Tailoring the genome: the power of genetic approaches

Andras Nagy[1], Norbert Perrimon[2], Suzanne Sandmeyer[3] & Ronald Plasterk[4]

doi:10.1038/ng1115

In the last century, genetics has developed into one of the most powerful tools for addressing basic questions concerning inheritance, development, individual and social operations and death. Here we summarize the current approaches to these questions in four of the most advanced models organisms: *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (worm), *Drosophila melanogaster* (fly) and *Mus musculus* (mouse). The genomes of each of these four models have been sequenced, and all have well developed methods of efficient genetic manipulations.

Almost 50 years after the discovery of the structure of DNA, we are able to 'read' and 'write' genetic code. With the sequencing of complete genomes underway and technologies to trace polymorphisms among individuals being improved, our ability to read genetic code is certain. But to what extent can we write DNA? If the challenge is to create a particular genetic change, we are almost there, but if the task is to produce a given property in an organism, then we have a long way to go. We do not yet fully understand the entire process of the 'natural construction' of a given organism from its genetic code, or 'blueprint'.

The history of each of the model organisms is slightly different. Beginning in the 1940s, yeast had proven to be useful as a classical genetic model and in the 1980s it began to be used as a molecular genetic model. In the last decade, yeast has been used for its genomic/genetic advantage. The worm is the youngest model system (just turning 40). It is a very simple animal, having only 959 somatic cells. The complete cell lineage and wiring of the worm's nervous system (302 cells) are known, and it was the first multicellular organism to have its genome sequenced. The fly has been a favored system for genetic studies for more than 90 years and has proven to be an excellent model to identify genes involved in evolutionarily conserved developmental and cellular processes. Finally, the mouse has been used as a model for genetic studies for as long as the fly. Fancy breeders of different coat color variants were the first mouse geneticists, whose work was quickly recognized by different scientific fields in biology. The range of capabilities for altering the genes and genomes of these four organisms is more or less at the same level, which allows them to be analyzed with comparable approaches.

Genetic approaches: definition matters

Classical genetics as pioneered by Mendel[1], continued by Morgan[2] for spontaneous variants and followed by Muller[3] for induced mutations, is based on detection of individuals with phenotypically apparent mutations. Starting with the phenotype and working toward identifying the affected gene sequences is referred to as 'forward genetics'.

The starting point of reverse genetics is a gene of interest, for which the sequence is usually known. This genetic approach involves inducing either a structural alteration of a gene (regulatory or coding mutation) or transgenic expression; the phenotypic effect of such a manipulation is then studied. Both forward and reverse genetics establish a connection between changes in genotype and changes in phenotype, with the aim of understanding how genes control inheritance.

The advent of new methods of genome manipulation and the availability of full genome sequences led to the development of innovative genetic approaches with unexpected effectiveness. New subcategories have been specified, such as 'forward mutant' or 'phenotype-sensitized screens', which focus on a phenotype of interest and may lead to the discovery of a diverse set of genes, such as those encoding kinases, transcription factors and membrane proteins, or genes affecting the same disease. Reverse genetics programs usually begin with a well defined class of genes (such as those encoding nuclear hormone receptors, transcription factors or receptor kinases), but may end in the illumination of diverse biological phenotypes. A variant of reverse genetics is 'target-selected gene changes'. This is considered reverse genetics because the starting point is a gene of known sequence. In this case, the gene is not changed in a directed fashion—for example,

[1]*Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada; Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada.* [2]*Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Ave., Boston, Massachusetts 02115, USA.* [3]*Department of Biological Chemistry, Medical Sciences I, D240, College of Medicine, University of California, Irvine, California 92697-1700, USA.* [4]*Hubrecht Laboratory, Center for Biomedical Genetics, Uppsalalaan 8, 3584 CT, Utrecht, the Netherlands. Correspondence should be addressed to A.N. (e-mail: nagy@mshri.on.ca).*

by targeting for a specific alteration—but by subjecting the model organism to random mutagenesis. Thus, individuals with changes in the gene of interest are recognized at the DNA level. The approach was pioneered in the fruit fly by Seymour Benzer and Kim Kaiser. It has since been widely applied in this model system, and is now also beginning to be used for the others. One example is the recent development of chemical mutagenesis in mouse embryonic stem (ES) cells and its combination with high-throughput sequencing of mutations in a gene of interest.

Rapidly emerging genetic technology has blurred the borders between forward and reverse genetics. In many cases, the simple forward versus reverse categorization no longer holds. The field is rapidly moving toward one form of genetics with many different approaches, all aimed at understanding the interplay among genes in determining the functions of a complex organism, from inheritance to both normal and disease-related development.

Source of variation

Variation among individuals of a species is a fundamental requirement of genetics, which deals with the inheritance of this variation. For inheritance, the nature of variation has to be represented at the level of the genome (DNA mutations). One source of mutations, or variants, is natural populations. Breeders have known this for centuries. These natural or 'classical' variants are still of enormous value, even though we are now able to produce 'modern' variants with mostly induced, single-locus changes.

Induced mutations fall into three classes: (i) small changes in nucleotide sequence induced by chemicals, (ii) major rearrangements (insertions, deletions, translocations and so on) induced by irradiation or chemicals[4] and (iii) gene disruptions associated with insertion of gene trap vectors[5] or mobile elements[6], or homologous recombination–based gene targeting[7,8]. The advantage of the first two classes is that chemicals and radiation are quite efficient in creating many random point mutations or deletions per genome without too much bias regarding the sites. Their drawback is the difficulty of identifying the mutations. However, the availability of the genome sequence of the four model systems, high density of genetic markers, deletion libraries, and characterization of single-nucleotide polymorphisms (SNPs) are making this huge task more feasible[9–13].

Random insertional mutagenesis of known DNA sequence has the advantage that the mutated locus (flanking genomic regions in all models or in a gene trap scenario, exon sequences with 5′ or 3′ rapid amplification of DNA ends (RACE); ref. 14) can be quickly identified. The insertion can be a transposon, transgene, retrovirus or even a restriction fragment (restriction enzyme–mediated integration (REMI); ref. 15). A classic case of insertion tagging occurs in the soma of mice, where retroviruses induce cancer by being inserted next to and activating an oncogene[16].

For obvious reasons, interest in targeted mutagenesis (gene knockouts) increased as genome sequences became available. Yet, the four model systems are not at the same stage with respect to this technology. Yeast and mouse are the most established; fly is still in its infancy, as is the youngest, worm. Nevertheless, even in the two latter organisms there are efficient ways to get around existing obstacles and achieve mutations in particular genes of interest.

The yeast

Ten years ago, yeast molecular genetics focused on individual genes or processes; these approaches are still among the most valuable for learning about a given process. In addition to using classical genetics coupled with mutagenesis and screening and cloning by complementation, more molecular-based approaches—such as positional mapping by chromosome blotting, targeted gene disruption, gapped plasmid rescue, Ty1 mutagenesis and oligonucleotide-based mutagenesis—were used[17]. A variety of vectors were available for different levels of expression under native or regulated promoters, and proteins were expressed and studied using epitope tags. At the global level, yeast cDNA and genomic libraries were available for studies of gene expression regulation.

