Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis

Elizabeth A. Winzeler,[1*] Daniel D. Shoemaker,[2*] Anna Astromoff,[1*] Hong Liang,[1*] Keith Anderson,[1] Bruno Andre,[3] Rhonda Bangham,[4] Rocio Benito,[5] Jef D. Boeke,[6] Howard Bussey,[7] Angela M. Chu,[1] Carla Connelly,[6] Karen Davis,[1] Fred Dietrich,[8] Sally Whelen Dow,[2] Mohamed El Bakkoury,[9] Françoise Foury,[10] Stephen H. Friend,[2] Erik Gentalen,[11] Guri Giaever,[1] Johannes H. Hegemann,[12] Ted Jones,[1] Michael Laub,[1] Hong Liao,[4] Nicole Liebundguth,[8] David J. Lockhart,[11] Anca Lucau-Danila,[10] Marc Lussier,[7] Nasiha M'Rabet,[3] Patrice Menard,[7] Michael Mittmann,[11] Chai Pai,[1] Corinne Rebischung,[8] Jose L. Revuelta,[5] Linda Riles,[13] Christopher J. Roberts,[2] Petra Ross-MacDonald,[4] Bart Scherens,[9] Michael Snyder,[4] Sharon Sookhai-Mahadeo,[6] Reginald K. Storms,[7] Steeve Véronneau,[7] Marleen Voet,[14] Guido Volckaert,[14] Teresa R. Ward,[2] Robert Wysocki,[10] Grace S. Yen,[1] Kexin Yu,[6] Katja Zimmermann,[12] Peter Philippsen,[8] Mark Johnston,[13] Ronald W. Davis[1†]

The functions of many open reading frames (ORFs) identified in genome-sequencing projects are unknown. New, whole-genome approaches are required to systematically determine their function. A total of 6925 *Saccharomyces cerevisiae* strains were constructed, by a high-throughput strategy, each with a precise deletion of one of 2026 ORFs (more than one-third of the ORFs in the genome). Of the deleted ORFs, 17 percent were essential for viability in rich medium. The phenotypes of more than 500 deletion strains were assayed in parallel. Of the deletion strains, 40 percent showed quantitative growth defects in either rich or minimal medium.

The budding yeast *S. cerevisiae* serves as an important experimental organism for revealing gene function. In addition to carrying out all the basic functions of eukaryotic cells, up to 30% of positionally-cloned genes implicated in human disease have yeast homologs (*1*). Determining the function of all yeast gene products will be an important step toward understanding their function in metazoans and lays the foundation for a more complete comprehension of cellular processes and pathways.

A powerful way to determine gene function is the phenotypic analysis of mutants missing the gene. Several genome-wide approaches have been proposed including genetic footprinting and random mutagenesis (*2, 3*). While genetic footprinting has the advantage that all genes can be tested for their contribution to fitness under a particular growth condition relatively quickly, it has the disadvantage that the mutant strains cannot be recovered. In addition, testing each additional condition is as time-consuming as the first. Random mutagenesis is relatively rapid, but the subsequent matching of phenotypes to genes is slower. In addition, with random approaches a certain fraction of genes may be missed, even with oversampling. These limitations can be overcome by deleting each gene in the genome in a directed fashion and by marking each yeast gene with a molecular "barcode" that allows the phenotypes of the mutant strains to be assayed in parallel.

The precise deletion of yeast genes can be efficiently accomplished using a polymerase chain reaction (PCR)–mediated gene disruption strategy that exploits the high rate of homologous recombination in yeast (*4*). For this method, short regions of yeast sequence [~50 base pairs (bp)] identical to those found upstream and downstream of a targeted gene are placed at each end of a selectable marker gene through PCR. The resulting PCR product, when introduced into yeast cells, can replace the targeted gene by homologous recombination. For most genes, >95% of the resulting yeast transformants carry the correct deletion (*5*). In addition, this method can be modified so as to introduce two molecular barcodes (UPTAG and DOWNTAG) into the deletion strain. The barcodes or "tags" are unique 20-base oligomer (20-mer) sequences that serve as strain identifiers (*6, 7*). We show that these barcodes allow large numbers of deletion strains to be pooled and analyzed in parallel in competitive growth assays. This direct, simultaneous, competitive assay of fitness increases the sensitivity, accuracy and speed with which growth defects can be detected relative to conventional methods.

To take full advantage of this approach and to accelerate the pace of progress, an international consortium was organized to generate deletion strains for all annotated yeast genes. Here, we report the construction of precise start-to-stop codon deletion mutants for 2026 ORFs (*8*).

Genes essential for viability in yeast, in particular those encoding proteins lacking human homologs, have been proposed to be the best targets for antifungal drugs. When spores from the 2026 heterozygous strains were germinated on YPD (yeast extract–peptone–dextrose) media at 30°C haploid deletants could not be recovered for 356 ORFs (see www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html for an exact list) (*9*). Despite the considerable interest in these genes as potential drug targets, only 56% of these ORFs had previously been shown to be essential for viability (*10*). Of the 2026 ORFs analyzed, 1620 were not essential for viability in yeast. For these one additional homozygous and two haploid deletants (Table 1) were also constructed.

A computational Smith-Waterman analysis indicated that 8.5% of the identified nonessential ORFs in the yeast genome have a closely related homolog elsewhere in the ge-

[1]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305–5307, USA. [2]Rosetta Inpharmatics Inc., 12040 115th Street NE, Kirkland, WA 98034, USA. [3]Universite Libre de Bruxelles, Laboratoire de Physiologie Cellulaire et de Genetique des Levures, Campus Plaine, Brussels CP244, Belgium. [4]Department of Molecular, Cellular & Developmental Biology, Yale University, New Haven, CT 06520–8103, USA. [5]Dipartimento de Microbiologia y Genetica, Universidad de Salamanca, Edificio Depatamental 323/CSIC, Campus Miguel de Unamuno, E-37007 Salamanca, Spain. [6]Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, 617 Hunterian Building, 725 North Wolfe Street, Baltimore, MD 21205–2185, USA. [7]Department of Biology, McGill University, Montreal, PQ, Canada H3A 1B1. [8]Biozentrum, Department of Molecular Microbiology, Biozentrum, University of Basel, Switzerland. [9]IRMW-ULB, Avenue E. Gryson, 1, B-1070 Brussels, Belgium. [10]FYSA-UCL, Place Croix du Sud, 2/20, 1348-Louvain-la-Neuve, Belgium. [11]Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051, USA. [12]Institut für Mikrobiologie, Geb. 26.12.01 Raum 64, Universitaetsstrasse 1, D-40225 Düsseldorf, Germany. [13]Department of Genetics, Washington University Medical School, St. Louis, MO 63110, USA. [14]Katholieke Universiteit Leuven, Laboratory of Gene Technology, Kardinaal Mercierlaan 92, B-3001 Leuven, Belgium.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dbowe@cmgm.stanford.edu

