Molecular Systems Biology 5; Article number 242; doi:10.1038/msb.2008.79
Citation: ***Molecular Systems Biology*** **5:242**

www.molecularsystemsbiology.com

molecular systems biology

Comprehensive analysis of *Arabidopsis* expression level polymorphisms with simple inheritance

Stephanie Plantegenet[1,7], Johann Weber[2,7], Darlene R Goldstein[3,4], Georg Zeller[5,6], Cindy Nussbaumer[1], Jérôme Thomas[2], Detlef Weigel[6], Keith Harshman[2] and Christian S Hardtke[1,*]

[1] Department of Plant Molecular Biology, University of Lausanne, Biophore Building, Lausanne, Switzerland, [2] Lausanne DNA Array Facility, Center for Integrative Genomics, University of Lausanne, Genopode Building, Lausanne, Switzerland, [3] École Polytechnique Fédérale de Lausanne (EPFL), Institut de mathématiques (IMA), Bâtiment MA, Lausanne, Switzerland, [4] Swiss Institute of Bioinformatics, Lausanne, Switzerland, [5] Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany and [6] Max Planck Institute for Developmental Biology, Department of Molecular Biology, Tübingen, Germany
[7] These authors contributed equally to this work
* Corresponding author. Department of Plant Molecular Biology, University of Lausanne, Biophore Building, Lausanne, CH-1015, Switzerland.
Tel.: +41 21 692 4251; Fax: +41 21 692 4195; E-mail: christian.hardtke@unil.ch

Received 4.9.08; accepted 18.12.08

In *Arabidopsis thaliana*, gene expression level polymorphisms (ELPs) between natural accessions that exhibit simple, single locus inheritance are promising quantitative trait locus (QTL) candidates to explain phenotypic variability. It is assumed that such ELPs overwhelmingly represent regulatory element polymorphisms. However, comprehensive genome-wide analyses linking expression level, regulatory sequence and gene structure variation are missing, preventing definite verification of this assumption. Here, we analyzed ELPs observed between the Eil-0 and Lc-0 accessions. Compared with non-variable controls, 5′ regulatory sequence variation in the corresponding genes is indeed increased. However, ~42% of all the ELP genes also carry major transcription unit deletions in one parent as revealed by genome tiling arrays, representing a >4-fold enrichment over controls. Within the subset of ELPs with simple inheritance, this proportion is even higher and deletions are generally more severe. Similar results were obtained from analyses of the Bay-0 and Sha accessions, using alternative technical approaches. Collectively, our results suggest that drastic structural changes are a major cause for ELPs with simple inheritance, corroborating experimentally observed indel preponderance in cloned Arabidopsis QTL.
Molecular Systems Biology 17 February 2009; doi:10.1038/msb.2008.79
Subject Categories: functional genomics; plant biology
Keywords: Arabidopsis; eQTL; expression level polymorphism; heritability of expression; QTL; structural gene variation

Introduction

Recent advances in high throughput technologies have had a major impact on quantitative genetic analyses, enabling the interrogation of whole genomes for characteristics such as, gene expression levels, single nucleotide polymorphisms (SNPs) or structural genome variation (Keurentjes *et al*, 2008). Among these approaches, microarray-based discovery of genetically controlled gene expression level differences has identified numerous expression quantitative trait loci (eQTL) in humans and model organisms (Brem *et al*, 2002; Morley *et al*, 2004; Doss *et al*, 2005; Li *et al*, 2006; West *et al*, 2007; Stranger *et al*, 2007b; Potokina *et al*, 2008). eQTL can be divided principally into two classes (Gibson and Weir, 2005; Rockman and Kruglyak, 2006; Hansen *et al*, 2008). *Trans*-acting eQTL (*trans*-eQTL) control the expression of other loci, whereas *cis*-acting eQTL (*cis*-eQTL) coincide with the loci whose expression varies. The latter represent ~20–50% of eQTL in various systems (Morley *et al*, 2004; Li *et al*, 2006; Stranger *et al*, 2007b; Potokina *et al*, 2008) and, additively, often explain significant portions of observed phenotypic variability (Li *et al*, 2006; Petretto *et al*, 2006; Keurentjes *et al*, 2007; Wentzell *et al*, 2007; Stranger *et al*, 2007b).

In this study, we focused on expression level polymorphisms (ELPs) that are already observed between parental lines and display simple, single locus inheritance. Such loci constitute a highly heritable subset of *cis*-eQTL and, because of their simple inheritance, can be exploited as markers (Doss

et al, 2005; Petretto *et al*, 2006; West *et al*, 2006, 2007; Keurentjes *et al*, 2007; Stranger *et al*, 2007b; Potokina *et al*, 2008). They can, for instance, replace SNPs in genotyping, a particularly interesting application in systems with poorly characterized genomes (West *et al*, 2006; Potokina *et al*, 2008). Despite the abundance of ELPs with simple inheritance, little is known about their molecular basis. In principle, they could represent *trans*-eQTL that are tightly linked to the locus they control, and this scenario might account for a significant fraction of heritable ELPs in large, complex genomes that are difficult to analyze at high resolution. Generally, however, it appears more likely that ELPs with simple inheritance represent large effect *cis*-acting polymorphisms in individual genes (Ronald *et al*, 2005; Stranger *et al*, 2007b; Hansen *et al*, 2008). These could include polymorphisms that affect gene expression at the transcriptional, post-transcriptional or post-translational level. For instance, mutations might alter transcript stability, or activity of the encoded protein, which could in turn affect RNA levels in cases of auto-regulatory feedback. Generally, however, *cis*-eQTL and thus ELPs with simple inheritance are assumed to reflect sequence variation in regulatory elements of the corresponding genes (Jansen and Nap, 2001; Cowles *et al*, 2002; Schadt *et al*, 2003; Pastinen and Hudson, 2004; Ronald *et al*, 2005; Williams *et al*, 2007), although only few studies have addressed this issue systematically (Cowles *et al*, 2002; GuhaThakurta *et al*, 2006).

Somewhat counter to the idea that regulatory polymorphisms are major determinants of phenotypic variability, in *Arabidopsis thaliana*, quantitative trait locus (QTL) cloning over the last years has often identified knockout mutations that affect the transcript and/or protein as the underlying molecular cause (e.g. Aukerman *et al*, 1997; Grant *et al*, 1995; Johanson *et al*, 2000; Kliebenstein *et al*, 2001; Kroymann *et al*, 2003; Kroymann *et al*, 2001; Mouchel *et al*, 2004; Werner *et al*, 2005). Even if many of these loci represent ELPs, generally, a preponderance of indels, whether in regulatory or transcript regions, is observed among these drastic mutations (Koornneef *et al*, 2004). However, because of the considerable sequence polymorphisms distinguishing naturally occurring isogenic Arabidopsis strains (so-called accessions), identification of the precise change underlying a QTL is often difficult, and structural changes are the easiest to discover. Thus, the successful reports of QTL isolation might reflect a bias in the ease with which such changes are detected. Indeed, recent studies that exploited recombinant inbred line (RIL) populations created from Arabidopsis accessions have identified numerous eQTL by microarray analyses (Keurentjes *et al*, 2007; West *et al*, 2007), including a varying portion of *cis*-eQTL. Among the *cis*-eQTL, a sizable fraction of loci represented parental ELPs with simple inheritance, which are strong QTL candidates to explain morphological or physiological variation between the parental lines. In this study, we analyzed the molecular basis of such ELPs in greater detail, by comprehensive comparison of gene expression, sequence variation and gene structure. Corroborating the experimental evidence from published reports of QTL cloning, we found again a preponderance of sizable indels, suggesting that QTL representing more subtle regulatory polymorphisms might be less common than anticipated.

