Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology

Quantitative genetics in the age of omics

Joost JB Keurentjes[1,2,3], Maarten Koornneef[1,4] and Dick Vreugdenhil[2]

The use of natural variation in the genetic dissection of quantitative traits has a long-standing tradition. Recent advances in high-throughput technologies for the quantification of biological molecules have shifted the focus in quantitative genetics from single traits to comprehensive large-scale analyses. So-called omic technologies now enable geneticists to take a look in the black box that translates genetic information into biological function. These processes include transcriptional and (post) translational regulation as well as metabolic signaling pathways. The progress made in analytical and statistical techniques now allows the construction of regulatory networks that integrate the different levels of the biological information flow from gene-to-function.

Addresses
[1] Laboratory of Genetics, Wageningen University, Arboretumlaan 4, NL-6703 BD Wageningen, The Netherlands
[2] Laboratory of Plant Physiology, Wageningen University, Arboretumlaan 4, NL-6703 BD Wageningen, The Netherlands
[3] Centre for Biosystems Genomics, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands
[4] Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Corresponding author: Keurentjes, Joost JB (joost.keurentjes@wur.nl)

Current Opinion in Plant Biology 2008, **11**:123–128

This review comes from a themed issue on
Genome Studies and Molecular Genetics
Edited by Juliette de Meaux and Maarten Koornneef

Available online 6th March 2008

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2008.01.006

Introduction

For most organisms, including plants, variation between individuals of the same species is observed in nature, which can partly be explained by genetic differences [1]. Natural variation among different genotypes (accessions, varieties, etc.) can be classified as qualitative or quantitative. Qualitative traits are characterized by distinct phenotypic classes and are often a result of differences at single genes. Such traits can relatively easily be dissected genetically because of their clear segregation pattern in the progeny of crosses. Quantitative traits, on the contrary, often display a more continuous variation in phenotypes because of a multiplicity of genes involved and a relatively large effect of environmental factors on the expression of the trait. Recombination of genes results in a large number of phenotypic classes, which cannot unambiguously be associated with genotypic classes [2] because various genes can contribute positively or negatively to a quantifiable trait. The complexity of quantitative traits is further enhanced by the presence of epistatic interactions and interactions between genes and the environment [3••].

Quantitative natural variation controls adaptive strategies of organisms to cope with biotic and abiotic influences and its understanding can provide insight into ecological mechanisms and the evolutionary history of plants [4]. Moreover, it is the basis of variation for many agronomic traits [5]. *Arabidopsis thaliana* has proven to be a very efficient model plant because of a number of biological properties and available genetic resources that make genetic and molecular analyses very efficient [6]. These advantages also make *A. thaliana* very suitable for the genetic analysis of natural variation [7]. Because of the growing impact of large-scale molecular detection techniques (collectively nicknamed 'omics' technologies) in the dissection of complex traits, we aim to present a brief overview of the key technological advances and some of the recent findings in the field with emphasis on *A. thaliana*.

Genetic analysis of natural variation in quantitative traits

Despite the complexity in genetic regulation of quantitative traits much progress has been made over the past decades in dissecting these traits using molecular markers. The increasing ease by which molecular markers can be generated [8] in combination with the application of sophisticated mapping methods [9•] has led to a strong interest in the use of natural variation for studying quantitative traits [10]. Specific advantages are associated with the study of multiple natural perturbations in the same mapping population. This allows for the genetic analysis of an almost indefinite number of traits in the same genetic resources [9•]. For this type of study so-called immortal mapping populations, consisting in most cases of homozygous genotypes that can be tested in replicates and in different experiments, have proven to be very useful. Although various types of mapping populations have been developed for a variety of species, the relative ease of generating recombinant inbred lines (RILs) has led to their favorable use for quantitative trait locus (QTL) analysis in *A. thaliana* and many other plants [11]. However, especially for the study of differences between less related material, introgression or backcross inbred lines have proven to be very useful too [12]. Another genetic approach that makes use of a much larger part of the available genetic variation within a species is association or linkage disequilibrium (LD) mapping using historical recombination events [13].