The yeast genome sequence was completed[18] in 1996 and enabled the expansion of a variety of single-gene and single-process studies to a genomic scale. The availability of genome-wide information catalyzed our ability to connect information about individual genes and proteins, to better understand biological processes and to access the large collection of genes about which, until then, almost nothing was known.

Exogenous DNA is typically introduced into yeast cells either by exposing the cells to lithium acetate or by electroporation. Where it is desirable for the purposes of introducing one plasmid or genetic marker to a large number of strains, investigators have turned to mating in high-throughput format, creating a systematic array of matings. This approach has been used in two-hybrid studies to examine all possible combinations of fusion proteins for interaction[19]. Another application of this mating matrix approach has been to look for the 'synthetic' phenotypes of combinations of mutations[20].

Gene mapping, the oldest sense in which we manipulate genes, has been affected by the availability of the yeast genome sequence. In particular, strains that are significantly polymorphic can be distinguished using oligonucleotide arrays such as the Affymetrix GeneChip[11]. If these strains also differ in some property that can be quantified through a number of backcrosses, then it is possible to map the regions that correlate with that property. The power of this approach will be seen in its application to studying multigenic traits or random mutations with unusual properties. This power has yet to be fully realized, in part because of the expense of these types of arrays.

The availability of the yeast genome sequence introduced the possibility of massive, parallel genome disruption and gene tagging, and with it the possibility of learning about many genes simultaneously. Ongoing developments in miniaturization, robotics and informatics contribute to the success of this type of approach. In a broad sense, global yeast genome disruption strategies are an extension of strategies used to study individual genes. These strategies fall into two categories: directed and random. Directed strategies typically rely on PCR-based amplification of fragments that are integrated by homologous recombination mediated by short, terminal sequences that are homologous with the borders of the target sequence, resulting in replacement of the target sequence with the exogenous fragment. Depending on the intent of the experiment, the sequences included in the fragment to be introduced can function as (i) unique identifiers, or 'bar codes', for the newly created strain or a disruption site[21]; (ii) mediators of site specific recombination, allowing partial looping out of the insertion; (iii) selectable markers for forward and/or reverse selection of cells containing insertions (as for *HIS3*, *URA3* and others); (iv) reporters of gene or protein activity or location (green fluorescent protein or *lacZ*); or (v) epitope tags (such as HA, His, TAP).

A consortium of yeast investigators has produced close to a comprehensive collection of strains[21], deleted for ORFs greater than 100 codons which is now available from Open Biosystems (www.openbiosystems.com). Each strain can be distinguished by unique sequence bar codes, the gene for kanamycin resistance that is incorporated at the site of the disruption and the lack of one of the ORFs that is longer than 100 codons.

Two different approaches have been used to create collections of strains randomly disrupted at different loci. Early on, Ty1 transposon elements were used as tools for disruption[22]. It is now known that they are not efficient as random mutagens, as originally hoped, because of a biased insertion pattern. But Brown and colleagues[23] successfully used this approach to footprint genes that were essential for growth under certain conditions. It is possible to mobilize bacterial transposable elements that are relatively random and then use them to mutagenize DNA *in vivo* or *in vitro* (for example, using mini-Tn3[24] or Tn7[25] derivatives respectively). This approach has been used to create a mutagenized yeast library by Snyder and colleagues[26,27]. In their design, the random insertions of mini-Tn3, which contained a reporter gene, selectable markers and an epitope tag, helped to identify over 300 expressed but previously non-annotated genes. The availability of non-biased insertions, particularly insertions into 3′ regions, allowed essential genes to be recovered and also allowed reporter fusions for a number of genes to be used for the localization of gene products.

Concomitant with the development of disruption strain collections came the need to develop phenotypic assays that could be used effectively with these libraries. These assays were carried out in microtiter dish format. Phenotypes lending themselves to high-throughput analysis include growth rate under various conditions, ability to sporulate and sensitivity to a wide variety of chemicals[27]. Coupled with informatics, particularly clustering, such studies have led to new insights about drug activity and predictions of functions of unknown genes. Analysis of sets of disruption strains has been done for cells grown under several standardized conditions[21,28]. The presence in each disruption strain of unique sequence bar codes allows these strains to be grown in mixes and analyzed for competitive advantage under different conditions. This approach could potentially enhance the sensitivity of analysis well beyond the signal that can be measured by judging differences in growth rate among individual strains under a specified condition. Moreover, it has the potential to definitively identify genes that are essential and, therefore, potential drug targets. Notably, this approach has already shown that genes that are highly regulated under a given condition are not necessarily essential for survival under that condition. Analysis of the random collection, which is larger and more complex, may offer more information than the knockout collection about function because of the multiple positions at which individual genes are disrupted.

In addition to these global strategies of studying expressed gene functions, reporter genes can be targeted for studying chromatin functions. Targeting reporters to telomeric and ribosomal DNA has improved understanding of the stochastic nature of silencing. To monitor switching between expressed and non-expressed sequences, selectable markers that offer both forward and reverse selections to monitor expression status, such as *URA3*, are useful[29].

The worm

Encouraged by the successful creation of the new science of molecular biology, researchers at the Medical Research Council Laboratory of Molecular Biology considered what to do next. This institute had seen the discovery of the triplet code, nonsense codons, mRNA, the double helix DNA structure and protein as well as nucleic acid sequencing technology. Several lessons were learned from these successes. First, a model organism should be as simple as possible but as complex as required. Second, the genetic code is better studied in bacteriophages, with a replication time of 20 minutes, than in whales. In addition, to find important answers, one needs to ask important questions. What was important? Brenner stated in 1963 that the problems of genetics were solved (this was before introns or the role of DNA methylation had been discovered) and that biology had only two interesting questions remaining: how do we develop and how do we think? To answer these important questions, one needed to choose an organism that thought and developed while remembering the phage lesson: the model should be as simple as possible. Brenner set out to find the bacteriophage of the animal kingdom and returned with the nematode *C. elegans*[30].

This animal has muscles, a nervous system of 302 cells, a digestive tract and sexual reproduction. One sex is the male, the other the hermaphrodite. As the latter is self-fertilizing, one animal can produce a brood of 200 progeny within three days. Thus, one can toothpick an animal clone onto a fresh agar plate, as if it were a bacterial colony or bacteriophage plaque. This is what Brenner did—an animal for the price of a bacterial colony. Crosses between individuals can be initiated by adding a male worm. Hermaphrodites have two sex chromosomes and males have one; thus, self-progeny of hermaphrodites is hermaphroditic (unless chromosome non-disjunction takes place), and cross-progeny is 50% male, 50% hermaphrodite. An experimental advantage foreseen by Brenner was that a cross-section of a complete worm would fit on a grid for analysis by an electron microscope. This later allowed the reconstruction of the complete wiring of the worm brain.

