46. Maiato, H., Rieder, C. L. & Khodjakov, A. Kinetochore-driven formation of kinetochore fibers contributes to spindle assembly during animal mitosis. *J. Cell Biol.* **167**, 831–840 (2004).
47. Knop, M. & Schiebel, E. Spc98p and Spc97p of the yeast γ-tubulin complex mediate binding to the spindle pole body via their interaction with Spc110p. *EMBO J.* **16**, 6985–6995 (1997).
48. Knop, M. & Schiebel, E. Receptors determine the cellular localization of a γ-tubulin complex and thereby the site of microtubule formation. *EMBO J.* **17**, 3952–3967 (1998).
49. Sawin, K. E., Lourenco, P. C. & Snaith, H. A. Microtubule nucleation at non-spindle pole body microtubule-organizing centers requires fission yeast centrosomin-related protein mod20p. *Curr. Biol.* **14**, 763–775 (2004).
50. Venkatram, S. *et al.* Identification and characterization of two novel proteins affecting fission yeast γ-tubulin complex function. *Mol. Biol. Cell* **15**, 2287–2301 (2004).
51. Samejima, I., Lourenco, P. C., Snaith, H. A. & Sawin, K. E. Fission yeast mto2p regulates microtubule nucleation by the centrosomin-related protein mto1p. *Mol. Biol. Cell* **16**, 3040–3051 (2005).
52. Venkatram, S., Jennings, J. L., Link, A. & Gould, K. L. Mto2p, a novel fission yeast protein required for cytoplasmic microtubule organization and anchoring of the cytokinetic actin ring. *Mol. Biol. Cell* **16**, 3052–3063 (2005).
53. Zimmerman, S. & Chang, F. Effects of γ-tubulin complex proteins on microtubule nucleation and catastrophe in fission yeast. *Mol. Biol. Cell* **16**, 2719–2733 (2005).
54. do Carmo Avides, M. & Glover, D. M. Abnormal spindle protein, Asp, and the integrity of mitotic centrosomal microtubule organizing centers. *Science* **283**, 1733–1735 (1999).
55. Takahashi, M., Yamagiwa, A., Nishimura, T., Mukai, H. & Ono, Y. Centrosomal proteins CG-NAP and kendrin provide microtubule nucleation sites by anchoring γ-tubulin ring complex. *Mol. Biol. Cell* **13**, 3235–3245 (2002).
56. Casenghi, M. *et al.* Polo-like kinase 1 regulates Nlp, a centrosome protein involved in microtubule nucleation. *Dev. Cell* **5**, 113–125 (2003).
57. Terada, Y., Uetake, Y. & Kuriyama, R. Interaction of Aurora-A and centrosomin at the microtubule-nucleating site in *Drosophila* and mammalian cells. *J. Cell Biol.* **162**, 757–763 (2003).
58. Zimmerman, W. C., Sillibourne, J., Rosa, J. & Doxsey, S. J. Mitosis-specific anchoring of γ tubulin complexes by pericentrin controls spindle organization and mitotic entry. *Mol. Biol. Cell* **15**, 3642–3657 (2004).
59. Anders, A., Lourenco, P. C. & Sawin, K. E. Non-core components of the fission yeast γ-tubulin complex. *Mol. Biol. Cell* **17**, 5075–5093 (2006).
60. Verollet, C. *et al. Drosophila melanogaster* γ-TuRC is dispensable for targeting γ-tubulin to the centrosome and microtubule nucleation. *J. Cell Biol.* **172**, 517–528 (2006).
61. Vogt, N., Koch, I., Schwarz, H., Schnorrer, F. & Nusslein-Volhard, C. The γTuRC components Grip75 and Grip128 have an essential microtubule-anchoring function in the *Drosophila* germline. *Development* **133**, 3963–3972 (2006).
62. Vogel, J. *et al.* Phosphorylation of γ-tubulin regulates microtubule organization in budding yeast. *Dev. Cell* **1**, 621–631 (2001).
63. Stumpff, J., Kellogg, D. R., Krohne, K. A. & Su, T. T. *Drosophila* Wee1 interacts with members of the γTURC and is required for proper mitotic-spindle morphogenesis and positioning. *Curr. Biol.* **15**, 1525–1534 (2005).
64. Starita, L. M. *et al.* BRCA1-dependent ubiquitination of γ-tubulin regulates centrosome number. *Mol. Cell. Biol.* **24**, 8457–8466 (2004).

DATABASES
The following terms in this article are linked online to:
UniProtKB: http://ca.expasy.org/sprot
Centrosomin | CG-NAP | GCP2 | GCP3 | GCP6 | kendrin | Mto1 | Mto2 | NEDD1 | ninein | Pcp1

FURTHER INFORMATION
Tim Stearns's laboratory: http://stearnslab.stanford.edu
Access to this links box is available online.

INNOVATION

Towards the plant metabolome and beyond

Robert L. Last, A. Daniel Jones and Yair Shachar-Hill

Abstract | Methods for network-wide analysis are increasingly showing that the textbook view of the regulation of plant metabolism is often incomplete and misleading. Recent innovations in small-molecule analysis have created the ability to rapidly identify and quantify numerous compounds, and these data are creating new opportunities for understanding plant metabolism and for plant metabolic engineering.

Plant biologists hope to harness the availability of genome-wide data on DNA, RNA and proteins to develop an integrated understanding of the behaviour of complex metabolic networks. This is an increasingly important goal for our global economy as we rely on plant-derived materials to replace fossil fuels as energy sources and industrial feedstocks, and for their traditional and crucial roles in food, fibre and pharmaceuticals. Reaching such an understanding will require cataloguing all the metabolites that are synthesized, identifying the enzymes that are responsible for the synthesis of each product, and characterizing the dynamic behaviour of the metabolic pathways and networks. Developing a predictive level of understanding of metabolism is an especially daunting task in plants because they produce orders of magnitude more metabolites than animals.

Although the analysis of metabolites in plants has a long history, the development of high-throughput methods for measuring large numbers of compounds has arisen in less than a decade. This is a result of rapid improvements in mass spectrometry (MS)-based methods and in computer hardware and software that is capable of interpreting large datasets. Studies in which bioinformatics and data-analysis methods are applied to such data are beginning to contribute to functional genomics studies and to generate new hypotheses about how plant metabolism is controlled, but such systems approaches are still in their infancy.

The use of high-throughput analytical strategies to identify and measure the concentration of as many metabolites as possible is termed metabolomics. For systems-wide analyses, accurate identification and quantification of small molecules are the essential first steps. Despite much recent progress, metabolomics of plant systems is not yet '-omic' in scale, so it is useful to consider the prospects for progress in this direction.