Genetical genomics: variation in genome sequence and expression

In *A. thaliana* as well as in other species, genome-wide analyses of genomic polymorphisms in a large collection of accessions have revealed extensive sequence variation [1,14,15,16•]. Polymorphisms, when converted to molecular markers, are indispensable for (fine) mapping of quantitative traits in experimental populations. When surveyed in natural populations at high density, polymorphisms will enable high-resolution mapping through linkage disequilibrium [17]. The best marker, however, is the polymorphism causal for the observed variation. By definition, natural genetic variation is a result of genomic differences and therefore the extent of variation in quantitative traits is largely dependent on the level of DNA sequence variation. Although many of the polymorphisms will be functionally neutral, it leaves little doubt that the study of quantitative traits can benefit enormously from genomic analyses [18]. Nonsynonymous polymorphisms in coding sequences of genes might alter protein function or stability, introducing phenotypic variation. Polymorphisms in regulatory sequences on the other hand might result in differences in transcriptional efficiency of genes. It is therefore conceivable that genetically controlled expression differences, or variation in mRNA stability, contribute to natural variation in *A. thaliana* [19]. Given the extensive variation in phenotype and genomic sequence within *A. thaliana*, it is therefore not surprising that for many genes expression differences are observed between accessions [20,21,22•].

The genetic regulation of natural variation in gene expression should not be different from any other 'classical' quantitative trait and therefore, all statistical tools of quantitative genetics can be applied. The combination of linkage analysis (genetics) and expression profiling (genomics) was coined 'genetical genomics' [23••] and experiments were first reported in yeast [24•], soon followed by data of higher eukaryotes [25]. Because of the available high quality mapping populations and the commercially available genome-wide microarrays, *A. thaliana* is ideally suited for these kinds of analyses and several studies in various RIL populations have indicated extensive genetic regulation of gene expression [26–29].

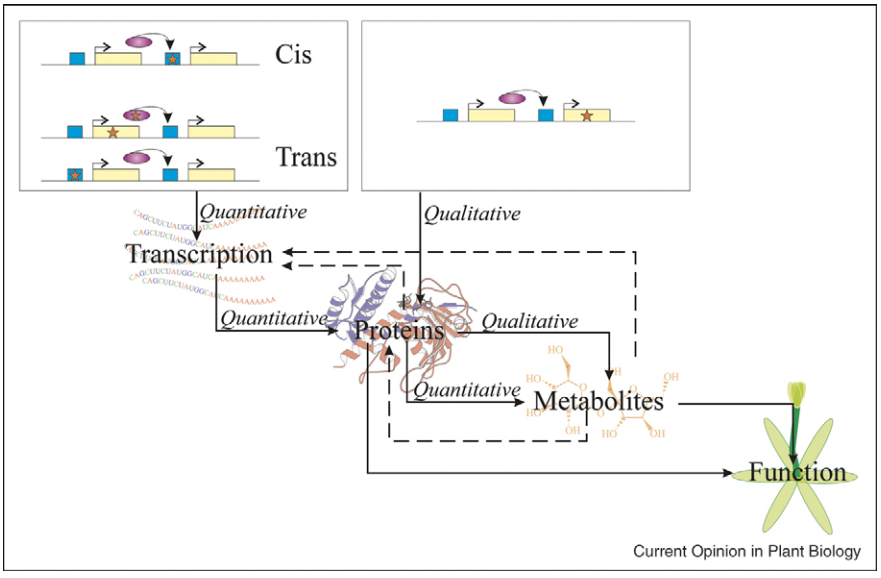
Genome-wide expression analysis of fully sequenced genomes, like *A. thaliana*, offers the unique possibility to compare genomic positions of genes with the map positions of the QTL(s) affecting the expression of these genes (eQTLs). Such comparative analyses reveal either local or distant regulation of gene expression. Local regulatory variation is often a result of polymorphisms in *cis*-acting regulatory elements affecting transcriptional activity. Distant regulatory variation most probably acts *in trans*, that is polymorphisms in another gene (e.g. a transcription factor) affect transcription of the gene for which the distant eQTL was detected. Nonetheless, other mechanisms of local and distant regulation, both *in cis* and *in trans*, are imaginable [30••]. Experimental data show that approximately half of the eQTLs map at the position of the gene itself (*cis*) and the other half at other loci (*trans*). However, when significance thresholds are made less stringent the proportion of *trans*-eQTLs increases [29]. This indicates that, in contrast to major-effect *cis*-eQTLs, many small-effect eQTLs act *in trans*.

