15. Ohta, T. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. USA* **90,** 10676–10680 (1993).
16. Vogel, F., Kopun, M. & Rathenberg, R. in *Molecular Anthropology* (eds Goodman, M. & Tashian, R. E.) 13–33 (Plenum, New York, 1976).
17. Burt, D. W. & Cheng, H. H. The chicken gene map. *Inst. Lab. Anim. Res. J.* **39,** 229–236 (1998).
18. Hastie, T. J. & Tibshirani, R. J. *Generalised Additive Models* (Chapman and Hall, London, 1990).
19. Hurvich, C. M., Simonoff, J. S. & Tsai, C.-L. Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. *J. R. Statist. Soc.* **60,** 271–293 (1998).
20. Kosambi, D. D. The estimation of map distance from recombination values. *Ann. Eugen.* **12,** 172–175 (1944).
21. Morton, N. E. Parameters of the human genome. *Proc. Natl. Acad. Sci. USA* **88,** 7474–7476 (1991).
22. Wienberg, J. & Stanyon, R. Comparative chromosome painting of primate genomes. *Inst. Lab. Anim. Res. J.* **39,** 77–91 (1998).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London Editorial office of *Nature*.

Acknowledgements

We thank members of the EC ChickMAP project and H. Cheng for providing data prior to publication, and J. Burt for helpful comments. Genome research at the Roslin Institute is supported by the Ministry of Agriculture, Fisheries and Food, the Biotechnology and Biological Sciences Research Council and the Commission of the European Communities.

Correspondence and requests for materials should be addressed to D.W.B. (e-mail: Dave-Burt@bbsrc.ac.uk).

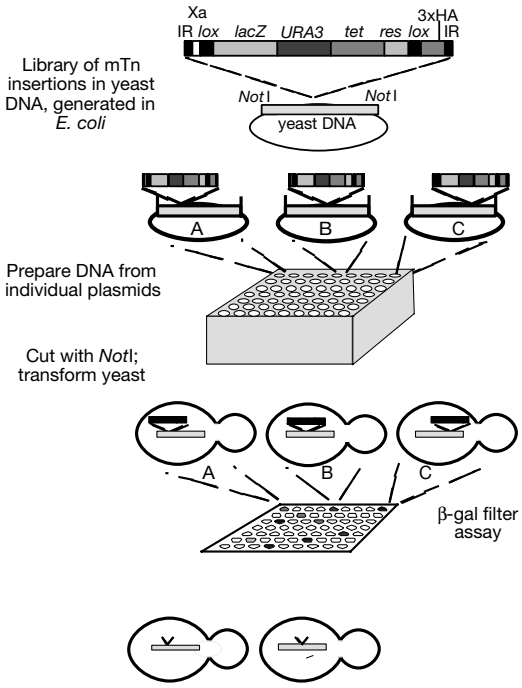

Figure 1 The mTn insertion project. Most steps were performed using a Robbins Hydra 96-channel dispenser; all strains are maintained in a 96-well format.

Large-scale analysis of the yeast genome by transposon tagging and gene disruption

Petra Ross-Macdonald*, Paulo S. R. Coelho*, Terry Roemer*, Seema Agarwal*, Anuj Kumar*, Ronald Jansen†, Kei-Hoi Cheung‡, Amy Sheehan*, Dawn Symoniatis*, Lara Umansky*, Matthew Heidtman*, F. Kenneth Nelson*, Hiroshi Iwasaki*, Karl Hager§, Mark Gerstein†, Perry Miller‡, G. Shirleen Roeder* & Michael Snyder*†

* *Department of Molecular, Cellular and Developmental Biology, Yale University, PO Box 208103, New Haven, Connecticut 06520-8103, USA*
† *Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA*
‡ *Center for Medical Informatics, Department of Anesthesiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA*
§ *Keck Foundation Biotechnology Resource Laboratory, 295 Congress Avenue, Yale University, New Haven, Connecticut 06520, USA*

Economical methods by which gene function may be analysed on a genomic scale are relatively scarce. To fill this need, we have developed a transposon-tagging strategy for the genome-wide analysis of disruption phenotypes, gene expression and protein localization, and have applied this method to the large-scale analysis of gene function in the budding yeast *Saccharomyces cerevisiae*. Here we present the largest collection of defined yeast mutants ever generated within a single genetic background—a collection of over 11,000 strains, each carrying a transposon inserted within a region of the genome expressed during vegetative growth and/or sporulation. These insertions affect nearly 2,000 annotated genes, representing about one-third of the 6,200 predicted genes in the yeast genome[1,2]. We have used this collection to determine disruption phenotypes for nearly 8,000 strains using 20 different growth conditions; the resulting data sets were clustered to identify groups of functionally related genes. We have also identified over 300 previously non-annotated open reading frames and analysed by indirect immunofluorescence over 1,300 transposon-tagged proteins. In total, our study encompasses over 260,000 data points, constituting the largest functional analysis of the yeast genome ever undertaken.

