Mutant power: using mutant allele collections for yeast functional genomics

Kaitlyn L. Norman and Anuj Kumar

Corresponding author: Anuj Kumar, Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA. Tel.: 734-647-8060; Fax: 734-647-0884; E-mail: anujk@umich.edu

Abstract

The budding yeast has long served as a model eukaryote for the functional genomic analysis of highly conserved signaling pathways, cellular processes and mechanisms underlying human disease. The collection of reagents available for genomics in yeast is extensive, encompassing a growing diversity of mutant collections beyond gene deletion sets in the standard wild-type S288C genetic background. We review here three main types of mutant allele collections: transposon mutagen collections, essential gene collections and overexpression libraries. Each collection provides unique and identifiable alleles that can be utilized in genome-wide, high-throughput studies. These genomic reagents are particularly informative in identifying synthetic phenotypes and functions associated with essential genes, including those modeled most effectively in complex genetic backgrounds. Several examples of genomic studies in filamentous/pseudohyphal backgrounds are provided here to illustrate this point. Additionally, the limitations of each approach are examined. Collectively, these mutant allele collections in *Saccharomyces cerevisiae* and the related pathogenic yeast *Candida albicans* promise insights toward an advanced understanding of eukaryotic molecular and cellular biology.

Key words: yeast; *Saccharomyces cerevisiae*; *Candida albicans*; functional genomics; transposon; overexpression

Introduction

In this post-genome sequencing era of biological discovery, reverse genetic studies have increasingly come to utilize genome-wide collections of mutant alleles as a fundamental resource. This paradigm is exemplified in analyses of budding yeasts, particularly with respect to the genetic workhorse *Saccharomyces cerevisiae*. The virtues of *S. cerevisiae* for functional genomics are readily evident in light of its genetic tractability, evolutionarily conserved cellular pathways and complement of nearly 1000 genes orthologous to disease-associated genes in humans [1, 2]. The *S. cerevisiae* genome was sequenced in 1996 [3], providing the genetic template to identify and subsequently mutate each gene within the genome. The generation of this genome-wide deletion collection, representing to date the first and only complete collection of deletion alleles in a eukaryote, has been reviewed in full previously and will not be discussed

at length here [4]. The gene deletion collection has unquestionably stimulated research to help elucidate primary functions for the majority of yeast genes; however, there is still a need to explore multifaceted gene functions and discover complex interactions within and between pathways, for which the gene deletion collection as constructed may be insufficient [4–6].

A diverse set of reagents for functional genomics in the yeasts *S. cerevisiae* and *Candida albicans* have been constructed coincident with and subsequent to the construction of the initial gene deletion collection in *S. cerevisiae*. These powerful mutant allele collections have enabled high-throughput and genome-wide studies of essential and nonessential genes, with many yeast allele collections applicable to the analyses of gene functions and gene–gene relationships in nonstandard genetic

Kaitlyn L. Norman received her BS in biology and microbiology from New Mexico State University. Kaitlyn is currently a PhD candidate in the Department of Molecular, Cellular, and Developmental Biology at the University of Michigan. Her research utilizes genomics to identify signaling pathways regulating pseudohyphal growth and hyphal development in *Saccharomyces cerevisiae* and *Candida albicans*.

Anuj Kumar is an Associate Professor in the Department of Molecular, Cellular, and Developmental Biology at the University of Michigan. His research encompasses the application of genomics, proteomics and bioinformatics to dissect signaling pathways regulating eukaryotic stress responses, including filamentous growth, in the budding yeasts *S. cerevisiae* and *C. albicans*.

backgrounds. Here, we review these collections along with their respective advantages, disadvantages and applications in yeast.

Plasmid-based transposon insertion alleles

Transposable elements have long been used as laboratory reagents, and, even with the availability of a genome-wide deletion collection, transposon insertion alleles continue to be a useful resource for functional genomics. Transposon insertions can be generated easily on a large scale, facilitating genome-wide studies. Importantly, plasmid-based insertion allele libraries can be introduced into any number of genetic backgrounds, facilitating phenotypic analyses that cannot be undertaken using the standard S288C background in which the deletion collection was constructed. Additionally, the ability to introduce insertion alleles in a genetic background with a preexisting mutation can ease the workload in generating strains for large-scale analyses of synthetic phenotypes.

Transposon insertion libraries for genomic studies in *S. cerevisiae* have been constructed typically by mutagenesis of a plasmid-based genomic DNA library and subsequent transformation of the library into a desired yeast strain. The transposon-mutagenized DNA integrates into the genome by homologous recombination, yielding a collection of yeast mutants suitable for a variety of applications from phenotypic analyses to studies of protein localization, depending upon transposon design [7–9]. Here, we will very briefly review the widely used transposon insertion libraries generated in Mike Snyder's laboratory.

For these libraries, a bacterial transposon is modified to generate a mini-transposon (mTn) containing only the terminal ends that are necessary for transposition, a *lacZ* reporter, *lox* recombination sites and an epitope tag [7] (Figure 1). The transposon cassette also contains markers for selection in *E. coli* and *S. cerevisiae*. By this design, the transposon can be used to generate gene disruptions and epitope-tagged alleles upon Cre-lox recombination. Transposon insertion libraries generated using Tn3- and Tn7-based elements have been used for large-scale phenotypic screening of full-length insertion alleles and protein localization studies of mTn-encoded epitope-tagged alleles [8, 10–12]. Much of the data from these studies have been made available through the *Saccharomyces* Genome Database and the organelle database Organelle DB [13].