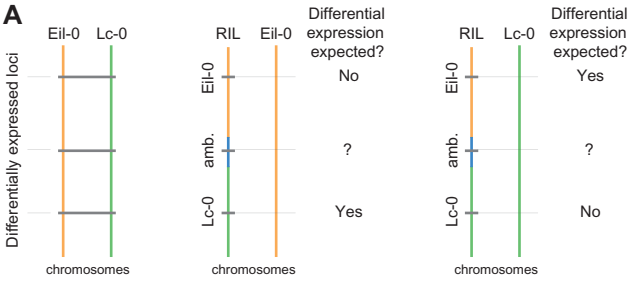
Results and discussion

Expression level polymorphisms between the Eil-0 and Lc-0 accessions

To identify parental ELPs, we determined transcript level variation between Eil-0 and Lc-0 seedlings by microarray analyses. Arrays based on short oligonucleotide probes are particularly sensitive to SNPs in parental transcripts, resulting in spurious eQTL and overestimation of *cis*-eQTL (Doss *et al*, 2005; Alberts *et al*, 2007), although this appears to depend on various factors, such as array design (Luo *et al*, 2007). In the absence of detailed genomic sequence information on the Eil-0 and Lc-0 accessions, in this study, we used arrays based on gene-specific probes of 150–500 bp lengths (Allemeersch *et al*, 2005). Genomic DNA hybridizations have previously shown that such two-color arrays are largely insensitive to potential hybridization efficiency biases introduced by minor sequence polymorphisms (Keurentjes *et al*, 2007). Moreover, they also offer the advantage of direct sample comparison, allowing immediate ELP assessment rather than ELP inference from statistical comparison of single sample hybridizations of oligonucleotide-based arrays (West *et al*, 2006). Nevertheless, in this study, we chose to follow established recommendations for the analysis of two-color arrays, which includes a statistical component (Shi *et al*, 2006). Based on duplicate dye swap comparison of three independent RNA samples, 499 ELPs ($P<0.005$ with Benjamini–Hochberg false discovery rate multiple testing correction and fold change $\geqslant 2$) representing 480 genes distributed across the genome were observed (Supplementary Table 1). Comparable numbers of parental ELPs have been found for other pairs of Arabidopsis accessions (West *et al*, 2006; Keurentjes *et al*, 2007).

Determination of expression level polymorphisms with simple inheritance by microarray analysis of recombinant inbred lines

To determine which of these ELPs show simple inheritance over several generations, we took advantage of a RIL population that had been derived by single-seed descent over seven generations, starting with F2 individuals from an Eil-0 (♀) × Lc-0 (♂) cross (Sibout *et al*, 2008). Notably, it was evident from earlier studies that detection of parental ELPs with simple inheritance does not require full-scale eQTL analysis of RIL populations, as they represent a subgroup of *cis*-eQTL that display firm allele-dependent inheritance of differential expression through all generations starting from the parents. Thus, dye swap comparisons between a few RILs and their two parents in microarray analyses were sufficient for their detection (Figure 1A). The RILs were chosen to represent the genetic diversity of the population based on genotyping data from 79 segregating genome-wide SNP markers (Warthmann *et al*, 2007; Sibout *et al*, 2008) (Supplementary Table 2), such that each locus would be derived typically from the same parent in at least three RILs. Thus, seven RILs were chosen for detailed analyses. RNA from these lines was hybridized against RNA from either parent in a dye swap layout. To assess the heritability of parental ELPs, we compared expected and observed differential expression,



B

		Observed/predicted expression level difference			
Hybridized against:		Lc-0		Eil-0	
Diff. expr. expected:		Yes	No	Yes	No
Recombinant inbred line	EL 1024	231/327 (70.6%)	72/79 (91.1%)	51/79 (64.6%)	298/327 (91.1%)
	EL 1050	117/229 (51.1%)	151/169 (89.4%)	94/ 69 (55.6%)	204/229 (89.1%)
	EL 1088	131/176 (74.4%)	172/186 (92.5%)	128/186 (68.8%)	158/176 (89.8%)
	EL 2011	79/147 (53.7%)	276/313 (88.2%)	132/313 (42.2%)	133/147 (90.5%)
	EL 2019	140/262 (53.4%)	180/198 (90.9%)	136/198 (68.7%)	233/262 (88.9%)
	EL 2040	182/322 (56.5%)	118/128 (92.2%)	79/128 (61.7%)	306/322 (95.0%)
	EL 2076	201/318 (63.2%)	114/120 (95.0%)	69/120 (57.5%)	295/318 (92.8%)
	Average:	60.4% ± 3.5	91.3% ± 0.8	59.9% ± 3.5	91.0% ± 0.8

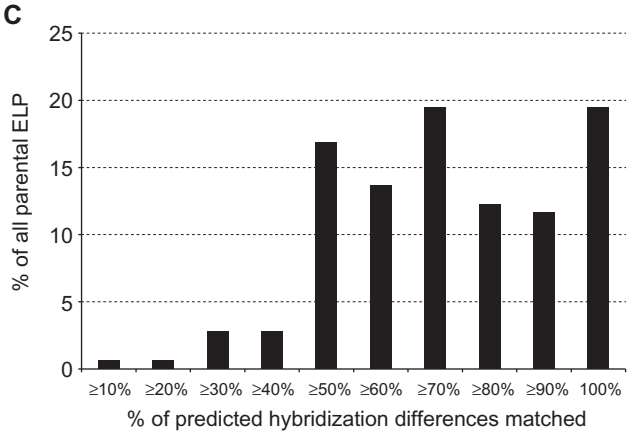


Figure 1 Assessment of ELP heritability by microarray analyses. RIL from a cross between Eil-0 and Lc-0 were genotyped with a set of 79 genome-wide SNP markers (Warthmann *et al*, 2007), defining the parental origin of chromosome segments. (**A**) Principles for the assessment of the heritability of ELPs observed between the Eil-0 and Lc-0 parents. Genotyped RIL from the S6 generation were compared with both parents in dye swap replicates. Based on the RIL genotype for a particular chromosome segment as determined by the flanking SNP markers, differential expression of a parental ELP locus on this chromosome segment was not expected in hybridizations of the RIL against the parent from whom the segment was inherited. However, differential expression (>2-fold) was expected in hybridizations against the other parent. ELPs located in regions of ambiguous genotype, i.e. heterozygous regions or segments spanning recombination breakpoints, were omitted from the analysis of that particular RIL. (**B**) Summary of parental ELP behavior in the hybridizations of the seven RILs (EL lines) against the two parent lines based on the principles outlined in (A). (**C**) Percentage of parental ELPs matching predictions across all RIL-parent hybridizations at a given frequency (100 or the 10% intervals below).