nome ($P < 1.0e-150$), whereas only four (1%) of the essential genes (*PYK1, YDR341C*, *PRP22*, and *MYO2*) encoded proteins that were homologous to another predicted protein in the genome. The redundancy may be why more genes in the yeast genome are not essential. The essential genes were distributed fairly evenly across the chromosome but were slightly biased toward being located near other essential genes (60% of essential genes were within 5 kb of another essential gene, whereas 47% of nonessential genes were). Essential genes were generally not found within 50 kb of the telomeres (Fig. 1A). Essentials were also more heavily transcribed. Transcripts were detected for >99% of essential genes versus 90% of nonessential genes (*11*). The average number of transcripts per cell for all essential genes was 70% higher than for all nonessential genes. The functional classification of the essential genes versus the nonessential genes is shown in Fig. 1B.

The phenotypic analysis of the deletion strains, in particular those whose cognate protein is not essential to life, is a formidable task. The role of many genes will likely be manifested only under very specialized growth conditions, necessitating the examination of many different conditions. Previous work demonstrated that the barcodes allowed the relative abundance of their respective strains to be measured when 12 strains were grown competitively for many generations (*6*). The barcoding scheme thus has the potential to accelerate the phenotypic analysis of the deletion strains by allowing the growth rates of all strains to be assayed simultaneously. The first 558 homozygous deletion strains constructed were pooled (*12*) and grown in rich and minimal media for about 60 generations. During this time, aliquots were removed from the two pools. The tags were amplified, and hybridized to high-density arrays containing the tag complements (Fig. 2A) (*13*). The hybridization data were used to calculate the relative growth rates for each deletion mutant in the population (*14*). It was expected that the growth rate for each strain obtained independently with the UPTAG and DOWNTAG signals would agree. For strains in which both the UPTAG and DOWNTAG signal were at least threefold over background, the correlation for growth rates measured with the UPTAG and DOWNTAG were 0.97 in rich medium, and 0.94 in minimal medium (Fig. 2, B and C). The weakest correlations were observed for strains that were the most growth impaired (growth rate <0.6 of that of the wild type) because sufficient signal was detected only for first few time points.

As expected, the strains disappearing at the highest rates in minimal medium but which grew relatively normally in rich medium (at >98% of the pool growth rate) included all of the strains carrying deletions in genes known to be essential for growth in minimal medium including *ade1* [tag average <0.50 in minimal (M); 0.98 in rich (R)]; *arg5,6* (<0.50, M; 0.98, R); *yer068c-a* (<0.50, M; 0.99, R; overlaps *arg5,6*); *ilv1* (<0.50, M; 1.01, R); *arg4* (<0.50, M; 1.0, R); *gcn4* (0.53, M; 0.98, R); *hom3* (0.54, M;1.0, R); and *ade4* (0.56, M; 1.01, R). In addition, the *gyp1* (0.78, M; 0.99, R) deletant showed a minimal medium-specific growth defect (*15*). *GYP1* (*YOR070C*) is a GTPase activating protein for Sec4p, a protein in the secretion pathway (*16*). Mutants of *lys7* also showed a growth defect in minimal medium but also grew somewhat slowly in rich medium (0.52, M; 0.88, R). Six strains exhibiting a specific growth defect in rich medium were also identified. These strains included *cin8* (0.80 R; 0.95, M); *erg6* (0.71, R; 0.96, M); *rpl39* (0.85, R; 0.98, M); *yml193C-A* (0.85, R; 0.98, M; overlaps ribosomal protein *rpl36a*); *esc1* (0.83, R; 0.97, M) and *yml013w* (0.78, R; 0.95, M).

