Genome-wide mutant collections: toolboxes for functional genomics

Paulo SR Coelho, Anuj Kumar and Michael Snyder*

The sequencing of entire genomes has led to the identification of many genes. A future challenge will be to determine the function of all of the genes of an organism. One of the best ways to ascertain function is to disrupt genes and determine the phenotype of the resulting organism. Novel large-scale approaches for generating gene disruptions and analyzing the resulting phenotype are underway in the budding yeast *Saccharomyces cerevisiae* and other organisms including flies, *Mycoplasma*, worms, plants and mice. These approaches and mutant collections will be extremely valuable to the scientific community and will dramatically alter the manner in which science is performed in the future.

Addresses
Department of Molecular, Cellular and Developmental Biology, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
*e-mail: michael.snyder@yale.edu

Current Opinion in Microbiology 2000, **3**:309–315

1369-5274/00/$ – see front matter

Abbreviations

BDGP	Berkeley *Drosophila* Genome Project
β-gal	β-galactosidase
ES	embryonic stem
HA	hemagglutinin
ORF	open reading frame
TMV	tobacco mosaic virus
VIGS	virus-induced gene silencing

Introduction

The sequencing of entire genomes of prokaryotic and eukaryotic organisms has resulted in the identification of most protein coding genes [1,2]. Although gene function can sometimes be deduced by comparison of the predicted gene products to those previously characterized, more often than not, the sequence information is unique and provides little information as to gene function [3]. For this reason, a number of experimental methods have been developed that link the molecular information disclosed by sequenced genomes with biological function [4,5].

Arguably the best way to determine gene function is to define the phenotype resulting from gene mutation. In yeast, as well other genetic model systems, mutagenesis has been traditionally performed by treatment of cells with chemicals or irradiation followed by screening for a phenotype of interest (e.g. no growth in medium containing glycerol as the sole carbon source is indicative of mutations in genes involved in aerobic respiration) [6]. The technique is highly saturable (i.e. one can easily obtain mutations for most of the genome) and the gene can be eventually cloned and identified after complementation of the mutant phenotype with a wild-type genomic DNA library. This process, however, can be very time consuming and is not applicable on a genomic scale.

Table 1

The advantages of insertional mutagenesis and targeted mutagenesis.

Technique	Advantages
Insertional mutagenesis	Prior knowledge of gene location is not required. Many different alleles can be generated. By designing the insertional mutagen, information about gene expression and protein function can be deduced. Relatively inexpensive.
Targeted mutagenesis	Every gene can be mutated and equally represented. Null mutations can often be generated and analyzed.

Recently, new approaches have been developed that allow the mutagenesis of large numbers of genes. These procedures are of two types: first, 'random' insertional mutagenesis in which insertion mutations are randomly generated throughout the genome followed by identification of the gene(s) affected by comparing the sequence adjacent to the insertions with the genome sequence or expressed sequence tags (EST), and second, targeted mutagenesis in which specific genes are deleted or analyzed. The advantages of these random and directed approaches are shown in Table 1.

One or both of these approaches have been applied to a wide variety of organisms including microorganisms, such as yeast and mycoplasma, and multicellular organisms such as flies, worms, mice and plants. Below we review some of these projects. In particular, much emphasis is placed on the budding yeast, *Saccharomyces cerevisiae*. With its small genome (predicted to encode roughly 6200 proteins) and its tractable genetics, *Saccharomyces* has played a prominent role in the development of many methodologies for functional genomics [7]. At present, projects to generate comprehensive gene disruption libraries in yeast — both by random and targeted approaches — have advanced rapidly and, in some cases, are nearing completion.

S. cerevisiae

Insertional mutagenesis

One of the most economical and easy ways to construct libraries of tagged mutations involves transposon insertional mutagenesis. This type of mutagenesis was first developed in *Escherichia coli* using bacteriophage transposons [8] and is now broadly used in bacteria [9],

fungi [10,11], plants [12•], *Caenorhabditis elegans* [13], and *Drosophila* [14••] .

In *S. cerevisiae*, two different systems of transposon insertional mutagenesis have been used for large-scale analysis of the genome. In a procedure known as 'genetic footprinting', the endogenous retrotransposon Ty1 has been modified and expressed under an inducible promoter and used to generate a population of yeast strains containing insertions mutations throughout the genome [15,16]. PCR analysis using a gene-specific primer and one complementary to the modified transposon can be used to deduce the relative prevalence of insertion alleles for a particular gene. If the gene is important for growth under a particular condition, yeast strains containing insertion mutations within that gene will be less prevalent after growth under this condition [15,16]. The advantages of this procedure are that it allows the analysis of any genomic sequence and that the selection is carried out on a population of cells. The disadvantage is that a stable population of mutations is not maintained and available.

