Widespread aneuploidy revealed by DNA microarray expression profiling

Timothy R. Hughes, Christopher J. Roberts, Hongyue Dai, Allan R. Jones, Michael R. Meyer, David Slade, Julja Burchard, Sally Dow, Teresa R. Ward, Matthew J. Kidd, Stephen H. Friend & Matthew J. Marton

Expression profiling using DNA microarrays holds great promise for a variety of research applications, including the systematic characterization of genes discovered by sequencing projects[1,2]. To demonstrate the general usefulness of this approach, we recently obtained expression profiles for nearly 300 *Saccharomyces cerevisiae* deletion mutants[3]. Approximately 8% of the mutants profiled exhibited chromosome-wide expression biases, leading to spurious correlations among profiles. Competitive hybridization of genomic DNA from the mutant strains and their isogenic parental wild-type strains showed they were aneuploid for whole chromosomes or chromosomal segments. Expression profile data published by several other laboratories also suggest the use of aneuploid strains. In five separate cases, the extra chromosome harboured a close homologue of the deleted gene; in two cases, a clear growth advantage for cells acquiring the extra chromosome was demonstrated. Our results have implications for interpreting whole-genome expression data, particularly from cells known to suffer genomic instability, such as malignant or immortalized cells.

Using a two-colour competitive hybridization DNA microarray protocol[4,5], we recently generated expression profiles for nearly 300 *S. cerevisiae* deletion mutants[3], mostly obtained through the *Saccharomyces* Genome Deletion Consortium[6]. We observed an unexpected transcript profile similarity (r=0.63; Fig. 1*a*) between mutants harbouring null mutations in *ERG4* and *ECM18* (refs 7,8). Many of the shared upregulations corresponded to genes located on chromosome VII, but not other chromosomes (Fig. 1*b*,*c*). A plot of the mean of the expression ratios for all genes on a particular chromosome revealed that, on average, the expression of all genes on chromosome VII was higher in the *erg4Δ* and *ecm18Δ/ecm18Δ* mutants than in the parental wild-type control with which the mutant was compared (Fig. 1*d*,*e*, red circles). To determine whether this increased expression could be explained by increased gene dosage, genomic DNA from the mutant and parental wild-type strains was isolated, labelled and hybridized to DNA microarrays, and the results plotted in the same manner (Fig. 1*d*,*e*, blue squares). These data indicate that the mutants possess more genomic DNA from chromosome VII than does the wild-type control. Because the elevated gene expression and genomic DNA ratios include essentially all genes on the chromosome, the simplest model explaining these observations is that the mutant strains contain an additional copy or copies of chromosome VII.

The discovery of a spurious correlation resulting from aneuploidy in two independent yeast mutants not known to suffer chromosome instability prompted a search for additional examples of aneuploidy in our collection of expression profiles[3]. Plots of the mean expression ratio for each chromosome for all mutants profiled revealed that expression profiles from approximately 8% of the mutants (22/290) contained at least 1 chromosome that displayed a mean chromosomal ratio bias greater than 0.1 in log space and that

