JCB Article

Transcriptome profiling to identify genes involved in peroxisome assembly and function

Jennifer J. Smith,[1] Marcello Marelli,[1] Rowan H. Christmas,[1] Franco J. Vizeacoumar,[2] David J. Dilworth,[1] Trey Ideker,[1] Timothy Galitski,[1] Krassen Dimitrov,[1] Richard A. Rachubinski,[2] and John D. Aitchison[1,2]

[1]The Institute for Systems Biology, Seattle, WA 98103
[2]Department of Cell Biology, University of Alberta, Edmonton, Alberta, T6G 2H7

Yeast cells were induced to proliferate peroxisomes, and microarray transcriptional profiling was used to identify *PEX* genes encoding peroxins involved in peroxisome assembly and genes involved in peroxisome function. Clustering algorithms identified 224 genes with expression profiles similar to those of genes encoding peroxisomal proteins and genes involved in peroxisome biogenesis. Several previously uncharacterized genes were identified, two of which, *YPL112c* and *YOR084w*, encode proteins of the peroxisomal membrane and matrix, respectively. Ypl112p, renamed Pex25p, is a novel peroxin required for the regulation of peroxisome size and maintenance. These studies demonstrate the utility of comparative gene profiling as an alternative to functional assays to identify genes with roles in peroxisome biogenesis.

Introduction

Global transcriptional profiling is a powerful tool that can expose expression patterns to define cellular states or to identify genes with similar expression patterns. Here, we apply this technique to the identification of novel peroxisomal proteins and peroxins, proteins involved in peroxisome biogenesis.

Peroxisomes are organelles found in organisms from yeasts to mammals and in most cell types. They compartmentalize several oxidative reactions, including fatty acid β-oxidation, and the enzymes catalase and superoxide dismutase that detoxify peroxides and superoxides (Keller et al., 1991; for review see Purdue and Lazarow, 2001; Kira et al., 2002). Peroxisomes are essential for human survival; inherited mutations that disrupt the formation of peroxisomes lead to severe neuropathologic defects and often death in early infancy (Gould and Valle, 2000). Because the signals on proteins (peroxisomal targeting signals [PTSs]*) that function to direct them to peroxisomes and the receptors that recognize these signals (for review see Titorenko and Rachubinski, 2001) have been well conserved through evolution, the study of model organisms has led to tremendous advances in our understanding of the molecular mechanisms underlying these disorders.

To date, 24 *PEX* genes encoding peroxins required for peroxisome assembly or maintenance have been characterized (Purdue and Lazarow, 2001; Titorenko and Rachubinski, 2001; Tam and Rachubinski, 2002). In yeast, most of these peroxins have been identified using screens to isolate mutants that do not have functional peroxisomes. These include screens to identify mutants unable to utilize fatty acids or alcohols (two carbon sources metabolized by peroxisomes) and screens to identify mutants unable to import proteins into peroxisomes (for review see Subramani, 1998; Snyder et al., 1999).

In many organisms, the development of peroxisomes is a regulated process. Coincident with the transcriptional induction of genes encoding peroxisomal proteins, the size and number of peroxisomes in a cell change in response to environmental cues. In mammals, hypolipidemic drugs, cell differentiation, and certain carcinogens such as plasticizers induce peroxisome proliferation (for review see Gonzalez et al., 1998; Fajas et al., 2001). In yeast, the development of peroxisomes can be induced by metabolic shifts. Cells grown in the presence of glucose or glycerol have a few small peroxisomes, but when these cells are transferred to medium that contains the fatty acid oleate as the sole carbon source peroxisomes rapidly proliferate and increase in size (Veenhuis et al., 1987).

To identify components of peroxisomes and the peroxisome biogenesis program, we have taken advantage of this peroxi-

The online version of this article contains supplemental material.

Address correspondence to John D. Aitchison, Institute for Systems Biology, 1441 N. 34th St. Seattle, WA 98103-8904. Tel.: (206) 732-1344. Fax: (206) 732-1299. E-mail: jaitchison@systemsbiology.org

J.J. Smith and M. Marelli contributed equally to this work.

T. Ideker's present address is The Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142.

*Abbreviations used in this paper: DsRed, red fluorescent protein from *Discosoma sp.*; ORE, oleate response element; PTS, peroxisomal targeting signal; SOM, self-organizing map.

Key words: microarray; clustering algorithms; peroxin; PEX25; PEX11

© The Rockefeller University Press, 0021-9525/2002/07/259/13 $5.00

some induction process to identify novel genes whose expression patterns match those encoding known peroxins and peroxisomal proteins during metabolic shifts to induce or repress peroxisomes. Using oligonucleotide-based whole genome microarrays, we compared the yeast transcriptome during growth on glycerol to that at various time points during growth on oleate (peroxisome induction) or glucose (peroxisome repression) and created an expression profile for each gene. Combining the results from the different clustering analyses identified over 200 genes whose expression patterns match those of known peroxins and peroxisomal proteins. Based on this criterion, we expect that previously uncharacterized members of this group are involved in peroxisome biogenesis and/or function. Here we present evidence that two such candidates, Yor084p and Ypl112p, are indeed peroxisomal proteins and further demonstrate that Ypl112p is a novel membrane peroxin required for regulating peroxisomal size and maintenance. We term this protein Pex25p.

Results

Microarray analysis of peroxisome generation

To identify novel genes involved in peroxisome biogenesis or function, the yeast transcriptome was interrogated under conditions that induce or repress peroxisome assembly and proliferation. Importantly, whereas increased steady-state transcript levels during growth on oleate compared with glucose are characteristic of genes involved in peroxisome biogenesis or function, this property is shared with many unrelated genes such as those involved in heat shock or stress responses (Kal et al., 1999). Thus, to discriminate expression fingerprints among genes with these different functions several conditions were explored. Messenger RNAs were isolated from duplicate samples of cells either maintained in glucose or glycerol or after a shift from glycerol to oleate medium for 0.5, 1, 3, 6, 9, or 26 h. cDNAs from pairs of cell populations were labeled with either Cy3 or Cy5 fluorescent dye and hybridized to microarrays representing the complete yeast genome (~6,200 genes). For each gene, the levels of each dye were measured from 16 replicate spots, and the average expression ratio was calculated. The data from eight different experiments (listed in Table I) were combined to create a profile of expression (represented by mRNA expression ratios) for each gene.

Genes were filtered first to retain only those whose expression showed significant changes in at least one of the eight experiments (Ideker et al., 2000). The resulting dataset (containing 3,031 genes; see Online supplemental material and

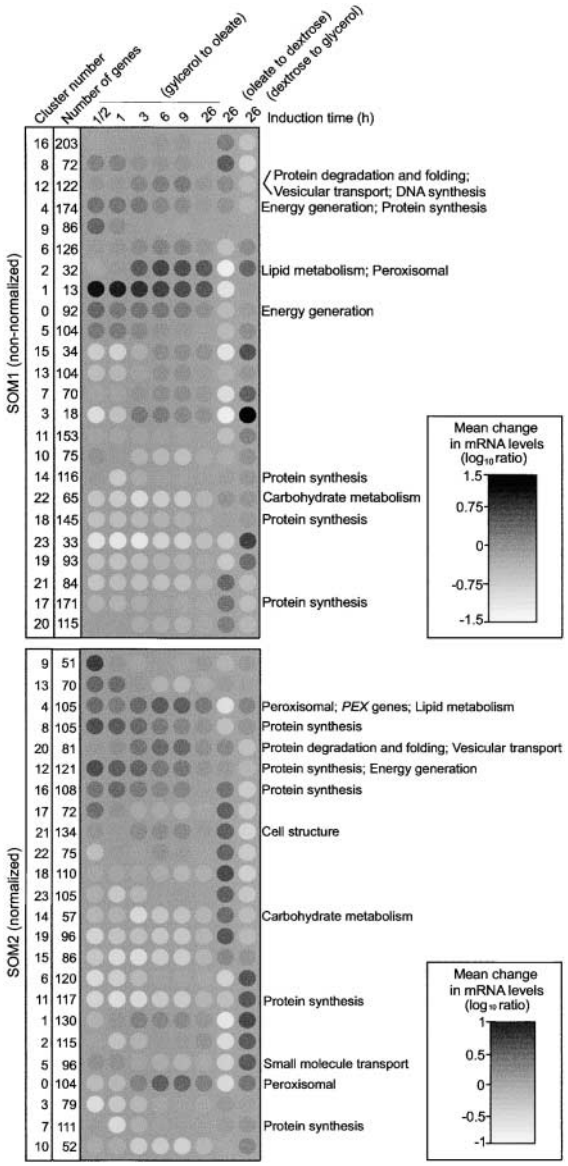

Figure 1. **GeneCluster was used to organize the filtered dataset (2,300 genes) into a two-dimensional 6 × 4 SOM (SOM1).** The clusters were then organized into a one-dimensional SOM using Cluster software for display (top). The mean $\log_{10}$ expression ratio for each cluster is represented by a shaded spot. The cluster numbers and genes per cluster of the two-dimensional SOM are at the left. Genes involved in lipid, fatty acid, and sterol metabolism and genes encoding peroxisomal proteins are enriched in cluster 2. A second SOM (SOM2) was generated the same way (bottom) except the $\log_{10}$ expression ratios for each gene were first normalized to have a mean of zero and a standard deviation of one. *PEX* genes are enriched in cluster 4.

Table I. **Microarray experiments**

Experiment number	Growth conditions compared (condition 1 versus condition 2)[a]
1	Oleate (YPBO)-induction time course 0 h versus 1/2 h
2	Oleate (YPBO)-induction time course 0 h versus 1 h
3	Oleate (YPBO)-induction time course 0 h versus 3 h
4	Oleate (YPBO)-induction time course 0 h versus 6 h
5	Oleate (YPBO)-induction time course 0 h versus 9 h
6	Oleate (YPBO)-induction time course 0 h versus 26 h
7	Oleate (YPBO) versus glucose (YPBD)
8	Glucose (YPBD) versus glycerol (YPBG)

[a]For the oleate-induction time course, 0 h time points are glycerol (YPBG)-grown cells.

