UC Berkeley
UC Berkeley Previously Published Works

Title

A universal TagModule collection for parallel genetic analysis of microorganisms

Permalink

https://escholarship.org/uc/item/0nf6j20m

Journal

Nucleic Acids Research, 38(14)

ISSN

0305-1048

Authors

Oh, Julia
Fung, Eula
Price, Morgan N
et al.

Publication Date

2010-08-01

DOI

10.1093/nar/gkq419

Peer reviewed

A universal TagModule collection for parallel genetic analysis of microorganisms

Julia Oh[1,2], Eula Fung[2], Morgan N. Price[3,4], Paramvir S. Dehal[3,4], Ronald W. Davis[1,2], Guri Giaever[5], Corey Nislow[6,*], Adam P. Arkin[3,4,*] and Adam Deutschbauer[3,4,*]

[1]Department of Genetics, Stanford University, Mail Stop-5120, Palo Alto, CA 94305, [2]Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304, [3]Physical Biosciences Division, [4]Virtual Institute for Microbial Stress and Survival, 1 Cyclotron Road, MS 977-152, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA, [5]Department of Pharmaceutical Sciences, University of Toronto, 160 College St. Toronto, ON M5S3E1 and [6]Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S1A8, Canada

ABSTRACT

Systems-level analyses of non-model microorganisms are limited by the existence of numerous uncharacterized genes and a corresponding over-reliance on automated computational annotations. One solution to this challenge is to disrupt gene function using DNA tag technology, which has been highly successful in parallelizing reverse genetics in *Saccharomyces cerevisiae* and has led to discoveries in gene function, genetic interactions and drug mechanism of action. To extend the yeast DNA tag methodology to a wide variety of microorganisms and applications, we have created a universal, sequence-verified TagModule collection. A hallmark of the 4280 TagModules is that they are cloned into a Gateway entry vector, thus facilitating rapid transfer to any compatible genetic system. Here, we describe the application of the TagModules to rapidly generate tagged mutants by transposon mutagenesis in the metal-reducing bacterium *Shewanella oneidensis* MR-1 and the pathogenic yeast *Candida albicans.* Our results demonstrate the optimal hybridization properties of the TagModule collection, the flexibility in applying the strategy to diverse microorganisms and the biological insights that can be gained from fitness profiling tagged mutant collections. The publicly available TagModule collection is a platform-independent resource for the functional genomics of a wide range of microbial systems in the post-genome era.

INTRODUCTION

A fundamental goal in microbiology is the functional annotation of uncharacterized genes identified by genomic sequencing. Due to the anticipated explosion in clonal microbial genome sequences and metagenomes as a result of next-generation sequencing, there is an urgent need for methodologies that rapidly and systematically determine gene function across a range of diverse microorganisms. One promising strategy to meet these challenges is high-throughput reverse genetics using DNA tag, or DNA barcode, technology, in which strains or samples are marked with unique DNA identifiers. One well-established application of DNA tags has been the creation of mutant libraries and their use in pooled phenotypic assays (1,2). These tagged pools allow the phenotypes of hundreds or thousands of mutants to be assayed simultaneously. A notable example is the yeast *Saccharomyces cerevisiae* deletion collection, in which each gene was knocked out via homologous recombination with a deletion cassette that contains a pair of tags flanked by universal priming sites, which allow amplification of the tags at the end of a pooled assay (3). Changes in tag abundance, which reflect the fitness of a mutant in the chosen condition, are determined by hybridization to a microarray that contains the complements of the tag sequences used (2).

These tagged mutants, when assayed in parallel, have been powerful tools for assessing gene function, genetic interactions and drug target or mechanism of action (4,5). This well-established system can be adapted to a wide range of microorganisms, which would advance understanding of their biology and facilitate development of new treatment strategies for pathogenic organisms.

While efforts have been ongoing for several species, the tagging of microorganisms other than *S. cerevisiae* has been limited in size and scope due to technical and biological constraints. First, the method used to create the *S. cerevisiae* deletion collection requires that a unique set of primers and tagged deletion cassettes be generated for each strain. Second, homologous recombination is inefficient for many organisms and requires prior knowledge of genome sequence. An alternative used primarily in bacteria, signature-tagged mutagenesis, has been limited by low numbers of available tags, or limited detection capabilities (1,6). Recently, high-throughput sequencing of a saturated transposon library used flanking genomic sequence as a 'tag' (7), but this approach does not generate the single-mutant archived collections necessary for assay validation and in-depth studies.

To address these issues, we have created a sequence-verified, Gateway-compatible TagModule collection that can be readily adapted to any DNA tagging strategy, as it is organism and platform-independent. In conjunction with transposon mutagenesis, this collection can be a powerful tool to rapidly generate tagged mutants in a range of microorganisms. Here, we describe our use of TagModule-based transposon mutagenesis for the genetic analysis of two organisms, *Shewanella oneidensis* MR-1 and *Candida albicans*. *Candida albicans* is one of the most common causes of nosocomial infections (8) with a high rate of acquired drug resistance (9). *Shewanella oneidensis* MR-1 is a metabolically versatile bacterium with potential applications in bioremediation (10). While signature-tagged mutagenesis has been applied to *S. oneidensis* MR-1 to identify genes involved in soil survival (6), the low number of tags used for detection limits the scope of genome-wide studies of this microorganism to a small number of conditions. Similarly, large-scale genetic analysis and understanding of virulence of *C. albicans*, the most common fungal human pathogen, has been confounded due to its obligate diploid state, its lack of a traditional meiotic cycle and its diverse morphogenic forms. A proprietary partial (2868 strains) tagged deletion collection (11), while shown to be effective in identifying novel antifungal targets (12–14), was created based on criteria that would limit examination of *C. albicans*-specific genes, as strains were selected based on homology to genes essential in *S. cerevisiae* or conservation with other fungi and higher eukaryotes. Other transposon-mutagenized collections are untagged (15) or have only partially identified insertions (16). Additional resources are needed to identify novel drug targets and to understand the genetic basis of *C. albicans* virulence.

MATERIALS AND METHODS

All strains, plasmids and primers are detailed in Supplementary Tables 7–9. Details on reaction conditions and cloning methods are available upon request. Please contact AMD or APA for instructions on how to obtain the TagModule collection. Initially, our lab will freely supply the TagModule collection as *Escherichia coli* glycerol stocks in 384-well format. If demand is high, we will seek a third-party distributor for the TagModule collection. Under both scenarios, the entire TagModule collection will be available at low cost for the non-limited use by any academic or non-for-profit investigator.

Creating the TagModule resource

Each TagModule contains two unique 20-bp sequences (uptag and downtag) flanked by common polymerase chain reaction (PCR) priming sites. All TagModules were cloned into pENTR_TagModule, a modified Gateway entry vector. To make pENTR_TagModule, an 83-bp double-strand linker (5′-GTAGTACCTCAATGC GCGGCCGCATGCCCTGCAGGATGCATGCATGC GGCGCGCCATGCGCGATCGCATGCTAGCCTATC GTA-3′) was cloned into vector pCR 8/GW/TOPO (Invitrogen) using TOPO TA Cloning methodology. The 83-bp linker includes restriction enzyme sites for SbfI and AscI (underlined). We used two methods for cloning the 4280 TagModules into pENTR_TagModule. The first, used to construct the first 2400 TagModules, utilized standard synthesized oligonucleotides and individual reactions/sample tracking for each TagModule (referred to as the 'Standard' method). The second, used to construct the remaining 1880 TagModules, was accomplished using a random cloning method with long oligonucleotides synthesized on a microarray (referred to as the 'Agilent' method). Regardless of the method used, all TagModules are identical Gateway entry vectors except for a unique 20-bp uptag and a unique 20-bp downtag sequence. A comparison of the two methods is shown in Supplementary Figure 1.

For the Standard method, the TagModules were constructed by PCR using two unique oligonucleotides as the template and two common primers containing SbfI and AscI restriction sites for cloning into pENTR_TagModule. The 76-bp TagModule_UPTAG oligonucleotide contains a unique uptag flanked by the common uptag priming sites U1 and U2_revcomp and a 20-bp adaptamer region at the 3′ end. The 79-mer TagModule_DOWNTAG oligonucleotide contains a unique downtag flanked by the common downtag priming sites D1 and D2_revcomp and 20 bp of reverse complement to the TagModule_UPTAG 3′ adaptamer region. PCR with TagModule_UPTAG and TagModule_DOWNTAG as template and primers UPTAG_fix and DOWNTAG_fix produces a 175-bp product containing the entire TagModule. TagModule PCR products were purified, SbfI/AscI-digested, ligated into an SbfI/AscI-digested pENTR_TagModule, transformed into chemically competent *E. coli* (Edge Biosystems) and selected on LB + spectinomycin (100 μg/ml). Cloned TagModules were verified by colony PCR with primers vec_for2 and vec_rev2; PCR products were sequenced with primer GW3. A custom Perl program was used for TagModule sequence analysis, which identified clones with no mutations in either the tags or common priming sites. Sequence-verified clones for each TagModule were picked into 384-well plates for storage in LB + 10% glycerol.

For the Agilent method, we took advantage of the ability to synthesize thousands of long oligonucleotides on the surface of a microarray (17). Using oligo library synthesis (Agilent), we obtained a pool of 5911 unique 135-mer oligonucleotides of the structure (5′-<u>GATGTCC ACGAGGTCTCT</u>-UPTAG-<u>CGTACGCTGCAGGTCG ACCATGGTGGTCAGCTGGAATTGAAAACGAGC TCGAATTCATCG</u>-DOWNTAG-<u>CTACGAGACCGA CACCG</u>-3′), where the underlined sequences represent the common tag priming sites U1, U2_revcomp, D2 and D1_revcomp flanking the unique 20-bp uptag and downtag sequences. The oligonucleotide pool was amplified using the same primers and PCR conditions described for the Standard TagModule amplification. The correct size PCR product of 160 bp was gel purified, SbfI/AscI digested, and cloned into SbfI/AscI-digested pENTR_TagModule. We used the same colony PCR and sequencing protocol described for the Standard method to identify mutation-free TagModules.

