 3:125–131.

Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nature Genetics*, 32:261–266.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Collnayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., Linsley, P. S., Mao, M., Stoughton, R. B., and Friend, S. H. (2003). Genetics of gene expression surveyed in maize, mouse, and man. *Nature*, 422:297–302.

Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H., and Davis, R. W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature*, 416:326–330.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *arabidopsis thaliana*. *Nature*, 408:796–815.

Wayne, M. L. and McIntyre, L. M. (2002). Combining mapping and arraying: an approach to candidate gene identification. *Proceedings of the National Academy of Sciences of the United States of America*, 99:14903–14906.

West, M. A. L., Kim, K., Lam, C., Walker, R., Doerge, R. W., Michelmore, R. W., and St. Clair, D. A. (2004). Survey in *arabidopsis thaliana* for potential expression polymorphisms associated with quantitative disease resistance pathways. Poster presented in Plant and Animal Genome XII Conference.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science*, 297:1143.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R., and Kruglyak, L. (2003). *Trans*-acting regulatory variation in *saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics*, 35:57–63.

Zeng, Z.-B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America*, 90:10972–10976.

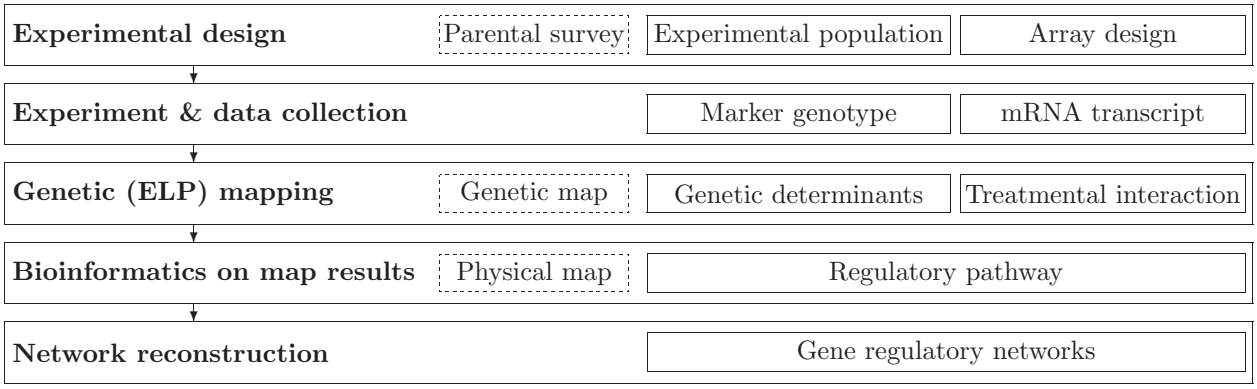

Figure 1: **Framework for expression level polymorphism analysis.** At each step (top to bottom), a newly added input(s) (in addition to inputs from the previous steps) is presented in a dashed box whereas an output(s) is in a solid box.

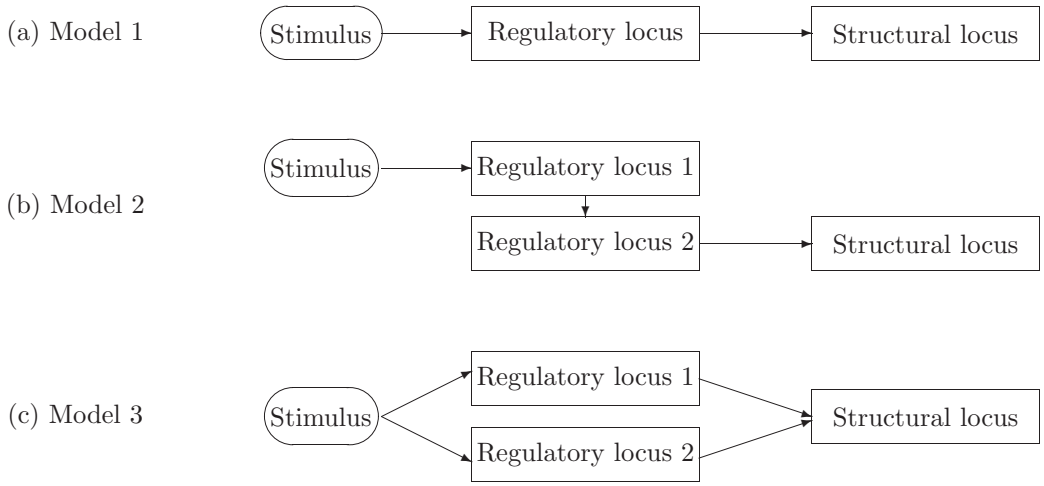

Figure 2: **Biological models for simple regulatory pathway models.** Although there is great diversity in biological models of regulatory pathways, three simple models are illustrated for a single structural gene with one or two regulatory ELP(s). Note that this figure was taken from Kim et al. (2005).

