Received: 10 May 2023 | Revised: 4 January 2024 | Accepted: 16 January 2024

DOI: 10.1002/yea.3927

RESEARCH ARTICLE

Yeast WILEY

An ultra high-throughput, massively multiplexable, single-cell RNA-seq platform in yeasts

Leandra Brettner[1] | Rachel Eder[1,2] | Kara Schmidlin[1,2] | Kerry Geiler-Samerotte[1,2]

[1]Biodesign Institute Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

[2]School of Life Sciences, Arizona State University, Tempe, Arizona, USA

Correspondence
Leandra Brettner, Biodesign Institute Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA.
Email: leandra.brettner@asu.edu

Funding information
National Institutes of Health, Grant/Award Number: R35GM133674; National Science Foundation, Biological Integration Institution, Grant/Award Number: 2119963; Alfred P. Sloan Research Fellowship in Computational and Molecular Evolutionary Biology, Grant/Award Number: FG-2021-15705

Abstract

Yeasts are naturally diverse, genetically tractable, and easy to grow such that researchers can investigate any number of genotypes, environments, or interactions thereof. However, studies of yeast transcriptomes have been limited by the processing capabilities of traditional RNA sequencing techniques. Here we optimize a powerful, high-throughput single-cell RNA sequencing (scRNAseq) platform, SPLiT-seq (Split Pool Ligation-based Transcriptome sequencing), for yeasts and apply it to 43,388 cells of multiple species and ploidies. This platform utilizes a combinatorial barcoding strategy to enable massively parallel RNA sequencing of hundreds of yeast genotypes or growth conditions at once. This method can be applied to most species or strains of yeast for a fraction of the cost of traditional scRNAseq approaches. Thus, our technology permits researchers to leverage "the awesome power of yeast" by allowing us to survey the transcriptome of hundreds of strains and environments in a short period of time and with no specialized equipment. The key to this method is that sequential barcodes are probabilistically appended to cDNA copies of RNA while the molecules remain trapped inside of each cell. Thus, the transcriptome of each cell is labeled with a unique combination of barcodes. Since SPLiT-seq uses the cell membrane as a container for this reaction, many cells can be processed together without the need to physically isolate them from one another in separate wells or droplets. Further, the first barcode in the sequence can be chosen intentionally to identify samples from different environments or genetic backgrounds, enabling multiplexing of hundreds of unique perturbations in a single experiment. In addition to greater multiplexing capabilities, our method also facilitates a deeper investigation of biological heterogeneity, given its single-cell nature. For example, in the data presented here, we detect transcriptionally distinct cell states related to cell cycle, ploidy, metabolic strategies, and so forth, all within clonal yeast populations grown in the same environment. Hence, our technology has two obvious and impactful applications for yeast research: the first is the general study of transcriptional phenotypes across many strains and environments, and the second is investigating cell-to-cell heterogeneity across the entire transcriptome.

KEYWORDS

Candida albicans, high-throughput, low-cost, *Saccharomyces cerevisiae*, single-cell RNA sequencing, yeast

1 | INTRODUCTION

Yeasts are probably one of the most well-researched lifeforms and serve as workhorse model organisms for exploring genetics, evolutionary, cell, and systems biology. They are arguably the most genetically tractable eukaryotes with the easy construction of comprehensive gene deletion collections (Giaever et al., 2002) and, more recently, libraries of thousands of mutant strains in a single genetic editing experiment (Levy et al., 2015; Sharon et al., 2018). There is also a wealth of well-annotated, natural genetic diversity available for study (Liti et al., 2009; Peter, 2018). However, measuring the phenotypic effects of the myriad genetic differences that separate these strains is challenging because most phenotyping approaches require each strain to be studied separately. Further, to comprehensively understand the differences between each genotype, each strain should be phenotyped in multiple conditions to grasp the effects of environmental context. This scale quickly multiplies into a number of experiments that are hard-to-impossible to manage.