Genetic regulation of complementary omic traits

The impact of variation in gene expression on quantitative traits is now widely acknowledged and the use of high-throughput genomic analyses has become an important tool in genetic analyses of natural variation [31]. An important mechanism in controlling transcriptional activity is through epigenetic modulation of *cis*-regulatory elements by cytosine methylation. With the recent development of genome-wide detection techniques [32], comprehensive genetic analyses of variation in methylation are at hand (Justin Borevitz, unpublished data). Transcription, however, is only a first link in the chain from genotype to phenotype and successive entities like proteins and metabolites (quality and quantity) are probably sources for natural phenotypic variation but have been largely under-exploited. Yet, high-throughput technologies, that is proteomics and metabolomics, have shown that much variation is observed upon physiological perturbation and between genetic variants [33,34]. Moreover, small-scale targeted analyses and subsequent QTL analysis revealed strong genetic regulation [35,36].

Analogous to genetical genomics, the combination of high-throughput proteomics and metabolomics and multifactorial genetic analyses would therefore allow studying the functional consequences of natural genetic variation at a much larger scale [37]. However, full-scale analyses for proteins and metabolites, equivalent to genome-wide expression analysis, are not available yet. This is mainly because proteins and metabolites are much more diverse in their properties than nucleic acids, making it difficult to extract and analyze all different classes using a single protocol. Even for a fully sequenced genome one cannot predict all protein variants and metabolites that a plant may contain. Moreover, the dynamic range of protein and metabolite abundance is far greater than for nucleic acids and no amplification techniques are available for these entities, making sample volume and detection range (sensitivity versus saturation) critical limitations. Nevertheless, several complementing high-throughput technologies covering together a large part of the proteome [38] and metabolome [39–41] have been developed. Noteworthy, more or less comprehensive platforms for large-scale analyses of plant elemental content and enzyme activity are now available [42,43]. Applications in experimental mapping populations have not been reported yet for most platforms but several

Figure 1



Natural variation affecting the various interconnected transducers of the biological information flow. Variation in gene expression because of polymorphisms in *cis*-regulatory elements (blue boxes) or expression differences of regulators (*trans*) may cause quantitative differences in protein content and metabolic fluxes resulting in altered function. Polymorphisms in coding regions of genes (yellow boxes) may result in qualitative differences affecting molecular functioning. Feedback and metabolic signaling mechanisms further complicate the delicate regulation of quantitative traits.

studies involving enzyme activity and proteomics are underway (author's unpublished work).

The progress made in metabolomics already enabled large-scale genetic analyses, which has first been demonstrated for primary metabolites [44•]. However, variation in secondary metabolism is probably more extensive and may determine much of the phenotypic variation. In *A. thaliana* alone already hundreds of secondary metabolites representing numerous chemical classes have been discovered [45]. Given the wide global distribution range of *A. thaliana* and the diverge range of site plants have been collected, it is conceivable that metabolites play an important role in local adaptation strategies. It is therefore not surprising that the high level of natural variation in *A. thaliana* is also reflected in metabolite composition and content [46]. Indeed, the successful combination of large-scale untargeted metabolomics and quantitative genetic approaches has revealed extensive genetic control of, and high flexibility in metabolic profiles [47••].

Regulatory network construction

To functionally link the large data sets obtained in 'omic' experiments as an order of events that ultimately result in a specific phenotype, network construction provides a useful tool. Biological networks describe relationships between individual components of a biological process [48]. Such components can either be genes, proteins, metabolites, or a combination thereof (Figure 1). Depending on the data source, networks can be constructed in various ways but all aim at resolving the complex regulation of biological processes.

