SYSTEMATIC GENOME-WIDE SCREENS OF GENE FUNCTION

Anne E. Carpenter and David M. Sabatini

By using genome information to create tools for perturbing gene function, it is now possible to undertake systematic genome-wide functional screens that examine the contribution of every gene to a biological process. The directed nature of these experiments contrasts with traditional methods, in which random mutations are induced and the resulting mutants are screened for various phenotypes. The first genome-wide functional screens in *Caenorhabditis elegans* and *Drosophila melanogaster* have recently been published, and screens in human cells will soon follow. These high-throughput techniques promise the rapid annotation of genomes with high-quality information about the biological function of each gene.

TRANSCRIPTIONAL PROFILING The expression of thousands of genes can be measured simultaneously by spotting an array of DNA sequences on a glass slide and hybridizing a cell population's fluorescently labelled mRNA (or reverse-transcribed cDNA) to the slide. The fluorescence intensity of each spot corresponds to the prevalence in the cells of that nucleic acid species.

Whitehead Institute for Biomedical Research, MIT Department of Biology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. Correspondence to D.M.S. e-mail: sabatini@wi.mit.edu
doi:10.1038/nrg1248

The success of genome sequencing projects has allowed biologists to identify almost all the genes that are responsible for producing the biological complexity of several model organisms. The next important task is to assign a function to each of these genes. Several genomic approaches have recently been developed for this purpose, but these methods are indirect and usually based on gene-sequence or expression patterns[1]. For example, genes with a similar expression pattern, as shown by TRANSCRIPTIONAL PROFILING, often function in the same biological processes. Function can also be inferred when a protein is found to physically interact with a protein of known function in a YEAST TWO-HYBRID ASSAY or through co-precipitation in purified complexes. Furthermore, gene function can be predicted by identifying, *in silico*, pairs of genes whose evolution is correlated between organisms, or whose homologues are fused into a single gene in other organisms. Although these strategies have yielded large amounts of useful information when used in a high-throughput fashion, these data sets as a whole should be treated with caution[2]. False-positives are a substantial problem[1], possibly resulting from technical limitations or the predictive nature of these strategies, although the reliability of predictive data can be improved notably by combining two or more independent techniques to predict function[1,3].

Observing the effects of perturbed genes on cells or organisms using forward or reverse genetics is more reliable than predictive approaches and has been a standard strategy for biological research. It has recently become possible to test every gene in an organism's genome systematically owing to the availability of collections of gene-perturbing reagents or of organisms with known gene knockouts. Examples of these gene-targeted collections include a set of yeast strains wherein each strain has one gene that is knocked out, and sets of RNA INTERFERENCE (RNAi) reagents that allow the expression of each *Drosophila melanogaster* or *Caenorhabditis elegans* gene to be decreased.

The use of these collections to conduct genome-wide screens has several important advantages compared with screening randomly generated mutants. First, each member of the collection of gene-targeted organisms or reagents is already sequenced and is easily and instantly identifiable. Therefore, changes in phenotype (or lack thereof) can be recorded for every gene, rather than painstakingly following up on a handful of 'hits'. A second important advantage of systematic screens is that every gene is tested in a finite number of samples. By contrast, a collection of random mutants must contain many times the total number of genes in the organism to cover most of the genome. Even then, statistical sampling, and more importantly, the non-random nature of 'random' mutagenesis means that some genes are always missed. In a remarkable example of this phenomenon, a traditional screen of 29,000 randomly generated yeast

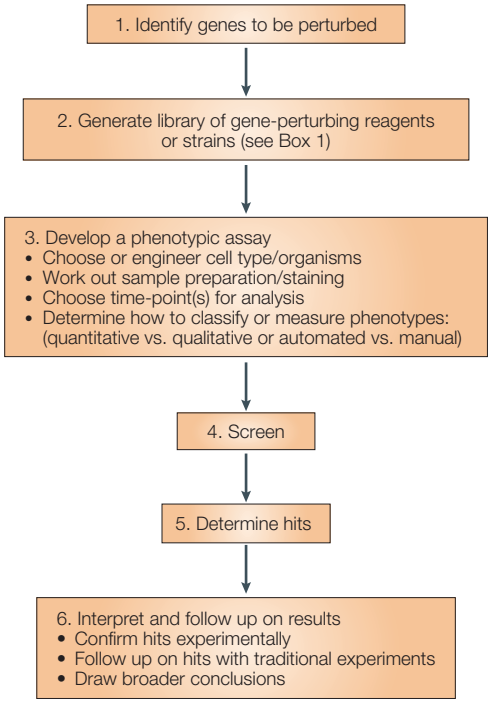


Figure 1 | **Steps involved in the design and implementation of a systematic genome-wide functional screen.** See the main text for details of each step.

mutants found three important genes that were involved in catabolite degradation[4], whereas a genome-wide screen of 5,000 systematically generated yeast mutants found the original three as well as six new genes[5]. Systematic screens therefore save labour and materials and give a more complete picture of the genes that underlie a biological phenomenon.

The overall design and implementation of a systematic genome-wide functional screen (FIG. 1) is described in this review. The first step, to identify the genes to be perturbed, relies heavily on bioinformatics to convert raw genome-sequence data to a list of predicted genes. The second step involves generating the library of gene-perturbing reagents or strains. In the third step, expertise in cell biology and instrumentation is required to develop a biologically meaningful phenotypic assay, preferably in a high-throughput format. After the gene-perturbed collection is screened (the fourth step), the data are analysed statistically to identify hits (the fifth step). In the final step, the resulting data are analysed using bioinformatics to generate broader conclusions, and often the results are followed up experimentally. Coupled with a new crop of high-throughput technologies, systematic genome-wide functional screens allow biologists to undertake experiments of unprecedented scale, in which the entire set of genes in the genome of a model organism is tested for biological function. These screens, as applied to multicellular organisms, have been published in the past few months and are likely to become a routine tool of modern biology over the next decade.

YEAST TWO-HYBRID ASSAY
One protein is fused to a transcriptional activation domain and the other to a DNA-binding domain, and both fusion proteins are introduced into yeast. Expression of a reporter gene with the appropriate DNA binding sites upstream of the promoter indicates that the two proteins physically interact.

RNA INTERFERENCE
(RNAi). The process by which the introduction or expression within cells of single- or double-stranded RNA leads to the degradation of the encoded mRNA and therefore to gene suppression.

LAMELLA
The dense, actin-rich structure that extends the leading edge of a migrating cell.

microRNAs
Tiny, noncoding RNAs that are probably involved in gene regulation.

Identifying genes to be perturbed

The set of genes to be targeted must be identified and catalogued, unless a collection of gene-perturbing reagents is already available. This information is typically available at each model organism community's main web site (see online links box). Although many genomes have been sequenced, the precise identification of genes that are expressed is still a work in progress for all organisms, because the sequence features that govern transcription, splicing and translation are not fully understood[6–8]. Therefore, although the term 'genome scale' probably would have been more appropriate for the screens that have been conducted so far, the term 'genome-wide' is generally used loosely to indicate that most known genes have been targeted (>85%). If genome-wide collections are not available for an organism or if sample preparation is laborious, interesting screens can be conducted by focusing on a set of candidate genes. For example, 50 de-ubiquitylating enzymes were screened by RNAi for involvement in the nuclear factor of κB (NF-κB) pathway in human cells[9], 100 putatively constitutively active GTPases were tagged with cyan fluorescent protein and observed for their localization and effects on actin morphology in mammalian cells[10], and selected cytoskeleton and motor-related *Drosophila* genes were screened for effects on LAMELLA formation and mitotic microtubule function[11,12]. Non-protein coding genes, such as those encoding microRNAs, have begun to be identified but methods to systematically target them are not available.

Generating gene-perturbing reagents or strains

Once the set of genes has been identified, they can then be targeted by several gene-perturbing strategies (BOX 1) to generate genome-wide collections (TABLE 1). The time between sequencing a genome and producing a genome-wide collection has decreased rapidly even as more complex genomes are tackled. Although technological advances deserve much of the credit, the decreasing cost of oligonucleotides has made production of genome-wide collections more feasible, as most methods for generating gene-targeting reagents rely heavily on DNA oligonucleotides and PCR.

Homologous recombination. The most direct way to assess gene function is to knock out the gene at the DNA level in the genome. This type of knockout is very difficult to produce in most organisms[13] but is easily done in yeast and is possible in cultured mouse-embryonic stem (ES) cells by homologous recombination (BOX 1). The first available genome-wide collection of any kind was a set of ~6,000 *Saccharomyces cerevisiae* knockout (deletion) strains that were generated in this manner[14]. The complete loss-of-function alleles obtained through knockouts simplify the interpretation of phenotypes. It should be noted, however, that the partial and gain-of-function alleles generated by random mutagenesis, although more difficult to characterize, are often informative. This type of information cannot be obtained with the complete knockout strategy.