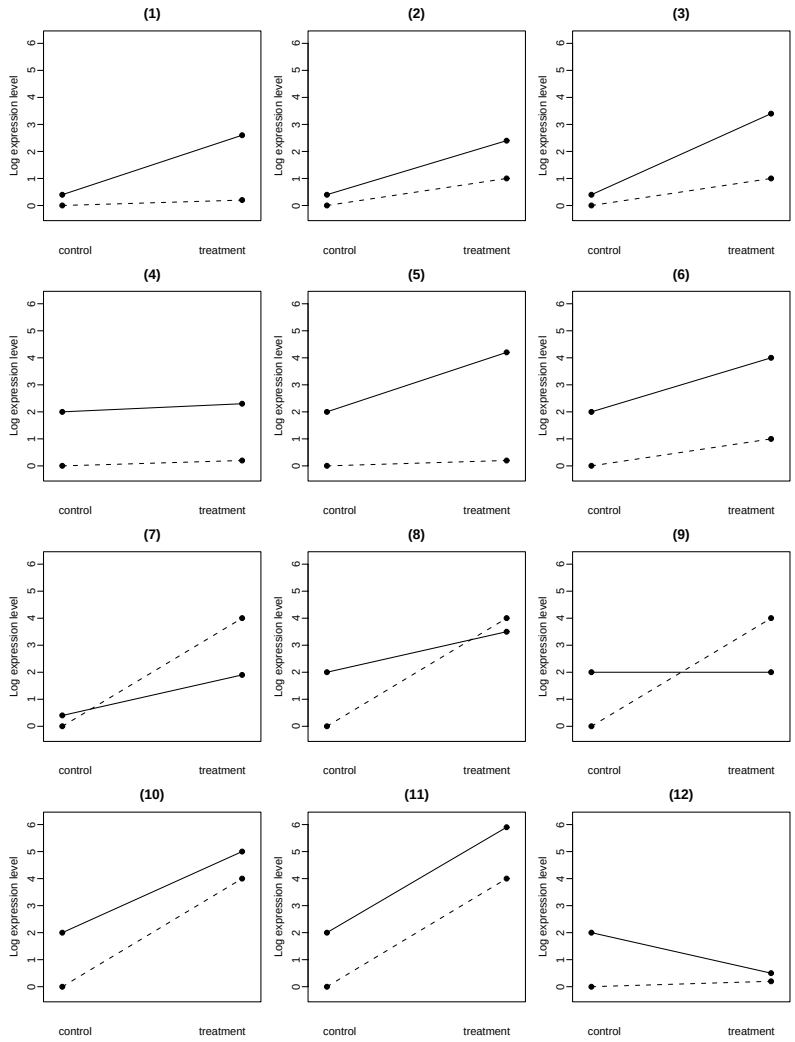

Figure 3: **Examples of gene expression profiles with potential ELP.** Dashed lines indicate one parental line (reference line); solid lines indicate the other parental line. The vertical axis is for (relative) expression levels where the expression level for the reference line with no treatment (control) was adjusted to 0. The cutoff value for potential expression level polymorphisms was assumed to be adjusted to 2-fold change.