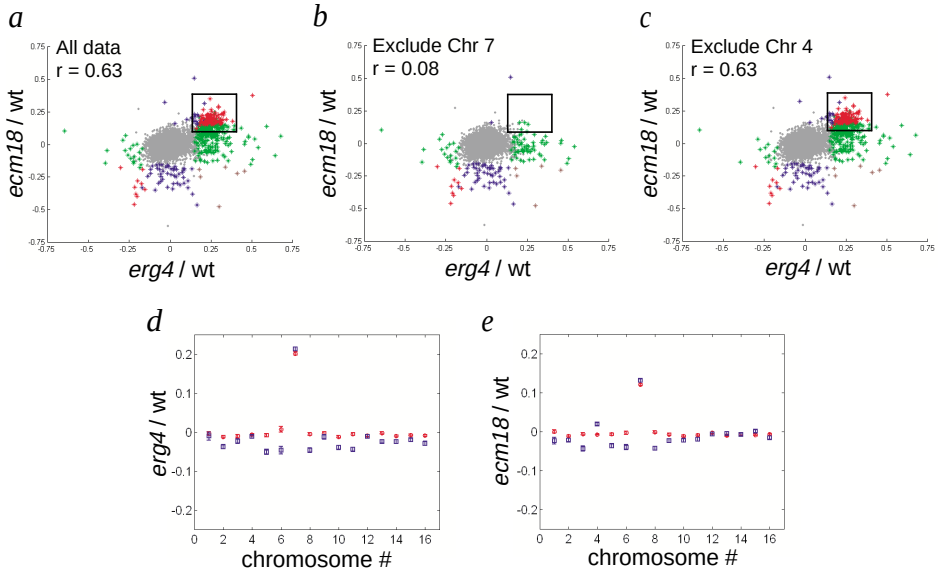


Fig. 1 Detection of a chromosome VII expression bias in *erg4Δ* and *ecm18Δ/ecm18Δ* mutants by expression profiling, and confirmation of aneuploidy by two-colour hybridization of genomic DNA to DNA microarrays. ***a***, Scatter plot comparing the $\log_{10}$ (expression ratios) from *erg4Δ* and *ecm18Δ/ecm18Δ* mutants compared with wild-type control in a two-colour cDNA microarray hybridization[3]. Genes flagged as statistically significantly regulated at the 99% confidence level in both experiments, in only the *ecm18Δ/ecm18Δ* experiment, or in only the *erg4Δ* experiment are denoted by red, blue and green dots, respectively. The correlation coefficient for all genes is r=0.63. The boxes in (*a*–*c*) highlight dots representing genes primarily from chromosome VII. ***b***, The same scatter plot with genes on chromosome VII excluded from analysis. Correlation coefficient with these genes excluded is r=0.08. ***c***, Same scatter plot with genes on chromosome IV excluded as a negative control. Correlation coefficient with these genes excluded is r=0.63. The mean of the $\log_{10}$ (expression ratios) of all genes on an individual chromosome (red circles) and the mean of the $\log_{10}$ (genomic content signal ratios) of all genes on an individual chromosome (blue squares) are shown in the *erg4Δ* (***d***) or *ecm18Δ /ecm18Δ* (***e***) strains. Error bars represent error of the mean $\log_{10}$(ratio). Expression ratios from chromosome VII were 58% higher in *erg4Δ* and 35% higher in *ecm18Δ/ecm18Δ* compared with parental wild-type control strains. Genomic DNA ratios from chromosome VII were 66% higher in *erg4Δ* and 41% higher in *ecm18Δ /ecm18Δ* compared with parental wild-type strains.

Rosetta Inpharmatics, Inc., Kirkland, Washington, USA. Correspondence should be addressed to M.J.M. (e-mail: mmarton@rii.com).

Table 1 • Aneuploidy among 25 yeast mutants

Genotype	Rosetta strain no.	Di- or trisomy chrom. no.	Di- or trisomy s.d.	Monosomy chrom. no.	Monosomy s.d.	Possible explanation
Single aneuploidy						
ecm1Δ/ecm1Δ	2311	3	21	–	–	–
ecm18Δ/ecm18Δ	4719	7	39	–	–	–
erg4Δ	7363	7	46	–	–	–
ste20Δ/ste20Δ	2012	11	27	–	–	–
rml2Δ/rml2Δ	1852	13	22	–	–	–
rpd3Δ/rpd3Δ	320	13	23	–	–	–
yhr011wΔ/yhr011wΔ	2018	14	30	–	–	–
pfd2Δ/pfd2Δ	1778	14	37	–	–	–
yor051cΔ/yor051cΔ	2083	14	38	–	–	–
mcm1Δ/MCM1	120	–	–	3	15	–
mcm1Δ/MCM1	121	–	–	3	16	–
yap3Δ/yap3Δ	2010	–	–	3	24	–
yor080wΔ/yor080wΔ	2103	–	–	3	15	–
Multiple aneuploidy						
bim1Δ/bim1Δ	406	15	26	1	11	spindle defects
bub1Δ	277	2, 10	36, 26	–	–	mitotic checkpoint defect
bub3Δ/bub3Δ	1040	2, 13	16, 11	1	13	mitotic checkpoint defect
sin3Δ/sin3Δ	3432	5, 11	14, 16	–	–	–
Aneuploidy potentially resulting from selection for deleted gene						
rnr1Δ	777	9	10	–	–	selection for *RNR3*
rpl27aΔ/rpl27aΔ	2017	4	30	–	–	selection for *RPL27B*
rpl34aΔ/rpl34aΔ	382	9	21	–	–	selection for *RPL34B*
rps24aΔ/rps24aΔ	985	9	12	–	–	selection for *RPS24B*
rps27bΔ/rps27bΔ	2024	11	30	–	–	selection for *RPS27A*
Segmental aneuploidy						
rad27Δ/rad27Δ	1184	–	NA	18 ORFs	NA	enhanced recombination
rpl20aΔ/rpl20aΔ	2373	56 ORFs	NA	–	NA	selection for *RPL20B*
top3Δ	9379	28 ORFs	NA	–	NA	strain construction