HOMOLOGOUS RECOMBINATION
The process by which segments of DNA are exchanged between two DNA duplexes that share high sequence similarity.

MOLECULAR BAR CODES
Short, unique, engineered DNA sequences that are used as tags. For example, the bar code on each yeast deletion strain allows the identity of the strain to be determined by sequencing the code or by hybridizing DNA from the strain onto a microarray.

S2 CELLS
A cell line that is isolated from dissociated *Drosophila melanogaster* embryos. The cell line is phagocytic, which might contribute to its susceptibility to RNAi.

INTERFERON RESPONSE
A primitive antiviral mechanism that triggers sequence-nonspecific degradation of mRNA and downregulation of cellular protein synthesis.

A powerful feature of the yeast collection, that should be included in future collections in other organisms, is that each strain has a unique molecular bar code that allows highly parallel screens to be done in one culture vessel (BOX 1). In this approach, a selection or a competitive growth condition is applied to a pooled culture containing approximately equal proportions of each strain. A microarray is then used to measure the enrichment or diminishment of each bar code (and, therefore, each gene) in the initial versus the final culture (FIG. 2c). This allows for very efficient functional screening of the entire yeast genome without laborious sample handling.

RNA interference. The breakthrough and improvements in RNA interference have allowed systematic knockouts, or, more accurately, 'knockdowns', in new organisms. Several recent reviews discuss the mechanisms, biological functions and practical considerations of using RNAi reagents[15–20], which can be delivered to cells in several ways (BOX 1). RNAi differs from permanent knockout technologies because the protein is not necessarily eliminated from the cells completely (hence the term 'knockdown'): a small amount of mRNA might escape degradation and if the rate of protein turnover is slow, some protein can remain in the cell for a long time, which might weaken or mask the true phenotype. This can actually be a useful phenomenon, comparable to a hypomorphic allele, because it allows the study of gene knockdowns that would be lethal as knockouts. The temporal induction of RNAi allows cells to be monitored at time-points after the amount of the protein of interest begins to decline so that initial, direct effects can be observed. For essential genes, this means that RNAi allows the observation of phenotypic effects before death.

Unfortunately, RNAi against some genes is not uniformly effective for unknown reasons. For some genes, cells might have mechanisms to sense low amounts of certain proteins and boost expression levels accordingly. Also, some cell types, such as neurons, seem to be refractory to certain methods of RNAi[21]. The RNAi techniques that depend on transfection of cultured cells usually result in a mixed population of cells, some of which escape the RNAi effect, although the transfection efficiency of short interfering RNA (siRNA) is typically higher than that of plasmid DNA[18] (BOX 1).

Other gene-perturbing strategies. Random insertional mutagenesis, or gene trapping, can be used to perturb gene function or to attach an epitope or reporter tag, such as green fluorescent protein (GFP), to a gene. Although they are randomly generated, insertional mutants can be sequenced and placed into a collection to

Box 1 | **Gene-perturbing strategies**

Homologous recombination

In the process of HOMOLOGOUS RECOMBINATION, the cell's machinery replaces an endogenous gene with an introduced cassette flanked by regions that are homologous to sequences upstream and downstream of the targeted gene. For the genome-wide set of *Saccharomyces cerevisiae* knockouts, each gene from the start to stop codon was replaced with a cassette containing a kanamycin resistance gene and two unique MOLECULAR BAR CODES. Homologous recombination is not nearly as straightforward as this in most multicellular eukaryotes, so knockout collections in other organisms have lagged behind (TABLE 1).

RNA interference (RNAi)

RNA interference, first discovered in *Caenorhabdtitis elegans* and published in 1998 (REF. 86), has been rapidly adopted throughout the scientific community and has spread to many experimental organisms.

C. elegans worms can be injected with, bathed in, or fed a solution of long (> 300 base pairs (bp)) double-stranded RNA (dsRNA), or they can be fed bacteria that contain a plasmid that expresses two RNA strands that hybridize *in vivo*. The false-negative rate can be high using this latter method[47]; to improve matters, a mutant strain of *C. elegans* can be used that is particularly sensitive to RNAi (rrf-3, which is defective in a putative RNA-directed RNA polymerase)[47].

In *Drosophila*, long dsRNA can be injected into embryos, transfected into cells that are grown in culture, or, for some phagocytic cell types, added as a solution onto cells that are grown in culture. In particular, treatment of *Drosophila* S2 CELLS with dsRNA seems to be reliable (all 26 knockdowns were successful in two recent studies[11,12]). The dsRNA is typically generated *in vitro* by transcription from PCR templates that are generated either from genomic DNA or cDNA. Whereas each *in vitro* transcription reaction produces a limited supply of dsRNA, the gene-specific oligonucleotides that are used to generate the templates are typically available in such quantity that they could be conveniently shared between laboratories.

In most mammalian cells, long dsRNAs induce the INTERFERON RESPONSE, so researchers use short interfering RNAs (siRNAs) instead. siRNAs are dsRNAs, with fewer than 30 nucleotides, that suppress gene expression without inducing the interferon response. The short length of siRNAs allows the targeting of a region that lacks homology to other genes, although nonspecific effects are still a concern[87]. An important disadvantage of siRNA versus long dsRNA is that the efficacy of siRNA varies on the basis of the precise portion of the gene that is chosen, so several siRNAs must be tested. Although the two strands of an siRNA can be chemically synthesized and hybridized, these can also be produced *in vivo*: viruses or plasmids introduced into cells by transfection or transduction can express two short RNA molecules that hybridize to form an siRNA, or express a single RNA that self-hybridizes to form a short hairpin RNA (shRNA). Compared with synthetic siRNA, virus or plasmid-based RNAi produces a longer-lasting effect, is cheaper and is more easily shared within the research community. The *C. elegans* SID-1 protein (systemic RNA interference-deficient) was recently proposed to be responsible for the ready uptake of dsRNA without the need for transfection in *C. elegans*, and therefore might be a useful tool in simplifying the delivery of functional RNAi reagents into *Drosophila* and mammalian cells[88].

Table 1 | **Publicly available genome-wide collections of mutated strains and gene-perturbing reagents**

Organism	Collection type*	Percentage genome coverage	URL‡	Refs
Saccharomyces cerevisiae	A non-redundant set of deletion strains. Two independent haploids have been made (one of each mating type), as well as heterozygous and homozygous diploids	96%	http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html	14
	A partially redundant set of 5,442 yeast strains harbouring epitope-tagged alleles generated by insertional mutagenesis of 1,917 different genes	31%	http://ygac.med.yale.edu/default.stm	27
	A set of >13,000 yeast strains harbouring epitope-tagged alleles of 3,565 different genes. Most strains were generated by insertional mutagenesis, but some were generated by directed cloning of PCR-amplified ORFs into a yeast tagging/expression vector	58%	http://ygac.med.yale.edu/default.stm	96
	A non-redundant set of yeast strains harbouring TAP epitope-tagged genes§, although one-third of the strains did not produce detectable protein by Western blot	98%	http://yeastgfp.ucsf.edu/ and www.openbiosystems.com	61
	A non-redundant set of yeast strains harbouring GFP-tagged genes, although one-third of the strains did not show detectable fluorescence	97%	http://yeastgfp.ucsf.edu/ and www.openbiosystems.com	97
Caenorhabditis elegans	A set of dsRNA-expressing plasmids that induce RNAi when grown in *E. coli* and fed to *C. elegans*	86%	www.hgmp.mrc.ac.uk/geneservice/reagents/products/descriptions/Celegans.shtml	44
	A set of ORFs from *C. elegans* that have been cloned into Gateway™ vectors to allow convenient recombination into desired expression vectors	34%	http://worfdb.dfci.harvard.edu/stats.html	100
	A set of more than 700 knockout strains	4%	www.celeganskoconsortium.omrf.org	—
Drosophila melanogaster	A set of dsDNA templates for *in vitro* transcription of dsRNA constructed by a group of researchers at the University of California, San Francisco	~50%	www.openbiosystems.com	—
	A collection of primer pairs used to PCR-amplify templates for *in vitro* transcription of dsRNA	91%	http://flyrnai.org/	M. Boutros and N. Perrimon, unpublished observations
	A set of plasmids from the *Drosophila* Gene Collection Release 1, which can be used as templates for the *in vitro* transcription of dsRNA	43%	www.fruitfly.org/DGC/	62
Arabidopsis thaliana	A partially redundant set of mutant lines generated by insertional mutagenesis	~74%	www.arabidopsis.org/abrc/	101
Mus musculus	A number of partial collections of transposon-mutated mouse embryonic stem cell lines and knockout mice that are available from several sources	—	—	See table in REF. 102
Homo sapiens	Several groups of non-profit institutions have announced plans to make RNAi collections for the entire human genome, including Netherlands Cancer Institute/Cancer Research UK, Vienna's Research Institute of Molecular Pathology/EMBL/Sanger Institute, Cold Spring Harbor Laboratories, and The RNAi Consortium‖	—	—	—
	Several partial-genome collections of overexpression plasmids are publicly available	—	Open BioSystems (www.openbiosystems.com); Invitrogen (http://clones.invitrogen.com/cloneinfo.php?clone=gs); NIH (some of their Mammalian Gene Collection is in expression vectors; http://mgc.nci.nih.gov)	—