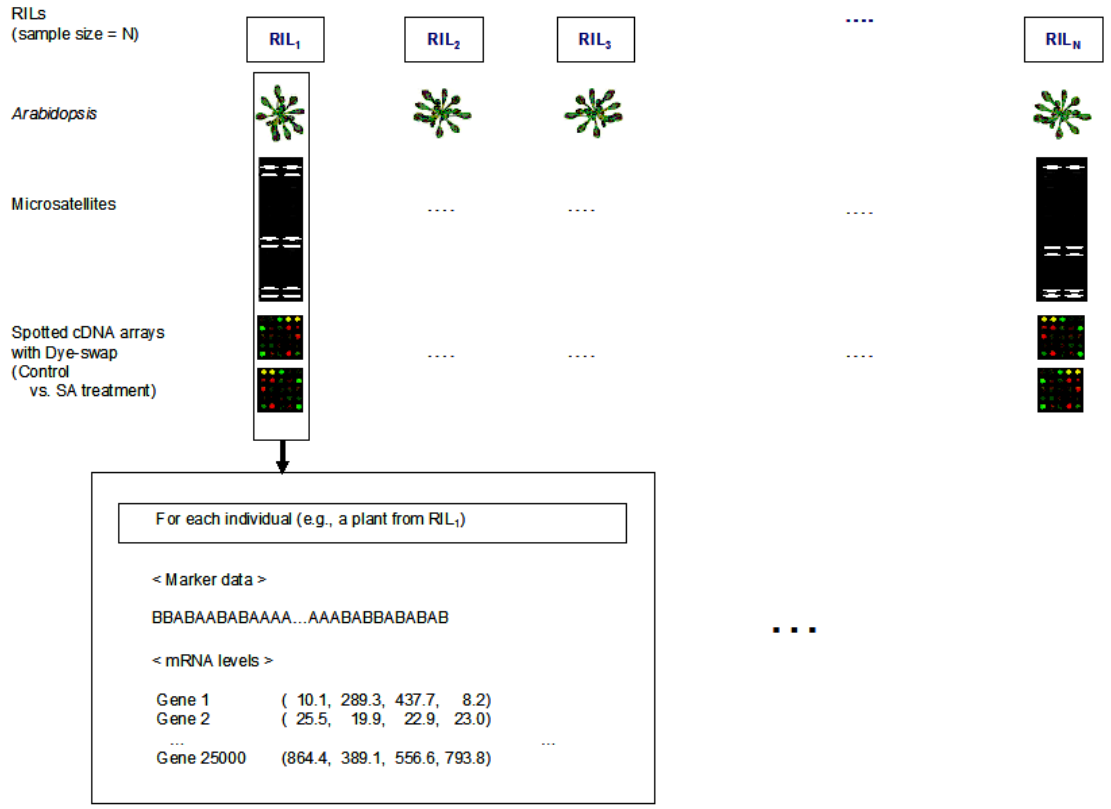

Figure 4: **Experimental details and data collection for expression level polymorphism analysis.** 'RIL' denotes a recombinant inbred line. For each individual RIL, microsatellites are employed to collect marker genotype data, and 'A' and 'B' stand for homozygous genotypes of either parent, respectively. Two spotted cDNA microarrays are used to measure transcript abundance levels under two treatment conditions (control vs. SA) with dye-swap design. For each gene, four numbers indicate transcript abundance levels relative to combinations of two treatment conditions and two dyes.

