EMERGING TECHNOLOGIES IN YEAST GENOMICS

Anuj Kumar and Michael Snyder

The genomic revolution is undeniable: in the past year alone, the term 'genomics' was found in nearly 500 research articles, and at least 6 journals are devoted solely to genomic biology. More than just a buzzword, molecular biology has genuinely embraced genomics (the systematic, large-scale study of genomes and their functions). With its facile genetics, the budding yeast *Saccharomyces cerevisiae* has emerged as an important model organism in the development of many current genomic methodologies. These techniques have greatly influenced the manner in which biology is studied in yeast and in other organisms. In this review, we summarize the most promising technologies in yeast genomics.

FUNCTIONAL GENOMICS
The development and systematic application of experimental methodologies to analyse gene function on a genome-wide scale.

Department of Molecular, Cellular and Developmental Biology, Yale University, PO Box 208103, New Haven, Connecticut 06520-8103, USA. Correspondence to M.S. e-mail: michael.snyder@yale.edu

The genomic era began in earnest during the late 1980s. Catalysed by advances in DNA-sequencing technology, early genomic studies were principally large-scale projects to completely sequence whole genomes. These DNA-sequencing projects, although labour intensive, were a profound success. In 1995, a team headed by Craig Venter published the complete 1,830-kilobase (kb) genomic sequence of the bacterium *Haemophilus influenzae*[1]. In 1996, an international consortium of 600 scientists released the complete 12,000-kb sequence of the simple eukaryote *Saccharomyces cerevisiae*[2]. In the years since, whole-genome sequencing projects have exploded in number such that, at present, at least 44 complete genome sequences are publicly available, with nearly 800 sequencing projects now underway (see link to NCBI's Entrez Genome site).

These projects have provided a wealth of gene sequences; however, the challenge ahead lies in understanding gene function and the manner in which genes are regulated. Furthermore, data sets from large numbers of genes need to be integrated to determine effectively how networks of genes carry out biological processes. At present, innovative experimental methods are being developed to address these needs on a genome-wide scale. The continued development of these methods is essential to the blossoming field of FUNCTIONAL GENOMICS[3].

Yeast as a genomic model

In its brief history, functional genomics has benefited greatly from analysis of the baker's yeast *Saccharomyces cerevisiae*. Long recognized as an informative model organism in traditional genetic studies, *S. cerevisiae* also presents an ideal model genome for large-scale functional analysis. Relative to other eukaryotes, *S. cerevisiae* has a compact genome: ~70% of its total (non-ribosomal DNA) genetic complement is protein-coding sequence. Encompassing 16 chromosomes, the 12-megabase (Mb) yeast genome is predicted to encode ~6,200 genes, with 1 gene per 2 kb of genomic sequence[2]. The genes of higher eukaryotes typically contain introns; however, only 263 of yeast genes do[4], which simplifies the process of computer-based gene identification. Experimental manipulation of yeast is equally straightforward. *S. cerevisiae* can be grown easily in the laboratory and is stable in both the haploid and diploid state — an advantage for studying recessive mutations and for characterizing gene function. Moreover, transforming DNA tends to integrate in yeast by homologous recombination, greatly facilitating gene cloning and reverse genetics. Importantly, *S. cerevisiae* is an informative predictor of human gene function; nearly 50% of human genes implicated in heritable diseases have yeast homologues[5–8].

Despite the popularity of yeast as a model system, only one-third of all predicted yeast genes had been functionally characterized when the complete

© 2001 Macmillan Magazines Ltd

sequence of the yeast genome first became available[9]. At present, 3,780 yeast genes have been characterized by genetic or biochemical means, and an additional 560 yeast genes have homologues in other organisms, which provides some indication of their functions. However, ~1,900 yeast genes still encode proteins of unknown function[4]. There exists, therefore, a need to better characterize gene function in *S. cerevisiae*. Genomic techniques provide a route by which this might be accomplished in a systematic, high-throughput manner. In this review, we highlight emerging genomic technologies in yeast — nascent approaches that hold tremendous promise as future tools for biological discovery in yeast and in other organisms.

Transposon tagging and mutagenesis

Transposable elements have long been used in prokaryotes and eukaryotes as mutagenic agents in traditional genetic studies[10–12]; however, TRANSPOSONS constitute an equally powerful tool for functional genomics[13,14]. Transposons can be engineered to carry various reporter genes, EPITOPE tags and regulatory elements, thereby serving as multipurpose tools by which gene function might be investigated on a genome-wide scale.

For use in a genome-wide study of gene expression, gene disruption and protein localization in yeast, we constructed a series of multifunctional transposons[15,16] (FIG. 1a), which were engineered from either the bacterial transposons Tn*3* (REF. 17) or Tn*7* (REF. 18). Each transposon was modified to carry a reporter gene (typically a modified form of *lacZ* that lacks both its start codon and promoter — the expression of which is dependent on the transposon being introduced in-frame into a transcribed and translated region of the genome). Transposons were also modified to contain a *lox* site near each end; *lox* sequences are targets of the site-specific DNA recombinase, Cre. One *lox* site is positioned adjacent and internal to the sequence that encodes three tandem copies of an epitope from the influenza virus haemagglutinin protein (the HA epitope). By inducing Cre expression, the 6-kb transposon can be reduced to a 93-codon read-through insertion element that encodes three copies of the HA epitope[16].

Transposons are introduced into the yeast genome by SHUTTLE MUTAGENESIS[19] (FIG. 1b). A plasmid library of yeast genomic DNA is initially mutagenized by transposon insertion, and insertion alleles are subsequently introduced into a diploid strain of yeast by DNA transformation. By homologous recombination, each insert integrates at its corresponding genomic locus, thereby replacing its chromosomal copy. Transposon insertion in coding sequence generates a *lacZ* gene fusion; *lacZ* acts as a marker of gene expression and as a gene trap, which can be used to isolate novel coding sequences. Transposon insertion also truncates the host genes, generating disruption alleles for phenotypic analysis. In yeast strains that show β-galactosidase activity (which is encoded by *lacZ*), Cre-*lox* recombination can be used to generate proteins that

TRANSPOSON
Mobile DNA elements that can relocate within the genome of their hosts; transposons can be used for various applications, including insertional mutagenesis, gene identification, gene tagging and DNA sequencing.

EPITOPE
Part of a protein (antigen) that combines with the antigen-binding site of an antibody; the incorporation of an epitope-encoding sequence into a target gene is called epitope-tagging.

SHUTTLE MUTAGENESIS
A method in which cloned yeast genes are mutated by bacterial transposition in *Escherichia coli*; mutant alleles are subsequently introduced ('shuttled') into yeast where they integrate at their corresponding genomic loci by homologous recombination.

