Theor Appl Genet (2008) 116:683–690
DOI 10.1007/s00122-007-0701-3

ORIGINAL PAPER

Combining classical trait and microarray data to dissect transcriptional regulation: a case study

Dong Wang · Dan Nettleton

Received: 10 August 2007 / Accepted: 13 December 2007 / Published online: 9 January 2008
© Springer-Verlag 2008

Abstract The selective transcriptional profiling approach involves selecting an optimal subset of individuals to microarray from a larger set of individuals for which relatively inexpensive quantitative trait and molecular marker data are available. The goal of the selection and subsequent analyses is to identify genes whose expression is associated with a quantitative trait or quantitative trait locus (QTL). In this paper, we applied the selective transcriptional profiling approach to data sets concerning flowering time and gene transcription levels of *Arabidopsis* recombinant inbred lines. Our results confirm that the selective transcriptional profiling approach can achieve much greater power for uncovering associations than standard approaches that ignore information from classical traits. In addition, we show that selective transcriptional profiling can achieve power similar to standard approaches at a fraction of the cost and effort. We also identified three groups of genes which show distinctive patterns with regard to gene expression levels, QTL genotype, and a classical trait. This study represents the first application of selective transcriptional profiling to real data and serves as a template for dissecting gene regulation networks related to a classical trait using the selective transcriptional profiling approach.

Communicated by M. Sorrells.

D. Wang (✉)
Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall North, Lincoln, NE 68583-0963, USA
e-mail: dwang3@unl.edu

D. Nettleton
Department of Statistics, Iowa State University, 111A Snedecor Hall, Ames, IA 50011, USA
e-mail: dnett@iastate.edu

Introduction

Recently, there has been great interest in combining quantitative trait locus (QTL) mapping technique and microarray technology to dissect regulation networks for gene expression. In genetical genomic (Jansen and Nap 2001) studies, a gene's transcript abundance measured by microarrays is treated as a quantitative trait to map expression QTLs (eQTL, Schadt et al. 2003). Published results involving yeast, mice, human, and *Arabidopsis* have revealed complex inheritance mechanisms for gene transcription levels (Brem et al. 2002; Schadt et al. 2003; Yvert et al. 2003; Bystrykh et al. 2005; Chesler et al. 2005; Hubner et al. 2005; DeCook et al. 2006; West et al. 2007). Both *cis*- and *trans*-acting eQTLs have been described for regulatory loci that do or do not colocalize with the gene targeted for regulation. While true *cis*-eQTL are believed to be genes that regulate their own expression, evidence in several organisms suggests the existence of eQTL hot spots with multiple adjacent *trans*-eQTLs that control a large number of transcripts. Studies using the genetical genomics approach have also led to successful mapping of *cis*-regulating genes. Examples include Brem et al. (2002), Rockman and Wray (2002), Lan et al. (2003), Schadt et al. (2003), Pastinen and Hudson (2004), Doss et al. (2005), GuhaThakurta et al. (2006), Hughes et al. (2006), Kiekens et al. (2006), Zhang et al. (2006), Liang et al. (2007) and Luo et al. (2007). Readers are also referred to the review article of Sieberts and Schadt (2007) for more discussion. However, there remain considerable obstacles in

deciphering the pathways that eQTLs use to modulate various phenotypes.

The selective transcriptional profiling approach proposed by Wang and Nettleton (2006) focused on a new aspect of identifying gene regulation pathways. The approach is motivated by the fact that a research project is often focused on a specific classical quantitative trait. If a major QTL for this classical trait has been identified, it is often desirable to test whether this QTL is also associated with the transcription level of any genes, which will provide clues as to which genes belong to the pathway that the QTL uses to modulate the classical trait. Wang and Nettleton (2006) argue that because the transcriptional abundance of genes that are regulated by the QTL is likely to be correlated with the trait level after accounting for the QTL effect, incorporating the trait data on extra individuals can significantly increase the power for detecting the association between the QTL and gene expression levels. Thus in a study based on selective transcriptional profiling, a panel of several dozen to several hundred individuals is used to map the QTL associated with the trait of interest. After a major QTL has been identified, a subset of individuals is selected for microarray analysis to measure gene expression levels; then the microarray data and the trait data on all individuals are combined to perform tests on the association of the QTL with gene expression levels. Because microarray experiments are only performed on a subset of individuals, the financial and human cost related to a large number of microarrays can be greatly reduced. At the same time, utilizing trait data on extra individuals in selective transcriptional profiling can potentially enhance the power of the test even with fewer microarrays.