taking into account the genotyping data (Figure 1A). On average, ~60% of predicted ELPs were recovered in a given RIL versus parent hybridization (Figure 1B, Supplementary Table 3), similar to proportions found in other studies (Keurentjes *et al*, 2007). The absence of differential expression was an even better predictor, matching ~91% of observations. This discrepancy is likely due to the fact that the 2-fold change in expression represents a rather stringent but also arbitrary selection criterion. Overall, predictions of presence and absence of differential expression matched better, if data were treated according to 5% false discovery rate. However, as an extensive study of two-color microarray hybridizations recommended a 2-fold change in conjunction with false discovery rate for scoring differential signals (Shi *et al*, 2006), we used the analysis of our data according to those criteria, as the baseline in the following. Similar to earlier studies (West *et al*, 2006; Keurentjes *et al*, 2007), the parental ELPs could be used for RIL genotyping, delivering higher resolution than the SNP data (Supplementary Figure 1).

Comparison of the patterns of individual genes corresponding to parental ELPs, across all RIL hybridizations, enabled us to classify them according to the frequency at which predictions were met. This analysis identified a group representing ~20% of parental ELPs that perfectly matched predictions (Figure 1C, Supplementary Table 4) and, thus, can be considered to have simple *cis*-inheritance. Notably, as many of the other loci frequently missed our cutoff criteria for differential expression only narrowly, particularly the 2-fold criterion (see above), this is a conservative estimate. Overall, ELPs whose hybridization pattern matched 80% of predictions or more represented ~44% of all parental ELPs.

Sequence analysis of regulatory regions in genes representing ELPs with simple inheritance

To determine whether ELPs with simple inheritance are associated with increased sequence variation in regulatory regions, as observed in other systems (Cowles *et al*, 2002; GuhaThakurta *et al*, 2006), we compared a sample of 61 genes chosen from the ELP group that matched at least 90% of predictions with a control group of 85 genes that displayed very low variability and differential expression across all microarray experiments (see Methods). Notably, in Arabidopsis, regulatory elements controlling gene expression are generally found in the 5′ vicinity of the transcription start sites and the 5′ leader sequences (Lee *et al*, 2006). Thus, we isolated 1 kb fragments immediately upstream of the start codon for each of the sample and control group genes from both Eil-0 and Lc-0. Sequence information was obtained for ~44 kb of stably heritable ELP loci and ~62 kb of control loci (Supplementary Table 5, Supplementary sequence alignments). Sequence diversity between Eil-0 and Lc-0 was considerably higher in the ELPs with simple inheritance as compared with the control group (Figure 2A, Supplementary Table 5). Overall, SNP frequency was increased >4.5-fold, indel number >4.7-fold and the number of bp affected by indels >9.0-fold (Figure 2B and C). Generally, SNPs were biased towards the promoter as compared with the leader sequences. These results support the idea that ELPs with

A

		Total bp	SNPs	Indels #	Indels bp	
Non-variable (85)	Total sequence	61'972	360	80	283	tot.
		729 ± 16	4.2 ± 0.7	0.9 ± 0.2	3.3 ± 1.0	av.
		774	2	1	1	me.
	Promoter sequence	44'830	319	58	249	tot.
		567 ± 26	4.0 ± 0.8	0.7 ± 0.1	3.2 ± 1.0	av.
		658	1	0	0	me.
	5' leader sequence	13'940	19	17	26	tot.
		188 ± 23	0.3 ± 0.1	0.2 ± 0.1	0.4 ± 0.2	av.
		101	0	0	0	me.
ELPs with simple inheritance (61)	Total sequence	44'397	1'177	273	1'848	tot.
		728 ± 18	19.3 ± 4.8	4.5 ± 1.2	30.3 ± 9.5	av.
		765	5	2	3	me.
	Promoter sequence	24'813	605	130	1'040	tot.
		577 ± 32	14.1 ± 3.7	3.0 ± 0.8	24.2 ± 11.6	av.
		630	3	1	1	me.
	5' leader sequence	6'427	93	25	264	tot.
		169 ± 29	2.5 ± 1.0	0.7 ± 0.3	7.0 ± 5.1	av.
		94	1	0	0	me.

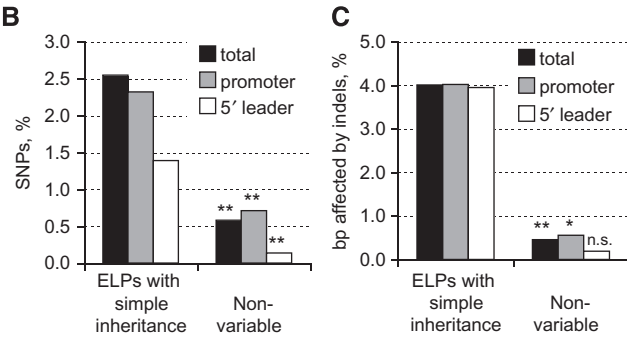


Figure 2 Sequence analysis of regulatory regions of a sample of 61 genes representing parental ELPs with simple inheritance and a control group of 85 genes, which displayed very low variability and differential expression (see Supplementary Materials and methods) in the array experiments ('controls'). For the ELPs with simple inheritance, only genes which perfectly matched predictions (see Figure 1C), and for which at least 10 precise predictions could be made (i.e. loci located in unambiguous chromosome segments in at least five RIL) were included. (**A**) Summary of sequence analyses of regulatory regions from 61 ELPs with simple inheritance and 85 control genes. Observed total absolute values (tot. line), per gene average values (av. line) and median values (me. line) are indicated. Note that numbers for promoter sequences and 5′ leader sequences do not add up to the total, because leader sequences were not defined for all genes investigated. (**B**) Relative abundance of SNPs (based on total sequence investigated). (**C**) Relative amount of bp affected by indels (based on total sequence investigated). Asterisks indicate *t*-test significance between the ELPs with simple inheritance and the control group (*$P<0.05$; **$P<0.01$; NS, not significant).

simple inheritance are associated with increased sequence diversity in the regulatory regions of the corresponding genes.

