



ELSEVIER

Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants

Francisco A Cubillos, Vincent Coustham and Olivier Loudet

Even if considerable progress has been achieved towards the understanding of natural variation in plant systems, the contribution of transcript abundance variation to phenotypic diversity remains unappreciated. Over the last decade, efforts to characterise the genome-wide expression variation in natural accessions, structured populations and hybrids have improved our knowledge of the contribution of non-coding polymorphisms to gene expression regulation. Moreover, new studies are helping to unravel the role of expression polymorphisms and their orchestrated performance. Recent advances involving classical linkage analysis, GWAS and improved eQTL mapping strategies will provide a greater resolution to determine the genetic variants shaping the broad diversity in plant systems.

Address

Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech, F-78000 Versailles, France

Corresponding author: Loudet, Olivier (Olivier.Loudet@versailles.inra.fr)

Current Opinion in Plant Biology 2012, **15**:192–198

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Yves Van de Peer and J. Chris Pires

Available online 20th January 2012

1369-5266/\$ – see front matter

© 2012 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.pbi.2012.01.005](https://doi.org/10.1016/j.pbi.2012.01.005)

Introduction

Phenotypic traits of ecological, agronomical and economical interest are polygenic, with many components outlining natural variation between individuals. One of the central goals in classical genetics is to understand how the distinct allelic variants can shape phenotypic diversity through multiple genetic and environmental interactions [1]. The advent of new genotyping technologies (e.g. arrays and DNA-sequencing) has heightened the mapping of quantitative trait loci (QTLs) [2]. Genome-wide association studies (GWAS) and linkage mapping strategies have identified countless numbers of QTLs in plants; however, only a small number of essentially large-effect loci originally detected as QTLs have been molecularly characterised [3,4]. Moreover, many of the resolved QTLs correspond to changes that dramatically alter the encoded protein sequence, probably because

transcriptional regulation differences are more difficult to identify, target and complement. The lack of direct approaches to relate gene expression variation to phenotypic variation has therefore left its actual contribution to natural variation unappreciated.

To date, most expression QTL (eQTL) mapping studies have aimed to resolve networks of genes in order to identify candidate genetic elements that contribute to complex phenotypes [5,6]. In contrast, only a handful of studies were intended to detect signatures of selection [7,8] and to understand the role of evolutionary forces on heritable expression traits [9]. In this review, we focus on genomic studies within the plant community that have assessed the role of non-coding regions as a source of natural phenotypic variation, demonstrating how the identification of transcriptional modules can aid the dissection of complex biological processes. In addition, as the understanding of the regulation of expression traits is still in its infancy, we provide an overview of some future approaches that could complement our knowledge of transcript abundance variation.

The genetic architecture of transcript accumulation

Gene transcription is a primary intermediate between the information encoded in the genome and the final phenotype. In this regard, variation in expression levels may be an important source of phenotypic variation [4,8]. A fundamental step towards understanding the effects of genetic variation in gene expression levels and its influence on trait adaptation is the characterisation of transcript abundance variation at the genome scale [7]. The detection of loci exhibiting differential expression levels between accessions was initially achieved in plants through the use of microarrays in the model system *Arabidopsis thaliana*, given its timely sequence availability [10]. A series of pioneer studies (reviewed elsewhere [11]) initially described thousands of eQTLs and demonstrated the polygenic character of expression traits in plants ([12,13]; Cubillos, Loudet *et al.*, unpublished results).

The advent of next generation sequencing technologies, genotyping approaches and the generalisation of transcriptome arrays has allowed genome-wide expression profiling within natural accessions and structured populations (RILs, Recombinant Inbred Lines [14]) to be carried out in several plant systems, many of which are industrially or economically relevant [6,15^{••},16,17].