Exemplifying the utility of transposon-mutagenized insertion libraries for the analysis of phenotypes within nonstandard backgrounds, several groups have introduced mTn-based mutant allele collections into strains of *S. cerevisiae* capable of undergoing filamentous growth. Some strains of *S. cerevisiae* can transition from the typical yeast growth form to a filamentous

form, where cells remain connected after cytokinesis in extended pseudohyphae [14]. The S288C strain, from which the original gene deletion collection was derived, is nonfilamentous; instead, derivatives of the strain Σ 1278b are routinely used for studies of pseudohyphal growth. Mösch and Fink introduced a mTn3-derived transposon insertion library into a strain of the Σ 1278b background modified to contain both mating type alleles and a high-copy plasmid carrying *PHD1*, a transcription factor that enhances pseudohyphal growth [15]. By this approach, the screened transformants identified 16 genes yielding decreased pseudohyphal growth. In Jin et al. [16], 3627 transposon insertion alleles were systematically introduced into the filamentous Σ 1278b strain by DNA transformation in 96-well format. The resulting mutants were screened for pseudohyphal growth phenotypes, identifying 309 mTn-disrupted genes yielding increased or decreased surface-spread pseudohyphal filament formation. Suzuki et al. [17] screened mTn3-derived insertion mutants in the Σ 1278b background under conditions of nitrogen limitation for expressed genes exhibiting loss-of-function invasive growth phenotypes. Additionally, Lorenz et al. [18] utilized transposon mutagenesis to determine key regulators of filamentous growth by screening for mTn insertion mutants that do not undergo filamentation in response to 1-butanol. Wild-type *S. cerevisiae* exhibits a pseudohyphal morphology in liquid and on solid media in the presence of butanol, and this screen identified numerous genes encoding polarized growth regulators and mitochondrial proteins.

Applications of transposon insertion libraries in *C. albicans*

Transposon insertion collections have been utilized to great effect in the diploid human fungal pathogen *C. albicans*. Relative to targeted gene deletions, transposon insertion alleles offer several advantages for the large-scale generation of mutants; further, transposons are particularly useful for mutagenesis in *C. albicans* as compared against its closely related yeast *S. cerevisiae*. *C. albicans* does not display a traditional meiotic cycle, thereby eliminating mutant generation strategies that necessitate a stable haploid [19]. Transforming DNA, such as PCR-amplified DNA for the generation of gene deletion mutants, does not integrate by homologous recombination in *C. albicans* with the efficiency observed in *S. cerevisiae*, and high copy 2 μ plasmids, used routinely for overexpression studies in *S. cerevisiae*, do not replicate autonomously in *C. albicans* [20]. Transposon insertion alleles, however, can be generated easily in *C. albicans* using shuttle mutagenesis approaches. Consequently, transposon mutagenesis has been

Figure 1. Summary of transposon-based approaches for functional genomics in *S. cerevisiae* and *C. albicans*. Diagrammatic representation of a Tn7-derived transposon insertion library for mutagenesis of a desired yeast genetic background. Transposon features are not drawn to scale.

used to a significant extent in *Candida* for studies of haploinsufficiency.

Haploinsufficiency refers to a mutant phenotype evident upon deletion or mutation of one allele at a given locus in a diploid cell. Uhl *et al.* [21] constructed a transposon-based heterozygous mutant collection, encompassing 18000 insertions. This collection has been analyzed for phenotypes resulting in altered filamentous growth and hyphal development—growth transitions that are necessary for the virulence of *C. albicans*. From this work, 146 genes were identified; notably, a third of these genes lack orthologs in *S. cerevisiae* [21]. Transposon mutagenesis has been used recently to construct strains for the analysis of complex haploinsufficiency in *C. albicans* (Figure 2). Complex haploinsufficiency is a synthetic genetic interaction where a strain that is heterozygous for mutant alleles at two separate loci exhibits a more severe phenotype than either individual heterozygous mutant alone [22, 23]. Genes exhibiting complex haploinsufficiency typically function in a common cellular structure or functional pathway in cells. When used to identify components of signaling pathways or parallel pathways, this genetic approach can sometimes enable the identification of gene–gene relationships and points of pathway crosstalk. For this purpose, Bharucha *et al.* [22] generated an insertional library by Tn7 transposon mutagenesis of a strain containing a heterozygous deletion of *CBK1*, a gene encoding a key protein involved in the regulation of Ace2 and morphogenesis (RAM) pathway regulating cell morphogenesis, polarized growth and hyphal development. The resulting library consists of 6528 double heterozygous mutants and was screened for hyphal growth phenotypes indicative of complex haploinsufficiency on nutrient-deficient Spider medium [22]. The work identified pathway crosstalk between the RAM and protein kinase A (PKA) signaling networks and represents the first large-scale synthetic genetic interaction analysis in *C. albicans*.

Despite the advantages of transposon-based approaches for the generation of genome-wide mutant collections, transposon insertion libraries constructed in *S. cerevisiae* and *C. albicans* do possess technical limitations. Biases in transposition can make it difficult to achieve insertions evenly over an entire genome. Further, identifying insertion sites in mutants of interest can be challenging; typically, inverse or vectorette PCR approaches have proven successful, as have rapid amplification of cDNA ends (RACE)-based approaches for the identification of insertions in transcribed regions of the genome [22]. Recently, we used high-throughput sequencing to identify insertion sites in genomic DNA extracted from a set of mutants exhibiting a phenotype of interest. On balance, these limitations are offset by the advantages in utilizing insertional libraries, particularly for the generation of mutants in nonstandard genetic backgrounds.