*Details for each collection are based on the original publication; more clones might be available. ‡The web sites listed usually describe the construction of the collection as well as methods for obtaining a set for research use (fees and/or restrictions might apply). For other projects in progress in *Arabidopsis thaliana*, see www.arabidopsis.org/info/2010_projects/2003_Report.pdf; § TAP tags consist of several epitope tags, which can be used to purify protein complexes from cell lysates; ‖The RNAi Consortium (Harvard, MIT, the Broad Institute, the Dana–Farber Cancer Institute and the Whitehead Institute for Biomedical Research); dsRNA, double-stranded RNA; EMBL, European Molecular Biology Laboratory; GFP, green fluorescent protein; MIT, Massachusetts Institute of Technology; NIH, National Institutes of Health; ORF, open reading frame; RNAi, RNA interference; TAP, tandem affinity purification.

be screened systematically; however, these collections are typically not genome-wide, even in yeast. Another technology that is potentially applicable to all eukaryotes involves tagging genes with DEGRONS, which so far has been done for only ~100 yeast genes. As degrons produce a much faster knockout of protein function than RNAi or genomic knockouts, direct effects of protein loss can be observed, particularly in essential genes[22]. To complement technologies that eliminate or decrease gene function, partial-genome collections of cDNA expression

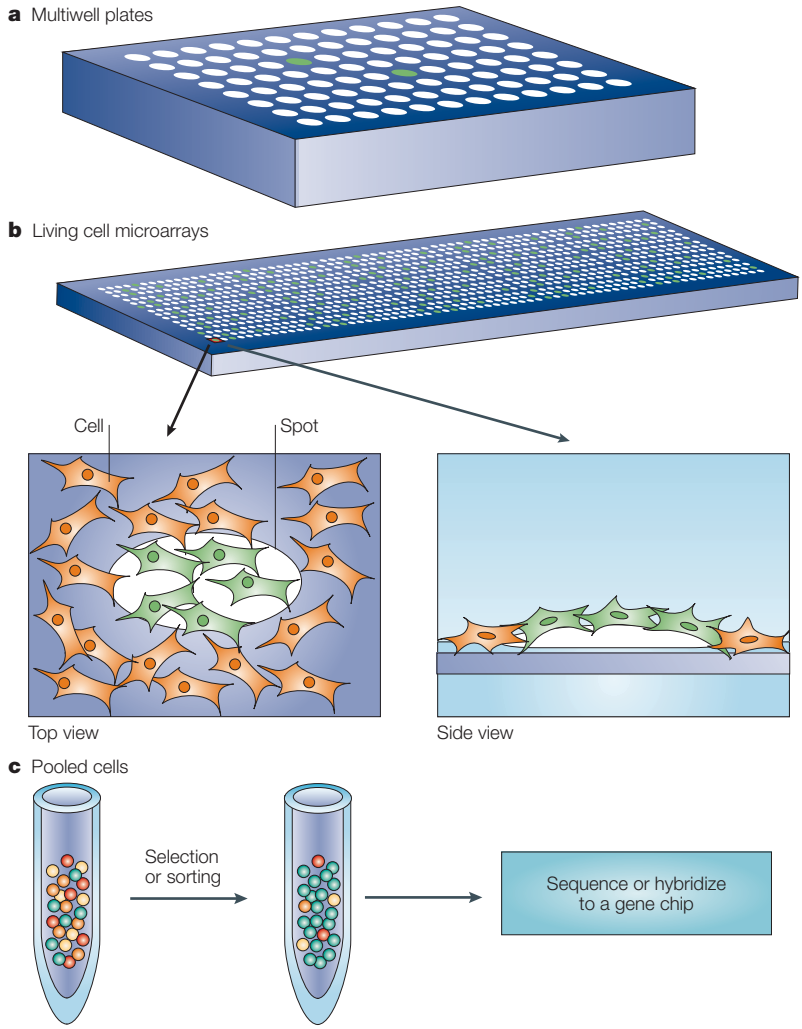


Figure 2 | **High-throughput formats for screening. a** | Multiwell plates. In this screening method, each well contains a different gene-perturbing reagent (for example, plasmid or RNAi reagent), and cells are added to the wells. Alternatively, each well can contain an organism with a different gene that is permanently perturbed (such as a yeast deletion strain). Green cells indicate those that have produced a positive response in a hypothetical assay. **b** | Living cell microarrays. To generate transfected cell microarrays, thousands of spots are printed onto a glass slide with a microarrayer (top diagram). Each spot contains a different gene-perturbing reagent (a plasmid or RNAi reagent, for example), and cells are plated on top (bottom diagram). Green cells indicate those that have produced a positive response in a hypothetical assay. **c** | Pooled cells. Cells can be screened in pools if each gene perturbation is marked, for example, with molecular bar codes. For example, a mixture of yeast deletion strains, initially present in equal proportions (indicated in the diagram by different coloured cells), can be subjected to some selective pressure. Cells that proliferate more slowly or die are underrepresented when the pooled genomic DNA is analysed using microarrays at the end of the experiment.

plasmids are becoming available to allow the effects of too much gene product to be determined (TABLE 1).

Small molecule and PEPTIDE APTAMER INHIBITORS[23] of protein function are useful research tools. High-throughput phenotypic screens have been conducted using libraries of these potential inhibitors, but systematic sets with known protein targets are not available. This is because it is extremely difficult to identify or develop a specific inhibitor for any protein, and it is impossible to identify inhibitors for proteins with unknown function. New strategies, such as engineering a kinase to be inhibited by an engineered small molecule while knocking out the endogenous kinase protein[24], are not yet suitable for high-throughput approaches. So, these types of inhibitor will probably not be used soon in systematic genome-wide screens, although they are useful in the pharmaceutical industry, in which chemical inhibitors are not just research tools but are also potential therapeutic agents.

DEGRON TAG
A degron (degradation) tag attached to a protein of interest specifically targets the protein for rapid proteolysis if the cells are grown at high temperature (37°) and in the presence of an overexpressed ubiquitin-associated protein that recognizes the tag.

Developing a phenotypic assay

As the genome-wide collection of reagents or strains is created or acquired, design of the phenotypic assay(s) to be used in the screen can begin. Phenotypic assays are certainly not new, but their application to thousands of gene perturbations in parallel has required substantial technical and economic limitations to be overcome. Initial genome-wide screens have examined a simple phenotype: cell proliferation versus cell death. Screens have become progressively more complex as sample preparation has become more elaborate (for example, by using enzymatic reagents or antibody staining), the cells have been engineered in some way (for example, by using mutants or having a reporter gene or fluorescent marker inserted), multicellular organisms are used (such as *C. elegans*), and more complex phenotypes are measured (such as intracellular localization, life span, general morphology and growth rate). Nearly all screens completed so far have relied on traditional, manual techniques; high-throughput technologies are just beginning to be used.

Measuring traditional cellular phenotypes. A surprising variety of biological processes has been studied using cell proliferation and/or death as a phenotype. In yeast, cell growth and/or proliferation can be scored by viewing colonies on agar plates for colony size or density. Alternatively, a pooling strategy (FIG. 2c) can be used: this method has successfully identified *S. cerevisiae* genes that are involved in nutrient and metabolism pathways[14,25–30], resistance and sensitivity to irradiation[31–34] and various chemicals[34–41], including antifungals[42,43]. In *C. elegans*, lethality has been scored in several partial-genome and genome-wide screens[44–47]. In addition, a systematic partial-genome screen in *C. elegans* identified genes that, when knocked down by RNAi, increase lifespan: death was measured at time-points based on lack of movement of worms in response to prodding[48]. In the human HeLa cell line, a collection of synthetic siRNAs targeting 510 human kinases was screened to identify genes involved in TRAIL-induced apoptosis, using viability as the readout[49].

While still using cell proliferation and/or death as the measurable phenotype, some relatively straightforward changes in the design of a screen can allow other interesting questions to be addressed. For example, *S. cerevisiae* genes that are required for NON-HOMOLOGOUS END-JOINING were recovered by selecting for cells that successfully repaired a linearized plasmid that contained a selectable marker[50]. Another particularly powerful twist on screens that involved cell proliferation has been termed 'synthetic genetic array', in which all genes are tested for SYNTHETIC LETHAL interactions with a mutation

PEPTIDE APTAMER INHIBITOR
Synthetic proteins that can bind and inhibit protein function.