An unanticipated advantage was that the genome was relatively small, allowing the worm to be the first animal to have its complete genome sequenced. It is true, however, that from the start the approach in the *C. elegans* field was to do the whole thing—not the wiring of some neurons but of all 5,000 synapses of the complete nervous system; not the lineage of one tissue but development described by tracing the complete cell lineage. This spirit was later visible when Sulston and Coulson[31] decided to make a complete cosmid clone map of the worm genome and then sequence it (the success of this effort was important in the decision to sequence the complete human genome).

Brenner's agenda was to do for development and neurobiology what the phage had done for DNA replication: carry out saturation mutagenesis and determine all the genes required. The approach does not differ essentially from the one described above for yeast and below for the fly. One important difference is the self-fertilizing nature of the animal. Mendel's laws dictate that when a P0 animal is exposed to a mutagen, the F1 progeny will contain numerous heterozygous mutations, and each of those will segregate as a homozygote at 25% in the F2 generation. So, no laborious family crosses are required, and—depending on the phenotype of interest—one can simply visually inspect F2 cultures for mutants or even carry out a conditional selection (such as for drug resistance) by plating batches of F2 animals and picking surviving F3 animals. Once mutants are obtained, they can be mapped. This was done initially by using visible physical markers; here the worm has its limitations, not having fancy wings with curly edges and numerous antennae. Not all mutations of interest can be scored in combination with the common marker mutations that affect locomotion (*unc* for uncoordinated) or morphology (*dpy* for dumpy or *lin* for lineage defect). In recent years, large sets of SNPs have become available that allow easier gene mapping[10]. Once a gene has been mapped, the contiguous clone map allows selection of a set of cosmids or bacterial artificial chromosomes (BACs) that cover the mutation; rescue after transgenesis can be done to sort out which clone contains the gene of interest.

Transgenesis is commonly done by injection of DNA into the syncitial parts of the gonads[32]. The injected DNA forms long concatamers that can segregate to progeny like plasmids

in bacteria; they are stable but are lost at some frequency. This is usually sufficient to recognize rescue. For more subtle purposes, such multi-copy transgenes can be made to integrate into chromosomes. With the genome sequence available and DNA sequencing becoming cheaper, an alternative last step to identifying a mutated gene is to skip rescue by transgenesis entirely and sequence candidate genes. Another alternative, developed most recently, is to compare the mutant phenotype with available (or newly generated) RNA interference (RNAi) data for candidate genes, and thus guess the gene identity before confirming it by DNA sequence analysis.

This basic method of gene discovery can be embedded in more subtle schemes–for example, for second-site suppressors and enhancers of another mutation, conditional mutants (thermosensitive mutants are popular: worms are commonly grown at 20 °C; the permissive temperature is 15 °C and the non-permissive is 25 °C), maternal effect mutations and so on. In essence, this does not differ from the toolbox in yeast or flies. Common alternative mutagens to the frequently used chemical ethyl methane sulfonate (EMS) are other chemicals and transposons (usually the transposon Tc1). The latter allow quick gene identification, as a transposon serves as a molecular tag, again similar to Ty in yeast, P elements in flies and retroviruses in mice.

For years this was the only approach available to link genotype to phenotype. More recently, the genome sequence has given a major push toward reverse genetics[33]. Direct gene inactivation by homologous disruption has not yet become a standard technique in the worm, but is getting close, now that ballistic transformation using DNA-coated gold beads allows easy generation of many stable transformed animals. Also, the route of entry of DNA after particle bombardment may differ in important ways from that after DNA microinjection into gonads, but this is not yet clear. Two alternatives have caught on, however. One is target-selected gene inactivation. Worms are mutagenized, either by transposon-jumping or by a chemical that generates DNA deletions, such as psoralene. After brief culturing, samples from cultures are analyzed at the DNA level to detect alterations in a gene of interest. After detection, mutants of interest are cloned out. Hundreds of worm genes have been knocked out using this approach. The approach is aided by a property of the worm that should not be underestimated: cryopreservation. Glycerol stocks of worms can survive in a –70 °C freezer, and many years of tedious stock-keeping have thus been avoided. In this context, worm freezing has been used to generate permanent mutant libraries of mutagenized worms plus a corresponding set of DNA samples; knockouts are now dug out of a freezer after DNA analysis[34].

A second approach, RNAi, recently became available. RNAi was discovered in the worm[35] and has since been applied to many other organisms. Double-stranded RNA (dsRNA) can be applied to worms by microinjection, but one can also administer it by soaking the worms in it[36]. A genome-wide approach has become available through the generation of a library of dsRNA-producing plasmids in *Escherichia coli*[37]. Worms are commonly fed *E. coli* in the laboratory; if the bacterium contains a plasmid that carries a worm gene plus two promoters, one for each strand of RNA, and if the worm eats the bacteria, this gives sufficient exposure to elicit an RNAi response. Over 16,000 of the predicted 19,000 *C. elegans* genes are represented in this library, gridded and frozen in microtiter plates. A search for genes involved in a biological process can now be carried out by growing worms on 16,000 different food sources and inspecting their progeny. The approach has advantages as well as disadvantages when compared with classic genetics. RNAi does not work for every gene; it is estimated that around 60% of the genes in current protocols give full or partial loss of gene function on RNAi. Thus, genes are certainly missed. On the other hand, the partial effect of RNAi also works the other way: essential genes are easily missed or underrepresented (by rare viable alleles) in genetic screens, whereas the leaky RNAi approach may visualize viable phenotypes for essential genes. The main advantage of RNAi is that the identity of all RNAi clones is known, so that gene identification is direct and requires no laborious mapping steps. Thus, the first genome-wide RNAi screens yielded more new genes underlying the phenotypes under study than the entire field of worm genetics had generated in forty years[38].

Once genes that are involved in some aspect of biology have been identified, there are many ways to study them in more detail. Conditional gene expression can be achieved using tissue-specific promoters or inducible promoters (the most popular is the heat-shock promoter). Mosaic analysis can be used to test in which cells a gene is required for its function. The expression patterns of genes can be followed with antibodies, but a quick initial analysis is to generate a transgene that has the gene fused to GFP (this application of GFP was first developed in the worm[39] and has since been used for other organisms).

The tendency to 'do the whole thing' has also applied to bioinformatics. Wormbase[40] contains not only all of the predicted genes and their annotation but also expression patterns and other information. There is an extensive database containing gene expression data derived from DNA array experiments and an extensive analysis of cumulative data in 'expression mountains'[41]. The genome of the related nematode *Caenorhabditis briggsae* has been sequenced, allowing recognition of conserved sequences. Genome-wide yeast-two-hybrid studies are underway to assay protein–protein interactions on a global basis[42].

What is missing in the worm? Gene traps. In general, transposon-mediated transformation does exist[43], but is not very efficient and is not yet commonly used. Conditional tissue-specific knockouts using Cre-*lox*P technology are being developed. Other missing technologies include a two-component Gal-UAS system, easy and routine homologous gene disruption and permanent cell lines (short-term cell cultures do exist). Re-generation of animals from cultured stem cells, as done in mice, also does not exist. Not all of these tools will be developed, because they are not all equally essential in the worm field.