One type of network does not rely on experimental data but rather predicts *in silico* connections based on genome-wide sequence information. Most notably are genome-scale metabolic connectivity networks, where metabolites are connected when the genome contains a gene encoding an enzyme able to catalyze the conversion of one of the metabolites into the other [49]. However, genetic networks have also been predicted *in silico* by analyzing regulatory elements of genes for binding sites of known transcription factors [50]. Although powerful in hypothesis formation such studies require empirical data for confirmation of predicted pathways and interactions. Therefore, many approaches for network construction are based on experimental data, which also allows the identification of relationships unable to be predicted from genomic information only. Protein–protein interactions for instance, are difficult to deduce from sequence information but require immuno-precipitation or two-hybrid screens. Similar analyses, like chromatin immuno-precipitation (ChIP-Chip), can also be used to identify and confirm transcriptional regulation of target genes by transcription factors or other known regulators [51]. In yeast, much progress in regulatory network construction was made by expression and metabolic profiling of deletion strains [52,53] and genetic interaction analyses using double mutants (synthetic lethals) [54]. However, for most higher eukaryotes such genome-wide analyses

are not realistic because of the much higher gene number, the presumably more complex genetic architecture, and aspects of subcellular and tissue specific compartmentation. Many attempts in regulatory network construction therefore rely on indirect approaches of establishing associations between network components.

A straightforward approach is correlation analysis over a large set of data compiled from numerous perturbation experiments [55]. Exemplary are the widely applied gene coexpression analyses, where correlation in gene expression patterns is surveyed under a large number of diverse conditions [56•,57]. The rationale for this kind of analysis is that genes participating in the same biological process are often coregulated and hence exhibit similar expression patterns. Following the same line of reasoning, metabolic correlation networks have been constructed [58]. However, the reliability of, and information contained in constructed networks would gain much strength from integrated analyses of interdisciplinary approaches [59,60]. Such integrated studies can either combine experimental data with *in silico* analyses [61] or benefit from multiparallel analyses of diverse biological samples [62–64]. These approaches already enabled the identification of novel regulatory steps in metabolic biosynthesis pathways [65].

Although demonstrably effective, correlation analyses depend on large compendia of publicly available data or suffer from the limited number of physiological conditions that can be analyzed in dedicated experiments. However, sometimes coregulation is displayed only in particular conditions [57], which may remain undiscovered, even in large data sets because of dilution effects. The largest drawback of correlation analyses, however, is that no information can be retrieved about the nature of the underlying genetic regulation. Correlation does not necessarily imply functional relatedness nor does it address causality issues. Correlation may be a result from coregulation by a common regulator or because of independent pathways that occur in parallel, possibly because of developmental or spatial control. Otherwise, a highly correlated cluster of biological elements, such as genes, proteins, and metabolites, can also result from downstream effects of the regulation of a single member but no information about cause and consequence can be extracted from genetic correlations.

Mapping populations combine a high number of genetic perturbations by which numerous quantitative traits segregate in a single experiment. Moreover, genetic analysis offers the unique possibility of identifying genomic loci causal for observed variation in, and possible correlation between traits. When applied to genome-wide expression analysis or other large-scale 'omic' analyses this therefore allows the identification of true gene-to-gene or gene-to-function regulation. The successful (re)construction of metabolic and genetic regulatory networks [29,47••] has shown the usefulness of combining quantitative genetics and large-scale omic analyses. Unfortunately, mapping resolution is often not high enough to identify causal genes underlying detected QTLs directly and will require further analysis such as fine mapping, the study of overexpressors and mutants of candidate genes, etc. However, differently regulated genes, because of polymorphic *cis*-regulatory elements, are obvious candidates and coregulated traits can effectively be identified through colocation of detected QTLs [47••]. Yet, coinciding QTLs not necessarily represent the same causal gene because effects of closely linked genes are difficult to distinguish from true pleiotropic effects of a single gene. Without further experimentation genetic interactions can be predicted computationally by comparing QTL profiles and correlation analyses [66]. However, the accuracy of constructed networks can benefit tremendously from the integration of additional information like gene ontology [29,67], sequence data [68] and related quantitative trait data for end traits including metabolites and plant performance [35,69].