Among other systems, eQTL studies were conducted in several important crops [16,18] and trees [19]. In agreement with findings in *Arabidopsis* [12,13,20^{*}] and other non-plant species [21], these studies reported that a large fraction of the expression differences segregating for a single gene was due to polymorphisms either within the gene's own sequence or near the encoded transcript (detected as local-eQTLs) and likely to act in *cis*. In contrast, some eQTLs mapped elsewhere in the genome (distant-eQTLs), and probably act in *trans*. Nevertheless, the influence of distal regulations is not trivial and numerous genes are controlled by distant-factors [22]. Although they explain a smaller fraction of the variation on a single transcript, approximately 70% of the eQTLs detected in maize [23], rice [6] and *Brassica rapa* [5] are distant. Furthermore, the distribution of distant-eQTLs along the genome is not random and many of them gather in hotspots, each comprising one or several regulatory loci [12]. Whether a hotspot is the consequence of many tightly linked loci or a master regulator controlling (directly or indirectly) the expression of many genes remains to be addressed. It could be expected that the relative effect of a *trans*-factor over a single transcript is minor compared to *cis*-changes, since sequence variation within large effect pleiotropic *trans*-factors affecting hundreds of transcripts may be deleterious in a population [23,24].

The rapid characterisation of polymorphisms between *Arabidopsis* accessions [25] provides an opportunity to individually consider larger sets of allelic variants. The utilisation of whole-genome genotyping arrays and sequencing data have proven fruitful for *Arabidopsis* GWAS at the species scale [20^{*}] and within reduced sets of accessions involved in complex crosses [15^{**}]. In contrast to eQTL studies in structured populations, GWAS for expression variation identified many more local associations than distant. Most of this potential *cis*-variation was attributed to polymorphisms within the promoter region and 5'-genic sequence [15^{**}]. Local variants exhibit, on average, moderate population allele frequencies [20^{*}], consistent with a reduced or neutral selection over *cis*-variants [26]. In contrast, *trans*-acting polymorphisms are found at lower frequencies within populations, which may be due to stronger selective forces [20^{*}]. Alternatively, this trend could be explained by a lack of resolution of GWAS for *trans*-regulators, for example due to epistasis or allelic/genetic heterogeneity.

Certainly, through the increased use of RNA-sequencing (RNA-seq) technologies and analysis pipelines, full-scale GWAS that use unbiased transcript quantification will soon be performed, expanding the repertoire of interacting local-acting and distant-acting variants. Moreover, local and distant effects have been extensively and independently quantified in all these studies; however, the principles by which polymorphic *cis*-elements and