The fly

So far, most discoveries in the fly have been made by taking a forward genetic approach, whereby genes are identified by their mutant phenotypes as part of a large-scale mutagenesis using either chemical mutagens, such as EMS, or P-element transposons. Two approaches in particular have been used to identify functional links between gene products. The first relies on the assumption that genes with similar mutant phenotypes encode molecules that act in the same biochemical pathway[44,45], a hypothesis particularly successful if the phenotype under analysis is rare. The second approach consists of screening for second-site modifiers (suppressors and enhancers) of a specific sensitized genetic background[46]. In the past ten years, a number of advances have been made that facilitate forward genetic analyses in *Drosophila* (reviewed in refs. 12, 13). In particular, a large collection of P-element insertions, established mostly by the Berkeley *Drosophila* Genome Project (BDGP; www.fruitfly.org) is now available and can be screened directly for mutant phenotypes. This collection is still expanding and encompasses various P-element derivatives, including some designed to act as GAL4-enhancer trap vectors that can generate GAL4 lines for misexpression screens.

The P-element insertion collection is also useful for mapping and cloning genes defined by mutations induced using EMS. EMS is the preferred mutagen in *Drosophila* because, unlike the P-element transposable elements, it shows little target specificity and induces mutations at a high frequency. A drawback of EMS is that the molecular lesions it induces are mostly single base-pair changes, making the mapping of EMS-induced mutations to specific genes quite tedious. One strategy to map and molecularly characterize these EMS mutations is to first use meiotic and deletion mapping to narrow the position of the EMS-induced mutations and then generate local jumps of P elements to recover an insertion in the gene of interest. This is an efficient method, as most genes are within 10–100 kb of a P-element insertion (Flybase Consortium, 1999; flybase.bio.indiana.edu), and local jumps of P elements occur at high frequency[47]. Recently, maps of SNPs have been generated; these molecular polymorphisms can be used to quickly define a small genomic region to which the EMS-induced mutations map[48–51]. Although more work is needed to expand the SNP collection, it is likely that this approach will become widely used to map EMS-induced mutations.

A number of tools, initially developed to either control gene expression (GAL4-UAS system)[52] or promote mitotic recombination between chromosomes (FLP site–specific recombinase system)[53], have been used to design a new generation of forward genetic screens (reviewed in refs. 13, 54). The GAL4-UAS system has been adapted to carry out modular misexpression screens[55], which use a P element that carries UAS sites at one end (EP elements). Gain-of-function phenotypes can be generated when an EP line, inserted in the promoter region of a gene, is crossed to a GAL4 driver. The FLP–FRT (FLP recombinase target) recombination system has been used extensively to generate clones of homozygous mutant cells in otherwise phenotypically wild-type animals. So far, many screens to systematically analyze the function of pleiotropic genes in various tissues have been conducted[54, 56–59]. This approach should continue to be fruitful, as it can be expanded to virtually every cell type in the animal. Finally, P elements have been constructed to conduct 'gene trap screens', in which a P element that contains the GFP gene is mobilized at random. When this element is inserted into an intron, a fusion protein between GFP and the endogenous protein can arise, generating GFP expression patterns[60] that mimic the expression of the trapped gene. Because the frequency of P-element insertions within intronic sequences is very low, an efficient method to identify these events is to use an embryo sorter that selects the GFP-expressing embryos[61].

Although forward genetic approaches will remain valuable for identifying *Drosophila* gene functions, the completion of the sequence of the *Drosophila* euchromatic regions has emphasized the need for efficient reverse genetics methodology in this organism. The analysis of the *Drosophila* genome by Celera and BDGP has led to the annotation of 13,676 genes[62,63]. Notably, the current literature discusses only approximately 20% of these genes, and only half of these have been characterized genetically. Thus, it is clear that a wealth of information has yet to be gathered from this model organism, and that new techniques that can quickly lead from sequence to function are needed. In addition, using *Drosophila* Affymetrix oligo chips[64] and *Drosophila* expressed-sequence tag (EST) microarrays[65,66], a large number of new genes are being implicated in various processes, and methods to quickly validate their loss of function phenotypes *in vivo* are urgently needed.

A number of methods, based on RNAi using dsRNAs and targeted gene replacement, have been developed towards this goal. RNAi has been shown to effectively and specifically block gene expression *in vivo*[67]. DsRNAs can be delivered following injection into embryos, but can also be synthesized *in vivo* by expressing hairpin transgenes. These hairpin constructs can be generated and expressed in a tissue-specific manner using the GAL4-UAS system to control where and when the dsRNA is produced[68–71]. In addition, an important discovery was made by Clemens *et al.*[72]; they reported that the simple addition of dsRNA to *Drosophila* culture cells reduced or eliminated the expression of target genes by RNAi and efficiently recapitulated the phenotypes of loss-of-function mutations achieved by direct gene manipulations. This is a powerful method that can be applied to study various questions of signal transduction and cell biology[73]. Moreover, the combination of RNAi and microarrays can be used effectively to dissect the organization of signaling pathways[74]. Recently, methods to conduct high-throughput RNAi screens in *Drosophila* cells cultured in 384-well plates have been developed (A. Kiger, M. Boustros & N.P., unpublished data). Using this methodology, genome-wide RNAi screens can be conducted to quickly identify the function of all predicted *Drosophila* open reading frames in specific cell-based assays.

A targeted gene replacement method also has been established, allowing the substitution of a wild-type gene with a modified copy (refs. 75, 76 and reviewed in ref. 12). The introduced copy can either destroy or alter the activity of the target gene. This method requires the generation of a line that contains the targeting construct as a transgene. This is followed by excision of the targeting construct using FLP recombinase–mediated excision, and subsequent cleavage of the circular plasmid using the rare cutter I-*Sce*I enzyme[77]. Although further improvements are needed to make this method user-friendly, it is a promising approach for silencing and modifying genes.

The mouse

Modern genetics in the mouse started with the search for genes underlying classical mutations with obvious candidates, such as mutations causing deficiencies in structural proteins[78] or enzymes[79]. Positional cloning and fine genetic mapping of these types of mutations resulted in many newly identified genes with profound roles in developmental and disease processes. Gene knockouts also helped by frequently identifying genes responsible for long-known classical mutations[80,81]. Adaptations of methods developed in other model organisms further increased the power of genetic approaches to reach beyond the limited range of classical mutations. These advances made it possible to saturate the entire mouse genome by chemical and insertional (gene trap) mutations. From the pioneering work of William R. Russell, who established the foundations of induced mutagenesis in the mouse[82], techniques developed to generate point mutations with ease by the use of N-ethyl-N-nitrosourea (ENU)[83]. Several centers have been formed to carry out large-scale chemical mutagenesis screens in the mouse. Even at this early stage, they provide useful information about genes involved in critical developmental and physiological processes[9].