The ability to sequence entire genomes has resulted in an abundance of raw sequence data[3]; however, relatively few methods exist to assess gene function on a genomic scale[4–8]. We have developed a transposon-based method for the large-scale accumulation of expression, phenotypic and protein localization data in yeast without bias towards previously annotated genes. Our approach utilizes a multipurpose minitransposon (mTn) derived from the bacterial transposable element Tn3 (ref. 9). The minitransposon mTn–3xHA/*lacZ* (Fig. 1) contains a *lacZ* reporter gene lacking an initiator methionine and upstream promoter sequence. Introduction of this transposon into yeast results in production of β-galactosidase (β-gal) if the mTn is present within a transcribed and translated region of the genome, typically corresponding to an in-frame fusion of *lacZ* to the yeast protein-coding sequence. Additionally, mTn–3xHA/*lacZ* contains a *lox* site near each Tn3 end; adjacent to one *lox* site is DNA encoding three copies of a haemagglutinin (3xHA) epitope tag. Production of the Cre recombinase in yeast containing this minitransposon results in recombination of the *lox* sites, reducing the mTn to a 274 base pair (bp) element (the HAT tag)[9]. As five bases of genomic DNA are duplicated during Tn3 insertion, the net result is a 279-bp insertion encoding a 93-codon open reading frame (ORF) which includes the 3xHA sequence. When the mTn's *lacZ* reporter has been fused in-frame to a yeast coding region, creation of the HAT tag allows production of a full-length, epitope-tagged protein from that gene.

We have used mTn–3xHA/*lacZ* in conjunction with high throughput methods to generate a large collection of yeast strains, each containing an insertion at a known location within the genome (Fig. 1). Briefly, a yeast genomic DNA library was mutagenized in *Escherichia coli* with mTn–3xHA/*lacZ*. Using a 96-well format, individual plasmids were prepared, digested with *Not*I and transformed into a diploid yeast strain[10]. By homologous recombination, each fragment should integrate at its corresponding genomic locus, thereby replacing its genomic copy. To verify this process, polymerase chain reaction (PCR) analysis was performed on 48 independent yeast transformants, each carrying one of six different mTn insertion alleles. Using a primer in the mTn and a primer external to

flanking genomic DNA, we confirmed that the allele had substituted properly in 47 out of 48 cases.

Transformants carrying the mTn-borne *URA3* marker were selected and assayed for β-gal activity after growth on rich medium and sporulation medium[10]. Strains were also treated to induce the Cre recombinase. Following this Cre-mediated excision event, we used Ura$^-$ colonies for immunolocalization of epitope-tagged proteins with anti-HA antibodies. The bacterial clone that gave rise to each yeast strain in the collection was also stored. Plasmid DNA from these bacteria allowed transformation of insertion alleles into a haploid strain for analysis of disruption phenotypes.

Using this procedure, we have carried out 92,544 plasmid preparations and yeast transformations, allowing us to identify 11,232 strains containing *lacZ* fusions expressed during vegetative growth; the precise site of mTn insertion within the yeast genome has been determined by DNA sequence analysis for 6,358 strains. As expected, most of the strains within our collection contain transposon insertions affecting annotated ORFs. Of these 6,358 insertion alleles, 5,442 lie in or within 200 bp of an annotated ORF. In total, these insertions affect 1,917 ORFs distributed over all 16 chromosomes of the yeast genome (Fig. 2).

Most of these transposon insertions represent in-frame fusions to coding sequence: we have identified in-frame fusion events in 1,346 annotated ORFs (Fig. 2). However, reproducibly high levels of β-gal activity were also found in around 480 strains carrying a transposon located within an annotated ORF, but in either the +1 or −1 reading frame. In a study of 287 such insertions, we found that an alternative in-frame start codon is lacking in 75% of the sequences, indicating that a frameshift may occur during translation of the annotated ORF. Translational frameshifting within the yeast genome may, therefore, occur more frequently than has been reported.

Figure 2 Distribution and phenotypic analysis of 1,917 mTn-mutagenized ORFs within the *S. cerevisiae* genome. Each chromosome is represented as a linear map with start and end coordinates indicated to the left and right, respectively. Mutagenized ORFs are shown as proportionally sized boxes within either the Watson (upwards) or Crick (downwards) strand of each chromosome. Common gene names have been used when available; otherwise, ORFs are designated by the third through sixth characters of their standardized names. All in-frame mTn3 fusions are indicated with asterisks. A subset of putative NORFs located within intergenic regions are shown in their corresponding chromosomal positions; these NORFs were identified by multiple mTn insertion events. NORFs included have been designated N400 through N442 (see ref. 4). Macroarray analysis has identified disruption phenotypes associated with 407 ORFs; these ORFs have been colour-coded as indicated (see Table 1). Complete data sets can be accessed at http://ycmi.med.yale.edu/ygac/triples.htm.

In analysing mTn insertion events, we have identified a large number of previously non-annotated ORFs (NORFs) within the yeast genome. When annotating the *S. cerevisiae* genome, an arbitrary lower size limit of 100 amino acids was adopted as convention unless additional information suggested otherwise. Also, ORFs of more than 100 codons that fully or partially overlapped even longer ORFs were usually not annotated[4,11]. Such NORFs are, however, capable of being translated, as indicated in previous studies[4,8]. Using an arbitrary lower limit of 50 codons, we find that 328 of 6,358 mTn insertions represent in-frame fusions to NORFs. These NORFs range in size up to 247 codons. A detailed analysis of 229 such insertions revealed three classes: 52% are in NORFs that fully or partially overlap an annotated ORF in the antisense direction; 15% fall in NORFs that overlap an annotated ORF in the same orientation but in a different frame; and 33% of these insertions occurred in NORFs previously classified as intergenic regions. Many NORFs were identified by multiple mTn insertion events. For example, a 124-codon NORF in the antisense