An alternative method is to use *E. coli* transposons to mutagenize yeast DNA in *E. coli* or *in vitro* and then 'shuttle' the mutated DNA into yeast where the DNA substitutes for the genomic copies [17,18]. This has the advantage that unlike Ty1, which is strongly biased toward inserting into the promoter regions of genes transcribed by RNA polymerase III [19], many *E. coli* transposons are less biased in selection of their insertion site. In particular, mTn*3* multipurpose transposons have enjoyed widespread utility (Figure 1) [18]. These transposons contain a *lacZ* reporter gene lacking an initiator methionine and upstream promoter sequences, which allows for the identification of in-frame transposition events in transcriptionally and translationally active regions of the genome. Near the ends of the transposon are *lox* sites which are recognized by the Cre recombinase. Adjacent to one of these elements is a coding sequence for three copies of the hemagglutinin (HA) epitope tag [20]. After integration of the transposon in the genome, one can excise the central region by inducing Cre expression and, thereby leave behind a 93 codon in-frame HA tag. Thus, this transposon allows the analysis of three different aspects of gene function using a single cassette: first, gene expression by monitoring β-galactosidase (β-gal) activity, second, protein localization by immunofluorescence with anti-HA antibodies and third, disruption phenotypes of strains carrying either the large mTn insertion or the smaller HA tag allele [18].

Recently, a miniTn*3* multipurpose transposon has been used for the large-scale analysis of the yeast genome [21••]. This transposon was used to mutagenize a library of yeast sequences in *E. coli*. Individual mutant clones from this bank were prepared and transformed into diploid yeast to identify strains that expressed β-gal during vegetative growth and sporulation. A set of ~12,000 mutant strains, each containing an expressed insertion at a known location in the genome, was generated. Gene expression was analyzed during vegetative growth and sporulation, generating expression data for over 2000 genes. Epitope tagged proteins were localized for over 6000 strains. Finally, approximately 8000 insertion alleles were transformed into haploid yeast, and the resulting strains were analyzed for 20 different phenotypes. Mutations of approximately 400 open reading frames (ORFs) resulted in a detectable growth defect.

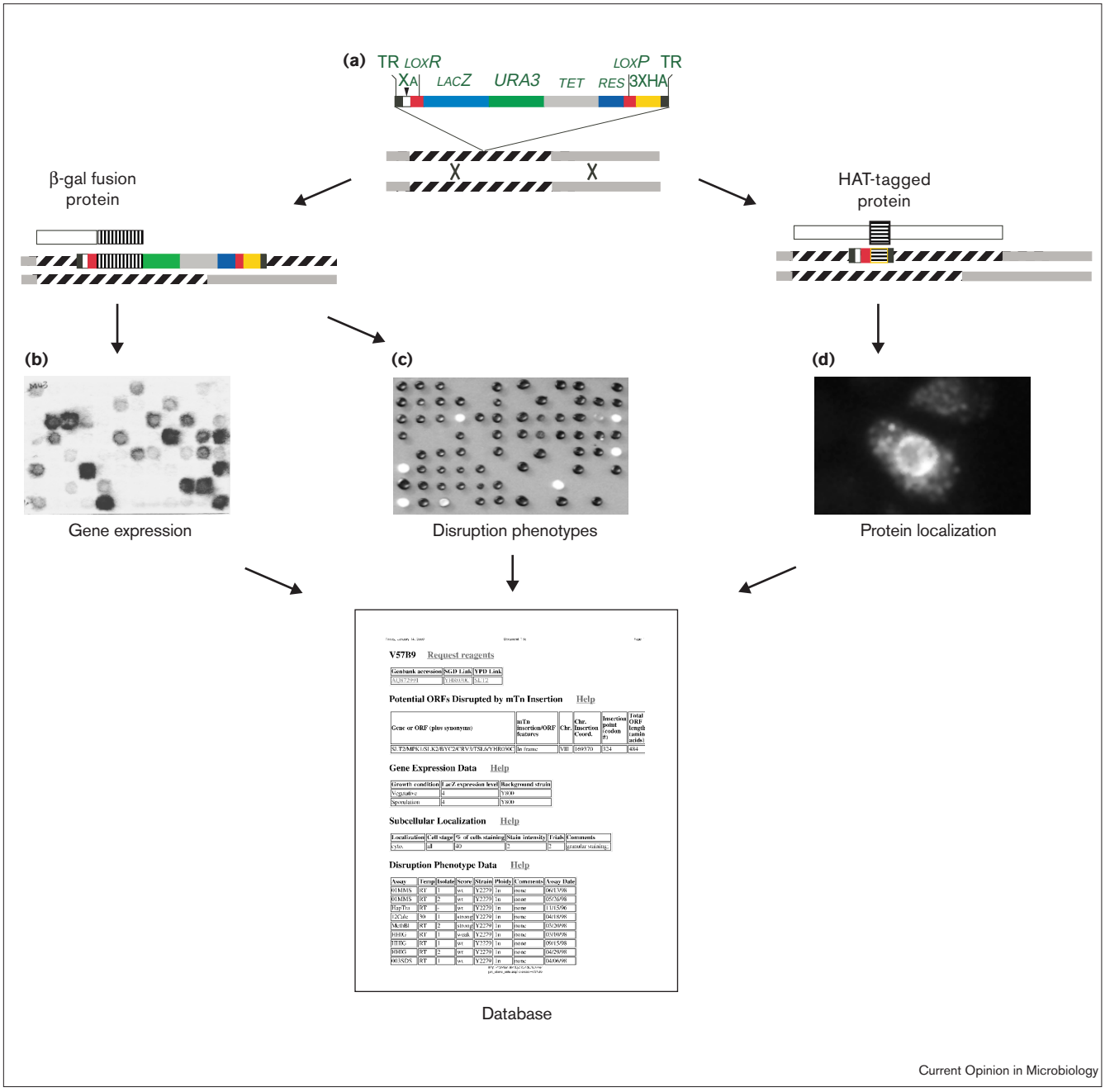
One important aspect of the insertional mutagenesis is that it identifies sequences that have not been annotated, but are capable of being expressed. Over three hundred small ORFs that are capable of being expressed were identified; since annotation was arbitrarily implemented for ORFs greater than 100 codons in length, many functionally significant ORFs were missed [22]. In addition, it was discovered that ORFs in an antisense direction in annotated ORFs were capable of being expressed, as were many regions that lie out-of-frame with annotated coding regions [21••]. These observations indicate that much more of the yeast genome is capable of being expressed than was previously realized. Presumably, expression of much of the genome is important from an evolutionary standpoint; the expression of large amounts of coding sequences is useful if any of the sequences prove beneficial under selective conditions.