Genome tiling array analyses of the Eil-0 and Lc-0 genomes

Analyses of Arabidopsis genome variation have discovered unexpectedly high levels of accession-specific indels, which often impair gene function (Clark *et al*, 2007; Zeller *et al*, 2008). Such indels can, for instance, be identified by probing whole genome tiling arrays with genomic DNA (Hinds *et al*, 2006; Clark *et al*, 2007; Yazaki *et al*, 2007). As we failed to amplify the 5′ regions of at least one parent for ~34% of all loci initially targeted for sequencing in the ELP group and ~12% in the control group, we sought to determine whether this could be explained by indels. To this end, duplicate samples of genomic Col-0, Eil-0 and Lc-0 DNA were hybridized to Affymetrix Arabidopsis tiling 1.0R arrays, which represent the Col-0 genome as a tile of 25mer oligonucleotides with 10 bp spacing. Thresholds for detection of deletions (≥2.8-fold drop in hybridization signal over ≥35 bp, maximum allowed gap 150 bp) in Eil-0 and Lc-0 were determined empirically. This was done using deletions identified in the sequencing data (Figure 3). These threshold criteria consistently allowed detection of indels greater than ~30 bp, whereas at the same time ruling out the possibility that deletion calls could represent spurious differential signals because of SNPs or smaller indels (Figure 3) as detected in other studies (Li *et al*, 2006; West *et al*, 2006; Alberts *et al*, 2007; Borevitz *et al*, 2007; Clark *et al*, 2007). The genes representing parental ELPs as well as the control genes were inspected individually and only indels that were detected consistently in both replicate hybridizations were considered real. Even using these stringent criteria, numerous indels of various sizes were identified in both Eil-0 and Lc-0 (e.g. Figure 4; bar files for viewing tiling paths are provided in the Supplementary information). However, although ~42% of all parental ELP genes displayed indels when comparing their structure in Eil-0 versus Lc-0, only 9% of control genes did (Figure 5A; Supplementary Table 6); thus representing a >4-fold enrichment. Moreover, in the control group, deletions were usually small and affected mostly intron or leader sequences. As it appeared possible that the low expression variability of the control group genes could reflect the effect of purifying selection, we also analyzed a non-redundant random set of genes, which yielded essentially quantitatively similar results (Supplementary Table 6). By contrast, in the ELP group, generally multiple indels per gene were detected, and these were often larger and frequently affected exons. Moreover, in the ELP group, gene deletions (defined as uninterrupted deletion detection signal spanning >50% of the transcript region) were observed for nearly 10% of loci. Gene deletions were never observed in either control group.

The majority of genes representing ELPs with simple inheritance display uni-parental structural changes

Analysis of deletions according to ELP class with respect to matched predictions revealed a clear trend towards more severe indel types in ELP loci with simple inheritance. For instance, in the class of ELPs that perfectly matched predictions, 20% of loci displayed uni-parental gene deletions, whereas 25% of loci carried deletions in exons (Figure 5B). Still within the group of ELPs that matched at least 80% of predictions, the majority of loci displayed major uni-parental deletions. By contrast, the proportion of loci, for which no structural difference was observed between Eil-0 and Lc-0,

continuously increased in the parental ELP classes that matched predictions less and less faithfully, accompanied by a decrease in the severity of deletions observed. Thus, indels that are likely to impair or even abolish gene function appear to be much more frequent in genes representing ELPs, with simple inheritance, than in genes representing less heritable parental ELPs or invariable (or random) controls. These data suggest that the majority of ELPs with simple inheritance reflect a uni-parental impairment or even loss of gene function.

Importantly, in the vast majority of cases, over 90%, deletions were in phase with the direction of expression difference between the parents, such that the allele that carried deletions was expressed at a lower level. This observation would be consistent with the idea that the majority of deletions negatively affect gene function, thus leading to a loss of selection on gene maintenance and consequently gene expression. Supporting this notion, those parental ELPs that carried indels in their coding region also displayed a higher level of sequence variation in their 5′ regulatory regions (Figure 5C). However, the observation that alleles carrying deletions were expressed at lower levels could also simply reflect a difference in hybridization signal because of deletions in one allele. Although this appears likely for loci that displayed uni-parental gene deletion, this explanation might not be generally applicable to loci that carried partial deletions. Such loci might still yield detectable although potentially aberrant transcripts, even if those would not encode functional protein. In fact, deletions were not evident from our expression arrays, as documented by the signal strength distribution of parental ELPs, which resembles the one for all genes (Figure 5D–E). Moreover, as background noise is difficult to define in the two-color array hybridizations employed in our study, absence of hybridization signal is hard to establish, in particular for genes that are expressed at low levels (Czechowski *et al*, 2004). Finally, an earlier study used two-color arrays as well, and the authors entertained the notion that ELPs might reflect deletions (Keurentjes *et al*, 2007). To test this, they hybridized their arrays with competing genomic DNA from the parental accessions, Ler and Cvi-0, to identify a total of 159 indels. Of those, 14 coincided with *cis*-eQTL that mostly reflected ELPs with simple inheritance observed between the parents. However, as their study identified 922 parental ELPs, this would mean that there are either