By the 1980s, the experimental embryology and molecular biology of the mouse had reached a point where their parallel development allowed a fruitful fusion. Sufficient understanding of gene structure and *in vitro* DNA sequence tailoring, combined with the efficient *in vitro* manipulation of the mouse embryo at pre-implantation stage, led to the first production of transgenic mice expressing a viral[84] and a human[85] gene. Introduction of exogenous, actively expressed transgenes into the mouse was the first and most extensively used genetic approach. It has provided invaluable information about the role of oncogenes and genes involved in development or cell function in normal and disease processes. Most of these transgenic studies were gain-of-function

approaches. The advent of mouse ES cells[86,87] opened up the possibility of loss-of-function forward and reverse genetics. These useful cells can be kept undifferentiated in permanent cultures but retain the capability of differentiating into almost any cell type, even after extensive *in vitro* culture, antibiotic drug selection and subcloning, when they are placed back into an embryonic environment by injection into or aggregating with a host embryo. Most importantly, they can contribute to the germ cells of the resulting chimeric mice. Moreover, they can be used to produce a completely ES cell–derived embryo or animal when they are aggregated with a tetraploid mouse embryo host[88]. In this scenario, the diploid ES cells outcompete the tetraploid cells in the entire embryo and allow the tetraploid cells to provide essential extra-embryonic membranes such as the placenta. Recent developments in this field showed that ES cells derived from F1 hybrid embryos, particularly from those made between C57BL/6 and 129 inbred strains, have superior developmental potential to that of cell lines derived from inbred embryos[89]. These F1 ES cell lines promote the survival of completely ES cell–derived animals at a very high efficiency, and they retain this potential after intensive *in vitro* culturing and subcloning for genetically modified lines. All of these features provide a straightforward route to the mouse for genome alterations, achieved *in vitro* in the ES cell culture. In fact, ES cells have become the method of choice for manipulating the mouse genome.

Gene trapping in mouse ES cells started as a forward genetic approach[90]. The idea, as we have seen above, was first born in the *Drosophila* research[5], but soon became a popular and powerful method of insertional mutagenesis in mouse ES cells (reviewed in ref. 91). Random insertion of a promoterless splice acceptor/reporter construct into a gene allows simultaneous detection of gene transcription and mutation of the host gene. The insertion of exogenous DNA into the ES cell genome is fairly random[92], but with specific designs, desired biases can be achieved toward certain gene categories or insertion types. Most of the vectors are trapping genes, which are expressed in ES cells[93]. There are, however, strategies and corresponding trapping vectors for genes that are not expressed in undifferentiated cells, but become active in certain differentiation pathways or in the presence of transcription factors[94,95]. Other vectors specifically trap genes encoding secreted proteins[96] or are simply designed to identify a representative piece of sequence of a trapped gene[97]. Today, gene trap screens may have an increasing reverse genetic component: the approach starts with the identification of the trapped gene, and on the basis of this information, a decision is made whether further phenotype analysis will follow[97]. This is one of the best examples of how the border between forward and reverse genetics is diminishing. According to a current estimate, there are about 9,000 mouse genes already trapped in the public domain; this number will probably reach 40,000–50,000 within the next three years. Analyzing this number of mutations will be challenging. It is thus likely that at least certain 'branches' of gene traps in the mouse will turn to ES cell–based *in vitro* prescreens, as methods have been developed to achieve homozygosity in mutant ES cells[98,99].

Not all genome-wide, random mutagenesis approaches are designed for negative interference of gene expression or gene function. Similar to the fly[55], a specifically designed gene trap insertion in the mouse can create random gain-of-function alterations as well. In this scenario, if the insertion occurs into the first exon or intron, the promoter sequence in the trap vector takes over the expression of the 3′ positioned trapped gene[100].

Homologous recombination–based gene targeting[7,8] has had a large impact on our understanding of gene function in the mouse. It is difficult to estimate the number of mouse genes that have been knocked out so far by gene targeting, but the number certainly exceeds 5,000. Clearly, we have discovered much about gene function in normal development and disease processes. We have also learned the limitations of creating null alleles, however, as many genes have multiple functions that shift in different developmental stages. Often, the lack of an early function compromises an individual or embryo at a level that interferes with the investigation of later functions. To dissect multiple functions, other tools have to be combined with gene targeting. Of these new tools, the site-specific Cre recombinase of the P1 phage has been the most powerful. It efficiently recombines between two consensus sites (*lox*P) of 34 bp that are not represented in the mouse genome, but can be placed there by transgenic or targeted insertion.

Increasing numbers of designs combine transgenes, homologous recombination–based targeting and Cre/*lox*P site–specific recombinase systems to achieve a wide range of genomic alterations, from introduction of a point mutation through small or large genomic deletions to reciprocal translocations[101]. Two other site-specific DNA recombining enzymes, FLP recombinase[102,103] and ϕC31 integrase[104], are joining the Cre recombinase system in providing possibilities for further sophistication in manipulating the mouse genome. Our current capabilities in the mouse have reached the point where we can generate practically any 'genocopy' of mutations identified, for example, in human disease.

Yet, we can do even more: *lox*P sites can be placed to flank essential exons of a gene in a way that does not interfere with the gene function in the cell (producing a conditional allele; reviewed in refs. 105, 106). When Cre recombinase is expressed in the same cell, it removes the flanked pivotal exons and creates a null allele. With a simple breeding scheme combining the conditional allele and a transgene expressing the Cre recombinase in a spatially or temporally restricted manner, it is possible to achieve cell type/tissue–specific or developmental stage–specific gene modification (such as knockouts) or genome alterations. For a temporal control, the tetracycline-inducible transgene expression system[107], applied on induction of Cre recombinase expression, serves as an example[108]. This is one of the most powerful genetics tools in the mouse system for dissecting multiple gene functions.

What lies ahead

The yeast. The potential for probing the yeast genome has barely been tapped. In every dimension of this relatively simple cell, much remains to be learned—from DNA to RNA to protein; from organelle to cell to colony and environment—as we struggle to document and understand the complex web of interactions and their regulation[109]. Powerful proteomics is being developed on the foundation of yeast genomic research, as the current trends of technology continue toward smaller samples, greater instrument sensitivity and parallel analysis in roboticized systems. In terms of genomic manipulations, it is likely that more use will be made of labeled or polymorphic strains to understand the quantitative effects of collections of genes on particular processes. One of the most exciting challenges is to develop computational tools with which to simulate these processes. The challenge of continuing to develop 'technomics', however, pales next to the challenge of integrating and using the information collected.