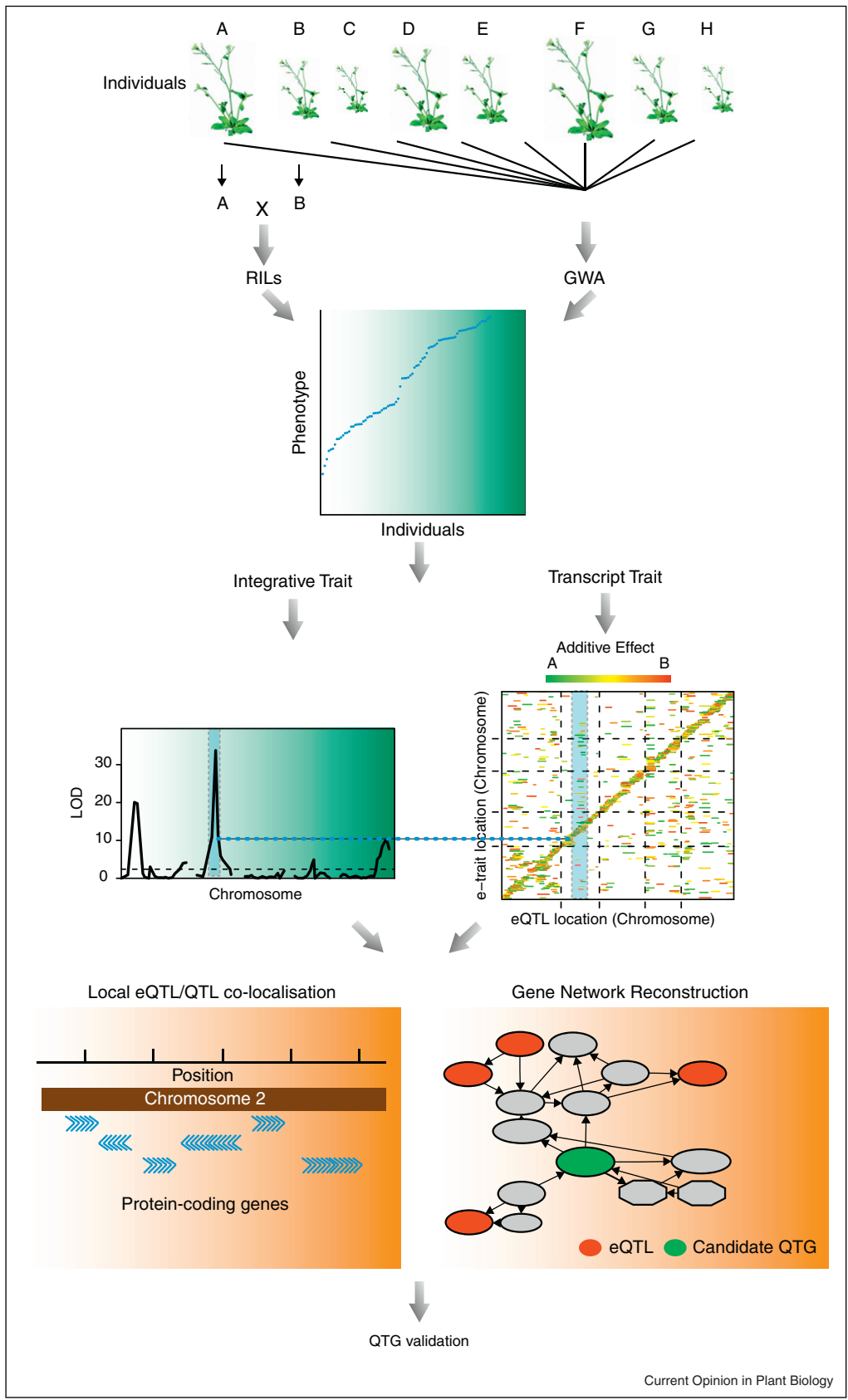
trans-factors interact and how transcript abundance variation is modulated still remain largely unknown.

Transcript abundance and phenotypic variation

While many QTLs accounting for large effect phenotypes have been shown to be the result of major alterations in gene expression such as knockdowns [4], it also became clear that more subtle allelic variation in gene expression is a major contributor to phenotypic evolution [27,28^{**},29^{*}]. The utilisation of comparative genomics has been successfully applied for the identification of polymorphisms within non-coding regions underlying phenotypic variation. For example, a single nucleotide polymorphism at the same position within a conserved *cis*-regulatory element in *Brassica* [27] and domesticated rice [30] is able to alter gene expression of the two orthologous genes (*REPLUMLESS* and *qSH1* in *Brassica* and rice, respectively) and consequently the control of seed shattering, suggesting convergent domestication processes between unrelated species. Similarly, expression variation at *TEOSINTE BRANCHED 1* (*TB1*), a major QTL candidate for apical dominance in maize, has been reported as a causal variant for this key step in the domestication of the species. This variation in expression is mediated by a transposable element inserted in an upstream regulatory region of *TB1* [31^{*}].