Generation of a genome-wide gene deletion collection in a filamentous strain of *S. cerevisiae*

The construction of a genome-wide set of precise gene deletions in a nonstandard genetic background of *S. cerevisiae*, such as in the aforementioned filamentous $\Sigma 1278b$ strain, provides a powerful resource for functional genomic analysis of processes that cannot be effectively studied in S288C derivatives. Recently, Ryan *et al.* [24], principally in Charlie Boone's laboratory at the University of Toronto, constructed a genome-wide collection of gene deletion mutants in both haploid and diploid $\Sigma 1278b$ backgrounds. The deletion collections were used for systematic phenotypic analysis of pseudohyphal growth, invasive growth and biofilm formation, identifying partially overlapping sets of genes required for these modes of wild-type filamentous growth [24]. The start codon-to-stop codon deletions obviate concerns regarding partially functional alleles generated by random insertional mutagenesis; however, the labor and expense in undertaking such a project is very substantial, counterbalancing the benefit gained from generating precise gene deletions.

Collections of loss-of-function mutations in essential genes

The yeast gene deletion collection has been a valuable resource for genomics studies; however, nearly 19% of yeast genes are necessary for survival/colony formation under laboratory growth conditions and cannot be deleted because of their functions in essential processes, including DNA replication, translation, transcription and cell wall integrity [5]. Not surprisingly, this gene set is biologically important, as 38% of essential genes are conserved in genes associated with human disease [25]. Consequently, there exists a standing need to better characterize essential genes, and, typically, conditional alleles are used for this purpose. To generate a conditional allele, a mutation or regulatable promoter is introduced into the targeted essential gene locus, yielding an allele that is functional under permissive conditions and nonfunctional under another set of conditions, termed restrictive. Several gene collections utilizing conditional mutations have been created to facilitate analyses of essential genes (Figure 3). These collections can be used on a small scale to study individual genes or on a large scale to explore genetic interactions utilizing techniques such as synthetic genetic array (SGA) analysis [25, 26]. Morphological phenotypes resulting from research with these collections has been compiled and can be viewed using the database PhenoM [27].

Figure 2. Application of shuttle mutagenesis to generate double heterozygous mutants for large-scale analysis of complex haploinsufficiency in *C. albicans*. In this illustration, complex haploinsufficiency is assessed with respect to hyphal growth. YFG1, Your Favorite Gene.

Figure 3. Overview of approaches to systematically generate hypomorphic alleles of essential yeast genes. (A) Diagram illustrating a degron-based method to conditionally target a protein of interest for degradation. (B) DAmP constructs for the generation of hypomorphic alleles through reduced transcript stability. A variation of the approach utilizing a TAP tag and the Degron cassette is also illustrated. DNA elements are not drawn to scale.

Tetracycline repressible gene collection

Tetracycline (tet) repressible strain collections contain alleles in which an essential gene's native promoter is replaced with the *Escherichia coli* promoter, TetO. The first and largest sets of tet-repressible strains utilized the TetO₇-promoter and encompassed two collections: the European Consortium collection containing 99 strains, and the Yeast Tet-Promoters Hughes Collection (YTHC) containing over 800 mutant strains [28, 29]. A collection of 173 strains using the TetO₂ promoter has also been constructed [30], containing a gene to express the tetracycline-controlled activator (tTA) that binds to the TetO promoters [29, 30]. Genes under control of the TetO promoters are repressed by the addition of doxycycline, a tetracycline derivative, which inhibits tTA binding [29].

These tet-repressible collections have been used by the yeast research community to study cell cycle progression [31], connect the chromatin structure remodeling (RSC) complex to the maintenance of the nuclear envelope [32] and identify the first proteins implicated in peroxisome biogenesis and assembly [33]. Additionally, tet-repressible alleles have been used to implicate 47 essential genes in preventing DNA damage and random genome rearrangements [34]. These collections have not only been used to study the functions of essential genes in yeast but have proven informative for the discovery of novel yeast host genes that affect RNA replication of the tomato bushy stunt virus [35]. Utilization of the tet-repressible collections is advantageous for many reasons: the genomic context/position and sequence of the targeted yeast genes are unperturbed; doxycycline does not exhibit off-target effects in yeast and does not affect gene expression levels outside of the TetO promoter; and the use of titratable levels of doxycycline allows for the manipulation of expression levels for the gene of interest [29, 36]. When using the tet-repressible promoter system, it is important to keep in mind some key issues. Until repression with doxycycline, the TetO promoter is constitutively active, which can result in gene expression levels that differ from wild type [36]. This constant expression may also eliminate any fluctuating transcriptional regulation that should occur naturally, therefore potentially disrupting complex expression relationships [36]. Large-scale genome-wide screens can be problematic to implement with inducible promoters because an intermediate level of expression may be more difficult to obtain [29]. Furthermore, the tet-promoters have been placed directly in the genome and therefore are only applicable in the strain in which they were created; these alleles are not easily adaptable for use in nonstandard strains. These collections are available in the R1158 background, matching the genetic background of the yeast deletion collection.