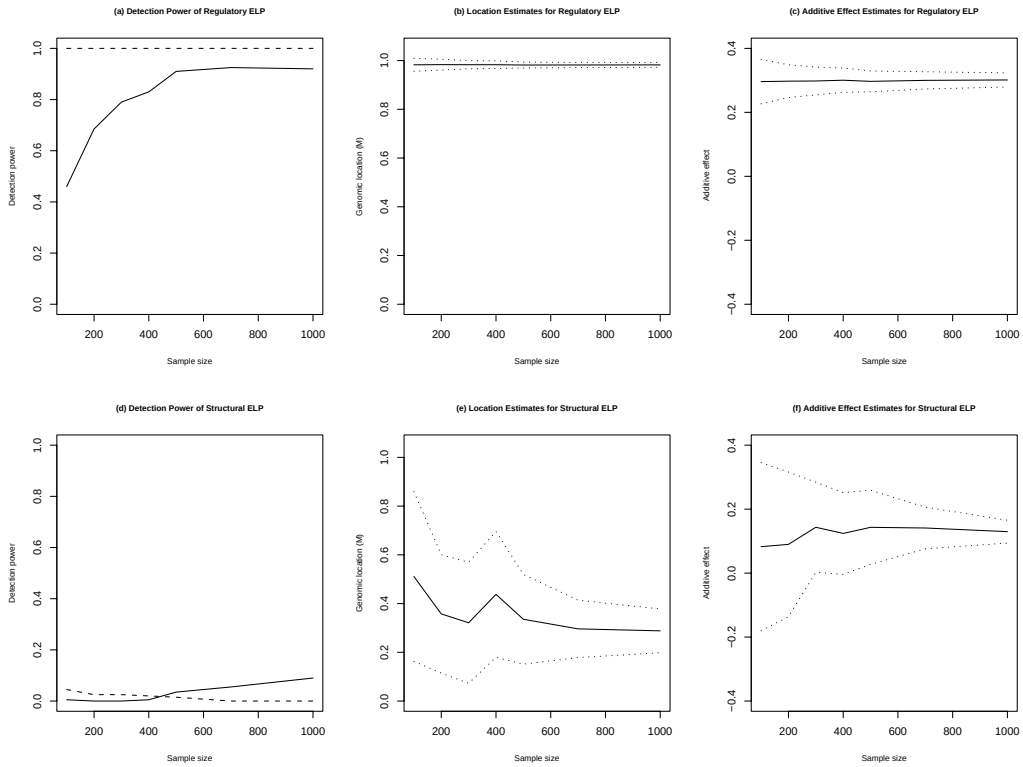

Figure 5: **Power and estimates according to sample size.** Results based on 200 simulation runs for configuration 1. True locations for a regulatory and a structural genes are 98cM and 27cM, respectively. Additive effect was 0.1 for both genes. For (a), (d) the solid lines represent detection power for ELP determinants; For (b), (c), (e), (f) the solid lines represent the sample mean of the estimates of location and additive effect; dotted lines represent the corresponding 1-standard error limits. Note that this figure was taken from Kim et al. (2004).

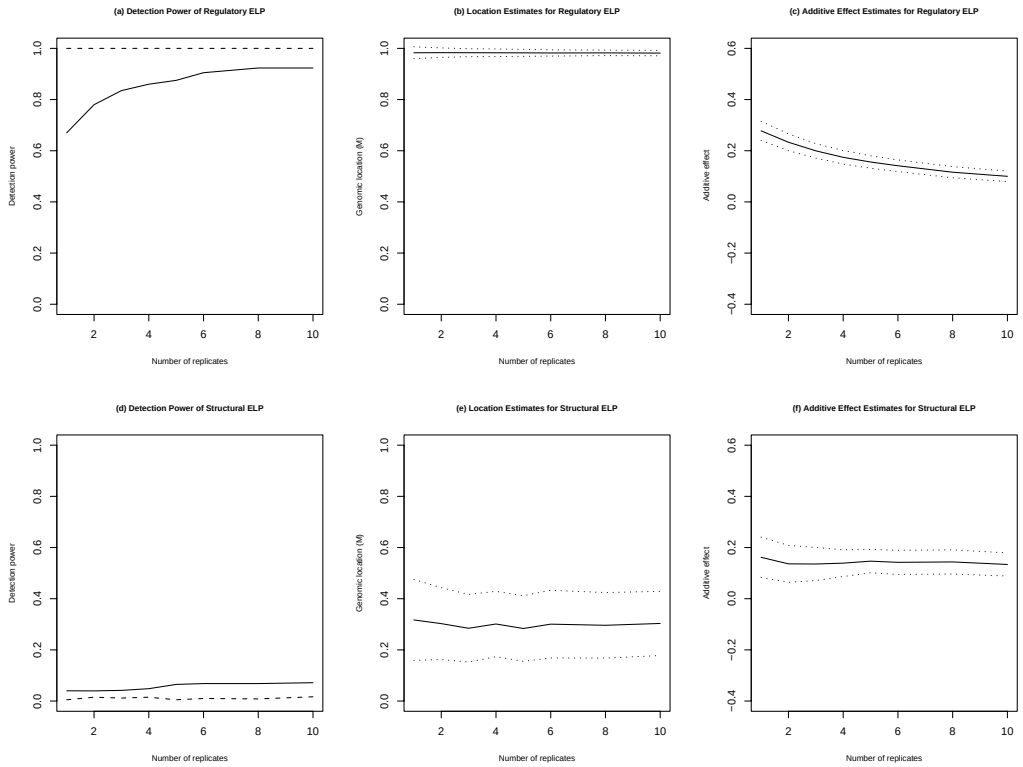

Figure 6: **Power and estimates according to replicates.** Results based on 600 simulation runs with sample size of 500 for configuration 5. True locations for a regulatory and a structural genes are 98cM and 27cM, respectively. Additive effect was 0.1 for both genes. For (a), (d) the solid lines represent detection power for ELP determinants; dashed lines represent detection power for treatmental interaction. For (b), (c), (e), (f) the solid lines represent the sample mean of the estimates of location and additive effect; dotted lines represent the corresponding 1-standard error limits.