Altogether, almost 40% of the deletants

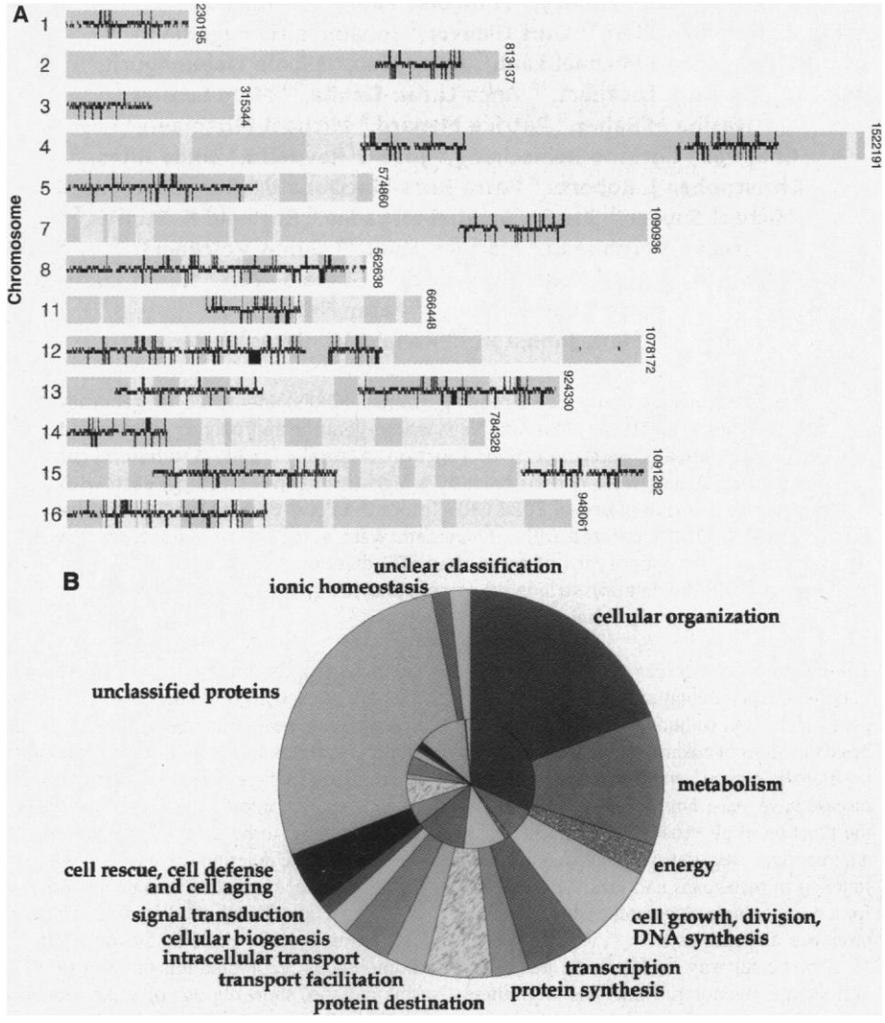

Fig. 1. (**A**) Genomic locations of 1620 nonessential (short black bars) and 356 essential genes (tall black bars). Deletions were generated in consecutive groups on multiple chromosomes. A lighter gray background indicates the location of chromosomal duplication blocks (*23*). For 15 of the 356 essential ORFs, a haploid null mutant had been previously described. These inconsistencies may be due to differences in strain background or in the conditions used for germination of spores. For example, the previously-constructed *yhr101c* (*big1*) and *ybr196c* (*pgi1*) null mutants show glucose sensitivity, and the *ybr256c* (*rib5*) deletion mutant requires a riboflavin supplement for viability (*24*). In several cases, haploid null mutants were reported to have slow-growth phenotypes [*ymr308c* (*PSE1*), *yol022c*, *ypl243c* (*srp68*), *ypl210c* (*srp72*), and *ydr353w* (*trr1*) (*25*)] or to be temperature sensitive [*ydr113c* (*pds1*) and *ygr216c* (*gpi1*) (*26*)]. In some cases, haploid deletions strains were constructed for genes previously determined to be essential by others [*HKR1*, *RNR1*, and *FUN9* (*27*)]. (**B**) Distribution of functional classes of essential (inner circle) and nonessential (outer circle) ORFs using criteria from the Munich Information Centre for Protein Sequences (MIPS) (*10*).

examined showed some sort of growth defect in the competitive growth assay (Fig. 3): 24 (5%) at less than 75% of the pool doubling time; 27 (5%) from 75 to 85%; 80 (15%) at 80 to 98%; and 71 (14%) at 98 to 100%. Strains that grew poorly in rich medium generally also grew poorly in minimal medium aside from the exceptions (*e.g. rnr1, hem14*) described above. The phenotypic profiles for the deletion strains were in good accordance with those obtained using other methods (*17*).

It is often assumed that if a gene is expressed under a particular set of conditions, then that gene is important for growth under those conditions. Deletion of the up-regulated gene would then be expected to cause a decrease in growth. These data offered a unique opportunity to correlate changes in gene expression with deletion phenotypes. The tran-

Table 1. Genotypes of strains used in study. For the YD strains, in a few cases deletions were generated in the BY4730 and BY4739 parent strains. These haploid deletants, as well as the resulting homozygous and heterozygous diploids, are *HIS*$^+$.

Strain	Genotype	Reference
BY4730	*MAT***a** *leu2Δ0 met15Δ0 ura3Δ0*	(*29*)
BY4739	*MATα leu2Δ0 lys2Δ0 ura3Δ0*	(*29*)
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(*29*)
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	(*29*)
BY4743	*MAT***a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 met15Δ0/MET15 lys2Δ0/LYS2 ura3Δ0/ura3Δ0*	(*29*)
YD00000-YD09999	*MAT***a** *orfΔ::kanMX4 his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	This work
YD10000-YD19999	*MATα orfΔ::kanMX4 his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	This work
YD20000-YD29999	*MAT***a**/α *orfΔ::kanMX4/ORF his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 lys2Δ0/LYS2 met15Δ0/MET15 ura3Δ0/ura3Δ0*	This work
YD30000-YD39999	*MAT***a**/α *orfΔ::kanMX4/orfΔ::kanMX4 his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 lys2Δ0/LYS2 met15Δ0/MET15 ura3Δ0/ura3Δ0*	This work