Validation of TagModule collection

Escherichia coli stocks of all 4280 TagModule entry clones were combined in equivalent amounts, and pooled plasmid DNA was isolated. The uptags and downtags were PCR-amplified with U1′ and BTEG-U2′, and D1′ and BTEG-D2′, respectively. Tag PCR products were hybridized to an Affymetrix 16K TAG4 array as previously described (18). For analysis of tag performance, outliers were masked and discarded based on standard criteria (18), and then raw intensity values for the remaining replicates of each uptag and downtag were averaged. Background was calculated as the median of unused tag probes on the array. Individual uptags and downtags with raw intensity values <5× background (300 intensity units) were flagged as unusable. For analysis of cross-hybridization, the sequences of all unused tags with raw intensity values over 300 were compared for one, two, three or four or more mismatches to tags currently used in the tag module collection. The majority of the observed cross-hybridization was due to the presence of 'repaired' tags (19) on the array that closely resemble the sequences of an uptag or downtag contained in our TagModule collection.

Construction of *S. oneidensis* MR-1 mutant collection

We used two different transposon systems to generate insertion mutations in *S. oneidensis* MR-1. The first was a mini-Tn5 delivery system carried on the plasmid pRL27 (20). The second was a modified mariner system carried on the plasmid pMiniHimar_RB1 (21). Both systems have previously been shown to function in *S. oneidensis* MR-1 (21,22). To make pRL27 into a Gateway destination vector, we PCR-amplified Reading Frame B from the Gateway vector conversion system (Invitrogen), digested the product with KpnI, ligated into KpnI-digested pRL27 and transformed into λ*pir ccd*B-survival cells. To make pMiniHimar_RB1 into a Gateway destination vector, we cut with NsiI to remove a ~300-bp fragment (including one of the IR ends of the miniHimar transposon) and cloned in a 60-bp linker fragment with an SbfI site.

Reading Frame B was amplified, cut with SbfI and cloned into the SbfI-cut linker plasmid. The IR end removed during NsiI digestion was added onto one of the primers used to amplify Reading Frame B. The resulting vectors, pRL27_Dest and pMiniHimar_RB1_Dest, contains *att*R1 and *att*R2 sites flanking the *ccd*B toxin gene (see Supplementary Figure S3 for an outline of this process).

pENTR_TagModule_0001 to pENTR_TagModule_ 1824 were recombined with NcoI-digested pRL27_Dest via the LR clonase reaction following the manufacturer's guidelines. The LR recombination reactions were transformed into electrocompetent WM3064 *E. coli* and selected on LB + 50 μg/ml kanamycin + 300 μM diaminopimelic acid (DAP). LR reactions were performed either individually for each TagModule or in pools of 12 different TagModules. pENTR_TagModule _1825 to pENTR_ TagModule_2400 were recombined with BglII-digested pMiniHimar_RB1_Dest following the same protocols as for pRL27_Dest, with pools of 12 TagModules used in each LR reaction. The final products, pRL27_ TagModule_XXXX and pMiniHimar_RB1_TagModule_ XXXX, are transposon delivery vectors containing 1 of 2400 different TagModules.

We mutagenized *S. oneidensis* MR-1 with tagged transposons by conjugation with WM3064. A pool of donor WM3064 (carrying a mixture of different TagModules) was grown to saturation in LB + 50 μg/ml kanamycin + 300 μM DAP. The recipient *S. oneidensis* MR-1 was grown to saturation in LB. Equal volumes of the donor and recipient were mixed and 5 μl was spotted onto a 0.2 μM filter overlaid on an LB + 300 μM DAP agar plate. After 4–5 h of mating, the filters were added to 100 μl of LB, which was vortexed and plated onto LB + 50 μg/ml kanamycin. Putative transposon mutants were picked into 96-well microplates containing LB + 7.5% glycerol + 50 μg/ml kanamycin and grown in a microplate platform shaker at 30°C. After overnight growth, four 96-well mutant plates were rearrayed into a single 384-well storage plate and into a 96-well template plate for transposon insertion site mapping.

To map the gene disrupted and the TagModule associated with the disruption for each of our mutants, we used a previously described two-step arbitrary PCR method (23). Round 1 of the arbitrary PCR used template lysed at 95°C for 15 min and primers MR-1_ARB1, MR-1_ARB3, and pRL27_IE_rev1 (for pRL27 _TagModule_XXXX) or D1 (for pMiniHimar_RB1 _TagModule_XXXX). Round 2 used 1 ul of round 1 PCR product as template and primers ARB2 and U2_revcomp (for pRL27_TagModule_XXXX) or D2_revcomp (for pMiniHimar_TagModule_XXXX). Round 2 PCR products were sequenced using U2_revcomp (for pRL27_TagModule_XXXX) or D2_revcomp (for pMiniHimar_TagModule_XXXX). We used a custom Perl program to identify both the genome insertion location and the TagModule identity for each transposon mutant. A summary of our 7387 mutants is found in Supplementary Table 2.

Pooling and growth of *S. oneidensis* MR-1 mutants

From the set of 7387 transposon mutants, we made a single pool of 1761 uniquely tagged mutants, representing insertions in 1646 unique genes (Supplementary Table S3). To make the pool, we grew selected strains to saturation in LB at 30°C, mixed equal volumes of each mutant culture and froze pool aliquots as 10% glycerol stocks at −80°C. For pooled experiments, we first grew ∼2×10^9 cells of the pool in LB for 3 h at 30°C (∼two doublings) to allow the cells to recover from −80°C. The recovered pools were pelleted, washed twice with 1X phosphate-buffered saline and used to inoculate 48-well microplates for screening in LB, LB + 300 mM NaCl, and MR-1 minimal media (per liter: 0.3 g NaOH, 1.5 g NH$_4$Cl, 0.1 g KCl, 0.6 g NaH$_2$PO$_4$, 0.2 g Na$_2$SO$_4$, 30 mM PIPES, d,l-lactate 30 mM and a trace element mixture, pH 7.0). The screens were done with robotic transfer as described (24). For the LB and LB + 300 mM NaCl experiments, we collected samples at 5, 10, 15 and 20 population doublings. For the minimal media experiment, we collected samples at 4, 8, 12 and 16 population doublings. For all samples, we extracted genomic DNA (Qiagen DNeasy kit) and amplified the uptags and downtags as previously described (18). Array hybridization to TAG4 arrays was as described except that 10 μl of both the uptag and downtag PCR products were used for hybridization (18). For individual mutant confirmations, we grew selected strains in LB, LB + 300 mM NaCl, and minimal media in a 96-well microplate. Doubling times for individual mutants were calculated according the AvgG metric of St. Onge *et al.* (25) using OD600 measurements taken every 15 min in a microplate reader (Tecan GENios).

Analysis *of S. oneidensis* MR-1 tag microarray data

Our analysis was based on the identification of individual strains with reduced relative fitness as inferred by the loss of tag signal during a time-course experiment. A linear regression model implemented in the statistical program R was used for this purpose. For each sample, we first extracted the MEAN probe intensity from each Affymetrix CEL file. Array normalization was performed separately for the uptags and downtags using the normalizeAverage function from the bioconductor package aroma.light (http://www.bioconductor.org/pack ages/2.4/bioc/html/aroma.light.html).
Post-normalization, the five replicate probes for each tag complement were averaged to a single value as described (19). Linear regression of the log$_2$ probe intensity for each tag module as a function of pool doubling times (5, 10, 15 and 20 for LB and LB + 300 mM salt; 4, 8, 12 and 16 for minimal media) was implemented using the linear model function (lm) in R. A zero time point was not used because we found that some strains in the pool had slow −80°C recovery times, which could result in spurious identification of slow growth in these strains (data not shown). We used both the uptag and downtag values in the regression model because they produce similar results as when used separately for calculating strain fitness (Supplementary Figure 5b). Prior to regression, the uptag and downtag measurements were normalized to a mean of zero, as

any differences in the average intensity between uptag and downtag is not biologically meaningful. We used the regression slope to determine which mutants in the pool had a fitness defect. A negative slope indicates that the tag signal decreases over time, and the corresponding mutant has a relative growth defect compared to the pool average. The regression slopes were added to 1 to calculate the final FITNESS values for each mutant (referred to in the main text as 'relative fitness value'). For the calculation of LB fitness (FITNESS_LB), two replicate time-course experiments were used in a single linear regression because we observed that each individual experiment gives very similar results (Supplementary Figure 5c). A single time-course experiment was used in the calculation of strain fitness for the LB + 300 mM NaCl (FITNESS_LB_Salt) and minimal media (FITNESS_Minimal) experiments. To correct for multiple testing, the *P*-values from each regression were adjusted using the false discovery rate ('fdr') option of the R function p.adjust.