The worm. Brenner's dream was to understand a simple brain and the resulting behavior derived from its DNA. This requires a full understanding of all intermediate steps, from protein folding to development and nervous system action. It seems likely that this will never be achieved; the more realistic goal is that for each

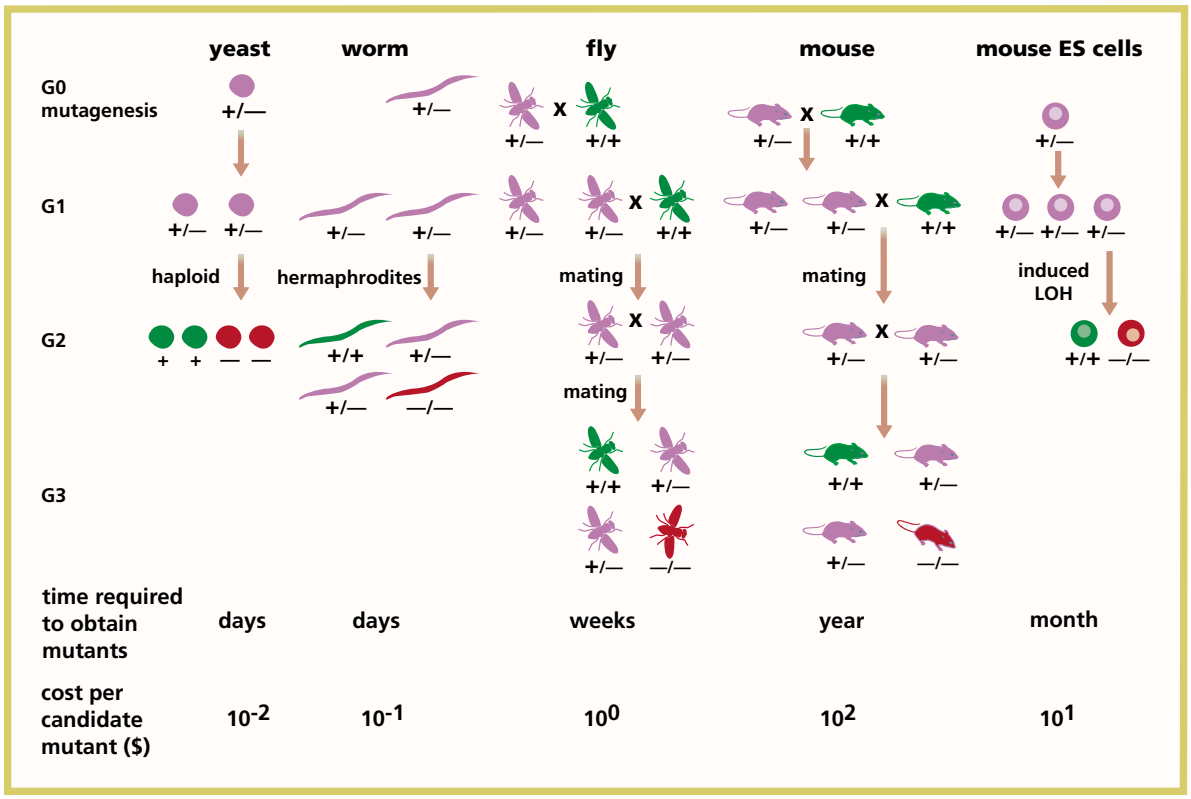

Fig. 1 Schematics of typical recessive chemical mutagenesis screens in the model systems discussed: yeast, worm, fly, mouse and ES cell. Rough estimates of the time required to obtain mutants and of the cost ($) per candidate mutant before any phenotyping was done are also indicated. +/+(green), wild-type; +/– (yellow), heterozygous carrier; –/– (red), homozygous mutant.

fundamental process, a few examples will be analyzed in depth, after which one can conclude that the rest works more or less the same. The genetic analysis of apoptosis in *C. elegans* was recently recognized by a Nobel prize (for Robert Horvitz and Brenner and Sulston). Lifespan, taste and smell, genome stability, the first asymmetry in early development, organogenesis and sexual behavior—these are just a few of the many aspects of animal life that will ultimately be solved in the *C. elegans* system.

The fly. The advances made in the past ten years provide an excellent example of the effectiveness of new technologies for discovering gene functions, as perhaps best illustrated by the advances made possible by the RNAi method. What lies ahead? First, in the next few years, there will be significant expansion in the availability of tools that facilitate forward genetic studies. More extensive collections of the P-element insertions, as well as other transposable elements with different insertional specificity (such as Piggyback), will become available. In addition, it is likely that more complete SNP maps will be constructed, facilitating the molecular characterization of EMS-induced mutations. Second, more effective methods for targeted gene replacement may be developed. The current methodology is cumbersome, as it requires generation of transgenic animals. Thus, methods in which a targeting construct or molecule is simply injected into the developing embryo would be preferable. As such, studies on Group II introns that can be designed to insert into specific DNA targets are of considerable interest[110]. Third, there will be an abundance of information from the interpretation of the *Drosophila* genome sequence and its derived applications. We can expect that a large amount of information will be generated from expression genomics (microarrays) and functional genomics (RNAi). Furthermore, proteomics such as the construction of maps of protein–protein interactions built from either large-scale two hybrid screens or mass spectometry will probably become a major focus of study. Finally, we can also expect that 'chemical genetics', identifying compounds on cell-based assays (see for example ref.111), will be combined with high-throughput RNAi screens to identify protein targets of specific compounds. Considering the upcoming wealth of information that will be generated from various large-scale approaches, it is crucial that the primary data be carefully recorded in well designed databases and that powerful computational tools allowing one to extract biological information be developed.

The mouse. Modern mouse genetics builds upon the ES cell–based manipulations of genes or the genome and has thus gained theoretically unlimited capabilities. There is a practical limitation, however: money (Fig. 1). Even if genome manipulation is done on ES cells, it is not difficult to see what an impossible task it may be to keep the results of a genome-wide mutagenesis project in mouse cages. Fortunately, mouse ES cells can provide more than simply germline transmission. These cells are similar to those of yeast in that they are cultured as individual cells in huge numbers, and are easy to mutate with chemical mutagens[112,113]. As positive and negative selectable marker genes are operational, it is easy to introduce DNA into the cells. Homologous recombination can be used for gene targeting. Heterologous enzymes, such as Cre, Flp site–specific recombinases and ϕC31 integrase, are useful for post-integrational modifications of transgenes or insertions. ES cells even act *in vitro* beyond simply proliferating as stem cells, as they can differentiate into different cell types depending on the culture conditions.

What is missing? The haploid screen, as haploid ES cells do not exist. Several ways to get around this problem are now emerging, however. ES cell deletion libraries are being generated[114], which provide starting reagents for haploid screens of relatively small genomic regions. A whole chromosome could be made available for haploid screen–equivalent approaches by making the ES cell mutations homozygous in a chromosome-specific manner[99,115]. RNAi, which is capable of silencing gene expression in diploid cells, may provide a solution as well. Many laboratories are establishing stable transgenic systems for expressing short interfering RNAs with the *Pol*III promoter in a reliable manner in ES cells. There are still a few steps waiting for optimization, but high-throughput *in vitro* screens, similar to those taking advantage of the haploid nature of the yeast, will soon be on the way.

Conclusion

The abundance of tools available for manipulating the genome may raise the question of whether we will need all of these methods in the future. The answer is yes. None of these approaches is better than the others. They all complement each other and cooperate in untangling the complexity of genetic determinants. At certain times, one method will dominate the others, but this dominance will be temporary as new technologies are introduced.