One disadvantage of insertional mutagenesis is that it is not random. In the study above, on average there are three independent insertions for every gene, with some genes containing as many as 20 insertion alleles. This bias presumably comes from a number of sources, including the transposon itself and the size of the gene. Nevertheless, from a relatively modest effort, it was possible to prepare a large collection of mutated yeast strains affecting many genes and generate a wealth of information concerning these genes. Importantly, all of the information derived from the insertion points, gene expression and protein localization are integrated in a database accessible to the scientific community (http://ygac.med.yale.edu) [23]. Of equal importance, all strains and insertion alleles generated in this study are available to the research community as well [23].

PCR-based gene deletions

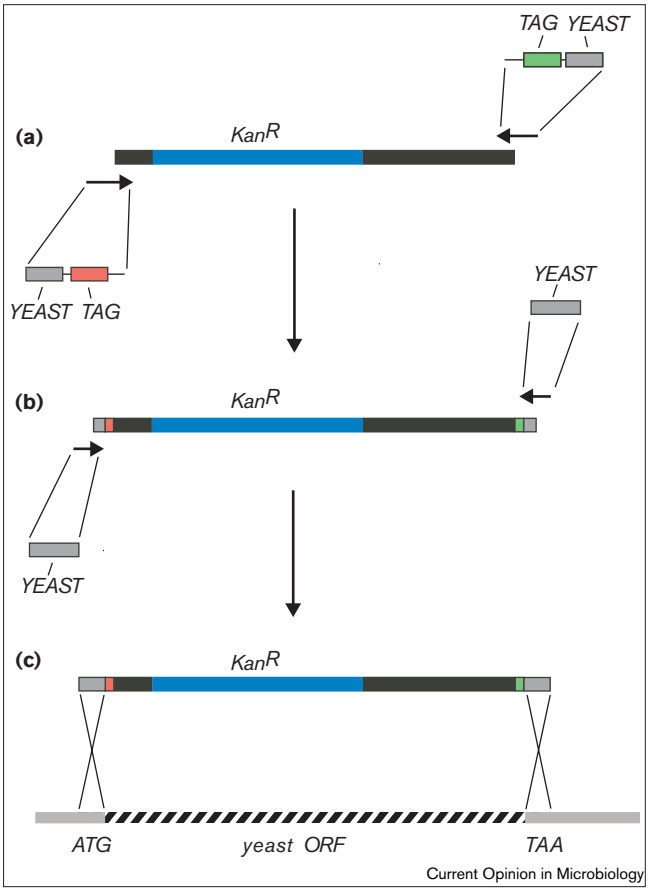
Its relatively small genome and high rate of homologous recombination make yeast an ideal organism for the construction of mutant collections via targeted mutagenesis. In this procedure, a PCR product is prepared that contains at each end 45–50 bp of sequence homology flanking a gene of interest [24]; the central region contains a selectable marker that ideally lacks homology to any genomic DNA sequence [25]. This PCR product substitutes properly at its correct region in the genome in more than 95% of the cases [26]. Thus, it is possible to prepare a deletion cassette, without any cloning step, by PCR using tailor-made oligonucleotides. One important feature of this approach is that it deletes the entire coding region, thereby generating a null mutation. A European group, the