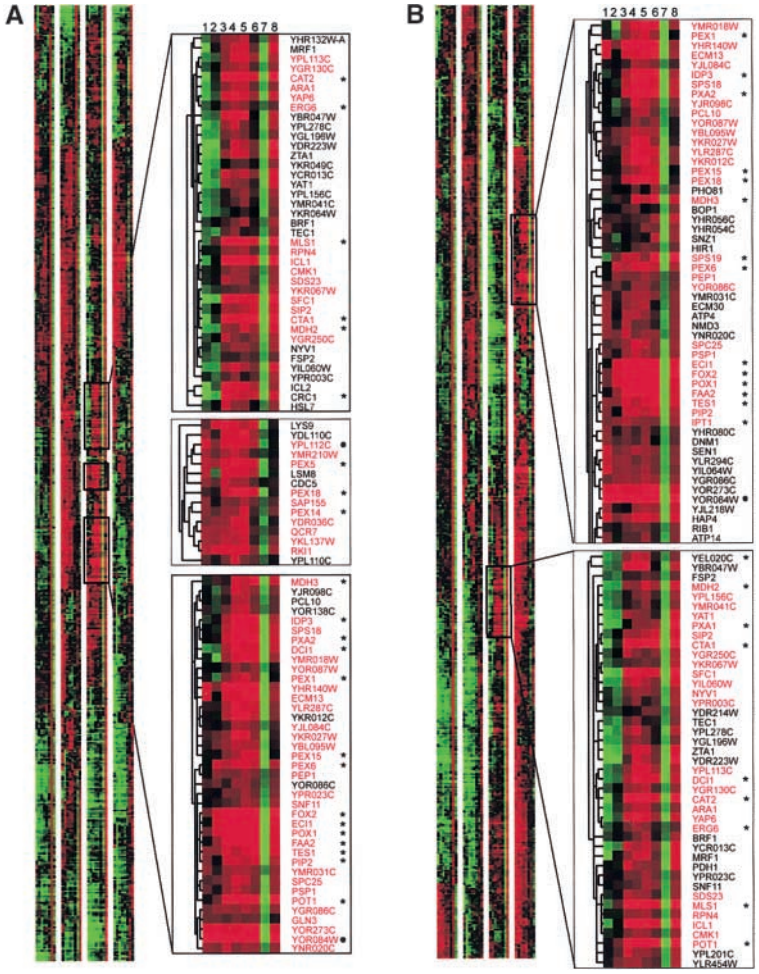

Figure 2. **Unsupervised hierarchical clustering.** Clustered display of gene expression profiles generated using Cluster and Tree view software with normalized (A) and nonnormalized (B) data. Columns represent the expression ratios for experiments 1–8 (Table I). Rows represent $\log_{10}$ expression ratios of individual genes, with red and green indicating positive and negative ratios, respectively. All genes in the filtered dataset are represented at the left. Candidate genes are those in branches containing *PEX* genes or genes encoding exclusively peroxisomal proteins, in which all pair-wise correlation coefficients are >0.95. Regions of the tree containing selected branches (red text) are magnified, and genes that encode peroxisomal proteins and/or proteins involved in lipid, fatty acid, and sterol metabolism are marked with an asterisk. Bullets mark *YOR084w* and *YPL112c* (renamed *PEX25*).

Table S1 available at http://www.jcb.org/cgi/content/full/jcb.200204059/DC1) was again filtered to exclude genes that showed little or no change in expression ratios across the experiments. This was done by rejecting genes for which the difference between the maximum and minimum $\log_{10}$ expression ratios was <0.5. Whereas the first filter was based on the significance of gene expression changes within each experiment, the second was based on the difference in expression ratios between experiments. The resulting dataset contained 2,300 genes.

Interpretation of gene expression profiles by comparative expression analysis

Three complementary clustering algorithms were used to classify genes into groups reflecting common behaviors across the various conditions and to identify genes whose profiles matched those of genes known to be involved in peroxisome function or biogenesis.

Self-organizing maps

The expression profiles of the genes in the filtered dataset were clustered based on Euclidean distance between their $\log_{10}$ expression ratios over all eight experiments using GeneCluster software (Tamayo et al., 1999). A six row by four column self-organizing map (SOM)1 (see Online supplemental material and Table S1 [available at http://www.jcb.org/cgi/content/full/jcb.200204059/DC1] for data) was generated. This SOM geometry was chosen to maximize the number of clusters with distinct profiles without generating clusters with redundant profiles (Tamayo et al., 1999). Genes encoding peroxins and peroxisomal proteins and genes with known cellular roles were annotated based on their classifications in the literature and in the Proteome database (Costanzo et al., 2000). To establish if genes with similar classifications were enriched in any cluster, the probabilities that the observed distributions would be found by chance were determined by calculating the hypergeometric distributions for each classification in each cluster. Classifications that were enriched in any cluster (P value $< 5 \times 10^{-4}$) are shown in Fig. 1 (top). Genes encoding peroxisomal proteins had a strong enrichment in cluster 2 (P value of 3×10^{-21}). Interestingly, although genes involved in lipid, fatty acid, and sterol metabolism (cellular roles requiring peroxisomes) were also strongly enriched in cluster 2 (P value of 1.7×10^{-14}), PEX genes were not enriched in any cluster in this analysis. This is because most PEX genes identified to date are not induced to the same extent as enzymes directly involved in lipid metabolism (Kal et al., 1999).

A second SOM (SOM2; see Online supplemental material and Table S1 [available at http://www.jcb.org/cgi/content/full/jcb.200204059/DC1] for data) was therefore generated after normalizing the $\log_{10}$ expression ratios for each gene to have a mean of zero and a standard deviation of one.

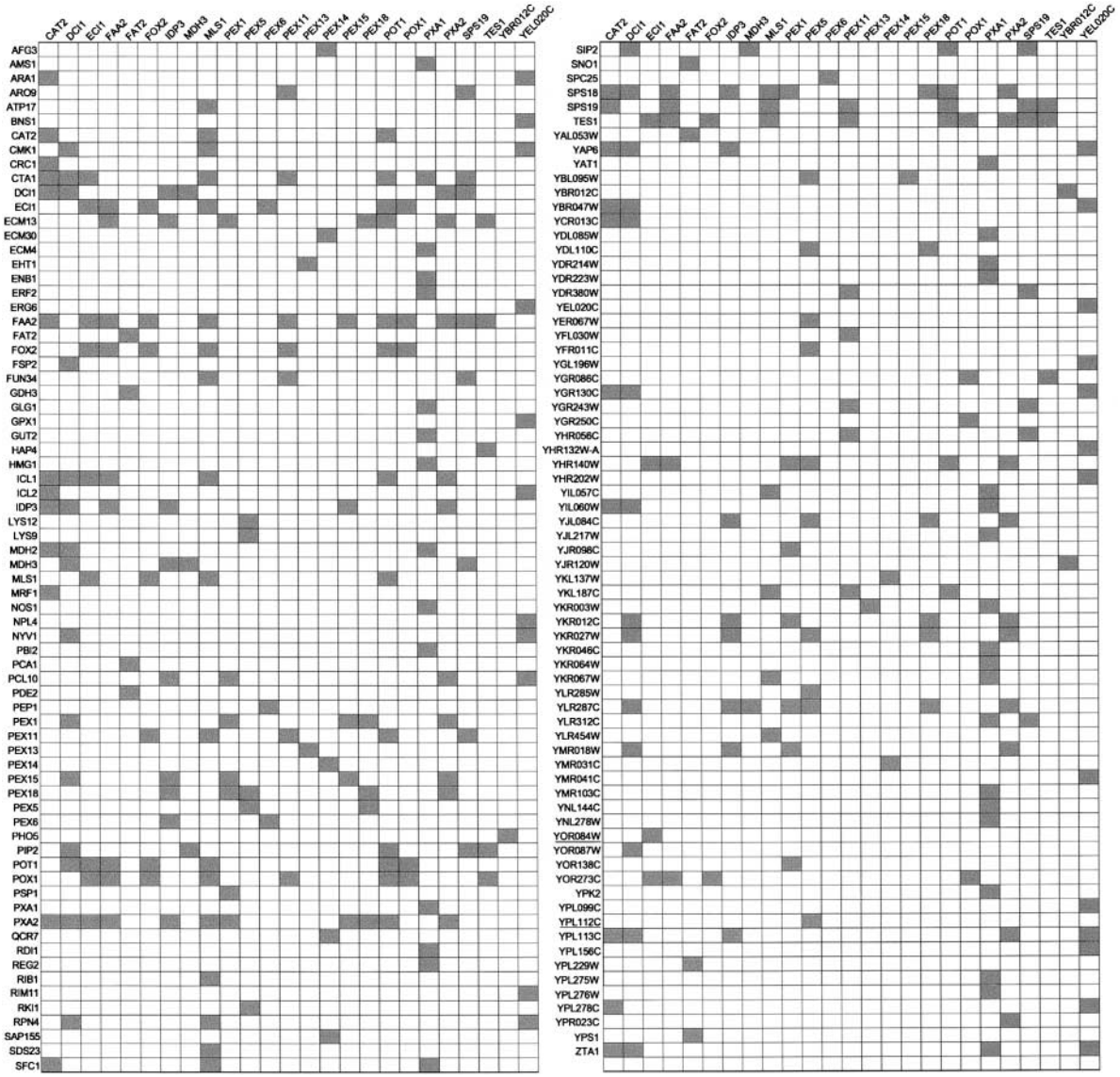

Figure 3. **Nearest Neighbor analysis.** The expression profiles of *PEX* genes and genes encoding exclusively peroxisomal proteins were compared with the expression profiles of all other genes in the filtered dataset using Nearest Neighbor analysis (GEAP; Galitski et al., 1999). Neighboring genes meeting the correlation coefficient/geometric distance cutoffs outlined in the Materials and methods were selected as candidate genes. Queried genes with at least two neighbors are at the top, and neighboring candidate genes are indicated by shaded boxes. *YOR084w* and *YPL112c* (renamed *PEX25*) are underlined.

Whereas SOM1 was set up to organize genes based on their expression profiles and absolute expression levels, the normalized data were organized in SOM2 based solely on the shapes of the expression profiles. The genes were annotated, and distribution probabilities were calculated as for SOM1. In SOM2 (Fig. 1, bottom), genes encoding peroxisomal proteins and genes involved in lipid, fatty acid, and sterol metabolism coenriched in cluster 4 (P values of 10^{-6} and 2.9×10^{-4}, respectively). However, unlike in SOM1, PEX genes were also enriched in this cluster (P value of 1.10^{-5}).