This article primarily focuses on recent innovations in MS-based analytical chemistry that have created the opportunity to measure the abundances of hundreds of metabolites and also provides information on the powerful and complementary techniques of nuclear magnetic resonance (NMR) (BOX 1). The challenges of using the large amounts of data that are generated by metabolomics are discussed. We then show, by using examples from the recent literature, how metabolomics can be used for discovering new metabolic genes and enzymes.

Measuring molecules: analytical methods
There are three general approaches to analyse small molecules, and these differ in the number of compounds they analyse, in the level of structural information they obtain, and in their sensitivity[1–4]. The most common approach, metabolite profiling, is the analysis of small numbers of known metabolites in specific compound classes (for example, lipid, amino acid or sterol). At the other extreme, metabolic fingerprinting detects many compounds, but their structures are rarely identified. Metabolomics methods typically allow the measurement of hundreds of molecules, with a small number being definitively

Metabolites are almost entirely made of hydrogen, carbon, nitrogen, oxygen and phosphorus. These elements all have isotopes (^{1}H, ^{13}C and ^{15}N being the most important isotopes for metabolomics) that yield nuclear magnetic resonance (NMR) signals when placed in a strong magnetic field and pulsed with radiofrequency electromagnetic radiation. NMR spectra are useful for plant metabolomics because they yield detailed information about the quantities and identities of the metabolites present in extracts or *in vivo*.

Spectra can be recorded from crude, even one-step, extracts without derivatization, making the composition of the samples closer to the actual levels in tissues. Furthermore, the signals in a spectrum are proportional to the amount of each metabolite in a sample, allowing absolute quantification of different metabolites without the use of separate standards for each of them.

NMR provides detailed structural information about metabolites. This is useful as many plant metabolites are not identified in mass spectra. Sugars, amino acids and organic acids can be identified and 20–40 compounds are routinely identified in crude extracts. Although this is fewer than in a typical mass spectrum, NMR methods, especially two-dimensional ones, are the 'gold standard' for the structural elucidation of newly isolated compounds.

NMR is non-destructive, allowing the sample to be subsequently analysed by other spectroscopy techniques. It is also possible to record NMR spectra from living cells and tissues, permitting the measurement of levels of certain labile metabolites without perturbation during extraction. Also, information on subcellular compartmentalization of metabolites can be obtained.

Depending on the type of spectrum and instrumentation used, NMR can detect metabolites in the sample at micromolar to low millimolar concentration ranges in spectra acquired in seconds or minutes — a much lower sensitivity than mass spectrometric analyses. Its insensitivity, combined with the fact that the most sensitive NMR spectra (one-dimensional ^{1}H) also have the poorest resolution, is the reason why NMR is used less in plant metabolomics than is mass spectrometry (MS). Nevertheless, the advantages listed above make NMR a powerful complementary adjunct to MS for profiling and a useful stand-alone method for metabolite fingerprinting.

identified, a larger number being identified as belonging to particular compound classes, and many remaining unidentified. Most metabolomics studies use MS, although NMR spectroscopy[5–7] (BOX 1) and high-performance liquid chromatography (HPLC) with electrochemical or photodiode array detection (see Glossary), offer additional useful information, as does Fourier-transform infrared (FT-IR) spectroscopy. As technology has progressed, there are more studies in which many plant metabolites are identified and quantified. This is an exciting development, as such studies are beginning to make significant contributions to our understanding of plant gene functions.

MS can detect and resolve a broad range of metabolites[8,9] with speed, sensitivity and accuracy. At the heart of the method is the ionization of molecules and the measurement of ion abundances and mass-to-charge ratios. The mass spectrometer can be used to analyse molecules in a complex mixture without prior separation (often referred to as infusion or flow-injection analysis MS) (FIG. 1). This approach relies on accurate measurements of mass-to-charge ratios to distinguish between different molecules of the same nominal mass and to indicate the elemental composition of each metabolite in the mixture. Separation of the components of complex mixtures by

gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis before samples enter the mass spectrometer has important advantages (FIG. 2). For example, the separation properties usually reflect the type of molecule being measured (for instance, polar versus hydrophobic or positively versus negatively charged). Fractionation also leads to a reduction of the number of ions being measured at a given time, which improves the analytical properties of the method.

The number of ions that can be detected in a plant extract depends on the sample, the method of ionization and the resolution of the mass spectrometer. High mass resolution Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometers have detected several thousand distinct mass peaks in extracts from 50–100 mg of plant material[10,11]. However, it is important to keep in mind that the number of observed molecular species does not equal the number of metabolites. A single metabolite can yield peaks at several masses, owing to the presence of stable isotopes, the attachment of different cations (such as hydrogen, sodium, or potassium ions), the formation of non-covalent dimer ions, and the fragmentation of the parent molecule. For these reasons, the observation of thousands of ions of different masses might correspond to the detection of hundreds of metabolites.

In the absence of a costly FT-ICR instrument, metabolic fingerprinting analyses can also be done at intermediate mass resolution on less expensive time-of-flight (TOF) mass spectrometers[12] or on the new Orbitrap mass analyser[13]. Furthermore, fast TOF mass spectrometers now combine short analysis times with the resolution of several hundred known and unknown metabolites in a single analysis[14].

Besides MS, the method most often used for fingerprinting and profiling metabolites in plants is NMR spectroscopy[5,6,15] (BOX 1). Recent studies illustrate the advantages of NMR, particularly for generating unbiased, quantitative metabolic fingerprints of plant tissues. A good example is the discrimination and phenotyping of *Arabidopsis* ecotypes by Ward and colleagues[4]. In this study, one-dimensional proton (^{1}H) NMR spectra taken of one-step extracts of 15 mg of lyophilized plant tissue were sufficient to discriminate among 9 ecotypes and to identify the metabolites that accounted for the differences in their metabolite profiles. Compounds that contribute to the differences can be identified through an examination of the loadings of the components in principal components analysis (see Supplementary information S1 (box) and REF. 16).

NMR can yield much more information on plant metabolism than is available from simple one-dimensional ^{1}H spectra[7], including increased resolution and structural information by using two-dimensional methods[17,18] and by coupling NMR spectroscopy to chromatographic separation[19]. The powerful synergy between NMR and MS — especially for metabolite identification, discovery and quantification in plant tissues and exudates — has been well demonstrated[17,20] but remains underused in plant metabolomics.

Quantitative metabolite measurements

One of the many challenges faced by metabolomics studies is the issue of quantitative analysis. Signal intensities in mass spectra are only indirectly related to the absolute levels of the metabolites that they represent, as they depend on the instrument used, the chemical nature of the analytes and solvents, and on the other compounds in the sample (so-called matrix effects). Therefore, absolute quantification in MS usually requires the inclusion of authentic standards (ideally, isotopically labelled ones). However, such standards are not available for most compounds in a complex mixture, and many signals in metabolite extracts

remain to be identified. Therefore, metabolite levels are often normalized to known amounts of one or more reference compounds that have been added to the extracts or to the total observed signal to give a relative quantification. By contrast, the signals in routine NMR spectra are directly proportional to the concentration of each compound, therefore greatly facilitating metabolite quantification. However, the relatively low sensitivity and resolution in such spectra mean that only 20–40 metabolites are usually quantified.