Table 1 Phenotypes scored by macroarray analysis

Assay/growth conditions	Number*	Colour code‡	Abbreviation§
YPD + 8 mM caffeine	27	Purple	Caff
Cycloheximide hypersensitivity: YPD + 0.08 μg ml^{-1} cycloheximide at 30 °C	28†	Pink	Cyc^S
White/red colour on YPD	39	Yellow	W/R
YPGlycerol	54	Yellow	YPG
Calcofluor hypersensitivity: YPD + 12 μg ml^{-1} calcofluor at 30 °C	66	Purple	Calc^S
YPD + 46 μg ml^{-1} hygromycin at 30 °C	136	Purple	Hyg
YPD + 0.003% SDS	109	Purple	SDS
Benomyl hypersensitivity: YPD + 10 μg ml^{-1} benomyl	67	Green	Ben^S
YPD + 5-bromo-4-chloro-3-indolyl phosphate at 37 °C	35	Purple	BCIP
YPD + 0.001% methylene blue at 30 °C	12	Purple	MB
Benomyl resistance: YPD + 20 μg ml^{-1} benomyl	11†	Green	Ben^R
YPD at 37 °C	29	Cyan	YPD^{37}
YPD + 2 mM EGTA	30	Black	EGTA
YPD + 0.008% MMS	16	Pink	MMS
YPD + 75 mM hydroxyurea	21	Pink	HU
YPD at 11 °C	20	Cyan	YPD^{11}
Calcofluor resistance: YPD + 0.3 μg ml^{-1} calcofluor at 30 °C	4†	Purple	Calc^R
Cycloheximide resistance: YPD + 0.3 μg ml^{-1} cycloheximide	2†	Pink	Cyc^R
Hyperhaploid invasive growth mutants	25	Orange	HHIG
YPD + 0.9 M NaCl	13	Black	NaCl

All tests were carried out at room temperature (~24 °C) unless otherwise noted.
* The total number of ORFs conferring a given phenotype upon disruption is indicated.
† Data sets are incomplete for assays of cycloheximide hypersensitivity (5,760 transformants screened), cycloheximide resistance (3,456 transformants screened), calcofluor resistance (2,208 transformants screened) and benomyl resistance (5,760 transformants screened).
‡ Colours correspond to those shown in Fig. 2.
§ The abbreviations for each growth condition are used as labels in Fig. 4.
The tree to the left of the table was generated by cluster analysis and indicates growth conditions that were found to result in most similar phenotypes for the entire data set (see also Fig. 4).

region of the rDNA has more than 50 insertions. In addition, we have observed mutant phenotypes resulting from transposon insertions within putative NORFs: for example, mTn insertion at base number 198,816 of chromosome XII (identifying a putative intergenic NORF of 86 amino acids) results in hypersensitivity to the microtubule-depolymerization drug benomyl. Furthermore, three NORFs encode proteins exhibiting a distinct pattern of subcellular localization. Transposon insertion at base number 1,236,754 of chromosome IV identifies a NORF of 97 amino acids; using the mTn-encoded epitope tag, we have localized the encoded protein to the nucleus. These findings substantiate the biological importance of NORFs, raising the possibility that they encode previously uncharacterized proteins. Complete data sets for all strains carrying mTn insertions identifying putative NORFs may be accessed from our web site at http://ycmi.med.yale.edu/ygac/triples.htm.

To analyse phenotypes on a genomic scale, we transformed 7,680 mTn-insertion alleles into a haploid strain. Most transformants were viable. In the 14.1% (1,082) that were inviable, the transposon typically affected a gene known to be required for cell viability. Gene functions associated with these essential genes may also be analysed using our multipurpose minitransposon. By *cre-lox*-mediated reduction of mTn–3xHA/*lacZ*, we have generated viable mutants containing in-frame insertions of HAT-tags within essential genes. We estimate that 30–45% of small insertion alleles in essential genes retain gene function.

Transformed haploid strains were scored for 20 phenotypes after growth under the test conditions indicated in Table 1 and Fig. 2. Transformants were screened on a large scale using 'phenotypic macroarrays' compatible with 96-well formats. Yeast strains grown in 96-well arrays were transferred to test medium using a 96-pin tool; 576 strains were simultaneously screened for a given phenotype (Fig. 3).

Tranposon insertions in 407 genes resulted in a phenotype significantly different from the wild type (excluding those insertion alleles causing inviability). Many mutant alleles caused phenotypes that might be expected. For example, strains carrying insertions in *TUB3* (ref. 12), *CIK1* (ref. 13), and *CIN1* (ref. 14)—genes known to be involved in microtubule functions—are sensitive to benomyl. In other cases, insertions within characterized genes resulted in unexpected phenotypes, indicating that these genes may have previously unrecognized roles or affect multiple processes indirectly. For example, the serine/threonine phosphatase Rts1p (ref. 15) is involved in the control of stress-related responses and cell-cycle regulation, whereas *YMA5* (ref. 16) encodes a hydrophilic subunit of a vacuolar H^+-ATPase; mutations affecting both these ORFs caused sensitivity to the cell wall-binding dye calcofluor, indicative of genes involved in cell wall biogenesis[17]. Other insertions caused mutant phenotypes in a variety of apparently unrelated assays: strains containing insertions in genes for nuclear pore complex proteins and for certain transcription factors were sensitive to hydroxyurea, benomyl and calcofluor. We speculate that genes whose mutations result in highly pleiotropic phenotypes are probably involved in