The performance of the selective transcriptional profiling approach has been demonstrated with simulated data sets in Wang and Nettleton (2006) and through hypothetical examples in Nettleton and Wang (2006). Until now, however, the effectiveness of the approach has never been investigated using trait and expression data from an actual experiment. In this paper, we illustrate the first ever real-data application of selective transcriptional profiling using data concerning flowering time under short photoperiod on 398 recombinant inbred lines derived from two *Arabidopsis thaliana* accessions (Bayreuth-0 and Shahdara), as well as microarray data on 142 of these recombinant inbred lines (RILs). We demonstrate that the selective transcriptional profiling approach does possess the proposed advantage in identifying genes whose transcriptional abundance is associated with a quantitative trait or QTL. Our analyses also revealed three distinctive groups of genes that are of interest for studying expression regulation related to a specific trait. Methods used in this paper can serve as a template for future research using selective transcriptional profiling.

Materials and methods

Selective transcriptional profiling

The experimental design and data analysis for selective transcriptional profiling has been discussed in Wang and Nettleton (2006). For a classical trait of interest, suppose a panel of N individuals was used to carry out QTL mapping and a QTL has been identified. For convenience, also suppose the QTL has two genotypes, x and y; N_x of the N individuals are of genotype x and N_y individuals are of genotype y. Because of financial or time constraints, only n_x individuals of genotype x and n_y individuals of genotype y will be chosen for microarray analysis to measure the expression level of each of thousands of genes. First we consider the model for the expression of only one gene, though the analysis will be implemented separately on thousands of genes.

Let X_{i1} and X_{i2} denote the quantitative trait and gene expression measure for the ith individual of genotype x. We assume that $(X_{i1}, X_{i2})'$ has (perhaps after suitable transformation) a bivariate normal distribution. There are n_x complete pairs of data for expression and trait,

$$(X_{11}, X_{12}), (X_{21}, X_{22}), , (X_{n_x1}, X_{n_x2}),$$

and $N_x - n_x$ observations of trait only,

$$X_{(n_x+1)1}, X_{(n_x+2)1}, , X_{N_x1}.$$

Similarly, Y_{j1} and Y_{j2} are trait and expression measures of the jth individual of genotype y. We assume $(Y_{j1}, Y_{j2})'$ are bivariate normal with the same covariance matrix as $(X_{i1}, X_{i2})'$. There are n_y pairs of (Y_{j1}, Y_{j2}) and $N_y - n_y$ observations of Y_{j1} only. The difference between the gene expression means of the two genotypes, δ, is estimated by the maximum likelihood estimator, $\hat{\delta}$, with the likelihood function defined on both the complete data pairs of expression and trait and also the observations of trait only. Wang and Nettleton (2006) derived the following expression for the asymptotic variance of $\hat{\delta}$:

$$\mathrm{Var}(\hat{\delta}) \cong \left\{ \left(\frac{1}{N_x} + \frac{1}{N_y} \right) \left(\frac{(1-\rho^2)\sigma_{\mathrm{t}}^2}{(n_x+n_y)\sigma_{\mathrm{st}}^2} + \rho^2 \right) + \left(\frac{1}{n_x} + \frac{1}{n_y} \right) (1-\rho^2) \right\} \sigma_{\mathrm{e}}^2 + \frac{\{(\mu_{sx}-\mu_{sy}) - (\mu_x - \mu_y)\}^2 \sigma_{e|t}^2}{(n_x+n_y)\sigma_{\mathrm{st}}^2}, \quad (1)$$