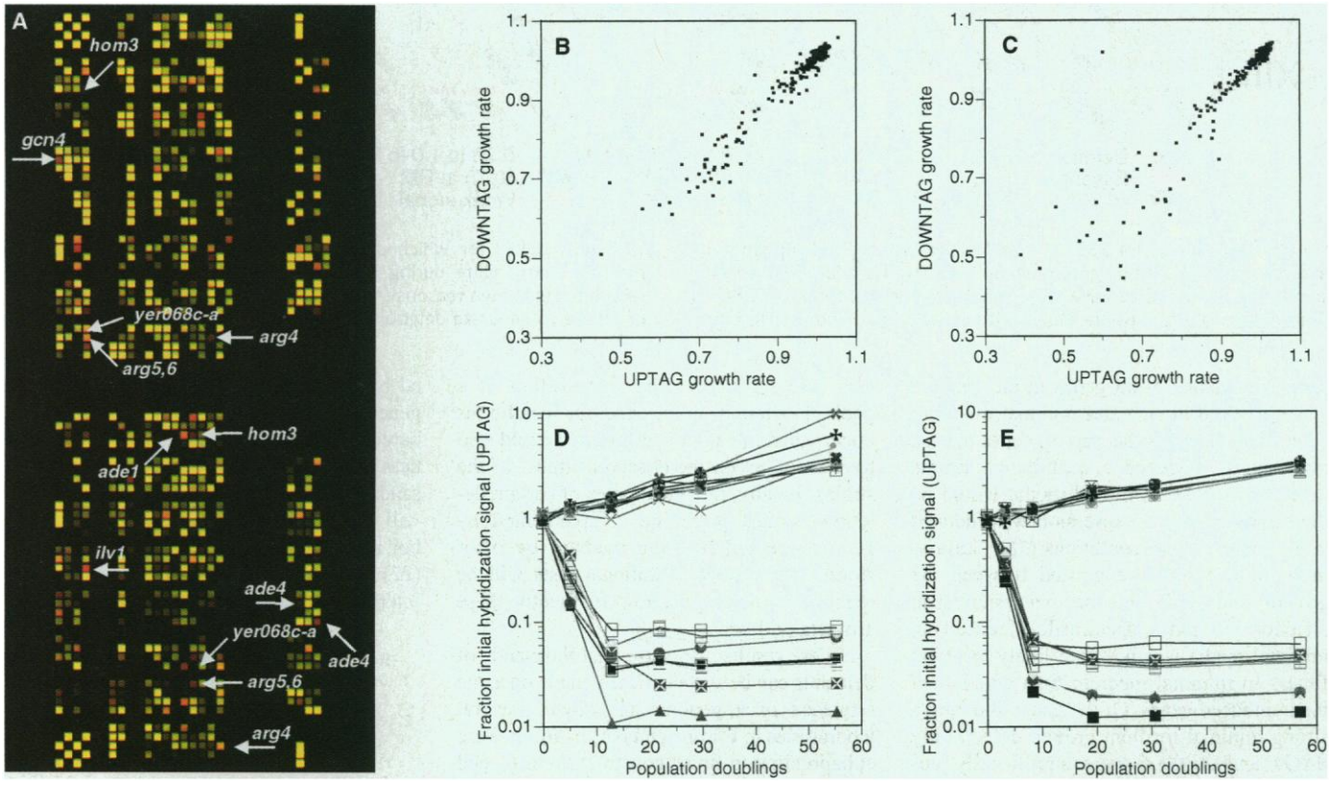

Fig. 2. Analysis of 558 homozygous diploid strains in rich and minimal medium. Pools were grown for ~60 generations in minimal medium (SD) supplemented with histidine, uracil, and leucine, or in rich medium (YPD) at 30°C. The batch-transfer method (*2, 3*) was used to ensure that the cell densities did not exceed 4×10^7 cells/ml during the growth study. At least 500,000 cells were transferred during each dilution to avoid unequal representation of strains. Samples (approximately 5×10^7 cells) were taken at six different time points for each growth study and genomic DNA was extracted. The two tags for each strain were amplified separately with biotin-labeled primers and were hybridized together to the arrays. (**A**) Two-color comparison of scanned images of high-density oligonucleotide arrays hybridized with fluorescently-labeled tags amplified from 558 strains grown for 0 (red channel) or 6 hours (green channel) in minimal medium. Approximately 4000 different features, each containing more than 10^7 20-mers of a specific sequence, were synthesized in a specific physical location on the array by photolithography and photosensitive oligonucleotide chemistry (*6, 28*). Only a portion of the array is shown. The features containing probes to the bar codes for deletion strains that exhibit a growth defect in minimal medium, but not in rich medium, are labeled and indicated with an arrow. The array sequences were assigned sequentially to different deletion strains. A second-generation array has been designed, which contains enough tag complement sequences for every gene in the entire genome. [(B) and (C)] Correlation of growth rate data obtained with UPTAG and DOWNTAG sequences for strains grown in rich (**B**) and minimal (**C**) media. Data are shown for 331 strains (of the 401 strains that contained both an UPTAG and a DOWNTAG sequence) which had $t = 0$ UPTAG and DOWNTAG hybridization signals that were both at least threefold over background. More frequent sampling during the initial growth period should improve the correlation. (**D**) Normalized hybridization intensity data for the 10 slowest-growing (*yer014c-a, rnr1, hem14, mot2, pfk2, rpl27a, yer044c, rp50a, ymr188c, yel029c*) and the 10 fastest-growing strains in rich medium. (**E**) Hybridization data for the 10 slowest-growing (*hem14, arg5,6, ade1, yer068c-a, ilv1, yer014c-a, yer044c, lys7, gcn4, rnr1*) and 10 fastest growing strains in minimal medium.