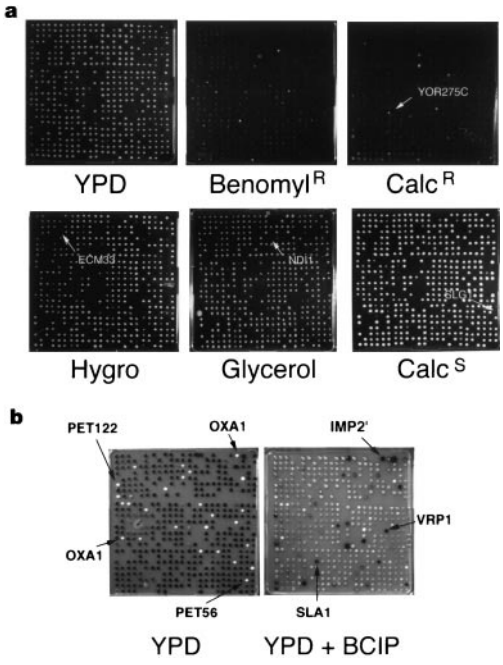

Figure 3 Phenotypic macroarray analysis. **a**, Examples of 21 cm × 21 cm macroarray plates scoring growth on test media: YPD, YPD supplemented with 20 μg ml^{-1} benomyl, YPD with 66.7 μg ml^{-1} calcofluor, YPD with 46 μg ml^{-1} hygromycin, YPGlycerol, YPD supplemented with 12 μg ml^{-1} calcofluor. Arrows indicate strains mutated for genes functioning in cellular respiration (*NDI1*) or cell-wall biogenesis (*ECM33*, *SLG1*, *YOR275C*)[17]. **b**, Macroarray analysis of yeast mutants deficient in oxidative phosphorylation and cell-wall maintenance. Mitochondrial mutants unable to carry out oxidative phosphorylation were identified as white (rather than red) colonies on YPD medium; red pigment formation within *ade2* mutant strains requires oxidative phosphorylation[28]. Representative respiratory genes identified within our screen are labelled with arrows (*PET122*, *PET56*, *OXA1*). Genes functioning in cell-wall maintenance were characterized through macroarray analysis of mutants grown on YPD overlaid with agar containing BCIP. BCIP is a chromogenic substrate of alkaline phosphatase, which when released from vacuoles during cell lysis cleaves BCIP, thereby staining lysed mutants blue. Arrows highlight a sample of genes (*IMP2'*, *VRP1*, *SLA1*)[17] involved in cell-wall maintenance.

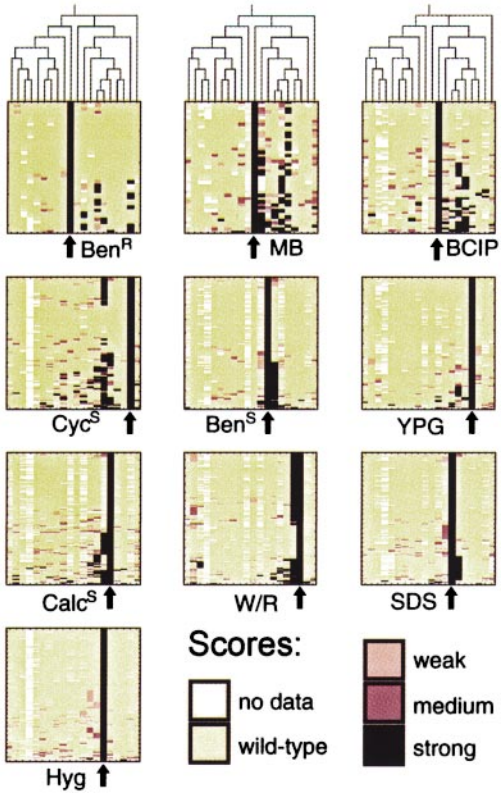

Figure 4 Graphical representation of clustered phenotypic data. Examples of data clusters exhibiting a predominant phenotype are shown (all 21 clusters may be viewed at http://bioinfo.mbb.yale.edu/genome/phenotypes). Vertical columns represent growth conditions; each horizontal row represents a different transformant. Transformants in each cluster are sorted from top to bottom according to the distance from the cluster average (centroid). The abbreviation below each square indicates the predominant phenotype of each cluster (see Table 1). The arrow below each square points to the column representing this predominant phenotype. In addition to clustering transformant phenotypes (the rows of the data set), we have also clustered vertical data columns (representing assayed growth conditions). The tree generated by column clustering is shown above the first row of clusters (as well as to the left of Table 1).

Table 2 Observed immunofluorescence patterns for in-frame HAT fusions

Pattern	Number*
Discrete	
Nuclear	
general	41
granular	64
nucleolus	22
Nuclear rim/endoplasmic reticulum	29
Mitochondrial	37
Spindle pole body/microtubules	5
Cell periphery	11
Cytoplasmic patch/dots	10
Cell neck	2
Vacuole	2
General cytoplasmic	
Uniform, finely speckled	3
Granular, fibrous	189
Background	925
Total	1,340

*All strains exhibiting discrete cellular or cytoplasmic localizations were tested at least twice. Complete data sets for all tested clones can be accessed at http://ycmi.med.yale.edu/ygac/triples.htm.

general cellular processes. Finally, many strains carrying insertions within uncharacterized ORFs exhibit phenotypes, thereby providing clues as to the function of these genes (for example, *YBR077C*, *YDR101C*, *YIR016W* and *YIL129W*).