Data analysis and interpretation

The application of metabolomics data can be divided into three categories that are conceptually different and that typically involve different data-analysis procedures. The most common application is the phenotypic description of one or more species or genetic variants of a species. This involves cataloguing and comparing datasets to determine whether the results are the same or different; metabolic fingerprinting is often sufficient for this task. The second application is to generate and/or address hypotheses about plant function by more detailed statistical analysis of differences between metabolomics datasets; this requires compound identification. The third application, systems biology, involves the integrated consideration of metabolomics and other '-omics' data, such as transcript measurements. A further subset of the third application of analysis is the mechanistic modelling of plant functional networks, in which simulations of transport and metabolism are computed and compared to quantitative data on metabolite levels. However, this is not yet widely used in plant studies[21,22], although it is a key goal in systems biology (see WTEC Panel Report on International Research and Development in Systems Biology in Further information).

The first stage of data handling — experimental design — precedes data generation. Methods used to extract and process metabolite data and the chromatographic and spectrometric parameters used in their analysis influence data, how comparable the data generated will be to other datasets and what libraries and other existing information about metabolite identity and quantification will be applicable for data interpretation. Variability between measurements of biological replicates can be large in plant metabolomics studies, and this has led some groups to favour pooling several biological replicate samples
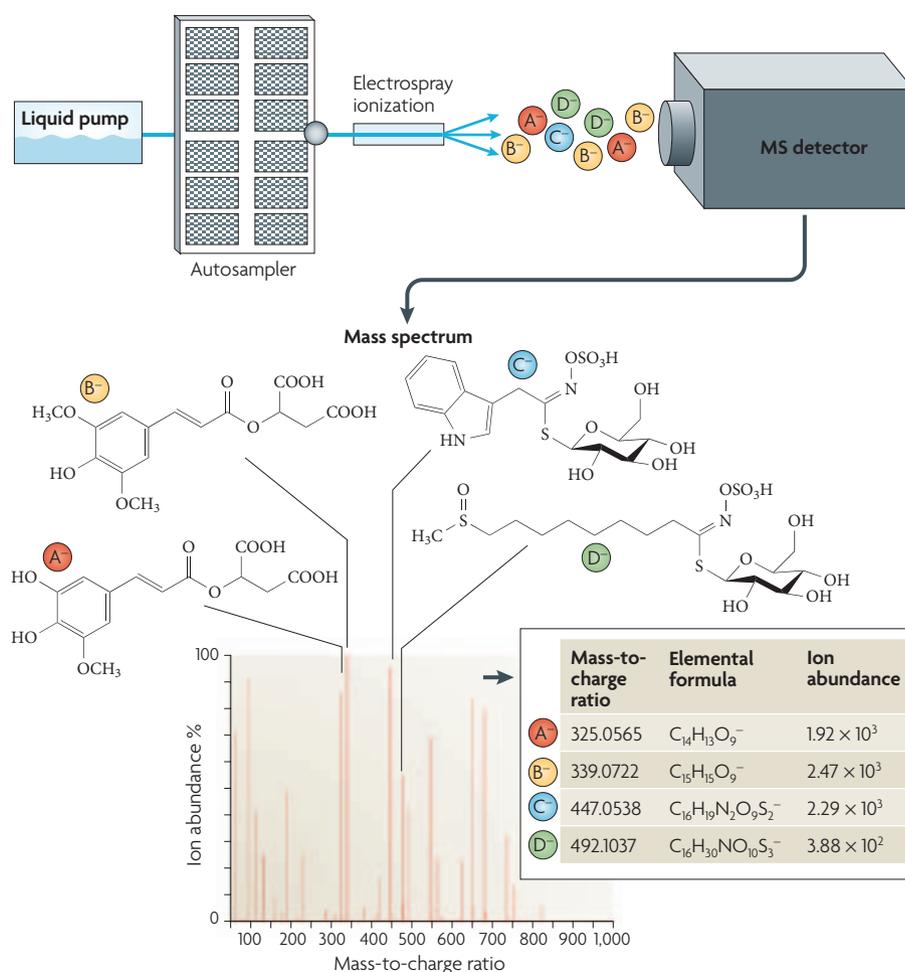

Figure 1 | **Small-molecule analysis by mass spectrometry without chromatographic separation.** Rapid mass spectrometry (MS)-based analyses can be performed on complex mixtures without chromatographic separation. Extracts are introduced into a mass spectrometer by infusion or flow injection. The mass spectrometer ionizes metabolites and measures ion abundances. Electrospray ionization (ESI) is commonly used for these analyses. Higher mass resolution mass spectrometers (for example, time-of-flight or Fourier-transform ion cyclotron resonance (FT-ICR) MS), can often predict the correct elemental formula for each detected mass, increasing the chance of inferring the correct molecular structure. The FT-ICR approach offers the highest mass resolution and accuracy, but at greater expense and complexity. These methods generate metabolite-signal patterns, such as mass versus abundance lists, that can provide complex phenotypic information and can be used to distinguish among plant genotypes or environmental perturbations. Because such metabolic fingerprinting approaches rarely provide absolute quantitative measurements (metabolite standards are scarce), classifications are based on relative ion abundances.

to reduce the differences between measurements[23]. This can greatly reduce the number of analyses needed to distinguish between two populations of samples (for example, a wild-type line from a mutant or transgenic line), but hides the biological variability between individual samples and prevents other analyses from being performed with statistical validity. In studies that use true biological replicates, it has been found that numerous replicates might be needed to distinguish subtle differences in populations (for example, 30–40 replicate plants per genotype in REF. 14).