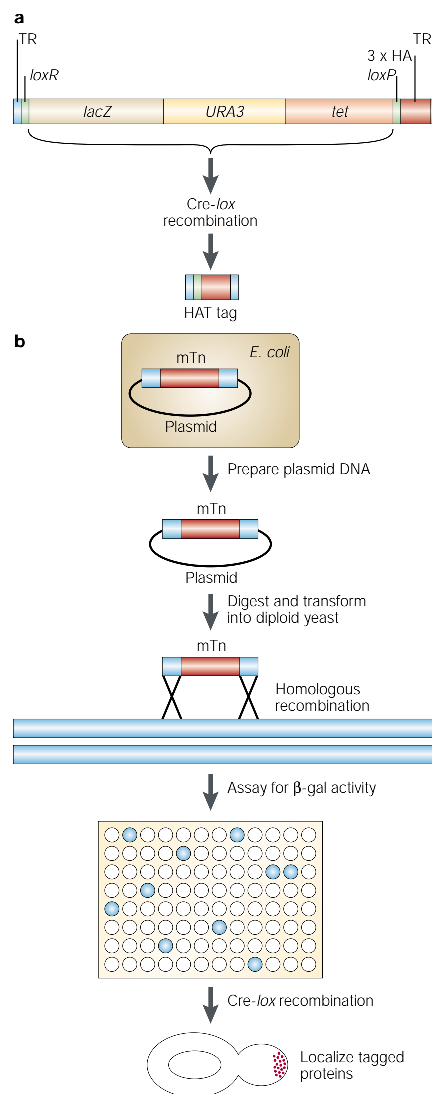

Figure 1 | **Multipurpose transposons. a** | Diagram of a full-length 6-kb multipurpose transposon (not drawn to scale), which can be reduced by Cre-*lox* recombination to a 93-codon haemagglutinin epitope tag (HAT tag). Constructs are annotated as follows: TR, transposon terminal repeat; *loxR*, *loxP*, target sites for Cre recombinase; 3 × HA, sequence that encodes three tandem copies of the HA epitope; *lacZ*, reporter gene; *URA3*, yeast selectable marker; *tet*, bacterial selectable marker. **b** | Shuttle mutagenesis with multipurpose transposons. All procedures are done in 96-well format. Transposon (mTn) insertions are generated in *Escherichia coli* and are then introduced into yeast by DNA transformation. Cre-*lox* recombination is induced in yeast strains that express β-galactosidase (β-gal) to generate epitope-tagged proteins for various studies, such as immunolocalization, as shown.

© 2001 Macmillan Magazines Ltd

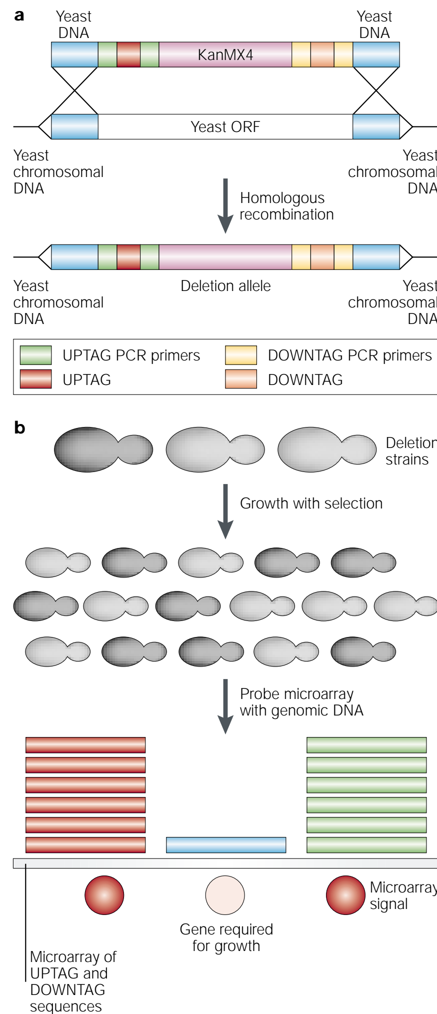

Figure 2 | **PCR-based gene-deletion strategy. a** | Deletion cassettes are generated by PCR such that each cassette is flanked by two 45-bp regions of yeast DNA sequence that correspond to the intended deletion target. These short regions of homology direct the integration of the deletion cassette to its intended genomic locus, resulting in a precise start-to-stop codon gene replacement. (ORF, open reading frame); KanMX4, marker that confers resistance to the antibiotic geneticin (G418).) **b** | A population of barcoded deletion strains can be assayed for fitness by comparative hybridization of their genomic DNA to a microarray of UPTAG and DOWNTAG sequences. Strains deleted for a gene that is required for growth under a given condition will be underrepresented after selection, yielding lower intensity signals upon hybridization to the microarray. As individual barcodes can show differing hybridization intensities, microarray signals do not reflect the absolute levels of a given mutant within a pool; instead, the fitness of a mutant is determined by relative changes in signal intensity after growth under selection (the analysis of a single time point is shown here).

CONDITIONAL MUTATIONS
Mutations that generate an observable mutant phenotype under a given set of growth conditions (restrictive conditions), but no mutant phenotype (or a reduced phenotype) under a separate set of conditions (permissive conditions).

ESSENTIAL GENE
A gene that is indispensable for cell viability under defined growth conditions; complete loss of the function of an essential gene is lethal.

carry the transposon-encoded HA epitope. These epitope-tagged proteins might be immunolocalized or immunoprecipitated using antibodies directed against HA[16,20]. Furthermore, transposon-encoded epitope tags provide a genome-wide mechanism for generating CONDITIONAL MUTATIONS, which are particularly useful for studying ESSENTIAL GENES[15]. This approach is highly scalable[21]. By adopting most protocols to a 96-well format, we have carried out nearly 200,000 plasmid preparations and yeast transformations, which have generated a collection of ~20,000 strains that contain transposon insertions that affect ~50% of annotated yeast genes[16].

As illustrated in this approach, genomic studies benefit from the ease with which transposons can randomly generate large numbers of marked insertions in a population of target DNA, while generating a more diverse array of alleles than can be obtained by chemical mutagenesis or ultraviolet irradiation. Transposon tagging has also led to the discovery of over 350 previously unidentified yeast gene-coding sequences by gene-trap analysis. Additionally, this strategy provides an effective means of uncovering new gene functions not identified by other techniques. In a recent study of 31 meiotic genes identified by transposon tagging[16], only two-thirds were similarly identified from a microarray-based expression profile of the yeast genome during sporulation[22]. Transposon-based shuttle mutagenesis can also be applied to other organisms (such as to mouse embryonic stem cells and species of *Candida*), provided they have a tractable genetic system in which DNA tends to integrate by homologous recombination. Furthermore, this particular methodology generates plasmid-borne insertion alleles and other reagents of widespread use to the yeast community[23]. However, transposon-based mutagenesis has some limitations. Insertions are essentially generated at random; therefore, it is very difficult to mutagenize all genes within a genome by transposon mutagenesis alone. Furthermore, these technologies are sensitive to transposon-specific biases in target-site selection: for reasons not fully understood, transposons such as Tn*3* insert non-randomly into certain regions of target DNA[24].

Genome-wide gene deletions

Considering the above caveats, transposon mutagenesis might be best complemented by methods of directed mutagenesis. Adopting a PCR-based approach, an international consortium has undertaken a continuing project to delete systematically each annotated gene in the yeast genome[25]. This gene-disruption approach, modified from a previous strategy[26], is based on the high rate with which DNA integrates by homologous recombination in yeast. Using two rounds of PCR amplification, an individual 'deletion cassette' is constructed for each annotated gene (FIG. 2a). Each end of this PCR product contains 45 bp of sequence that is identical to the region upstream and downstream of the targeted gene. Upon its introduction into yeast, this short region of homology is sufficient to direct the cassette to its corresponding genomic locus, which results in a precise start-to-stop codon gene replacement.