Data[3] were analysed by determining the mean of the $\log_{10}$ (expression ratios) for genes grouped by chromosome. Expression profiles that exhibited a mean chromosomal expression ratio greater than 0.1 ($|\log_{10}(\text{ratio})|>0.1$) and whose offset was at least 10 s.d. from the mean ($P<10^{-20}$) were identified. A strain is listed above only if its genomic DNA hybridization data were consistent with aneuploidy. Not listed are several mutants (*mad1Δ*, *cin8Δ* and *ase1Δ*) for which genomic content data did not confirm aneuploidy predicted by the expression bias; presumably this is because these mutants are known or suspected to have increased rates of chromosome loss that might result in individual colonies with different aneuploidies.

was at least 10 s.d. from the mean. We confirmed each case by hybridizing genomic DNA from the mutant strains to microarrays (Table 1). Although several of these mutants (*bub1*, *bub3* and *bim1*) are known to have defects in chromosome segregation[9,10], most are not thought to be directly involved in genome stability.

To determine whether the high frequency of aneuploidy we observed was peculiar to our strains or a more general phenomenon, we examined all publicly available *S. cerevisiae* expression profiling data. We found several cases of chromosome-wide expression biases, including the published *tup1Δ* data[4] (Fig. 2*a*), *rpb1Δ187* data[11] (Fig. 2*b*) and *hhf2* depletion expression data[12] (Fig. 2*c*). In addition, an expression profile of a *pip2Δ oaf1Δ* double mutant was recently determined by SAGE analysis[13], and although the number of sequence tags for the mutant is small, there appears to be a chromosome-wide expression bias in that data as well (Fig. 2*d*). Because genomic DNA hybridizations were not performed on the strains used in these expression analyses, it is not possible to distinguish whether the mutant or the wild-type control exhibited the aneuploidy, or to rule out other explanations.

In five of our aneuploid mutants (Table 1), the additional chromosome harboured a gene encoding a highly related (80–99% identical) protein. For example, the *rps24aΔ/rps24aΔ* and *rnr1Δ* strains profiled both contained extra copies of chromosome IX, which contains *RPS24B* (97% identical to *RPS24A*) and *RNR3* (80% identical to *RNR1*). In all five cases, the deletions resulted in a slow-growth phenotype (data not shown), suggesting that gain of the entire chromosome may have been a result of a selection for increased growth rate by increasing gene dosage of the paralogue of the deleted gene. When we streaked slow-growing colonies of our *rps24aΔ/rps24aΔ* and *rnr1Δ* mutants for single colonies on solid medium, fast-growing colonies were seen (Fig. 3*a*). Comparative hybridization of genomic DNA from pooled large colonies versus pooled small colonies for each mutant revealed that the large