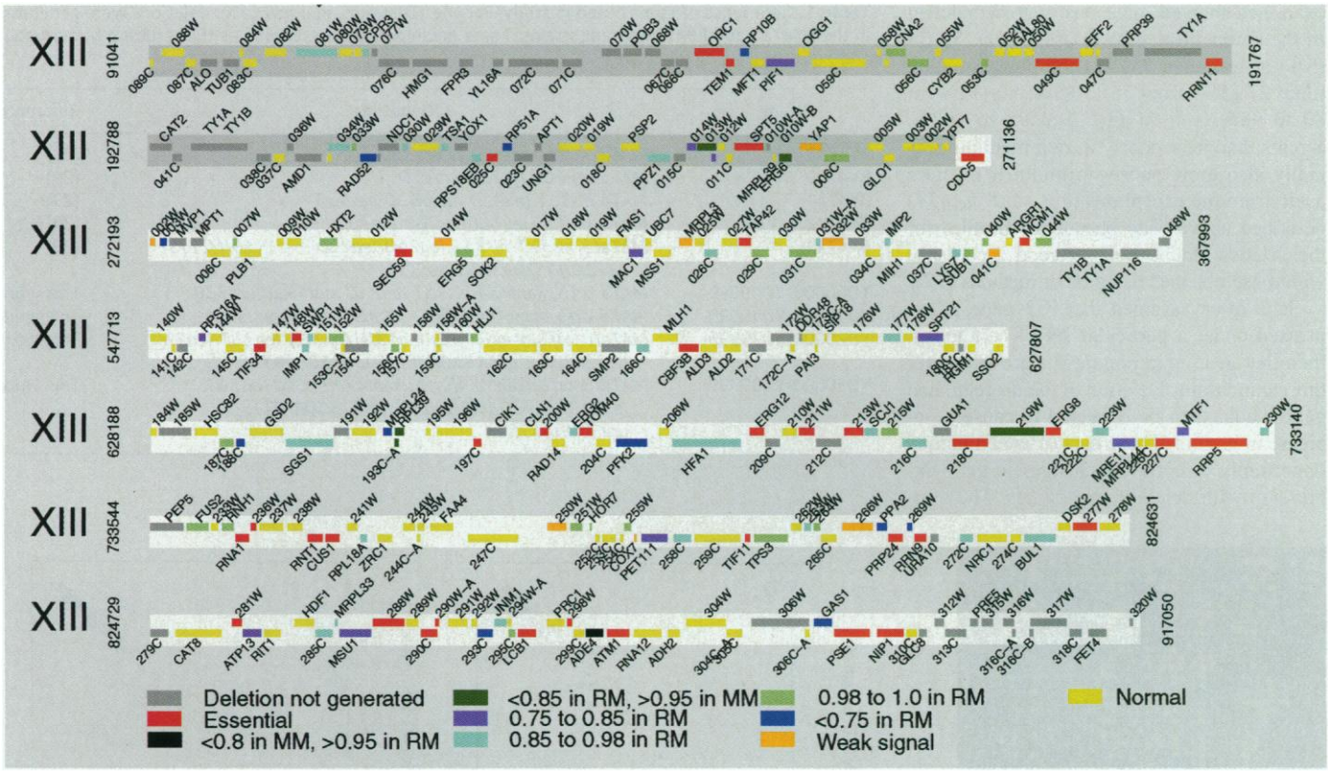

Fig. 3. Deletion map for 336 ORFs and the results of parallel phenotypic analysis for 226 ORFs on chromosome XIII. Data for additional chromosomes can be found at www.sciencemag.org/feature/data/1040380.shl. Chromosome right arms are shown with white backgrounds and left arms with gray. ORFs for which deletions were not generated (gray bars) resulted from failure during PCR or oligonucleotide synthesis (5.2%), failure for unknown reasons (4%), failure to pick unique primers (3.3%), or failure to generate deletions in all four strains (2.5%).

script abundance of all genes in the genome was measured in rich and minimal media to determine whether the set of genes whose inactivation produced a quantitative fitness defect in rich and minimal media would be the same set whose expression was induced under these growth conditions (*18*). Surprisingly little correlation existed between the growth study data and the expression data. Deletion of genes specifically induced in minimal media was no more likely to affect fitness in minimal medium than deletion of the uninduced genes. Of the genes showing a strong minimal medium growth defect, only *ARG4* and *ARG5,6* were significantly upregulated (greater than twofold) in minimal medium relative to rich medium. Similarly, only one of the six genes that showed a rich medium–specific growth defect was upregulated more than 2.5-fold in rich medium relative to minimal media. These data indicate the importance of multiple approaches in genome-wide functional analysis studies.

The results we present demonstrate that quantitative fitness data can be rapidly obtained under various conditions. Although the presence of the KanMX4 gene has been shown to have no effect on the fitness of some deletion strains (*19*), it is theoretically possible that the encoded neomycin phosphotransferase could have an impact on a particular deletion strain. The composition of a pool of deletion strains and the conditions under which the pool was cultured could also have an effect on the observed fitness of the strains. Finally, the phenotypes of certain deletions strains might be complemented by factors released into the medium by other strains in the pool. Additional tests will be required to determine how frequently these artifacts will occur.

These results also show that thousands of deletants can be systematically made once the sequence of a genome is known. Several laboratories in Europe and North America are collaborating to finish construction of tagged deletions for all annotated *S. cerevisiae* ORFs within 1 year (*20*). Currently, more than three-quarters of the ORFs in the yeast genome have been deleted (5100 ORFS and more than 15,300 strains generated). Whereas a significant amount of work was expended to construct the strains, in contrast to other methods for generating functional data (*2, 3*), the strains provide a lasting resource. In addition, the availability of a consistent set of isogenic strains should provide a better way for researchers to compare their results with those of others, easing the task of curating the functional assignments that hitherto have been made in various strain backgrounds. Finally, while other efforts have been mounted by a European consortium and others to generate deletion strains (*21*), the inclusion of barcodes significantly enhances the usefulness of the strains. The ability to assess thousands of strains quantitatively and in parallel will significantly decrease the amount of labor and materials needed for fitness screens (*22*) and increase the reliability of the data interpretation and functional classifications.

References and Notes

1. D. E. Bassett Jr. *et al.*, *Nature Genet.* **15**, 339 (1997); F. Foury, *Gene* **195**, 1 (1997).
2. V. Smith, D. Botstein, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6479 (1995); N. Burns *et al.*, *Genes Dev.* **8**, 1087 (1994); P. Ross-Macdonald, A. Sheehan, G. S. Roeder, M. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 190 (1997).
3. V. Smith, K. N. Chou, D. Lashkari, D. Botstein, P. O. Brown, *Science* **274**, 2069 (1996).
4. A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, C. Cullin, *Nucleic Acids Res.* **21**, 3329 (1993); A. Wach, A. Brachat, R. Pohlmann, P. Philippsen, *Yeast* **10**, 1793 (1994); M. C. Lorenz *et al.*, *Gene* **158**, 113 (1995).
5. A. Wach, A. Brachat, C. Alberti-Segui, C. Rebischung, P. Philippsen, *Yeast* **13**, 1065 (1997).
6. D. D. Shoemaker, D. A. Lashkari, D. Morris, M. Mittmann, R. W. Davis, *Nature Genet.* **14**, 450 (1996).
7. M. Hensel *et al.*, *Science* **269**, 400 (1995).
8. To construct deletion strains, two long oligonucleotide primers are synthesized, each containing (3′ to 5′) 18 or 19 bases of homology to the antibiotic resistance cassette, KanMX4 (U1, D1), a unique 20-bp tag sequence, an 18-bp tag priming site (U2 or D2), and 18 bases of sequence complementary to the region upstream or downstream of the yeast ORF