Figure 1



Random insertional mutagenesis in yeast using a multipurpose minitransposon. By shuttle mutagenesis, a genome-wide collection of yeast mutants has been generated in which each constituent mutant carries a single insertion of a novel multifunctional minitransposon; this Tn*3*-derived minitransposon may be used to monitor protein production, localization, and function as follows. **(a)** The Tn*3*-derived multifunction transposon contains a *lacZ* reporter. Near the ends of the transposon are *lox* sites and next to one of these is the coding sequence for three copies of HA. To facilitate shuttle mutagenesis, this transposon carries selectable markers for bacteria (*TET*, encoding tetracyclin resistance) and yeast (*URA3*). Transposon mutagenesis is carried out in *E. coli* on a plasmid-based library of yeast genomic DNA. Transposon-mutagenized yeast DNA is subsequently transformed into *S. cerevisiae*; by homologous recombination, the mutagenized DNA will integrate at its corresponding chromosomal locus, thereby replacing genomic copy. **(b)** Transposon-encoded *lacZ* serves as a reporter of gene expression; *lacZ* fusions may be detected by a simple filter-based assay for β-galactosidase activity. **(c)** Transposon insertion generates a truncation of the host gene, thereby generating disruption alleles for phenotypic analysis. This analysis is amenable to high-throughput methods, as specific disruption alleles may be transformed in 96-well format into any desired yeast background for subsequent growth under a variety of test conditions. One such array of haploid transformants in an *ade2* background is shown. **(d)** Finally, Cre-*lox* recombination may be used to reduce this transposon construct in yeast to sequence encoding a 93-codon tag containing three copies of the HA epitope (HAT tag). Corresponding HAT-tagged proteins may be localized by immunocytology with anti-HA antibodies. All accumulated data is subsequently integrated into an on-line database accessible from the URL http://ygac.med.yale.edu.

Figure 2



Bar-coded gene deletions in yeast. **(a)** Per deletion, two oligonucleotide primers are synthesized, each containing sequence complementary to the region upstream or downstream of the targeted yeast ORF (including start and stop codons) as well as an internal 20 bp 'bar code' (a unique DNA sequence tag serving as a strain identifier). Each oligonucleotide contains a unique bar code sequence; therefore, two bar codes are ultimately incorporated into each deletion strain. The bar-coded primers also contain a 3′ region complementary to an antibiotic-resistance cassette (encoding resistance to kanamycin). PCR amplification subsequently generates a deletion cassette carrying the Kan^R marker flanked by two unique bar codes internal to yeast genomic DNA sequence upstream and downstream of the targeted ORF. **(b)** This deletion cassette is subsequently reamplified using 45 bp primers complementary to genomic DNA upstream and downstream of the targeted ORF. This second round of PCR ensures that each end of the deletion cassette carries a sufficiently large region of yeast sequence identical to that found flanking the target gene. **(c)** PCR-amplified deletion cassettes are transformed into yeast; by homologous recombination, the deletion cassette integrates at its targeted site, generating a precise deletion of the desired ORF.

EUROFAN, has disrupted more than 2000 protein coding genes using this approach [27]. The disrupted strains are being distributed to several laboratories to assess gene expression and define phenotypes by carrying out several hundred assays in parallel; strains are being assayed for hypersensitivity or resistance to various drugs under a number of different stress conditions [28,29]. Information on disruption phenotypes and gene expression generated from this EUROFAN study can be found at http://www.mips.biochem.mpg.de/proj/eurofan/index.htm.

Recently a variation of this approach has been used by a consortium of laboratories in North America and Europe to disrupt ~85% of the 6200 yeast genes [30••]. The PCR products in addition to containing flanking yeast gene sequences also contain a 'bar code', a unique 20 bp sequence that can be used to identify the insertion allele [31]. The bar code is amplified by PCR using primers that flank the bar code. The PCR product is hybridized to a microarray containing sequences complementary to both strands of the bar code. Typically, two unique bar codes are incorporated within each PCR product (Figure 2). Using this approach, it is possible in principle to simultaneously measure the presence of all 6200 deletions in a single experiment. The strains are mixed and subjected to selection. The bar codes are then amplified and hybridized to the arrays [30••,32•].