Both MS and NMR metabolomics data are generated in the form of digitized signals, and the first stage of data analysis is processing to identify compounds in the samples and to calculate absolute concentrations (when possible) or relative signal levels. This data pre-processing involves automated or semi-automated chromatographic peak alignment, baseline correction, noise suppression, signal deconvolution and quantification of peaks (FIG. 3). These computer-aided data-processing steps are not always done, but they have been shown to significantly aid in generating
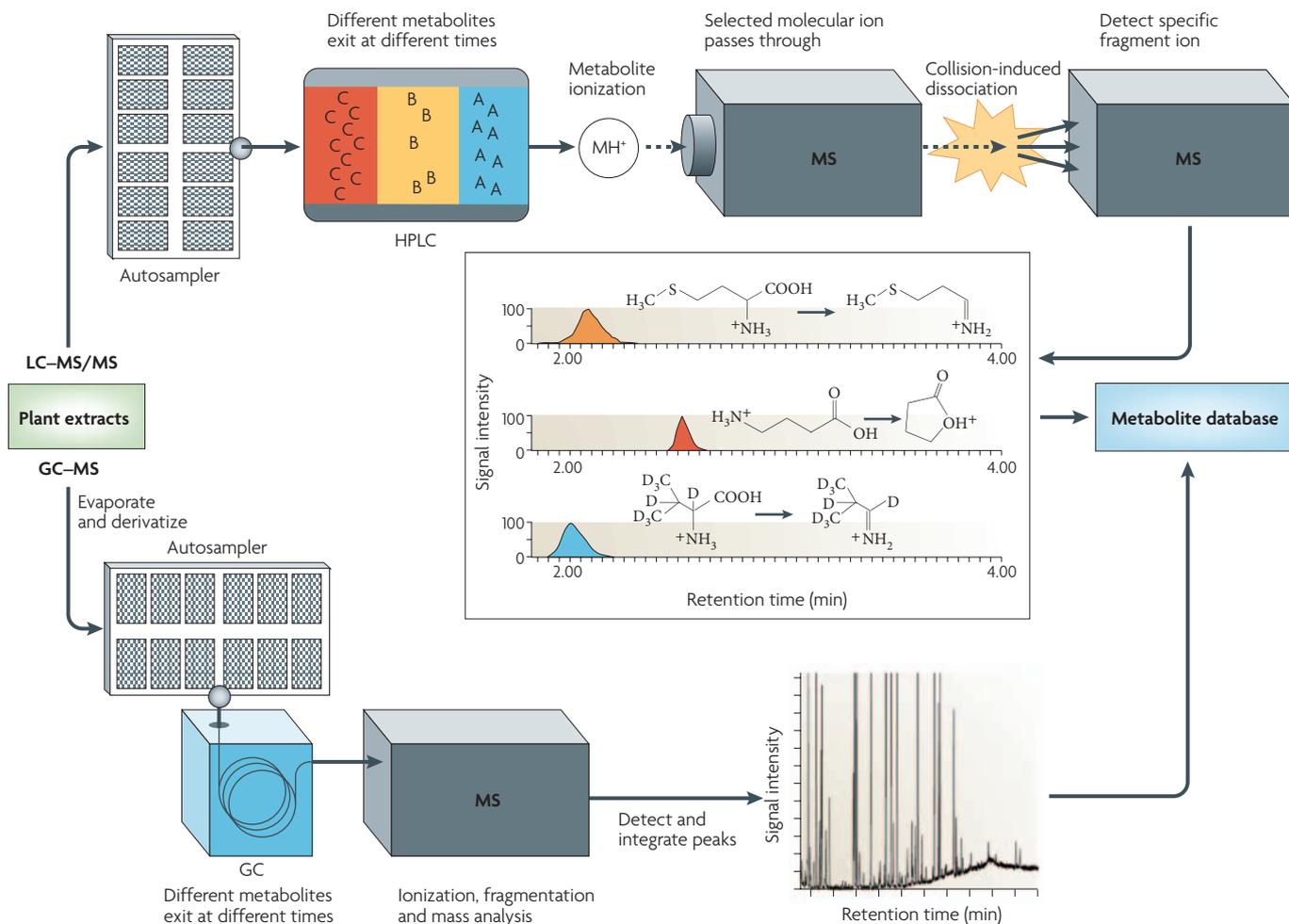
Figure 2 | **Small-molecule analysis by mass spectrometry with chromatographic separation.** Coupling of liquid chromatography (LC) or gas chromatography (GC) separations to a mass spectrometer provides physical separation of metabolites, introducing different compounds into the mass spectrometer at different times. Chromatography decreases analysis speed but offers two main benefits: metabolites with identical masses but different retention times can be distinguished, and the separation of metabolites from interfering substances allows for improved quantitative accuracy. The sample extract is injected onto a GC or LC column. For GC, chemical conversion of metabolites to less polar forms (derivatization) allows for detection of more metabolites. Flow of gas or liquid carries compounds through the column. The column retains metabolites based on physical properties, and different metabolites reach the mass spectrometer at different times. Ion abundances for individual masses are measured relative to the time the extract was introduced to the chromatography column (retention time). Measurement of the area under the chromatographic peak for a particular mass and retention time provides a measure of the amount of the metabolite that elutes at that time. For targeted metabolic profiling analyses, two stages of mass spectrometry (MS/MS) filter out interfering signals after molecular ions are broken into fragments between mass-analysis steps. Reductions in background interference by the MS/MS technique allow for more rapid and accurate metabolite analyses, which are needed for large-scale metabolite-profiling studies[46]. HPLC, high-pressure liquid chromatography.

reproducible replicates and in identifying peaks that differ between spectra[23–25]. Various software tools are used for data pre-processing, including the software that is supplied by the manufacturers of the instruments used to acquire the metabolomics data.

The next step in using data is the reduction of the data to a more manageable size. At one extreme (metabolite profiling), this can involve generating a list of compounds and their concentrations in the sample. At the other extreme (metabolic fingerprinting), spectra are converted to a list of the spectral amplitudes in the form of tens to tens of thousands of intensity numbers without compound identification or quantification.

Once pre-processed and reduced, metabolomics datasets are analysed with one or more statistical tools. These are used to find significant differences or commonalities between different plant treatments or mutants and to identify individual peaks or compounds that contribute to such differences (see Supplementary information S1 (box) and FIG. 4). The most commonly applied approach is principal components analysis, and its relatives, independent components analysis[26] and hierarchical clustering[14]. Less commonly used techniques are neural networks[27], self-organizing maps[28], fuzzy cognitive maps[29], partial least squares and linear discriminant analysis. A recent review by Janes and Yaffe[16] provides a useful overview of the theory and methods behind hierarchical clustering, principal components analysis and partial least squares.

Because metabolomics studies produce large and complex datasets, it is difficult to interpret them without visual help.

Visualization tools often represent correlations among levels of different signals — usually as some sort of two- or three-dimensional distance map or cluster diagram. Other tools place metabolomics, gene expression and proteomic/enzyme-activity data in the context of metabolic pathways (See MapMan[30] and Pathexplore in Further information). Although existing tools can be extremely useful, there is a continuing need for data exploration and visualization methods that are user friendly (easy to set-up, import data and run). Web-based tools analogous to Genevestigator[31] (see Further information), which easily allows the user to analyse large amounts of public domain Affymetrix GeneChip data, would be ideal for this purpose.

Metabolomics applications
Understanding biological processes.
Metabolomics is an attractive phenotypic assay because it can be used to detect many molecules from biosynthetically unrelated pathways. Despite many published studies documenting the measurement of specific classes of small molecule in plant biochemistry and biochemical genetics studies, there are surprisingly few published examples in which broad metabolite profiling or metabolomics were used to answer questions in plant biology. The general approach to using metabolomics to reveal phenotypic differences that are due to genetic variation or environmental perturbation is shown in FIG. 4.