Temperature-regulated mutant allele collections

Temperature-sensitive (ts) alleles have been used as a genetic tool long before the sequencing of *S. cerevisiae* [37], although initial applications were undertaken on a small scale, introducing mutations individually by random mutagenesis or plasmid shuffling [38]. As techniques improved, the production of temperature sensitive alleles of essential genes became more efficient, resulting in a collection of over 900 strains covering 65% of essential genes in *S. cerevisiae* [25, 39]. In these mutants, ts alleles are generated at the endogenous locus of each essential gene and contain the *kanMX* resistance cassette in the 3' untranslated region (UTR) [25, 39]. The ts collection contains specific nucleotides that can be used as molecular barcodes for the identification and analysis of individual strains within a

pool of mutants [25]. Since its completion in 2011 by the Boone laboratory, this collection has been utilized for high-throughput studies of chemical-genetic interactions [40], SGA analyses to identify gene networks [26], and projects to determine the role of Parkinson's related genes in cellular pathways [41]. The Stirling laboratory recently constructed an additional barcoded ts allele collection, generating a 'diploid-shuffle' allele set that spans 600 essential genes [42]. This collection has been used to assess genes responsible for an increase in P-bodies [42].

There are many advantages in using ts alleles for yeast functional genomics, including the endogenous location of the gene of interest, the incorporation of barcoding elements for facile identification of strains and the control of genes under native promoters. There are also, however, drawbacks in the application of this collection since activation of the conditional allele is accomplished by shifting yeast to a higher temperature, subsequently inducing heat shock and affecting other physiological functions. Like the tet-repressible collection, these strains are integrated in the genome, restricting experiments to the background strain of the collection.

Another temperature-regulated system for studying essential genes involves the use of a heat-inducible degron cassette that is fused to the amino terminus of an essential gene (Figure 3A) [43–45]. This cassette targets the host protein for proteolysis when grown at 37°C [43, 44]. The degron is acted upon by the N-end-rule pathway that determines a protein's half-life by the amino acid residue in the N terminus [46]. The degron cassette encodes a temperature-sensitive mutant of the mouse protein dihydrofolate reductase (DHFR) containing a ubiquitin molecule that is removed after translation, revealing an unstable arginine residue that is involved in the initiation of ubiquitylation at 37°C [43, 44, 47]. This process rapidly targets the protein for degradation. The degron collection covers 94% of yeast essential genes and is available in both a haploid and diploid GAL1-UBR1 strain that produces Ubr1 protein for recognition of the degron at shifting temperatures [45].

Heat-inducible degrons have been used to study many essential gene functions in cytokinesis [48], spindle assembly [49] and the cell cycle [50]. Though these degrons have proven useful in advancing our knowledge of essential gene functions, they also possess several disadvantages for large-scale applications. Like ts mutants, the shift in temperature to activate the degron induces heat shock, which can have adverse effects on the cell. Furthermore, the large 25 kDa degron cassette can only be placed at the N-terminus, as it utilizes the N-end-rule, possibly impairing proper function and folding of certain proteins. A recently described system, the auxin-inducible degron, has been generated to overcome problems associated with a heat-inducible degron, including a smaller cassette that can be placed at either the N or C terminus and is activated by the plant hormone auxin instead of temperature [51, 52]. This system has been used to replace heat-inducible degrons in many studies but has not yet been used to construct a full collection of conditional alleles.

Hypomorphic alleles and the DAmP collection

Hypomorphic, or partial loss-of-function, alleles present a useful tool toward the analysis of essential gene functions and complex interactions. To generate these alleles for essential genes, a method known as decreased abundance by mRNA perturbation (DAmP) has been utilized, resulting in the generation of a collection of hypomorphic alleles spanning 82% of essential genes in both diploid and haploid backgrounds [53, 54]. DAmP

Table 1. Overexpression libraries for functional genomics in *S. cerevisiae*

Description	Features	Applications/advantages	Reference
GST-tagged library	N-terminal GST tag; expression regulated by GAL1 promoter; encompasses 85% of annotated yeast genes	Strong induction and overexpression in media containing galactose; phenotypic analysis in nonstandard yeast strains	[60]
barFLEX arrays	FLEX overexpression plasmids driven from a galactose-inducible promoter; barcoded yeast strains; FLEX plasmids are fully sequence-verified	Applicable for pooled multiplex analysis; barFLEX arrays utilize a genetic background with incorporated barcodes	[61]
MORF library	MORF-based 2 μ plasmids are Gateway-compatible; genes are expressed from the galactose-inducible GAL1 promoter; genes encode C-terminal fusions to His6, HA and Protein A tags; the MORF library encompasses 5854 genes	Well suited for high-throughput biochemical studies and analyses requiring protein purification; can be introduced into nonstandard genetic backgrounds	[62]
Genomic tiling library	Genomic regions cloned in a high-copy vector; the library encompasses 1588 plasmids in total; genes are present in genomic DNA with native sequence and promoters	The tiling collection can be easily introduced into a variety of genetic backgrounds; overexpression is not galactose-dependent; overexpression alleles exhibit low toxicity	[63]

alleles are constructed by placing a drug resistance marker in the 3'UTR of the gene of interest, destabilizing the resulting mRNA and lowering levels anywhere from 2- to 10-fold depending upon the gene of interest [53] (Figure 3B). For an even stronger reduction in protein levels, this collection can also be combined with a C-terminal tandem affinity purification (TAP) tag-degron [53]. The resulting low level of expression still supports cell survival under normal growth conditions [54]. The original DAmP collection has been barcoded, allowing for further analysis of mutant strains as a pooled sample [55]. The bar-coded DAmP collection encompasses alleles for 958 essential genes along with 444 nonessential, but slow growing, genes; each allele contains two unique oligonucleotide sequences that can be amplified from a set of universal primers [55].