TRAIL
A member of the tumour necrosis factor superfamily that preferentially induces apoptosis in tumour cells while leaving normal cells intact.

NON-HOMOLOGOUS END-JOINING
(NHEJ). One of two cellular DNA-repair pathways that are involved in the repair of double-strand breaks.

SYNTHETIC LETHAL
Synthetic interactions are identified if mutations in two separate genes produce a different phenotype from either gene alone, and indicate a functional association between the two genes. Two genes have a synthetic lethal relationship if mutants in either gene are viable but the double mutation is lethal.

CALCOFLUOR
A chemical that binds the chitin-rich bud scars that remain on the cell surface after cytokinesis.

SYNTHETIC DOSAGE LETHALITY
This type of genetic interaction is detected when overexpression of a gene is lethal only if another, normally nonlethal, mutation is present.

COULTER PARTICLE COUNTER
An instrument that measures the size of particles on the basis of changes in the electrical voltage as they pass through an orifice.

EPISTATIC TESTS
These can place genes in the same or different pathways and can establish the order of gene function if they are in a single, linear pathway. Gene A is epistatic to gene B if the phenotype that results from mutation of both genes matches the phenotype that results from gene A alone (and does not match that of gene B alone).

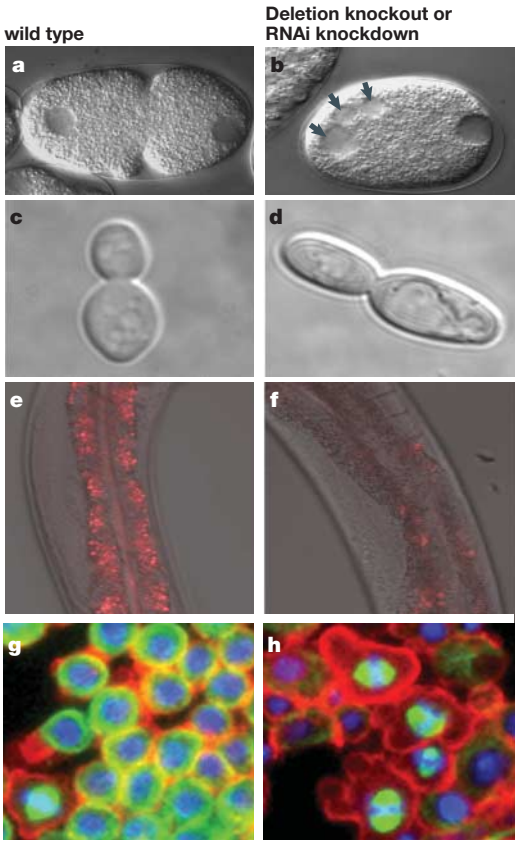


Figure 3 | **Examples of scorable phenotypes from various screens.** In addition to studies in which several classical phenotypes were scored by eye in *Caenorhabditis elegans*[44,45,47], visual screens have begun to use microscopy. Yeast genome-wide visual screens have identified genes that are important for sporulation (by microscopy of unstained cells)[93], cell-cycle progression (by microscopy of unstained cells)[94], bud site selection (by fluorescence microscopy of CALCOFLUOR-stained cells)[68], mitochondrial morphology (by fluorescence microscopy of antibody-stained cells)[29] and endocytosis (by fluorescence microscopy of lucifer-yellow-stained cells)[95]. In addition, the subcellular localization of about two-thirds of *Saccharomyces cerevisiae* genes has been determined by tagging genes with green fluorescent protein (GFP) or an epitope[96,97]. **a**, **b** | A partial-genome screen in *C. elegans*, which used time-lapse microscopy of whole worms, identified genes that are required for proper cell division[89]. RNAi against C16A3.9, a 40S ribosomal protein, resulted in several female pronuclei rather than one (arrowheads). **c**, **d** | A genome-wide screen of yeast deletion mutants identified genes that are required for normal shape and size, including phenotypes that are classified as 'elongated'[14]. **e**, **f** | A full-genome RNAi screen in *C. elegans*, which examined the pattern and intensity of Nile-red staining of fat storage droplets, identified genes that are involved in fat metabolism[98]. RNAi against choline/ethanolamine phosphotransferase reduced fat staining. **g**, **h** | A partial-genome RNAi screen in *Drosophila*-cultured cells identified genes that are involved in cytoskeletal organization, viability, attachment, cell-cycle progression and cytokinesis[99]. RNAi against fizzy, a protein that is involved in cyclin catabolism, produced cells with an increased frequency of mitotic spindles. Panels **a** and **b** reproduced with permission from REF. 89 © (2000) MacMillan Magazines Ltd; Panels **c** and **d** reproduced from REF. 14 © (2002) MacMillan Magazines Ltd; Panels **e** and **f** reproduced from REF. 98 © (2003) MacMillan Magazines Ltd; panels **g** and **h** reproduced from REF. 99 © (2003) BioMed Central.

in a gene of interest[51–53]. Although not yet used genome-wide, screens for other types of genetic interaction, such as SYNTHETIC DOSAGE LETHALITY[54], should prove useful as well. Any of these concepts for a screen can be applied using various combinations of new technologies for gene perturbation, including using RNAi-induced 'mutations'. These screens identify genes with redundant functions and pathways that buffer each other, even in *S. cerevisiae*[55]; their application to higher organisms with presumably more redundancy should prove even more useful.

Screens with a straightforward, but much more quantitative, output in *S. cerevisiae* identified genes that are required for the proper maintenance of cell size[56,57]. In these studies, the sizes of cells from each yeast deletion strain were determined by measuring each strain one by one using a COULTER PARTICLE COUNTER. The quantitative output of this screen allowed powerful synthetic/EPISTATIC TESTS to flesh out a network for several of the signalling proteins that are involved in maintaining proper cell size[57].

If it is feasible to perform sample preparation for each gene-perturbed cell population, the amounts of a protein or metabolite can be measured. Staining colonies of yeast deletion strains, with an antibody for example, has identified genes that are involved in glycogen storage[58], vacuolar protein storage[59], spore-wall maturation[60] and glucose-induced degradation of fructose-1,6-bisphosphatase (FBPase)[5]. Western blots of epitope-tagged proteins can show expression levels, as was done for approximately two-thirds of *S. cerevisiae* proteins[61]. By measuring ATP levels, a recent genome-wide RNAi screen in *Drosophila* identified genes that are involved in cell proliferation, metabolism, growth and/or viability (M. Boutros and N. Perrimon, unpublished observations).

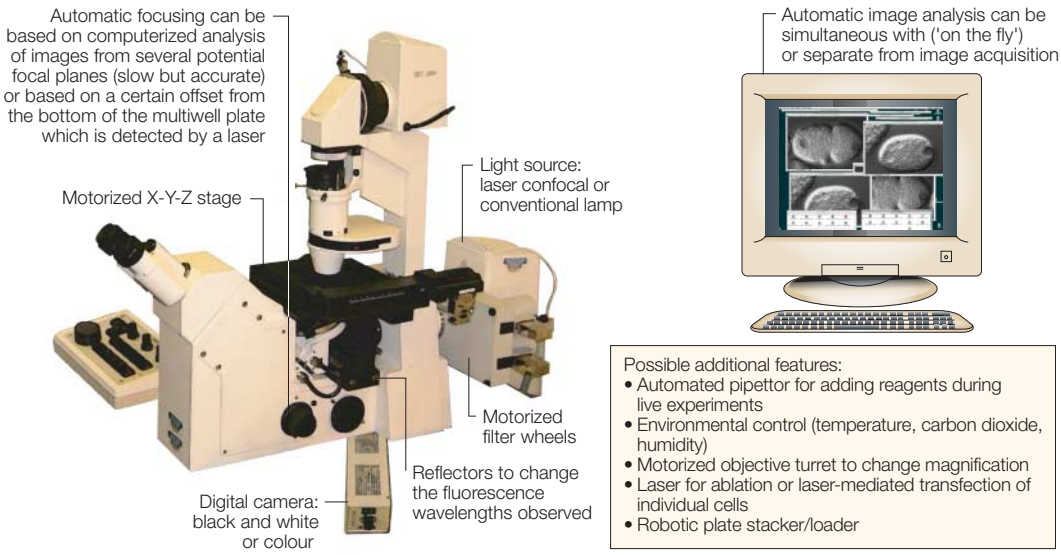
Using reporter genes. The screens described so far have mainly focused on the responses of unengineered cells to the disruption of genes. The use of reporter genes can conveniently show phenotypes that would otherwise be laborious or challenging to detect, and multi-well plate readers are available to detect fluorescent or luminescent reporter products (FIG. 2a). In the first published systematic partial-genome RNAi screen in *Drosophila*, a reporter gene was transiently transfected into cultured cells to identify genes that are involved in modulating the response to the Hedgehog signalling pathway[62]. In overexpression screens, plasmids that encode 11,000–20,000 human genes have been tested for their effects on reporter genes in mammalian cell lines[63–65]. Variations of reporter assays can also be used to examine biological functions other than transcription. In *C. elegans*, a specially designed LacZ cassette was used as a reporter in a genome-wide RNAi screen to identify genes that protect against mutations in the genome[66]. Insertion of a transposon into a muscle gene in *C. elegans* causes a twitching phenotype that was used as a reporter in a genome-wide RNAi screen to identify genes involved in the removal of the transposon from the gene[67].

