The availability of complete genomic sequences of many model organisms has made it possible to perform highly informative genome-wide functional analyses. For multicellular organisms (including the nematode Caenorhabditis elegans, the fruit fly Drosophila melanogaster, the plants Arabidopsis thaliana and rice, as well as mouse), phenotypic analysis of genetic mutations is still one of the most effective ways to explore the function of a gene. Collections of strains with mutations in nearly every gene are now available, making it possible to analyze the phenotypes of a large number of independent strains. However, conventional analytic approaches, such as high-magnification microscopy at the single-cell level, require manual manipulation of samples and screening by eye, thus limiting throughput and presenting bottlenecks to large-scale genetic studies in multicellular organisms. Therefore, development of high-throughput methods, including automation in phenotyping and screening, is a strategy that is now coming to fruition [1]. Systematic large-scale phenotyping efforts have begun to generate information on a previously unattainable scale. For example, it was recently shown that even a highly dynamic process such as the division of human cells can be studied on a genome-wide scale by live imaging [2].

Cultured cells have also proved amenable to high-throughput phenotyping [2]. Although more challenging, the study of living organisms can provide insights into biological pathways, regulatory networks and/or cellular activity and behavior not obtainable from cultured cells [3-6]. Large-scale acquisition of phenotypic data can then predict important biological outputs, such as the roles of individual genes in development. Thus, high-throughput phenotyping approaches (that is, phenomics) can encompass a broad range of model systems and techniques aimed at understanding the link between genotype and phenotype.

A good example of the evolution of high-throughput phenotyping is provided by RNA interference (RNAi) screens in the worm C. elegans, where recent advances in robotic sample preparation have facilitated high-throughput screens. However, C. elegans is only one of many systems in which innovative technologies for high-throughput studies are being developed. Indeed, the development and use of robotic platforms has also enabled high-throughput phenotypic analysis of plant growth and development at a larger physical scale. Here, we use C. elegans and Arabidopsis as the primary examples of the exciting new wave of approaches to functional genomics [7-10]. We focus on current advances in high-throughput phenotyping (HTP) for the analysis of C. elegans and Arabidopsis, as lessons learned from these organisms can be broadly applied to other animal and plant species.

RNAi and high-throughput phenotyping in C. elegans
Reverse genetic screening has proved a powerful method to identify gene function [11,12]. RNAi is a well-conserved phenomenon observed in many different organisms [13-23]. It was originally discovered in plants, and became one of the first genome-wide techniques used to study loss-of-function phenotypes in several model systems and in mammalian cell culture [24-27]. RNAi screens have become invaluable tools in assessing genotype-phenotype relationships [28,29], and several
large-scale RNAi libraries have been generated to identify essential genes and those with novel functions [16,30-32]. For example, an RNAi library of 750 ovary-enriched genes was generated to study the function of genes involved in embryogenesis [31]. RNAi genome-wide screens in *Drosophila* have been performed using cell culture [12,15]. The genome-wide collection of transgenic constructs that has been prepared for *in vivo* screening has underpinned a number of studies, including a screen that led to the identification of the sex-peptide receptor of *Drosophila* [33,34]. Large-scale mutagenesis and phenotyping projects are also under way in mammalian cells, and are likely to yield similarly important results [23,35].

Over the past few years, increasingly sophisticated image-analysis tools have facilitated RNAi screens. Initially, high-throughput RNAi phenotyping focused on endpoint observations, such as worm morphology and viability, and thus were unable to distinguish between the primary and secondary effects of gene silencing. It is now possible to perform rapid and accurate phenotyping of embryonic lethality in different *C. elegans* developmental stages by analyzing high-throughput image data [36]. The image-analysis system DevStaR uses a hierarchical approach, in which the output of one step is the input for the next, for automatic classification of the developmental stages of worms from a population of mixed stages (including adult, larval and embryonic stages; Figure 1). The system consists of several layers that result in the identification of an area of interest: a segmentation of pixels within this region; a model-based component that breaks the pixel regions into object parts; and finally, a categorization of those objects using a machine-learning approach. This multi-layered object-recognition software offers the computational flexibility for generalized object-recognition problems and, therefore, is not limited to high-throughput worm screens [36].

New computer-aided visualization methods, which automatically distinguish images of worms grown in agar plates, are also available [37]. In addition, automated phenotyping based on machine-learning methods of images obtained from movie frames can also be used to study embryo development [38]. These systems overcome previous bottlenecks in image analysis by scoring image data in a fully automated manner and providing rapid quantitative output that would not be obtainable at high-throughput by manual scoring. Because high-throughput phenotyping generates a large volume of data, which need to be standardized, normalized and analyzed, statistical and bioinformatics approaches are also becoming increasingly available.