A separate regression analysis was used to identify mutants with significantly different fitness in minimal media relative to LB and in LB + 300 mM NaCl relative to LB. For each strain in the pool, we compared regression lines for the two conditions by computing the *F*-statistic. As mentioned above, the mean of the log$_2$ tag values for the four sample time points of all tags were scaled to 0 to control for differences in variance in the tag measurements. A *P*-value for each comparison of regression lines (termed pvalue_LB_and_LB_Salt and pvalue_LB_and_Minimal) was derived from the *F*-distribution. Low *P*-values indicate that the slopes of the regression lines (and hence, relative fitness) are significantly different between the two conditions. These *P*-values were corrected for multiple testing using the same fdr method as described for the FITNESS calculations. We used the following criteria to define strains with slow growth in LB + 300 mM NaCl relative to LB; FITNESS_LB > 0.97, FITNESS_LB_Salt < 0.95; pvalue_FITNESS_LB_Salt < 0.05, pvalue_LB_and_LB_Salt < 0.01. These criteria select for strains with normal fitness in LB, significantly reduced fitness in LB + 300 mM NaCl, and a significantly different fitness value in the two conditions. The 25 mutants that fit these criteria are listed in Supplementary Table 5. To identify strains with reduced fitness in minimal media relative to LB, we used the following criteria: FITNESS_LB > 0.97, FITNESS_Minimal < 0.90; pvalue_FITNESS_Minimal < 0.05, pvalue_LB_and_Minimal < 0.01. The 38 strains with specific growth defects in minimal media are listed in Supplementary Table S4.

Construction of *C. albicans* mutants

Our approach to mutagenize *C. albicans* was based on prior reports of *in vitro* mutagenesis of a *C. albicans* genomic library, followed by fragment excision and homologous recombination to create a final heterozygous mutant (15,16). Our strategy is outlined in Supplementary Figure S5. To increase coverage, we used multiple genomic libraries as targets for the transposon mutagenesis, which were either purchased (Open

Biosystems) or created for this study using DNA isolated from *C. albicans* strain BWP17 (26). pENTR_TagModule or pUC19 (fitted with a polylinker amplified using polylinker_EcoRI_L and polylinker_EcoRI_R) were used as vector backbones, and a variety of 6-bp cutters (XbaI, EcoRV, SpeI or XbaI/SpeI) were used to digest the genomic DNA for the construction of the genomic library. Ligations of digested genomic DNA into either of the vectors were electroporated into Transformax EC100 Electrocompetent *E. coli* (Epicentre Biotechnologies). Each library contained on average 20 000 clones with average size ranging from 2 to 8 kb.

For the mutagenesis, we initially used the Tn7-*UAU1*, kindly provided by A. Mitchell, reported in Davis *et al.* (15) FseI-cut Tn7-*UAU1* was ligated to a Gateway conversion cassette A (Invitrogen) flanked with FseI restriction sites by PCR amplification with primers FSEI-F and FSEI-R. Due to technical difficulties in reliably recovering insertions and sequencing them, we switched to the EZ-Tn5 Transposase system (Epicentre Biotechnologies). The kanamycin resistance gene and the *UAU1* cassette from Tn7-*UAU1* were cloned into EZ-Tn5 pMOD-3 (Epicentre Biotechnologies), and the Gateway conversion cassette C.1 cloned using SnaBI into pMOD-3+Kan+*UAU1*, creating the transposon destination vector, Tn5-*UAU1*-C.1.

Tag transfer with the Tn7-*UAU1*-A or the Tn5-*UAU1*-C.1 transposon destination vector was performed according to manufacturer's guidelines with subpools of pENTR_TagModule_XXXX entry clones ranging in number from 100 to 1000 plasmids per pool. Following the LR clonase reaction, transposons containing TagModules (Tn7-*UAU1*-TM or Tn5-*UAU1*-TM) were electroporated and a minimum of 5000 colonies were recovered and combined to create sublibraries of tagged transposons. Mutagenesis with Tn7-*UAU1*-TM pools was performed according to manufacturer's guidelines for the GPS Mutagenesis System (New England Biolabs) using 20 ng of Tn7-UAU1-*TM* and 80 ng of genomic library. Mutagenesis using pools of PshAI-cut, gel-purified Tn5-*UAU1*-TM was performed according to manufacturer's guidelines using 200 ng each of genomic library and transposon. Mutagenized products were electroporated and recovered on LB + kanamycin (100 µg/ml) + carbenicillin (50 µg/ml) plates. Individual colonies were then picked and plasmid-prepped in 384-well format using Seqprep (Edge Biosystems). Inserts were sequenced using D2_revcomp for Tn7 insertions or U1 for Tn5 insertions.

We generated custom Perl scripts to analyze each sequence to determine (i) the gene disrupted, and (ii) the TagModule associated with the disruption. Each sequence was analyzed with BLAST-N against Assembly 21 of the *C. albicans* sequence. Percentage of gene disrupted was calculated as 1 − (# bp pairs from transposon junction to gene start)/(total gene length). Genes were then sorted to maximize the number of unique TagModules and unique gene insertions. In cases of multiple tag-gene pairs, the pair with the highest percentage gene disrupted was selected. Selected insertions were then plasmid prepped with REAL 96 Plasmid Prep (Qiagen) or

Seqprep 96 (Edge Biosystems). The genomic fragment containing the tagged insertion was excised with the appropriate enzyme and chemically transformed into BWP17 in 96-well format.

Pooled growth and analysis of *C. albicans* mutants

We arrayed all transformants to agar plates and scraped and combined them in SC-arginine + uridine plus 15% glycerol to create the *C. albicans* pool, which we then stored in 50 µl aliquots at −80°C. Twenty-generation pooled growth assays were performed with 1% DMSO or 50 µM clotrimazole (Sigma) in YPD as described (24), minus the 10-generation recovery time. PCR amplification of the uptags and downtags and hybridization to TAG4 microarrays was as described (18). For each array, outliers were masked and removed, and the average of the unmasked replicateswas calculated for each tag. Uptags and downtags were mean-normalized, and low-quality tags (poor up and downtag correlation, or low signal intensity in control set) were removed as described (18).To estimate reproducibility of the *C. albicans* pool, we calculated the correlation of the unnormalized, averaged tag intensities from two random arrays grown in YPD + 1% DMSO (Supplementary Figure S7a). We also calculated the correlation between raw intensity values of the uptags and downtags of a zero-time-point aliquot of the pool (Supplementary Figure S7b). To determine sensitive strains, the experimental array was compared to a matched control set comprised of 13 no-drug arrays (4). We then calculated the $\log_2$ ratio of the control intensity over the treatment intensity and used this as an indicator of sensitivity, where highly sensitive strains have a relatively larger positive $\log_2$ ratio with respect to unaffected or resistant strains.

For individual strain confirmations, mutants were picked from agar and grown to saturation in selective media. Each strain was then diluted to 0.0625 OD_{600} in YPD + 75 µM clotrimazole or YPD + 1% DMSO and grown for ~24 h at 30°C in a Tecan GENios microplate reader. OD_{600} measurements were taken every 15 min, and average doubling time (AvgG) was calculated according to Lee *et al.* (24). Briefly, the time the culture took to reach a five-generation time point from the starting OD_{600} was measured and divided by the number of generations (five), and strain sensitivity was determined by the $\log_2$ ratio of the control AvgG to the experimental AvgG.

RESULTS

A universal Gateway TagModule collection

As the tags, primers, and Affymetrix TAG4 microarrays used with the current *S. cerevisiae* deletion collection have been widely accepted (18,19), we chose to use the same tags and primer sets for the construction of our TagModule collection. However, rather than PCR-amplifying the tags from the yeast deletion mutants for reuse [as up to 20% of its tags contain mutations (27)], we re-synthesized them in an attempt to reduce the number of mutations in the tags. Each TagModule,

cloned into a modified Gateway *att*L entry vector, contains two unique tags (termed uptag and downtag) flanked by a set of common priming sequences (Figure 1a). We used two parallel approaches to synthesize the TagModules (Supplementary Figure S1). The first 2400 TagModules were synthesized from pairs of individual long oligonucleotides, one containing the uptag and the second containing the downtag. An overlapping adaptamer region allowed PCR amplification to link the two long oligonucleotides to create a 175-bp TagModule, which was then individually cloned, recovered and sequence-verified. A second set of putative TagModules were synthesized on the surface of a microarray as 135-mers and were recovered, PCR-amplified and cloned as a pool (17). Sequence analysis of individual clones allowed us to recover 1880 mutation-free TagModules, for a total of 4280 unique and sequence-verified TagModules (Supplementary Table S1).
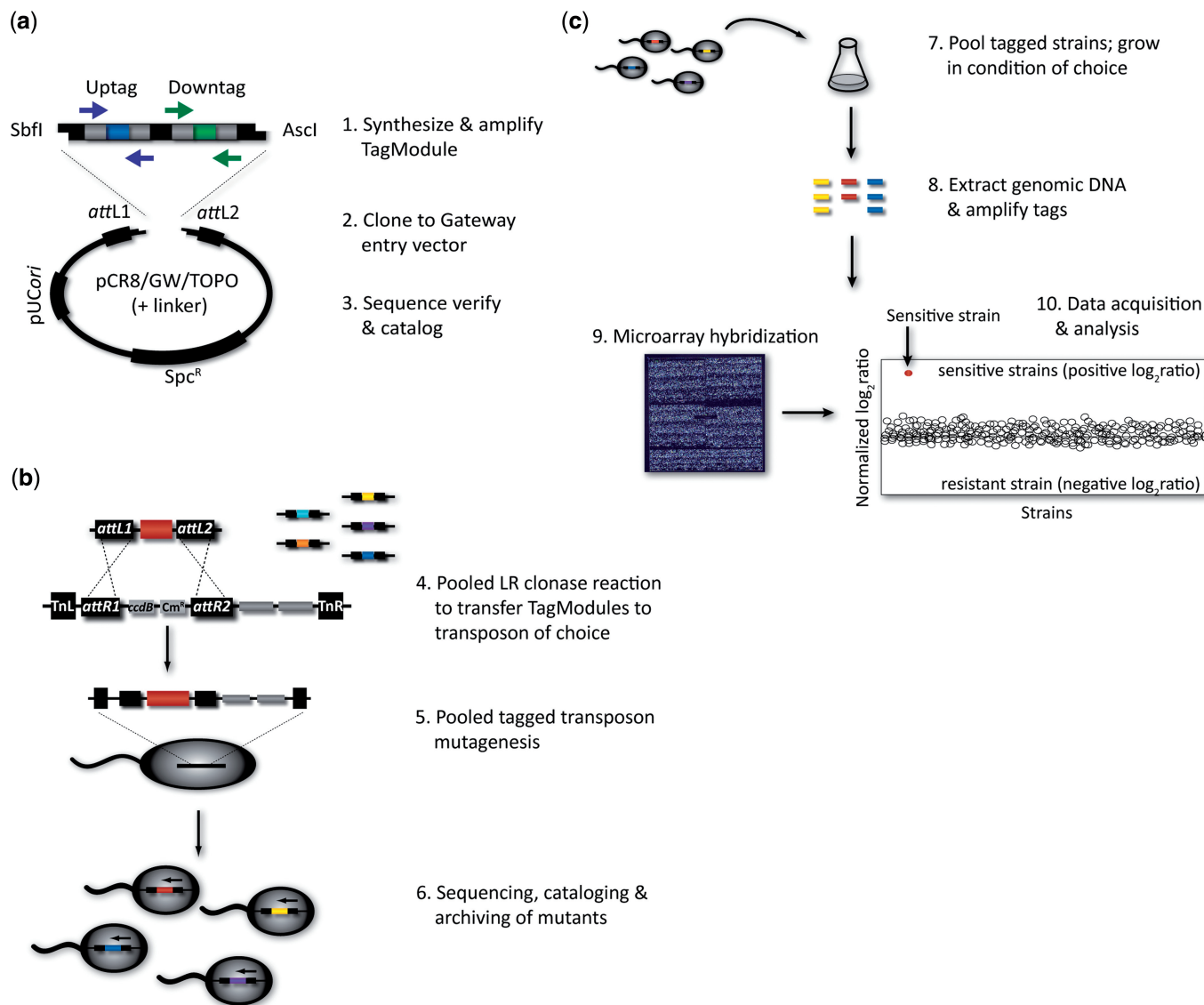