where ρ denotes the within-genotype correlation between expression and trait; ${\sigma^2}_{\mathrm{t}}$ denotes the within-genotype trait variance; ${\sigma^2}_{\mathrm{st}}$ denotes the within-genotype trait variance of the selected individuals; ${\sigma^2}_{\mathrm{e}}$ denote the within-genotype expression variance; μ_x and μ_y denote the trait means for the two genotypes; μ_{sx} and μ_{sy} denote the trait means for

the selected individuals of the two genotypes and $\sigma^2_{\rm elt}$ denotes the variation of expression for individuals with a common trait value. The unknown parameters in Eq. (1) can be estimated from the data using the maximum likelihood method to obtain $\widehat{\rm Var}(\hat{\delta})$, an estimate of $\mathrm{Var}(\hat{\delta})$. The Wald statistic, $W = \hat{\delta}^2/\widehat{\rm Var}(\hat{\delta})$, is used to conduct a test for association between gene expression and QTL genotype that appropriately accounts for the process of selecting individuals for measurement with microarrays.

Examination of Eq. (1) suggests an optimal strategy for minimizing $\mathrm{Var}(\hat{\delta})$ and maximizing power to detect association between gene expression and QTL genotype. Note that only the parameters $\sigma^2_{\rm st}$, μ_{sx} and μ_{sy} are affected by the selection strategy. To minimize variance, it is clearly preferable to have

$$\{(\mu_{sx} - \mu_{sy}) - (\mu_x - \mu_y)\}^2 = 0, \qquad (2)$$

which can be achieved by equating $\mu_{sx} - \mu_{sy}$ to $\mu_x - \mu_y$. This indicates that the difference in trait means across genotypes for the selected individuals should be identical to the difference in trait means across genotypes for all individuals. Since $\sigma^2_{\rm st}$ should preferably be maximized at the same time, the optimal strategy involves selecting the individuals with the most extreme traits within each genotype class. Specifically, within each genotype class, we should select a equal number of individuals with the highest trait values as well as individuals with the lowest trait values. Note that this differs from the popular approach of selecting only the individuals with the most extreme trait values regardless of QTL genotype. That strategy is actually expected to perform poorly because the difference in means for selected individuals will be artificially exaggerated relative to the difference in means for all individuals.

It also should be noted that, for sufficiently large sample sizes, random selection within each QTL genotype class will be expected to perform nearly as well as the optimal selection strategy described above. Random selection satisfies the condition of Eq. (2), at the same time, the effect of the size of $\sigma^2_{\rm st}$ will be negligible if N_x and N_y are very large relative to n_x and n_y, as is easily observed by examining Eq. (1). Results from simulation studies in Wang and Nettleton (2006) show that random selection can achieve almost identical power when compared to the optimal strategy, and both have performance significantly better than tests using only microarray data on those $n_x + n_y$ individuals.

To understand the genetic networks that underlie quantitative variation in the trait, it is also very important to discover genes whose expression is correlated with the trait after accounting for the known effects of the QTL on the trait. Many of these genes may have expression that is associated with QTL genotype, and would therefore be identified as important via the tests described above. Other genes, however, may have expression values that are correlated with the trait but unassociated with genotype at the QTL. Wang and Nettleton (2006) also developed a new approach for identifying genes in the second category by testing $H_0{:}\rho = 0$ versus $H_a{:}\rho \neq 0$. This method is analogous to the Fisher's transformation, but it also utilizes the trait information on extra individuals as opposed to using only individuals with both trait and expression data. Similar to the Wald test described earlier, we will show that this test is more powerful in testing for a non-zero correlation coefficient within genotypes when compared to tests using only individuals chosen for microarray experiments.

Data sets and analyses

A population of 420 RILs was generated from the cross between two genetically distant ecotypes of *A. thaliana*, Bay-0 and Shahdara (Loudet et al. 2002). A set of 38 physically anchored microsatellite markers was used to construct a genetic map from these RILs. Loudet et al. (2002) performed QTL mapping for flowering time under both long and short photoperiod, and two of the detected QTLs colocalize very precisely with *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) genes. Data sets for genomic markers and flowering time are accessible through the website http://www.genenetwork.org.