The DAmP hypomorphic allele collections have been utilized effectively in studying epistatic relationships and in generating genetic-interaction maps of essential genes [54]. These alleles have also been pivotal in discovering essential genes that are necessary for the maintenance of telomeres [56] and in discovering a novel pathway for alkylation resistance [57]. Hypomorphic alleles from this library have also been utilized in the study of essential genes involved in arsenic sensitivity [58].

The DAmP collection is subject to many of the advantages and disadvantages discussed in regards to other applications for the large-scale analysis of essential gene function. The genetic background of the alleles is fixed and is consistent with that of the gene deletion collection. While this does promote comparative and parallel studies with deletion mutants, the genetic background is inappropriate for some phenotypic analyses [53]. Additionally, these hypomorphic alleles are not conditional, and expression levels cannot be reverted back to wild-type. However, the strains do enable analysis without necessitating a shift to stressful temperatures or environments and present another informative method through which essential gene function may be addressed systematically or in parallel.

Gene overexpression libraries for large-scale analysis

Gene overexpression is a classic genetic approach for the analysis of gene functions and relationships, complementing the use of loss-of-function mutations or gene deletions. Gene functions

can often be assessed informatively through overexpression and subsequent determination of consequences or benefits to the cell. Additionally, gene overexpression alleles may be useful in instances where the deletion of a gene may not yield an observable phenotype, due to compensatory effects or the absence of a necessary environmental/nutritional stimulus. Overexpression studies can be implemented in a multitude of ways. Suppressor studies can be conducted in which an overexpressed gene negates the phenotype of a mutant gene. Enhancer studies identify a multiplied phenotype resulting from overexpression of a gene in a mutant strain. Genetic interaction networks, epistatic relationships and even drug targets [59] can be examined with overexpression strains. This section will discuss several collections that can be utilized to overexpress both essential and nonessential genes (Table 1). Relative to deletion mutants, these overexpression collections are highly versatile, as constituent plasmids can be introduced by transformation into any number of desired genetic backgrounds of *S. cerevisiae*.

GST-tagged alleles

In 2006, a glutathione S-transferase (GST)-tagged overexpression library was constructed, encompassing 5280 strains with alleles spanning 85% of all yeast genes [60]. Each strain in this collection contains a multi-copy plasmid carrying a single gene coding sequence that is N-terminally tagged with GST and regulated by a galactose-inducible GAL1 promoter [60, 64]. These strains can be grown using standard yeast media and then shifted to galactose-containing media in order to activate overexpression of the specific gene. Despite the clear utility in applying overexpression alleles for genetic screening, this overexpression library does exhibit limitations. In particular, approximately a quarter of the genome is unrepresented in this collection. The GST tag itself may interfere with proper protein folding, function and targeting to certain pathways. Additionally, glucose is a preferred and more readily utilized carbon source that yields higher growth rates than those occurring on alternative carbon sources, such as galactose. Consequently, it can be suboptimal to assay overexpression phenotypes in media containing galactose due to the effects on growth rate, metabolism and gene regulation.

Strains from the GST-tag collection have been instrumental in studies utilizing nonstandard strains such as wine yeast strains. These plasmids have provided a means to overexpress targets of the Rsp5p-Bul1/2p ubiquitin ligase complex and reveal a possible target for evolutionary adaptation against alcoholic fermentation necessary in the production of wine [65].

barFLEX arrays

Douglas *et al.* [61], principally from the Andrews laboratory at the University of Toronto, constructed a collection of bar-coded yeast strains with overexpression plasmids from the Full Length EXpression ready collection (FLEX), yielding a unique overexpression reagent set of over 5500 *S. cerevisiae* genes known as barFLEX. The barFLEX collection cleverly utilizes bar-coding methods that had been used previously in the deletion collection [66] and in the MoBY-ORF [67] library in which each open reading frame (ORF) is cloned in a centromeric plasmid with two unique nucleotide sequences that can be amplified with universal primers and sequenced to identify specific genes. The barcodes thus enable researchers to pool the barFLEX strains for analysis in genetic screens. The bar-FLEX collection is noteworthy as one of the few overexpression collections to be fully sequence-verified [61]. The utility of this collection has been demonstrated for the discovery of toxic mutants as well as for identifying synthetic dosage lethal interactions [61].

Movable ORF (MORF) alleles

The Movable ORF (MORF) collection is currently the most complete overexpression collection in existence, with 5854 unique gene coding sequences represented [62]. It is considered movable because each ORF is flanked by att sequences that allow the contained sequence to be shuttled to any vector that also contains att sites [62]. This allows for evolving diversity in genomics as new vectors for a variety of interactions continue to be developed. The MORF collection is a high copy 2 μ plasmid-based system, with the ORF under transcriptional control of a GAL1 promoter. Each gene coding sequence encodes C-terminal His6, HA and protein A tags, facilitating protein purification and detection [62]. The inclusion of C-terminal tags in the MORF constructs is noteworthy, as the folding or trafficking of many proteins, including genes involved in the secretory pathway, is often inhibited by N-terminal tags [62]. Indeed, the MORF allele collection has been used to study processes of post-translational modification in yeast, identifying 454 novel candidate proteins for glycosylation [62]. The MORF collection has also been used for the further characterization of membrane proteins [68]. Though this system is applicable toward a wide array of experiments, the encoded C-terminal tags are likely to interfere with protein function in some cases. Approximately half of the MORF constructs have been sequence verified to date [62].