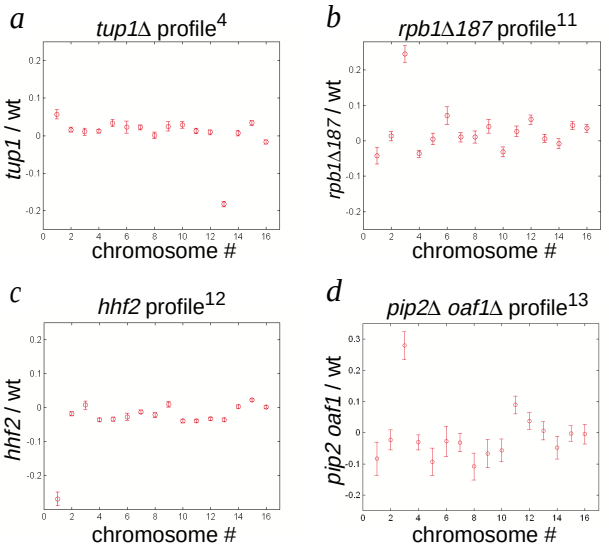


Fig. 2 Expression profiling data in the literature consistent with aneuploidy. *a–d*, The mean of the $\log_{10}$ (expression ratios) of all genes on an individual chromosome in published *tup1Δ* (***a***; ref. 4; chromosome XIII mean $\log_{10}(\text{ratio})=0.18$; $P<10^{-123}$), *rpb1Δ187* (***b***; ref. 11; chromosome III mean $\log_{10}(\text{ratio})=0.25$, $P<10^{-23}$), *hhf2* (***c***; ref. 12; chromosome I mean $\log_{10}(\text{ratio})=0.27$, $P<10^{-33}$) and *pip2Δ oaf1Δ* (***d***; ref. 13; chromosome III mean $\log_{10}(\text{ratio})=0.29$, $P<10^{-10}$) strains. The size of the error bar in (***a–c***) is the error of the mean $\log_{10}$ (ratio), computed from the spread of the data, taking into account the error of each point and the number of data points. The error bar in (***d***) is computed from the square root of the number of tags.

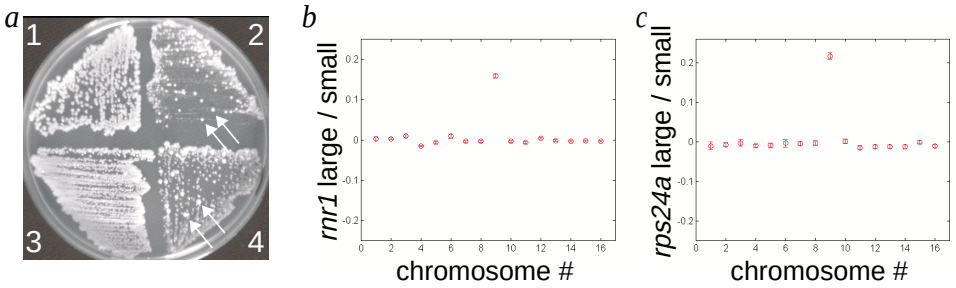


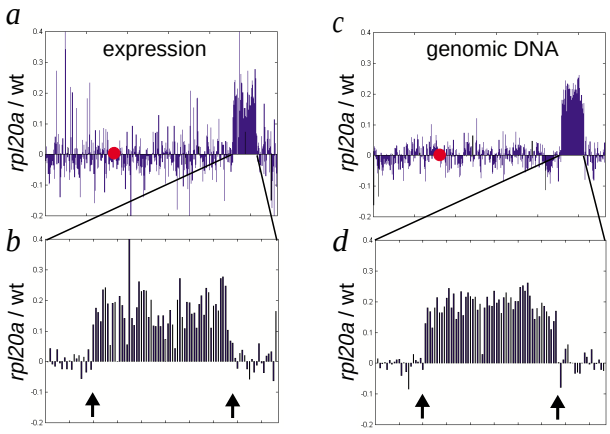
Fig. 3 Selection for aneuploidy in *rnr1Δ* and *rps24aΔ/rps24aΔ* mutants. ***a***, Slow-growing *rnr1Δ* (sector 2) or *rps24aΔ/rps24aΔ* (sector 4) cells or their isogenic parental wild-type cells (sectors 1 and 3) were streaked onto solid medium and incubated at 30 °C for 2–4 d. Fast-growing colonies, shown to harbour an extra copy of chromosome IX, are highlighted by arrowheads. *b*,*c*, The mean of the $\log_{10}$ (genomic content signal ratios) of all genes on an individual chromosome in large *rnr1Δ* colonies compared with small *rnr1Δ* colonies (***b***) or large *rps24aΔ/rps24aΔ* colonies compared with small *rps24aΔ/rps24aΔ* colonies (***c***). Error bars represent error of the mean $\log_{10}$(ratio).

colonies contained an additional copy or copies of chromosome IX (Fig. 3*b*,*c*), suggesting the extra chromosome provided a selective growth advantage.