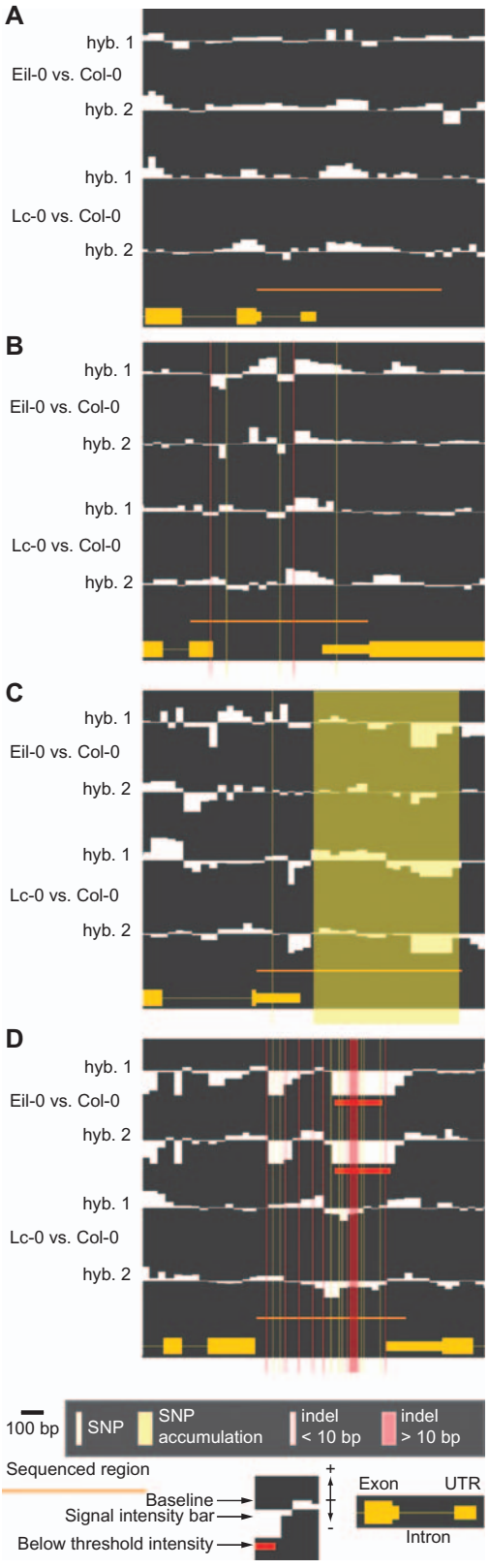


Figure 3 Genomic tiling array analysis of the Eil-0 and Lc-0 genomic DNA hybridized against a tile of the Col-0 genome. Two independent hybridizations were performed for each genotype. For classification of deletions, thresholds were determined by an empirical approach based on the promoter sequencing data described in Figure 2. The deduced settings of a signal drop below 2.8-fold (−1.5 on log2 scale), a minimum run >35 and for maximum gap $\leqslant 150$ allowed detection of indels >30 bp, but detected neither smaller indels nor SNPs. Examples are shown for tiles of individual sequenced regions. (**A**) Promoter region of At1g29030. No polymorphisms were observed among Lc-0 or Eil-0 as compared with Col-0 or each other. (**B**) Promoter region of At1g05830. Sequencing revealed a few dispersed small indels and SNPs between the three genotypes. (**C**) Promoter region of At1g13650. Sequencing revealed an extended stretch of many small indels and SNPs. (**D**) Promoter region of At1g33480. Sequencing revealed several small indels and SNPs. Only a 39 bp deletion in Eil-0 is picked up as a positive (red horizontal bars) by our settings. Gene structure is shown at the bottom of each panel (thick yellow blocks, exons; small yellow blocks, UTRs; yellow lines, introns). Difference in hybridization signal between Lc-0 or Eil-0 versus Col-0 along the oligonucleotides representing the tiling path is shown as white vertical bars. Upward deviation from the base line indicates positive hybridization signal, downward deviation negative hybridization signal.

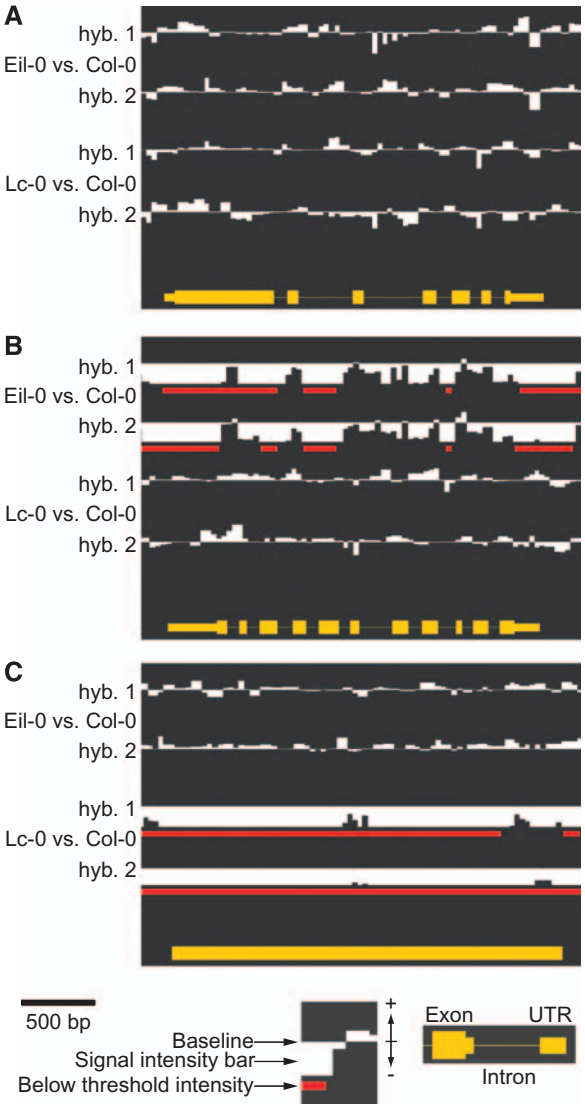


Figure 4 Indel analysis of the Eil-0 and Lc-0 genomes using genome tiling arrays. Genomic DNA of genotypes was hybridized against a tile of the Col-0 genome. Two independent hybridizations were performed for each genotype. Indels were deduced using threshold settings (signal drop $\leqslant$2.8-fold, min run $>$35, max gap $\leqslant$150) determined empirically as described in Figure 3. Examples are shown for tiles of individual genes. (**A**) At1g59900. No polymorphisms were observed in Lc-0 or Eil-0 as compared with Col-0 or each other. (**B**) At1g63900. Various deletions were detected in Eil-0 as compared with Lc-0 and Col-0. (**C**) At1g12220. A large-scale deletion likely covering the whole gene as indicated by a continuous detection bar was observed in Lc-0.

significantly fewer structural differences between the Ler and Cvi-0 genomes than between the Eil-0 and Lc-0 genomes, or that indels were underestimated as compared with our study.

Independent analysis of ELPs with simple inheritance between Bay-0 and Sha

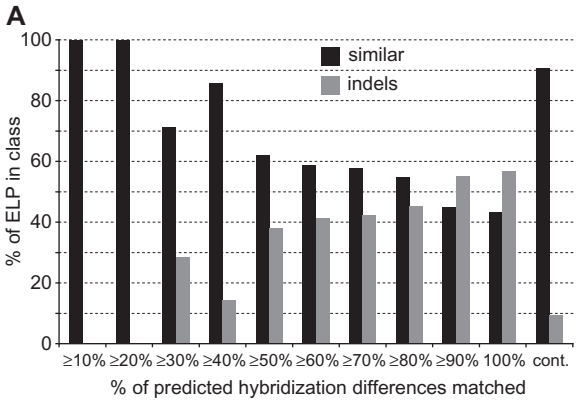
To corroborate independently the validity of our approach, we made use of other studies, in which expression differences between the Arabidopsis Bay-0 and Sha accessions were reported (West *et al*, 2006, 2007). Importantly, these data were extracted from full-scale single-feature polymorphism and e-QTL analysis of a population of more than 200 RILs, which, compared with our Eil-0xLc-0 analysis, was characterized using a different, short oligonucleotide microarray platform (Affymetrix) and a different conceptual approach to extract heritable gene expression differences. Thus, 187 genes representing parental ELPs with simple inheritance between Bay-0 and Sha were identified. We performed two independent hybridizations of genomic DNA of both Bay-0 and Sha to genome tiling arrays and analyzed the data as outlined above for Eil-0 and Lc-0. Again, we observed a strong preponderance of indels in the 187 ELPs with simple inheritance ($>$6-fold enrichment) as compared with the same control group used above (the two gene sets did not overlap; importantly, the control genes had been selected from the Eil-0xLc-0 analysis according to the indicated threshold criteria, but also according to the fact that they were monitored in all hybridizations, and that they were not part of gene expression markers in the Bay-0 $\times$ Sha eQTL analysis). (Figure 6A, Supplementary Table 7). Similar to our results for Eil-0 and Lc-0, the majority of the deletions in the ELPs with simple inheritance were observed at the level of exons ($\sim$33% of loci) or genes (18%) (Figure 6B).