One such study used metabolite profiling to characterize the freezing-tolerance response of *Arabidopsis thaliana* and to understand the function of the cold-response regulatory pathway that is regulated by the CBF transcription-factor family[32]. This study showed that the physiological process of cold acclimation significantly influenced the concentration of three-quarters of the >400 metabolite peaks that were detected by GC–TOF MS, and that the levels of most of these peaks were influenced by the CBF-mediated cold-response pathway. Because of the large number of compounds that respond to cold treatment, these findings need to be followed up with functional genomics approaches to determine which of the changes are useful in the plant's adaptation to cold.

In contrast to the large number of metabolites that respond to chilling, far fewer changes in the metabolic profile of *A. thaliana* were found following transgenic manipulation of the expression of diverse genes. For example, targeted metabolite analysis revealed that the overexpression of the
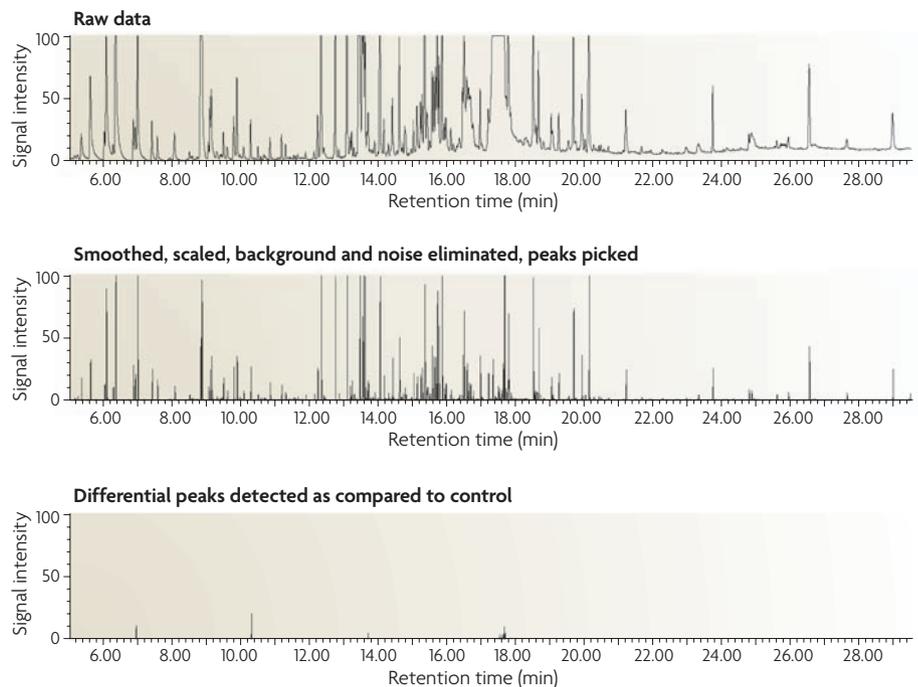

Figure 3 | **Approaches to data reduction.** Extracting and comparing metabolomic information from mass spectra is complicated by baseline variation, peak overlap, variation in retention times, noise and peak identification. The top row shows raw mass spectrometer data (a total ion chromatogram) from an extract of tomato fruit, the middle row shows the processed data produced by an automated software package that aligns the peaks, corrects the baseline, filters out noise, distinguishes between overlapping peaks and identifies peaks based on automatic comparisons of observed mass spectra with spectra in standard libraries. The bottom row shows a validation of this data-handling protocol by the extraction of three spiked compounds from an extract spectrum by comparison of samples (peaks in the figure). The y axis represents ion intensities, and the x axis represents retention time. The figure is modified with permission from Plant Research International and RIKILT, Wageningen; www.metalign.nl.

A. thaliana MYB transcription factor *PAP1* caused abnormally high accumulation of more than a dozen flavonoid family aromatic secondary compounds[11]. By contrast, non-targeted metabolite analysis of 1,800 anonymous metabolite peaks using FT-ICR MS revealed that *PAP1* overexpression did not measurably influence other classes of metabolite. These data, combined with RNA-array experiments, showed that PAP1 has a narrow effect on *A. thaliana* metabolism, and also led to the discovery of genes that encode previously uncharacterized flavonoid biosynthetic enzymes.

Although plant metabolic engineering has the potential to create healthier varieties of crops[33], the current political climate inhibits the acceptance of new transgenic crops. One argument for caution is that transgene expression might cause the accumulation of undesirable metabolites. A recent study[12] looked for unintended metabolic consequences in transgenic potatoes engineered to overproduce inulin-type fructans, compounds that seem to

promote the growth of beneficial intestinal flora. A combination of targeted and non-targeted methods was used to measure >250 metabolite peaks, and showed large differences in non-fructan metabolites among 12 conventionally bred (non-transgenic) potato varieties. Fructans were the only small molecules in the transgenic potatoes that had abundances that were found to be out of the normal range defined by the non-transgenic cultivars. This study and several NMR-based studies of tomato, tea and pea[34–36] provide evidence that the transgenic foods that have been studied are substantially equivalent to those from standard varieties that are considered safe for human consumption. A study of the metabolic consequences of introducing an insecticidal *Bacillus thuringiensis* toxin gene into corn also found no metabolic changes of concern for human health, but found altered free amino-acid levels that indicate possible changes in primary nitrogen metabolism[37]. Such studies should contribute to the scientific evaluation of transgenic food safety.

a Mutant or transgenic plants

b Analysis of metabolites

c Quantification of peaks from each sample

Peak 1　　Peak 2　　Peak 3　　Peak 4　　Peak 5　　Peak 6

Signal intensity

d

Wild type / Mutant — 23°
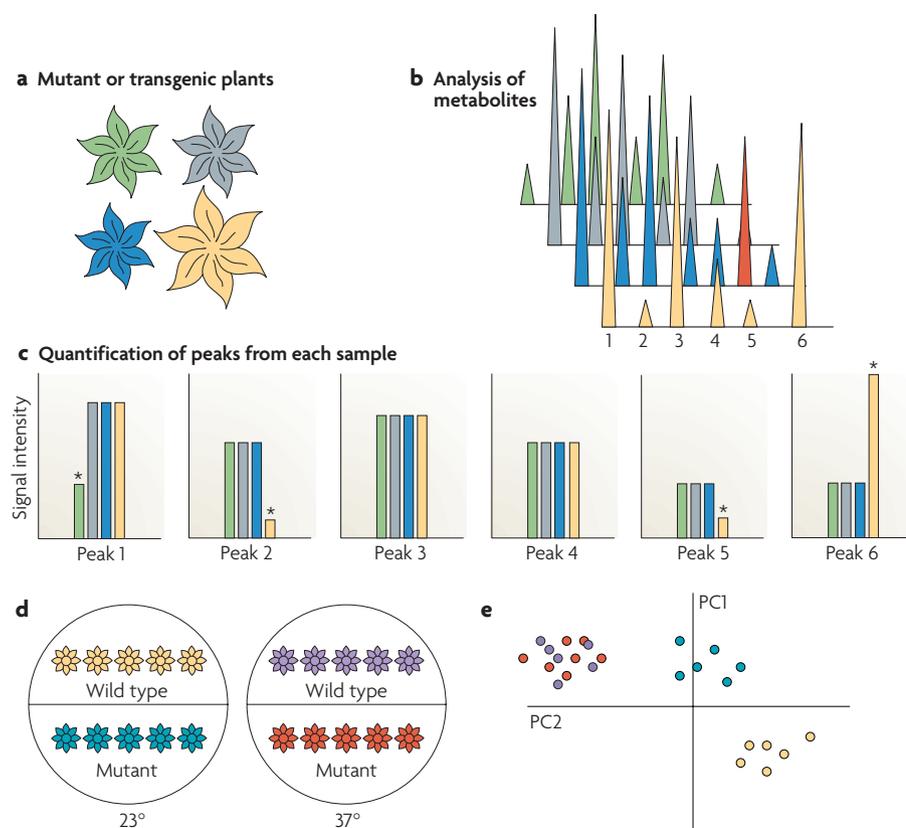
Wild type / Mutant — 37°