We also identified expression biases within chromosomes (Table 1). By plotting the expression ratio of each gene as a function of its chromosomal location, we noted an expression bias in a 56-ORF region on the right arm of chromosome XV in our *rpl20aΔ/rpl20aΔ* expression profile (Fig. 4*a*). The genomic content data (Fig. 4*c*) precisely mirror the expression data in this region, suggesting the duplication can completely explain the expression bias. This region (between ORFs *YOR290c* and *YOR343c*) is precisely flanked by retrotransposon LTRs (Fig. 4*b*,*d*) and contains *RPL20B*, which encodes a protein with 99% identity to Rpl20ap. It is tempting to speculate that the duplication was a result of a homologous recombination event and selection for increased dosage of *RPL20B*.

The presence of chromosome-wide expression biases in data from three other laboratories as well as in 8% of our strains indicates that whole-chromosome aneuploidy is widespread in laboratory yeast strains. Considering the number of cell divisions involved in strain construction and storage, however, the frequency of aneuploidy in our strains (excluding mutants with a clear growth defect or chromosome missegregation phenotype) agrees with previous estimates of mitotic chromosome loss[14] (data not shown). Our results have several implications. First, the data show that the mRNA abundance of nearly every gene on trisomic or monosomic chromosomes is altered, suggesting that in yeast there is no global dosage-compensation mechanism to normalize expression from each gene (or chromosome). Previous results[15] suggest that human genes may also generally lack homeostatic expression mechanisms. An expression profile therefore serves as a tool for the detection of aneuploidy, including even small deletions (data not shown) or duplications. Second, the fact that alterations in DNA copy number can lead to spurious correlations between expression profiles poses a potential hazard in drawing conclusions from gene-expression data, particularly from cell lines or tumour cells that have unstable genomes.

Aneuploidy may complicate use of a public GenBank-like resource of expression data: although biased chromosomes could be detected and masked, the downstream consequences of aneuploidy are unpredictable. For example, loss of one copy of chromosome III, which contains the heteroallelic *MAT*a/*MAT*α mating control locus, resulted in a false correlation between our *mcm1Δ/MCM1* and *yor080wΔ/yor080wΔ* mutants (Fig. 5). In contrast to the *erg4Δ ecm18Δ/ecm18Δ* correlation, which was dependent on a large number of small-magnitude expression changes arising from genes on the duplicated chromosome, the *mcm1Δ/MCM1-yor080wΔ/yor080wΔ* correlation was mostly due to changes in the expression of genes present on chromosomes other than the aneuploid chromosome, as might be expected when a key transcriptional regulator is affected directly by the aneuploidy (Fig. 5*a*). Finally, the potential for aneuploidy to surreptitiously mask or alter phenotypes of deleterious mutations is a more general concern for geneticists when interpreting their results. The observation that very large duplications are recovered as dominant suppressors of single-gene mutations might suggest that the prevalence of such duplications in nature, in evolution[16,17] and in cancer cells may be the result of a need to compensate for loss of function of other genes.



Methods

Strains and expression profiling. The genotypes of the nearly 300 strains for which we generated an expression profile, along with supporting expression and genomic DNA hybridization data, can be found at Rosetta's web site (http://www.rii.com/tech/pubs/natgen_hughes.htm and http://www.rii.com/tech/pubs/cell_hughes.htm). Essentially all are derived from strain BY4743 (*MAT*a/*MAT*α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 ura3Δ0/ura3Δ0 +/met15Δ0 +/lys2Δ0*), the parental strain for the international *Saccharomyces* Genome Deletion Consortium[6] (http://www sequence.stanford.edu/group/yeast_deletion_project/deletions3.html). We generated expression profiles as described[3].