As in the case of transposon-mediated shuttle mutagenesis, PCR-based approaches for the construction of genome-wide libraries are only possible in those organisms with a high rate of homologous recombination. Also, the investment in time and labor required to construct such a PCR-based collection is beyond the scope of most individual laboratories; instead, genome-wide targeted mutagenesis is best accomplished by a consortium of laboratories operating in concert.

Mycoplasma

Whether generated by targeted or random approaches, genome-wide mutant collections constitute a valuable resource for a variety of functional studies. For example, Hutchison *et al.* [33••] have recently utilized random transposon mutagenesis to generate a collection of viable *Mycoplasma* mutants as a means of identifying the minimum complement of genes essential for cellular life. *Mycoplasma genitalium* encodes the smallest gene complement of any known independently replicating cell; it had, therefore, been predicted to possess a nearly minimal genome required for growth under standard laboratory conditions. By global transposon mutagenesis, however, Hutchison *et al.* identified 243 *Mycoplasma* genes whose functions are not essential for cell viability. From this analysis (coupled with statistical and computational analysis), the authors conclude that 265–350 of the 517 genes in *M. genitalium* are essential under laboratory growth conditions.

While exemplifying the virtues of transposon mutagenesis as a rapid and cost-effective means of analyzing gene function on a genomic scale, this study illustrates its inherent limitations as well. Transposition is not a truly random process; biases in transposon target site selection complicate interpretation of these findings — particularly as essential genes were defined in this study, in part, by an absence of viable mutants containing a transposon insertion within a given coding region. Moreover, not all transposon insertions can be expected to disrupt gene function using a non-targeted approach.

Plants

With genomic sequence data from the mustard plant *Arabidopsis thaliana* accumulating at a rapid pace [34], plant geneticists are implementing several approaches applicable to the generation of genome-wide mutant libraries and subsequent analysis of gene function. Using a transposon-based approach, researchers at Pioneer Hi-Bred are mutagenizing the maize genome by random insertion of the endogenous transposable element *Mutator*, thereby generating from fertile maize plants a collection of seed stocks that possess potentially desirable gene disruptions [12•]. This adaptation of transposon mutagenesis, however, is subject to the same drawbacks discussed above: insertional bias and an inability to detect many disruption phenotypes.

In contrast to this random approach, plant biologists are also utilizing a novel site-specific method of generating gene disruptions with tailor-made chimeric oligonucleotides [35••]. These self-complementary hybrid oligonucleotides comprised DNA and modified RNA containing at least a single base mismatch to the endogenous target gene sequence. It has been hypothesized that this mismatch is recognized by DNA repair enzymes, resulting in the stable incorporation of a given site-specific mutation into the endogenous gene. This approach has already been used successfully to introduce single base substitutions and frame-shift mutations in *Arabidopsis*; its potential applicability on a genome-wide scale is currently being explored [35••].

Also of note, viral vectors, such as the tobacco mosaic virus (TMV), have recently been used as an effective means of introducing DNA into plant cells. Cloned genes may be introduced into TMV on a large-scale in order to generate libraries of TMV-infected tobacco plants for subsequent phenotypic analysis. In addition, TMV vectors have been used in a promising strategy known as virus-induced gene silencing (VIGS) [36•]. In VIGS, exons from a target gene are cloned into TMV, which is then used to infect a host tobacco plant. For reasons not yet understood, the exogenous RNA product induces the decay (silencing) of the host mRNA, inactivating production of the corresponding target protein as well as, in most cases, its homologs. As a result, VIGS may be useful in circumventing problems of gene redundancy encountered in studying members of large eukaryotic gene families. Although VIGS holds potential as a powerful genomic methodology, its utility may be limited; TMV is used to infect only full grown plants, potentially rendering it uninformative as a means of studying early plant development.

D. melanogaster

Drosophila has long served as an informative model for the study of genetics and development; not surprisingly, extensive projects are underway to mutate each of the 12,000 or so predicted genes encoded in the 165-Mb *Drosophila melanogaster* genome [5,37]. By random insertional mutagenesis using the transposable *P*-element, two groups — the Berkeley *Drosophila* Genome Project (BDGP) and the European *Drosophila* Genome Project (EDGP) — are constructing a comprehensive genome-wide library of mutant strains, each strain containing a single genetically engineered *P*-element that disrupts a unique ORF [37,38]. This *P*-element carries an enhancer trap — a reporter gene equipped with a minimal promoter that is expressed when inserted near an enhancer sequence. As expression of the reporter will mimic expression of the host gene, these *P*-elements are useful in studying patterns of gene expression.