© 2001 Macmillan Magazines Ltd

Box 1 | **DNA microarrays**

High-density arrays of ordered DNA can be generated by several methods. Short oligonucleotides (typically 25 nucleotides in length) can be synthesized *in situ* on a glass substrate using photolithographic methods developed by Affymetrix, Inc.[63] In this approach, hundreds of thousands of synthetic oligonucleotides can be arrayed in a 1.3 cm^2 area. Arrays of longer oligonucleotides can also be synthesized *in situ* using proprietary ink-jet technology developed at Rosetta Inpharmatics[64]. Alternatively, double-stranded DNA molecules (typically generated by PCR amplification) can be printed robotically onto glass slides or nylon membranes[33], such that ~10,000 DNA samples might be arrayed in a 2 cm^2 area.

DNA microarrays have classically been used in comparative hybridization-based assays to measure relative levels of target DNA in two experimentally derived populations of nucleic acid. Typically, each population of DNA, cDNA, or mRNA is labelled with a different fluorescent dye (usually either Cy3 or Cy5). Both labelled sample populations are co-hybridized to the DNA array; subsequently, the fluorescence intensity of each 'spot' on the array is measured by confocal laser-scanning microscopy. The ratio of Cy5 to Cy3 signal per spot provides a quantitative indication of the relative abundance of each target sequence in the two sample populations. In this way, genes expressed differentially under two growth conditions can be identified by comparative hybridization of cDNA probes derived from cells grown under each condition.

The PCR primers used in this project have been designed to incorporate the following elements into each deletion cassette[25]. All cassettes carry the KanMX4 marker, which confers resistance to the antibiotic geneticin (G418). This marker is flanked by unique 20-bp sequences that are not present in the yeast genome; each cassette has two such sequence tags that are specific to that particular cassette. These unique tags (called UPTAG and DOWNTAG) serve as strain identifiers, or 'molecular barcodes', by which a given deletion strain can be identified in a mixed population. The incorporation of two barcodes in each deletion cassette increases the likelihood that each deletion strain can be identified, even if a single tag carries a mutation. Each barcode is itself flanked by common 18-bp sequences, which serve as PCR-priming sites, such that a single pair of primers can amplify the UPTAG from every deletion, and a separate single pair of primers can amplify the DOWNTAG. The UPTAG and DOWNTAG sequences amplified from the genomic DNA can be hybridized to a microarray of complementary oligonucleotides, thereby providing a method by which individual deletion strains might be quantitatively analysed in parallel[27] (FIG. 2b).

'Barcoded' deletion strains exemplify the advantages of a targeted approach to genome-wide mutagenesis: gene deletions are as precise as possible and, therefore, deletion mutants should be null for each mutagenized gene. Additionally, the barcodes provide a convenient way to analyse the deletion mutants in parallel within a pooled population. However, the use of this collection is limited by several factors. A recent study of nearly 300 yeast deletion strains indicated that ~8% of these mutants were aneuploid for whole chromosomes or for chromosomal segments[28]. As this aneuploidy might mask phenotypes associated with deleterious gene deletions, phenotypic studies of the deletion strain collection will need to be correlated with analysis of chromosome content and gene dosage. More generally, this PCR-based strategy is dependent on the availability of correctly annotated gene sequence and is, therefore, best applied to organisms with a fully sequenced genome. Furthermore, in the above method, four long, gene-specific primers are required to generate each deletion cassette; four additional gene-specific primers are required to verify correct genomic integration of each cassette. The cost in money and labour to synthesize and use these primers on a large scale is prohibitively expensive to many researchers. At present, these deletions are provided in a single genetic background; this allows independent groups to integrate the results of different experiments, but most genes are best analysed in several strains. By cloning each deletion cassette and making this reagent publicly available, researchers could easily delete a given gene from a strain of their choice, thereby rendering this collection more informative.

Nonetheless, these deletion strains are extremely useful for characterizing disruption phenotypes[25]; in addition, several studies have used targeted gene deletions to characterize drug responses and drug targets on a genome-wide scale. Deletion mutants have been used to investigate the underlying genes and pathways that mediate yeast response to the immunosuppressive antibiotic rapamycin[29]. Rapamycin is known to inhibit two redundant proteins, the target of rapamycin (Tor) proteins, Tor1 and Tor2; TOR genes encode protein kinases that are implicated in various processes normally induced by nutrient starvation (such as G_1 cell-cycle arrest, glycogen accumulation, reduced protein synthesis and sporulation). To characterize genes that are involved in TOR-mediated signalling pathways, 2,216 haploid deletion strains and 50 HETEROZYGOUS DIPLOID deletion strains (each deleted for one of two copies of an essential gene) were individually screened for hypersensitivity or resistance to rapamycin. In total, this screen identified 106 genes involved in nutrient-dependent functions, transcriptional regulation, vacuolar biogenesis, ubiquitin-dependent proteolysis and microtubule-related functions. Heterozygous diploid deletion strains were also used in a study[30] to assay drug-induced HAPLOINSUFFICIENCY on a genome-wide scale in yeast. Using this approach, barcoded heterozygous deletion strains were pooled and grown competitively in sub-lethal concentrations of a given drug. Deletion strains were then quantitatively analysed in parallel by hybridization to an oligonucleotide array of UPTAG and DOWNTAG sequences; strains showing reduced fitness were heterozygous at loci that encode putative drug targets. In an analysis of 233 heterozygous diploid strains treated with the drug tunicamycin[30], this technique successfully identified one known target of tunicamycin and two hypersensitive loci, validating the use of genomics in drug-target identification and drug discovery.

HETEROZYGOUS DIPLOID
A diploid yeast cell with different alleles at a particular locus. Heterozygous diploid cells can be used to ascribe cellular functions to essential genes.

HAPLOINSUFFICIENCY
When loss of function of one gene copy leads to a mutant phenotype.

Expression profiling with DNA microarrays

Since its development in the mid-1990s[31,32], the DNA microarray has emerged as the pre-eminent tool for functional genomics. The ability to analyse thousands of DNA samples simultaneously by hybridization-based assay (BOX 1) has provided a popular method for analysing the relative levels of mRNA transcripts on a genome-wide scale. Although an exhaustive account of

© 2001 Macmillan Magazines Ltd

Box 2 | **Microarray data analysis**

Microarray studies generate tens of thousands of data points from a single experiment, which presents an imposing volume of results for subsequent interpretation and analysis. To render these results informative, microarray expression data is mined using one of several clustering algorithms designed to identify sets of genes that show similar expression patterns. For this purpose, Eisen *et al.*[65] have applied a hierarchical clustering method in which relationships between genes are presented graphically within a single tree, the branch lengths of which reflect the degree of similarity between individual expression profiles. Hierarchical clustering, however, can lead to artefacts, which can be overcome by first partitioning data into relatively homogenous groups by using self-organizing maps (SOMs)[66] or *k*-means clustering[16,67]. Both SOMs and *k*-means clustering generate a pre-set number of clusters; in using these algorithms, researchers must estimate the number of significant patterns in a data set before clustering. To bypass this guesswork, Heyer *et al.*[68] have developed the quality clustering algorithm, a partitioning method that eliminates many problems associated with hierarchical clustering, *k*-means clustering and SOMs.