These data suggest that genes involved in peroxisome function and genes involved in peroxisome biogenesis tend to be coregulated, albeit to different extents. Consequently, profiles of genes involved in lipid metabolism and genes that encode peroxisomal proteins showed more similarity if the profile shapes and the absolute expression levels were compared (SOM1), whereas PEX gene profiles showed more similarity if only the shapes of the profiles were compared (SOM2). Since SOM1 and SOM2 gave complementary results, the genes in cluster 2 of SOM1 and in cluster 4 of SOM2 were selected as candidate genes for further analyses.

Unsupervised hierarchical clustering

Cluster and Tree View software packages (Eisen et al., 1998) organize gene expression profiles into "phylogenetic trees" in which similar profiles are adjacent to one another and branch lengths are inversely proportional to the degree of similarity between profiles. This software was used to organize the gene profiles in the filtered dataset into two trees: Tree 1 (Fig. 2 A), constructed using centered (Pearson) correlation, was organized based on the shapes of the expression profiles. Tree 2 (Fig. 2 B), constructed using uncentered

Table II. **Candidate genes**

Gene	F	ORE	N	Gene	F	ORE	N	Gene	F	ORE	N	Gene	F	ORE	N	Gene	F	ORE	N
FAA2	p	1	31	CMK1	o		6	YGR243W	u		3	CAT8	o		1	SMF2	o		1
POX1	p	4	26	IPT1	p		6	YGR250C	o	1	3	CDA1	o		1	TIF4631	o		1
ECI1	p	1	25	MDH2	p	1	6	YLR312C	u		3	CDC5	o		1	WHI2	o		1
FOX2	p	1	25	PEX5	p		6	YMR041C	u		3	CYT1	o	1	1	YAL053W	u		1
TES1	p	1	25	PEX6	p	1	6	ACS2	o	1	2	CYT2	o		1	YDL085W	o		1
PXA2	p	1	24	RPN4	o		6	AFG3	o		2	DNM1	o		1	YDL219W	o		1
SPS18	o	1	22	SFC1	o		6	ECM30	o		2	ECM4	o		1	YDR205W	o		1
IDP3	p	3	20	SNF11	o		6	GPX1	o		2	ENB1	o		1	YDR214W	u		1
PIP2	p		19	ERG6	p		5	GUT2	o		2	ERF2	o		1	YDR223W	u		1
POT1	p	2	18	FUN34	u		5	HMG1	p		2	FRE1	o		1	YDR247W	u		1
YLR287C	u		17	PEX14	p		5	LYS9	o		2	FSP2	o		1	YGL101W	u		1
DCI1	p		15	YIL060W	u	N/D	5	PDR15	o		2	GAL11	o		1	YHR054C	u		1
ECM13	o		14	YPL112C	u	1	5	RIB1	o		2	GDH3	o		1	YHR080C	u		1
SPS19	p	1	14	ARO9	o		4	RIM11	o		2	GDS1	o		1	YHR132W-A	u	N/D	1
YKR012C	u	N/D	14	CIT2	p		4	RNC1	o		2	GLG1	o		1	YHR202W	u		1
CTA1	p	1	13	EHT1	o		4	YAT1	p		2	HIR1	o		1	YIL064W	u		1
ICL1	p		13	FAT2	p		4	YCR013C	u	N/D	2	INO1	p		1	YIL163C	u		1
PEX15	p	1	13	LYS1	p		4	YDR036C	u		2	KNS1	u		1	YJL112W	o		1
PEX18	p	1	13	PEP1	o		4	YER067W	u	N/D	2	LSM8	o		1	YJL218W	u		1
PSP1	o		13	PEX13	p	2	4	YFR011C	u		2	LYS14	o		1	YJR120W	o		1
SPC25	o	1	13	PXA1	p	1	4	YGL196W	u		2	MCM1	o		1	YKL054C	u		1
YHR140W	u	1	13	RKI1	o		4	YGR035C	u	1	2	MRF1	o		1	YKR046C	u		1
YMR018W	u		13	SAP155	o		4	YIL057C	u		2	MRS6	o		1	YKR064W	u		1
PEX1	p	1	12	SDS23	o		4	YJL217W	u		2	MSB1	o		1	YLR294C	u	N/D	1
YKR027W	u		12	YBR012C	u		4	YLR285W	u	1	2	MYO5	o		1	YLR454W	u		1
YPL113C	u		12	YBR047W	u		4	YMR210W	u		2	NMD3	o		1	YMR103C	u		1
YOR273C	o	1	11	YDR380W	u		4	YNL134C	u	1	2	NPL4	o		1	YNL140C	u	N/D	1
MLS1	p		10	YEL020C	p		4	YOR086C	u		2	NRP1	o		1	YNL144C	u		1
PEX11	p	1	10	YHR056C	o		4	YOR138C	u		2	NUP145	o		1	YNL205C	u	N/D	1
YAP6	o		10	YJR098C	u		4	YPL099C	u		2	NUP192	o		1	YNL212W	u		1
YGR086C	u		9	YKL137W	o		4	YPL156C	u		2	OYE3	p	1	1	YNL278W	u		1
YJL084C	u		9	YKL187C	u		4	YPL276W	u		2	PBI2	o		1	YOL027C	u	1	1
CAT2	p	1	8	YKR003W	p		4	YPL278C	u		2	PCA1	o		1	YOL087C	u		1
MDH3	p	1	8	YKR067W	p		4	YTA12	o		2	PDE2	o		1	YOR186W	u		1
YMR031C	u		8	ZTA1	u		4	ADR1	o		1	PDR12	o	1	1	YOR240W	u	1	1
YOR084W	u	1	8	CRC1	p	1	3	AMS1	o		1	PET9	o		1	YPK2	o		1
YBL095W	u		8	HAP4	o		3	ARG1	o		1	PHO5	o		1	YPL095C	u	1	1
GLN3	o		7	ICL2	o		3	ATP14	o		1	PHO84	o		1	YPL110C	u	1	1
PCL10	o		7	LYS12	o		3	ATP17	o		1	PPA2	o		1	YPL207W	u	N/D	1
SIP2	u		7	NOS1	o		3	ATP3	o		1	PTC4	o		1	YPL229W	u		1
YGR130C	u	1	7	NYV1	o		3	ATP4	o		1	RDI1	o		1	YPL275W	u		1
YNR020C	u		7	QCR7	o		3	ATP7	o		1	REG2	o		1	YPR003C	o		1
YOR087W	o	N/D	7	SNO1	o		3	BNS1	u		1	RPR2	o		1	YPR117W	u		1
YPR023C	o		7	YDL110C	u		3	CAF17	o		1	SEN1	o		1	YPS1	o		1
ARA1	o		6	YFL030W	u		3	CAR2	o		1	SKY1	o	1	1				

The 224 candidate genes are listed. Column F contains functional classifications: p, genes having a role related to peroxisomes (i.e., involved in lipid, fatty acid, and sterol metabolism or encoding a protein that is either peroxisomal or a peroxin); o, genes with other cellular roles that have not been classified as "p"; u, unknown. Annotations are based on their classifications in the literature and in the Proteome database (Costanzo et al., 2000). Column ORE contains the number of OREs found upstream of each gene using ScanACE software (version 1.3) (Hughes et al., 2000a). Only elements that were <3 SDs from the motif average are reported. N/D, not determined. Column N contains the number of "hits" each gene received by all clustering methods. For a description of each "hit," refer to the supplementary material. Two candidate genes of interest, *YPL112c* (renamed PEX25) and *YOR084w*, are underlined.

(modified Pearson) correlation, was based on both shape and the absolute levels of expression. Candidate genes chosen from these analyses are genes in branches that contain visually uniform expression profiles (Eisen et al., 1998) and contain PEX genes or genes that encode proteins that are exclusively peroxisomal. All candidate-containing branches chosen yielded correlation coefficients >0.95.

Nearest Neighbor clustering

The Nearest Neighbor component of GEAP software (Galitski et al., 1999) compares the expression profile of a user-selected gene to the profiles of other genes in a dataset, lists "neighbors" with similar profiles, and determines the similarity that neighbors share with the selected gene. We used this analysis to identify a maximum of 20 nearest neighbors (meeting the similarity threshold described below) of known *PEX* genes and genes encoding proteins that are exclusively peroxisomal. Analyses were performed using the Pearson and Euclidian metrics to find neighbors based on expression profile shape and on both shape and profile amplitude, respectively. For the Pearson-/Euclidian-based analyses, candidate genes were selected that had correlation coefficients/geometric distances above/be-

low a threshold established with randomized data to give, on average, <0.5 false positive neighbors to each selected gene. The results of these analyses are presented in Fig. 3.

Selection of candidate genes

Of the 46 genes encoding proteins that have been experimentally characterized to be peroxisomal or peroxins (according to the literature and the Proteome database [Costanzo et al., 2000]), 29 (63%) showed significant gene expression changes induced by the metabolic shifts. The remaining 17 genes were filtered out because little or no gene expression changes were detected. Among these genes, some have low transcript levels (for example, PEX8, which encodes a protein of very low abundance [Rehling et al., 2000]), whereas others, like ANT1 (Geraghty et al., 1999), are not induced significantly by oleate. Nevertheless, the profiles of the genes remaining after the filters served as beacons to identify similarly expressed genes as candidates involved in peroxisome biogenesis or function.

Table II summarizes the 224 candidate genes chosen by combining the three clustering methods. Genes are ordered by the number of "hits" each received by all clustering methods. The genes are annotated with a functional classification (as described in the Proteome database [Costanzo et al., 2000]) and include 80 genes that have not yet been characterized and 105 genes that have proposed cellular roles other than those relating to peroxisomes. Putative oleate response elements (OREs; Einerhand et al., 1993; Filipits et al., 1993), which have been shown to control transcription in response to oleate, were identified in the candidate genes and in entire genome using the CompareACE and ScanACE software packages (Hughes et al., 2000a). Candidate genes containing OREs with scores <3 standard deviations away from the ORE consensus motif average are annotated in Table II. Putative OREs meeting this condition were approximately six times more abundant in the candidate genes than in the entire genome. Interestingly, not all genes that were induced by oleate contained a recognizable ORE, and not all genes that contained an ORE showed the predicted induction, suggesting the existence of additional regulatory elements.