Several examples of natural variation in quantitative gene expression demonstrated a strong adaptive value in response to endogenous or external stimuli. Plant responses to fungal diseases were shown to vary quantitatively in barley [32]. The *ACTIN DEPOLYMERIZING FACTOR 3* (*ADF3*) localised proximal to the main QTL involved in fungal resistance, was previously excluded as a candidate given the lack of non-synonymous substitutions within the coding sequence. Nonetheless, a later eQTL study in the same barley lines identified *ADF3* as a major candidate mediating resistance to *Puccinia graminis*, due to a lower expression in resistant lines, validating the role of strong expression polymorphisms [17]. A direct link between growth and gene expression has been confirmed in rice, in which grain size and yield of different varieties is dependent on *GS5*, a putative serine carboxypeptidase [29^{*}], although the gene's transcript accumulation showed complex genotype \times tissue (or stage) interactions. The authors identified three haplotypes (corresponding to sequence variation in the promoter of *GS5*) resulting in different levels of expression that are positively correlated with these traits of agronomic importance. In *A. thaliana*, the *cis*-regulatory allelic variation of a Na⁺ transporter named *ATHKT1;1* has been linked to differential salinity tolerance, suggesting local adaptation to coastal and/or saline soils in Europe [28^{**}]. Similarly, allelic contribution to gene expression variation was shown in *Arabidopsis* for *FLOWERING LOCUS T* (*FT*) and *FLOWERING LOCUS C* (*FLC*), two key

Figure 1



regulators in the network of pathways that control flowering initiation. Many QTL analyses reported that these two genes often contribute to the variation in flowering time, and most of the natural allelic variation seems to affect their expression levels, allowing a fine-tuning of the flowering response [33,34]. Interestingly, *FT* and *FLC* genes are under epigenetic regulation [34,35] that could possibly be influenced by the sequence variation of these polymorphic non-coding regions. Further evidence for the role of epigenetic polymorphism was detected between economically important *Populus* hybrids [36]. Clones located in different locations exhibited distinct genome-wide transcriptional patterns in response to drought, potentially due to divergent epigenomes as a way of adaptation to the local environment. Sequence variation could, therefore, underlie differential chromatin states leading to changes in expression patterns and ultimately phenotypic differences between accessions.

Transcriptional networks as a tool for dissecting complex traits

The mapping of large sets of differentially expressed alleles is essential for understanding complex phenotypes and can support the construction of complex regulatory networks, particularly when a number of genes are controlled by a common eQTL [13,37,38]. Additionally, the integration of eQTL and QTL data can help to identify candidate genes within QTLs for classical phenotypes (Figure 1). Although *cis*-regulation exhibits substantial robustness in different tissues and organs, *trans*-factors play a significant role especially in response to distinct environmental conditions [19,39–41]. This was demonstrated using distant-eQTL hotspots in *Populus* through the identification of tissue-specific transcriptional networks, demonstrating their role in differential plant development and phenotypic variation [19]. Similarly, the presence of co-localising eQTL and QTL hotspots in *Arabidopsis* has been associated with pleiotropic genes that can act as network hubs [42]. Different studies conducted in this model organism demonstrated that the utilisation of eQTL databases could also lead to the identification of master regulatory loci and genes. For example, the identification of polymorphisms between two accessions in the *trans*-factor *ELF3* revealed the differential expression of related genes and its link with shade avoidance [43]. Similarly, the utilisation of co-expression networks in combination with GWAS can help to decrease the high number of false-positive associations usually detected by this approach and improve the mapping of correct candidate genes [44]. These studies emphasise the need to study genes as transcriptional

modules [45]. The identification of co-expressed allelic variants with minor individual effects can also provide novel insights into the biological processes underpinning a specific trait, stressing the need for further eQTL studies seeking the identification of large transcriptional units.