The microarray data sets used in the analysis were described in Kliebenstein et al. (2006). Of the total of 420 RILs, 148 RILs were included in the microarray experiment. For each biological replicate, five plants from each of the RILs were grown under short day conditions. At the end of growth period, all rosette leaves from three plants were harvested to extract RNA samples. Two biological replicates were processed for each RIL. Each RNA sample was used for cDNA synthesis, and biotinylated cRNA was synthesized and hybridized to an Affymetrix ATH1 GeneChip representing 22,810 *A. thaliana* genes. These microarray data sets are available at http://elp.ucdavis.edu. GeneChips corresponding to RILs with missing phenotype (flowering time) or genotype data were removed from the subsequent analyses. Data from the remaining GeneChips were normalized using the robust multichip average (RMA) method (Bolstad et al. 2003). The measurements from the two biological replicates of each RIL were averaged to give a single transcript measurement per gene and RIL. This process yielded a data set consisting of genotype and flowering time data for 398 RILs and expression levels for 142 of these 398 RILs.

Selective transcriptional profiling was carried out according to Wang and Nettleton (2006) with flowering time during short days as the classical trait of interest and

the flowering time QTL on chromosome four identified by Loudet et al. (2002) as the locus of primary interest. Because the position of this QTL is very close to that of marker MSAT4.8, genotypes of MSAT4.8 were used to represent the QTL genotype. As shown in the simulation studies of Wang and Nettleton (2006), a small portion of mispredicted QTL genotypes has little effect on the test results.

Results

To assess the performance of the selective transcriptional profiling approach, the Wald test was performed on each gene represented on the Affymetrix ATH1 GeneChip, and the results were compared to those obtained by two-sample *t* test using only microarray data. For each gene, expression levels and flowering time for 142 RILs (106 with Bay-0 genotype, 36 with Shahdara genotype) as well as the flowering time for 256 additional RILs (141 with Bay-0 genotype, 115 with Shahdara genotype) were used to construct the Wald statistic following the method described by Wang and Nettleton (2006). The *t* test for each gene, on the other hand, was performed only on the expression levels of the 142 RILs for which microarray data were available. To further evaluate the effect of the proportion of individuals chosen for microarray experiments, we also randomly selected 36 RILs with Bay-0 genotype and 35 RILs with Shahdara genotype, respectively, among those with microarray data, and carried out the Wald test on expression levels and flowering time of these 71 lines plus the flowering time measurements on the remaining 327 lines. Two-sample *t* tests were also performed on the expression levels on these 71 lines. After conducting tests on every gene, the *q* value method described in Storey and Tibshirani (2003) was used to control the false discovery rate (FDR) at specified levels.

Table 1 shows the number of genes whose transcriptional abundance differs significantly between the two genotypes and are thus likely to be associated with the MSAT4.8 locus, based on the Wald test or the two-sample *t* test. It is clear that at every specified FDR level, the Wald test detected more significant genes than the two-sample *t* test. Specifically, the Wald test using expression level on 142 RILs and flowering time on all 398 lines detected all the genes that were detected by the *t* tests and, in addition, detected as much as 31% more significant genes (at FDR = 0.01). More prominently, the Wald test detected 61% more genes using expression data on 71 lines and flowering time on all 398 lines (FDR = 0.01) in addition to all the genes detected by the *t* test performed on the expression levels of 71 lines only. It is also notable that the number of genes detected by the Wald test using expression levels on 71 lines plus flowering time on 398 lines approached the number detected by the two-sample *t* test using microarray data on all 142 lines. Using only half the microarray resources, the Wald test found about 80% of the genes detected by the *t* tests plus additional associations close to 10% of the total detected by the *t* test. These results confirm the conclusion of Wang and Nettleton (2006) that the selective transcriptional profiling approach can achieve much greater power using the same number of microarrays by incorporating data on the classical trait of interest from additional individuals, or the approach can be used to achieve similar power using many fewer microarrays thus resulting in considerable savings in money and time.