Characterization of candidates

To test the validity of our screen, two candidate genes, *YOR084w* and *YPL112c*, were analyzed for their involvement in peroxisome function. Both encoded proteins have no known function. Yor084p shows similarity to two *Candia albicans* proteins. It has no obvious extended similarities to proteins in other organisms but shares conserved motifs with members of the prokaryotic lipase-esterase family. Also, in agreement with our microarray analysis, SAGE analysis showed that the expression of *YOR084w* is induced by growth of cells on oleate (Kal et al., 1999). Ypl112p also shows similarity to a protein from *C. albicans*, but interestingly a search using RPS-BLAST (Altschul et al., 1997) identified a region of similarity between Ypl112p (aa 160–300) and Pex11p (aa 46–177).

Viable yeast stains containing individual deletions of *YPL112c* or *YOR084w* constructed by an international consortium of laboratories (Winzeler et al., 1999) (obtained from Resgen) were assayed for growth in the presence of oleate (requiring functional peroxisomes and mitochondria for metabolism) or acetate (requiring only mitochondria) as the sole carbon source. Both *yor084Δ* and *ypl112Δ* cells grew on both carbon sources (unpublished data). Although *yor084Δ* cells appeared to grow at or near wild-type rates, *ypl112Δ* cells displayed a distinct growth defect on oleate, suggesting that it plays an important but nonessential role in oleate metabolism. Based on the data presented below, we have renamed *YPL112c*, *PEX25* and termed the encoded protein Pex25p.

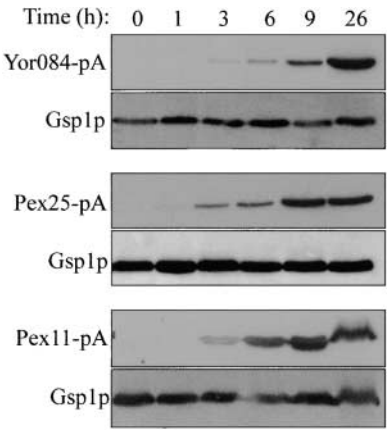

Figure 4. **Yor084-pA and Pex25-pA are induced by oleate.** Yeast strains were grown in oleate, and lysates were prepared at the indicated times. Samples containing equal protein were analyzed by Western blotting to visualize the protein A fusions. Polyclonal antibodies directed against Gsp1p (Leslie et al., 2002) were used to confirm equal protein loads.

Yor084p and Pex25p are induced by oleate

Genomically encoded protein A chimeras of Yor084p and Pex25p were monitored to analyze the expression of *YOR084w* and *PEX25*, respectively, under the control of their endogenous promoters. Yeast strains synthesizing Pex25-pA, Yor084-pA, or Pex11-pA were grown in oleate, and total cell lysates were prepared at the indicated times after induction. Changes in protein levels were analyzed by Western blotting (Fig. 4). Pex11-pA and Pex25-pA reached maximal levels within 9 h of induction, whereas Yor084-pA differed slightly, continuing to increase in abundance over the 26-h induction period. These data are also consistent with a previous study, demonstrating that *YOR084w* transcript levels are higher in oleate-grown cells than those grown in glucose (Kal et al., 1999).

Yor084p and Pex25p encode peroxisomal proteins, and Yor084p may be targeted as an oligomer

A COOH-terminal PTS (PTS1) or an NH_2-terminal signal (PTS2) is sufficient to direct reporter proteins to peroxisomes (for review see Purdue and Lazarow, 2001). Thus, to visualize peroxisomes and evaluate peroxisomal protein import (see below) we generated fluorescent chimeras that illuminate peroxisomes. Gene fusions were designed to append a PTS1 (Ser-Lys-Leu) (Gould et al., 1989) to the COOH terminus of *Discosoma sp.* red fluorescent protein (DsRed) (Matz et al., 1999) and a PTS2 (from Pot1p) (Glover et al., 1994b) to the NH_2 terminus of GFP (Chalfie et al., 1994).

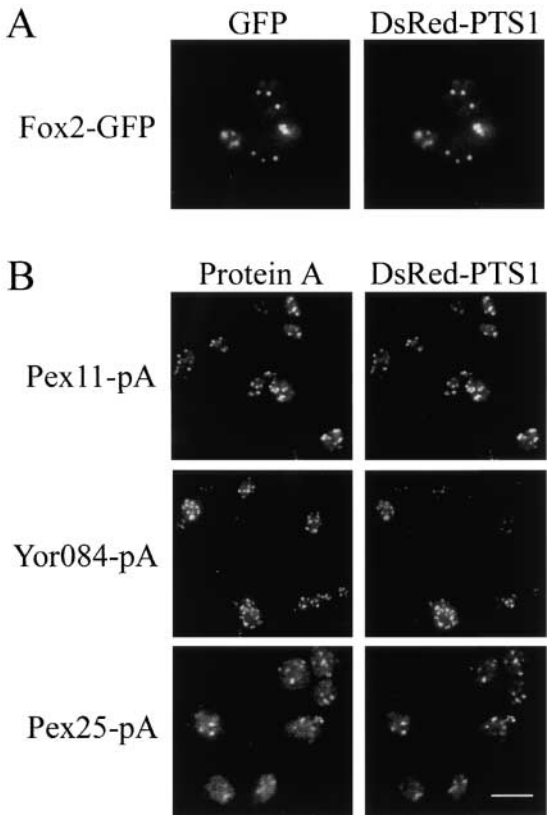

Figure 5. **Cellular distributions of Pex25p and Yor084p.** (*A*) The DsRed-PTS1 reporter colocalizes with Fox2–GFP chimera to peroxisomes. (B) The subcellular distributions of protein A chimeras were compared with that of DsRed-PTS1 in oleate-induced haploid cells (Pex11-pA and Pex25-pA) and heterozygous diploid cells (Yor084-pA) by double-labeling immunofluorescence microscopy. Bar, 10 μm.

Both of these reporters localized appropriately to peroxisomes. When expressed in a strain containing the peroxisomal multifunctional enzyme, Fox2p, genomically tagged with GFP (Fox2–GFP), the DsRed and GFP signals yielded identical overlapping patterns (Fig. 5 A). Furthermore, GFP–PTS2 similarly colocalized with DsRed-PTS1 (unpublished data). Thus, both chimeras mark the positions of peroxisomes and can be used in double fluorescence microscopy experiments to localize novel proteins in vivo.

Genomically encoded protein A chimeras appended to Yor084p, Pex25p, or Pex11p were localized in oleate-induced cells by indirect immunofluorescence microscopy combined with direct fluorescence from DsRed-PTS1 to identify peroxisomes (Fig. 5 B). In heterozygous diploid cells, Pex11-pA, Yor084-pA, and Pex25-pA colocalized with DsRed-PTS1 to small punctate structures, characteristic of peroxisomal proteins, suggesting that, like Pex11p, Yor084p and Pex25p are peroxisomal.

The localization of each chimera was also investigated in haploid cells where the only copy of each gene was fused to protein A. The pattern of Pex25-pA was the same as that observed in diploid cells. However, Yor084-pA displayed a diffuse signal characteristic of cytosolic proteins (unpublished data). This suggests that under these conditions Yor084-pA was not targeted to peroxisomes and, further, that the mistargeting of the chimera resulted from both the appended epitope tag and the absence of a wild-type version of the protein. This phenomenon is reminiscent of examples of oligomer import into peroxisomes in which proteins lacking a functional targeting signal can be imported by "piggybacking" on an interacting partner that contains a functional signal (Glover et al., 1994a) and suggests that Yor084p is imported as a homooligomer. It also seems likely that Yor084p import is mediated by a functional derivative of PTS1. The COOH-terminal tripeptide of Yor084p, QKL, is strikingly similar to the COOH-terminal PTS1 consensus, (S/A/C)-(K/R/H)-(L/M/I), and the COOH-terminal position of the epitope tag disables other PTS1 signals (unpublished data; Gould et al., 1989). Together, these data suggest that Yor084p is imported into peroxisomes as an oligomer and is targeted by a PTS1.

Pex25p is a peripheral peroxisomal membrane protein

Subcellular fractionation and extraction were used to establish if Pex25p and Yor084p are stably associated with peroxisomes and to determine their suborganellar localizations. For this analysis, Pex25-pA, Yor084-pA, Pex7-pA, and Pex13-pA strains were grown in oleate medium, cells were fractionated, and postnuclear supernatants were separated by centrifugation into supernatants enriched for cytosol (20kgSs) and crude organellar pellets (20kgPs). Equal portions of each fraction were analyzed by Western blotting (Fig. 6 A). Pex25-pA and about half of Yor084-pA were found in the 20kgP fraction, whereas protein A fused to Pex7p, a protein that is both peroxisomal and cytoplasmic (for review see Purdue and Laz-

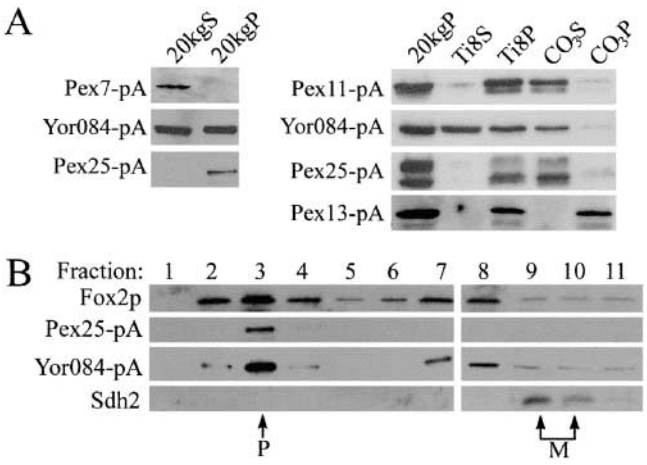

Figure 6. **Pex25p-pA is a peroxisomal peripheral membrane protein, and Yor084-pA is a peroxisomal matrix protein.** (A) A postnuclear supernatant was separated by centrifugation into a supernatant (20kgS) enriched for cytosol and a pellet (20kgP) enriched for peroxisomes and mitochondria. The 20kgP was treated with Ti8 to lyse peroxisomes releasing matrix proteins (Ti8S). The membrane-containing pellet fraction (Ti8P) was treated with Na_2CO_3 to separate peripherally associated membrane proteins (CO_3S) from integral membrane proteins (CO_3P). Western blotting detected chimeras shown and included Pex11p (peripheral) and Pex13p (integral) as controls. (B) Pex25-pA and Yor084-pA cofractionated with peroxisomes. Organelles in the 20kgP were separated by isopycnic centrifugation. Proteins were identified in each fraction by Western blotting, and peak peroxisomal (P), and mitochondrial (M) fractions (arrows) were identified with antibodies against Fox2p and Sdh2p, respectively.

arow, 2001), was in the 20kgS fraction. Peroxisomes were isolated from 20kgP fractions of Pex25-pA and Yor084-pA strains by isopycnic centrifugation. The gradients were fractionated, and equal portions of each fraction were analyzed by Western blotting (Fig. 6 B). Both Pex25-pA and Yor084-pA coenriched with the peroxisomal protein, Fox2p, and not with the mitochondrial protein, Sdh2p, suggesting that they are stably associated with peroxisomes.