The genomic tiling collection

The overexpression libraries above are unquestionably useful; however, each collection is also relatively large. With each plasmid representing a different gene, the transformation of over 5000 individual plasmids into a background of interest is required to utilize the collections described above, and this process can be laborious. To alleviate this concern, a systematic yeast overexpression library was created in the Prelich

laboratory, with each insert representing a genomic DNA region as opposed to an individual gene [63]. For this collection, high-copy plasmids were constructed containing genomic DNA regions in an overlapping tiling pattern, generating a library of 1588 vectors covering greater than 95% of genes with known functions [63]. As validated in an overexpression screen, only 1.5% of the plasmids in this collection resulted in a toxic or lethal phenotype, as compared to 15% of the constructs in the GST-tagged ORF collection [60, 63]. This difference is likely due to toxicity caused by the GST tag and GAL1 promoter that are absent in this collection. The genomic tiling overexpression library has been utilized for complementation studies to validate temperature-sensitive mutants of essential genes [69] and to determine mechanisms of survival during fatty acid toxicity [70].

Functional genomics and systems biology

Applications of mutant allele collections provide valuable insights toward understanding not only specific gene functions but also functions for entire biological pathways, cellular processes and development [71, 72]. The field of systems biology integrates numerous large-scale analyses, including genomics, metabolomics and proteomics, to study and effectively model cell processes and function as a whole. Mutant allele collections have proven useful for system-wide analyses. For example, the DAmP collection has been used to study the relationship between the processes of fitness, pleiotropy and phenotypic robustness [73]. Yeast deletion and mutant collections are useful in creating genome-scale metabolic reconstructions (metabolic GENRES), which provide valuable information in characterizing flux and phenotype-genotype relationships [74, 75]. Furthermore, by understanding gene function, we can more efficiently select targets for subsequent mutation and analysis; results can be used to profile system-wide effects of metabolic pathways, as implemented to study the network for galactose-utilization in yeast [76].

Outlook

Its compact genome and genetic tractability have established *S. cerevisiae* at the forefront of organisms for functional genomics, and the years subsequent to the sequencing of the yeast genome have witnessed the construction of many clever and informative mutant allele collections for large-scale analyses of gene and protein function. These resources are notable for the breadth of studies they enable, and in the foreseeable future, it is reasonable to expect further growth in the data types and insights generated through yeast-based research. It is also notable that equivalent mutant allele collections are largely unavailable for analyses in other eukaryotes; consequently, yeast functional genomic studies continue to provide templates for the eventual analysis of other eukaryotic systems. These ongoing studies need to be supported by the greater scientific community, as the resulting promise of data and resources for the modeling of cellular processes is substantial. Yeast genomics is informing our understanding of complex processes, relevant, for example, in identifying the mechanisms by which hundreds of genes can induce a precisely integrated pseudohyphal differentiation response as discussed here. The importance of furthering this basic understanding of cell signaling and function should not be

shortchanged in immediate pursuit of disease treatments, as the era of functional genomics in *S. cerevisiae* and *C. albicans* continues.

Key Points

- Collections of mutant alleles, extending beyond gene deletions, continue to be used for yeast-based functional genomics.
- Transposon insertion libraries are effective tools for genome-wide mutagenesis in nonstandard genetic backgrounds of *S. cerevisiae* and *C. albicans*.
- Systematic gene deletion collections constructed in nonstandard genetic backgrounds of *S. cerevisiae* can be used to analyze processes, such as pseudohyphal growth, that cannot be studied effectively in S288C-based strains.
- Essential gene functions can be studied systematically using collections of temperature-sensitive mutants as well as through other allele collections employing conditional transcriptional or translational regulation.
- Plasmid-based overexpression alleles have been used for large-scale studies of cell signaling, encompassing individual genes cloned under inducible promoters and high-copy plasmids containing genomic regions.

Funding

This work was supported by the National Institutes of Health (1R01-A1098450-01A1, R21-A1-114837) (to A.K.); and the March of Dimes (1-FY11-403) (to A.K.).