Current advances and future directions

Given the large availability of genomic tools, the study of transcript abundance towards the understanding of natural variation can now be performed in a wide range of plant species. Over the past 10 years microarray platforms have been at the forefront of gene expression quantification studies and were successfully used for the identification of eQTLs. However, this technique presents several drawbacks such as, insufficient sensitivity, a lack of reproducibility, and susceptibility to be misled by sequence and structural variation. The swift improvement of high-throughput sequencing technologies for large-scale transcriptome analyses provides a means of eliminating microarray deficiencies and extending the analysis beyond the expected reference transcriptome [46]. Although RNA-seq sensitivity is highly dependent on the sequencing depth and its associated costs [47], this strategy has already proven fruitful in a series of organisms, including maize [48] and Brassica [49]. The sequencing of 1001 *Arabidopsis* accessions and several of its relatives (*Arabidopsis lyrata*, *Capsella rubella*, *B. rapa*, among others) is now starting to yield high resolution data, providing evidence on how sequence variants within coding and non-coding regions affect gene functions [15,25,27,50]. The clear benefit from gathering huge quantities of genome sequences relies on the opportunity to infer the impact of gene expression variation from a wide genetic basis. Although DNA regulatory sequences are difficult to recognise since they tolerate a high degree of sequence degeneration while preserving robustness of expression [41], there are several approaches that could optimise the study of non-coding regions in plants and their link with phenotypic variation. Obtaining the transcriptome of multiple individuals within a species allows the identification of differentially expressed genes between accessions and the mapping of potential *cis*-acting variants [15]. A more comprehensive assessment of *cis*-variation is also feasible: the utilisation of F1 hybrids for systematic allele-specific expression (ASE) assays through RNA-sequencing has already been successfully applied in yeast, mouse and *Drosophila* [51–53] and still needs to be developed in plants.

Another upcoming challenge is to understand the extent of the epigenetic effects on global variation in gene

(Figure 1 Legend) QTL validation strategies by combining quantitative genetics approaches for classical and expression traits. The utilisation of cooperative strategies supporting the identification of candidate genes is based on the integration of expression trait mapping into classical linkage and/or genome-wide association studies (GWAS). A mapping set (either from RILs or wild accessions) exhibiting a quantitative phenotypic variation can be studied, for example, for an integrative trait, which together with transcriptome data can facilitate the identification of local (potentially *cis*-acting) variants or central hub controllers underlying the variation for the observed phenotype (QTG = Quantitative Trait Gene).

expression. One epigenetic modification that has been extensively studied is DNA methylation, which directly targets the DNA sequence at cytosines. In *Arabidopsis*, it has been shown that spontaneous epigenetic variation in DNA methylation may generate new allelic states that alter transcription and could have major impacts on phenotypic diversity [54,55]. Gene methylation appears to be polymorphic so that only half of all gene methylation on any one chromosome is shared between natural accessions collected from around the world, although it would only account for changes in expression for a fraction of the protein-coding genes [56]. Therefore, depicting the global methylome/epigenome (beyond simple DNA methylation) in natural accessions could assist the identification of the complete molecular basis of expression differences between strains showing positive ASE.

Finally, the rapid development of RNA-sequencing technologies together with the recent accessibility to datasets comprising hundreds of accessions from resequencing projects and fine-scale genotyped recombinant populations will allow the combination of classical linkage, GWAS and eQTL mapping strategies at a previously unimaginable scale. Consequently, the identification of genome-wide allelic expression differences, in interaction with environmental perturbations, will improve our understanding of how variation within regulatory elements affects transcript accumulation, allowing the prediction of the mechanisms by which standing genetic variation within these genetic elements shapes phenotypic diversity in natural populations.