The Bay-0 and Sha accessions were also part of a recent genome re-sequencing effort using high-density oligonucleotide arrays that interrogate SNP polymorphisms at every single base of the Arabidopsis genome (Clark *et al*, 2007). These data offered us the opportunity to independently verify our results. To this end, we analyzed the ELPs with simple inheritance and control group genes by a recently developed algorithm (Zeller *et al*, 2008) to identify polymorphic region predictions (PRPs), i.e. reduced hybridization signal over extended tracts of sequence. Such PRPs could result from an accumulation of SNPs or indels. Matching our tiling array analysis, PRPs were dramatically more frequent and generally more extended in the ELPs with simple inheritance as compared with the controls (Supplementary Table 7). This is illustrated by comparison of the combined PRP lengths in the Bay-0 versus the Sha alleles, which also revealed a marked asymmetry in PRP size in the genes representing ELPs with simple inheritance (Figure 6C), but not in the control genes (Figure 6D). In nearly all cases, increased PRP size matched the presence of deletions as detected by the tiling array approach.

Conclusions

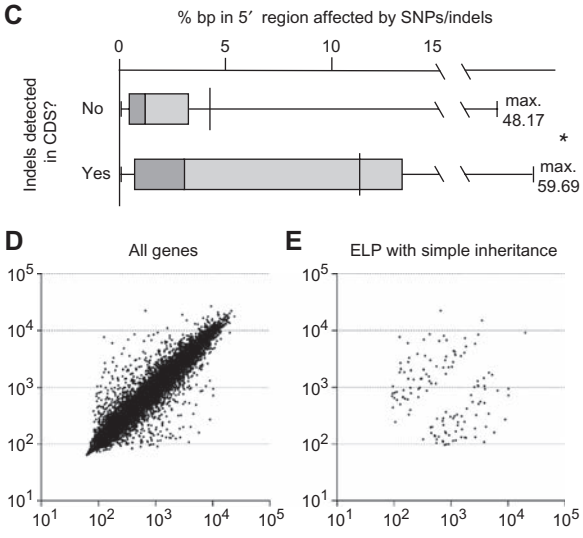
In summary, our data suggest that ELPs with simple inheritance in Arabidopsis primarily reflect the consequences of structural differences in the corresponding genes, rather than variation in regulatory elements, even if such a variation is observed. Notably, association of increased SNP variability and proximal deletions has also been observed in the human genome (Hinds *et al*, 2006). The large majority of deletions detected in ELPs with simple inheritance affected open reading frames or even complete genes, suggesting that they could frequently lead to loss of gene function. Moreover, we repeatedly observed major deletions of flanking regulatory

regions. Even if those deletions leave transcription units intact, they might lead to reduced or abolished gene expression, resulting in a *de facto* loss of gene function.



B

indels in Eil-0 vs. Lc-0 / ELP class	None	Introns	UTRs	Exons	Whole gene
Controls (97)	88 (91%)	4 (4%)	3 (3%)	2 (2%)	0 (0%)
=100% (97)	42 (43%)	2 (2%)	10 (10%)	24 (25%)	19 (20%)
<100, ≥90% (58)	26 (45%)	0 (0%)	8 (14%)	13 (22%)	11 (19%)
<90, ≥80% (62)	34 (56%)	3 (5%)	7 (11%)	15 (25%)	2 (3%)
<80, ≥70% (97)	56 (58%)	1 (1%)	9 (9%)	25 (26%)	6 (6%)
<70, ≥60% (68)	40 (59%)	2 (3%)	2 (3%)	21 (31%)	3 (4%)
<60, ≥50% (84)	52 (62%)	3 (4%)	7 (8%)	19 (23%)	3 (4%)
<50, ≥40% (14)	13 (87%)	2 (13%)	0 (0%)	0 (0%)	0 (0%)
<40, ≥30% (14)	9 (69%)	0 (0%)	2 (15%)	2 (15%)	0 (0%)
<30, ≥20% (3)	3 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<20, ≥10% (3)	3 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)



It remains to be determined whether Arabidopsis suffers from a particularly heavy mutational load because of inbreeding, as suggested before (Bustamante *et al*, 2002), or whether our findings apply more broadly. The similarity in ELP behavior across systems and the finding that copy number variation can explain significant portions of quantitative traits (Cutler *et al*, 2007; Stranger *et al*, 2007a) suggests that this could be the case. Finally, although functional variation in *cis*-regulatory elements contributes clearly to phenotypic variation (Bentsink *et al*, 2006; Rus *et al*, 2006; Sibout *et al*, 2008), large-effect changes that impact the integrity of transcribed regions should be considered as an equally valid explanation for expression variation. Indeed, such mutations have been shown to underlie phenotypic variation in natural strains of Arabidopsis (Grant *et al*, 1995; Aukerman *et al*, 1997; Johanson *et al*, 2000; Kliebenstein *et al*, 2001; Kroymann *et al*, 2001, 2003; Koornneef *et al*, 2004; Werner *et al*, 2005). Finally, the prevalence of indels in ELPs with simple inheritance mirrors the preponderance of indels with a drastic effect on gene integrity underlying cloned QTL, suggesting that the latter do not reflect a technical bias in the ease of detection. Thus, Arabidopsis QTL representing more subtle regulatory polymorphisms might be less common than anticipated.