Figure 1. Overview of TagModule design and tagged transposon mutagenesis strategy. (**a**) Each TagModule contains a unique uptag (blue) and downtag (green) flanked by universal priming sites (arrows). TagModules were cloned into pCR8/GW/TOPO, a Gateway-compatible entry vector and sequence verified with a primer on the pCR8/GW/TOPO backbone. Spc^R indicates spectinomycin resistance; pUC*ori* indicates the origin of replication; *att*L1 and *att*L2 are the Gateway recombination sites. (**b**) To use the TagModules for tagged transposon mutagenesis, first the transposon delivery plasmid must be converted into a Gateway compatible destination vector by addition of an *att*R1-*ccdB*-Cm^R-*att*R2 cassette. Addition of LR clonase to a pool of TagModule entry clones (a colored bar represents a pair of tags) and a transposon destination vector induces recombination at the *att* sites, replacing the Gateway conversion cassette with the TagModule. The resulting pool of tagged transposons can then be used to mutagenize the organism of interest. TnL and TnR are the ends of the transposon, and each TagModule is represented by a different color rectangle.(**c**) Tagged mutants can then be pooled and grown competitively in a condition of choice. Strains lacking genes essential for optimal growth in the test condition will grow slower and become underrepresented in the pool; resistant strains grow faster and will be overrepresented. Recovery, amplification and hybridization of the tags to an array will yield information on the abundance of each tag, which can then be converted to a relative fitness value or sensitivity score for each mutant.

We tested the hybridization properties of these TagModules by amplifying the tags from a pool of the tagged Gateway entry vectors and hybridizing them to an Affymetrix TAG4 microarray containing the tag complements (Figure 2a). Given that the correlation of uptag and downtag hybridization intensities is affected by the different common priming sites, separate PCR reactions used to amplify the tags, and differences in tag hybridization efficiency, we observed reasonably correlated unnormalized uptag and downtag intensities ($R = 0.40$, $P < 10^{-16}$). More importantly, the majority of the 8560 individual tags were detected with a robust hybridization signal, verifying the quality of our tag collection. Only 0.4% of the tags did not perform to expectations, with 37 having hybridization signal intensities under 5X of median background intensity. Five of these were a matched uptag/downtag pair, likely indicating very low abundance or absence of the TagModule. The remaining 32 were flagged as a defective tag of a pair, likely the result of an undetected mutation in the common priming site or the tag sequence itself. This new TagModule collection showed a significant improvement in performance and correlation over the tags used in the current *S. cerevisiae* deletion collection. When tags were amplified from a frozen aliquot of a combined heterozygous essential/homozygous deletion pool (5), we observed that 11.3% of tags exhibited hybridization signals below 5X background. For the remaining tags, we observed a correlation of 0.20, $P < 10^{-16}$ between uptag and downtag probe intensities (Supplementary Figure S2). Correlation
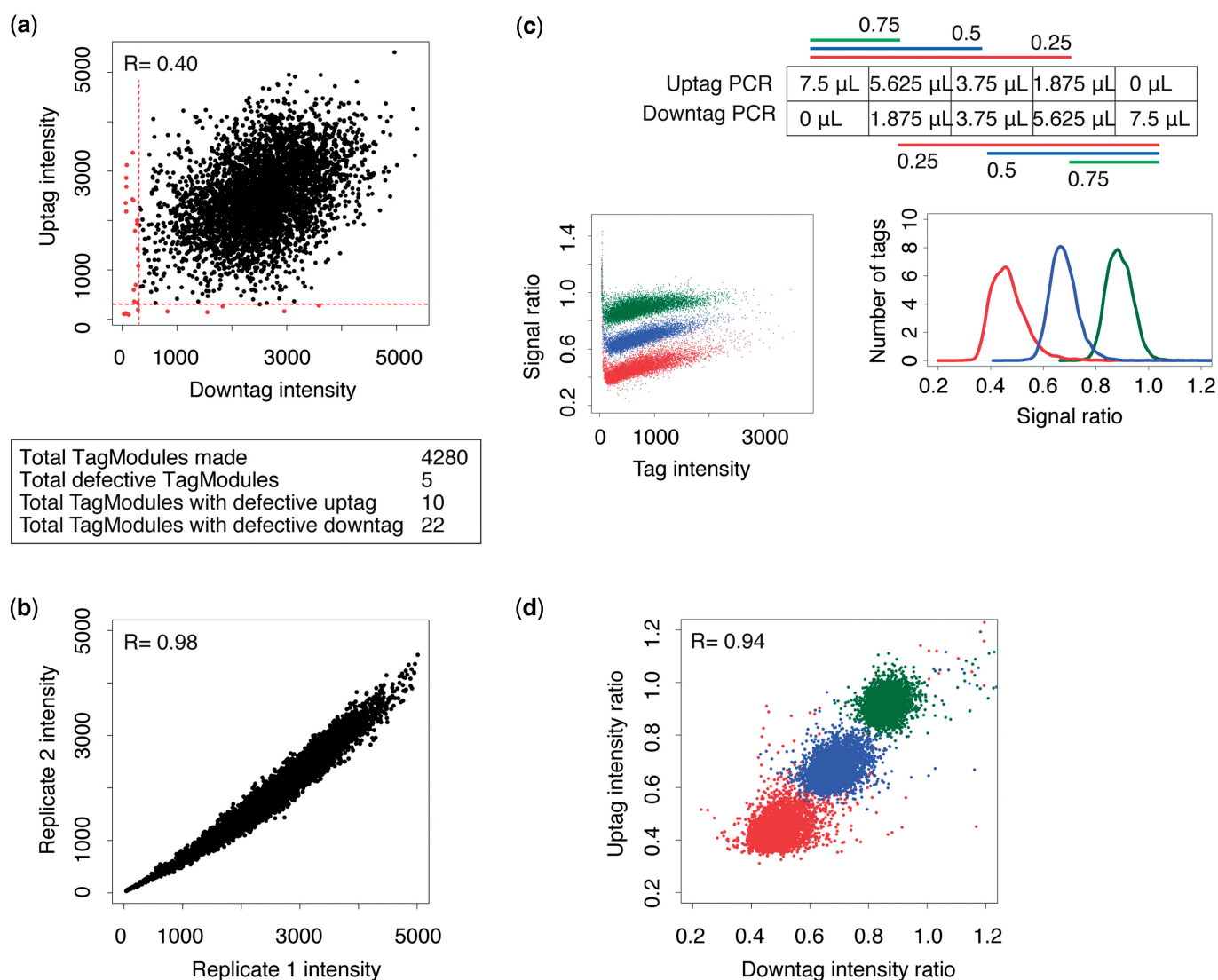

(a)

R= 0.40

Total TagModules made	4280
Total defective TagModules	5
Total TagModules with defective uptag	10
Total TagModules with defective downtag	22

(b)

R= 0.98

(c)

	0.75	0.5	0.25	
Uptag PCR	7.5 µL	5.625 µL	3.75 µL	1.875 µL	0 µL
Downtag PCR	0 µL	1.875 µL	3.75 µL	5.625 µL	7.5 µL
	0.25	0.5	0.75	

(d)

R= 0.94

Figure 2. Validation of TagModule activity. (**a**) TagModules were pooled, and uptags and downtags were amplified and hybridized to an Affymetrix TAG4 array. Tags with raw intensity values below a cutoff of 5-fold above median background (red line) were flagged for exclusion (red points). Pearson's correlation is indicated in the upper left corner, $P < 10^{-16}$. (**b**) Independently amplified TagModule pools are highly reproducible even lacking array or tag normalization. Pearson's correlation is in the upper left. (**c**) Uptag and downtag PCRs were hybridized to arrays in defined ratios, using a total of 7.5 µl of PCR product (top). (Left) Actual tag ratios are plotted corresponding to their expected ratio, and right, density plots of corresponding color highlights the resolution of the tag ratios. (**d**) Comparison of calculated uptag and downtag ratios from (c) with colors corresponding to their expected ratio; Pearson's correlation is indicated in the upper left, $P < 10^{-16}$.

between independently amplified, unnormalized replicates of our TagModule collection was also high (Figure 2b, $R = 0.98$, $P < 10^{-16}$), demonstrating that technical replicate hybridizations are not necessary for the TagModule collection.