MITOTIC INDEX
The percentage of cells in the mitotic phase of the cell cycle.

Visual phenotypes. Many biological questions can only be answered visually — that is, by scoring a visible phenotype at the level of the entire organism or in individual cells, and many yeast screens that use microscopy have already been completed (FIG. 3). Because cells from higher organisms are larger and more specialized, there are tremendous opportunities for interesting visual screens that have only begun to be tapped (FIG. 3). Until recently, visual screens required highly trained scientists to spend many hours examining samples with a microscope and yielded only qualitative results. Now, however, images can be collected by motorized microscopes and then rapidly viewed later by a researcher, or quantitatively analysed using computerized software (BOX 2). Dozens of systems for automated imaging have become commercially available over the past several years, making it feasible to carry out genome-wide screens with complex and subtle visual outputs (TABLE 2).

Screening for multiple phenotypes. There are methods other than automated imaging that can be used to multiplex samples — that is, to extract more than one data point per sample in a given experiment. For example, the activity of two different reporter genes can be measured simultaneously. Two systematic genome-wide phenotypic screens can also be run in parallel and their results compared, and can provide new insights. For example, screening most of the genes in *S. cerevisiae* showed a correlation between cell morphology and bud site selection[68]. In another study, 8,000 supernatants from cells that were transfected with human cDNA expression plasmids, that encode predicted secreted proteins, were tested in many cell-based assays relating to diabetes and T-cell function[69,70]. The combination of data from such a variety of assays provides a more complete understanding of the biological effects of each protein.

Box 2 | **Advances in automated imaging**



Automated imaging, or high-throughput microscopy, allows the rapid acquisition of a large number of images from multiple samples. Unlike other types of assay, image data provides information on multiple phenotypes simultaneously, because several parameters such as shape, size, texture and staining intensity can be measured, using multiple fluorophores to label different targets within each cell. Images can also be reanalysed years later from a different perspective. For example, images from a screen to look for changes in the size of nuclei can later be reanalysed for differences in DNA texture should that parameter become interesting. In addition to these advantages, a unique property of image-based assays is that the phenotype of interest can often be measured on a per-cell basis, so that the phenotype of a particular subset of the cells in the dish can be analysed, using the remaining cells as controls. For example, the length of neurites can be measured from among the mixture of cells that are present in a primary brain cell culture if the neurons are labelled with a specific antibody.

Automated microscopes, which are becoming staples of academic laboratories, have motorized control of stages, filter wheels and cameras to allow the automated acquisition of images (see figure). The real power of automated imaging comes when image processing is also automated, allowing phenotypes to be scored by a computer. Automated image analysis is now being adapted to several assays. It is relatively straightforward to measure the brightness of fluorophores on a per-cell basis: this allows comparison, for example, of green fluorescent protein (GFP) expression levels or amounts of protein (stained with a fluorescent antibody). Other common automated imaging assays show the extent of co-localization between two proteins, MITOTIC INDEX, apoptosis, cell spreading, cell migration, the growth of neurites from neurons and the changes in intracellular localization that many signalling proteins experience on activation. Many studies that use automated imaging have recently begun and the next few years should see a flurry of publications using this technology, particularly as genome-wide collections become available for mammalian cells.

Future high-throughput technologies. Several powerful technologies have yet to be applied to systematic genome-wide screens of gene function. For example, FLUORESCENCE PLATE READERS have yet to be used (for example, in fluorescent staining or FLUORESCENCE RESONANCE ENERGY TRANSFER (FRET)), although they have already been widely used in the pharmaceutical industry for high-throughput screens of chemical compounds. Poor signal-to-noise could be an important problem when using fluorescence plate readers with cells that are grown in monolayers on the bottom of the well. The ability to use FLUORESCENCE ACTIVATED CELL SORTERS (FACS) to rapidly screen entire genomes of mammalian or *Drosophila* cells awaits the availability of reagents to generate a pooled collection of gene-perturbed cells — for example, with LENTIVIRUSES. Decreasing costs of robotic equipment could also give rise to new advances. Although pharmaceutical companies that screen tens to hundreds of thousands of compounds per day use extensive automation, robotics are not routinely used in many of the academic laboratories that conduct systematic genome-wide screens. CELL MICROARRAYS can ease throughput issues, as conventional microarrayers can deposit up to 10,000 different gene-perturbing reagents in spots on a single slide (FIG. 2b). If cells are plated onto the slide, those that attach on one of the spots take up the corresponding reagent[71]. Initially demonstrated using a mixture of transfection reagent and overexpression plasmids (in a process known as reverse transfection), the format is also compatible with spots of synthetic siRNA[72,73]. Similarly, the format should be adaptable to other strategies: spots of RNAi-inducing double-stranded RNA (dsRNA) could be used for *Drosophila* cells and siRNA expressed from a purified virus or from a plasmid mixed with transfection reagent could be used for mammalian cells.

Screening

Practical, economic and technical issues become relevant when dealing with the large number of samples that are involved in systematic genome-wide phenotypic screens. These are discussed in BOX 3.

FLUORESCENCE PLATE READERS
Instruments that read the fluorescence in each well of a multiwell plate.

FLUORESCENCE RESONANCE ENERGY TRANSFER
(FRET). A phenomenon by which the energy from an excited fluorophore is transferred to an acceptor molecule at short (<100 Å) distances, leading to decreased fluorescence of the donor and increased fluorescence of the acceptor. The efficiency of energy transfer depends strongly on the distance between the donor and acceptor molecules.

FLUORESCENCE ACTIVATED CELL SORTERS
(FACS). The separation of cells or chromosomes by their fluorescence and light-scattering properties, which are measured as the particles flow in a liquid stream as they pass through laser beams. The stream is then broken into droplets, and selected droplets are electrically charged and deflected into collection vessels as they pass through an electric field.

LENTIVIRUS
A type of retrovirus that can transduce overexpression and RNAi-inducing constructs to dividing and non-dividing mammalian cells.

CELL MICROARRAYS
An array of gene-perturbing reagents (such as plasmids plus a transfection reagent) that is spotted onto a glass slide. Cells that are are plated onto the slide and that land on a spot are affected by the reagent.

Table 2 | **Automated imaging companies and products**

Company	Instrument	Web site
Acumen Bioscience/TTP LabTech	Acumen Explorer™	www.acumenbioscience.com
Amersham Pharmacia	IN Cell Analyzer 1000 & 3000	www4.amershambiosciences.com
Amnis	ImageStream™	www.amnis.com
Aperio	ScanScope®	www.aperio.com
Applied BioSystems	FMAT®/8200	www.appliedbiosystems.com
Atto Bioscience	Pathway HT	www.atto.com
Automated Cell	ACI	www.automatedcell.com
Axon Instruments	ImageXpress® 5000A	www.axon.com
BioGenex	GenoMx VISION	www.biogenex.com
CellaVision	DiffMaster™ & MICRO21®	www.cellavision.com
Cellomics	ArrayScan®/KineticScan®	www.cellomics.com
ChromaVision	ACIS®	www.chromavision.com
Compucyte	Laser Scanning Cytometer & iCyte™	www.compucyte.com
Cyntellect	LEAP	www.cyntellect.com
Cytokinetics	Cytometrix™	www.cytokinetics.com
Evotec Technologies	Opera	www.evotec-technologies.com
Imstar	Pathfinder™	www.imstar.fr
Intelligent Imaging Innovations	Everest™/Marianas™	www.intelligent-imaging.com
Metasystems GmbH	Metafer/Metacyte	www.metasystems.de
Olympus BioSystems	LUI	Product information not yet available
Q3DM	EIDAQ 100	www.q3dm.com
QED Imaging	Custom cell scanner	www.qedimaging.com
Universal Imaging Corporation™	Discovery-1	www.image1.com

Apart from the companies listed, several software packages can be used to automate nearly any off-the-shelf motorized microscope. In addition, image analysis can be done completely separately from image collection using programs such as NIH Image, ImagePro, MetaMorph, MATLAB, Mathematica, Mathcad, IDL (Interactive Data Language) and LabView. Some of these packages have add-in modules to allow sophisticated image analysis, but all require expertise in image analysis and/or programming. Our laboratory is currently developing publicly available, flexible software to allow cell biologists to analyse thousands of cell images automatically for any phenotype (A.E.C. and D.M.S., unpublished observations). There is a continuum between microscopes with some motorized components and simple software for automation, and a 'turn-key' integrated system with full automation and pre-programmed image-analysis algorithms. We have focused on the latter, but this list is probably incomplete. Note that some of the instruments do not have fluorescence capabilities and are intended for clinical applications (such as for analysing tissue or blood samples). Some instruments are not commercially available (they are either still in development or have been developed for the company's own drug development) but might be available as a service or collaboration.