Materials and methods

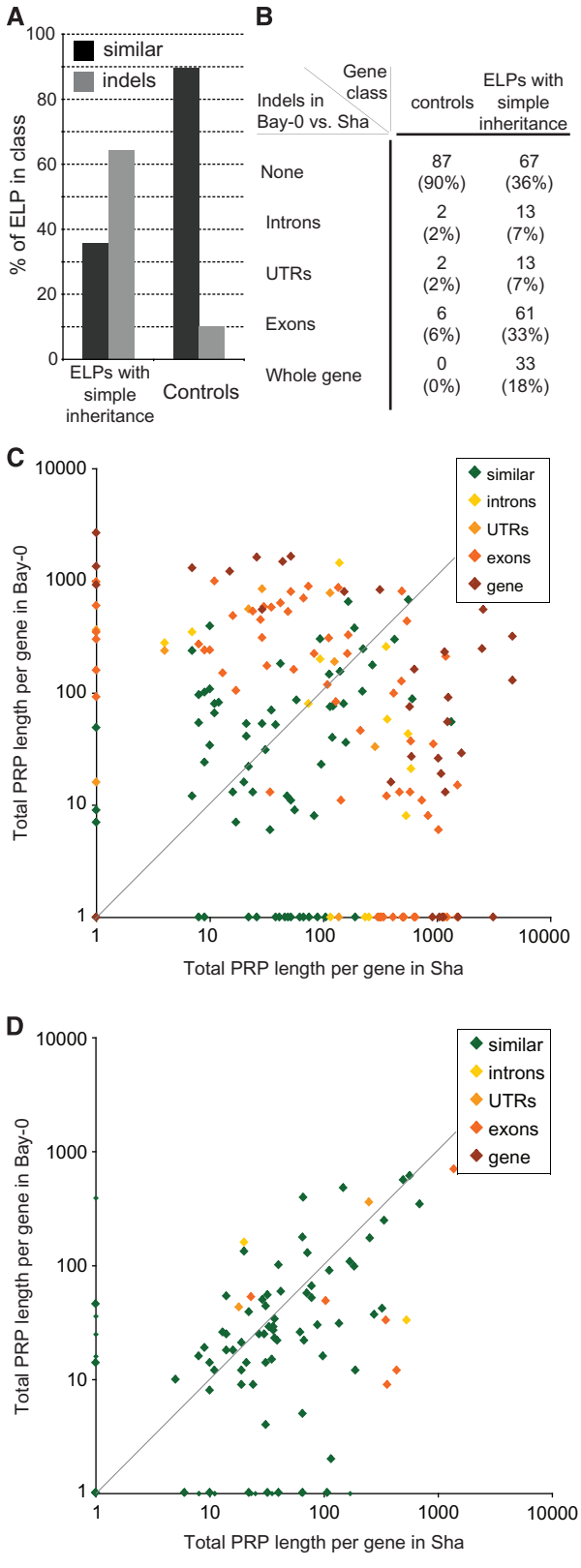
Plant materials

Seeds of Arabidopsis accessions were obtained from the Arabidopsis Biological Resources Center (Ohio State University, USA). Sterilized seeds were stratified for 48 h at 4°C, and seedlings were germinated and grown in tissue culture on a basic solid medium with macro and micronutrients (0.5 × MS) and 0.9% agar (Duchefa, the Netherlands), supplemented with 2% sucrose at 21°C under continuous light of 130 μE intensity. The Eil-0 × Lc-0 RIL population was derived from a cross between those parents in which Eil-0 served as the mother, after seven generations of single-seed descent starting from the segregating F2 generation (Sibout *et al*, 2008). Plant material for RNA analysis was harvested at 9 days after germination, typically from pools of 20 seedlings per line.

Figure 5 Summary of indel analyses. (**A, B**) Indel analysis of genes representing parental ELPs between the Eil-0 and Lc-0 accessions. (A) Correlation between strict ELP heritability (matching of hybridization predictions, see Figure 1C) and presence of deletions in the corresponding genes in one of the parents. Percentage of genes in each class displaying structural changes between Eil-0 and Lc-0 ('indels') or not ('similar'). Controls represent an extended group of 97 genes as described in Figure 2. (B) Detailed classification of the parental ELPs and controls shown in (A). None: no indels detected in Eil-0 as compared with Lc-0; introns: indel(s) detected in intron(s) of one parent as compared with the other; UTRs: indel(s) detected in UTR(s) or UTR(s) and intron(s) of one parent as compared with the other; exons: indel(s) detected in exon(s) or exons, UTR(s) and/or intron(s) of one parent as compared with the other; whole gene: >50% of gene deleted or duplicated in one parent as compared with the other. (**C**) Correlation between the presence of indels in the coding region and increased sequence variation in the corresponding 5′ regulatory regions in the parental ELP genes. The quartiles as well as the average (wider line) are indicated. The distribution between the two groups is statistically significant ($P<0.0390$, *t*-test). (**D**) Expression microarray hybridization signal distribution of all genes in the Eil-0 versus Lc-0 parent comparison. (**E**) As in (D), shown for the parental ELPs with simple inheritance.

SNP Genotyping

Genomic DNA of the EL RIL F5 population was isolated with Plant DNeasy™ kits (QIAGEN, the Netherlands) according to the manufacturer's instructions. Genotyping with a set of 289 SNPs was carried out by Genaissance Pharmaceuticals, Inc. (New Haven, CT, USA). Of those SNP, 79 were polymorphic between Eil-0 and Lc-0.



Microarray hybridizations

Total RNA was isolated using the Plant RNeasy™ kit (QIAGEN, the Netherlands) according to the manufacturer's instructions. Total RNA from the seedling pools was amplified using the MessageAmp™ aRNA II kit (Ambion, TX, USA). Five micrograms of amplified RNA were reverse transcribed into cyanin 3- or cyanin 5-labeled cDNA, purified with Qiaquick™ columns (Qiagen, the Netherlands) and hybridized on microarrays produced by the Lausanne DNA Array Facility (GEO accession number GPL6147) containing 25 000 gene-specific tags for the *A. thaliana* genome (Hilson *et al*, 2004). In order to analyze ELPs between the accessions Eil-0 and Lc-0, three independently grown seedling pools were analyzed by two-color co-hybridization of the labeled cDNAs in dye swap experiments, giving a total of six slides. These experiments can be found in the GEO database under entry GSE13628.

Statistical analyses

Statistical analyses of gene expression measures were carried out with open source R software packages available as part of the BioConductor project (http://www.bioconductor.org). Raw data from the microarrays were normalized by print tip lowess normalization (Yang *et al*, 2002), without applying background subtraction. To identify differentially expressed genes, we computed single gene moderated *t*-statistics (Smyth, 2004) using the limma package (Smyth, 2005). Genes were ranked according to their mod-t *P*-value and a cutoff was set at a maximum false discovery rate (Benjamini–Hochberg multiple testing correction, (Benjamini and Hochberg, 1995) of 0.005. From these genes, those with a minimum 2-fold expression difference between Eil-0 and Lc-0 qualified as parental ELP. For the analysis of RIL gene expression, each RIL sample was co-hybridized with each parent (Eil-0 and Lc-0) in a dye swap, resulting in two slides per parent versus RIL comparison. Genes with large mod-t and an expression difference of at least 2-fold in the RIL-parent comparison were considered as expressed differentially. In order to select genes that show no ELP (control genes), we selected genes, which had a maximum fold change of 1.3 in at least 13 out of 17 conditions tested (all RIL versus parent comparisons and Eil-0 versus Lc-0 comparisons). These genes were ranked according to their signal intensities and genes with an A-value <8 were excluded. From the remaining medium to high-intensity genes (134), 97 were selected for promoter and tiling array analysis.

Sequencing

For sequence analyses of regulatory elements, 1 kb fragments of 85 control genes and 65 stable ELP genes spanning the region 5′ to the start codon were isolated by PCR with KOD Hot Start Polymerase® (Novagen™) following the manufacturer's instructions. PCR-amplified fragments were purified using QiaQuick columns (Qiagen, the Netherlands) and sequenced by Macrogen Inc. (Republic of South Korea). Obtained sequences were analyzed using MacVector™ 7.2.2

Figure 6 Indel analysis and polymorphic region prediction (PRP) analysis of genes representing ELPs with simple inheritance and controls between the Bay-0 and Sha accessions. (**A**) Percentage of genes representing ELPs with simple inheritance or controls (same group Figure 5) that carry indels in one parent as compared with the other or that display similar gene structure. (**B**) Detailed classification of genes shown in (A), categories similar to Figure 5B. (**C, D**) PRP predictions. The graphs (logarithmic scale) represent total PRP size observed in a given gene (in bp, equaling sum of all individual PRPs with respect to the gene model Atxgyyyyy.1, TAIR 7.0 annotation) detected in one accession plotted against the same value for the other accession. Classification of genes is similar to Figure 5B. (C) Genes representing ELPs with simple inheritance. (D) Control genes.

software. The sequences have been submitted to the GenBank database (accession numbers FJ441298-FJ441589).