At present, the BDGP has generated a library of 1045 strains containing disruptions in more than 900 genes required for adult viability [14••]. Progress towards completion of this collection has been hindered by an insertional bias associated with *P*-element transposition. Further complicating matters, only one-third of all *Drosophila* genes generate readily detectable phenotypes upon disruption. To address these obstacles, the BDGP now plans to identify disruptive *P*-element insertions by precisely localizing transposon insertion sites in all mutated strains through DNA sequencing (rather than solely by phenotype). This molecular mapping strategy is likely to be implemented using a new set of *P*-element misexpression vectors capable of disrupting genes as well as potentially generating informative gain-of-expression alleles [39].

C. elegans

With its recently sequenced genome, the nematode *C. elegans* is an increasingly attractive model for functional genomics. *C. elegans* can be grown in liquid cultures and is viable when frozen — characteristics rendering it amenable to a methodology involving the use of microtiter plates and consequent high-throughput techniques. Recently, Liu *et al.* [40••] devised an interesting means of generating genome-wide deletion mutants in *C. elegans* using random chemical mutagenesis coupled with PCR screening in 96-well microtiter plates. Specifically, the authors employ a series of chemical agents to create mutagenized nematode populations subsequently arrayed directly into microtiter plates. Genomic DNA is prepared from each worm population; 96 genomic DNA samples are pooled together into a microtiter plate. DNA samples from multiple plates are arrayed into a new microtiter plate such that each well corresponds to pooled DNA from a single original genomic DNA plate. This microtiter plate of pooled DNA is screened for specific deletion amplicons by nested PCR. The original genomic DNA plate carrying a candidate deletion is subsequently screened by PCR in order to identify a specific microtiter well housing the candidate deletion. Finally, corresponding frozen worms are subjected to PCR analysis as a means of identifying heterozygous and homozygous mutant strains. In theory, this rapid and scalable method of gene disruption is applicable to the analysis of any organism amenable to manipulation in high-density arrays, provided sufficient sequence data and an appropriate mutagen are available. Such random

chemical mutagenesis, however, may be problematic: extraneous mutations generated in the background of a selected target gene may complicate analysis of resulting mutant phenotypes.

Potentially, other approaches may also be used to construct genome-wide mutant collections in *C. elegans*. In particular, random insertional mutagenesis by the endogenous transposon Tc-1 [13] and gene silencing mediated by RNAi [41] are promising methodologies for genome-wide mutagenesis. Accordingly, a consortium has recently been formed to disrupt all genes in *C. elegans* by RNAi/PCR-based mutagenesis.

Mus musculus

Mouse embryonic stem (ES) cells constitute a powerful mammalian system for the potential generation of genome-wide mutant collections by tagged random mutagenesis. For example, researchers at Lexicon Genetics [42••] have used random insertional mutagenesis with gene trap vectors as a means of constructing a library of ES cells from which sequence-tagged mutations in over 2000 genes have already been identified. Gene trapping — random mutagenesis with a truncated reporter or marker expressed only upon insertion into a transcriptionally active gene — is a particularly effective method of insertional mutagenesis within genomes possessing a large fraction of non-coding DNA [43]. For this project, Zambrowicz *et al.* [42••] designed novel gene trap vectors such that a gene need not be transcribed in order to be trapped; this expression-independent method of sequence acquisition should facilitate the generation of a more comprehensive set of ES disruption clones. Even when implemented with improved vectors, however, gene trapping is still subject to target site bias affected by vector design, target gene size, and gene accessibility [42••].

Conclusions

Genome-wide mutant collections are fast becoming a concrete reality; in organisms ranging from *Mycoplasma* to *Mus musculus*, extensive projects are underway to generate comprehensive libraries of mutagenized genes as a means of investigating gene function on a genomic scale. Although the construction of these collections — whether by targeted or random mutagenesis — represents a significant feat in itself, the profound impact of such a library rests in its utility as a laboratory reagent. Traditional genetic studies have followed a paradigm whereby time-consuming approaches such as complementation analysis or genetic mapping have typically been necessary to identify an affected gene in a mutant of interest. In contrast, researchers will now be able to immediately identify affected genes within clones of interest — either by directly screening a collection of defined mutants or by conveniently recovering DNA flanking the site of mutation within a given clone. As a result, these genome-wide mutant libraries promise to dramatically expedite processes of gene analysis in the immediate future.