Genomic DNA extraction, labelling and hybridization to microarrays. We extracted genomic DNA from 5-ml saturated cultures grown in YPD medium as described[18]. Genomic DNA (2 µg) was denatured, annealed to random hexamers (1 µg) and labelled at 37 °C in 15-µl reactions (containing 1×NEB buffer 2, 7 units Klenow fragment of DNA polymerase I, 500 µM dATP, dCTP and dGTP, 200 µM dUTP and 100 µM Cy-dUTP). We conducted cDNA microarray production, hybridizations, washing and image analysis using a two-colour procedure as above. We scanned the arrays on either a General Scanning ScanArray3000 or a Genetic Microsystems 418 Array Scanner. To look for aneuploidy in small colonies versus large colonies, we streaked cells

Fig. 4 Segmental aneuploidy in an *rpl20aΔ/rpl20aΔ* mutant. ***a***,***b***, Chromosomal view of $\log_{10}$ (expression ratios) of all genes on chromosome XV in the *rpl20aΔ/rpl20aΔ* mutant expression profile plotted as a function of chromosomal location. ***c***,***d***, Chromosomal view of $\log_{10}$ (genomic content signal ratios) of all genes on chromosome XV in the *rpl20aΔ/rpl20aΔ* mutant plotted as a function of chromosomal location. Vertical bars represent the logarithmic expression ratio of an individual gene in the *rpl20aΔ/rpl20aΔ* mutant relative to the wild-type control. The leftmost and rightmost bars in (*a*) and (*c*) represent the ORFs nearest the left and right telomeres (*YOL166c* and *YOR394w*), respectively, and the red circle denotes the position of the centromere. Arrows denote the position of LTRs of retrotransposons.

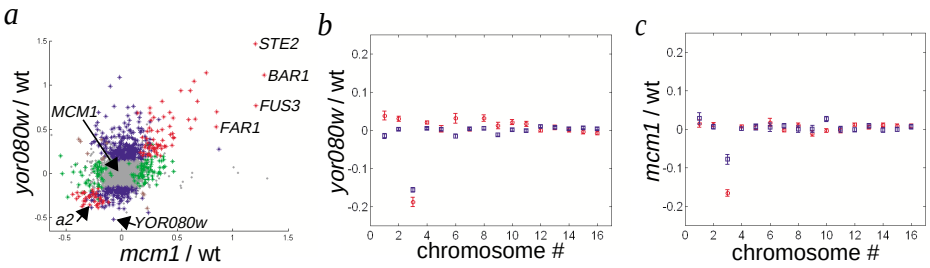


Fig. 5 Spurious correlation between two mutants displaying a large transcriptional signature resulting from aneuploidy. ***a***, Scatter plot comparing the $\log_{10}$ (expression ratios) from *mcm1Δ/MCM1* and *yor080wΔ/yor080wΔ* mutants. Correlation coefficient for all genes is r=0.55. Genes flagged as statistically significantly regulated at the 99% confidence level in both experiments, in only the *yor080wΔ/yor080wΔ* experiment or in only the *mcm1Δ/MCM1* experiment are denoted by red, blue and green dots, respectively. ***b***,***c***, The mean of the $\log_{10}$ (expression ratios) of all genes on an individual chromosome (red circles) and the mean of the $\log_{10}$ (genomic content signal ratios) of all genes on an individual chromosome (blue squares) in the *yor080wΔ/yor080wΔ* (***b***) and *mcm1Δ/MCM1* (***c***) mutants. Error bars represent error of the mean $\log_{10}$(ratio). The false correlation between *yor080wΔ/yor080wΔ* and *mcm1Δ/MCM1* profiles was not affected by masking genes from the monosomic chromosome (r=0.53; data not shown). Chromosome III harbours the heteroallelic mating type (*MAT*) locus: in wild-type diploids, one copy of chromosome III harbours *MAT***a**, and the second copy harbours *MATα*. *MATα* encodes the α2 transcriptional regulator, which represses mating type **a**-specific (for example, *STE2*, *BAR1*) and haploid-specific (for example, *FUS3*, *FAR1*) genes[24]. Thus, the observed expression changes would be expected from a diploid monosomic specifically for the *MAT***a** chromosome III.

on five plates, and picked ~2,000 small colonies or 50 large colonies by toothpick and resuspended them directly in lysis buffer[18] for DNA extraction.