Further extraction of peroxisome-containing fractions suggests that Yor084p is a peroxisomal matrix protein, whereas Pex25p is peripherally associated with the peroxisomal membrane (Fig. 6 A). 20kgP fractions from strains containing protein A fusions of Yor084p, Pex25p, Pex11p, or the integral membrane protein Pex13p (Gould et al., 1996) were hypotonically lysed to liberate matrix proteins. The membranes and associated components (Ti8P) were harvested by centrifugation. Chimeras of Pex25p and the membrane proteins Pex11p and Pex13p were enriched in the pellet fractions, whereas the majority of Yor084-pA was extracted, a behavior similar to that of peroxisomal matrix proteins (Eitzen et al., 1997). The Ti8P fractions were then extracted with sodium carbonate. This treatment liberates proteins associated with, but not integral to, the peroxisomal membrane (Fujiki et al., 1982). Under these conditions, Pex25-pA and Pex11-pA were extracted into the soluble fraction, whereas Pex13-pA was not extracted. These data suggest that Pex25p, as shown previously for Pex11p (Marshall et al., 1996), is peripherally associated with the peroxisomal membrane.

pex25Δ cells are impaired in matrix protein import and have large peroxisomes

To evaluate a potential role for Pex25p or Yor084p in PTS1 or PTS2 targeting, DsRed-PTS1 and PTS2–GFP were localized in the deletion strains of *PEX25* and *YOR084w* (Fig. 7). In both cases, the reporters localized to punctate structures. In *yor084Δ* cells, the structures appeared indistinguishable from wild-type peroxisomes (compare with Fig. 5 A); however, in *pex25Δ* cells two observations were made indicating that Pex25p is a peroxin. First, many *pex25Δ* cells contained few large peroxisomes in contrast to the numerous small peroxisomes seen in wild-type cells, a phenotype similar to that observed in *pex11Δ* cells. Interestingly, however, in contrast to *pex11Δ* cells this phenotype varied between individual cells in the population; some *pex25Δ* cells had no detectable peroxisomes, whereas others had numerous peroxisomes (unpublished data). By comparison, all *pex11Δ* cells contained a small number of detectable peroxisomes. The second phenotype observed was an increase in the cytoplasmic staining of both fluorescent matrix protein reporters, suggesting that in these cells peroxisomal matrix protein import was compromised and that this defect is not specific for either the PTS1 or PTS2 reporter. This defect was not observed in *pex11Δ* cells (Fig. 7) (Erdmann and Blobel, 1995).

The phenotypes of *pex25Δ* cells observed by fluorescence microscopy led us to observe the morphology of these cells by thin section EM (Fig. 8). Wild-type cells showed several characteristic small peroxisomes, each bound by a single unit membrane. In contrast, several *pex25Δ* cells contained very large peroxisomes, whereas others contained normal looking peroxisomes. Compared with wild-type cells, many more sections of *pex25Δ* cells did not exhibit detectable peroxisomes. This agrees with the fluorescence data because the presence of one or a few peroxisomes per cell would make it less likely to detect peroxisomes in single sections by EM. Cells lacking *PEX11* also contain few large peroxisomes; however, the phenotype of *pex11Δ* cells appears more uniform (Erdmann and Blobel, 1995; Marshall et al., 1995), and *pex11Δ* cells were observed to contain multiple buds that large peroxisomes were unable to enter (Erdmann and Blobel, 1995). We did not observe this phenotype in *pex25Δ* cells.

To investigate the fate of *pex25Δ* cells lacking fluorescently detectable peroxisomes, cells were incubated in the presence of oleate, to a point where peroxisomes were not

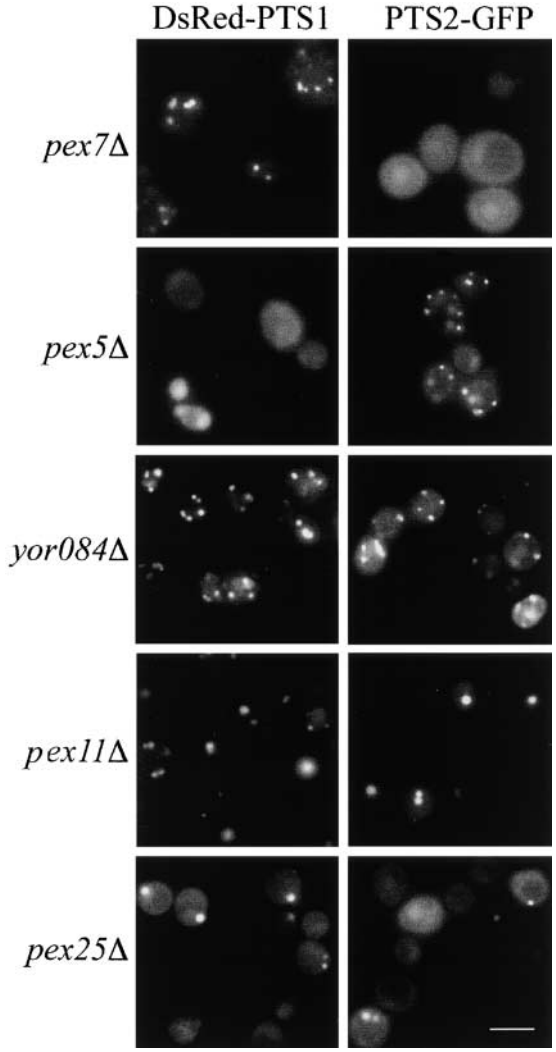

Figure 7. **Subcellular localization of PTS1- and PTS2-targeted proteins in *yor084wΔ* and *pex25Δ* cells.** DsRed-PTS1 and PTS2–GFP reporters were examined by fluorescence microscopy. DsRed-PTS1 localized to numerous small peroxisomes in *pex7Δ* and *yor084Δ* yeast cells. A similar localization was seen with PTS2–GFP in *pex5Δ* and *yor084Δ* cells. A diffuse cytoplasmic localization of these reporters was observed in strains lacking components necessary for the targeting of either PTS1 (*pex5Δ*) or PTS2 (*pex7Δ*). In *pex11Δ* cells, both reporters accumulated in a few large peroxisomes as reported previously (Erdmann and Blobel, 1995). In *pex25Δ* cells, a heterogeneous population was observed; some cells displayed a diffuse localization of the reporters, and some cells showed accumulation of the reporters to a few large peroxisomes. Bar, 10 μm.

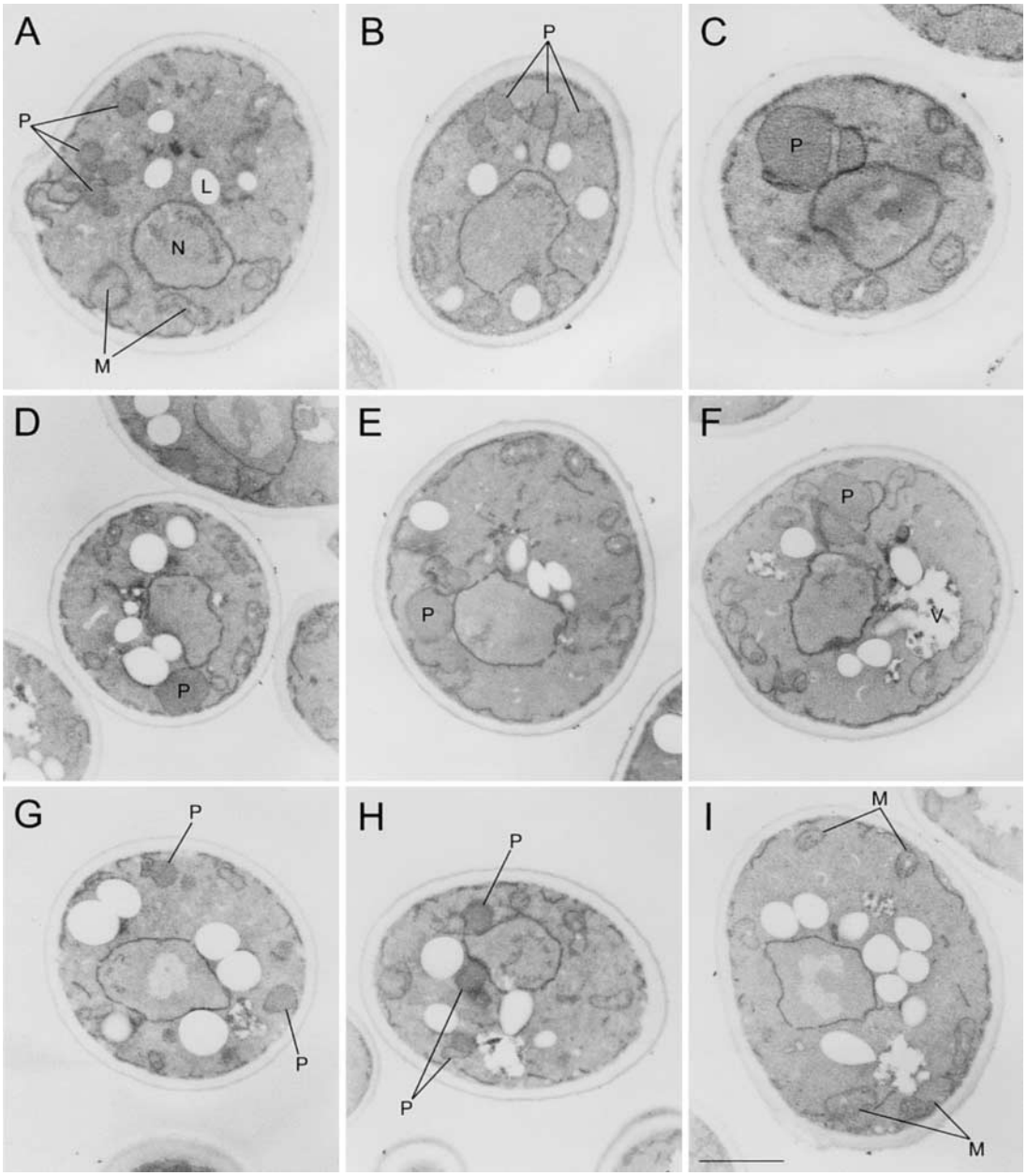

Figure 8. **Morphological analysis of *pex25Δ* cells.** Ultrastructure of wild-type (A and B) and *pex25Δ* (C–I) cells. Thin sections of cells prepared after 8 h induction in oleate. Note the heterogeneity in the peroxisome population in *pex25Δ* cells: large sized, (C–F); normal sized (G and H); and apparently absent (I). P, peroxisome; M, mitochondrion; L, lipid droplet; N, nucleus; V, vacuole. Bar, 1 μm.

detected in ~30–40% of the cells, and then plated onto oleate-containing medium to assess their viability. All cells in the population were capable of forming new colonies. Therefore, it is unlikely that these cells lacked peroxisomes altogether. These cells either contained small peroxisomes or peroxisomal precursors that were masked by the accumulation of matrix proteins in the cytoplasm, or they synthesized peroxisomes de novo. Further experiments are required to distinguish between these possibilities.