Table 1 Comparing the power of the Wald test and two-sample *t* test

FDR	142 lines		71 lines	
	t test	Wald	*t* test	Wald
0.0005	112	145	76	105
0.001	131	163	88	117
0.005	207	255	112	169
0.01	255	334	142	229

The number of genes whose transcriptional abundance is associated with the marker MSAT4.2 according to the Wald test or two-sample *t* test is listed with the corresponding false discovery rate (FDR) level. The Wald test was performed on both microarray data (from 142 or 71 RILs) as well as the flowering time measurements of all 398 RILs; the *t* test was performed on microarray data only

Moreover, we applied the popular approach of selecting only the RILs with the most extreme trait values (36 with the longest flowering time, 35 with the shortest flowering time) regardless of QTL genotype. Two-sample *t* tests performed on the expression levels of these 71 lines detected less than 62% of the significant genes detected by *t* tests on the same number of randomly selected lines. Wald tests using the expression data on these 71 lines and the flowering time on all 398 lines performed essentially the same as the *t* tests (results omitted). This confirmed the observation in Wang and Nettleton (2006) that selecting individuals with the most extreme trait values without taking into account the genotype is in fact undesirable.

We also used methods described in Wang and Nettleton (2006) to detect genes whose expression level is correlated with the trait after accounting for the effects of the QTL at the MSAT4.2 locus. Table 2 compared the performance of the method of Wang and Nettleton (2006) and that of Fisher's *z* test, which uses data only from individuals selected for microarray experiments. Again, when using the method of Wang and Nettleton (2006), more genes can be detected in addition to all the genes detected by Fisher's *z* test. Thus, by taking into account the flowering time information on plants not selected for microarray experiments, the selective transcriptional profiling approach can

Table 2 Testing for significant correlation coefficient within QTL genotype

FDR	142 lines		71 lines	
	Fisher	Wang and Nettleton	Fisher	Wang and Nettleton
0.0005	124	381	24	44
0.001	145	457	28	56
0.005	275	774	55	98
0.01	344	1,021	68	132

The numbers of genes whose transcriptional abundance is correlated with flowering time after accounting for the effect of the flowering time QTL [according to the method of Wang and Nettleton (2006) or Fisher's z test] are listed with the corresponding FDR levels. Wang and Nettleton's method utilizes both microarray data (from 142 or 71 RILs) as well as the flowering time measurements of all 398 RILs; Fisher's z test uses microarray data only

significantly increase the power for detecting genes whose expression level is correlated with the trait within the QTL genotype.

It is notable that genes whose expression was associated with the flowering time QTL and genes whose expression was correlated with flowering time within QTL genotype do not totally overlap. Table 3 shows the results for 12 genes with different patterns. For the first four genes (Group A) in Table 3, the transcriptional abundance is significantly associated with the flowering time QTL at MSAT4.2, and the within QTL genotype correlation coefficient between the transcription level and flowering time is also highly significant. For these genes, the difference between q values obtained using the t test and using the Wald test is very notable.

For the four genes in the middle (Group B) of Table 3, the within QTL genotype correlation coefficient is close to zero, though the association between the transcription level and the flowering time QTL is highly significant. In general, the difference between the q values obtained by using the two different methods for Group B genes tend to be smaller than those for Group A genes. This is consistent with the expectation that the Wald test will provide the most substantial increase in power when within QTL genotype correlation is large. Interestingly, one of the genes in this category is *FRI*, the gene generally thought to be the underlying gene for the QTL at this position. Another interesting fact is that, without accounting for the genotype, the correlation coefficient between the expression level of *FRI* and flowering time will be a significant 0.38, which demonstrates that a significant unconditional correlation might be due to the QTL effect on both trait and expression rather than correlation between trait and expression. Thus, in this case, we see little evidence that the expression level of *FRI* affects flowering time or vice versus. Rather, it appears that both flowering time and the expression level of *FRI* are associated with the flowering time QTL.