Although these clustering methods are important tools for data mining, more sensitive approaches are required to exploit microarray-based expression data fully. Recently, Brown *et al.*[69] have applied a supervised computer learning method to predict gene function from microarray expression data. Tavazoie *et al.*[70] have correlated gene expression with function by comparing *k*-means clustered sets of genes with functional categories curated by the Munich Information Centre for Protein Sequences (MIPS) database (see BOX 3); in this way, any statistically significant overlap between expression clusters and functional categories can be assessed. Similar methods to correlate expression data with pre-existing literature are extremely promising avenues by which gene function might be inferred from microarray data.

established microarray technology is beyond the scope of this review, we briefly discuss here several microarray-based studies that have contributed to our understanding of cellular processes in yeast. Typically, DNA microarrays have been used to identify genes, the expression of which is either induced or repressed during specific cellular responses. For example, DeRisi *et al.*[33] used DNA microarrays to monitor relative changes in mRNA levels during the shift from anaerobic fermentation to aerobic respiration in yeast. Microarrays have also been used to identify genes differentially expressed during sporulation[22], as well as genes periodically expressed during the cell cycle[34,35]. Jelinsky and Samson[36] used oligonucleotide arrays to identify over 400 genes that are either induced or repressed in response to the DNA-damaging, alkylating agent methyl methanesulphonate. These and other microarray-based studies have identified genes that putatively function in common regulatory pathways; such pathways are also being delineated by transcriptional profiling of strains mutated for key regulatory components. For example, 33 putative downstream targets of a developmental mitogen-activated protein kinase (MAPK) signalling pathway were identified through expression profiling of yeast strains that had been mutated for transcription factors known to act in this pathway[37]. By profiling strains in which key components of the yeast transcription-initiation machinery had been mutated, Holstege *et al.*[38] identified distinct sets of genes that were activated (directly or indirectly) by the RNA polymerase II holoenzyme, the general transcription factor TFIID, and the SAGA chromatin modification complex. In addition, the same study defined new functions for the proteins Srb5 and Srb10 in yeast mating and nutrient sufficiency.

To identify new gene function by transcriptional profiling, expression data must be sorted and analysed systematically. Typically, expression data is mined using 'clustering' algorithms (BOX 2) that are designed to group together genes that show similar patterns of expression (and presumably similar functions). Gene function can also be inferred from microarray data by comparing whole-genome expression profiles. These profiles provide a 'snapshot' of genes that are expressed in a given strain under a particular set of growth conditions, which can be diagnostic of that particular strain under such conditions. A function can be assigned to an uncharacterized gene if a strain deleted for that gene gives an expression profile identical to that generated by deleting a gene of known function.

Using this logic, Hughes *et al.*[39] initiated a pilot study in yeast, from which they developed a compendium of 300 expression profiles — each profile was generated from experiments in which transcript levels of a yeast deletion mutant or compound-treated culture were compared with transcript levels obtained from a wild-type or mock-treated culture. All experiments were done on strains that were derived from a single genetic background and that had been cultured under identical growth conditions. This expression profile compendium was compared with profiles generated from eight yeast mutants, each deleted for a gene of unknown function; from this comparison, functions were correctly assigned to all eight previously uncharacterized genes. This approach can also be used to identify drug targets. Yeast treated with the topical anaesthetic dyclonine showed an expression profile that closely resembled the profile shown by a strain deleted for *erg2*; *ERG2* functions in ergosterol biosynthesis in yeast and was confirmed as a target of dyclonine through further genetic analysis.

The feasibility of this compendium approach is based on several assumptions. The expression profile shown by a mutant must accurately predict the function of the mutated gene. Mutation of a given cellular pathway must yield a uniquely characteristic expression profile, and disruption of separate components in a single pathway must yield a similar transcriptional response. To facilitate comparisons, this technique also requires a large number of available expression profiles (each generated using mutants that were derived from a single background strain and grown under a single set of conditions). Despite these potential limitations, compendium-based profiling is promising — particularly as a means of identifying signalling pathway components[40] and unidentified drug targets both in yeast and in higher eukaryotes.

Mapping binding sites of chromosomal proteins

Microarray-based expression profiles can also be used to identify transcription-factor-binding sites, as these sequences will be statistically enriched upstream of genes upregulated in response to the activation of a certain transcription factor. For example, by computational analysis of motifs upstream of coregulated genes that are induced in response to DNA damage, Jelinsky *et al.*[41] identified a consensus regulatory sequence bound by Rpn4, a proteome-associated yeast protein. Although

© 2001 Macmillan Magazines Ltd

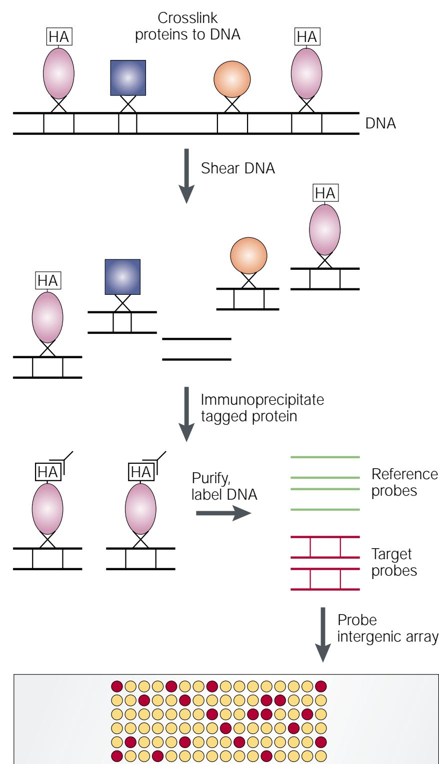

Figure 3 | **Genome-wide identification of protein–DNA interactions.** Protein–DNA interactions are 'captured' *in vivo* by crosslinking proteins to their genomic binding sites. Crosslinked DNA is subsequently extracted, sheared and purified by immunoprecipitation with antibodies directed against an epitope-tagged protein of interest (such as the influenza virus haemagglutinin protein (HA epitope)). Purified DNA fragments are subsequently amplified and fluorescently labelled for use as target probes; labelled reference probes (shown in green) are often prepared from a strain deleted for the protein of interest. Probes are co-hybridized to an array of intergenic regions. The ratio of target probe to reference probe at each array 'spot' provides an indication of the frequency with which each corresponding genomic locus is bound by the tagged protein. (See animation online.)