Interestingly, the exact site of transposon insertion within an ORF determines its disruption phenotype. As a result, our transposon can be used effectively to map functional domains within a given protein. For example, we have identified 17 transformants containing mTn insertions at 11 sites within *IMP2′*, a nuclear gene encoding a transcription factor involved in sugar utilization[18]. Tranposon insertions within the amino terminus and central region of Imp2 (at amino acids 46, 69, 230 and 270) result in mutants unable to grow on glycerol, a non-fermentable carbon source. However, strains carrying an insertion at codon 334 (12 codons upstream of the translational stop codon) still grow normally on YPGlycerol. This same insertion, though, results in a defect in cell-wall synthesis (Fig. 3b) shared by strains containing insertions at residues 46, 230, 263 and 319. Therefore, nearly the entire length of Imp2 appears to be involved in cell-wall biogenesis[17], although its carboxy terminus is not required for the control of sugar utilization. These findings illustrate the effectiveness of our collection in assigning specific functions to individual protein domains.

Many insertional mutants exhibit phenotypes under a specific subset of test conditions: these mutants appear to 'cluster' into discrete classes. To consider these clusters more rigorously, we have developed and implemented an algorithm by which haploid transformants may be grouped by common disruption phenotypes (see Methods). Based upon the tested phenotypes listed in Table 1, our collection of viable haploid transformants may be subdivided into 21 clusters: one cluster for each indicated growth condition as well as an additional cluster of mutants exhibiting wild-type phenotypes. Examples of these clusters are shown in Fig. 4. Transformants within a single cluster typically share a common mutant phenotype (Fig. 4) and can be subdivided further into smaller clusters. Interpreted judiciously, these phenotypic clusters provide a means of predicting cellular functions associated with a given ORF. For example, cluster 'YPG' contains transformants characterized most prominently by an inability to grow on YPGlycerol. This cluster is highly enriched in transformants carrying a transposon insertion within genes involved in cellular respiration (for example, *IFM1*, *IMP2′*, *SHY1* and *NDI1*). A number of strains within this cluster carry transposon insertions in previously uncharacterized genes (for example, *YLR294C*, *YJL131C* and *YGR101W*); from our clustering analysis, we can reasonably infer that these genes probably encode proteins involved in cellular respiration as well. Similar functional inferences may be drawn from each remaining data cluster (accessible in full at http://bioinfo.mbb.yale.edu/genome/phenotypes).

Additionally, our algorithm may be used to 'double cluster'

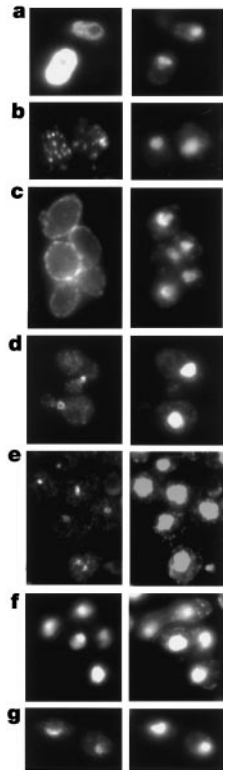

Figure 5 Immunolocalization of epitope-tagged proteins. Left, examples of immunofluorescence patterns in vegetative cells stained with monoclonal antibody against HA. Right, the same cells stained with the DNA-binding dye 4′,6-diamidino-2-phenylindole (DAPI). **a**, Diffuse cytoplasmic staining in a transformant HAT-tagged through an in-frame mTn insertion event in *YLR249W*. **b**, A punctate pattern of cytoplasmic staining resulting from HAT-tagging of Crn1, an actin-binding protein that bundles actin filaments[29]. **c**, Localization to the plasma membrane. **d**, Localization of HAT-tagged Bni4 to a ring in the mother-bud neck. **e**, Staining of the spindle body in a strain containing a hAT-tagged allele of the spindle pole protein Spc105. **f**, Nuclear staining observed in transformants HAT-tagged by mTn insertion in *NSR1*, a gene whose protein product binds nuclear localization sequences[30]. **g**, Localization of HAT-tagged Nan1 to the nucleolus.

observed phenotypic data: we can use this algorithm to group phenotypic assays (rather than transformants) based upon the ability of these assays to identify strains exhibiting similar phenotypic profiles. By double-clustering the data presented above, we can identify sets of assays that are effective in selecting functionally related proteins with a greater degree of confidence. To illustrate this, consider the clustered assays presented in the dendrogram to the left of Table 1 (reproduced at the top of Fig. 4). Note the tight grouping of assays (for example, hypersensitivity to calcofluor, hygromycin and SDS) that are expected to identify mutants defective in cell-wall biogenesis and maintenance[17]. Similarly, assays for growth in hydroxyurea and MMS (methyl methanesulphonate) are closely associated; these assays would be expected to preferentially identify mutants defective in DNA metabolism. The use of such assays in combination would, therefore, constitute a more effective means of identifying a pool of mutants enriched in a common phenotype.

Further indication of cellular function may be inferred from the subcellular localization of a given protein. To determine these localization patterns, we have examined 1,340 diploid strains carrying in-frame HAT tag insertions by indirect immunofluorescence with antibodies directed against the HA-epitope. Localization of the HAT-tagged protein to a discrete cellular site was observed and confirmed for 201 strains; 214 additional strains showed cytoplasmic localization. A variety of localization patterns was detected within these strains: tagged proteins were localized to the nucleus, nucleolus, mitochondria, plasma membrane, cell neck and

spindle pole body (Table 2 and Fig. 5). In many cases, strains that show a distinct localization pattern contain a HAT tag in a known protein whose localization had previously been determined (for example, Top1–HAT[19] and Rap1–HAT[20] localize to the nucleus). In other cases, however, our data constitute information concerning proteins that have not been studied or localized previously (such as the YHR196W gene product—a protein of unknown function—which localizes to the nucleolus).