Discussion

The identification of coexpressed genes using microarray technology has a wide range of applications from the characterization of transcription factors and elements to the prediction of gene function. However, the comparison of two cell types or growth conditions often leads to the identification of several genes with altered expression without an indication of which genes are coregulated (for review see Schulze and Downward, 2001). This problem can be overcome by comparing several different conditions and through merging the results of different microarray experiments (Eisen et al., 1998; Hughes et al., 2000b). Based on this principle, we developed a method to predict gene involvement in peroxisome biogenesis or function. The change in the yeast transcriptome was measured at various time points during peroxisome induction and repression by microarray analysis. These data were

combined to generate an expression profile for each gene, and candidate genes were identified by pattern matching the profiles of genes known to be involved in peroxisome biogenesis or function.

Pattern matching was done using three complementary clustering algorithms with both normalized and nonnormalized data. Thus, in addition to highly induced genes, genes that have lower levels of induction but similar induction shapes were identified. Clusters enriched with genes known to be involved in peroxisome biogenesis or function were consistently generated using different algorithms and data transformations; however, no particular method identified the same group of candidate genes. Therefore, to maximize the number of novel genes identified that are involved in these processes the results of all six analyses were combined to generate a composite list of candidate genes involved in peroxisome biogenesis or function.

This transcriptomic approach is complementary to classic genetic screens that, to date, have been successful in identifying many peroxins. These genetic approaches have been designed to identify mutants unable to utilize oleate or other carbon sources metabolized by functional peroxisomes. The rationale for the transcriptomic approach lies in the fact that over two-thirds of all yeast genes are not essential. This is partly because it appears that cells have evolved an inherent "buffering capacity" that enables them to withstand the loss of individual components without catastrophic consequences (Kirschner and Gerhart, 1998; Hartman et al., 2001). Thus, we expected that many genes involved in peroxisome assembly or function may not be absolutely required for cells to grow on oleate. Consistent with this idea, of the viable candidate gene deletion strains, ~90% were able to grow in the presence of oleate as a sole carbon source, and ~7% of these strains displayed impaired growth rates under this condition (unpublished data). With respect to the candidates presented here, although both Yor084p and Pex25p are peroxisomal and induced on oleate, neither is required, and only *pex25Δ* cells showed slower growth rates on oleate, thereby explaining why these two genes had not been identified by functional growth-based assays.

In addition, 19 of the 24 *PEX* genes characterized to date have predicted roles in protein targeting and import rather than other aspects of peroxisome biogenesis such as proliferation and membrane biogenesis (Purdue and Lazarow, 2001; Tam and Rachubinski, 2002). It may be that the "classical" screen mentioned above and targeting assays developed subsequently (for review see Subramani, 1998; Snyder et al., 1999) are biased toward identifying genes with this particular function. Furthermore, *PEX* genes could have functions in other cellular processes rendering them essential, and these would be difficult to identify using growth-based assays. Screens like that presented here are not based on functional assays of nonessential genes and thus do not have the same biases. 11 of the genes in our candidate list are essential. It remains to be determined which of these are involved in peroxisome function. Therefore, we expect that the data presented here will lay the foundation for the characterization of more proteins that are involved in other aspects of peroxisome function and of proteins that may have shared roles in other cellular processes.

To assess the quality of our screen, we investigated two genes, *YPL112c* and *YOR084w*, for their linkage to peroxisome function. Yor084p is a peroxisomal matrix protein, but we have yet to define a phenotype associated with its deletion. On the other hand, Ypl112p is peripherally associated with the peroxisomal membrane, and based on its involvement in peroxisome maintenance we have renamed it Pex25p. Although cells in a population lacking *PEX25* display different degrees of phenotype, Pex25p is required to regulate peroxisome size. EM revealed that many *pex25Δ* cells contained one or two large peroxisomes, and some sections had no identifiable peroxisomes. Consistent results were obtained by immunofluorescence microscopy with PTS1- and PTS2-containing peroxisomal marker proteins. Many cells of the *pex25Δ* strain had fewer and larger peroxisomes than cells of the wild-type strain, and in up to half of cells peroxisomes were not detected at all and the reporters were cytoplasmic.

The observed heterogeneity in the population of *pex25Δ* cells suggests that they are unable to correctly apportion peroxisomes to both the mother and daughter cells during mitosis. The partitioning of vacuoles and mitochondria in *Saccharomyces cerevisiae* during mitosis involves directed movement, a "motor," a target site in the bud, and an "anchor" to retain a portion of the organelles in the mother cell and organelle fission and fusion events (Yaffe, 1999; Catlett and Weisman, 2000). It has been demonstrated recently that peroxisome partitioning during cell division engages the dynamin-like Vps1p to control peroxisome number (possibly through fission), actin filaments to direct peroxisome movement, and the Myo2p motor (Hoepfner et al., 2001).

Disruption of components of the peroxisome-partitioning machinery would likely lead to the phenotype observed in *pex25Δ* cells: large peroxisomes would accumulate and would not be segregated efficiently to daughter cells. However, the mechanisms of Pex25p action remain to be established. It is interesting that deletion mutants of *VPS1* (Hoepfner et al., 2001) and *PEX11* (Erdmann and Blobel, 1995) do partition peroxisomes into the bud and have not been observed to generate daughter cells without detectable peroxisomes. It has been suggested by Hoepfner et al. (2001) that this may be because multiple fission mediators exist or because mechanical fission occurs. In this case, the force generated by the actin-based motor Myo2p toward the emergent bud would stress intact peroxisomes tethered to the mother cell. It is unclear why large peroxisomes would be partitioned to daughters in *pex11Δ* cells but not in *pex25Δ* cells, but it is attractive to speculate that *pex25Δ* cells complete cytokinesis without fission because they have become uncoupled from the machinery responsible for mechanical movement and fission. Nevertheless, it appears clear that *pex25Δ* cells are not completely "cured" of peroxisomes. It is likely that in cells without detectable peroxisomes, precursor structures exist that mature into larger organelles but would not be readily detected by the standard methods employed here. The nature and origin of such precursors remain somewhat controversial; however, this *pex25* strain may prove useful to evaluate this maturation process.

Materials and methods

Strains and culture conditions

Strains used in this study are BY4743 and a deletion strain library (Winzeler et al., 1999) (obtained from Resgen). All strains were grown as described (Sherman et al., 1986) in YPD (2% yeast extract, 1% peptone, 2% glucose), synthetic minimal media (SM) containing the necessary amino acids and nucleotides, or YPBD (YPB [0.3% yeast extract, 0.5% potassium phosphate, pH 6.0, 0.5% peptone] containing 2% glucose) unless otherwise stated.

Collection of samples for gene expression profiles

For the time course of oleate induction, cells were grown in YPBG (YPB, 3% [wt/vol] glycerol) overnight to a density of 3×10^7 cells/ml, pelleted by centrifugation, transferred to YPBO (YPB, 0.12% [wt/vol] oleate, 0.2% [wt/vol] Tween 40), and grown for 9 h to a density of 3×10^7 cells/ml. Cells were harvested at 0, 0.5, 1, 3, 6, and 9 h after induction. For steady-state profiles of gene expression, cells were grown in YPBO, YPBG, or YPBD for 26 h to a final cell density of $\sim 3 \times 10^7$ cells/ml. After harvesting, cells were frozen in liquid nitrogen and stored at –80°C.

Extraction of RNA, generation of cDNA, and microarray analysis

Total RNA was isolated by hot acid phenol extraction (Ausubel et al., 2002). PolyA+ RNA was isolated from total RNA, and cDNAs labeled with Cy3 or Cy5 dCTP were generated by reverse transcription as described (Ideker et al., 2000). cDNAs from pairs of cell populations, differently labeled with Cy3 and Cy5, were combined and used to hybridize yeast whole-genome oligonucleotide microarrays. Hybridization was for 14 h at 37°C in DIG Easy Hybe (Roche) containing 0.5 mg yeast tRNA/ml and 100 μg salmon sperm DNA/ml. Microarray slides were washed as described (Ideker et al., 2000) and scanned using a ScanArray 5000 microarray analyzer (Packard BioScience). Images were processed using the microarray spot finding and quantification software Dapple (http://www.cs.wustl.edu/~jbuhler/research/dapple/). Fluorescent spots were located, and for each dye background intensity was estimated and subtracted from the median intensity within each spot area. Values were then normalized so that the medians of all Cy3 and all Cy5 intensities were equal (Ideker et al., 2000, 2001). MAGE-ML data are available at http://www.systemsbiology.org.

Summary of microarray experiments and experimental repetitions

Each microarray experiment is a comparison of the cDNA derived from yeast grown under two different conditions. The eight experiments that were performed are listed in Table I. For every pair of cDNAs compared, the hybridization was repeated with cDNAs derived from cells of a separate culture. In addition, each microarray slide contained four replicate spots per gene, and each hybridization was repeated with the reverse labeling scheme. Therefore, for each experiment the expression levels of each gene were measured from 16 individual spots.