Our TagModule synthesis design and cloning strategy attempted to make use of all of the ~16-K tags available on the TAG4 array, including many not currently used in the yeast deletion system. We considered the possibility that the 8499 tags we failed to clone could have 'contaminated' our collection. We also wished to validate that the TagModules did not show significant cross-reactivity with other spots on the array. We examined the signal intensities of all unused tags following hybridization of the tags amplified from the tag collection. Of the spots corresponding to 8499 unused tags, 8257 (97.2%) had intensities below 5X background. Of the 242 tags with above 5X background intensities, we found that 187 corresponded to previously repaired tags (19). The remaining 55 had no significant sequence similarity to tags in use, indicating that the false positive rate of our TagModule collection is ~0.6%. We believe the hybridization signal of these 55 unexpected tags is due to cross-hybridization or contamination either in the PCR reaction or in the TagModule collection itself.

Finally, we confirmed that the TagModules were able to accurately measure differences in tag abundance. We hybridized fixed ratios of tags to an array, varying the amounts of uptag and downtags while keeping the total amount of PCR product hybridized constant (Figure 2c). We found that the three ratios were well resolved, although, as the tag intensity increased, the ability of the tags to accurately reflect their expected ratios decreased, as previously observed (19). Moreover, the observed ratios calculated separately for the uptags and downtags were highly correlated (Figure 2d, $R = 0.94$, $P < 10^{-16}$). This indicates that both tags in each TagModule reflect tag abundance similarly and would permit subsequent experiments to use a single tag, thereby doubling the effective number of TagModules. We note that a typical *S. cerevisiae* hybridization uses 30 μl of uptag and 30 μl of downtag PCR, and that we used a quarter this amount. We found that this level maximized the ability of the tags to reflect actual tag abundance at the three levels, as increasing amounts of PCR tend to saturate the array and underestimate differences in tag concentration (data not shown). This adjustment in hybridization conditions is significant as this underestimation can mask subtle changes in tag abundance due to modest growth defects. In general, while we have shown that our TagModules have robust hybridization and can accurately measure tag abundance, we recommend customizing the hybridization conditions based on the size of the pool or the number of TagModules used.

Creation of tagged mutants via transposon mutagenesis

Our Gateway-compatible TagModule collection was designed for the rapid transfer of tags to any genetic tool that can be been modified for recombinational cloning. Here, we use the collection to create a set of tagged transposons suitable for high-throughput functional genomic studies (Figure 1b). Briefly, conversion of a transposon vector to include the Gateway-compatible *att*R recombination sites allows TagModule transfer to the transposon upon addition of LR clonase. We have observed that the LR clonase reaction is sufficiently robust such that TagModules can be pooled for transfer [(28); data not shown], resulting in a pool of uniquely tagged transposons. These tagged transposons can then be used for mutagenesis of the desired organism *in vivo* or *in vitro*. Individual mutants are then recovered and sequenced to identify both the site of the transposon insertion and the TagModule identity. Uniquely tagged mutants can be combined and screened in pooled growth assays using the TAG4 arrays or sequencing technologies (Figure 1c). To test the *in vivo* activity of the TagModules, and to validate the pooled, tagged transposon mutagenesis approach as a method to create tagged disruption collections, we mutagenized the Gram-negative haploid bacterium *S. oneidensis* MR-1 and the diploid yeast human pathogen *C. albicans*. An overview of our flexible tagged transposon mutagenesis strategy and its application to *S. oneidensis* MR-1 and *C. albicans* is illustrated in Supplementary Figure S3.

Verification of TagModule transposon mutagenesis of *S. oneidensis* MR-1

We extended our flexible TagModule technique to prokaryotes by generating a library of tagged transposon insertion mutants in the metal-reducing bacterium *S. oneidensis* MR-1. We Gateway-converted two transposons previously described for use with this organism: a broad-range mini-Tn5 (20,22) and a *mariner*-derived *Himar1* transposon (21). TagModules were transferred to the transposon-containing suicide plasmids via recombinational cloning and transformed into an *E. coli* conjugation donor strain. Conjugation with recipient *S. oneidensis* MR-1 resulted in mutagenized, tagged *S. oneidensis* MR-1 mutants. TagModules and genomic sequence flanking the transposon insertion were amplified using a two-step arbitrary PCR protocol (23). PCR products were sequenced to identify both the gene disrupted and the TagModule associated with the disruption. This mutagenesis scheme is detailed in Supplementary Figure S4.

From a library of 7387 mutants (Supplementary Table S2), we constructed a single pilot pool of 1761 uniquely tagged strains, representing insertions in 1646 different genes (of 4467 in the genome) (Supplementary Table S3). To confirm TagModule quality *in vivo*, relative abundance of all strains in the pool, and the success of our strain sample tracking, we hybridized uptags and downtags amplified from the pool. Of the 3522 total tags, only eleven fell below 16X background intensity (Supplementary Figure S5a). Furthermore, when we examined the probe signals of unused TagModules, we only detected two of these unused tags above 16X background, indicating that there is little cross-hybridization between tags, and that our sample

tracking, from initial individual strain cultivation and sequencing to pooling, is robust.

To demonstrate the utility of our technique to quantitatively measure phenotypes and uncover new biology in bacteria, we profiled the pool of 1761 mutants under three conditions: rich media (LB), rich media with a NaCl stress (LB + salt), and minimal media. For each of these three conditions, we determined a relative fitness value for all 1761 strains in the pool by analyzing the change in tag abundance during growth (see 'Supplementary Methods'). The relative fitness value is based on tracking each tag's abundance over a

time-course experiment and fitting a line to the points. The slope of the regression (adjusted so that no strain inhibition equals a relative fitness value of 1, and values <1 correspond to a relative fitness defect) reflects the rate at which a strain's abundance in the pool decreases as a result of the experimental treatment, which in turn reflects the strain's relative fitness in the condition. By comparing each strain's relative fitness values in LB to that in minimal media, we identified 38 mutant strains with a relative fitness defect specific to minimal media (Figure 3a; Supplementary Table S4; see 'Supplementary Methods' for criteria).
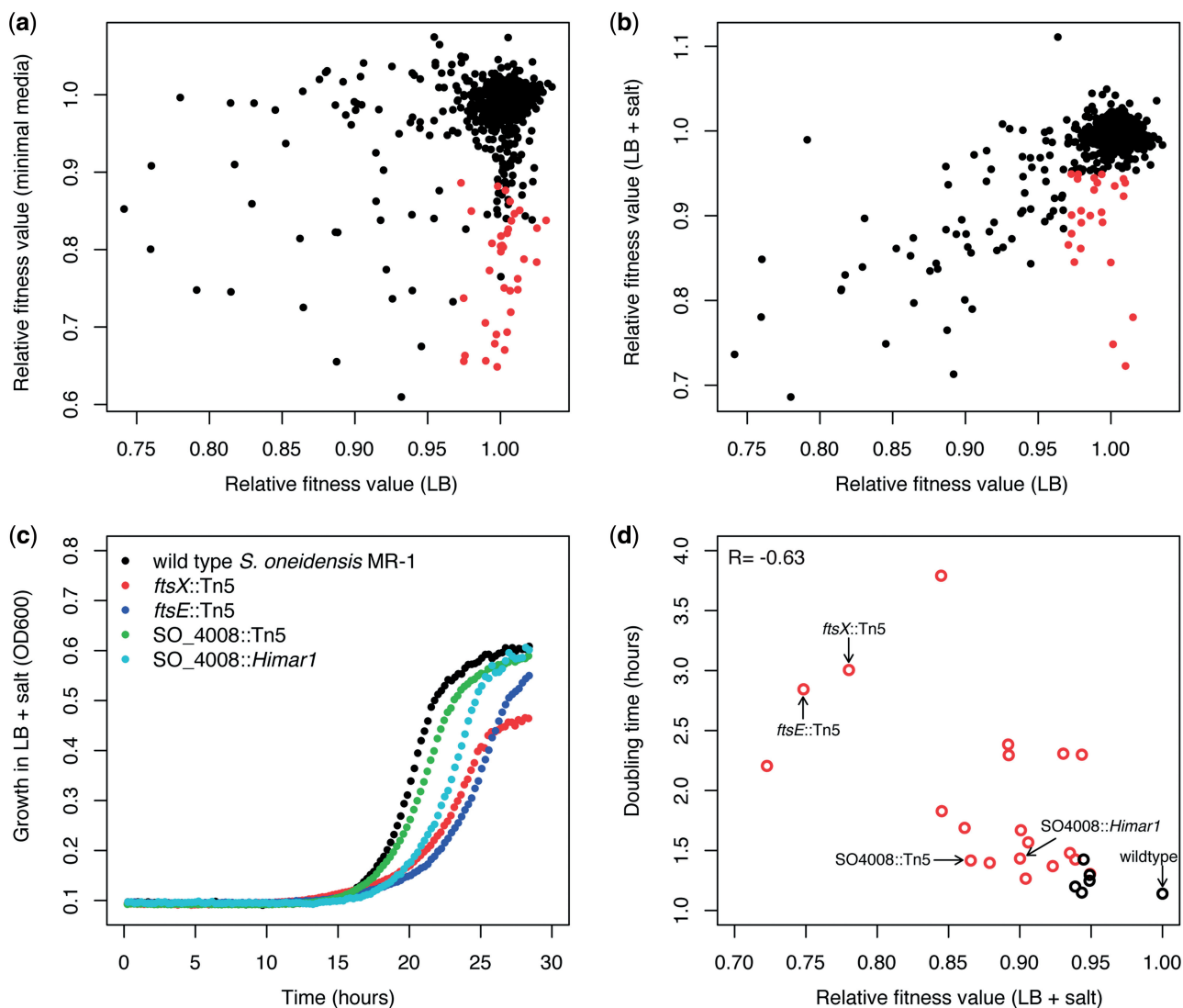