References

- Heinicke S, Livstone MS, Lu C, et al. The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. *PLoS ONE* 2007;2:e766.
- Botstein D, Fink GR. Yeast: an experimental organism for 21st century biology. *Genetics* 2011;189:695–704.
- Dujon B. The yeast genome project: what did we learn? *Trends Genet* 1996;12:263–70.
- Giaever G, Nislow C. The yeast deletion collection: a decade of functional genomics. *Genetics* 2014;197:451–65.
- Giaever G, Chu AM, Ni L, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002;418:387–91.
- Oliver SG. From DNA sequence to biological function. *Nature* 1996;379:597–600.
- Kumar A. Multipurpose transposon insertion libraries for large-scale analysis of gene function in yeast. In: Osterman A, Gerdes S (eds). *Microbial Gene Essentiality: Protocols and Bioinformatics*. Totowa, NJ: Humana Press, 2008, 117–29.
- Kumar A, Siringhaus M, Biery MC, et al. Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon. *Genome Res* 2004;14:1975–86.
- Xu T, Bharucha N, Kumar A. Genome-wide transposon mutagenesis in *Saccharomyces cerevisiae* and *Candida albicans*. *Methods Mol Biol* 2011;765:207–24.
- Ross-Macdonald P, Coelho PSR, Roemer T, et al. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 1999;402:413–18.
- Manoil C, Traxler B. Insertion of in-frame sequence tags into proteins using transposons. *Methods* 2000;20:55–61.
- Kumar A, Agarwal S, Heyman JA, et al. Subcellular localization of the yeast proteome. *Genes Dev* 2002;16:707–19.
- Wiwatwattana N, Landau CM, Cope GJ, et al. Organelle DB: an updated resource of eukaryotic protein localization and function. *Nucleic Acids Res* 2007;35:D810–14.
- Dickinson JR. Filament formation in *Saccharomyces cerevisiae*—a review. *Folia Microbiol (Praha)* 2008;53:3–14.
- Mosch HU, Fink GR. Dissection of filamentous growth by transposon mutagenesis in *Saccharomyces cerevisiae*. *Genetics* 1997;145:671–84.
- Jin R, Dobry CJ, McCown PJ, et al. Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. *Mol Biol Cell* 2008;19:284–96.
- Suzuki C, Hori Y, Kashiwagi Y. Screening and characterization of transposon-insertion mutants in a pseudohyphal strain of *Saccharomyces cerevisiae*. *Yeast* 2003;20:407–15.
- Lorenz MC, Cutler NS, Heitman J. Characterization of alcohol-induced filamentous growth in *Saccharomyces cerevisiae*. *Mol Biol Cell* 2000;11:183–99.
- Noble SM, Johnson AD. Genetics of *Candida albicans*, a diploid human fungal pathogen. *Annu Rev Genet* 2007;41:193–211.
- Molero G, Diez-Orejas R, Navarro-Garcia F, et al. *Candida albicans*: genetics, dimorphism and pathogenicity. *Int Microbiol* 1998;1:95–106.
- Uhl M, Biery M, Craig N, et al. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C. albicans*. *EMBO J* 2003;22:2668–78.
- Bharucha N, Chabrier-Rosello Y, Xu T, et al. A large-scale complex haploinsufficiency-based genetic interaction screen in *Candida albicans*: analysis of the RAM network during morphogenesis. *PLoS Genet* 2011;7:e1002058.
- Haarer B, Viggiano S, Hibbs MA, et al. Modeling complex genetic interactions in a simple eukaryotic genome: actin displays a rich spectrum of complex haploinsufficiencies. *Genes Dev* 2007;21:148–59.
- Ryan O, Shapiro RS, Kurat CF, et al. Global gene deletion analysis exploring yeast filamentous growth. *Science* 2012;337:1353–6.
- Li Z, Vizeacoumar FJ, Bahr S, et al. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol* 2011;29:361–7.
- Kuzmin E, Sharifpoor S, Baryshnikova A, et al. Synthetic genetic array analysis for global mapping of genetic networks in yeast. In: Smith JS, Burke DJ (eds). *Yeast Genetics*. New York: Springer, 2014, 143–68.
- Jin K, Li J, Vizeacoumar FS, et al. PhenoM: a database of morphological phenotypes caused by mutation of essential genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2012;40:D687–94.
- Hughes TR, Marton MJ, Jones AR, et al. Functional discovery via a compendium of expression profiles. *Cell* 2000;102:109–26.
- Mnaimneh S, Davierwala AP, Haynes J, et al. Exploration of essential gene functions via titratable promoter alleles. *Cell* 2004;118:31–44.
- Wishart JA, Osborn M, Gent ME, et al. The relative merits of the tetO2 and tetO7 promoter systems for the functional analysis of heterologous genes in yeast and a compilation of essential yeast genes with tetO2 promoter substitutions. *Yeast* 2006;23:325–31.