Selection of differentially expressed genes within and between experiments

For each experiment, the 16 replicates were combined, and genes showing significant differential expression were established using the programs VERA and SAM (http://www.systemsbiology.org/VERAandSAM) ($\lambda > 50$) as described previously (Ideker et al., 2000). The data for each of the eight experiments (~6,200 genes each) were combined and then filtered to contain only those genes that showed significant differential expression for at least one of the eight experiments. The resulting data (containing 3,031 genes) were again filtered to exclude genes with little or no change in expression ratios across the eight experiments. For this filter, genes for which difference between the maximum and minimum $\log_{10}$ expression ratios was >0.5 were rejected. The $\log_{10}$ expression ratios of the twice-filtered dataset (containing 2,300 genes) were used for all clustering analyses.

Generation and analysis of SOMs

The expression profiles of the genes in the filtered dataset were clustered based on the Euclidean distance between their $\log_{10}$ expression ratios over all eight experiments using GeneCluster software (version 1.0; Tamayo et al., 1999). A six row by four column SOM (SOM1) was generated using the default parameters but running 10^7 iterations. In addition, $\log_{10}$ expression ratios for each gene over all eight experiments were normalized to have a mean of zero and a standard deviation of one, and a second SOM (SOM2) was generated in the same manner.

PEX genes, genes encoding peroxisomal proteins, and proteins with known cellular roles (based on their classifications in the literature and in the Proteome database [Costanzo et al., 2000]) were annotated. To establish if genes with the same annotation were enriched in any cluster, the probabilities that the observed distributions would be found by chance were determined by calculating the hypergeometric distributions for each annotation in each cluster. If $P < 5 \times 10^{-4}$, the annotation was considered to be enriched.

Unsupervised hierarchical clustering

The expression profiles of the genes in the filtered dataset were compared with those of genes encoding peroxins and exclusively peroxisomal proteins using Cluster and Tree View software (Eisen et al., 1998). For the generation of the normalized tree, the default parameters were used with the following specifications: the genes were organized into a one-dimensional SOM (using 47 nodes and running 2×10^5 iterations), experiments were weighted ($k = 0.6$; $n = 1$), and genes were clustered using average linkage clustering with centered correlation as the similarity metric. The nonnormalized tree was generated in the same way except uncentered correlation was used as the similarity metric.

Nearest Neighbor analysis

For genes encoding peroxins or exclusively peroxisomal proteins in the filtered dataset, the Nearest Neighbor component of GEAP software (Galitski et al., 1999) was used to obtain a list of genes (to a maximum of 20) with the most similar expression profiles using either the Pearson or Euclidian distance metric. For each metric used, a threshold was established for each gene queried by comparing its expression profile to those in a randomized dataset (containing 10 random within-gene permutations for each of the 2,300 genes). Candidate genes were chosen from each Nearest Neighbor set that had correlation coefficients (or geometric distances) that were above (or below) a threshold that yielded 5 false positives per 10 randomized datasets.

Plasmids

The following plasmids were described previously: pRS315 *CEN/LEU2* and pRS316 *CEN/URA3* (Sikorski and Hieter, 1989), pProtA/HIS5 (Rout et al., 2000), pBluescript II SK (Stratagene), and pGFP/HIS5 (Dilworth et al., 2001). pDsRed-PTS1 was constructed by PCR directed mutagenesis (Ausubel et al., 2002) to yield the *Discosoma sp. FP583* gene (DsRed; CLONETECH Laboratories, Inc.), containing a 3′ extension encoding a tripeptide PTS1 (Ser-Lys-Leu) flanked by the ORE-containing *FAA2* promoter (−1 to −398) and terminator (+2236 to +2690). For the construction of pPTS2–GFP, a fragment containing the 5′ untranslated region of the *S. cerevisiae POT1* gene encoding peroxisomal thiolase and the portion of the gene encoding the 19 NH_2-terminal amino acids (−475 to +57) was amplified and joined to the 5′ end of the sequence encoding *Aequoria victoria* GFP by PCR-based mutagenesis (Ausubel et al., 2002). The product was ligated into pRS315 to generate pPTS2–GFP.

Protein A and GFP tagging of candidate genes

Genes were genomically tagged with the sequences encoding *Staphylococcus aureus* protein A or GFP by homologous recombination using previously described PCR-based integrative transformation procedure into wild-type BY4743 diploid cells (Aitchison et al., 1995; Dilworth et al., 2001).

Microscopy

For fluorescence microscopy, yeast cells transformed with plasmids were grown in SM containing 0.1% (wt/vol) Tween 40 and 0.15% (wt/vol) oleate for 16 h at 30°C. Cells were collected, washed, and visualized either by direct or indirect fluorescence microscopy as described (Kilmartin and Adams, 1984) with modifications (Aitchison et al., 1995). Protein A fusions were detected with rabbit IgG and FITC-conjugated goat anti–rabbit antibodies. Yeast strains synthesizing genomic GFP chimeras were induced in YPBO for 16 h and observed directly. Fluorescent cells were visualized using an Axiophot II microscope (ZEISS), and optical sections were obtained with a confocal microscope (Leica). Images were obtained using Metaview software (Universal Imaging Corp.). EM was performed as described previously (Eitzen et al., 1997).

Subcellular fractionation and isolation of peroxisomes

Peroxisomes were isolated essentially as described in Bonifacino et al. (2000). Briefly, yeast cells were grown in YPD overnight and seeded into 1 liter of SCIM (0.5% yeast extract, 0.1% peptone, 0.79 g complete supplement mixture [Q-biogene]/L, 0.5% ammonium sulfate, 1.7 g yeast nitrogen base/L, 0.1% [wt/vol] Tween 40, 0.1% glucose, and 0.15% [wt/vol] oleate) and grown overnight at 30°C. Cells were harvested, washed, and converted to spheroplasts with 1 mg Zymolyase 100 T/g of cells for 1 h at 30°C. Spheroplasts were lysed by homogenization in MS buffer (0.65 M sorbitol, 5 mM MES, pH 5.5) containing 1 mM KCl, and PINS (1 mM EDTA, 0.2 mM

PMSF, 2 μg leupeptin/ml, 2 μg aprotinin/ml, and 0.4 μg pepstatin A/ml). Cell debris and nuclei were pelleted from the homogenate by centrifugation for 10 min at 2,000 *g* to generate a postnuclear supernatant, which was subjected to 20,000 g_{max} for 30 min to yield a supernatant (20kgS) and a pellet (20kgP). The 20kgP was resuspended in MS buffer, and a volume containing 1 mg of protein was overlaid onto a 12-ml step gradient consisting of 17, 25, 35, and 50% Nycodenz in MS buffer. Organelles were separated by isopycnic centrifugation at 116,000 *g* for 2 h in a NVT65 rotor (Beckman Coulter). Fractions of 1 ml were collected from the bottom of the gradient and analyzed by SDS-PAGE and Western blotting. Anti-Fox2p antibodies were raised in rabbits to Fox2p COOH-terminally fused to maltose binding protein as described previously (Eitzen et al., 1995). Antibodies to Sdh2p have been described previously (Dibrov et al., 1998).

Extraction peroxisomes

Peroxisomes were extracted as described by Fujiki et al. (1982) and Nuttley et al. (1990). Organelles in 40 μg of 20kgP were lysed by incubation in 10 vol of Ti8 buffer (10 mM Tris, pH 8) containing PINS on ice for 1 h and separated by centrifugation at 200,000 *g* for 1 h at 4°C in a TLA 100.2 rotor (Beckman Coulter) into a pellet (Ti8P) and a supernatant (Ti8S). Ti8P was resuspended in Ti8 buffer, and a fraction was extracted with 0.1 M Na_2CO_3, pH 11.3, for 30 min on ice and separated by centrifugation at 200,000 *g* and 4°C for 1 h into a supernatant (CO_3S) and pellet (CO_3P). Proteins in both Ti8S and CO_3S were TCA precipitated and washed with 90% methanol. Proteins in equal portions of each fraction were separated by SDS-PAGE and analyzed by Western blotting.

Online supplemental material

Table S1 (available at http://www.jcb.org/cgi/content/full/jcb.200204059/DC1) is a summary of clustering and array data. The 3,031 genes that showed significant differential expression for at least one of the eight experiments are listed. For each gene, the $\log_{10}$ expression ratios and the λ significance values for each of the eight experiments (described in Table I) are reported. The cluster numbers of all genes in each two-dimensional SOM (SOM1 and SOM2) are listed. Candidate genes selected using hierarchical and Nearest Neighbor clustering are annotated with a description of the "hits" each received by each of the two methods.

We are grateful to Honey Chan for her expert assistance in EM and to Bruz Marzolf and Allison Golden for preparing microarray slides, Bernard Lemire for anti-Sdh2p, and Vesteinn Thorsson and George Lake for stimulating discussions and insightful comments.

The authors acknowledge support for this research from Merck and the Institute for Systems Biology to J.D. Aitchison and T. Galitski, a Canadian Institutes for Health Research (CIHR) postdoctoral award to J.J. Smith, CIHR and Alberta Heritage Foundation for Medical Research studentship awards to D.J. Dilworth, and operating grants from the CIHR to R.A. Rachubinski and J.D. Aitchison. T. Galitski is a recipient of a Burroughs Wellcome Fund Career Award in the Biomedical Sciences. R.A. Rachubinski is an International Research Scholar of the Howard Hughes Medical Institute and Canada Research Chair in Cell Biology. We apologize to colleagues whose original research we were unable to cite due to constraints of article length.

Submitted: 11 April 2002
Revised: 7 June 2002
Accepted: 7 June 2002

References

Aitchison, J.D., M.P. Rout, M. Marelli, G. Blobel, and R.W. Wozniak. 1995. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore-membrane protein Pom152p. *J. Cell Biol.* 131:1133–1148.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, and K. Struhl. 2002. Introduction of a point mutation by sequential PCR steps. *In* Current Protocols in Molecular Biology. John Wiley & Sons, Inc., New York.

Bonifacino, J.S., M. Dasso, J. Lippincott-Schwartz, J.B. Harford, and K.M. Yamada. 2000. Isolation of oleate-induced peroxisomes using sucrose gradient step gradients. *In* Current Protocols in Cell Biology. John Wiley & Sons, Inc., New York. 3.8.32–3.8.36.