Acknowledgements

PSR Coelho was supported by the Fundacao de Amparo a Pesquisa do Estado de Sao Paulo, Brazil. A Kumar is supported by a postdoctoral fellowship from the American Cancer Society. Yeast genome research in the Snyder laboratory was supported by National Institutes of Health grant CA77808.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Mewes HW, Albermann K, Bahr M, Frishman D, Gleissner A, Hani J, Heumann K, Kleine K, Malieri A, Oliver SG *et al.*: **Overview of the yeast genome.** *Nature* 1997, **387**:7-65.

2. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L: **New goals for the US Human Genome Project: 1998–2003.** *Science* 1998, **282**:682-689.

3. Johnston M: **Towards a complete understanding of how a simple cell works.** *Trends Genet* 1996, **12**:242-243.

4. Hieter P, Boguski, M: **Functional genomics: it's all how you read it.** *Science* 1997, **278**:601-602.

5. Miklos GL, Rubin GM: **The role of the genome project in determining gene function: insights from model organisms.** *Cell* 1996, **86**:521-529.

6. Lawrence CW: **Classical mutagenesis techniques.** *Methods Enzymol* 1991, **194**:273-280.

7. Winzeler EA, Davis RW: **Functional analysis of the yeast genome.** *Curr Opin Genet Dev* 1997, **7**:771-776.

8. Casabadan MJ, Cohen SN: **Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: in vivo probe for transcriptional control sequences.** *Proc Natl Acad Sci* 1979, **76**:4530-4533.

9. Hensel M, Holden DW: **Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi.** *Microbiology* 1996, **142**:1049-1058.

10. Burns N, Grimwade B, Ross-Macdonald PB, Choi E-Y, Finberg K, Roeder GS, Snyder M: **Large-scale characterization of gene expression, protein localization and gene disruption in *Saccharomyces cerevisiae*.** *Genes Dev* 1994, **8**:1087-1105.

11. Kempken F, Kuck U: **Transposons in filamentous fungi — facts and perspectives.** *Bioessays* 1998, **20**:652-659.

12. • Martienssen RA: **Functional genomics: probing plant gene function and expression with transposons.** *Proc Natl Acad Sci USA* 1998, **95**:2021-2026.
An informative review of transposon-based approaches to functional genomics in *Arabidopsis*, maize, and other plant species with particular emphasis placed upon methods to screen mutant libraries by site-selected mutagenesis and systematic insertion site sequencing.

13. Zwaal RR, Broeks A, Meurs JV, Groenen JT, Plasterk RH: **Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank.** *Proc Natl Acad Sci USA* 1993, **90**:7431-7435.

14. •• Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, Misra S, Rubin GM: **The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes.** *Genetics* 1999, **153**:135-177.
An update as to the progress of the BDGP, which currently reports a collection of 1045 mutant strains containing *P*-element insertions disrupting more than 900 *Drosophila* genes essential for adult viability.

15. Smith V, Chou KN, Lashkari D, Botstein D, Brown PO: **Functional analysis of the genes of yeast chromosome V by genetic footprinting.** *Science* 1996, **274**:2069-2074.

16. Smith V, Botstein D, Brown PO: **Genetic footprinting: a genomic strategy for determining a gene's function given its sequence.** *Proc Natl Acad Sci USA* 1995, **92**:6479-6483.

17. Hoekstra MF, Burbee D, Singer J, Mull E, Chiao E, Heffron F: **A Tn3 derivative that can be used to make short in-frame insertions within genes.** *Proc Natl Acad Sci USA* 1991, **88**:5457-5461.

18. Ross-Macdonald P, Sheehan A, Roeder GS, Snyder M: **A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae.*** *Proc Natl Acad Sci USA* 1997, **94**:190-195.

19. Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, Boeke JD: **Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences.** *Cell* 1993, **73**:1007-1018.

20. Tyers M, Tokiwa G, Nash R, Futcher B: **The Cln3-Cdc28 kinase complex of *S. cerevisiae* is regulated by proteolysis and phosphorylation.** *EMBO J* 1992, **11**:1773-1784.

21. •• Ross-MacDonald P, Coehlo PSR, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung K-H, Sheehan A, Symoniatis D, Umansky L: **Large-scale analysis of the yeast genome by transposon tagging and gene disruption.** *Nature* 1999, **402**:413-418.
This work represents an extensive application of Tn*3*-mediated shuttle mutagenesis towards functional genomics in yeast. The approach utilizes a multifunctional minitransposon enabling the genome-wide analysis of disruption phenotypes, gene expression, and protein localization – a data set currently unattainable through any other single genomic approach.

22. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW: **Characterization of the yeast transcriptome.** *Cell* 1997, **88**:243-251.

23. Kumar A, Cheung K-H, Ross-MacDonald P, Coelho PSR, Miller P, Snyder M: **TRIPLES: a database of gene function in *S. cerevisiae*.** *Nucleic Acids Res* 2000, **28**:81-84.

24. Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C: **A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae.*** *Nucleic Acids Res* 1993, **21**:3329-3330.

25. Wach A, Brachat A, Pohlmann R, Philippsen P: **New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae.*** *Yeast* 1994, **10**:1793-1808.

26. Wach A, Brachat A, Alberti-Segui C, Rebischung C, Philippsen P: **Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae.*** *Yeast* 1997, **13**:1065-1075.

27. Dujon B: **European functional analysis network (EUROFAN) and the functional analysis of *Saccharomyces cerevisiae*.** *Electrophoresis* 1998, **19**:617-624.

28. Rieger KJ, El-Alama M, Stein G, Bradshaw C, Slonimski PP, Maundrell K: **Chemotyping of yeast mutants using robotics.** *Yeast* 1999, **15**:973-986.

29. Bianchi MM, Sartori G, Vandenbol M, Uccelletti D, Mazzoni C, DiRago JP, Carignani G, Slonimski PP, Frontali L: **How to bring orphan genes into functional families.** *Yeast* 1999, **15**:513-526.

30. •• Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H *et al.*: **Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis.** *Science* 1999, **285**:901-906.
This paper reports the progress of an international consortium of laboratories utilizing targeted PCR-based gene deletions to disrupt each open reading frame (ORF) within the yeast genome. At present, the authors report deletion of 2026 ORFs and subsequent phenotypic analysis of 500 deletion strains.

31. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW: **Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.** *Nat Genet* 1996, **14**:450-456.

32. • Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW: **Genomic profiling of drug sensitivities via induced haploinsufficiency.** *Nat Genet* 1999, **21**:278-283.
An interesting genomic approach to drug target identification in which diploid yeast strains made heterozygous for a single gene are screened for drug-induced sensitivity (haploinsufficiency) on high-density oligonucleotide arrays.

33. •• Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC: **Global transposon mutagenesis and a minimal Mycoplasma genome.** *Science* 1999, **286**:2165-2169.
This paper describes the application of genome-wide mutagenesis by Tn*4001* to the genomes of *M. genitalium* and *M. pneumoniae* as a means of identifying a minimal set of essential genes under laboratory growth conditions.

34. Meyerowitz EM: **Today we have naming of parts.** *Nature* 1999, **402**:731-732.

35. •• Beetham PR, Kipp PB, Sawycky XL, Arntzen CJ, May GD: **A tool for functional plant genomics: Chimeric RNA/DNA oligonucleotides cause in vivo gene-specific mutations.** *Proc Natl Acad Sci USA* 1999, **96**:8774-8778.
A newly developed technique with genomic potential – self-complementary DNA/RNA chimeric oligonucleotides are used to induce stable, site-specific base substitutions and frameshift mutations in plant cells *in vivo*.

36. • Baulcombe DC: **Fast forward genetics based on virus-induced gene silencing.** *Curr Opin Plant Biol* 1999, **2**:109-113.
An interesting review of virus-induced gene silencing – a promising technique for plant functional genomics in which a target gene cloned into tobacco mosaic virus induces the silencing of corresponding endogenous mRNA.

37. Rubin GM: **The *Drosophila* genome project: a progress report.** *Trends Genet* 1998, **14**:340-343.

38. Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM: **Gene disruptions using *P* transposable elements: An integral component of the Drosophila genome project.** *Proc Natl Acad Sci USA* 1995, **92**:10824-10830.

39. Rorth P: **A modular misexpression screen in Drosophila detecting tissue-specific phenotypes.** *Proc Natl Acad Sci USA* 1996, **93**:12418-12422.

40. •• Liu LX, Spoerke JM, Mulligan EL, Chen J, Reardon B, Westlund B, Sun L, Abel K, Armstrong B, Hardiman G *et al.*: **High-throughput isolation of *Caenorhabditis elegans* deletion mutants.** *Genome Res* 1999, **9**:859-867.
An application of random chemical mutagenesis to high-throughput genome-wide functional analysis. Chemically mutagenized nematode populations are screened by PCR for deletion mutants in microtiter plate arrays; manipulation of whole animals and genomic DNA in these arrays may provide a blueprint for similar mutagenic approaches in microbial genomes as well.

41. Montgomery MK, Fire A: **Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression.** *Trends Genet* 1998, **14**:255-258.

42. •• Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT: **Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells.** *Nature* 1998, **392**:608-611.
An extensive application of expression-independent gene trapping allowing for the automated identification of sequence tags from insertionally mutagenized genes.

43. Evans MJ, Carlton MBL, Russ AP: **Gene trapping and functional genomics.** *Trends Genet* 1997, **13**:370-374.