Analysis of Rosetta data. The relative expression level of a gene in a mutant relative to that in a wild-type control conveyed as a ratio is called the expression ratio. The correlation plot displays the expression ratio of each gene from one profile plotted versus its expression ratio in a second expression profile. The Cy3 and Cy5 channels are normalized by the mean signal intensity for all yeast ORF spots. Thus, by convention, the mean expression ratio for all spots is unity. The mean chromosomal ratio plots display in logarithmic scale the average of all expression ratios for each individual chromosome. The mean expression ratio for each chromosome is an error-weighted mean of all the ORFs present on that chromosome, with the error calculated based on the slide quality and the individual spot intensity. A chromosome was flagged as having a statistically significant chromosome-wide expression bias if the mean chromosomal ratio had an offset of greater than 0.1 in log space and was at least 10 s.d. from the mean ($P<10^{-20}$). P values were calculated from the number of standard deviations from the mean, assuming a Gaussian distribution, which was verified by analysis of 63 wild-type versus wild-type control experiments[3]. The estimated systematic bias of each chromosome with respect to the mean is at the level of 0.0016 of $\log_{10}$(ratio). The error bar of the mean ratio in log space is computed from the spread of the data, taking into account the error of each point and the number of data points. To explore expression profiling data for potential occurrences of segmental aneuploidy, we scanned data for instances in which 4 or more non-overlapping chromosomally adjacent genes were all up- or downregulated at a 0.05 significance threshold. We identified 22 potential cases of segmental aneuploidy (that is, there were 22 cases in which at least four adjacent genes were apparently coordinately regulated). We tested four cases and each was confirmed experimentally by genomic DNA hybridization. The *rpl20aΔ/rpl20aΔ* mutant contained a 56-ORF duplication from *YOR290c* to *YOR343c*, which in the wild type is flanked by retrotransposon LTRs and a Ty2 transposon on the centrometric and telomeric sides, respectively. The *top3Δ* mutant contained a 28-ORF duplication from *YLR228c* to *YOR256w* and in the wild type is flanked by LTRs and a Ty1 transposon on the centromeric and telomeric sides, respectively. The genomic DNA hybridization of the *rad27Δ/rad27Δ* mutant was consistent with an 18-ORF deletion from *YDR367w* to *YDR385w*. The centromeric side of the duplicated region is flanked by two LTRs and a Ty1 transposon, whereas the telomeric side has no obvious sequence features.

Analysis of data in the public domain. We downloaded data for the *tup1Δ* deletion mutant[4] (http://cmgm.stanford.edu/pbrown/explore/tup1data.txt) and used mutant/wild-type control expression ratios without applying an intensity threshold. We applied our error model to all public domain data, assuming similar data quality. Because the analysis suggested that the wild-type strain was the source of the aneuploidy in the *tup1Δ* strain, expression data were also downloaded for another experiment using a wild-type strain[4] (wild type compared with overexpression of *YAP1*). After normalization by total signal intensity, the mean chromosomal expression of the wild-type channels from the two independent hybridizations were compared, and were consistent with the hypothesis that the wild-type control in the *tup1* data had an additional copy or copies of chromosome XIII. This analysis, although suggestive, has limitations because it assumes the two independent hybridizations have correlated errors, that is, it treats a competitive two-colour experiment as if it were a one-colour experiment. We downloaded expression profiles for 16 mutants[11,12] including *rpb1Δ187* and *hhf2* (http://web.wi.mit.edu/young/pub/regulation.html). Most of these mutants were profiled in duplicate (that is, two mutant hybridizations and two wild-type control hybridizations); genes called 'absent' (that is, genes not expressed) in any of the four hybridizations were excluded from our analysis. As the *hhf2* strain (whose expression profile suggested the loss of chromosome I) was haploid, the simplest explanation is that the control strain was the source of the aneuploidy. Chromosome-wide expression biases were detected only in the *rpb1Δ187* and *hhf2* profiles. We also downloaded the *pip2Δ oaf1Δ* double mutant SAGE expression data[13] (http://www.molbiolcell.org/cgi/content/full/10/6/1859/DC1). As there were nearly three times as many tags for the wild-type control than for the mutant (14,367 and 5,419 tags for the wild type and mutant, respectively), we first normalized the data for number of tags sequenced. As suggested by the authors, tags more than 500 bases upstream of the 3′ end of an ORF were excluded from analysis. To generate the plot for each chromosome, we divided the sum of all tags from the mutant by the sum of all tags from the wild type, because random statistical fluctuations of values near zero makes examining expression ratios inappropriate, and because only considering genes having two or more tags excluded most genes (only 1,756 and 616 tags remain for the wild type and mutant, respectively). The error bars are estimated by the square root of the sample size. Several other studies contained data suggestive of aneuploidy, but the expression biases did not meet the criteria described above (0.1 bias in log space and at least 10 s.d. from the mean). For example, we noted an expression bias in data from strain E1, which underwent adaptive evolution during ~500 generations in glucose-limited media[19] (chromosome XIV mean $\log_{10}$(ratio)=0.07; 10 s.d. from the mean), and in two tetraploid strains[20] (chromosome VI of strain *MATα/MATα/MATα/MATα* had a mean $\log_{10}$(ratio)=0.16 and was 7 s.d. from the mean; chromosome I of strain *MATa/MATa/MATα/MATα* had a mean $\log_{10}$(ratio)= 0.17 and was 6 s.d. from the mean). We did not detect a chromosome-wide expression bias in the published *ndt80* expression profile[21]. Furthermore, expression profiles in which the same strain is profiled under two or more conditions, such as kinetic analyses (during diauxic shift[4], sporulation induction[21] or cell-cycle progression[22]), drug treatments (methyl methanesulfonate treatment[23]) or gene induction experiments (*GAL-CLB2* and *GAL-CLN3* experiments[22]) are not expected to be susceptible to this type of problem and did not exhibit chromosome-wide expression biases.