Acknowledgements

We would like to thank Helen North, Amy Upton and Sarah Sharp for useful comments on the manuscript. Work in our group is supported by the European Commission Framework Programme 7, ERC Starting Grant: ERC-2009-StG-243359-DECODE to OL. VC is supported by a post-doctoral fellowship from INRA-DGAP and by the European Commission Framework Programme 7, Marie Curie Reintegration Grant: PERG08-GA-2010-277068.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Mackay TF, Stone EA, Ayroles JF: **The genetics of quantitative traits: challenges and prospects.** *Nat Rev Genet* 2009, **10**:565-577.
 2. Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al.*: **Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines.** *Nature* 2010, **465**:627-631.
 3. Trontin C, Tisne S, Bach L, Loudet O: **What does *Arabidopsis* natural variation teach us (and does not teach us) about adaptation in plants?** *Curr Opin Plant Biol* 2011, **14**:225-231.
 4. Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, Koornneef M: **What has natural variation taught us about plant development, physiology, and adaptation?** *Plant Cell* 2009, **21**:1877-1896.
 5. Hammond JP, Mayes S, Bowen HC, Graham NS, Hayden RM, Love CG, Spracklen WP, Wang J, Welham SJ, White PJ *et al.*: **Regulatory hotspots are associated with plant gene expression under varying soil phosphorus supply in *Brassica rapa*.** *Plant Physiol* 2011, **156**:1230-1241.
 6. Wang J, Yu H, Xie W, Xing Y, Yu S, Xu C, Li X, Xiao J, Zhang Q: **A global analysis of QTLs for expression variations in rice shoots at the early seedling stage.** *Plant J* 2010, **63**:1063-1074.
 7. Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, Schadt EE: **Systematic detection of polygenic cis-regulatory evolution.** *PLoS Genet* 2011, **7**:e1002023.
 8. Bullard JH, Mostovoy Y, Dudoit S, Brem RB: **Polygenic and directional regulatory evolution across pathways in *Saccharomyces*.** *Proc Natl Acad Sci U S A* 2010, **107**:5058-5063.
 9. Rockman MV, Skrovaneck SS, Kruglyak L: **Selection at linked sites shapes heritable phenotypic variation in *C. elegans*.** *Science* 2010, **330**:372-376.
 10. Kliebenstein DJ, West MA, van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St Clair DA: **Genomic survey of gene expression diversity in *Arabidopsis thaliana*.** *Genetics* 2006, **172**:1179-1189.
 11. Kliebenstein D: **Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs.** *Annu Rev Plant Biol* 2009, **60**:93-114.
 12. West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA: **Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*.** *Genetics* 2007, **175**:1441-1450.
 13. Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci.** *Proc Natl Acad Sci U S A* 2007, **104**:1708-1713.
 14. Simon M, Loudet O, Durand S, Bérard A, Brunel D, Sennesal F-X, Durand-Tardif M, Pelletier G, Camilleri C: **QTL mapping in five new large RIL populations of *Arabidopsis thaliana* genotyped with consensus SNP markers.** *Genetics* 2008, **178**:2253-2264.
 15. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrandt KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT *et al.*: **Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*.** *Nature* 2011, **477**:419-423.
- This paper provides new annotated genomes and transcriptome data for 18 natural *A. thaliana* accessions. The authors demonstrated that sequence variation between strains was enriched in transposable elements and intergenic regions, meanwhile most polymorphisms within coding regions were compensated for, without altering the gene structure. At the transcriptome level, almost half of the actively transcribed genes were differentially expressed between any pair of accessions, mostly due to polymorphisms within promoter regions (likely *cis*-acting).
16. Holloway B, Luck S, Beatty M, Rafalski JA, Li B: **Genome-wide expression quantitative trait loci (eQTL) analysis in maize.** *BMC Genomics* 2011, **12**:336.
 17. Moscou MJ, Lauter N, Steffenson B, Wise RP: **Quantitative and qualitative stem rust resistance factors in barley are associated with transcriptional suppression of defense regulons.** *PLoS Genet* 2011, **7**:e1002208.
 18. Chen X, Hackett CA, Niks RE, Hedley PE, Booth C, Druka A, Marcel TC, Vels A, Bayer M, Milne I *et al.*: **An eQTL analysis of partial resistance to *Puccinia hordei* in barley.** *PLoS One* 2010, **5**:e8598.
 19. Drost DR, Benedict CI, Berg A, Novaes E, Novaes CR, Yu Q, Dervinis C, Maia JM, Yap J, Miles B *et al.*: **Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of *Populus*.** *Proc Natl Acad Sci U S A* 2010, **107**:8492-8497.
 20. Zhang X, Cal AJ, Borevitz JO: **Genetic architecture of regulatory variation in *Arabidopsis thaliana*.** *Genome Res* 2011, **21**:725-733.
- Using a large set of F1 *Arabidopsis* hybrids, the authors were able to describe a genetic landscape of local and distant polymorphisms

explaining the expression differences between accessions. This corresponds to the first large-scale study able to describe the heterogeneous nature of *cis*-variation in plants.

21. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast.** *Science* 2002, **296**:752-755.
22. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: **Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors.** *Nat Genet* 2003, **35**:57-64.
23. Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS: **Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids.** *Science* 2009, **326**:1118-1120.
24. Hansen BG, Halkier BA, Kliebenstein DJ: **Identifying the molecular basis of QTLs: eQTLs add a new dimension.** *Trends Plant Sci* 2008, **13**:72-77.
25. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C *et al.*: **Whole-genome sequencing of multiple *Arabidopsis thaliana* populations.** *Nat Genet* 2011, **43**:956-963.
26. Wray GA: **The evolutionary significance of cis-regulatory mutations.** *Nat Rev Genet* 2007, **8**:206-216.
27. Arnaud N, Lawrenson T, Ostergaard L, Sablowski R: **The same regulatory point mutation changed seed-dispersal structures in evolution and domestication.** *Curr Biol* 2011, **21**:1215-1219.
28. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M *et al.*: **A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *AtHKT1*;** *PLoS Genet* 2010, **6**:e1001193.
In order to understand local adaptation and leaf Na⁺ accumulation (an indicator of salinity tolerance), the authors performed a GWAS on 349 European accessions. Na⁺ accumulation was highly correlated with the geographic distribution, with coastal accessions being more tolerant. Phenotypic differences were mapped down to polymorphisms within the regulatory region of *HKT1* and *HKT1* transcript abundance is indicated as a main cause of the observed geographic pattern.
29. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J *et al.*: **Natural variation in *GS5* plays an important role in regulating grain size and yield in rice.** *Nat Genet* 2011.
This work provides an example of a major rice grain size regulator (*GS5*) identified by a QTL analysis in a doubled haploid population. Natural variation within *GS5* promoter affects its own expression, with high levels positively affecting grain size. Interestingly, this contrasts with the majority of characterized regulators of grain or fruit size that are usually negatively correlated with expression.
30. Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M: **An SNP caused loss of seed shattering during rice domestication.** *Science* 2006, **312**:1392-1396.
31. Studer A, Zhao Q, Ross-Ibarra J, Doebley J: **Identification of a functional transposon insertion in the maize domestication gene *tb1*.** *Nat Genet* 2011, **43**:1160-1163.
This paper describes a case of gene expression variation at the maize domestication gene *TB1* enhanced by a Hopscotch transposon inserted in an upstream regulatory region. Elevated *TB1* expression results in an increase in maize apical dominance compared to teosinte, and the authors demonstrate that this variation likely emerged before maize domestication (standing variation).
32. Steffenson BJ, Jin Y, Brueggeman RS, Kleinhofs A, Sun Y: **Resistance to stem rust race TTKSK maps to the *rpg4/Rpg5* complex of chromosome 5H of barley.** *Phytopathology* 2009, **99**:1135-1141.
33. Schwartz C, Balasubramanian S, Warthmann N, Michael TP, Lempe J, Sureshkumar S, Kobayashi Y, Maloof JN, Borevitz JO, Chory J *et al.*: **Cis-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*.** *Genetics* 2009, **183**: 723-732, 721S1-727S1.
34. Shindo C, Lister C, Crevillen P, Nordborg M, Dean C: **Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response.** *Genes Dev* 2006, **20**:3079-3083.
35. Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F: **cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of *FLOWERING LOCUS T* in *Arabidopsis*.** *Plant Cell* 2010, **22**:1425-1440.
36. Raj S, Brautigam K, Hamanishi ET, Wilkins O, Thomas BR, Schroeder W, Mansfield SD, Plant AL, Campbell MM: **Clone history shapes *Populus* drought responses.** *Proc Natl Acad Sci U S A* 2011, **108**:12521-12526.