being targeted (including the start codon or stop codon; see http://sequence-www.stanford.edu/group/yeast/yeast_deletion_project/new_deletion_strategy.html). These 74-mers are used to amplify the heterologous KanMX4 module, which contains a constitutive, efficient promoter from a related yeast strain, *Ashbya gosspii*, fused to the kanamycin resistance gene, *nptI* (*5*). Because oligonucleotide synthesis is 3′ to 5′ and the fraction of full-size molecules decreases with increasing length, improved targeting is achieved by performing a second round of PCR using primers bearing 45 bases of homology to the region upstream and downstream of a particular ORF. Transformation with the PCR product results in replacement of the targeted gene upon selection for G418 resistance. The unique 20-mer tag sequences are covalently linked to the sequence that targets them to the yeast genome, creating a permanent association and genetic linkage between a particular deletion strain and the tag sequence. The use of two tags increases confidence in the analyses, and the redundancy is useful in case one of the tags carries a mutation or performs poorly in hybridization assays. To verify correct integration of the deletion cassette, genomic DNA was prepared from the resistant strains and used as template in PCR reactions using two primers common to the KanMX4 module (KanB (5′-CTGCAGCGAGGAGCCGTAAT-3′) and KanC (5′-TGATTTTGATGACGAGCGTAAT-3′) and four ORF-specific primers (A, B, C, and D). A and D are from regions 200 to 400 bases upstream or downstream of the start codon, whereas B and C are from within the ORF (see http://sequence-www.stanford.edu/group/yeast_deletion_project/confirmation.html). For verification, both the A-KanB and the D-KanC PCR reactions were required to give the correct size product when analyzed by gel electrophoresis. If one of either the A-KanB or D-KanC reactions failed to yield a product, the identification of the correctly-sized AD product could suffice. In addition, haploid deletion strains were tested for the disappearance of the wild-type AB and CD products. All ORFs encoding proteins greater than 100 amino acids in size were initially selected for deletion. The deletion cassettes were designed to remove the entire coding sequence for a given ORF but to leave the start and stop codon intact. Although ~10% of ORFs in *S. cerevisiae* overlap one another, the positions of the deletions were not adjusted, nor was any attempt made to avoid essential genes, genes in which a previous deletion had been constructed, or genes with a well-defined function. Genes represented multiple times in the genome (telomeric ORF, Ty-elements) were usually not deleted as their targeted disruption would pose a challenge due to the conservation of upstream and downstream regions. Some smaller nonannotated ORFs (NORFs) will be deleted in the future. Transcripts from many of these NORFs have been detected in SAGE analysis, warranting their inclusion in the study [V. E. Velculescu *et al.*, *Cell* **88**, 243 (1997)]. All oligonucleotide primers (5 nmol scale) were synthesized on an automated multiplex oligonucleotide synthesizer [D. A. Lashkari, S. Hunicke-Smith, R. M. Norgren, R. W. Davis, T. Brennan, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7912 (1995)] in batches of 96 primers. Scripts were written to automate the selection of primers. Primer sequences and ORF locations were chosen from the Stanford Genome Database (http://genome-www.stanford.edu/Saccharomyces/) at various times over a 2-year period. The KanMX4 cassette was PCR-amplified, and the resulting PCR products were sent to participating laboratories where 1 μg of PCR product was used to transform yeast by means of a variation on the standard lithium acetate procedure (http://sequence-www.stanford.edu/group/yeast_deletion_project/protocols.html) in a 96-well format. Electronic records, accessible over the World-Wide Web, were kept for every strain constructed. *MAT***a** haploid strains were given record numbers of less than 10,000, *MAT*α haploid strains were given record numbers between 10,000 and 20,000, the heterozygous diploid, between 20,000 and 30,000, and the homozygous diploid, greater than 30,000. Each record consists of primer sequence information, the results of the different diploid tests that were performed, and notes about the phenotype. Data for completed strains are accessible from www-sequence.stanford.edu/group/deletion/index.html. These strains, frozen in 15% glycerol, can be obtained from Research Genetics (Huntsville, AL) or EUROSCARF (Frankfurt, Germany).