Acknowledgements

We thank C. Boone, A. Murray, F. Spencer, N. Hastie, L. Hartwell, R. Stoughton and D. Shoemaker for their comments on the manuscript; P. Paddison for discussions on the implications of aneuploidy in human cancer; and our colleagues in the academic community for making full data sets publicly available. This work was supported by Rosetta Inpharmatics, Inc.

Received 7 February; accepted 24 April 2000.

1. Pennisi, E. Worming secrets from the *C. elegans* genome. *Science* **282**, 1972–1974 (1998).
2. Somerville, C. & Somerville, S. Plant functional genomics. *Science* **285**, 380–383 (1999).
3. Hughes, T.R.*et al.* Functional discovery via a compendium of expression profiles. *Cell* (in press).
4. 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).
5. Marton, M.J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**, 1293–1301 (1998).
6. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
7. Lai, M.H. *et al.* The identification of a gene family in the *Saccharomyces cerevisiae* ergosterol biosynthesis pathway. *Gene* **140**, 41–49 (1994).
8. Lussier, M. *et al.* Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. *Genetics* **147**, 435–450 (1997).
9. Hoyt, M.A., Totis, L. & Roberts, B.T. *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell* **66**, 507–517 (1991).
10. Schwartz, K., Richards, K. & Botstein, D. *BIM1* encodes a microtubule-binding protein in yeast. *Mol. Biol. Cell* **8**, 2677–2691 (1997).
11. Holstege, F.C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
12. Wyrick, J.J. *et al.* Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418–421 (1999).
13. Kal, A.J. *et al.* Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Mol. Biol. Cell* **10**, 1859–1872 (1999).
14. Hartwell, L.H. & Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of S. cerevisiae. *Genetics* **110**, 381–395 (1985).
15. Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
16. Wolfe, K.H. & Shields, D.C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
17. Smith, N.G., Knight, R. & Hurst, L.D. Vertebrate genome evolution: a slow shuffle or a big bang? *Bioessays* **21**, 697–703 (1999).
18. Hoffman, C.S. & Winston, F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* **57**, 267–272 (1987).
19. Ferea, T.L., Botstein, D., Brown, P.O. & Rosenzweig, R.F. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl Acad. Sci. USA* **96**, 9721–9726 (1999).
20. Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S. & Fink, G.R. Ploidy regulation of gene expression. *Science* **285**, 251–254 (1999).
21. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
22. 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).
23. Jelinsky, S.A. & Samson, L.D. Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc. Natl Acad. Sci. USA* **96**, 1486–1491 (1999).
24. Johnson, A.D. Molecular mechanisms of cell-type determination in budding yeast. *Curr. Opin. Genet. Dev.* **5**, 552–558 (1995).