Figure 3. Validation of TagModule-based transposon mutagenesis in *S. oneidensis* MR-1. The relative fitness values of 1761 tagged transposon mutants in rich media (LB) was plotted versus their values in (**a**) minimal media or (**b**) LB + 300 mM NaCl as determined by pooled competitive growth followed by hybridization to a TAG4 array. A value of 1 corresponds to no growth defect relative to the pool average, whereas a value <1 indicates a growth defect in that condition. The red points in (a) correspond to 38 mutant strains with a specific growth defect in minimal media relative to rich media; the red points in (b) correspond to 25 mutant strains with a growth defect in LB + 300 mM NaCl (see 'Materials and Methods' section for selection criteria). (**c**) Representative single strain growth curves in LB + 300 mM NaCl for mutants with growth defects from (b). Wild type *S. oneidensis* MR-1 is included as a control (black). (**d**) Single strain doubling times for all 25 defective mutants from the red points in (b) were compared with the relative fitness values obtained from the pooled growth assay. Pearson correlation is in the upper left (*P* = 0.0006). Mutant stains with individual growth curves in (c) are indicated. The points represent the average of at least four independent growth curve assays for each strain. Red points represent mutant strains with significantly different doubling times compared to the wild-type *S. oneidensis* MR-1.

Two lines of evidence suggest that these findings are biologically meaningful. First, 22 of the 36 mutated genes (two independent insertions of *crp* and SO3180 are included in this set) are predicted auxotrophs, including *argD*, *aroF*, *thiI*, *hisD*, *ilvI*, *panE* and *lysC*. Second, 11 of the 36 unique genes are contained within the same predicted operon, including one putatively involved in cell wall synthesis (SO3179, SO3180, SO3182, SO3183, SO3185, SO3188 and SO3190). In addition, we identified mutations in SO1518 and SO1521 as having fitness defects in minimal media. These genes were recently shown to encode d-lactate dehydrogenase (SO1521) and a subunit of l-lactate dehydrogenase (SO1518) (29). These results are consistent because we used d,l-lactate as the sole carbon source in our minimal media. Furthermore, the phenotypes of our transposon mutants, in particular the growth defect in d,l-lactate media, are consistent with those observed for targeted gene deletions in the same genes (29). Lastly, four of the genes are hypothetical, suggesting that additional core biological functions required for growth in minimal media have yet to be discovered.

We also compared the fitness profiles of LB and LB + 300 mM NaCl, identifying 25 mutants with a specific growth defect in high salinity media (Figure 3b; Supplementary Table S5). Generally, the phenotypes of this class of mutants are less severe than those of auxotrophs grown in minimal media and should reflect more subtle, quantitative growth differences. For this reason, and to test the ability of the flexible TagModule system in detecting subtle differences in growth, we isolated these 25 defective mutants and determined their individual growth rates (representative curves are shown in Figure 3c). We verified that 20/25 had statistically different doubling times compared to wildtype *S. oneidensis* MR-1. Additional replicate growth curve experiments, or more quantitative measures of single strain fitness such as those based on flow cytometry (30), may be necessary to verify the phenotypes of the other five. Lastly, we observed a significant correlation between our microarray-based relative fitness values and individual strain doubling times, demonstrating that the results of our pooled assays reflect quantitative measures of strain fitness (Figure 3d).

The composition of the 25 LB + salt mutants suggests the importance of fatty acid biosynthesis (SO1599, SO1602, SO1742, SO1743 and SO1744), transporters (SO4693, *ftsX*, *ftsE*), and uncharacterized proteins (including two independent mutants in SO4008) in the response and adaptation of *S. oneidensis* MR-1 to high salinity. Interestingly, the *ftsX* and *ftsE* mutants in *S. oneidensis* MR-1 appear to have the opposite phenotype in *E. coli*, where high salt is required for these mutants to grow (31). These findings underscore the limitations of annotations based solely on homology and the value of direct experimental evidence gained in diverged bacteria.

Verification of TagModule transposon mutagenesis of *C. albicans*

Following the approach of previous *C. albicans* transposon mutagenesis studies (15,16), we mutagenized

C. albicans genomic libraries *in vitro* with a pool of tagged transposons created by transferring the pooled TagModule library to a modified EZ-Tn5 (Epicentre). This Tn5 was modified to contain a selectable *UAU1* marker (32), kanamycin resistance, and *att*R recombination sites. Recovered insertions were sequenced from the uptag priming site, crossing the TagModule and the transposon junction, allowing identification of the tag associated with a gene disruption. We sorted the results to maximize the number of unique genes disrupted with the maximum number of unique tags. In the case of multiple gene insertions, we chose those with the highest percentage of gene disrupted, as defined as the percentage of the gene region downstream of the transposon junction. To transfer the transposon insertion to *C. albicans*, we excised the genomic fragment containing the tagged transposon, and transformed them via homologous recombination into the *C. albicans* strain BWP17 (26). We created a preliminary pilot set of 1290 tagged heterozygous mutants representing 1246 unique genes of 6197 annotated ORFs (Supplementary Figure 6, Supplementary Table 6).

To validate the performance of these tagged mutants in a pool, we pooled roughly equivalent amounts of each of the 1290 mutants and performed a chemogenomic fitness screen with clotrimazole (Figure 4a), a well-characterized antifungal known to target the sterol 14α-demethylase (*ERG11*) in the ergosterol biosynthesis pathway. After 20 generations of growth in the presence or absence of 50 μM clotrimazole, we amplified the tags from the pool and hybridized them to a TAG4 array. Following array normalization for uptags and downtags and averaging of tag replicates and up and downtags (18), we calculated the $\log_2$ ratios of the tag intensities of the strains grown in a 1% DMSO control set versus those grown in clotrimazole. The magnitude of the $\log_2$ ratio is an indication of strain sensitivity, with increasing magnitude indicating an increasing fitness defect. Note that the calculation of fitness defect for *C. albicans* mutants differs from the relative fitness values calculated for *S. oneidensis* MR-1 mutants in that the intensity values only at the assay endpoint are used. We found that while the relative fitness values based on multiple hybridizations as part of a time-course experiment incorporate more data, the results are similar to those obtained with an endpoint assay (data not shown). The type of analysis comparing experimental $\log_2$ ratios to those from a control set is more typical when using the pooled growth assay to identify drug targets (4), as the magnitude of inhibition relative to that of other strains is more relevant than incremental changes in tag abundance over time.

We were able to confirm previous reports (11) of sensitivity to deletions of *ERG11* and *NCP1* (NADP-cytochrome P450 reductase in the ergosterol biosynthesis pathway). To corroborate the results from the pooled growth assay and to confirm that the TagModules quantitatively determines fitness defect *in vivo*, we selected the 37 most sensitive strains and measured their growth rate individually with and without clotrimazole. Representative growth curves (Figure 4b) for three strains heterozygous for genes in the ergosterol biosynthesis pathway, *NCP1*, *ERG11* and *ERG2* (C-8 sterol)
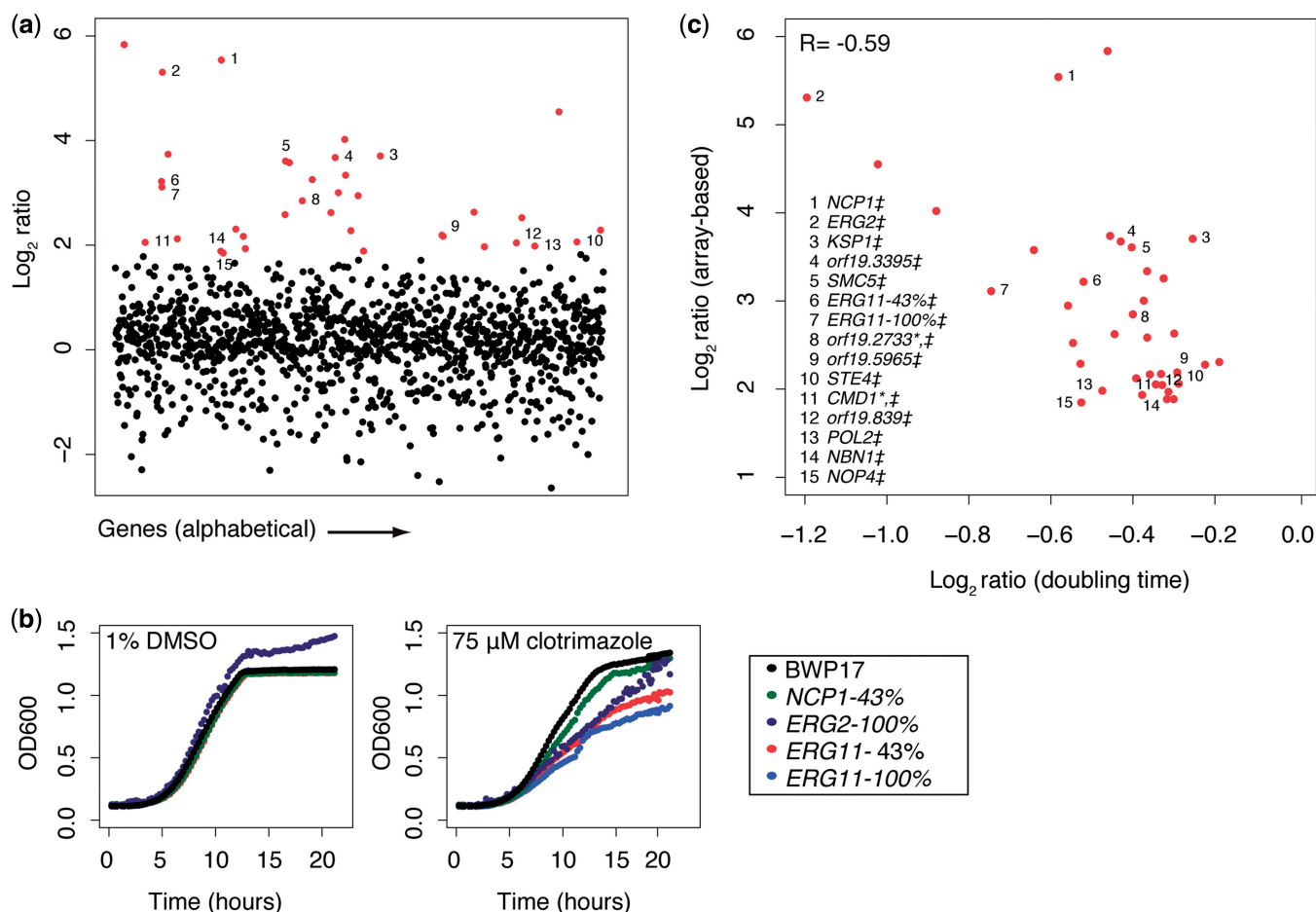
Figure 4. Validation of TagModule-based transposon mutagenesis in *C. albicans*. (**a**) A total of 1290 heterozygous disruption strains were grown for 20 generations in $50\,\mu M$ clotrimazole or 1% DMSO. Increasing $\log_2$ ratio (control versus experimental intensities) reflects increasing sensitivity. The 37 most sensitive strains (red points) were selected for confirmation; the 15 with an enriched GO annotation are marked. (**b**) Representative single strain growth curves for sensitive mutants from (a). Wild-type *C. albicans* (BWP17) is included as a control (black). For ERG11, two different disruption events were screened, a 43% and a 100% gene disruption. (**c**) Pearson's correlation (upper left) of $\log_2$ ratios obtained from TAG4 array hybridization and $\log_2$ ratios of growth rate (control versus experimental) obtained from (b) for sensitive strains from (a). Wild-type $\log_2$(growth rate) = 0. Genes mapping to GO function phosphatidylinositol metabolic and biosynthetic process are marked *; genes mapping to GO function response to stimulus are marked ‡.