Our transposon-based system also presents a means of defining gene expression profiles on a genomic scale. For example, to identify genes whose expression is strongly induced during sporulation, diploid transformants carrying mTn insertions were incubated on sporulation and control media and then assayed for β-gal activity. Of 106 insertions in yeast genomic DNA, 36 represent in-frame fusions in 31 different genes. Six of these genes have been previously shown to be induced in meiosis, eight are in named genes not known to be meiotically induced, and the remainder are in ORFs not yet characterized (Fig. 2). As noted above for fusion genes expressed during vegetative growth, mTn insertions are also in the antisense direction or in the +1 or −1 reading frame of annotated ORFs. In addition, several insertions are in NORFs, noncoding sequences or the ribosomal DNA.

In ref. 21, a DNA microarray representing all annotated ORFs in the yeast genome was screened for changes in gene expression during sporulation. Of the 31 meiotic genes we identified by in-frame *lacZ* fusions, only 17 were also found to be induced at least two-fold by microarray screening. For the remainder, microarray analysis indicated little or no induction during meiosis. One of the genes detected in our analysis but not by microarray screening is *MER1*, which has been shown to be strongly induced in meiosis both by northern blot hybridization and analysis of a *Mer1–β-gal* translational fusion gene[22]. This observation indicates that the β-gal assay may serve as a more sensitive indicator of meiotic gene expression and/or detect differences in protein levels more readily. In any case, our transposon-based approach is an effective means of identifying novel sporulation-induced genes; one gene identified in our screen is *YIL073C*, which we found to be important for spore viability and synaptonemal complex formation.

Our transposon insertion strategy allows the rapid generation of reporter gene fusions, epitope-tagging constructs and insertion alleles (both large and small), all from a single mutagenic event. As we do not bias this event towards a population of previously annotated genes, our approach effectively identifies new genes. This random approach, however, presents an obstacle in achieving saturation mutagenesis: small genes are less likely to be mutagenized by random mTn insertion than are large genes. Using our current approach, an additional 30,000 mTn insertions in yeast ORFs would be required to mutagenize 90% of the yeast genome. Once complete, however, we see our collection ushering in the 'new yeast genetics'—a new paradigm by which yeast genetic screens may be performed. Using a collection such as ours, researchers will be able to screen large numbers of mutant yeast strains for a given phenotype (by macroarray analysis or other methods) and instantly know the gene affected in a mutant of interest. Furthermore, the availability of alleles as a plasmid collection will allow introduction into any desired yeast strain background before screening. We have already distributed over 700 insertion alleles to the research community. These alleles will expedite gene analysis. □

Methods

Generation and analysis of yeast transformants

We used transposon mutagenesis in *E. coli* to generate $>10^6$ independent transformants; individual transformant colonies were selected and stored in 96-well plates[10]. Plasmids were prepared from these strains by alkaline lysis[23], digested with *Not*I and transformed into diploid strain Y800 (ref. 8). Following transformation, yeast cells were cultured as described[10]; after incubation under appropriate growth conditions, cells were permeabilized and assayed for β-gal activity[10]. Transformants displaying activity were arrayed in a 96-well format; each strain has a unique coordinate in this collection. The strains were subsequently re-tested for β-gal activity after vegetative growth and sporulation. The results of this retest are accessible at http://ycmi.med.yale.edu/ygac/triples.htm.

To determine the genomic site of mTn insertion within these clones, plasmid-borne insertion alleles were sequenced using a primer complementary to bases 74–97 of mTn–3xHA/*lacZ* (GenBank accession number U54828). The resulting sequence data were searched against the *S. cerevisiae* genome using BLAST[24].

Phenotypic macroarray analysis

DNA obtained from the insertion allele plasmid collection was digested with *Not*I and transformed into haploid strain Y2279 [*MAT***a** *ura3-52 trp1Δ1 ade2-101 lys2-801 leu2Δ98*]. Transformants were selected on SC-ura. Because phenotypes are often associated with mutations unlinked to the transposon insertion (M.S., unpublished data), we usually analysed two independent transformants. We scored yeast strains as exhibiting a defect only if both transformants displayed the defect or if two independent insertion alleles in the same gene produced the same defect. In many cases, haploid transformants were not obtained. We presume that in these cases either the affected gene was essential, or there was little transforming DNA. Transformants were grown in SC-ura medium overnight at 30 °C. A few microlitres of cell suspension were then transferred to selective growth medium using a 96-pin device for phenotypic macroarray analysis.

Cluster analysis

A combination of *k*-means clustering[25] and a heuristic algorithm was used to cluster the phenotypic macroarray data. The data were arranged into a 7,847 by 20 matrix $\boldsymbol{X}$, with each row representing one transformant and each column representing one growth condition. The qualitative phenotype observations 'wild-type', 'weak', 'medium' and 'strong' were mapped onto the numerical scores 1, 3, 9 and 27, which reflect the different weights of the observations. The matrix rows were clustered based on the Euclidean distance between them, which is defined as $d_{ij} = \sqrt{\Sigma_{k=1}^{N}(x_{ik} - x_{jk})^2}$ where x_{ik} and x_{jk} are the elements in rows i and j and column k of matrix $\boldsymbol{X}$ and $N = 20$. As initial cluster centroids we used 20 data rows that exhibit the phenotype 'strong' in only one of the growth conditions plus one data row with no 'strong' phenotype. The clustering was performed in several iterations until the cluster centroids remained unchanged.