Animated online

this strategy can be used to identify certain regulatory motifs, transcription-factor-binding sites in yeast can be mapped more effectively on a genome-wide scale through an approach that combines DNA microarrays with chromatin immunoprecipitation (FIG. 3). In this approach, yeast cells are treated with formaldehyde to crosslink DNA-binding proteins to their target sites *in vivo*. Crosslinked DNA is extracted and sheared by sonication; DNA bound to a transcription factor of interest is subsequently purified by immunoprecipitation using antibodies directed against the chosen transcription factor. After reversal of the crosslinks, immunoprecipitated chromatin is amplified and fluorescently labelled by PCR. This labelled DNA serves as a probe; a differently labelled reference probe is also generated, ideally from a strain that has been deleted for the desired transcription factor. The two probes are co-hybridized to a DNA microarray that contains all intergenic regions present within the yeast genome; alternatively, as yeast intergenic regions are typically smaller than the immunoprecipitated chromatin fragments, a standard microarray of coding sequences can also be used. The ratio of fluorescence intensities measured at each element in the array is indicative of the relative enrichment of each target sequence after immunoprecipitation and, therefore, indicates the extent to which each site is bound *in vivo*. As described in REF. 42, this approach has been used to identify over 200 previously unidentified targets of the G1/S-specific transcriptional activators SBF (a heterodimer of Swi4 and Swi6) and MBF (a heterodimer of Mbp1 and Swi6). This approach has also been used to identify 10 galactose-induced targets and 29 pheromone-induced targets of the yeast transcriptional activators Gal4 and Ste12, respectively[43].

Along similar lines, Gerton *et al.*[44] have used DNA microarrays to map regions of the yeast genome that show unusually high and low levels of meiotic recombination. In yeast, most recombination events are initiated through meiosis-specific double-stranded breaks (DSBs), which are catalysed in part by the topoisomerase-II-related protein, Spo11. Meiotic DNA fragments that are enriched for Spo11-binding can be obtained using a mutant yeast strain in which Spo11 remains covalently attached to broken DNA ends. Presumably, these Spo11-associated fragments will lie adjacent to recombination hot spots; conversely, these fragments should be deficient in regions adjacent to recombination cold spots. As these 2–3-kb DSB-enriched fragments are sufficiently large to extend beyond intergenic regions into neighbouring genes, they could be used as hybridization probes against a microarray that contains all 6,200 predicted yeast gene-coding sequences. The resulting hybridization data can be used to identify genes that are located near meiotic recombination hot spots and cold spots. From this analysis, Gerton *et al.*[44] found hot spots nonrandomly associated with regions of high G/C base content and cold spots preferentially located near centromeres and telomeres.

As most DSBs are believed to occur outside gene-coding sequence[44,45], genomic analysis of meiotic recombination can be carried out more directly by using a microarray of yeast intergenic regions. The construction (and subsequent manipulation) of such arrays requires a significant initial investment of time and money; however, the potential benefits of microarray technology far outweigh this shortcoming. Limited only by the obvious need for an informative set of probes, DNA microarrays can be used to characterize component molecules in nearly all heterogeneous populations of nucleic acids and will soon be used to map various functional sites in the yeast genome (such as origins of replication and other sites of protein–DNA interactions). Microarray technologies are also applicable to DNA from any organism and will greatly facilitate functional studies of the human genome.

© 2001 Macmillan Magazines Ltd

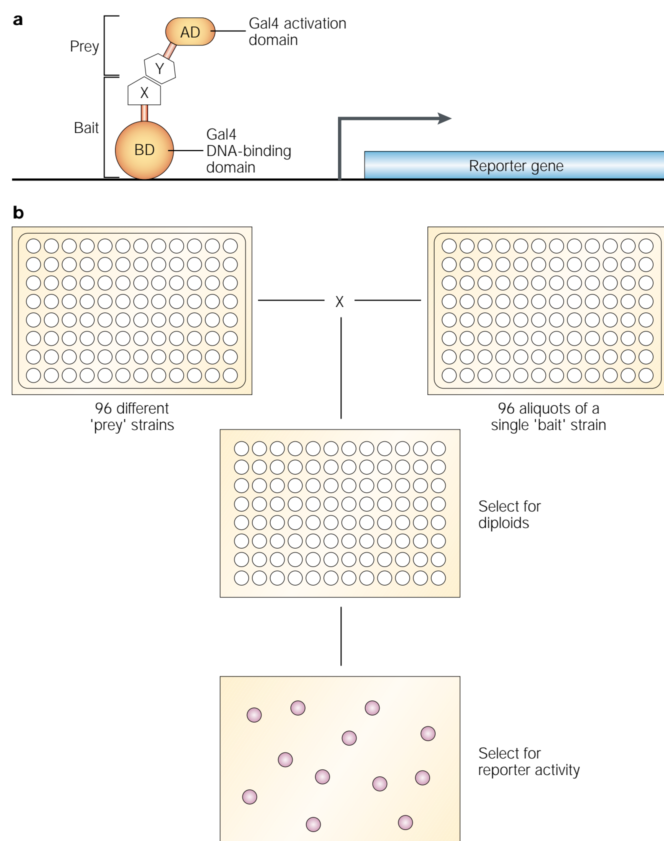

Figure 4 | **Two-hybrid assays. a** | In a typical two-hybrid assay, each putative interacting protein (X and Y) is fused to one of two functionally distinct domains of the transcription factor Gal4. The 'bait' comprises a protein fused to the Gal4 DNA-binding domain, and the 'prey', a protein fused to the Gal4 transcriptional activation domain. Physical interaction between bait and prey brings a DNA-binding domain and an activation domain of Gal4 into close proximity, thereby reconstituting a transcriptionally active Gal4 hybrid. Gal4 activity can be assayed by the expression of reporter genes and selectable markers. **b** | Interaction mating in an array format. To implement this two-hybrid method on a genome-wide scale, bait and prey plasmids are transformed into haploid yeast strains of opposite mating type; after mating, interacting protein pairs can be identified from diploid strains that show Gal4-mediated reporter gene expression[50]. As in REF. 48, an arrayed library of haploid prey strains are mated to an arrayed set of a single haploid bait strain. Resulting diploids are selected under appropriate growth conditions, and selected diploids are then scored on test plates for reporter activity. All transfers are done by an automated high-density replicating tool, which maintains the arrayed format and allows the identities of bait and prey hybrids in colonies expressing a reporter to be determined from the position of the colony in its array.

Large-scale two-hybrid studies

Many established experimental methods have been applied recently on a genome-wide scale, including the yeast two-hybrid assay[46–48]. Initially described in 1989 (REF. 49), the yeast two-hybrid system provides a means of identifying physical interactions between binary protein pairs. A typical two-hybrid assay is shown in FIG. 4a, and its implementation on a genome-wide scale in FIG. 4b.

PROTEOMICS
The development and systematic application of experimental methodologies to analyse the entire protein complement of an organism (its 'proteome').

Recently, several groups have systematically used two-hybrid assays to identify interactions between binary pairs of yeast proteins. In a pilot study, Ito *et al.*[47] cloned 5,700 annotated yeast open reading frames (ORFs) as both bait and prey *GAL4* fusions. In this approach, haploid strains carrying bait and prey plasmids were subdivided into pools for subsequent mating; resulting diploid strains were then screened for Gal4-dependent responses from three reporter genes. The analysis of two-hybrid constructs from the resulting positive clones led to the identification of 175 independent protein–protein interactions from 430 matings (representing 10% of all permutations to be examined). In a separate study, Uetz *et al.*[48] did two large-scale two-hybrid screens to map protein–protein interactions in yeast. In one screen, 192 haploid yeast strains that expressed bait constructs were individually mated against an array of ~6,000 haploid yeast strains of the opposite mating type that expressed prey constructs (see FIG. 4b). After mating, diploid strains that expressed appropriate markers were assayed for reporter gene activity. Proteins that produced a bait–prey interaction were identified by the position of the positive colony in its parent array. In a second screen, 6,000 strains carrying bait constructs were each mated in duplicate to a pooled library of 6,000 strains carrying prey constructs. Per mating, a maximum of 12 positive clones were selected and sequenced. In total, these two screens detected 957 putative interactions that involved 1,004 yeast proteins.