confirmed sensitivity. We note one case in which a full (100%, measured as distance to the start codon) and a partial (43%) disruption of *ERG11* show a difference in growth, indicating the potential of screening alternative disruption events to determine structure-function relationships within individual loci.

Correlation between the $\log_2$ ratios of the individual growth rates (in no drug and drug) and the array-based fitness screen for the 37 strains was significant (Figure 4c, $R = -0.59$, $P < 0.0001$). Having confirmed strain sensitivity, we used the Gene Ontology Term Finder from the Candida Genome Database (http://www.candidagenome .org/cgi-bin/GO/goTermFinder) to search for functional groupings among these 37 most sensitive genes. We found significant enrichment for phosphatidylinositol metabolic process (GO:0046488) and phosphatidylinositol biosynthetic process (GO:0006661) (both 5.6%, $n = 2$) when compared to the genome (0.1%, hypergeometric $P < 0.028$); genes are marked in Figure 4c. Both genes of this category, *CMD1* and the uncharacterized *orf19.2733*,

have characterized *S. cerevisiae* homologs (Sc-*CMD1* and Sc-*VPS30*) that function in stress response (33), autophagy (34) or vacuolar sorting (35). *Saccharomyces cerevisiae* haploinsufficiency assays with functionally related azoles (4) show marked sensitivity of Sc-*CMD1* (sensitive to itraconazole, miconazole and fluconazole) and Sc-*VPS30* mutants (sensitive to clotrimazole, miconazole and itraconazole), highlighting our assay's ability to find biologically consistent results. We also observed significant enrichment in the GO category response to stimulus (GO:0050896, 38.9%, $n = 14$) when compared to the genome (14.2%, hypergeometric $p < 0.033$); these genes are marked in Figure 4c. As clotrimazole's (and the azole drug class) known target is the ergosterol biosynthesis pathway, we hypothesize that these may be off-target effects part of a larger *C. albicans* multidrug or stress response, as this group includes a putative drug transporter, *orf19.3395* (36), and mutations in several of these genes (e.g. *KSP1*, *orf19.839*, *NOP4* and *CMD1*) were found to confer hypersensitivity in prior

C. albicans drug screens (11). In general, we confirm that *C. albicans* mutant strains that are highly sensitive in the fitness screen display drug sensitivity in individual tests and represent biologically meaningful results in a human pathogen. Taken together with results in *S. oneidensis* MR-1, we have validated our flexible TagModule mutagenesis strategy for the high-throughput functional genomic analysis of both eukaryotes and prokaryotes.

DISCUSSION

Genome-wide studies have proved useful in gaining insight into the function of genes and pathways for a handful of model microorganisms, most notably the worm *Caenorhabditis elegans*, *S. cerevisiae* and *E. coli*. However, the results of these systems-level studies are of limited use in interpreting the wealth of genome sequence information for non-model microorganisms. Given the power of genetics in elucidating gene function, the need for rapid and cost-effective approaches to generating large mutant collections in non-model microorganisms has increased. In addition, the screening of mutant collections in multiple conditions, often essential to identifying the function of a gene or pathway, requires multiplexing to keep such studies cost-effective. To this end, we have created a publicly available, Gateway-compatible TagModule resource that can be used to generate tagged mutant collections. Because the 4280-member TagModule collection was sequence-verified, the tags exhibit strong hybridization signal and low cross-hybridization, thereby ensuring that the system is highly quantitative in measuring changes in tag abundance in complex mixtures using Affymetrix TAG4 microarrays, as shown in this study, or with next-generation sequencing technology (Bar-seq) (27).

We observed robust and well-correlated performance of both tags in a module (Figure 2d, Supplementary Figures S5c and S7b). This demonstrates that one tag of a TagModule is sufficient to make an accurate measure of strain abundance. This is valuable because 8560 tags, rather than 4280, are then available for use. In practice, this would require that two separate pools be created, with one containing unique uptags and the other containing unique downtags. For pooled growth assays, each pool would be screened and tags amplified separately prior to combining the amplicons for microarray hybridization. This would be useful for tagging genomes with >4280 genes or for including multiple transposon insertion strains per gene (thus increasing the robustness of the pooled growth assay). In practice, we have already seen that this strategy is effective for mutant tagging in *C. albicans*, which has 6197 predicted genes (data not shown).

The TagModule collection has many potential applications. Most significantly, the TagModules are an *in vitro* system and thus can be applied to virtually any microorganism with a tractable genetic system, even with no prior knowledge of genome sequence. Transferred into an appropriate destination vector, the TagModules can also be applied to any application requiring sample multiplexing, such as tagging existing genome-wide mutant collections, creating tagged targeted deletions, strain or sample tracking, or as we have described here, a system to create tagged transposon insertion collections.

The TagModule method as described here for *S. oneidensis* MR-1 is similar in concept to previously described high-throughput methodologies for transposon monitoring in bacteria [e.g. TraSH (37), and more recently, HITS (38), TraDIS (39) and Tn-seq (7)]. The latter methods are powerful next-generation sequencing tools for the parallel genetic analysis of very large pools of bacterial mutants. While the TagModule method has the disadvantage of requiring more upfront work to make the collection and is less comprehensive across the genome (although we have currently mutated over 90% of the nonessential gene complements in *S. oneidensis* MR-1 and 78% of *C. albicans* open reading frames; data not shown), there are a number of advantages to the TagModule approach.

First, the TagModules can be used to barcode a variety of expression systems or mutant collections, for example, overexpression strains (40), untagged transposon collections (15) or targeted deletions. Second, the TagModules can be used in systems in which TraDIS-like techniques have not been demonstrated, such as *C. albicans*. Third, we are able to achieve a more quantitative measure of strain fitness by assaying fewer mutants. This in turn would permit sample multiplexing to a level currently not permitted with the 'deep' sequencing of hundreds of thousands of insertions required for the HITS, TraDIS, or Tn-seq methods. This makes multiplexed Bar-seq (27) an attractive option, given that profiling mutants in hundreds or thousands of conditions may be necessary to uncover phenotypes for most genes in a genome (4). Finally, our TagModule approach combines the beneficial features of archived, large-scale mutant collections, such as the *E. coli* KEIO collection and numerous clonal transposon mutant collections (23,41), with the ability to assay thousands of strains in parallel. This is especially important in instances where strain pooling is not advantageous (e.g. when measuring the abundance of a metabolic end-product) or targeted mutagenesis methods are not available for follow-up study. Finally, while we have used Sanger sequencing to identify mutants from our TagModule-based mutagenesis, this step can be bypassed by using a multiplexed high-throughput sequencing approach (42).

Although model organisms have been, and continue to be, valuable in understanding basic biological principles, models are inherently limited in their ability to represent related organisms of interest. Ultimately, studying non-model organisms will be more fruitful in identifying novel therapeutics for a pathogen or investigating pathways with potential for metabolic engineering in bacteria with industrial significance. In summary, we have described and validated the construction of a multi-purpose TagModule collection, which we believe will not only be of general use for tagging and multiplexing due to its Gateway-compatibility, but also will make high-throughput reverse genetics accessible to non-model microorganisms that are of medical, industrial or environmental importance.