The tree describing the relationship between growth conditions was calculated with a program in the PHYLIP package[26] that performs the Fitch–Margoliash least-squares algorithm[27] on a 20 by 20 matrix describing the Euclidean distance between the columns of $\boldsymbol{X}$. The Euclidean distance between columns is defined as in the formula above with $N = 7,847$, i and j representing columns, and k representing a row.

Production and immunolocalization of HAT-tagged strains

HAT tags were generated *in vivo* by *cre-lox* site-specific recombination[10]. HAT-tagged strains were analysed by indirect immunofluorescence as described[8,10], except that the primary antibody was mouse monoclonal anti-HA 16B12 (MMS101R, BAbCo) and the secondary antibody was Cy3-conjugated goat anti-mouse IgG (Jackson Labs).

Received 2 August; accepted 7 October 1999.

1. Goffeau, A. *et al*. Life with 6000 genes. *Science* **274,** 546–567 (1996).
2. Mewes, H. W. *et al*. Overview of the yeast genome. *Nature* **387,** 7–65 (1997).
3. Hieter, P. & Boguski, M. Functional genomics: it's all how you read it. *Science* **278,** 601–602 (1997).
4. Velculescu, V. E. *et al*. Characterization of the yeast transcriptome. *Cell* **88,** 243–251 (1997).
5. 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).
6. 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).
7. Smith, V., Chou, K. N., Lashkari, D., Botstein, D. & Brown, P. O. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274,** 2069–2074 (1996).
8. Burns, N. *et al*. Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.* **8,** 1087–1105 (1994).
9. Ross-Macdonald, P.-R., 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).
10. Ross-Macdonald, P. *et al*. Methods for large-scale analysis of gene expression, protein localization, and disruption phenotypes in *Saccharomyces cerevisiae*. *Methods Mol. Cell. Biol.* **5,** 298–308 (1995).
11. Cherry, M. *et al*. SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* **26,** 73–79 (1998).
12. Schatz, P. J., Pillus, L., Grisafi, P., Solomon, F. & Botstein, D. Two functional alpha-tubulin genes of the yeast *Saccharomyces cerevisiae* encode divergent proteins. *Mol. Cell. Biol.* **6,** 3711–3721 (1986).
13. Page, B. D. & Snyder, M. CIK1: a developmentally regulated spindle pole body-associated protein important for microtubule functions in *Saccharomyces cerevisiae*. *Genes Dev.* **6,** 1414–1429 (1992).
14. Stearns, T., Hoyt, M. A. & Botstein, D. Yeast mutants sensitive to antimicrotubule drugs define three genes that affect microtubule function. *Genetics* **124,** 251–262 (1990).
15. Evangelista, C. C. Jr, Rodriguez, T. A. M., Limbach, M. P. & Zitomer, R. S. Rox3 and Rts1 function in the global stress response pathway in baker's yeast. *Genetics* **142,** 1083–1093 (1996).
16. White, W. H. & Johnson, D. I. Characterization of synthetic-lethal mutants reveals a role for the *Saccharomyces cerevisiae* guanine-nucleotide exchange factor Cdc24p in vacuole function and Na^+ tolerance. *Genetics* **147,** 43–55 (1997).
17. Lussier, M. *et al*. Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. *Genetics* **147,** 435–450 (1997).
18. Lodi, T., Goffrini, P., Ferrero, I. & Donnini, C. IMP2, a gene involved in the expression of glucose-repressible genes in *Saccharomyces cerevisiae*. *Microbiology* **141,** 2201–2209 (1995).
19. Christman, M. F., Dietrich, F. S. & Fink, G. R. Mitotic recombination in the rDNA of *S. cerevisiae* is suppressed by the combined action of DNA topoisomerases I and II. *Cell* **55,** 413–425 (1988).
20. Shore, D. & Nasmyth, K. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell* **51,** 721–732 (1987).

21. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282,** 699–705 (1998).
22. Engebrecht, J. & Roeder, G. S. Mer1, a yeast gene required for chromosome pairing and genetic recombination, is induced in meiosis. *Mol. Cell. Biol.* **10,** 2379–2389 (1990).
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989).
24. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
25. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data* (Wiley, New York, 1990).
26. Felsenstein, J. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5,** 164–166 (1989).
27. Fitch, W. M. & Margoliash, E. Construction of phylogenetic trees. *Science* **155,** 279–284 (1967).
28. Adams, A., Gottschling, D. E., Kaiser, C. A. & Stearns, T. *Methods in Yeast Genetics* (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1997).
29. Goode, B. L. *et al.* Coronin promotes the rapid assembly and cross-linking of actin filaments and may link the actin and microtubule cytoskeletons in yeast. *J. Cell Biol.* **144,** 83–89 (1999).
30. Lee, W. C., Xue, Z. X. & Melese, T. The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. *J. Cell Biol.* **113,** 1–12 (1991).

Acknowledgements

P.S.R.C. is supported by the Fundacao de Amparo a Pesquisa do Estado de Sao Paulo, Brazil. This work was supported by an NIH grant (to G.S.R. and M.S.).