This study reveals how genetically identical *Populus* clones grown historically in contrasting environmental conditions, can exhibit distinct transcriptome profiles when exposed back to similar environments. Discrepancies in their gene-expression profiles in response to drought were attributed to differences in total DNA methylation patterns, highlighting the role of epi-polymorphism in shaping natural variation and local adaptation.
37. Delker C, Poschl Y, Raschke A, Ullrich K, Ettingshausen S, Hauptmann V, Grosse I, Quint M: **Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*.** *Plant Cell* 2010, **22**:2184-2200.
38. Terpstra IR, Snoek LB, Keurentjes JJ, Peeters AJ, van den Ackerveken G: **Regulatory network identification by genetical genomics: signaling downstream of the *Arabidopsis* receptor-like kinase *ERECTA*.** *Plant Physiol* 2010, **154**:1067-1078.
39. Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, Lam KC, Koka V, Kindmark A, Weiss ST, Tantisira K *et al.*: **Global analysis of the impact of environmental perturbation on cis-regulation of gene expression.** *PLoS Genet* 2011, **7**:e1001279.
40. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M *et al.*: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325**:1246-1250.
41. Barriere A, Gordon KL, Ruvinsky I: **Distinct functional constraints partition sequence conservation in a cis-regulatory element.** *PLoS Genet* 2011, **7**:e1002095.
42. Fu J, Keurentjes JJ, Bouwmeester H, America T, Verstappen FW, Ward JL, Beale MH, de Vos RC, Dijkstra M, Scheltema RA *et al.*: **System-wide molecular evidence for phenotypic buffering in *Arabidopsis*.** *Nat Genet* 2009, **41**:166-167.
43. Jimenez-Gomez JM, Wallace AD, Maloof JN: **Network analysis identifies *ELF3* as a QTL for the shade avoidance response in *Arabidopsis*.** *PLoS Genet* 2010:6.
44. Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ: **Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*.** *PLoS Biol* 2011, **9**:e1001125.
This paper reports a novel approach combining association mapping and co-expression network to identify candidates underlying glucosinolate traits. From the GWAS, the authors identified thousands of candidates, which after network filtering, were reduced to a smaller subset with genes closely linked to the glucosinolate biosynthetic network. This study demonstrates how data mining can help the identification of causal loci.
45. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR *et al.*: **Systems genetics of complex traits in *Drosophila melanogaster*.** *Nat Genet* 2009, **41**:299-307.
46. Oszolak F, Milos PM: **RNA sequencing: advances, challenges and opportunities.** *Nat Rev Genet* 2011, **12**:87-98.
47. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A: **Differential expression in RNA-seq: a matter of depth.** *Genome Res* 2011, **21**:2213-2223.
48. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR *et al.*: **The developmental dynamics of the maize leaf transcriptome.** *Nat Genet* 2010, **42**:1060-1067.
49. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F *et al.*: **The genome of the mesopolyploid crop species *Brassica rapa*.** *Nat Genet* 2011, **43**:1035-1039.

50. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H *et al.*: **The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change.** *Nat Genet* 2011, **43**:476-481.
51. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ: **Regulatory divergence in *Drosophila* revealed by mRNA-seq.** *Genome Res* 2010, **20**:816-825.
52. Babak T, Garrett-Engle P, Armour CD, Raymond CK, Keller MP, Chen R, Rohl CA, Johnson JM, Attie AD, Fraser HB *et al.*: **Genetic validation of whole-transcriptome sequencing for mapping expression affected by cis-regulatory variation.** *BMC Genomics* 2010, **11**:473.
53. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM: **A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data.** *Genome Res* 2011.
54. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR: **Transgenerational epigenetic instability is a source of novel methylation variants.** *Science* 2011, **334**:369-373.
55. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D: **Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome.** *Nature* 2011, **000**:1-6.
56. Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A *et al.*: **Epigenetic natural variation in *Arabidopsis thaliana*.** *PLoS Biol* 2007, **5**:e174.