e

PC1

PC2

Figure 4 | **Use of metabolomics to analyse plants perturbed by genetic or environmental changes.** Small-molecule analysis can be used to identify phenotypic differences in genetic variants generated by mutagenesis or transformation, or between natural populations (parts **a**–**c**). Four genetically different plants are indicated by different colours in part **a**. Although there are clear differences in the sizes of the blue and yellow plants, these morphological distinctions cannot be used to make inferences about the status of metabolic pathways. Part **b** shows a schematic representation of the intensities of six peaks (labelled 1–6) that were separated by gas or liquid chromatography and measured by the quantification of specific ions that eluted at each time. The peaks are colour coded to correspond to the plants shown in part **a**. There is a seventh peak labelled in red in the blue plant chromatogram that represents a novel metabolite present only in this mutant. Quantitative comparison of each of the six normally detected compounds from the four plants (part **c**). The asterisk indicates statistically significant differences in amounts of specified compounds from each plant. The compound that corresponds to peak 1 is present in lower amounts in the green plant, as are compounds 2 and 5 in the yellow plant, whereas this plant has an excess of compound 6. Metabolomics and metabolite profiling can be used to evaluate the responses of plants to environmental perturbations (parts **d**,**e**). In this example, wild-type and high-temperature tolerant mutant plants were continuously grown at 23°C or were shifted to 37°C (part **d**). Part **e** shows the results of principal components analysis of metabolites from these samples. The data reveal that the 23°C grown wild-type sample (yellow circles) is phenotypically distinct from the remaining samples. By contrast, the high-temperature tolerant mutant (turquoise and red) samples from both temperatures are similar to 37°C-treated wild-type plants (purple). PC, principal component.

Mutant analysis. The analysis of mutants that are altered in the accumulation of specific metabolites has greatly increased our knowledge of the structure and regulation of biosynthetic pathways in plants[38]. Many approaches for phenotypic analysis have been described, ranging from changes in whole plant phenotypes, such as altered floral, leaf or seed pigments[39,40], to direct metabolite analysis by thin-layer chromatography[38,41], GC[42], LC[33,43,44] or novel assays for detecting specific compounds[45].

Because mutant screens typically require assaying thousands of individual samples, they have been restricted to the analysis of compounds that can be quickly and inexpensively detected (such as fatty acids). Recent advances in chromatography coupled with MS have paved the way for genetic analysis of pathways that were not previously accessible and have allowed multiple compounds to be assayed simultaneously. For example, a 1.5-minute LC–MS/MS assay for 18 of the 20 amino acids allowed screening of 10,000

A. thaliana random mutant families for changes in the levels of free amino acids in seeds[46]. Genetic-mapping-based identification of the defective gene for one such mutant revealed that a defect in the threonine aldolase *THA1* gene led to increased levels of seed threonine, presumably by reducing the conversion of threonine to glycine. This enzyme was not previously suspected to have a central role in threonine homeostasis in plants, validating the usefulness of direct analytical chemical mutant screens for extending our understanding of even well studied biosynthetic pathways. Interestingly, a different approach — analysis of the metabolome of *Medicago truncatula* cell cultures treated with the stress-related hormone methyl jasmonate — identified a strong correlation between threonine and glycine levels[47]. This led the authors to suggest that a previously uncharacterized threonine aldolase is induced in *M. truncatula* cells following jasmonate treatment, illustrating the complementary nature of open-ended metabolomics analysis and directed metabolite-profiling approaches.

It would be desirable to switch from analysing mutants for specific classes of molecule to more global metabolomics approaches. This would allow the efficient discovery of genes that influence enzymes in multiple pathways and of relationships between different metabolites and between metabolites and other traits. Schauer and co-workers[48] reported an important step in this direction. This study took advantage of a set of genetically characterized lines derived by the introduction of specific chromosomal regions of the wild species *Solanum pennellii* into the genome of cultivated tomato (*Solanum lycopersicum*) to create nearly isogenic lines[49]. Seventy-four known metabolites were analysed from field-grown ripe fruit from these lines by GC–TOF MS, and correlations were sought between these metabolites and agronomically important morphological traits, such as the ratio of fruit to plant mass, plant weight and the number of seeds per fruit. In addition to establishing phenotypic correlations, it was possible to use genetic mapping to find nearly 900 quantitative trait loci that influence fruit metabolites. Because the nearly isogenic lines that were used in this study are an enduring resource, it should be possible for researchers to track down genes that influence many metabolic traits. An analogous study of metabolomics variation across *A. thaliana* accessions was recently published and included the identification of large numbers of quantitative trait loci for the mass peaks identified by LC–TOF MS[50].

Glossary

Allelopathic
The known or postulated ability of a plant-produced chemical to modify the physiology of a nearby plant, which constitutes a form of chemical communication between plants.

Capillary electrophoresis
Separation of compounds by the different times that they take to traverse a narrow column under the influence of an electric field.

Electrospray ionization
(ESI). A gentle method for ionizing compounds that applies high voltage to a solution of compounds that are sprayed through a capillary.

Flavonoid family
A large family of structurally diverse molecules that are produced by plants. They confer common colours to leaves, flowers and fruits, absorb ultraviolet light and have other known and postulated roles in plant growth and development.

Flow-injection analysis
(FIA). Introduction of a sample into a flowing medium that transports the sample components to a detector, such as a mass spectrometer, without chromatographic separation.

Fourier-transform-infrared (FT-IR) spectroscopy
Spectroscopy that enables the identification of classes of compound by analysing their interactions with infrared light. Mass resolution is the ability to distinguish ions of different masses, usually expressed as a ratio of the ion mass to the difference in mass that the mass spectrometer can reliably distinguish.

Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry
A method for obtaining accurate measurements of the mass-to-charge ratio of ions in a complex mixture, allowing the identification and measurement of the molecules.

Freezing-tolerance response
A measure of the ability of an organism to survive when exposed to below-freezing temperatures.

Gas chromatography
(GC). Method for the separation of volatile compounds in a gas phase through differential binding to a column at elevated temperatures.

High-performance liquid chromatography
(HPLC). Method for separating compounds in solution through their differential interactions with a column.

HPLC with electrochemical or photodiode-array detection
Method for the separation of metabolites by HPLC followed by the detection of the metabolites based on their redox behaviour or spectral absorption properties.

Mass-to-charge ratio
(m/z). The ratio of the mass of an ion (in Daltons, or atomic mass units) to the number of charges on the ion.

Nominal mass
The mass of a molecule, rounded off to the nearest integer value. Two compounds with different elemental formulas can have identical nominal masses, although their exact masses might differ by a small fraction of a Dalton (or atomic mass unit).

Orbitrap mass analyser
A high-resolution mass spectrometer that traps ions in an electric field and measures their abundances and mass-to-charge ratios based on their orbital motions in the ion trap.

Targeted metabolite analysis
Quantitative analysis of a pre-selected list of metabolites.

Thin-layer chromatography
Method for the separation of molecules through their differential interactions with a matrix on a glass or plastic plate.

Dynamics of the metabolome

The functions of many metabolites are dependent on their concentrations, which in turn are determined by synthesis and turnover. Therefore, changes in the amounts of some plant metabolites indicate the extent to which their functions are induced or repressed. Examples include metabolites that are involved in nutrient storage as well as metabolites with structural, defensive, signalling and allelopathic roles. However, many other metabolites are intermediates that function as precursors, metabolic buffers and catabolic substrates for biosynthesis, homeostasis, energy transduction and transport. The primary function of these metabolites is served by their turnover, and therefore the relationship between their steady-state levels and functional status is complex and hard to predict. At one extreme, changes in the level of a metabolite such as cyclic AMP directly alter the signalling that the metabolite mediates and so reflect a functional change. By contrast, the concentration of a pathway intermediate might not change in response to large perturbations in the metabolic flux of that pathway[51]. More account of this should be taken when analysing metabolomics datasets.

Because of the spatial and temporal variation in metabolite measurements, it is also crucial that scientists develop methods for going beyond taking snapshots of metabolite levels of mixtures of heterogeneous cell types. Complementary efforts are focusing on metabolite measurements for particular cell types or organelles. These include single-cell sampling by fine-tipped probes[52] and non-aqueous fractionation methods. Promising developments in both subcellular localization and time-dependent measurements of metabolites have been made by the engineering of protein-based fluorescent indicators of metabolite levels[53]. Established methods of *in vivo* NMR spectroscopy and imaging also allow temporal and spatial resolution in the measurement of metabolite levels and turnover rates[6]. Last, recently developed methods for determining multiple fluxes in the metabolic networks of plants[54] open the possibility of integrating flux information into the interpretation of steady-state metabolomics data.

Future directions

Systems biology depends on the integrated analysis of heterogeneous data types, such as metabolomics, transcript profiling, protein expression and protein-interaction mapping. Such datasets are now becoming available for plant systems[11] and there are tantalizing possibilities for teasing out causative relationships between the levels of metabolites, the activities of enzymes, and the levels of transcripts of known and unknown function. Tools for such analyses are being developed, although this endeavour is still in its infancy. One example is the Iowa Gene Expression Toolkit (see Further information). Examples from studies of microorganisms[55] show that this is an exciting field. It remains to be seen how far such correlations can be interpreted functionally, but it seems probable that the coming years will see the generation of fruitful ideas about the metabolic networks of plants using systems approaches.

Robert L. Last is at the Department of Biochemistry and Molecular Biology and the Department of Plant Biology, A. Daniel Jones is at the Department of Biochemistry and Molecular Biology and the Department of Chemistry, and Yair Shachar-Hill is at the Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA. Correspondence to R.L.L. email: lastr@msu.edu

1. Fiehn, O. Metabolomics — the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171 (2002).
2. Dunn, W. B. & Ellis, D. I. Metabolomics: current analytical platforms and methodologies. *Trends Anal. Chem.* **24**, 285–294 (2005).
3. Sweetlove, L. J., Last, R. L & Fernie, A. R. Predictive metabolic engineering: a goal for systems biology. *Plant Physiol.* **132**, 420–425 (2003).
4. Ward, J. L., Harris, C., Lewis, J. & Beale, M. H. Assessment of ¹H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **62**, 949–957 (2003).
5. Krishnan, P., Kruger, N. J. & Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **56**, 255–265 (2005).
6. Ratcliffe, R. G. & Shachar-Hill, Y. Probing plant metabolism with NMR. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 499–526 (2001).
7. Ratcliffe, R. G., Roscher, A. & Shachar-Hill, Y. Plant NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectros.* **39**, 267–300 (2001).
8. Glinski, M. & Weckwerth, W. The role of mass spectrometry in plant systems biology. *Mass Spectrom. Rev.* **25**, 173–214 (2006).
9. Villas-Bôas, S. G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* **24**, 613–646 (2005).

10. Aharoni, A. *et al.* Nontargeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. *OMICS* **6**, 217–234 (2002).

11. Tohge, T. *et al.* Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* **42**, 218–235 (2005).

12. Catchpole, G. S. *et al.* Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc. Natl Acad. Sci. USA* **102**, 14458–14462 (2005).

13. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass. Spectrom.* **40**, 430–443 (2005).

14. Weckwerth, W., Loureiro, M. E., Wenzel, K. & Fiehn, O. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl Acad. Sci. USA* **101**, 7809–7814 (2004).

15. Ratcliffe, R. G. & Shachar-Hill, Y. Revealing metabolic phenotypes in plants: inputs from NMR analysis. *Biol. Rev.* **80**, 27–43 (2005).

16. Janes, K. A. & Yaffe, M. B. Data-driven modelling of signal-transduction networks. *Nature Rev. Mol. Cell Biol.* **7**, 820–828 (2006).

17. Fan, T. W. M. *et al.* Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. *Phytochemistry* **57**, 209–221 (2001).

18. Shachar-Hill, Y. & Pfeffer, P. *Nuclear Magnetic Resonance in Plant Biology* (American Society of Plant Biologists, Rockville, 1996).

19. Bailey, N. J., Stanley, P. D., Hadfield, S. T., Lindon, J. C. & Nicholson, J. K. Mass spectrometrically detected directly coupled high performance liquid chromatography/nuclear magnetic resonance spectroscopy/mass spectrometry for the identification of xenobiotic metabolites in maize plants. *Rapid Commun. Mass Spectrom.* **14**, 679–684 (2000).