Correspondence and requests for materials should be addressed to M.S. (e-mail: michael.snyder@yale.edu).

Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast

John J. Wyrick*†, Frank C. P. Holstege*, Ezra G. Jennings*†, Helen C. Causton*, David Shore‡, Michael Grunstein§, Eric S. Lander*† & Richard A. Young†

* *Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA*
† *Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
‡ *Department of Molecular Biology, University of Geneva, Geneva 4, CH-1211, Switzerland*
§ *Department of Biological Chemistry, UCLA School of Medicine and Molecular Biology Institute, University of California, Los Angeles, California 90095, USA*

Eukaryotic genomes are packaged into nucleosomes, which are thought to repress gene expression generally[1–3]. Repression is particularly evident at yeast telomeres, where genes within the telomeric heterochromatin appear to be silenced by the histone-binding silent information regulator (SIR) complex (Sir2, Sir3, Sir4) and Rap1 (refs 4–10). Here, to investigate how nucleosomes and silencing factors influence global gene expression, we use high-density arrays to study the effects of depleting nucleosomal histones and silencing factors in yeast. Reducing nucleosome content by depleting histone H4 caused increased expression of 15% of genes and reduced expression of 10% of genes, but it had little effect on expression of the majority (75%) of yeast genes. Telomere-proximal genes were found to be de-repressed over regions extending 20 kilobases from the telomeres, well beyond the extent of Sir protein binding[11,12] and the effects of loss of Sir function. These results indicate that histones make Sir-independent contributions to telomeric silencing, and that the role of histones located elsewhere in chromosomes is gene specific rather than generally repressive.

We used a yeast strain (*Saccharomyces cerevisiae* UKY403) in which the sole copy of histone H4 is under the control of the *GAL1* promoter[13–15] to examine the effects of nucleosome depletion on expression of all protein-coding genes. Nucleosome depletion begins about one hour after switching from galactose to glucose medium, leading to a reduction in the nucleosomal content of yeast cells by around 50% within six hours[14,15]. Duplicate cultures of strain UKY403 and a control strain with a wild-type histone H4 gene were switched to glucose medium and harvested over a time course and labelled antisense RNA (cRNA) prepared from purified messenger RNA was hybridized to Affymetrix GeneChip arrays[16]. Interactive databases supporting this work, details of the experimental reagents and approaches and additional information can be obtained on the World-Wide Web at http://web.wi.mit.edu/young/chromatin/. For a summary of the data sets described here, see Supplementary Information.

The results of the time course of histone depletion are summarized in Fig. 1. Of the 5,894 genes whose mRNA levels were detected at the 6-h time point, the expression of 75% was not significantly altered, the expression of 15% increased more than three-fold, and the expression of 10% decreased more than three-fold during histone H4 depletion. It was notable that expression of most yeast genes was unaffected by histone depletion. Previous studies have shown that nucleosome loss in histone H4 depletion experiments results in a general reduction in nucleosome density[15], and this reduction occurs at all individual genes tested[14,17,18], so it is likely that most genes experience a significant reduction in nucleosome density. Reduced nucleosome density may not affect expression of genes that are already induced; inducible genes have been found to be sensitive to nucleosomes loss only when in their uninduced state[13,14]. It is also possible that nucleosomes have minimal roles in regulation of many genes; transcriptional activators and repressors may be dominant in regulation of such genes, as in prokaryotic cells.

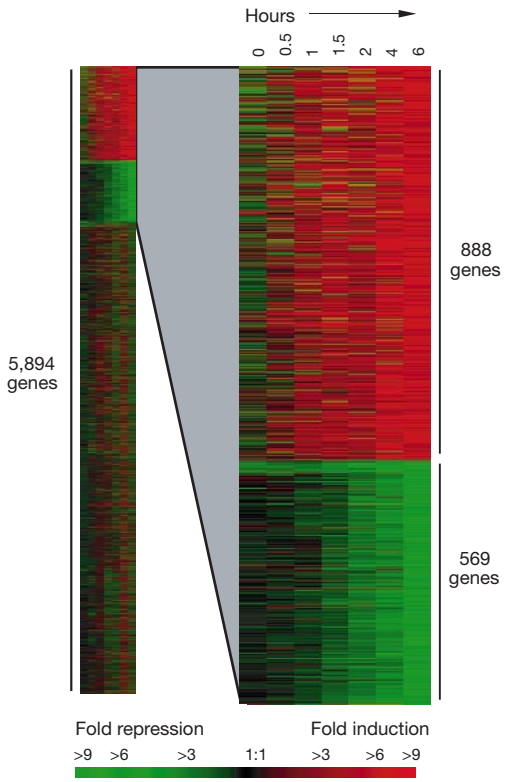

Figure 1 Effect of histone H4 depletion on genome-wide expression. The data for each gene are displayed as the ratio of mRNA in histone H4-depleted cells to that of wild-type cells harvested 0, 0.5, 1, 1.5, 2, 4 and 6 h after a shift from galactose to glucose medium. Each horizontal strip represents a single gene. The fold change is represented by a colour (see colour scale). The left panel shows genes whose mRNA levels increase (top 15% of diagram), genes whose mRNA levels decrease (next 10%), and genes which were relatively unaffected over the time course (bottom 75% of diagram). In the panel on the right, the subset of genes whose mRNA levels increased or decreased more than 3-fold at the 6-h time point are displayed. See ref. 30 for the display program.