31. Yu L, Pena Castillo L, Mnaimneh S, et al. A survey of essential gene function in the yeast cell division cycle. *Mol Biol Cell* 2006;17:4736–47.
32. Titus LC, Dawson TR, Rexer DJ, et al. Members of the RSC chromatin-remodeling complex are required for maintaining proper nuclear envelope structure and pore complex localization. *Mol Biol Cell* 2010;21:1072–87.
33. Perry RJ, Mast FD, Rachubinski RA. Endoplasmic reticulum-associated secretory proteins Sec20p, Sec39p, and Dsl1p are involved in peroxisome biogenesis. *Eukaryot Cell* 2009;8:830–43.
34. Cheng E, Vaisica JA, Ou J, et al. Genome rearrangements caused by depletion of essential DNA replication proteins in *Saccharomyces cerevisiae*. *Genetics* 2012;192:147–60.
35. Jiang Y, Serviène E, Gal J, et al. Identification of essential host factors affecting tombusvirus RNA replication based on the yeast Tet promoters Hughes Collection. *J Virol* 2006;80:7394–404.
36. Bauer CR, Li S, Siegal ML. Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. *Mol Syst Biol* 2015;11:773.
37. Hartwell LH, Culotti J, Reid B. Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proc Natl Acad Sci USA* 1970;66:352–9.
38. Sikorski RS, Boeke JD. In vitro mutagenesis and plasmid shuffling: from cloned gene to mutant yeast. *Methods Enzymol* 1991;194:302–18.
39. Ben-Aroya S, Coombes C, Kwok T, et al. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of *Saccharomyces cerevisiae*. *Mol Cell* 2008;30:248–58.
40. McCulley A, Haarer B, Viggiano S, et al. Chemical suppression of defects in mitotic spindle assembly, redox control, and sterol biosynthesis by hydroxyurea. *G3* 2014;4:39–48.
41. Chesi A, Kilaru A, Fang X, et al. The role of the Parkinson's disease gene PARK9 in essential cellular pathways and the manganese homeostasis network in yeast. *PLoS ONE* 2012;7:e34178.
42. Kofoed M, Milbury KL, Chiang JH, et al. An updated collection of sequence barcoded temperature sensitive alleles of yeast essential genes. *G3* 2015;5:1879–87.
43. Dohmen R, Wu P, Varshavsky A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* 1994;263:1273–6.
44. Dohmen RJ, Varshavsky A. Heat-inducible degron and the making of conditional mutants. *Methods Enzymol* 2005;399:799–822.
45. Kanemaki M, Sanchez-Diaz A, Gambus A, et al. Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature* 2003;423:720–4.
46. Varshavsky A. The N-end rule pathway of protein degradation. *Genes Cells* 1997;2:13–28.
47. Sanchez-Diaz A, Kanemaki M, Marchesi V, et al. Rapid depletion of budding yeast proteins by fusion to a heat-inducible degron. *Sci STKE* 2004;2004:PL8.
48. Nkosi PJ, Targosz B-S, Labib K, et al. Hof1 and Rvs167 have redundant roles in actomyosin ring function during cytokinesis in budding yeast. *PLoS ONE* 2013;8:e57846.
49. Ren Q, Liou L-C, Gao Q, et al. Bir1 deletion causes malfunction of the spindle assembly checkpoint and apoptosis in yeast. *Front Oncol* 2012;2:93.
50. Gallo-Fernández M, Saugar I, Ortiz-Bazán MÁ, et al. Cell cycle-dependent regulation of the nuclease activity of Mus81-Eme1/Mms4. *Nucleic Acids Res* 2012;40(17):8325–35.
51. Nishimura K, Kanemaki MT. Rapid depletion of budding yeast proteins via the fusion of an auxin-inducible degron (AID). *Curr Protoc Cell Biol* 2014;64:20.9.1–16.
52. Morawska M, Ulrich HD. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast (Chichester, England)* 2013;30:341–51.
53. Breslow DK, Cameron DM, Collins SR, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* 2008;5:711–18.
54. Schuldiner M, Collins SR, Thompson NJ, et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005;123:507–19.
55. Yan Z, Costanzo M, Heisler LE, et al. Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nat Methods* 2008;5:719–25.
56. Ungar L, Yosef N, Sela Y, et al. A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic Acids Res* 2009;37:3840–9.
57. Svensson JP, Pseudo LQ, Fry RC, et al. Genomic phenotyping of the essential and non-essential yeast genome detects novel pathways for alkylation resistance. *BMC Syst Biol* 2011;5:157.
58. Takahashi T, Satake S, Hirose K, et al. A screening for essential cell growth-related genes involved in arsenite toxicity in *Saccharomyces cerevisiae*. *J Toxicol Sci* 2011;36:859–61.
59. Bharucha N, Kumar A. Yeast genomics and drug target identification. *Comb Chem High Throughput Screen* 2007;10:618–34.
60. Sopko R, Huang D, Preston N, et al. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 2006;21:319–30.
61. Douglas AC, Smith AM, Sharifpoor S, et al. Functional analysis with a barcoder yeast gene overexpression system. *G3* 2012;2:1279–89.
62. Gelperin DM, White MA, Wilkinson ML, et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev* 2005;19:2816–26.
63. Jones GM, Stalker J, Humphray S, et al. A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. *Nat Methods* 2008;5:239–41.
64. Zhu H, Klemic JF, Chang S, et al. Analysis of yeast protein kinases using protein chips. *Nat Genet* 2000;26:283–9.
65. Mangado A, Tronchoni J, Morales P, et al. An impaired ubiquitin ligase complex favors initial growth of auxotrophic yeast strains in synthetic grape must. *Appl Microbiol Biotechnol* 2015;99:1273–86.
66. Giaever G, Chu AM, Ni L, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002;418:387–91.
67. Ho CH, Magtanong L, Barker SL, et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* 2009;27:369–77.
68. White MA, Clark KM, Grayhack EJ, et al. Characteristics affecting expression and solubilization of yeast membrane proteins. *J Mol Biol* 2007;365:621–36.
69. Li Z, Vizeacoumar FJ, Bahr S, et al. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotech* 2011;29:361–7.
70. Sec P, Garaiova M, Gajdos P, et al. Baker's yeast deficient in storage lipid synthesis uses cis-vaccenic acid to reduce unsaturated fatty acid toxicity. *Lipids* 2015;50:621–30.

71. Snyder M, Gallagher JEG. Systems biology from a yeast omics perspective. *FEBS Lett* 2009;**583**:3895–9.
72. Ge H, Walhout AJM, Vidal M. Integrating ‘omic’ information: a bridge between genomics and systems biology. *Trends Genet* 2003;**19**:551–60.
73. Bauer CR, Li S, Siegal ML. Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. *Mol Syst Biol* 2015;**11**:773.
74. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 2009;**5**:320.
75. Burgard AP, Nikolaev EV, Schilling CH, et al. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 2004;**14**:301–12.
76. Ideker T, Thorsson V, Ranish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;**292**:929–34.