Determining hits

So far, selecting hits in a screen has been fairly subjective; for example, by choosing a certain percentage of outliers as hits[56,57], or choosing a multiple of the standard deviation (M. Boutros and N. Perrimon, unpublished observations). Another method for confirming assay quality and setting a threshold above which samples are considered as hits is to compare the results of the screen with the published literature and comment on what percent of expected, known genes were uncovered in the screen. An important factor to consider when choosing the threshold is whether false-positives or false-negatives will be more detrimental during subsequent analysis. Because genome-wide screens are comparable to doing thousands of individual phenotypic experiments, the contributing factors to the false-negative and false-positive rates, as well as methods for normalizing data, depend on the particular assay and technology used. When developing and optimizing a phenotypic assay with a quantitative output, the Z-FACTOR provides a simple measurement of high-throughput assay 'quality'[74]. The Z-factor has not yet seen much use in genome-wide screens from academic laboratories, largely because few screens so far have a quantitative output.

Interpreting and following up on results

Confirming hits experimentally. For screens with a manageable number of hits, it is not unreasonable to repeat the original assay to confirm each hit, and this is sometimes done. It is less common to confirm each hit in an alternative way, although techniques are often available to do so. For example, a phenotype that is observed in a yeast deletion strain should usually be reversed by transforming the cells with a plasmid that expresses that gene. Another strategy used in yeast is to confirm that the observed phenotype co-segregates with the resistance gene that replaced the deleted gene, to ensure that the phenotype is not the result of an unknown secondary mutation that has spontaneously arisen in the strain. Such studies showed that phenotypes of 6.5% of the haploid strains that were generated by the *Saccharomyces* Genome Deletion Project (TABLE 1) did not co-segregate with the selectable marker appropriately[2]. When homozygous diploid strains are screened, this is not an important concern because two independently generated haploid strains were mated to generate the diploids, thereby decreasing the possibility of a spontaneous mutation in both alleles of an unknown gene.

Different methods can be used to confirm the results of RNAi screens. Although very important results could be confirmed by engineering a true knockout using traditional methods, this can be laborious. Guidelines for proper controls for RNAi experiments were recently suggested, most of which are currently feasible for confirming only a small number of hits[75]. For certain types of phenotype produced by RNAi (such as a decrease in cell size), it can be informative to transfect cells with a plasmid that overexpresses the candidate gene to determine whether the opposite phenotype is observed.

Z-FACTOR
A measurement that takes into account the dynamic range of the assay (how far apart the positive controls are from the negative controls), as well as data variability (how much variation is seen in the measurements of positive and negative controls).

Box 3 | **Practical challenges in high-throughput screening**

Pooling

It might be advisable to screen the genome in pools, depending on the time and money involved in sample preparation and analysis. Pooling more than one double-stranded RNA (dsRNA) in *Caenorhabditis elegans* reduces efficacy[89,90], but results with RNA interference (RNAi) in *Drosophila* have been more favourable[11,62]. Even so, hits are more easily identified when samples are not pooled, and for quantitative assays, obtaining a result for every individual gene is important for broader conclusions to be drawn. Pooling can be scientifically useful when several genes are thought to be redundant, because several genes can be knocked down simultaneously.

Repeating screens

It was initially considered sufficiently challenging to accomplish one pass through a collection of genes but, with time, it will probably be necessary to repeat screens before the data are considered rigorous and complete. Although the false-positive rate was negligible (<1%), a recent study in *C. elegans* showed false-negative rates of 10–30% between two rounds of screening for several classical phenotypes, even when conducted in the same laboratory[47]. It might also be useful to repeat screens in more than one cell type, as was done recently in *Drosophila* (M. Boutros and N. Perrimon, unpublished observations); any one cell type from a multicellular organism is likely to have a considerable portion of its genome silenced, which means that knockouts or knockdowns against these genes will have no effect.

Data handling

Storing, analysing and retrieving data that is generated from high-throughput screens of any kind can be a challenge, even with screens that generate a single quantitative output. For images, an academic-industry collaboration is working towards a universally accessible format and database, known as the Open Microscopy Environment (see online links box). This software, which will include integrated image-processing routines, allows storage, retrieval and analysis of image data[91].

Data sharing

Data from genome-wide screens, similar to those generated by genome sequencing projects and gene-expression microarrays, are potentially even more helpful after their initial publication, when they are analysed from a different perspective or in combination with new data. It is therefore imperative that these data be publicly accessible and in a standardized format. The main web site of each model organism (see online links box) typically provides a centralized repository for data from genome-wide screens[79,92], although data are sometimes available only from a laboratory's own database[80].

Another possibility is to use a specific chemical inhibitor to confirm a phenotype in those rare cases in which one is already available for the protein that is discovered in the screen. Of course, a genome-wide screen will typically include thousands of samples that produce negative results, thereby eliminating the need for negative controls that are designed to ensure that the RNAi protocol does not itself produce the phenotype of interest.

Experimental follow-up. In addition to simply confirming hits, it is fairly common to perform a series of traditional follow-up experiments on hits from screens, such as performing the original assay in a different cell type or strain background, or conducting a secondary screen to separate hits into mechanistic classes. For example, RNAi reagents against *C. elegans* genes, that caused embryonic lethality in initial screens[44,45], were screened by video microscopy in a secondary assay for defects in pronuclear migration[76]. It is often informative to combine two knockouts/knockdowns in the same cell to look for genetic interactions (such as epistasis, enhancement and suppression), either by treating cells with two RNAi reagents simultaneously or crossing two yeast deletion strains and analysing double mutants. This strategy can place proteins upstream and downstream of each other into pathways to deduce a network, as was accomplished for several yeast genes that are involved in cell growth and division[57]. Furthermore, dosage effects can be observed in yeast by comparing the phenotype of heterozygous diploids with haploids, and in RNAi-amenable organisms by varying the amount of RNAi reagent or assaying the phenotype at several time-points after RNAi treatment as protein levels drop.

Once a genetic perturbation is made, even by traditional mutagenesis, it has always been difficult to determine which changes in the organism are the direct result of the change and which are indirect or even compensatory for the perturbation[77,78]. From this point of view, certain technologies are better than others: RNAi offers better temporal resolution than permanent knockout strains. For example, cells can be examined at several time-points after the addition of the RNAi reagent to determine which phenotypes occur first. By contrast, knockout strains are permanently missing the protein of interest and have achieved some degree of homeostasis, possibly including compensatory changes.

Drawing broader conclusions. It is not atypical, particularly for initial genome-wide screens in an organism, for the primary screen to be the only experimental data in the study. Apart from the value of the list of genes that are implicated in the process of interest, several systems-wide conclusions can be drawn by looking at the data as a whole rather than focusing on individual genes. This perspective often shows valuable information about cellular physiology and about the relationships between classes of gene. The genes can be classified by several parameters, all of which can be informative: these include enzymatic function, signalling pathway, the presence of domains or motifs and genomic location. For example, a genome-wide screen showed that genes that produce a nonviable phenotype were found less often on the X chromosome or at the ends of chromosomes, indicating selective pressure against essential genes in those locations[44,46].

Functional categorization of genes that are found in screens can show important regulators of (or contributors to) cellular processes. For example, mitochondrial genes were found to be crucial for longevity in *C. elegans*[48]. This insight relied on combining the list of hits in the genome-wide screen with the WormBase online database (see online links box), to which researchers over the years have added notes about mitochondrial function to some of the genes. As another example, the discovery of a phenotype associated with RNAi against a predicted gene confirms that the gene is truly expressed, which adds to the annotation of the genome sequence (M. Boutros and N. Perrimon, unpublished observations). These examples show the cumulative effects of scientific research in action: as more is learned about each gene, more becomes known about other genes and the complex biological processes in which they are involved.

Genome-wide approaches should accelerate markedly the pace of this cumulative effect. Entire genome-wide data sets can be correlated with other published genome-wide data sets from various sources, including gene-expression microarrays, protein–protein interaction studies and *in silico* predictions of gene function. Meshing new genome-wide data with the increasingly large existing data sets can be daunting, but it is certainly possible[1,79,80]. For example, a study that measured expression of tandem affinity purification (TAP)-tagged proteins in yeast by Western blot was correlated with previous mRNA expression data to show a relatively constant ratio of protein to mRNA (~4,800) over a wide range of mRNA abundances[61].