The last four genes (Group C) in Table 3 have large within QTL genotype correlation coefficients, but their transcription levels do not seem to be associated with the flowering time QTL. This suggests that testing within QTL genotype correlation is useful even for genes whose transcriptional abundance is not associated with the QTL. In addition, without conditioning on the QTL genotype, the estimated correlation coefficient will be smaller. These three distinctive patterns are also evident in Fig. 1, in

Table 3 Distinctive patterns of three gene groups

Group	Probe ID	Gene description	Correlation	q values		
				Wang and Nettleton	t test	Wald
A	267509	*AGL20*	−0.692	0.000	0.000	0.000
A	246864	*GA3*	0.504	0.000	0.000	0.000
A	247469	Expressed protein	−0.565	0.000	0.001	0.000
A	264510	*PIF3*	−0.554	0.000	0.003	0.000
B	255634	*FRI*	−0.032	0.569	0.000	0.000
B	255558	*GLB1*	0.054	0.504	0.000	0.000
B	254909	Dirigent family protein	−0.000	0.644	0.001	0.000
B	254903	*TUF*	−0.092	0.382	0.002	0.001
C	264954	Carboxy-PEP mutase	0.653	0.000	0.532	0.609
C	245087	Expressed protein	−0.569	0.000	0.287	0.168
C	250063	*CSA1*	0.594	0.000	0.257	0.291
C	259879	Putative calmodulin	0.577	0.000	0.616	0.678

The within-genotype correlation coefficients between expression levels and flowering time, as well as q values for Wang and Nettleton's test for correlation, the Wald test and two-sample t test for association with the flowering time QTL, are shown for 12 genes. These genes belong to three different groups as described in the text. The correlation coefficients and q values are rounded to three decimal places