**Automated screening using worm-sorters**

Further advances combining RNAi and sample sorters have enabled rapid selection of organisms with

---

**Figure 1. Simplified illustration of the DevStaR system.** The input images are from 96-well plates containing a population of mixed stages of adult, larva and embryo worms. Each pixel within the wells is first grouped together (contrast measure). Pixels are then grouped into connected components based on a threshold value (pairwise symmetry score). Third, for the object categorization, a support vector machine (SVM) learning method assigned a score to each category. Finally, as a result of the segmentation and labeling, DevStaR distinguishes adult (blue), larva (red) and embryo (green) worms.
phenotypes of interest for a variety of assays, including genetic screens (Figure 2). Small-animal sorters, such as the BIOSORT/COPAS (complex object parametric analysis and sorter) machine, use a flow-through technique and a profiler system that can analyze up to 100 live animals per second and generate fluorescence emission profiles of the C. elegans body. COPAS has recently been used to analyze the expression pattern of 900 predicted C. elegans genes [39]. By analyzing large numbers of animals from a mixed-stage culture, Dupuy and colleagues [39] generated digitized chronograms of the intensity of gene expression throughout post-embryonic development. This machine allows researchers to study gene expression patterns in a large population of adult animals with a quantitative read-out. However, its sensitivity in sorting non-adult animals, such as embryos and larvae, is limited. Therefore, as a complementary approach, fluorescence-activated cell sorting (FACS) can be employed. By using embryo FACS (eFACS), large numbers of living embryos enriched in any desired embryonic stage can now be selected. Given the availability of different fluorescent marker genes, eFACS enables the assay of embryonic stage-specific gene expression in a high-throughput manner. Moreover, the need for a fast and reliable way of identifying phenotypic alterations in larvae, after modulating or eliminating genes, led researchers to develop a method to sort live C. elegans larvae (laFACS) [40]. Modifying a FACS machine enabled the collection of large quantities of live mutant worms from mixed populations, thereby expanding the arsenal of tools for high-throughput ‘sample preparation’ for genetic screens. Because these flow-cytometry-based systems sort animals only on one-dimensional intensity profiles, microfluidics chips have been developed to obtain single-cell resolution [41,42]. Microfluidic chips can be designed to function as small-scale sorters with channels and computer-controlled valves that control the environment surrounding the organism and restrict the worms’ movements. This technology, when combined with automated image processing, allows high-throughput, non-biased phenotyping, imaging and screening of multicellular organisms [43].

The resolution at which biological samples can be analyzed has greatly increased in recent years as fluorescence microscopy strategies have been developed to characterize gene expression at the single-cell level in C. elegans [1,44]. Methods to quantitatively measure gene-expression dynamics with cellular resolution are anticipated, and will be advantageous to functional genomic studies. However, the challenge of capturing high-resolution images that represent the entire sample remains formidable. Extensive high-throughput time-lapse fluorescence microscopy will only become a reality with improvements to the automation of microscopy imaging and the processing of large datasets.

**High-throughput phenotyping for plant biotechnology**

The identification of genes that underlie phenotypic variation for complex agronomic traits such as biomass and drought tolerance will be key to biotechnology-aided crop improvement. Because such traits are often controlled by many genes that are also heavily influenced by the environment, the discovery of their genetic basis often requires large-scale phenotyping strategies. Mutational methods such as chemical or fast neutron mutagenesis can be used in forward genetic screens, whereas insertional mutagenesis via T-DNA lines or transposons is used to generate libraries of loss-of-function mutants for reverse genetic screens. Arabidopsis has led the way in plant phenotypic profiling because insertional mutations of most genes are available [45-51]. Rice, as a leading experimental model for monocotyledonous crops, also
has a panel of insertional mutant lines [52]. Insertional mutagenesis has also been applied to other crops, including maize and *Medicago truncatula* [53,54]. However, advances in phenomics will be essential to fully realize the potential of these powerful genetic resources.

The investigation of complex traits such as root morphology, leaf size, plant height, flower shape or seed weight requires analyzing hundreds to thousands of plants, which poses a major challenge. Furthermore, gene response as a function of the environment must be accounted for. For this reason, tools specific for digital phenotyping together with automation of this process in controlled environments are necessary for high-throughput screening of plant phenotypes. Digital phenotyping offers the major advantage that data can be reanalyzed when new traits of interest or new types of measurements emerge. As the demand for digital image-acquisition technologies increases, several efforts have been made to generate software tools capable of producing objective and quantitative analyses of large image sets. Automated platforms have been developed for *Arabidopsis* and for crop plants to allow different aspects of automated visualization and image quantification. For example, the PHENOPSIS platform was used to dissect plant responses to soil water deficit in a collection of natural acces-sions of *Arabidopsis* [55]. The PHENODYN platform imposes drought scenarios and has been used to image maize and rice plants [56]. In addition, several efforts to improve aspects of automated visualization and image quantification for high-throughput phenotype scoring (for example, seed germination, hypocotyl growth, leaf area development and root growth dynamics) have been made for *Arabidopsis*. Specifically, the high-throughput seed-germination analysis platform GERMINATOR was used to screen for natural variation in a population of 165 recombinant inbred lines, which revealed several quantitative trait loci (QTLs) for salt tolerance [57]. High-resolution measurements of hypocotyl growth and shape have been obtained by automated quantification of time series of electronic images using HYPOTrace [58]. Other examples of fully or partially automated imaging platforms for non-destructive image-based phenotyping are LeafAnalyser, LAMINA and GROWSCREEN 3D [59-61]. These computer-based tools provide quantitative descriptors for leaf shape and size. A shortcoming of most of these tools is that they are designed to address very specific questions. Moreover, most traditional phenotype-scoring systems are based on endpoint analysis, and therefore do not easily capture the dynamic aspects of complex traits.