Depending on the technique used to perturb each gene, negative results can be meaningful as well, and should be incorporated into databases. This is particularly the case for knockout strains, such as the yeast collection, in which a negative result indicates that the gene is not required for the process. For RNAi techniques, caution must be used because it is possible that for any particular gene, the RNAi reagent was not functional for some reason. In cases in which the false-negative rate is too high to conclude anything about one particular gene, there might still be sufficient evidence to propose that a group of related genes is not required for a process if every member of the group produces a negative result.

Note that experiments that are conducted in the format of systematic genome-wide functional screens are not any less hypothesis-driven than the corresponding traditional experiments. In other words, if it is scientifically interesting to determine whether any one particular gene is required for a process, it is worthwhile answering the same question for all genes in parallel. In contrast to traditional experimentation, in which one gene is studied by many experimental methods, these screens use one experimental method to study all genes. If the results of these experiments are appropriately

added to gene annotation databases, conclusions can be drawn from the behaviour of many genes in a single genome-wide screen ('horizontal' data mining), but also for one particular gene across a set of screens ('vertical' data mining) — which is conceptually similar to traditional experimental design[80].

Conclusions

Systematic functional screens are fundamentally changing how biologists identify the molecular components that drive biological processes. These screens are becoming standard practice in *S. cerevisiae* and *C. elegans*, in which it is possible to undertake screens in the whole organism. In the near future, gene-perturbing reagents, such as retrovirally-expressed RNAi constructs and full-length cDNAs, will probably become available for all mouse and human genes. These tools will undoubtedly accelerate our understanding of the cell biological processes that can be observed in mammalian cells in culture. For example, certain aspects of vasculogenesis[81] can be reproduced *in vitro*, making this process amenable to genetic dissection with systematic screens.

Even more appealing is that genome-wide collections of reagents for mice raise the possibility of conducting screens in a whole mammalian organism. Although technological advances must be overcome to achieve this, the ability to study mammalian physiology *in vivo* while perturbing every gene in the genome promises a remarkable payoff. Transgenic RNAi constructs have already been shown to induce gene-specific knockdowns in a whole mouse[82–84]. As it is simpler to create loss-of-function mouse mutants with RNAi than with traditional homologous recombination, it might be feasible to generate a collection of mouse strains that express RNAi constructs for all genes. This would be a particularly powerful resource if gene knockdown could be controlled in a temporal- and tissue-specific fashion — this is a possibility as polymerase II promoters can control RNAi-triggering constructs[82,85].

To reduce the number of mice needed to screen the genome, exogenously delivered RNAi reagents, perhaps expressed in lentiviruses, could be screened by infecting mice with pools of lentiviruses *in vivo*. This approach would be particularly effective in organs such as the immune system, in which cells of interest can be removed from the organism, treated with RNAi *in vitro* to maximize gene suppression and returned to the organism, whose remaining endogenous cells have been ablated. Other organs that are easily manipulated might also be amenable to this approach (for example, liver regrowth after partial hepatectomy). As in all pooling strategies, there would be serious challenges in understanding which RNAi reagent caused which effect, a problem that could be overcome if mammalian RNAi reagents are barcoded in a way that is similar to the yeast deletion strains.

A screen of the entire genome for genes that are involved in mammalian physiological processes without pooling and without generating thousands of mouse strains might be possible through the delivery of a spatially defined two-dimensional array of RNAi-inducing viruses to the surface of an organ, such as the brain or skin. With this approach, the spatial location of the observed physiological effect, such as changes in neurite branching, would indicate which gene was involved. The demonstrated power of systematic screens in non-mammalian model organisms makes it highly likely that biologists will develop the approaches to harness this same power to study mammalian physiology.

1. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
2. Grunenfelder, B. & Winzeler, E. A. Treasures and traps in genome-wide data sets: case examples from yeast. *Nature Rev. Genet.* **3**, 653–661 (2002).
3. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
4. Hammerle, M. *et al.* Proteins of newly isolated mutants and the amino-terminal proline are essential for ubiquitin-proteasome-catalyzed catabolite degradation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 25000–25005 (1998).
5. Regelmann, J. *et al.* Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. *Mol. Biol. Cell* **14**, 1652–1663 (2003).
6. Snyder, M. & Gerstein, M. Genomics. Defining genes in the genomics era. *Science* **300**, 258–260 (2003).
7. Misra, S. *et al.* Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* **3**, RESEARCH0083 (2002).
8. Stein, L. Genome annotation: from sequence to biology. *Nature Rev. Genet.* **2**, 493–503 (2001).
9. Brummelkamp, T. R., Nijman, S. M., Dirac, A. M. & Bernards, R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB. *Nature* **424**, 797–801 (2003).
10. Heo, W. D. & Meyer, T. Switch-of-function mutants based on morphology classification of Ras superfamily small GTPases. *Cell* **113**, 315–328 (2003).
11. Goshima, G. & Vale, R. D. The roles of microtubule-based motor proteins in mitosis: comprehensive RNAi analysis in the *Drosophila* S2 cell line. *J. Cell Biol.* **162**, 1003–1016 (2003).
12. Rogers, S. L., Wiedemann, U., Stuurman, N. & Vale, R. D. Molecular requirements for actin-based lamella formation in *Drosophila* S2 cells. *J. Cell Biol.* **162**, 1079–1088 (2003).
13. Adams, M. D. & Sekelsky, J. J. From sequence to phenotype: reverse genetics in *Drosophila melanogaster*. *Nature Rev. Genet.* **3**, 189–198 (2002).
14. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
Describes the construction of ~6,000 yeast deletion strains, and a screen for proliferation under various growth conditions.
15. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
16. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. *Nature Rev. Mol. Cell Biol.* **4**, 457–467 (2003).
17. Shi, Y. Mammalian RNAi for the masses. *Trends Genet.* **19**, 9–12 (2003).
18. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genet.* **3**, 737–747 (2002).
19. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30**, 313–321 (2003).
20. Hutvagner, G. & Zamore, P. D. RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**, 225–232 (2002).
21. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. & Driscoll, M. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nature Genet.* **24**, 180–183 (2000).
22. Kanemaki, M., Sanchez-Diaz, A., Gambus, A. & Labib, K. Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*. *Nature* **423**, 720–724 (2003).
23. Geyer, C. R., Colman-Lerner, A. & Brent, R. 'Mutagenesis' by peptide aptamers identifies genetic network members and pathway connections. *Proc. Natl Acad. Sci. USA* **96**, 8567–8572 (1999).
24. Bishop, A. C., Buzko, O. & Shokat, K. M. Magic bullets for protein kinases. *Trends Cell Biol.* **11**, 167–172 (2001).
25. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
26. Steinmetz, L. M. *et al.* Systematic screen for human disease genes in yeast. *Nature Genet.* **31**, 400–404 (2002).
27. Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
28. Deutschbauer, A. M., Williams, R. M., Chu, A. M. & Davis, R. W. Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 15530–15535 (2002).
29. Dimmer, K. S. *et al.* Genetic basis of mitochondrial function and morphology in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **13**, 847–853 (2002).
30. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
31. Birrell, G. W., Giaever, G., Chu, A. M., Davis, R. W. & Brown, J. M. A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl Acad. Sci. USA* **98**, 12608–12613 (2001).
32. Bennett, C. B. *et al.* Genes required for ionizing radiation resistance in yeast. *Nature Genet.* **29**, 426–434 (2001).
33. Game, J. C. *et al.* Use of a genome-wide approach to identify new genes that control resistance of *Saccharomyces cerevisiae* to ionizing radiation. *Radiat. Res.* **160**, 14–24 (2003).
34. Hanway, D. *et al.* Previously uncharacterized genes in the UV- and MMS-induced DNA damage response in yeast. *Proc. Natl Acad. Sci. USA* **99**, 10605–10610 (2002).
35. Bianchi, M. M. *et al.* Large-scale phenotypic analysis reveals identical contributions to cell functions of known and unknown yeast genes. *Yeast* **18**, 1397–1412 (2001).