Catlett, N.L., and L.S. Weisman. 2000. Divide and multiply: organelle partitioning in yeast. *Curr. Opin. Cell Biol.* 12:509–516.

Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science.* 263:802–805.

Costanzo, M.C., J.D. Hogan, M.E. Cusick, B.P. Davis, A.M. Fancher, P.E. Hodges, P. Kondu, C. Lengieza, J.E. Lew-Smith, C. Lingner, et al. 2000. The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* 28:73–76.

Dibrov, E., S. Fu, and B.D. Lemire. 1998. The *Saccharomyces cerevisiae* TCM62 gene encodes a chaperone necessary for the assembly of the mitochondrial succinate dehydrogenase (complex II). *J. Biol. Chem.* 273:32042–32048.

Dilworth, D.J., A. Suprapto, J.C. Padovan, B.T. Chait, R.W. Wozniak, M.P. Rout, and J.D. Aitchison. 2001. Nup2p dynamically associates with the distal regions of the yeast nuclear pore complex. *J. Cell Biol.* 153:1465–1478.

Einerhand, A.W.C., W.T. Kos, B. Distel, and H.F. Tabak. 1993. Characterization of a transcriptional control element involved in proliferation of peroxisomes in yeast in response to oleate. *Eur. J. Biochem.* 214:323–331.

Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA.* 95:14863–14868.

Eitzen, G.A., J.D. Aitchison, R.K. Szilard, M. Veenhuis, W.M. Nuttley, and R.A. Rachubinski. 1995. The *Yarrowia lipolytica* gene PAY2 encodes a 42-kDa peroxisomal integral membrane protein essential for matrix protein import and peroxisome enlargement but not for peroxisome membrane proliferation. *J. Biol. Chem.* 270:1429–1436.

Eitzen, G.A., R.K. Szilard, and R.A. Rachubinski. 1997. Enlarged peroxisomes are present in oleic acid-grown *Yarrowia lipolytica* overexpressing the *PEX16* gene encoding an intraperoxisomal peripheral membrane peroxin. *J. Cell Biol.* 137:1265–1278.

Erdmann, R., and G. Blobel. 1995. Giant peroxisomes in oleic acid-induced *Saccharomyces cerevisiae* lacking the peroxisomal membrane protein Pmp27p. *J. Cell Biol.* 128:509–523.

Fajas, L., M.B. Debril, and J. Auwerx. 2001. Peroxisome proliferator-activated receptor-γ: from adipogenesis to carcinogenesis. *J. Mol. Endocrinol.* 27:1–9.

Filipits, M., M.M. Simon, W. Rapatz, B. Hamilton, and H. Ruis. 1993. A *Saccharomyces cerevisiae* upstream activating sequence mediates induction of peroxisome proliferation by fatty acids. *Gene.* 132:49–55.

Fujiki, Y., A.L. Hubbard, S. Fowler, and P.B. Lazarow. 1982. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* 93:97–102.

Galitski, T., A.J. Saldanha, C.A. Styles, E.S. Lander, and G.R. Fink. 1999. Ploidy regulation of gene expression. *Science.* 285:251–254.

Geraghty, M.T., D. Bassett, J.C. Morrell, G.J. Gatto, Jr., J. Bai, B.V. Geisbrecht, P. Hieter, and S.J. Gould. 1999. Detecting patterns of protein distribution and gene expression *in silico. Proc. Natl. Acad. Sci. USA.* 96:2937–2942.

Glover, J.R., D.W. Andrews, and R.A. Rachubinski. 1994a. *Saccharomyces cerevisiae* peroxisomal thiolase is imported as a dimer. *Proc. Natl. Acad. Sci. USA.* 91:10541–10545.

Glover, J.R., D.W. Andrews, S. Subramani, and R.A. Rachubinski. 1994b. Mutagenesis of the amino targeting signal of *Saccharomyces cerevisiae* 3-ketoacyl-CoA thiolase reveals conserved amino acids required for import into peroxisomes *in vivo. J. Biol. Chem.* 269:7558–7563.

Gonzalez, F.J., J.M. Peters, and R.C. Cattley. 1998. Mechanism of action of the nongenotoxic peroxisome proliferators: role of the peroxisome proliferator-activator receptor α. *J. Natl. Cancer Inst.* 90:1702–1709.

Gould, S.J., and D. Valle. 2000. Peroxisome biogenesis disorders: genetics and cell biology. *Trends Genet.* 16:340–345.

Gould, S.J., G.A. Keller, N. Hosken, J. Wilkinson, and S. Subramani. 1989. A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* 108:1657–1664.

Gould, S.J., J.E. Kalish, J.C. Morrell, J. Bjorkman, A.J. Urquhart, and D.I. Crane. 1996. Pex13p is an SH3 protein of the peroxisome membrane and a docking factor for the predominantly cytoplasmic PTS1 receptor. *J. Cell Biol.* 135: 85–95.

Hartman, J.L.T., B. Garvik, and L. Hartwell. 2001. Principles for the buffering of genetic variation. *Science.* 291:1001–1004.

Hoepfner, D., M. van den Berg, P. Philippsen, H.F. Tabak, and E.H. Hettema. 2001. A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in *Saccharomyces cerevisiae. J. Cell Biol.* 155:979–990.

Hughes, J.D., P.W. Estep, S. Tavazoie, and G.M. Church. 2000a. Computational identification of *cis*-regulatory elements associated with groups of function-

ally related genes in *Saccharomyces cerevisiae. J. Mol. Biol.* 296:1205–1214.

Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, et al. 2000b. Functional discovery via a compendium of expression profiles. *Cell.* 102:109–126.

Ideker, T., V. Thorsson, A.F. Siegel, and L.E. Hood. 2000. Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J. Comput. Biol.* 7:805–817.

Ideker, T., V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, and L. Hood. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science.* 292:929–934.

Kal, A.J., A.J. van Zonneveld, V. Benes, M. van den Berg, M.G. Koerkamp, K. Albermann, N. Strack, J.M. Ruijter, A. Richter, B. Dujon, et al. 1999. 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.

Keller, G.A., T.G. Warner, K.S. Steimer, and R.A. Hallewell. 1991. Cu,Zn superoxide dismutase is a peroxisomal enzyme in human fibroblasts and hepatoma cells. *Proc. Natl. Acad. Sci. USA.* 88:7381–7385.

Kilmartin, J.V., and A.E. Adams. 1984. Structural rearrangements of tubulin and actin during the cell cycle of the yeast *Saccharomyces. J. Cell Biol.* 98:922–933.

Kira Y., E.F. Sato, and M. Inoue. 2002. Association of Cu, Zn-type superoxide dismutase with mitochondria and peroxisomes. *Arch. Biochem. Biophys.* 399: 96–102.

Kirschner, M., and J. Gerhart. 1998. Evolvability. *Proc. Natl. Acad. Sci. USA.* 95: 8420–8427.

Leslie, D.M., B. Grill, M.P. Rout, R.W. Wozniak, and J.D. Aitchison. 2002. Kap121p-mediated nuclear import is required for mating and cellular differentiation in yeast. *Mol. Cell. Biol.* 22:2544–2555.

Marshall, P.A., Y.I. Krimkevich, R.H. Lark, J.M. Dyer, M. Veenhuis, and J.M. Goodman. 1995. Pmp27 promotes peroxisomal proliferation. *J. Cell Biol.* 129:345–355.

Marshall, P.A., J.M. Dyer, M.E. Quick, and J.M. Goodman. 1996. Redox-sensitive homodimerization of Pex11p: a proposed mechanism to regulate peroxisomal division. *J. Cell Biol.* 135:123–137.

Matz, M.V., A.F. Fradkov, Y.A. Labas, A.P. Savitsky, A.G. Zaraisky, M.L. Markelov, and S.A. Lukyanov. 1999. Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nat. Biotechnol.* 17:969–973.

Nuttley, W.M., A.G. Bodnar, D. Mangroo, and R.A. Rachubinski. 1990. Isolation and characterization of membranes from oleic acid-induced peroxisomes of *Candida tropicalis. J. Cell Sci.* 95:463–470.

Purdue, P.E., and P.B. Lazarow. 2001. Peroxisome biogenesis. *Annu. Rev. Cell Dev. Biol.* 17:701–752.

Rehling, P., A. Skaletz-Rorowski, W. Girzalsky, T. Voorn-Brouwer, M.M. Franse, B. Distel, M. Veenhuis, W.-H. Kunau, and R. Erdmann. 2000. Pex8p, an intraperoxisomal peroxin of *Saccharomyces cerevisiae* required for protein transport into peroxisomes binds the PTS1 receptor Pex5p. *J. Biol. Chem.* 275:3593–3602.

Rout, M.P., J.D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, and B.T. Chait. 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 148:635–651.

Schulze, A., and J. Downward. 2001. Navigating gene expression using microarrays—a technology review. *Nat. Cell Biol.* 3:E190–E195.

Sherman, F., G.R. Fink, and J.B. Hicks. 1986. Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 186 pp.

Sikorski, R.S., and P. Hieter. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae. Genetics.* 122:19–27.

Snyder, W.B., A. Koller, A.J. Choy, M.A. Johnson, J.M. Cregg, L. Rangell, G.A. Keller, and S. Subramani. 1999. Pex17p is required for import of both peroxisome membrane and lumenal proteins and interacts with Pex19p and the peroxisome targeting signal-receptor docking complex in *Pichia pastoris. Mol. Biol. Cell.* 10:4005–4019.

Subramani, S. 1998. Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiol. Rev.* 78:171–188.

Tam, Y.Y., and R.A. Rachubinski. 2002. *Yarrowia lipolytica* cells mutant for the *PEX24* gene encoding a peroxisomal membrane peroxin mislocalize peroxisomal proteins and accumulate membrane structures containing both peroxisomal matrix and membrane proteins. *Mol. Biol. Cell.* In press.

Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA.* 96:2907–2912.

Titorenko, V.I., and R.A. Rachubinski. 2001. The life cycle of the peroxisome. *Nat. Rev. Mol. Cell Biol.* 2:357–368.

Veenhuis, M., M. Mateblowski, W.H. Kunau, and W. Harder. 1987. Proliferation of microbodies in *Saccharomyces cerevisiae*. *Yeast.* 3:77–84.

Winzeler, E.A., D.D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J.D. Boeke, H. Bussey, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science.* 285:901–906.

Yaffe, M.P. 1999. The machinery of mitochondrial inheritance and behavior. *Science.* 283:1493–1497.