The limiting factor will be our understanding of how genetic changes affect phenotype. This knowledge requires an understanding of how primary sequences affect protein folding, how proteins and other factors interact in time and space in the cell and how molecular processes determine development, physiology and behavior. This comprehension may never be fully reached, and thus (contrary to popular belief) one should not confuse the advances of DNA technology with full control of the phenotype. But the tools described in this paper will contribute to partial elucidation of genotype–phenotype relations.

1. Mendel, G. Versuche uber Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brunn* **4**, 3–47 (1866).
2. Morgan, T. Sex-limited inheritance in *Drosophila*. *Science* **32**, 120–122 (1910).
3. Muller, H. Artificial transmutation of the gene. *Science* **66**, 84–87 (1927).
4. DeMarini, D.M. *et al*. Specific-locus mutations induced in eukaryotes (especially mammalian cells) by radiation and chemicals: a perspective. *Mutat. Res.* **220**, 11–29 (1989).
5. O'Kane, C.J. & Gehring, W.J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **84**, 9123–9127 (1987).
6. Craig, N.L., Craigie, R., Gellert, M. & Lambowitz, A.M. (eds) *Mobile DNA II*. (ASM, Washington DC, 2002).
7. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. Disruption of the proto-oncogene *int-2* in mouse embryo–derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
8. Doetschman, T., Maeda, N. & Smithies, O. Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **85**, 8583–8587 (1988).
9. Herron, B.J. *et al*. Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. *Nat. Genet.* **30**, 185–189 (2002).
10. Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H. & Plasterk, R.H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat. Genet.* **28**, 160–164 (2001).
11. Winzeler, E.A. *et al*. Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
12. Adams, M.D. & Sekelsky, J.J. From sequence to phenotype: reverse genetics in *Drosophila melanogaster*. *Nat. Rev. Genet.* **3**, 189–198 (2002).
13. St Johnston, D. The art and design of genetic screens: *Drosophila melanogaster*. *Nat. Rev. Genet.* **3**, 176–188 (2002).
14. Townley, D.J., Avery, B.J., Rosen, B. & Skarnes, W.C. Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Res.* **7**, 293–298 (1997).
15. Kuspa, A. & Loomis, W.F. Tagging developmental genes in Dictyostelium by restriction enzyme–mediated integration of plasmid DNA. *Proc. Natl. Acad. Sci. USA* **89**, 8803–8807 (1992).
16. Nusse, R. & Varmus, H.E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99–109 (1982).
17. Guthrie, C. & Fink, G.R. *Guide to Yeast Genetics and Molecular Biology.* (Academic, San Diego, 1991).
18. Goffeau, A. *et al*. Life with 6000 genes. *Science* **274**, 546, 563–547 (1996).
19. Uetz, P. *et al*. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
20. Tong, A.H. *et al*. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
21. Winzeler, E.A. *et al*. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
22. Garfinkel, D.J., Mastrangelo, M.F., Sanders, N.J., Shafer, B.K. & Strathern, J.N. Transposon tagging using Ty elements in yeast. *Genetics* **120**, 95–108 (1988).
23. Smith, V., Chou, K.N., Lashkari, D., Botstein, D. & Brown, P.O. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074 (1996).
24. Seifert, H.S., Chen, E.Y., So, M. & Heffron, F. Shuttle mutagenesis: a method of transposon mutagenesis for *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **83**, 735–739.. (1986).
25. Bachman, N., Biery, M.C., Boeke, J.D. & Craig, N.L. Tn7-mediated mutagenesis of *Saccharomyces cerevisiae* genomic DNA *in vitro*. *Methods Enzymol.* **350**, 230–247 (2002).
26. Burns, N. *et al*. Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.* **8**, 1087–1105 (1994).
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. Giaever, G. *et al*. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
29. van Leeuwen, F. & Gottschling, D.E. Assays for gene silencing in yeast. *Methods Enzymol.* **350**, 165–186 (2002).
30. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94.. (1974).
31. Coulson, A., Waterston, R., Kiff, J., Sulston, J. & Kohara, Y. Genome linking with yeast artificial chromosomes. *Nature* **335**, 184–186 (1988).
32. Mello, C. & Fire, A. DNA transformation. *Methods Cell. Biol.* **48**, 451–482 (1995).
33. Plasterk, R.H. Reverse genetics: from gene sequence to mutant worm. *Methods Cell. Biol.* **48**, 59–80 (1995).
34. Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T. & Plasterk, R.H. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci. USA* **90**, 7431–7435 (1993).
35. Fire, A. *et al*. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
36. Montgomery, M.K., Xu, S. & Fire, A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **95**, 15502–15507 (1998).
37. Kamath, R.S. *et al*. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
38. Zipperlen, P., Fraser, A.G., Kamath, R.S., Martinez-Campos, M. & Ahringer, J. Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* **20**, 3984–3992 (2001).
39. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
40. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. & Spieth, J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* **29**, 82–86.. (2001).
41. Kim, S.K. *et al*. A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
42. Davy, A. *et al*. A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2**, 821–828 (2001).
43. Bessereau, J.L. *et al*. Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. *Nature* **413**, 70–74 (2001).
44. Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
45. Siegfried, E., Wilder, E.L. & Perrimon, N. Components of wingless signalling in *Drosophila*. *Nature* **367**, 76–80 (1994).
46. Simon, M.A., Bowtell, D.D., Dodson, G.S., Laverty, T.R. & Rubin, G.M. Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell* **67**, 701–716 (1991).
47. Zhang, P. & Spradling, A.C. Efficient and dispersed local P element transposition from *Drosophila* females. *Genetics* **133**, 361–373 (1993).
48. Teeter, K. *et al*. Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *J. Exp. Zool.* **288**, 63–75 (2000).
49. Hoskins, R.A. *et al*. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* **11**, 1100–1113.. (2001).
50. Berger, J. *et al*. Genetic mapping with SNP markers in *Drosophila*. *Nat. Genet.* **29**, 475–481 (2001).
51. Martin, S.G., Dobi, K.C. & St Johnston, D. A rapid method to map mutations in *Drosophila*. *Genome Biol.* **2**, RESEARCH0036 (2001).
52. Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
53. Golic, K.G. & Lindquist, S. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**, 499–509 (1989).
54. Duffy, J.B., Harrison, D.A. & Perrimon, N. Identifying loci required for follicular patterning using directed mosaics. *Development* **125**, 2263–2271 (1998).
55. Rorth, P. A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. *Proc. Natl. Acad. Sci. USA* **93**, 12418–12422 (1996).
56. Xu, T., Wang, W., Zhang, S., Stewart, R.A. & Yu, W. Identifying tumor suppressors in genetic mosaics: the *Drosophila lats* gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
57. Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. Zygotic lethal mutations with maternal effect phenotypes in *Drosophila melanogaster*. II. Loci on the second and third chromosomes identified by P-element–induced mutations. *Genetics* **144**, 1681–1692 (1996).
58. Stowers, R.S. & Schwarz, T.L. A genetic method for generating *Drosophila* eyes composed exclusively of mitotic clones of a single genotype. *Genetics* **152**, 1631–1639 (1999).
59. Newsome, T.P., Asling, B. & Dickson, B.J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
60. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**, 15050–15055 (2001).
61. Furlong, E.E., Profitt, D. & Scott, M.P. Automated sorting of live transgenic embryos. *Nat. Biotechnol.* **19**, 153–156 (2001).
62. Adams, M.D. *et al*. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).