36. Chan, T. F., Carvalho, J., Riles, L. & Zheng, X. F. A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR). *Proc. Natl Acad. Sci. USA* **97**, 13227–13232 (2000).
37. Butcher, R. A. & Schreiber, S. L. A small molecule suppressor of FK506 that targets the mitochondria and modulates ionic balance in *Saccharomyces cerevisiae*. *Chem. Biol.* **10**, 521–531 (2003).
38. Page, N. *et al.* A *Saccharomyces cerevisiae* genome-wide mutant screen for altered sensitivity to K1 killer toxin. *Genetics* **163**, 875–894 (2003).
39. Zewail, A. *et al.* Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin. *Proc. Natl Acad. Sci. USA* **100**, 3345–3350 (2003).
40. Chang, M., Bellaoui, M., Boone, C. & Brown, G. W. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proc. Natl Acad. Sci. USA* **99**, 16934–16939 (2002).
41. Desmoucelles, C., Pinson, B., Saint-Marc, C. & Daignan-Fornier, B. Screening the yeast 'disruptome' for mutants affecting resistance to the immunosuppressive drug, mycophenolic acid. *J. Biol. Chem.* **277**, 27036–27044 (2002).
42. Gupta, S. S. *et al.* Antifungal activity of amiodarone is mediated by disruption of calcium homeostasis. *J. Biol. Chem.* **278**, 28831–28839 (2003).
43. Anderson, J. B. *et al.* Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. *Genetics* **163**, 1287–1298 (2003).
44. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
Describes the construction of a genome-wide collection of feedable, bacterially expressed dsRNA for *C. elegans*, and screening for classical phenotypes. This is the first genome-wide RNAi screen to be carried out in a multicellular organism.
45. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
46. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171–176 (2001).
47. Simmer, F. *et al.* Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveall novel gene functions. *PLOS Biology* **1**, E12 (2003).
48. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
49. Aza-Blanc, P. *et al.* Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12**, 627–637 (2003).
50. Ooi, S. L., Shoemaker, D. D. & Boeke, J. D. A DNA microarray-based genetic screen for nonhomologous end-joining mutants in *Saccharomyces cerevisiae*. *Science* **294**, 2552–2556 (2001).
51. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
Describes the first use of the synthetic genetic array strategy, in which a mutant of interest is crossed with all yeast deletion strains, allowing a genetic network to be constructed.
52. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.* **23**, 4207–4218 (2003).
53. Huang, D., Moffat, J. & Andrews, B. Dissection of a complex phenotype by functional genomics reveals roles for the yeast cyclin-dependent protein kinase Pho85 in stress adaptation and cell integrity. *Mol. Cell Biol.* **22**, 5076–5088 (2002).
54. Kroll, E. S., Hyland, K. M., Hieter, P. & Li, J. J. Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics* **143**, 95–102 (1996).
55. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
56. Zhang, J. *et al.* Genomic scale mutant hunt identifies cell size homeostasis genes in *S. cerevisiae*. *Curr. Biol.* **12**, 1992–2001 (2002).
57. Jorgensen, P., Nishikawa, J. L., Breitkreutz, B. J. & Tyers, M. Systematic identification of pathways that couple cell growth and division in yeast. *Science* **297**, 395–400 (2002).
58. Wilson, W. A., Wang, Z. & Roach, P. J. Systematic identification of the genes affecting glycogen storage in the yeast *Saccharomyces cerevisiae*: implication of the vacuole as a determinant of glycogen level. *Mol. Cell Proteomics* **1**, 232–242 (2002).
59. Bonangelino, C. J., Chavez, E. M. & Bonifacino, J. S. Genomic screen for vacuolar protein sorting genes in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **13**, 2486–2501 (2002).
60. Felder, T. *et al.* Dtrlp, a multidrug resistance transporter of the major facilitator superfamily, plays an essential role in spore wall maturation in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **1**, 799–810 (2002).
61. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
Reports the construction of ~6,000 yeast strains that contain TAP epitope-tagged genes in their natural genomic location. Protein expression was detected by Western blot in ~4,000 strains.
62. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
Reports the screening by dsRNA against ~43% of the *Drosophila* genome using a reporter assay.
63. Chanda, S. K. *et al.* Genome-scale functional profiling of the mammalian AP-1 signaling pathway. *Proc. Natl Acad. Sci. USA* **100**, 12153–12158 (2003).
64. Iourgenko, V. *et al.* Identification of a family of cAMP response element-binding protein co-activators by genome-scale functional analysis in mammalian cells. *Proc. Natl Acad. Sci. USA* **100**, 12147–12152 (2003).
65. Conkright, M. D. *et al.* TORCs: transducers of regulated CREB activity. *Mol. Cell* **12**, 413–423 (2003).
66. Pothof, J. *et al.* Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi. *Genes Dev.* **17**, 443–448 (2003).
67. Vastenhouw, N. L. *et al.* A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Curr. Biol.* **13**, 1311–1316 (2003).
68. Ni, L. & Snyder, M. A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **12**, 2147–2170 (2001).
69. Chen, C. *et al.* An integrated functional genomics screening program reveals a role for BMP-9 in glucose homeostasis. *Nature Biotechnol.* **21**, 294–301 (2003).
70. Fiscella, M. *et al.* TIP, a T-cell factor identified using high-throughput screening increases survival in a graft-versus-host disease model. *Nature Biotechnol.* **21**, 302–307 (2003).
71. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110 (2001).
72. Mousses, S. *et al.* RNAi microarray analysis in cultured mammalian cells. *Genome Res.* **13**, 2341–2347 (2003).
73. Kumar, R., Conklin, D. S. & Mittal, V. High-throughput selection of effective RNAi probes for gene silencing. *Genome Res.* **13**, 2333–2340 (2003).
74. Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen* **4**, 67–73 (1999).
75. Editorial. Whither RNAi? *Nature Cell Biol.* **5**, 489–490 (2003).
76. Le Bot, N., Tsai, M. C., Andrews, R. K. & Ahringer, J. TAC-1, a regulator of microtubule length in the *C. elegans* embryo. *Curr. Biol.* **13**, 1499–1505 (2003).
77. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295**, 1664–1669 (2002).
78. Greenspan, R. J. The flexible genome. *Nature Rev. Genet.* **2**, 383–387 (2001).
79. Bader, G. D. *et al.* Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends Cell Biol.* **13**, 344–356 (2003).
80. Vidal, M. A biological atlas of functional maps. *Cell* **104**, 333–339 (2001).
81. Michiels, F. *et al.* Arrayed adenoviral expression libraries for functional screening. *Nature Biotechnol.* **20**, 1154–1157 (2002).
82. Shinagawa, T. & Ishii, S. Generation of Ski-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter. *Genes Dev.* **17**, 1340–1345 (2003).
83. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33**, 401–406 (2003).
84. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33**, 396–400 (2003).
85. Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
86. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
87. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
88. Feinberg, E. H. & Hunter, C. P. Transport of dsRNA into cells by the transmembrane protein SID-1. *Science* **301**, 1545–1547 (2003).
89. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
90. Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**, RESEARCH0002 (2001).
91. Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K. Informatics and quantitative analysis in biological imaging. *Science* **300**, 100–102 (2003).
92. Issel-Tarver, L. *et al.* Saccharomyces genome database. *Methods Enzymol.* **350**, 329–346 (2002).
93. Enyenihi, A. H. & Saunders, W. S. Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*. *Genetics* **163**, 47–54 (2003).
94. Zettel, M. F. *et al.* The budding index of *Saccharomyces cerevisiae* deletion strains identifies genes important for cell cycle progression. *FEMS Microbiol. Lett.* **223**, 253–258 (2003).
95. Wiederkehr, A., Meier, K. D. & Riezman, H. Identification and characterization of *Saccharomyces cerevisiae* mutants defective in fluid-phase endocytosis. *Yeast* **18**, 759–773 (2001).
96. Kumar, A. *et al.* Subcellular localization of the yeast proteome. *Genes Dev.* **16**, 707–719 (2002).
97. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
Reports the construction of ~6,000 yeast strains that contain GFP-tagged genes in their natural genomic location. The localization of tagged product was detected by fluorescence microscopy in ~4,000 strains.
98. Ashrafi, K. *et al.* Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272 (2003).
99. Kiger, A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27 (2003).
100. Reboul, J. *et al. C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genet.* **34**, 35–41 (2003).
101. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
Insertional mutants in ~75% of *Arabidopsis* genes were made, although the entire set was not screened.
102. Beckers, J. & Hrabe de Angelis, M. Large-scale mutational analysis for the annotation of the mouse genome. *Curr. Opin. Chem. Biol.* **6**, 17–23 (2002).

Acknowledgements

The authors thank D. A. Guertin for helpful discussions, and M. Boutros and N. Perrimon for sharing unpublished work.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

ColiBase: http://colibase.bham.ac.uk
LacZ

MODEL ORGANISM DATABASES:

***A. thaliana* (TAIR):** http://www.arabidopsis.org
***C. elegans* (WormBase):** http://www.wormbase.org
***D. melanogaster* (FlyBase):** http://www.flybase.org
***H. sapiens* (Human genome resources):**
http://www.ncbi.nih.gov/genome/guide/human
***M. musculus* (Mouse genome informatics):**
http://www.informatics.jax.org; (Mouse genome server):
http://www.ensembl.org/Mus_musculus
***S. cerevisiae* (*Saccharomyces* genome database):**
http://www.yeastgenome.org

FURTHER INFORMATION

Open Microscopy Environment: www.openmicroscopy.org
Anne Carpenter's web site:
http://web.wi.mit.edu/sabatini/pub/AnneWeb/index.htm.
David Sabatini's laboratory: http://web.wi.mit.edu/sabatini.
Access to this interactive links box is free online.