Conclusions

The combination of genetic analyses and large-scale omic analyses in experimental mapping populations has shown to have great potential in unraveling meaningful biological regulatory networks. The dissection of the genetic architecture of quantitative traits will require multiparallel analyses of the different transducers of the biological information flow. The ultimate goal is to link genetic variation to phenotypic variation and the identification of the molecular pathway from gene-to-function. The first reports where natural variation at the metabolite level is linked to growth-related phenotypes have already been published [44•,70,71]. The recent progress made in humans by combining LD mapping and transcriptomics [72] holds great promises for high-resolution association mapping and identification of regulatory genetic factors. Future directions will include additional dimensions such as genotype × environment interactions and temporal and spatial control of regulation. Such studies will need ceaselessly advancing genetic resources, bioinformatics and cost-effective omic tools.

Acknowledgements

We apologize to all colleagues whose relevant work we could not cite because of space limitations, and thank Linus van der Plas for critically reading the manuscript. We acknowledge support from the Netherlands Organization for Scientific Research, Program Genomics (050-10-029) and the Centre for Biosystems Genomics (CBSG, Netherlands Genomics Initiative).

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R *et al.*: **The pattern of polymorphism in *Arabidopsis thaliana***. *PLoS Biol* 2005, **3**:e196.

2. Holland JB: **Genetic architecture of complex traits in plants**. *Curr Opin Plant Biol* 2007, **10**:156-161.

3. •• Kroymann J, Mitchell-Olds T: **Epistasis and balanced polymorphism influencing complex trait variation**. *Nature* 2005, **435**:95-98.
An enunciative example of the possible complexity of quantitative traits. The authors describe the effect of two tightly linked QTLs on fitness, which appeared to depend on interactions with the genetic background and hence were not detected in a segregating population.

4. Mitchell-Olds T, Schmitt J: **Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis**. *Nature* 2006, **441**:947-952.

5. Ross-Ibarra J: **Quantitative trait loci and the study of plant domestication**. *Genetica* 2005, **123**:197-204.

6. Somerville C, Koornneef M: **Timeline: a fortunate choice: the history of Arabidopsis as a model plant**. *Nat Rev Genet* 2002, **3**:883-889.

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

8. Borevitz JO, Chory J: **Genomics tools for QTL analysis and gene discovery**. *Curr Opin Plant Biol* 2004, **7**:132-136.

9. • Doerge RW: **Mapping and analysis of quantitative trait loci in experimental populations**. *Nat Rev Genet* 2002, **3**:43-52.
An excellent overview of quantitative genetic approaches and the use of natural variation in dissecting complex traits.

10. Slate J: **Quantitative trait locus mapping in natural populations: progress, caveats and future directions**. *Mol Ecol* 2005, **14**:363-379.

11. Jansen RC: **Quantitative trait loci in inbred lines**. In *Handbook of Statistical Genetics.* Edited by Balding DJ, Bishop M, Cannings C. John Wiley & Sons; 2003:445-476.

12. Zamir D: **Improving plant breeding with exotic genetic libraries**. *Nat Rev Genet* 2001, **2**:983-989.

13. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M: **Recombination and linkage disequilibrium in *Arabidopsis thaliana***. *Nat Genet* 2007, **39**:1151-1155.

14. Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B: **Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana***. *Genome Res* 2003, **13**:1250-1257.

15. Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT, Chen H, Werner JD, Nordborg M, Salt DE *et al.*: **Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana***. *Proc Natl Acad Sci U S A* 2007, **104**:12057-12062.

16. • Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA *et al.*: **Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana***. *Science* 2007, **317**:338-342.
The authors performed high-density array resequencing of 20 diverse Arabidopsis accessions and discovered more than 1 million single-nucleotide polymorphisms.

17. Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ *et al.*: **The extent of linkage disequilibrium in *Arabidopsis thaliana***. *Nat Genet* 2002, **30**:190-193.

18. Borevitz JO, Nordborg M: **The impact of genomics on the study of natural variation in Arabidopsis**. *Plant Physiol* 2003, **132**:718-725.

19. Chen WJ, Chang SH, Hudson ME, Kwan WK, Li J, Estes B, Knoll D, Shi L, Zhu T: **Contribution of transcriptional regulation to natural variations in Arabidopsis**. *Genome Biol* 2005, **6**:R32.

20. 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.