20. Fan, T. W. M., Higashi, R. M., Lane, A. N. & Jardetzky, O. Combined use of ¹H-NMR and GC–MS for metabolite monitoring and *in vivo* ¹H-NMR assignments. *Biochim. Biophys. Acta* **882**, 154–167 (1986).

21. Morgenthal, K., Weckwerth, W. & Steuer, R. Metabolic networks in plants: transitions from pattern recognition to biological interpretation. *BioSystems* **83**, 108–117 (2006).

22. Pettersson, G. & Ryde-Pettersson, U. A mathematical model of the Calvin photosynthesis cycle. *Eur. J. Biochem.* **175**, 661–672 (1988).

23. Tikunov, Y. *et al.* A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol.* **139**, 1125–1137 (2005).

24. Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X. & Sumner, L. W. MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Anal. Chem.* **78**, 4334–4341 (2006).

25. Jonsson, P. *et al.* High-throughput data analysis for detecting and identifying differences between samples in GC–MS-based metabolomic analyses. *Anal. Chem.* **77**, 5635–5642 (2005).

26. Scholz, M., Gatzek, S., Sterling, A., Fiehn, O. & Selbig, J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **20**, 2447–2454 (2004).

27. Goodacre, R., York, E. V., Heald, J. K. & Scott, I. M. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. *Phytochemistry* **62**, 859–863 (2003).

28. Hirai, M. Y. *et al.* Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **101**, 10205–10210 (2004).

29. Wurtele, E. S. *et al.* MetNet: software to build and model the biogenetic lattice of *Arabidopsis*. *Comp. Funct. Genomics* **4**, 239–245 (2003).

30. Thimm, O. *et al.* MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).

31. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. & Gruissem, W. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* **136**, 2621–2632 (2004).

32. Cook, D., Fowler, S., Fiehn, O. & Thomashow, M. F. A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **101**, 15243–15248 (2004).

33. Van Eenennaam, A. L. *et al.* Engineering vitamin E content: from *Arabidopsis* mutant to soy oil. *Plant Cell* **15**, 3007–3019 (2003).

34. Charlton, A. *et al.* NMR profiling of transgenic peas. *Plant Biotech. J.* **2**, 27–35 (2004).

35. Le Gall, G., Colquhoun, I. J. & Defernez, M. Metabolite profiling using ¹H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). *J. Agric. Food Chem.* **52**, 692–700 (2004).

36. Noteborn, H. P., Lommen, A., van der Jagt, R. C. & Weseman, J. M. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J. Biotechnol.* **77**, 103–114 (2000).

37. Manetti, C. *et al.* A metabonomic study of transgenic maize (*Zea mays*) seeds revealed variations in osmolytes and branched amino acids. *J. Exp. Bot.* **57**, 2613–2625 (2006).

38. Benning, C. Genetic mutant screening by direct metabolite analysis. *Anal. Biochem.* **332**, 1–9 (2004).

39. Rose, A. B., Li, J. & Last, R. L. An allelic series of blue fluorescent *trp 1* mutants of *Arabidopsis thaliana*. *Genetics* **145**, 197–205 (1997).

40. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* **126**, 485–493 (2001).

41. Chapple, C. C., Vogt, T., Ellis, B. E. & Somerville, C. R. An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. *Plant Cell* **4**, 1413–1424 (1992).

42. Somerville, C. & Browse, J. Plant lipids: metabolism, mutants and membranes. *Science* **252**, 80–87 (1991).

43. Cheng, Z. *et al.* Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *Plant Cell* **15**, 2343–2356 (2003).

44. Valentin, H. E. *et al.* The *Arabidopsis* vitamin E pathway gene 5–1 mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. *Plant Cell* **18**, 212–224 (2006).

45. Conklin, P. L., Saracco, S. A., Norris, S. R. & Last, R. L. Identification of ascorbic acid-deficient *Arabidopsis thaliana* mutants. *Genetics* **154**, 847–856 (2000).

46. Jander, G. *et al.* Application of a high-throughput HPLC–MS/MS assay to *Arabidopsis* mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality. *Plant J.* **39**, 465–475 (2004).

47. Broeckling, C. D. *et al.* Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *J. Exp. Bot.* **56**, 323–336 (2005).

48. Schauer, N. *et al.* Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnol.* **24**, 447–454 (2006).

49. Eshed, Y. & Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147–1162 (1995).

50. Keurentjes, J. J. *et al.* The genetics of plant metabolism. *Nature Genet.* **38**, 842–849 (2006).

51. Fell, D. A. Enzymes, metabolites and fluxes. *J. Exp. Bot.* **56**, 267–272 (2005).

52. Tomos, A. D. & Leigh, R. A. The pressure probe: a versatile tool in plant cell physiology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 447–472 (1999).

53. Lalonde, S., Ehrhardt, D. W. & Frommer, W. B. Shining light on signaling and metabolic networks by genetically encoded biosensors. *Curr. Opin. Plant Biol.* **8**, 574–581 (2005).

54. Ratcliffe, R. G. & Shachar-Hill, Y. Measuring multiple fluxes through plant metabolic networks. *Plant J.* **45**, 490–511 (2006).

55. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol.* **19**, 45–50 (2001).

FURTHER INFORMATION
Robert L. Last's homepage:
http://www.bch.msu.edu/faculty/last.html
A. Daniel Jones's homepage:
http://www.bch.msu.edu/faculty/jones.html
Yair Shachar-Hill's homepage:
http://www.plantbiology.msu.edu/shachar-hill.shtml
AraCyc *Arabidopsis* metabolic pathway annotations:
http://www.arabidopsis.org/biocyc/index.jsp
Genevestigator:
https://www.genevestigator.ethz.ch/at
Iowa Gene Expression Toolkit:
http://metnet.vrac.iastate.edu
KEGG Pathway Database:
http://www.genome.jp/kegg/pathway.html
***L. pennellii* Introgression Lines:**
http://zamir.sgn.cornell.edu/Qtl/il_story.htm
MapMan tool:
http://gabi.rzpd.de/projects/MapMan
Metabolic modeling:
http://www.hort.purdue.edu/cfpesp/models/models.htm
Metabolic Ontology Working Group:
http://msi-ontology.sourceforge.net
MetAlign tool for GC– or LC–MS data analysis:
http://www.rikilt.wur.nl/UK/services/MetAlign+download
Pathexplore tool:
http://www.cs.nyu.edu/~pathexp/index.html
Plant metabolites database:
http://prime.psc.riken.jp/?action=metabolites_index
Solcyc *Solanaceae* metabolic pathway annotations:
http://www.sgn.cornell.edu/tools/solcyc
The Golm Metabolome Database:
http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html
WTEC Panel Report on International Research and Development in Systems Biology:
http://www.wtec.org/sysbio/report/SystemsBiology.pdf

SUPPLEMENTARY INFORMATION
See online article: S1 (box)
Access to this links box is available online.