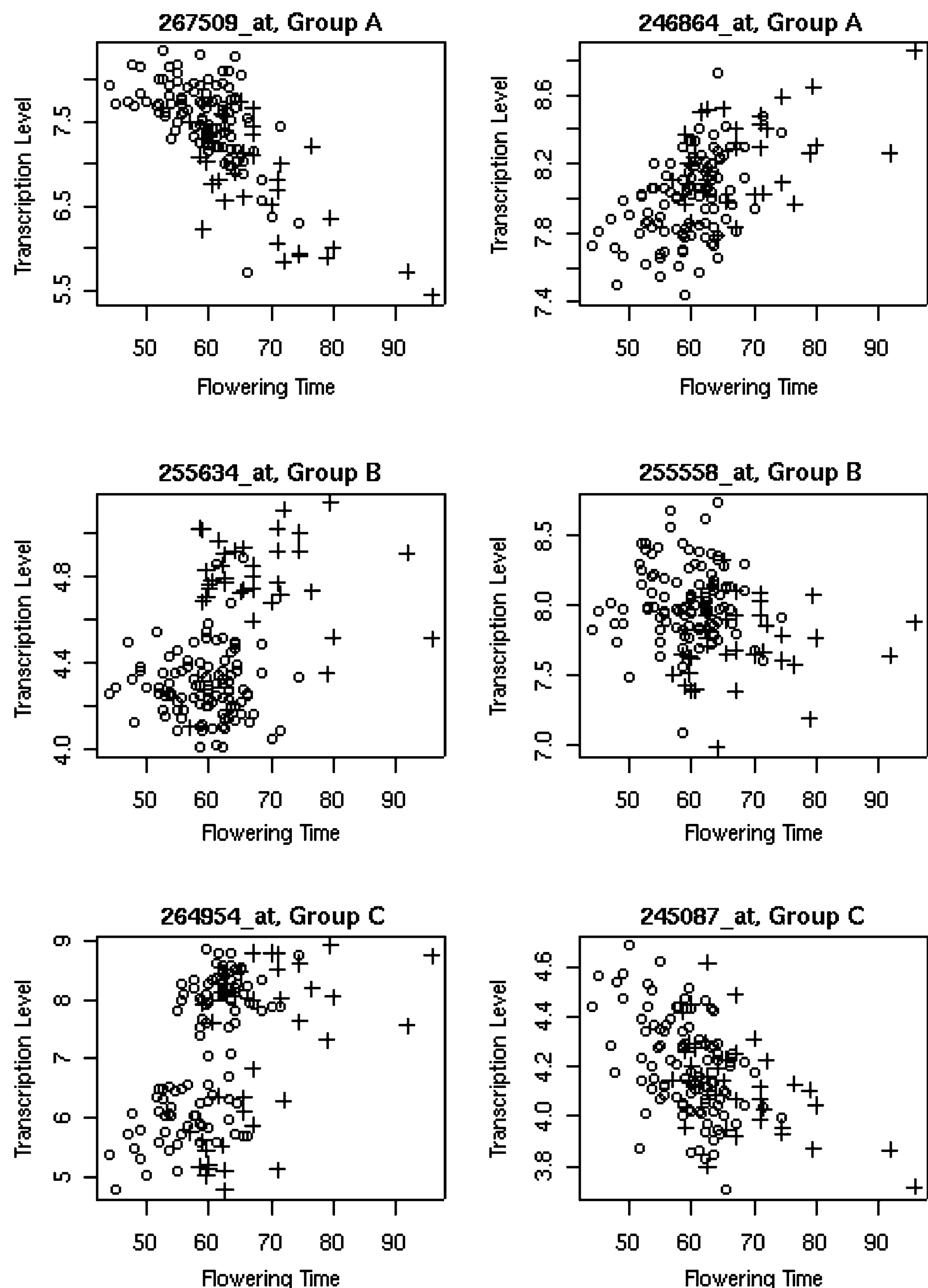

Fig. 1 Plots of gene expression levels and flowering time show different patterns regarding gene expression, QTL genotype, and trait. The transcription levels are plotted versus flowering time for six genes belonging to three groups described in the text. *Open circles* represent plants with the Bay-0 genotype at MSAT4.2, and *crosses* represent plants with the Shahdara genotype

which the transcription abundance is plotted versus flowering time for six genes in Table 3 using plants with microarray data.

Discussion

Combining QTL mapping and microarray analysis has great potential in providing detailed information on gene regulation networks. Recent studies on eQTLs treat transcript levels as quantitative traits, and have discovered eQTLs with multiple linkages colocalizing with classical QTLs associated with traits segregating in the population under study. As classical quantitative traits are often relatively inexpensive to obtain, the idea of selective transcriptional profiling, which uses classical trait information to enhance the power for detecting genes whose transcription level is associated with a classical QTL, is especially appealing when a researcher is interested in studying the transcriptional regulation mechanisms related to a specific trait. In this paper, we report the first application of selective transcriptional profiling on a real data set and confirm the superior power of

the selective transcriptional profiling approach. When microarray experiments cannot be performed on all plants, the selective transcriptional profiling approach can lead to detection power comparable to methods using only microarray data with considerably more microarrays. Though the results discussed here only involve one QTL with two genotypes, situations with more genotypes and/or more QTLs can be dealt with as discussed in Wang and Nettleton (2006), provided that there are enough individuals in each genotypic combination.

An interesting finding in our analyses is the existence of three distinctive groups of genes that are all of biological interest: genes whose expression levels are associated with the flowering time QTL and are also significantly correlated with flowering time after accounting for the QTL genotype, genes whose transcriptional abundance is associated with the flowering time QTL but not significantly correlated with the flowering time after accounting for the QTL genotype, and genes whose transcription levels are not associated with the flowering time QTL but are significantly correlated with flowering time within each QTL genotype. All of these three groups should be further considered in studying gene regulation networks for a classical trait. It would be of interest to determine whether there are distinct biological mechanisms associated with each of the three groups, which might provide insight into transcriptional regulation pathways.

The results presented in this paper also highlight the advantage of using recombinant inbred lines. Though ideally microarray experiments and the measurements of classical trait should be performed on the same plants or animals, we have shown that satisfactory results can be obtained with the selective transcriptional profiling approach using microarray data and classical trait data from different experiments performed by different researchers. Depositories such as genenetowork.org devote significant resources to collecting data on genotype, phenotype, and microarray measurements on recombinant inbred lines of various plant and animal species (Wang et al. 2003). This provides a ideal platform for applying the selective transcriptional profiling approach to analyze data from different sources in order to better understand gene regulation mechanisms.