21. Vuylsteke M, van Eeuwijk F, Van Hummelen P, Kuiper M, Zabeau M: **Genetic analysis of variation in gene expression in *Arabidopsis thaliana***. *Genetics* 2005, **171**:1267-1275.

22. • West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, Michelmore RW: **High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis**. *Genome Res* 2006, **16**:787-795.
A good example of the use of microarrays for simultaneous genotyping and expression profiling in experimental mapping populations.

23. •• Jansen RC, Nap JP: **Genetical genomics: the added value from segregation**. *Trends Genet* 2001, **17**:388-391.
The first paper explaining the concept and potential of genetical genomics.

24. • Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast**. *Science* 2002, **296**:752-755.
The first paper describing experimental data of mapping gene expression in a segregating population.

25. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G *et al.*: **Genetics of gene expression surveyed in maize, mouse and man**. *Nature* 2003, **422**:297-302.

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

27. 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.

28. Vuylsteke M, Daele H, Vercauteren A, Zabeau M, Kuiper M: **Genetic dissection of transcriptional regulation by cDNA-AFLP**. *Plant J* 2006, **45**:439-446.

29. Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, 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.

30. •• Rockman MV, Kruglyak L: **Genetics of global gene expression**. *Nat Rev Genet* 2006, **7**:862-872.
An excellent review of the genetic analysis of gene expression in segregating mapping populations.

31. Gibson G, Weir B: **The quantitative genetics of transcription**. *Trends Genet* 2005, **21**:616-623.

32. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE *et al.*: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis**. *Cell* 2006, **126**:1189-1201.

33. Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N, Rossignol M: **Proteomic investigation of natural variation between Arabidopsis ecotypes**. *Proteomics* 2004, **4**:1372-1381.

34. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: **Metabolite profiling for plant functional genomics**. *Nat Biotechnol* 2000, **18**:1157-1161.

35. Consoli L, Lefevre A, Zivy M, de Vienne D, Damerval C: **QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize**. *Plant Mol Biol* 2002, **48**:575-581.

36. Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T: **Genetic control of natural variation in Arabidopsis glucosinolate accumulation**. *Plant Physiol* 2001, **126**:811-825.

37. Jansen RC: **Studying complex biological systems using multifactorial perturbation**. *Nat Rev Genet* 2003, **4**:145-151.

38. Peck SC: **Update on proteomics in Arabidopsis. Where do we go from here?** *Plant Physiol* 2005, **138**:591-599.

39. De Vos RC, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD: **Untargeted large-scale plant metabolomics using liquid**

chromatography coupled to mass spectrometry. *Nat Protocol* 2007, **2**:778-791.

40. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR: **Gas chromatography mass spectrometry-based metabolite profiling in plants**. *Nat Protocol* 2006, **1**:387-396.

41. Ward JL, Harris C, Lewis J, Beale MH: **Assessment of 1H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana***. *Phytochemistry* 2003, **62**:949-957.

42. Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE, Guerinot ML, Harper JF, Ward JM, McIntyre L *et al.*: **Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana***. *Nat Biotechnol* 2003, **21**:1215-1221.

43. Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JH, Palacios N, Cross J, Selbig J, Stitt M: **A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness**. *Plant Cell* 2004, **16**:3304-3325.

44. • Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J *et al.*: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement**. *Nat Biotechnol* 2006, **24**:447-454.
One of the first papers showing a strong relationship between growth-related traits and metabolic profiles.

45. D'Auria JC, Gershenzon J: **The secondary metabolism of *Arabidopsis thaliana*: growing like a weed**. *Curr Opin Plant Biol* 2005, **8**:308-316.

46. Fiehn O: **Metabolomics — the link between genotypes and phenotypes**. *Plant Mol Biol* 2002, **48**:155-171.

47. •• Keurentjes JJB, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M: **The genetics of plant metabolism**. *Nat Genet* 2006, **38**:842-849.
One of the first publications of combining genetic analyses with large-scale untargeted metabolomics. The authors show the extensive natural variation in and genetic control of metabolic profiles in Arabidopsis.

48. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5**: 101-113.

49. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL: **The large-scale organization of metabolic networks**. *Nature* 2000, **407**:651-654.

50. Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E: **AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks**. *Plant Physiol* 2006, **140**:818-829.

51. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I *et al.*: **Transcriptional regulatory networks in *Saccharomyces cerevisiae***. *Science* 2002, **298**:799-804.

52. Hu Z, Killion PJ, Iyer VR: **Genetic reconstruction of a functional transcriptional regulatory network**. *Nat Genet* 2007, **39**:683-687.

53. Forster J, Gombert AK, Nielsen J: **A functional genomics approach using metabolomics and in silico pathway analysis**. *Biotechnol Bioeng* 2002, **79**:703-712.

54. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M *et al.*: **Global mapping of the yeast genetic interaction network**. *Science* 2004, **303**:808-813.

55. de la Fuente A, Bing N, Hoeschele I, Mendes P: **Discovery of meaningful associations in genomic data using partial correlation coefficients**. *Bioinformatics* 2004, **20**:3565-3574.

56. • Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules**. *Science* 2003, **302**:249-255.
The authors combined information of gene conservation over different evolutionary lineages and gene expression to identify clusters of functionally related genes.

57. Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P: **Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications**. *Plant Mol Biol* 2005, **58**:229-245.

58. Steuer R, Kurths J, Fiehn O, Weckwerth W: **Observing and interpreting correlations in metabolomic networks**. *Bioinformatics* 2003, **19**:1019-1026.

59. Fiehn O, Kloska S, Altmann T: **Integrated studies on plant biology using multiparallel techniques**. *Curr Opin Biotechnol* 2001, **12**:82-86.

60. Winnacker EL: **Interdisciplinary sciences in the 21st century**. *Curr Opin Biotechnol* 2003, **14**:328-331.

61. Segal E, Yelensky R, Koller D: **Genome-wide discovery of transcriptional modules from DNA sequence and gene expression**. *Bioinformatics* 2003, **19(Suppl 1)**:i273-i282.

62. Joosen R, Cordewener J, Supena ED, Vorst O, Lammers M, Maliepaard C, Zeilmaker T, Miki B, America T, Custers J *et al.*: **Combined transcriptome and proteome analysis identifies pathways and markers associated with the establishment of *Brassica napus* microspore-derived embryo development**. *Plant Physiol* 2007, **144**:155-172.

63. Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR: **Parallel analysis of transcript and metabolic profiles: a new approach in systems biology**. *EMBO Rep* 2003, **4**:989-993.

64. Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H *et al.*: **Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics**. *J Biol Chem* 2005, **280**:25590-25595.

65. Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K *et al.*: **Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis**. *Proc Natl Acad Sci U S A* 2007, **104**:6478-6483.

66. Lan H, Chen M, Flowers JB, Yandell BS, Stapleton DS, Mata CM, Mui ET, Flowers MT, Schueler KL, Manly KF *et al.*: **Combined expression trait correlations and expression quantitative trait locus mapping**. *PLoS Genet* 2006, **2**:e6.

67. Kliebenstein DJ, West MA, van Leeuwen H, Loudet O, Doerge RW, St Clair DA: **Identification of QTLs controlling gene expression networks defined a priori**. *BMC Bioinformat* 2006, **7**:308.

68. Hitzemann R, Malmanger B, Reed C, Lawler M, Hitzemann B, Coulombe S, Buck K, Rademacher B, Walter N, Polyakov Y *et al.*: **A strategy for the integration of QTL, gene expression, and sequence analyses**. *Mamm Genome* 2003, **14**:733-747.

69. Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ: **Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways**. *PLoS Genet* 2007, **3**:e162.

70. Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J *et al.*: **The metabolic signature related to high plant growth rate in *Arabidopsis thaliana***. *Proc Natl Acad Sci U S A* 2007, **104**:4759-4764.

71. Sergeeva LI, Keurentjes JJB, Bentsink L, Vonk J, van der Plas LHW, Koornneef M, Vreugdenhil D: **Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant analysis**. *Proc Natl Acad Sci U S A* 2006, **103**:2994-2999.

72. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M *et al.*: **A genome-wide association study of global gene expression**. *Nat Genet* 2007, **39**:1202-1207.