9. Four different types of strains, containing several genetic markers (Table 1), were generated for each ORF—two haploid strains, one of each mating type, and two diploid strains, heterozygous and homozygous for the deletion loci and kanamycin marker. The homozygous diploid was constructed by mating the two haploid strains, obtained from independent transformations. Typically, the heterozygous diploid and one of the haploid strains were obtained by direct transformation, while the other haploid strain and the homozygous diploid were obtained by sporulation and mating, respectively. Essential genes were identified by 2:2 segregation of viability in tetrads derived from the heterozygous diploid. Strains were sporulated by patching them on a fresh GNA plate (5% D-glucose, 3% Difco nutrient broth, 1% Difco yeast extract, 2% Difco Bacto agar) for 1 day at 30°C before transfer to liquid sporulation medium (1% potassium acetate and 0.005% zinc acetate supplemented with 0.1 mM uracil, 0.15 mM histidine-HCL, or 1.0 mM leucine as necessary). Sporulation cultures were incubated on a rollerwheel for 4 to 5 days at 25°C. If 2:2 segregation of viability was consistently observed in two independently-transformed heterozygous deletion strains, the gene was designated essential. All confirmed diploid strains obtained through mating were required to pass two of three tests: a diploid budding pattern, the ability to sporulate, and the inability to mate. In a few cases, essential genes overlapped other essential genes (20 pairs) or a gene whose viability status was unknown (four pairs), making it difficult to determine the cause of lethality.
10. H. W. Mewes, K. Albermann, K. Heumann, S. Liebl, F. Pfeiffer, *Nucleic Acids Res.* **25**, 28 (1997).
11. R. J. Cho *et al.*, *Mol. Cell* **2**, 65 (1998).
12. To construct the pools, each deletion strain was patched on YPD plates in the presence of 150 mg/liter G418. Approximately equal numbers of cells were harvested from the plate for each strain and combined. Aliquots of the pools were stored in the presence of 15% glycerol at −80°C.
13. The UPTAG and DOWNTAG sequences were separately amplified from genomic DNA by means of PCR using primers B-U2-comp (5′ biotin-GTCGACCTGCAGCGTACG-3′) and U1 (5′-GATGTCCACGAGGTCTCT-3′), or primers B-D2-comp (5′-biotin-CGAGCTCGAATTCATCG-3′) and D1 (5′-CGGTGTCGGTCTCGTAG-3′). In both cases, a twofold molar excess of biotinylated primers was used in the reactions. The amplified UPTAG and DOWNTAG sequences were combined and hybridized to high-density oligonucleotide arrays in 200 μl of 6×SSPE [1 M NaCl, 66 mM NaH_2PO^{-4}, 6.6 mM EDTA, (pH 7.4)], containing 0.005% Triton X-100 (SSPE-T), 200 pmol U1, 200 pmol U2 (5′-CGTACGCTGCAGGTCGAC-3′), 200 pmol D1, and 200 pmol primer D2 (5′-CGATGAATTCGAGCTCG-3′). The addition of complementary primers to the hybridization mix was shown to improve the signal-to-noise ratio [D. D. Shoemaker, thesis, Stanford University, Stanford, CA (1998)]. Samples were heated to 100°C for two min, and then cooled on ice before being applied to the array. Samples were hybridized for 1 hour at 42°C. The arrays were washed two times with six changes of 6×SSPE-T. The arrays were then stained at 42°C for 10 min with 6×SSPE-T containing 2 μg/ml phycoerythrin-streptavidin (Molecular Probes) and 1 μg/ml acetylated bovine serum albumin, washed two times with five changes 6×SSPE-T, and scanned at an emission wavelength of 560 nm using an Affymetrix GeneChip Scanner. Of the strains analyzed, 157 contained only a single tag sequence (UPTAG). Six of these strains were not detected in the hybridization mix. Of the strains represented with two-tag sequences, 98.5% exhibited either an UPTAG or DOWNTAG bar code hybridization signal that was greater than threefold over background. Sequencing of 186 deletion regions tags showed that 25% of mutations in the tags or tag priming sites resulted in a nonfunctional tag that could not be amplified or detected by hybridization, or both. In only 1.1% of cases was a complete lack of hybridization signal not associated with a mutation in the tag or tag priming site. Mutations were most often found in the tags or tag priming sites (0.85% per base) and were less frequent in the regions of yeast homology (0.25% per base), most likely due to selection against the mutated PCR products during the recombination event or to the two-step PCR strategy.
14. Grids were aligned to the scanned images using the known feature dimensions of the array. The hybridization intensities for each of the elements in the grid were determined using the 75th percentile method in the Affymetrix GeneChip software package. Subsequent analysis of the hybridization intensities consisted of two steps: adjustment of data to achieve approximate equality of background and maximal signals on each array, and analysis of the decrease (or increase) of the UPTAG and DOWNTAG signal strength over time. Equalization of signal strength relied on the fact that for most array sites, the amount of tag DNA present did not vary over time. A consensus score for these sites was obtained from the first principal component of the logarithms of the signals. The logarithms of the hybridization intensity for each element on the array were linearly transformed to make each array's overall signal approximately equal to the consensus. The growth rate for each strain was determined by using the model $\log_2(\text{signal}) = \max(a + b\,t, 0) + e$, where a and b are model parameters, t represents the number of population doublings, and e is a random error term. The growth rate is calculated as $1 + b$. For strains that have dropped out of the pool ($a + bt < 0$), the model describes the statistical distribution of background signals. When e is normally distributed, background signals have a lognormal distribution. This appeared to be roughly true in our data. However, we found that a small fraction of signals on each array are liable to have very high values, much larger than can be accounted for by a purely lognormal background model. To obtain a degree of robustness against these occasional outliers, we assumed that the error term e had a scaled t distribution with one degree of freedom. The use of this heavy-tailed distribution reduced the likelihood of false positive identifications of deficient strains due to occasional high signal levels at $t = 0$, at the expense of a possible reduction in the ability to detect marginally deficient strains.
15. The analysis was not as accurate for strains with growth rates of less than 0.5 that of the wild type, because generally only three data points were above background for these.
16. L. L. Du, R. N. Collins, P. J. Novick, *J. Biol. Chem.* **273**, 3253 (1998).
17. A comprehensive study of chromosome V genes using genetic footprinting (*3*) provided an opportunity to validate the data: the results generally agreed, with a few exceptions. For example, of the 52 genes whose disruption had no effect on strain fitness under all conditions tested by genomic footprinting, we detected a growth defect in deletants *yel033w* (0.68, R; 0.83, M), *yel050c* (0.73, R; 0.68, M) and *yer028c* (0.79, R; 0.69, M). The observed phenotype for *yel033w* probably results from interference with a neighboring gene (*HYP2*, encoding translation initiation factor eIF-5A). In addition, the *hem14* deletant showed a strong growth defect in rich medium, while genomic footprinting revealed a salt-specific defect, but no defect in rich medium. Of the 11 genes that had been shown by genomic footprinting to have a severe growth defect, in three cases our deletants appeared to have no discernible phenotype (*nrf1*, *gda1*, *pcl6*). These differences could be due to our use of diploid versus haploid strains, to the auxotrophies carried by our deletants, or to our using 30°C versus 25°C as the growth temperature. The effect of temperature is probably the cause of discrepancies for the *nrf1* strain, which grows slowly at 25°C but grows faster than normal at 36.5°C. Among other disparities, two (*YER082C* and *YEL026W*) of the 22 genes on chromosome V determined to be essential by tetrad

analysis were wild-type by footprinting (perhaps because of cross-feeding in the Ty pool or because the products are required for germination) and the minimal medium–specific growth defects we detected for *hom3* and *ilv1* mutants were not detected by genomic footprinting.

18. Cultures of BY4743 ho::KanMX4 were grown at 30°C in YPD or minimal media supplemented with histidine, uracil, lysine, methionine, and adenine, in the presence of G418, and were harvested at mid-log phase. cDNA was prepared from 20 μg of polyadenylated RNA from each sample, using a dT_{21} primer and Superscript II reverse transcriptase (GibcoBRL), according to the manufacturer's recommendation. cDNA was fragmented using DNaseI (GibcoBRL), biotinylated using ddATP (NEN) and Terminal Transferase (Boehringer), and hybridized to yeast full-genome arrays (Affymetrix) as described in L. Wodicka *et al.* [*Nature Biotechnol.* **15**, 1359 (1997)]. After scanning, the average signal from each array was normalized to the average signal strength of all eight chips.
19. F. Baganz *et al.*, *Yeast* **14**, 1417 (1998).
20. Contributing groups include all authors, G. Valle, S. Kelley, J. Strathern, and D. Garfinkel.
21. EUROFAN projects B0 and B9 at www.mips.biochem.mpg.de/proj/eurofan/index.html; R. Niedenthal *et al.*, *Yeast*, in press.
22. G. Giaever *et al.*, *Nature Genet.* **21**, 278 (1999).
23. K. H. Wolfe and D. C. Shields, *Nature* **387**, 708 (1997).
24. M. Bickle, P. A. Delley, A. Schmidt, M. N. Hall, *EMBO J.* **17**, 2235 (1998); J. Corominas *et al.*, *FEBS Lett.* **310**, 182 (1992); M. A. Santos, J. J. Garcia-Ramirez, J. L. Revuelta, *J. Biol. Chem.* **270**, 437 (1995).
25. T. Y. Chow, J. J. Ash, D. Dignard, D. Y. Thomas, *J. Cell Sci.* **101**, 709 (1992); M. R. Rad *et al.*, *Yeast* **13**, 281 (1997); J. D. Brown *et al.*, *EMBO J.* **13**, 4390 (1994); A. K. Machado, B. A. Morgan, G. F. Merrill, *J. Biol. Chem.* **272**, 17045 (1997).
26. A. Yamamoto, V. Guacci, D. Koshland, *J. Cell Biol.* **133**, 85 (1996); S. D. Leidich and P. Orlean, *J. Biol. Chem.* **271**, 27829 (1996).
27. R. K. Storms *et al.*, *Genome* **40**, 151 (1997); S. J. Elledge and R. W. Davis, *Genes Dev.* **4**, 740 (1990); S. Kasahara *et al.*, *J. Bacteriol.* **176**, 1488 (1994).
28. S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991); A. C. Pease *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5022 (1994).
29. C. B. Brachmann *et al.*, *Yeast* **14**, 115 (1998).
30. We thank D. Lashkari for establishing the oligonucleotide synthesis facility, T. Nguyen, M. Sigrist, and K. Tanner for help in tetrad analysis, S. Voegeli for DNA sequence analyses, P. Koetter for distribution of deletion strains, J. Rine for helpful advice, and M.Cherry and K. Wolfe for files. E.A.W. is supported by the John Wasmuth fellowship in Genomic Analysis (HG00185-02). Supported by NIH grants HG01633, HG01627, HG00198, by an operating grant from the Medical Research Council of Canada, by grants from the European Commission (BIO4-CT97-2294), by the Swiss Federal Office for Education and Science, and by the region de Bruxelles-Capital, Belgium.