Recently, a paper by Liang et al. (2007) provides an excellent example for combining QTL mapping methodology with microarray data analysis. In their study, recombinant inbred strains derived from crosses between B2 and D2 mice were used to identify a QTL on chromosome 3 associated with the size of the hematopoietic stem cell (HSC) population. Microarray experiments were then carried out to identify genes whose transcriptional abundance is associated with this QTL. The authors then focused on one of these genes, latexin (*Lxn*), which colocalizes with the QTL. By studying congenic strains and artificially overexpressing *Lxn* in marrow cells infected with retroviral vectors containing *Lxn*-GFP, the authors further confirmed that *Lxn* is the primary gene underlying the QTL associated with HSC numbers. The study of Liang et al. (2007) demonstrates the effectiveness of combining QTL mapping and microarray experiments, though it focused on identifying *cis*-acting eQTLs, and the QTL mapping and microarray data were analyzed separately. The methods described in this paper provide a natural way to combine classical trait data and microarray measurements for the detection of different patterns of transcriptional regulation.

Acknowledgments D. Wang acknowledges the financial support of National Science Foundation (Award Number 0701892).

References

Bolstad B, Irizarry RA, Astrand M, Speed T (2003) A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. Bioinformatics 19:185–193

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT et al (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using genetical genomics. Nat Genet 37:225–232

Chesler EJ, Lu L, Shou Y, Qu Y, Gu J et al (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 37:233–242

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in *Arabidopsis*. Genetics 172:1155–1164

Doss S, Schadt EE, Drake TA, Lusis AJ (2005) *Cis*-acting expression quantitative trait loci in mice. Genome Res 15:681–691

GuhaThakurta D, Xie T, Anand M, Edwards SW, Li G et al (2006) *Cis*-regulatory variations: a study of SNPs around genes showing cis-linkage in segregating mouse populations. BMC Genomics 7:235

Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H et al (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nat Genet 37:243–253

Hughes KA, Ayroles JF, Reedy MM, Drnevich JM, Rowe KC et al (2006) Segregating variation in the transcriptome: cis regulation and additivity of effects. Genetics 173:1347–1355

Jansen RC, Nap JN (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Kiekens R, Vercauteren A, Moerkerke B, Goetghebeur E, Van Den Daele H et al (2006) Genome-wide screening for *cis*-regulatory variation using a classical diallel crossing scheme. Nucleic Acids Res 34:3677–3686

Kliebenstein DJ, West MAL, van Leeuwen H, Loudet O, Doerge RW et al (2006) Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics 7:308

Lan H, Rabaglia ME, Stoehr JP, Nadler ST, Schueler KL et al (2003) Gene expression profiles of nondiabetic and diabetic obese mice suggest a role of hepatic lipogenic capacity in diabetes susceptibility. Diabetes 52:688–700

Liang Y, Jansen M, Aronow B, Geiger H, Van Zant G (2007) The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice. Nat Genet 39:178–188

Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. Theor Appl Genet 104:1173–1184

Luo ZW, Potokina E, Druka A, Wise R, Waugh R et al (2007) SFP genotyping from affymetrix arrays is robust but largely detects *cis*-acting expression regulators. Genetics 176:789–800

Nettleton D, Wang D (2006) Selective transcriptional profiling for trait based eQTL mapping. Anim Genet 37:s13–s17

Pastinen T, Hudson TJ (2004) *Cis*-acting regulatory variation in the human genome. Science 306:647–650

Rockman MV, Wray GA (2002) Abundant raw material for *cis*-regulatory evolution in humans. Mol Biol Evol 19:1991–2004

Schadt EE, Monks SA, Drake TA, Lusis AG, Che N et al (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Sieberts SK, Schadt EE (2007) Moving toward a system genetics view of disease. Mamm Genome 18:389–401

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. Proc Natl Acad Sci USA 100:9440–9445

Wang D, Nettleton D (2006) Identifying genes associated with a quantitative trait or quantitative trait locus via selective transcriptional profiling. Biometrics 62:504–514

Wang J, Williams RW, Manly KF (2003) WebQTL: web-based complex trait analysis. Neuroinformatics 1:299–308

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW et al (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. Genetics 175:1441–1450

Yvert G, Brem RB, Whittle J, Akey JM, Foss E et al (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat Genet 35:57–64

Zhang Z-Y, Ober JA, Kliebenstein DJ (2006) The gene controlling the quantitative trait locus epithiospecifier modifier1 alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. Plant Cell 18:1524–1536