These pilot studies represent promising steps towards the completion of a comprehensive protein–protein interaction map in yeast. They validate the feasibility of large-scale, two-hybrid experiments and highlight the increased sensitivity of array-based methods relative to library screens[48]. As has long been recognized, two-hybrid systems can be used very effectively to survey tight binary interactions that have an equilibrium dissociation constant of less than 10^{-6} M (REF. 50). Two-hybrid assays are also easy to do and are readily applicable to other eukaryotic systems. However, data from two-hybrid experiments must be interpreted with caution, prone as they are to identifying spurious interactions. Additionally, not all proteins are suitable for two-hybrid analysis. Many proteins fail to fold properly when fused to Gal4 domains; others (such as transcription factors) can autonomously activate reporter-gene expression when incorporated into a bait or prey hybrid. Therefore, two-hybrid methods cannot be used to identify all protein–protein interactions in an organism, although future two-hybrid technologies might overcome some of these present-day limitations.

Biochemical genomics

PROTEOMICS is an emerging field; however, genome-wide investigations into protein function require a ready supply of purified proteins that correspond to the entire protein complement of an organism. In yeast, several groups have generated reagents to address this need. In a pilot study[51], a topoisomerase-I-mediated cloning strategy was used to tag 1,553 yeast genes at their carboxyl termini with the V5 epitope and polyhistidine (His) tag. These genes were cloned under the transcriptional control of the galactose-inducible *GAL1* promoter, so that

© 2001 Macmillan Magazines Ltd

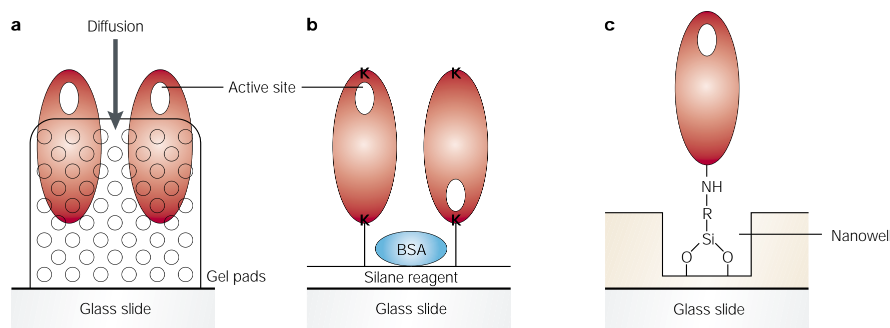

Figure 5 | **Protein microrrays.** Proteins not drawn to scale. **a** | A simplified representation of a gel-pad array, in which protein samples (shown in red) are allowed to diffuse into miniature polyacrylamide gel chambers. Proteins are immobilized within the modified polyacrylamide gel, which provides a matrix for subsequent immunoassays and enzymatic reactions. **b** | Protein microarrays printed onto glass slides. Protein samples are spotted onto a glass slide that has been treated with an aldehyde-containing silane reagent. Proteins are immobilized on the slide through covalent bonding between α-amine groups in the protein and aldehydes in the silane reagent. As proteins typically contain many α-amine groups within lysine (K) residues and at their amino termini, proteins will be bound to the slide in various conformations, with some molecules showing surface-accessible active sites. A buffer containing bovine serum albumin (BSA) blocks nonspecific protein binding during subsequent analysis. **c** | Nanowell protein microarray technology. Miniature wells are cast in a disposable silicone elastomer; the elastomer sheet is then mounted onto a glass slide for ease of handling. Proteins are covalently attached to the wells using a crosslinking compound as shown.

galactose induction could be used to drive the overexpression of each $(His)_6$-tagged gene for subsequent protein purification by metal affinity chromatography. In a larger study[52], ~6,000 yeast ORFs were cloned as gene fusions to glutathione *S*-transferase (GST) under the control of the copper-inducible promoter, P_{CUP1}. By incubating these strains with copper, expression of the GST-fusion genes is induced to generate proteins that can be purified by glutathione affinity chromatography.

Yeast strains that carry these expression constructs can be used to generate proteins for various functional studies. Using an approach termed 'biochemical genomics', Martzen *et al.*[52] purified GST-tagged proteins from pooled sets of yeast strains that express GST-fusion constructs. They screened the pooled precipitates for new biochemical activities; any pools that showed activities of interest were subsequently deconvoluted to identify the source strains. Although this analysis is complicated by the possibility that a given activity is due to a co-precipitated protein, it, nonetheless, underscores the use of these tagged proteins for elucidating protein function.

Protein microarrays

Genome-wide sets of purified proteins are a prerequisite for the development of protein microarrays (a technology by which thousands of proteins can be processed simultaneously for any number of informative assays, including the identification of potential drug targets). Several groups have recently developed microarrays potentially applicable to proteins purified from any source. Arenkov *et al.*[53] have generated arrays of functionally active proteins that are immobilized in tiny gel pads dotted across a glass surface (FIG. 5a). These gel pads act as miniature chambers in which protein assays can be done, providing a hydrated environment to minimize protein denaturation. Gel-pad arrays have been used to detect antigen–antibody interactions and to measure enzyme kinetics[53]; however, these gel pads are difficult to prepare and might not be suitable for all applications (such as the identification of interacting proteins by co-precipitation).

Several groups[54,55] have developed more accessible microarray technologies. One group used a standard CONTACT PRINTING robot to deliver nanolitre volumes of purified protein onto a glass slide that has been precoated with an aldehyde-containing silane reagent[54] (FIG. 5b). Protein samples (deposited in 'spots' 200 μm in diameter) are immobilized on the silane-treated slide through covalent bonding between the aldehydes and primary amines; primary amine groups are found in lysine residues and at the amino terminus of each protein. As typical proteins contain many lysines, protein samples should covalently bind to the slide in various orientations with at least some molecules showing surface-accessible active regions. The printed slide is immersed in a buffer to quench unreacted aldehydes and to reduce nonspecific protein binding. The feasibility of this technology was proved through pilot studies to identify interacting proteins and small molecules.

Another approach has used miniaturized wells cast in a disposable sheet of silicone elastomer (FIG. 5c), which is mounted onto a standard glass microscope slide, to generate a matrix for high-throughput protein analysis[55]. Proteins are covalently attached to the array using an amino crosslinker. This nanowell technology has been used to analyse protein-kinase–substrate specificity in yeast; 32 kinases that preferentially phosphorylate one or two substrates were identified, as well as 27 kinases that phosphorylate tyrosine *in vitro*. Importantly, most proteins bound to the nanowells in this manner were enzymatically active.