26 March 1999; accepted 7 July 1999

Early Neocortical Regionalization in the Absence of Thalamic Innervation

Emily M. Miyashita-Lin,[1]* Robert Hevner,[1]* Karen Montzka Wassarman,[2]† Salvador Martinez,[3] John L. R. Rubenstein[1]‡

There is a long-standing controversy regarding the mechanisms that generate the functional subdivisions of the cerebral neocortex. One model proposes that thalamic axonal input specifies these subdivisions; the competing model postulates that patterning mechanisms intrinsic to the dorsal telencephalon generate neocortical regions. *Gbx-2* mutant mice, whose thalamic differentiation is disrupted, were investigated. Despite the lack of cortical innervation by thalamic axons, neocortical region–specific gene expression (*Cadherin-6*, *EphA-7*, *Id-2*, and *RZR-beta*) developed normally. This provides evidence that patterning mechanisms intrinsic to the neocortex specify the basic organization of its functional subdivisions.

The mammalian neocortex is organized into regionally distinct functional subdivisions. There are two proposed mechanisms for neocortical regionalization. The protocortex hypothesis postulates that thalamic afferent fibers play an important role in neocortical regional development (*1*). On the other hand, the protomap hypothesis (*2*) postulates that regionalization of the neocortex is due to molecular differences within the neocortical ventricular zones (*3–7*). We investigated mice with a mutation of the *Gbx-2* gene (*8*, *9*) that disrupts thalamic histogenesis, which in turn blocks formation of thalamocortical projections.

Analysis of *Gbx-2*, *Id-4*, and *Lef-1* expression showed that *Gbx-2*–deficient mice have abnormal thalamic development. *Gbx-2* expression in the dorsal thalamus (DT) could be detected in the subventricular (SVZ) and mantle zones (MZ) (*10*). In the *Gbx-2* mutant, *Gbx-2* mRNA expression was maintained in the SVZ but was lost in the MZ (Fig. 1, E and F), suggesting that the mutant is unable to produce normal DT postmitotic cells (*Gbx-2* transcripts were detected in the mutant because the 5′ end of this gene has not been altered). This hypothesis was supported by the expression pattern of *Id-4*. The *Id-4* helix-loop-helix gene (*11*), which normally is expressed throughout the MZ of the DT, was not detectable in the DT of the *Gbx-2* mutant (Fig. 1, I and J). In addition to the apparent defect in the MZ, there was ectopic expression of *Lef-1* in the mutant. At embryonic day (E) 14.5, the *Lef-1* high mobility group box transcription factor (*12*) was expressed in the MZ of the pretectum (PT) and DT. In the *Gbx-2* mutants, *Lef-1* expression in the PT was unaffected, whereas its expression in the MZ was lost in the DT, and ectopic periventricular expression was observed (Fig. 1, M and N). Thus, by E14.5, loss of *Gbx-2* function disrupted differentiation in most, or all, of the DT.

To verify the thalamic abnormalities, we studied thalamic anatomy at postnatal day (P) 0, the day when these animals die. Histological abnormalities in the P0 *Gbx-2* mutant thalamus were even more pronounced. *Gbx-2* expression marked the lateral dorsal (LD), centromedian (CM), dorsal lateral geniculate (DLG), and reuniens (Re) nuclei (Fig. 1G). Several other nuclei also expressed *Gbx-2*, including the ventral lateral, the midline, the anteromedial, the mediodorsal, and the medial geniculate (MG) nuclei (*13*). In the mutant, there was only weak *Gbx-2* expression, which was restricted to a periventricular region (see asterisk in Fig. 1H).

Some thalamic regions of the *Gbx-2* mutants retained *Id-4* and *Lef-1* expression. *Id-4* expression marked most nuclei in the DT and the reticular nucleus (RT) of the ventral thalamus (Fig. 1K). In contrast, the *Gbx-2* mutant had greatly reduced *Id-4* expression in most areas of the DT (Fig. 1L) but retained expression in the RT, which does not express *Gbx-2* (*10*). Limited *Id-4* expression was detectable in some dorsal thalamic areas that, on the basis of their location, may correspond to the ventrobasal complex (VB) and DLG nuclei (Fig. 1L). *Lef-1* expression marked specific DT subdivisions including the DLG, VB, ventromedial (VM), and posterior (Po) nuclei (Fig. 1O). In the *Gbx-2* mutant, *Lef-1* expression was lost from most of the MZ, but unlike the wild type, its expression was found near the ventricle (arrow in Fig. 1P). Some *Lef-1*

[1]Nina Ireland Laboratory of Developmental Neurobiology, Department of Psychiatry, School of Medicine, University of California San Francisco, San Francisco, CA 94143–0984, USA. [2]Department of Anatomy and Program in Developmental Biology, School of Medicine, University of California, San Francisco, CA 94143–0452, USA. [3]Department of Morphological Sciences, Faculty of Medicine, University of Murcia, Institute of Neurosciences, University Miguel Hernandez, Alicante, Spain.

*These authors contributed equally to this work.
†Present address: Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health Building 18T, Room 101, 9000 Rockville Pike, Bethesda, MD 20892, USA.
‡To whom correspondence should be addressed. E-mail: jlrr@cgl.ucsf.edu