63. Misra, S. *et al*. Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* **3**, RESEARCH0083.0081–0083.0022 (2002).
64. McDonald, M.J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
65. White, K.P., Rifkin, S.A., Hurban, P. & Hogness, D.S. Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**, 2179–2184 (1999).
66. Furlong, E.E., Andersen, E.C., Null, B., White, K.P. & Scott, M.P. Patterns of gene expression during *Drosophila* mesoderm development. *Science* **293**, 1629–1633 (2001).
67. Kennerdell, J.R. & Carthew, R.W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
68. Lam, G. & Thummel, C.S. Inducible expression of double-stranded RNA directs specific genetic interference in *Drosophila*. *Curr. Biol.* **10**, 957–963 (2000).
69. Kennerdell, J.R. & Carthew, R.W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18**, 896–898 (2000).
70. Fortier, E. & Belote, J.M. Temperature-dependent gene silencing by an expressed inverted repeat in *Drosophila*. *Genesis* **26**, 240–244 (2000).
71. Kalidas, S. & Smith, D.P. Novel genomic cDNA hybrids produce effective RNA interference in adult *Drosophila*. *Neuron* **33**, 177–184 (2002).
72. Clemens, J.C. *et al*. Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl. Acad. Sci. USA* **97**, 6499–6503 (2000).
73. Ramet, M., Manfruelli, P., Pearson, A., Mathey-Prevot, B. & Ezekowitz, R.A. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* **416**, 644–648 (2002).
74. Boutros, M., Agaisse, H. & Perrimon, N. Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev. Cell.* **3**, 711–722 (2002).
75. Rong, Y.S. & Golic, K.G. Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013–2018 (2000).
76. Rong, Y.S. & Golic, K.G. A targeted gene knockout in *Drosophila*. *Genetics* **157**, 1307–1312 (2001).
77. Bellaiche, Y., Mogila, V. & Perrimon, N. I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in *Drosophila*. *Genetics* **152**, 1037–1044 (1999).
78. Roach, A., Takahashi, N., Pravtcheva, D., Ruddle, F. & Hood, L. Chromosomal mapping of mouse myelin basic protein gene and structure and transcription of the partially deleted gene in shiverer mutant mice. *Cell* **42**, 149–155 (1985).
79. Yokoyama, T. *et al*. Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. *Nucleic Acids Res.* **18**, 7293–7298 (1990).
80. Hebert, J.M., Rosenquist, T., Gotz, J. & Martin, G.R. FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* **78**, 1017–1025 (1994).
81. Pennisi, D. *et al*. Mutations in *Sox18* underlie cardiovascular and hair follicle defects in ragged mice. *Nat. Genet.* **24**, 434–437 (2000).
82. Russell, W.L. *et al*. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. USA* **76**, 5818–5819 (1979).
83. Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B. & Bradley, A. Mouse ENU mutagenesis. *Hum. Mol. Genet.* **8**, 1955–1963 (1999).
84. Brinster, R.L. *et al*. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* **27**, 223–231 (1981).
85. Palmiter, R.D. *et al*. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* **300**, 611–615 (1982).
86. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* **78**, 7634–7638 (1981).
87. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
88. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. Derivation of completely cell culture–derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
89. Eggan, K. *et al*. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. USA* **98**, 6209–6214 (2001).
90. Gossler, A., Joyner, A.L., Rossant, J. & Skarnes, W.C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1990).
91. Stanford, W.L., Cohn, J.B. & Cordes, S.P. Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2**, 756–768 (2001).
92. Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A. & Gruss, P. Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucleic Acids Res.* **25**, 1531–1536 (1997).
93. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
94. Forrester, L.M. *et al*. An induction gene trap screen in embryonic stem cells: identification of genes that respond to retinoic acid *in vitro*. *Proc. Natl. Acad. Sci. USA* **93**, 1677–1682 (1996).
95. Medico, E., Gambarotta, G., Gentile, A., Comoglio, P.M. & Soriano, P. A gene trap vector system for identifying transcriptionally responsive genes. *Nat. Biotechnol.* **19**, 579–582 (2001).
96. Skarnes, W.C., Moss, J.E., Hurtley, S.M. & Beddington, R.S. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA* **92**, 6592–6596 (1995).
97. Hicks, G.G. *et al*. Functional genomics in mice by tagged sequence mutagenesis. *Nat. Genet.* **16**, 338–344 (1997).
98. Lefebvre, V. *et al*. Characterization of primary cultures of chondrocytes from type II collagen/β-galactosidase transgenic mice. *Matrix Biol.* **14**, 329–335 (1994).
99. Liu, P., Jenkins, N.A. & Copeland, N.G. Efficient Cre-*lox*P–induced mitotic recombination in mouse embryonic stem cells. *Nat. Genet* **30**, 66–72 (2002).
100. Harrington, J.J. *et al*. Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat. Biotechnol.* **19**, 440–445 (2001).
101. Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99–109 (2000).
102. Dymecki, S.M. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **93**, 6191–6196 (1996).
103. Rodriguez, C.I. *et al*. High-efficiency deleter mice show that FLPe is an alternative to Cre-*lox*P. *Nat. Genet.* **25**, 139–140 (2000).
104. Belteki, G., Gertsenstein, M., Ow, D.W. & Nagy, A. Site-specific cassette exchange and germline transmission with mouse ES cells expressing PhiC31 integrase. *Nat. Biotechnol.* (in press) (2003).
105. Lewandoski, M. Conditional control of gene expression in the mouse. *Nat. Rev. Genet.* **2**, 743–755 (2001).
106. Kwan, K.M. Conditional alleles in mice: practical considerations for tissue-specific knockouts. *Genesis* **32**, 49–62 (2002).
107. Gossen, M., Bonin, A.L. & Bujard, H. Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends Biochem. Sci.* **18**, 471–475 (1993).
108. Saam, J.R. & Gordon, J.I. Inducible gene knockouts in the small intestinal and colonic epithelium. *J. Biol. Chem.* **274**, 38071–38082 (1999).
109. Schwikowski, B., Uetz, P. & Fields, S. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **18**, 1257–1261 (2000).
110. Guo, H. *et al*. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289**, 452–457. (2000).
111. Stockwell, B.R., Haggarty, S.J. & Schreiber, S.L. High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem. Biol.* **6**, 71–83. (1999).
112. Chen, Y. *et al*. Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat. Genet.* **24**, 314–317 (2000).
113. Munroe, R.J. *et al*. Mouse mutants from chemically mutagenized embryonic stem cells. *Nat. Genet.* **24**, 318–321 (2000).
114. You, Y. *et al*. Chromosomal deletion complexes in mice by radiation of embryonic stem cells. *Nat. Genet.* **15**, 285–288 (1997).
115. Lefebvre, L., Dionne, N., Karaskova, J., Squire, J.A. & Nagy, A. Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nat. Genet.* **27**, 257–258 (2001).