CONTACT PRINTING
A method of microarray generation in which samples are spotted onto a slide using specialized spring-loaded printing tips; liquid is drawn up into the printing tip by capillary action and subsequently deposited on contact with the surface of the slide.

© 2001 Macmillan Magazines Ltd

Box 3 | **Online resources for yeast genomics**

Curated databases	
• *Saccharomyces* Genome Database (SGD Function Junction)	http://genome-www.stanford.edu/Saccharomyces/
• Yeast Proteome Database (YPD)	www.proteome.com
• Munich Information Centre for Protein Sequences (MIPS)	www.mips.biochem.mpg.de/proj/yeast
Functional genomic projects	
• Transposon mutagenesis/tagging	http://ygac.med.yale.edu
• Yeast genome deletion project	http://sequencewww.stanford.edu/group/yeast_deletion_project/deletions3.html
• Microarray data sets	http://cellcycle-www.stanford.edu http://genome-www4.stanford.edu/MicroArray/SMD/ http://web.wi.mit.edu/young/pub/regulation.html
• Two-hybrid mapping	http://portal.curagen.com/ http://depts.washington.edu/sfields/projects/YPLM/
• Yeast protein function assignment	http://www.doe-mbi.ucla.edu

At present, protein microarrays are prototypes that require further modification and technical innovation. The nanowell technology described above is being refined to improve sample density by several orders of magnitude. Most microarray technologies could benefit from a more precise means of immobilizing protein on a solid support (so that a greater proportion of proteins are enzymatically active), preferably by attaching each protein through its affinity tag. Nevertheless, even in their present forms, protein microarrays hold tremendous promise as genomic tools. High-density arrays will allow the simultaneous analysis of thousands of proteins, thereby generating a broader level of data than is at present obtainable through traditional methods. Additionally, miniaturized assays require only a small quantity of protein sample (of practical value when reagents, such as PEPTIDE LIBRARIES, are precious).

Proteomics and mass spectrometry

Several high-throughput technologies are now used to resolve large and complex protein mixtures. The composition of heterogeneous protein samples can be analysed by using two-dimensional (2D) gel electrophoresis[56,57], in which proteins are separated by ISOELECTRIC POINT in one dimension and by molecular weight in a second dimension. Proteins in 2D gel 'spots' can be identified by amino-acid analysis, peptide sequencing and mass spectrometry (MS).

In particular, MS methods are rapidly emerging as powerful approaches for identifying components of complexes and subcellular organelles, and for analysing differential protein expression. Upon purification of a structure of interest, its molecular constituents can be deciphered in one of several ways. Initial studies involved separating the proteins in one-dimensional gels and digesting individual bands with trypsin; the resulting fragments were resolved using matrix-assisted laser desorption/ionization (MALDI) MS, which accurately measures molecular mass. By comparing the molecular mass of the fragments with those predicted from the putative translation product of each gene within the genome, it is possible to find sufficient peptide matches to accurately identify the protein. Using this approach, many components of an isolated yeast spindle pole body[58] and of nuclear pore complexes have been identified[59]. More recent studies illustrate the ease of using mass spectrometry without gel electrophoresis. Protein complexes are proteolysed and the resulting peptide fragments are fractionated by liquid chromatography before analysis by tandem MS. Using this approach, 56 ribosomal proteins in the yeast 80S ribosome have been characterized, including one previously unknown protein[60]. This approach promises to be extremely powerful for rapid analysis of complex mixtures.

Differential protein expression can also be analysed by MS using differential isotopic labelling techniques. One clever approach[61] uses proteins from two samples, which are isolated and labelled with either an isotope-coded affinity tag (ICAT) that contains eight deuteriums, or an isotopically light ICAT tag, which contains no deuterium. The ICAT reagent has a thiol and a biotin group that allows peptide fragments containing cysteine to be isolated. Heavy and light protein samples are mixed, digested and the cysteine-containing fragments are analysed by chromatography followed by tandem MS. Because only cysteine-containing fragments are analysed, the complexity of the mixture is reduced so that peptides representing many proteins can be tracked. After MS, the ratio of heavy to light ICAT-labelled peptide is compared for each individual peptide to evaluate its relative quantity within the samples. This technology has been applied to a comparison of protein expression in yeast grown with either ethanol or galactose as a carbon source[61]. The analysis of 800 yeast proteins by this method revealed an expected 200-fold increase in levels of *GAL1* and *GAL10* after growth in the presence of galactose; conversely, the alcohol dehydrogenase *ADH2* was induced 200-fold after growth on ethanol-containing medium. Both results indicate that ICAT labelling can be used to

PEPTIDE LIBRARIES
A collection of small polypeptides that might be used to assay protein function.

ISOELECTRIC POINT
The pH at which a molecule is electrically neutral (the sum of its positive charges equals the sum of its negative charges).

© 2001 Macmillan Magazines Ltd

quantify accurately proteins in complex mixtures. In the future, it is likely that the automatic coupling of several rounds of chromatography to tandem MS will allow the rapid analysis of thousands of proteins and perhaps nearly the entire proteome in a complex mixture without gel electrophoresis. This is extremely powerful, as many types of proteins (such as membrane proteins, and some highly charged proteins) are not readily resolved using 2D gel electrophoresis.

Mass spectrometry is also a powerful tool by which post-translational modifications can be detected: for example, phophorylation and accetylation events can be identified by determining the mass of a purified protein. Finally, MS can also be used as a genomic tool to quantify other macromolecules (such as metabolites in yeast mutants) as illustrated by Raamsdonk *et al.*[62]; their metabolic analysis revealed phenotypes that result from mutations that generate no observable defects in growth rate or other fluxes.

Conclusions

The approaches presented here provide a sound framework for biological discovery; continued innovation will undoubtedly expand this framework, providing new types of genomic data. In the immediate future, advancements in protein genomics will yield global methods to characterize enzymatic activities, protein–protein interactions and post-translational modifications. Microarray technologies will drive many of these approaches, as well as providing a continued means of assaying gene expression and protein–DNA interactions. Technological advances, coupled with increased practical familiarity, will render genomic approaches more accurate (an absolute necessity to tap the potential in bioinformatics and computational biology). Bioinformatic resources, in the form of curated databases and on-line data sets (BOX 3), provide an invaluable community-based repository of information, hastening the rate at which technological advances are developed in yeast. Ultimately, these genomic methods and resources might offer the most expedient and practical route to our end-goal: a comprehensive understanding of eukaryotic cellular function. Towards this end, *S. cerevisiae* will continue to have a key role both as a genomic model and as a source of biological information germane to further advancement of the human condition.