Tiling arrays

Genomic DNA was extracted from pools of three plants for each accession (Col-0, Eil-0, Lc-0 Bay-0, Sha) with Plant DNeasy™ kits (QIAGEN, the Netherlands) according to the manufacturer's instructions. Biotin-labeled target DNA was generated from this genomic DNA as described (Borevitz, 2006). Labeled targets were hybridized on Affymetrix GeneChip® Arabidopsis Tiling 1.0R Arrays and processed according to the supplier's protocols. CEL files were processed by Affymetrix tiling analysis software to generate normalized signal bar files. Tiling analysis software settings were quantile normalization and a bandwidth for probe analysis of 50 bp. To determine structural variations in the genomes of Eil-0, Lc-0, Bay-0 and Sha, two independent DNA isolates of each accession were compared with Columbia DNA. The resulting bar files were loaded into the Affymetrix integrated genome browser software and analyzed manually for the genes of interest. To qualify as deletions, the integrated genome browser signals had to be below cutoff—1.5 (log2 scale) and the settings for min run was >35 and for max gap $\leqslant 150$. These parameters were determined empirically (see text and Figure 3). The TAIR Arabidopsis genome annotation version 7.0 was used for analysis. The tiling array raw data have been deposited at the ArrayExpress database under accession number E-MEXP-1888.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We would like to thank Dr K Osmont for helpful comments on the manuscript, O Hagenbüchle and A Paillusson for Affymetrix tiling array hybridizations and E Farmer and P Reymond for the PCR products used to make the custom-spotted DNA microarrays. Contributions: CSH, KH, SP and JW conceived this study and analyzed the data together with DRG and GZ. CSH wrote the manuscript with help from KH, SP, JW, DRG and DW. Recombinant inbred lines were contributed by CSH and SP. All molecular biology experiments except microarray hybridizations were performed by SP. Microarray hybridizations were performed by CN and JT. Statistical analyses of microarray experiments were performed by DRG, JW and GZ. DRG was funded by the Swiss National Science Foundation National Centre for Competence in Research (Plant Survival). This work was supported by the University of Lausanne, by Swiss National Science Foundation Grant 3100A0-107631 to CSH and by the SystemsX 'Plant growth in a changing environment' project funding for CSH.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alberts R, Terpstra P, Li Y, Breitling R, Nap JP, Jansen RC (2007) Sequence polymorphisms cause many false cis eQTLs. *PLoS ONE* **2:** e622

Allemeersch J, Durinck S, Vanderhaeghen R, Alard P, Maes R, Seeuws K, Bogaert T, Coddens K, Deschouwer K, Van Hummelen P, Vuylsteke M, Moreau Y, Kwekkeboom J, Wijfjes AH, May S, Beynon J, Hilson P, Kuiper MT (2005) Benchmarking the CATMA microarray. A novel tool for Arabidopsis transcriptome analysis. *Plant Physiol* **137:** 588–601

Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, Amasino RM, Sharrock RA (1997) A deletion in the PHYD gene of the Arabidopsis Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant cell* **9:** 1317–1326

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57:** 289–300

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci USA* **103:** 17042–17047

Borevitz J (2006) Genotyping and mapping with high-density oligonucleotide arrays. *Methods Mol Biol (Clifton, NJ)* **323:** 137–145

Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT, Chen H, Werner JD, Nordborg M, Salt DE, Kay SA, Chory J, Weigel D, Jones JD, Ecker JR (2007) Genome-wide patterns of single-feature polymorphism in Arabidopsis thaliana. *Proc Natl Acad Sci USA* **104:** 12057–12062

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

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in Arabidopsis. *Nature* **416:** 531–534

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Scholkopf B, Nordborg M, Ratsch G, Ecker JR, Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science* **317:** 338–342

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. *Nat Genet* **32:** 432–437

Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD (2007) Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res* **17:** 1743–1754

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* **38:** 366–379

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

Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends Genetic* **21:** 616–623

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* **269:** 843–846

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

Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends Plant Sci* **13:** 72–77

Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, Chardakov V, Cognet-Holliger C, Colot V, Crowe M, Darimont C, Durinck S, Eickhoff H, de Longevialle AF, Farmer EE, Grant M *et al.* (2004) Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Res* **14:** 2176–2189

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* **38:** 82–85

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

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* **290:** 344–347

Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using

genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci USA* **104:** 1708–1713

Keurentjes JJ, Koornneef M, Vreugdenhil D (2008) Quantitative genetics in the age of omics. *Curr Opin Plant Biol* **11:** 123–128

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell* **13:** 681–693

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. *Annu Rev Plant Biol* **55:** 141–172

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. *Proc Natl Acad Sci USA* **100** (Suppl 2)**:** 14587–14592

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol* **127:** 1077–1088

Lee JY, Colinas J, Wang JY, Mace D, Ohler U, Benfey PN (2006) Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *Proc Natl Acad Sci USA* **103:** 6055–6060

Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC, Breitling R, Kammenga JE (2006) Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. *PLoS Genet* **2:** e222

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

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* **430:** 743–747

Mouchel CF, Briggs GC, Hardtke CS (2004) Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev* **18:** 700–714

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

Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, Fischer J, Maatz H, Kren V, Pravenec M, Hubner N, Aitman TJ (2006) Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet* **2:** e172

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008) Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J* **53:** 90–101

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. *Nat Rev Genet* **7:** 862–872

Ronald J, Brem RB, Whittle J, Kruglyak L (2005) Local regulatory variation in Saccharomyces cerevisiae. *PLoS Genet* **1:** e25

Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E, Salt DE (2006) Natural variants of AtHKT1 enhance Na + accumulation in two wild populations of Arabidopsis. *PLoS Genet* **2:** e210

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422:** 297–302

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM *et al.* (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24:** 1151–1161

Sibout R, Plantegenet S, Hardtke CS (2008) Flowering as a condition for xylem expansion in Arabidopsis hypocotyl and root. *Curr Biol* **18:** 458–463

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3

Smyth GK (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds). NewYork: Springer pp 397–420

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET (2007a) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315:** 848–853

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavare S, Deloukas P, Dermitzakis ET (2007b) Population genomics of human gene expression. *Nat Genet* **39:** 1217–1224

Warthmann N, Fitz J, Weigel D (2007) MSQT for choosing SNP assays from multiple DNA alignments. *Bioinformatics* **23:** 2784–2787

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ (2007) Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet* **3:** 1687–1701

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci USA* **102:** 2460–2465

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

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res* **16:** 787–795

Williams RB, Chan EK, Cowley MJ, Little PF (2007) The influence of genetic variation on gene expression. *Genome Res* **17:** 1707–1716

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30:** e15

Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol* **10:** 534–542

Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Ratsch G (2008) Detecting polymorphic regions in Arabidopsis thaliana with resequencing microarrays. *Genome Res* **18:** 918–929