REFERENCES

1. Hensel,M., Shea,J.E., Gleeson,C., Jones,M.D., Dalton,E. and Holden,D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science*, **269**, 400–403.
2. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387–391.
3. Shoemaker,D.D., Lashkari,D.A., Morris,D., Mittmann,M. and Davis,R.W. (1996) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet.*, **14**, 450–456.
4. Hillenmeyer,M.E., Fung,E., Wildenhain,J., Pierce,S.E., Hoon,S., Lee,W., Proctor,M., St Onge,R.P., Tyers,M., Koller,D. *et al.* (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**, 362–365.
5. Hoon,S., Smith,A.M., Wallace,I.M., Suresh,S., Miranda,M., Fung,E., Proctor,M., Shokat,K.M., Zhang,C., Davis,R.W. *et al.* (2008) An integrated platform of genomic assays reveals small-molecule bioactivities. *Nat. Chem. Biol.*, **4**, 498–506.
6. Groh,J.L., Luo,Q., Ballard,J.D. and Krumholz,L.R. (2005) A method adapting microarray technology for signature-tagged mutagenesis of Desulfovibrio desulfuricans G20 and Shewanella oneidensis MR-1 in anaerobic sediment survival experiments. *Appl. Environ. Microbiol.*, **71**, 7064–7074.
7. van Opijnen,T., Bodi,K.L. and Camilli,A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods*, **6**, 767–772.
8. Wisplinghoff,H., Bischoff,T., Tallent,S.M., Seifert,H., Wenzel,R.P. and Edmond,M.B. (2004) Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.*, **39**, 309–317.
9. Akins,R.A. (2005) An update on antifungal targets and mechanisms of resistance in Candida albicans. *Med. Mycol*, **43**, 285–318.
10. Heidelberg,J.F., Paulsen,I.T., Nelson,K.E., Gaidos,E.J., Nelson,W.C., Read,T.D., Eisen,J.A., Seshadri,R., Ward,N., Methe,B. *et al.* (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. *Nat. Biotechnol.*, **20**, 1118–1123.
11. Xu,D., Jiang,B., Ketela,T., Lemieux,S., Veillette,K., Martel,N., Davison,J., Sillaots,S., Trosok,S., Bachewich,C. *et al.* (2007) Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in Candida albicans. *PLoS Pathog.*, **3**, e92.
12. Rodriguez-Suarez,R., Xu,D., Veillette,K., Davison,J., Sillaots,S., Kauffman,S., Hu,W., Bowman,J., Martel,N., Trosok,S. *et al.* (2007) Mechanism-of-action determination of GMP synthase inhibitors and target validation in Candida albicans and Aspergillus fumigatus. *Chem. Biol.*, **14**, 1163–1175.
13. Jiang,B., Xu,D., Allocco,J., Parish,C., Davison,J., Veillette,K., Sillaots,S., Hu,W., Rodriguez-Suarez,R., Trosok,S. *et al.* (2008) PAP inhibitor with in vivo efficacy identified by Candida albicans genetic profiling of natural products. *Chem. Biol.*, **15**, 363–374.
14. Xu,D., Sillaots,S., Davison,J., Hu,W., Jiang,B., Kauffman,S., Martel,N., Ocampo,P., Oh,C., Trosok,S. *et al.* (2009) Chemical genetic profiling and characterization of small-molecule compounds that affect the biosynthesis of unsaturated fatty acids in Candida albicans. *J. Biol. Chem.*, **284**, 19754–19764.
15. Davis,D.A., Bruno,V.M., Loza,L., Filler,S.G. and Mitchell,A.P. (2002) Candida albicans Mds3p, a conserved regulator of pH responses and virulence identified through insertional mutagenesis. *Genetics*, **162**, 1573–1581.
16. Uhl,M.A., Biery,M., Craig,N. and Johnson,A.D. (2003) Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C.albicans*. *EMBO J.*, **22**, 2668–2678.
17. Cleary,M.A., Kilian,K., Wang,Y., Bradshaw,J., Cavet,G., Ge,W., Kulkarni,A., Paddison,P.J., Chang,K., Sheth,N. *et al.* (2004) Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. *Nat. Methods*, **1**, 241–248.
18. Pierce,S.E., Davis,R.W., Nislow,C. and Giaever,G. (2007) Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. *Nat. Protoc.*, **2**, 2958–2974.
19. Pierce,S.E., Fung,E.L., Jaramillo,D.F., Chu,A.M., Davis,R.W., Nislow,C. and Giaever,G. (2006) A unique and universal molecular barcode array. *Nat. Methods*, **3**, 601–603.
20. Larsen,R.A., Wilson,M.M., Guss,A.M. and Metcalf,W.W. (2002) Genetic analysis of pigment biosynthesis in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. *Arch. Microbiol.*, **178**, 193–201.
21. Bouhenni,R., Gehrke,A. and Saffarini,D. (2005) Identification of genes involved in cytochrome c biogenesis in Shewanella oneidensis, using a modified mariner transposon. *Appl. Environ. Microbiol.*, **71**, 4935–4937.
22. Thormann,K.M., Saville,R.M., Shukla,S., Pelletier,D.A. and Spormann,A.M. (2004) Initial Phases of biofilm formation in Shewanella oneidensis MR-1. *J. Bacteriol.*, **186**, 8096–8104.
23. Jacobs,M.A., Alwood,A., Thaipisuttikul,I., Spencer,D., Haugen,E., Ernst,S., Will,O., Kaul,R., Raymond,C., Levy,R. *et al.* (2003) Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc. Natl Acad. Sci. USA*, **100**, 14339–14344.
24. Lee,W., St Onge,R.P., Proctor,M., Flaherty,P., Jordan,M.I., Arkin,A.P., Davis,R.W., Nislow,C. and Giaever,G. (2005) Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genet.*, **1**, e24.

25. St Onge,R.P., Mani,R., Oh,J., Proctor,M., Fung,E., Davis,R.W., Nislow,C., Roth,F.P. and Giaever,G. (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.*, **39**, 199–206.

26. Wilson,R.B., Davis,D. and Mitchell,A.P. (1999) Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J. Bacteriol.*, **181**, 1868–1874.

27. Smith,A.M., Heisler,L.E., Mellor,J., Kaper,F., Thompson,M.J., Chee,M., Roth,F.P., Giaever,G. and Nislow,C. (2009) Quantitative phenotyping via deep barcode sequencing. *Genome Res.*, **19**, 1836–1842.

28. Gillette,W.K., Esposito,D., Frank,P.H., Zhou,M., Yu,L.R., Jozwik,C., Zhang,X., McGowan,B., Jacobowitz,D.M., Pollard,H.B. *et al.* (2005) Pooled ORF expression technology (POET): using proteomics to screen pools of open reading frames for protein expression. *Mol. Cell Proteomics*, **4**, 1647–1652.

29. Pinchuk,G.E., Rodionov,D.A., Yang,C., Li,X., Osterman,A.L., Dervyn,E., Geydebrekht,O.V., Reed,S.B., Romine,M.F., Collart,F.R. *et al.* (2009) Genomic reconstruction of Shewanella oneidensis MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc. Natl Acad. Sci. USA*, **106**, 2874–2879.

30. Breslow,D.K., Cameron,D.M., Collins,S.R., Schuldiner,M., Stewart-Ornstein,J., Newman,H.W., Braun,S., Madhani,H.D., Krogan,N.J. and Weissman,J.S. (2008) A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods*, **5**, 711–718.

31. de Leeuw,E., Graham,B., Phillips,G.J., ten Hagen-Jongman,C.M., Oudega,B. and Luirink,J. (1999) Molecular characterization of Escherichia coli FtsE and FtsX. *Mol Microbiol*, **31**, 983–993.

32. Enloe,B., Diamond,A. and Mitchell,A.P. (2000) A single-transformation gene function test in diploid Candida albicans. *J. Bacteriol.*, **182**, 5730–5736.

33. Cyert,M.S. (2001) Genetic analysis of calmodulin and its targets in Saccharomyces cerevisiae. *Annu. Rev. Genet.*, **35**, 647–672.

34. Kihara,A., Noda,T., Ishihara,N. and Ohsumi,Y. (2001) Two distinct Vps34 phosphatidylinositol 3-kinase complexes function in autophagy and carboxypeptidase Y sorting in Saccharomyces cerevisiae. *J. Cell. Biol.*, **152**, 519–530.

35. Seaman,M.N., Marcusson,E.G., Cereghino,J.L. and Emr,S.D. (1997) Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J. Cell. Biol.*, **137**, 79–92.

36. Gaur,M., Puri,N., Manoharlal,R., Rai,V., Mukhopadhayay,G., Choudhury,D. and Prasad,R. (2008) MFS transportome of the human pathogenic yeast Candida albicans. *BMC Genomics*, **9**, 579.

37. Sassetti,C.M., Boyd,D.H. and Rubin,E.J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl Acad. Sci. USA*, **98**, 12712–12717.

38. Gawronski,J.D., Wong,S.M., Giannoukos,G., Ward,D.V. and Akerley,B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl Acad. Sci. USA*, **106**, 16422–16427.

39. Langridge,G.C., Phan,M., Turner,D.J., Perkins,T.T., Parts,L., Haase,J., Charles,I., Maskell,D.J., Peters,S.E., Dougan,G. *et al.* (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res.*, **19**, 2308–2316.

40. Ho,C.H., Magtanong,L., Barker,S.L., Gresham,D., Nishimura,S., Natarajan,P., Koh,J.L., Porter,J., Gray,C.A., Andersen,R.J. *et al.* (2009) A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat. Biotechnol.*, **27**, 369–377.

41. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006 0008.

42. Goodman,A.L., McNulty,N.P., Zhao,Y., Leip,D., Mitra,R.D., Lozupone,C.A., Knight,R. and Gordon,J.I. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe*, **6**, 279–289.