Links

DATABASE LINKS **Tor1 | Tor2 | MAPK | TFIID | Srb5 | Srb10 | *erg2* | Rpn4 | Swi4 | Swi6 | Mbp1 | Gal4 | Ste12 | Spo11 | *GAL1* | GST | *GAL10* | *ADH2* | *Saccharomyces* Genome Database (SGD) (Function Junction) | Yeast Proteome Database (YPD) | Munich Information Centre for Protein Sequences (MIPS) | Transposon mutagenesis/tagging | Yeast Genome Deletion Project | Stanford cell-cycle microarray data set | Stanford Mircroarray Database | Rick Young's microarray data set | Curagen's Gene Scape Portal | Stan Field's two-hybrid data set | Yeast protein function assignment**

FURTHER INFORMATION **NCBI's Entrez Genome site | Affymetrix, Inc. | Rosetta Inpharmatics**

ENCYCLOPEDIA OF LIFE SCIENCES ***Saccharomyces cerevisiae*: Applications | Mass spectroscopy in biology | Multispot array technologies**

1. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
2. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
3. Hieter, P. & Boguski, M. Functional genomics: it's all how you read it. *Science* **278**, 601–602 (1997).
4. Costanzo, M. C. *et al.* The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73–76 (2000).
5. Bassett, D. E., Boguski, M. S. & Hieter, P. Yeast genes and human disease. *Nature* **379**, 589–590 (1996).
6. Bassett, D. E. *et al.* Genome cross-referencing and XREFdb: implications for the identification and analysis of genes mutated in human disease. *Nature Genet.* **15**, 339–344 (1997).
7. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
8. The Genome International Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
9. Oliver, S. G. From DNA sequence to biological function. *Nature* **379**, 597–600 (1996).
10. Devine, S. E. & Boeke, J. D. Efficient integration of artificial transposons into plasmid targets *in vitro*: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.* **22**, 3765–3772 (1994).
11. Garfinkel, D. J. Insertional mutagenesis by Ty elements in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **53**, 227–237 (1996).
12. Gwinn, M. L., Stellwagen, A. E., Craig, N. L., Tomb, J.-F. & Smith, H. O. *In vitro* TN*7* mutagenesis of *Haemophilus influenzae* Rd and characterization of the role of atpA in transformation. *J. Bacteriol.* **179**, 7315–7320 (1997).
13. Burns, N. *et al.* Large-scale characterization of gene expression, protein localization and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.* **8**, 1087–1105 (1994).
14. Smith, V., Botstein, D. & Brown, P. O. Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl Acad. Sci. USA* **92**, 6479–6483 (1995).
15. Ross-Macdonald, P., Sheehan, A., Roeder, G. S. & Snyder, M. A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **94**, 190–195 (1997).
16. Ross-MacDonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
The use of shuttle mutagenesis with multipurpose transposons in a genome-wide analysis of disruption phenotypes, gene expression and protein localization.
17. Hoekstra, M. F. *et al.* A Tn*3* derivative that can be used to make short in-frame insertions within genes. *Proc. Natl Acad. Sci. USA* **88**, 5457–5461 (1991).
18. Biery, M. C., Stewart, F. J., Stellwagen, A. E., Raleigh, E. A. & Craig, N. L. A simple *in vitro* Tn*7*-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res.* **28**, 1067–1077 (2000).
19. 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).
20. Ross-Macdonald, P., Sheehan, A., Friddle, C., Roeder, G. S. & Snyder, M. Transposon mutagenesis for the analysis of protein production, function, and localization. *Methods Enzymol.* **303**, 512–532 (1999).
21. Kumar, A., des Etages, S. A., Coelho, P. S. R., Roeder, G. S. & Snyder, M. High-throughput methods for the large-scale analysis of gene function by transposon tagging. *Methods Enzymol.* **328**, 550–574 (2000).
22. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
23. Kumar, A. *et al.* TRIPLES: a database of gene function in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **28**, 81–84 (2000).
24. Davies, C. J. & Hutchison, C. A. Insertion site specificity of the transposon Tn*3*. *Nucleic Acids Res.* **23**, 507–514 (1995).
25. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
A report on the progress of an international consortium that disrupted each annotated open reading frame in the yeast genome using a targeted PCR-based gene-deletion strategy.
26. 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.* **21**, 3329–3330 (1993).
27. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
28. Hughes, T. R. *et al.* Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genet.* **25**, 333–337 (2000).
29. Chan, T.-F., Carvalho, J., Riles, L. & Zheng, X. F. S. A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR). *Proc. Natl Acad. Sci. USA* **97**, 13227–13232 (2000).
30. Giaever, G. *et al.* Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nature Genet.* **21**, 278–283 (1999).
31. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).

© 2001 Macmillan Magazines Ltd

32. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
33. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
34. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
35. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
36. Jelinsky, S. A. & Samson, L. D. Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc. Natl Acad. Sci. USA* **96**, 1486–1491 (1999).
37. Madhani, H. D., Galitski, T., Lander, E. S. & Fink, G. R. Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatures of signaling mutants. *Proc. Natl Acad. Sci. USA* **96**, 12530–12535 (1999).
38. Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
39. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
40. Roberts, C. J. *et al.* Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–878 (2000).
41. Jelinsky, S. A., Estep, P., Church, G. M. & Samson, L. D. Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol. Cell. Biol.* **20**, 8157–8167 (2000).
42. Iyer, V. R. *et al.* Genomic binding distribution of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
43. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
References 42 and 43 describe a new genomic approach that combines chromatin immunoprecipitation and microarray hybridization to map the binding sites of chromosomal proteins *in vivo*.
44. Gerton, J. L. *et al.* Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA* **97**, 11383–11390 (2000).
45. Baudat, F. & Nicolas, A. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl Acad. Sci. USA* **94**, 5213–5218 (1997).
46. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277–282 (1997).
47. Ito, T. *et al.* Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA* **97**, 1143–1147 (2000).
48. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae. Nature* **403**, 623–627 (2000).
References 47 and 48 describe large-scale applications of the two-hybrid approach towards identifying protein–protein interactions in yeast.
49. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
50. Brent, R. & Finley, R. L. Understanding gene and allele function with two-hybrid methods. *Annu. Rev. Genet.* **31**, 663–704 (1997).
51. Heyman, J. A. *et al.* Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* **9**, 383–392 (1999).
52. Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
A genome-wide biochemical approach to identifying gene function.
53. Arenkov, P. *et al.* Protein microchips: use for immunoassay and enzymatic reactions. *Anal. Biochem.* **278**, 123–131 (2000).
54. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
55. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283–289 (2000).
References 53–55 describe three new protein microarray methodologies for high-throughput protein analysis.
56. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368 (1999).
57. Perrot, M. *et al.* Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis* **20**, 2280–2298 (1999).
58. Wigge, P. A. *et al.* Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141**, 967–977 (1998).
59. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
60. Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**, 676–682 (1999).
61. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
A powerful approach to quantitative protein analysis using isotope-coded affinity tags and mass spectrometry.
62. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol.* **19**, 45–50 (2001).
63. Pease, A. C. *et al.* Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA* **91**, 5022–5026 (1994).
64. Marton, M. J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**, 1293–1301 (1998).
65. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
66. Tamayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci. USA* **96**, 2907–2912 (1999).
67. Everitt, B. *Cluster Analysis* Vol. 122 (Heinemann, London, 1974).
68. Heyer, L. J., Kruglyak, S. & Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* **9**, 1106–1115 (1999).
69. Brown, M. P. *et al.* Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA* **97**, 262–267 (2000).
70. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nature Genet.* **22**, 281–285 (1999).

Acknowledgements

A.K. is supported by a postdoctoral fellowship from the American Cancer Society.

© 2001 Macmillan Magazines Ltd