Recent approaches to capture these aspects have incorporated time-course data acquisition so that transient events and subtle temporal changes can be observed. However, the challenge of observing dynamic growth processes and responses to environmental stimuli, through the combination of automated time-lapse imaging with automated image analysis, remains [62]. Many image-analysis-based software tools have focused on quantifying root growth rates and root structure. Advances in machine vision and computation of automatic trait evaluation have facilitated digital reconstruction of root systems and have potentially increased the levels of throughput for phenotyping in plants. Examples of software that allow higher-throughput phenotyping are RootTrace [63], KineRoot [64], SmartRoot [65], RootLM [66], Phytomorph [67,68], RootFlow [69] and WinRhizo [70].

Many high-throughput methods have been developed for *Arabidopsis*, aided by its small size. For crop plants, an automatic imaging system has been applied to monitoring rice growth [71]. Moreover, a foundation for high-throughput automatic phenotyping for QTL analysis of root system architecture (RSA) traits of crop plants has been laid recently. To capture the root-system topologies of diverse rice cultivars, inbred lines were grown in a transparent gel substrate and imaged at high resolution. The resulting images were combined in an analysis pipeline that automatically extracted RSA measurements. Using a machine-learning approach,
therefore, there is a pressing need for clearly defined experimental parameters and quantitative benchmarks. Include the lack of standardized vocabulary terms, current limitations to the reuse and sharing of such data easily compared and shared between labs. However, be of great benefit if large-scale phenotypic data could be accessed to store and search these large datasets. It would and analyzed. This increases the need for community standards and terms agreed upon by a given community. To achieve this goal, databases that contain phenotypic information and, especially, integration of phenomic and other genome-wide data are required. Multi-organism phenotype-genotype databases that facilitate cross-species identification of genes associated with orthologous phenotypes are now becoming available (for example, PhenomicDB) [83,84]. In the next few years, the ability to harvest the full benefit of such large datasets can only be obtained by combining the genomic, epigenomic, transcriptomic, proteomic, metabolomic and phenomic data into shared databases. This resource will be invaluable for the investigation and eventual elucidation of molecular mechanisms regulating the biology of multicellular organisms, and will form a comprehensive description of the whole organism, opening new paths into systems biology.

Acknowledgements
We are grateful to Tim Horn and Chris Topp for providing the ‘Root Imaging Platform’ design. We thank Ryan Baugh and Brad Martsberger, the members of the Benfey lab and the reviewers for their helpful comments. Our work in this area is funded by the NIH as well as the NSF Arabidopsis 2010 program and NSF Plant Genome Research Program.

Published: 28 March 2011

References
1. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, McKay S, Okada HM, Pan J, Schulz AK, Tu D, Wong K, Zhao Z, Alexeyenko A, Burglin T, Sonnhammer E, Schnabel R, Jones SJ, Marra MA, Baillie EL, Moerman DG: High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. PLoS Biol 2007, 5:237.
2. Neumann B, Walter T, Hériché KJ, Bulkescher J, Eiffe H, Conrad C, Rogers P, Poser I, Helé M, Liebel U, Cetin C, Schicknann F, Pau G, Käble R, Wünsche A, Satagopam V, Schmitz MH, Chapuis C, Gerlich DW, Schneider R, Els R, Huber W, Peters JM, Hyman AA, Durbin R, Pepperkok R, Ellenberg J: Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 2010, 464:721-727.
3. Hamilton B, Dong Y, Shindo M, Liu W, Odell I, Ruark G, Lee SS: A systematic RNAi screen for longevity genes in C. elegans. Genes Dev 2005, 19:1544-1555.
4. Kambris Z, Brun S, Jang IH, Nam HJ, Romeo Y, Takahashi K, Lee WJ, Ueda R, Lemaître B: Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation. Curr Biol 2006, 16:808-813.
5. Mummery-Widmer JL, Yamazaki M, Stoecker T, Novatchkova M, Bhalerao R,
collection: an Arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. \(\text{Plant Res} 2007, 120:157-165.\)

50. Rosso MG, Li Y, Strzhov N, Reiss B, Dekker K, Weisshaar B: An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. \(\text{Plant Mol Biol} 2003, 53:247-259.\)

51. Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Weigel D, Ahn JH, Blázquez MA, Borevitz JO, Christensen SK, Fankhauser C, Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B: A new algorithm for computational image analysis of deformable motion at high spatial and temporal resolution applied to root growth. Roughly uniform elongation in the meristem and also, after an abrupt acceleration, in the elongation zone. \(\text{Plant Physiol} 2003, 132:1138-1148.\)

52. Sozzani R, Benfey PN: High-throughput phenotyping of multicytoplasmic organisms: finding the link between genotype and phenotype. \(\text{Genome Biol} 2011, 12:219.\)