Single-cell RNA sequencing (scRNAseq), or the profiling of RNA expression in individual cells, has become a powerful tool that can enable the high-throughput interrogation of transcriptional phenotypes across multiple strains and multiple conditions simultaneously (Dixit et al., 2016; Jackson et al., 2020; Rodriguez-Fraticelli et al., 2020). While several recent papers have been published applying scRNAseq to yeast, these papers utilize scRNAseq methods involving physical isolation of cells and accumulate expense both in the specialized equipment needed and the number of cells processed (Kolodziejczyk et al., 2015; Liu & Trapnell, 2016; Urbonaite et al., 2021). Recently developed combinatorial barcoding methods (SPLiT-seq.; Rosenberg et al., 2018; sciSeq; Cao et al., 2017). solve these issues by allowing cells from multiple strains or environments to be phenotyped while pooled, and are thus more easily scalable. SPLiT-seq also only requires basic benchtop tools. However, to date, SPLiT-seq has been applied to a variety of organisms (Cao et al., 2017; Kuchina et al., 2021; Rosenberg et al., 2018) but, to our knowledge, has not been adapted for fungi. Here we present a yeast-optimized version of SPLiT-seq (Kuchina et al., 2021; Rosenberg et al., 2018).

SPLiT-seq has many advantages as it allows for the sequencing of transcriptomes of many many cells, giving us more power to profile many yeast strains and conditions as well as phenotypic heterogeneity. For example, our yeast-optimized SPLiT-seq method can process approximately 400,000 cells for approximately $2000 (a full cost breakdown is provided in the methods), while droplet-based methods would cost approximately $2000 per sample, with typically up to 10 K cells per sample and up to eight samples, based on price listings on the University of Kansas Medical Center, Boston University Medical Center Sequencing Core, and Cornell Institute of Biotechnology websites (10X Pricing, 2024; 10X Genomics single-cell libraries, 2024; Pricing, 2024). Note that this calculation does not include the large cost of the instrument ($65K–350K), (Pricing & Quoting – 10x Genomics, 2024) but assumes that researchers are submitting their samples to sequencing cores or companies. And, as SPLiT-seq implements combinatorics, it can differentiate as many samples (i.e., genotypes or environmental conditions) as there are first barcodes (e.g., 96 or 384 per plate per experiment depending on the type of multi-well plates used). Given the strength of the yeast system is the ability to study diverse genotypes and environments, optimizing SPLiT-seq for yeast unlocks the power of this model organism for studying transcriptomics.

Take-away

- We adapted the high-throughput, multiplexable, and relatively low-cost method for performing single-cell RNA sequencing, SPLiT-seq, to yeast.
- This adaptation works successfully in multiple ploidies and species of yeast including haploid and diploid *Saccharomyces cerevisiae* and *Candida albicans*.
- Species identity, ploidy, and environmental conditions can all be revealed through basic transcriptomics analyses.

2 | TECHNIQUE OVERVIEW

In opposition to isolation-based scRNAseq methods that typically utilize microfluidic droplets to contain the RNA from a single cell (Dohn, 2021; Macosko et al., 2015; McNulty et al., 2021; Zheng et al., 2017), the SPLiT-seq protocol uses the cell itself as a container for its own RNA, and all enzymatic reactions are performed in situ (Figure 1; Kuchina et al., 2021; Rosenberg et al., 2018). Fixed and permeabilized cells are loaded into a multiwell plate where they undergo in situ reverse transcription with well-specific barcoded random hexamer and poly-dT primers. They are then pooled and split into another 96-well plate where each well contains a short, unique barcode sequence that anneals to the first barcode via a linker strand. A ligation reaction covalently bonds these two pieces of DNA at the single-stranded nick created. The cells are subsequently pooled and split into a new plate where the process is repeated, adding a third barcode. Cells that received the same first barcode are unlikely to receive the same second and third barcodes. This process is completed n times depending on the population size, as unique barcode combination possibilities scale exponentially with each additional round (e.g., 96 barcodes, n split-pools = 96^n possible barcode combinations). Each cell is thus uniquely labeled by probabilistically biasing the outcome such that it takes its own path through the barcode plates. The basic SPLiT-seq protocol includes one round of reverse transcription and two rounds of ligation, generating cell barcodes from combinations of three oligo additions. Finally, the cells are lysed, and the extracted cDNA is prepped for sequencing (Figure 1). After sequencing, combinatorial barcodes are used to computationally resolve through which wells a cell has traveled to get single-cell data (Kuchina et al., 2021; Rosenberg et al., 2018).

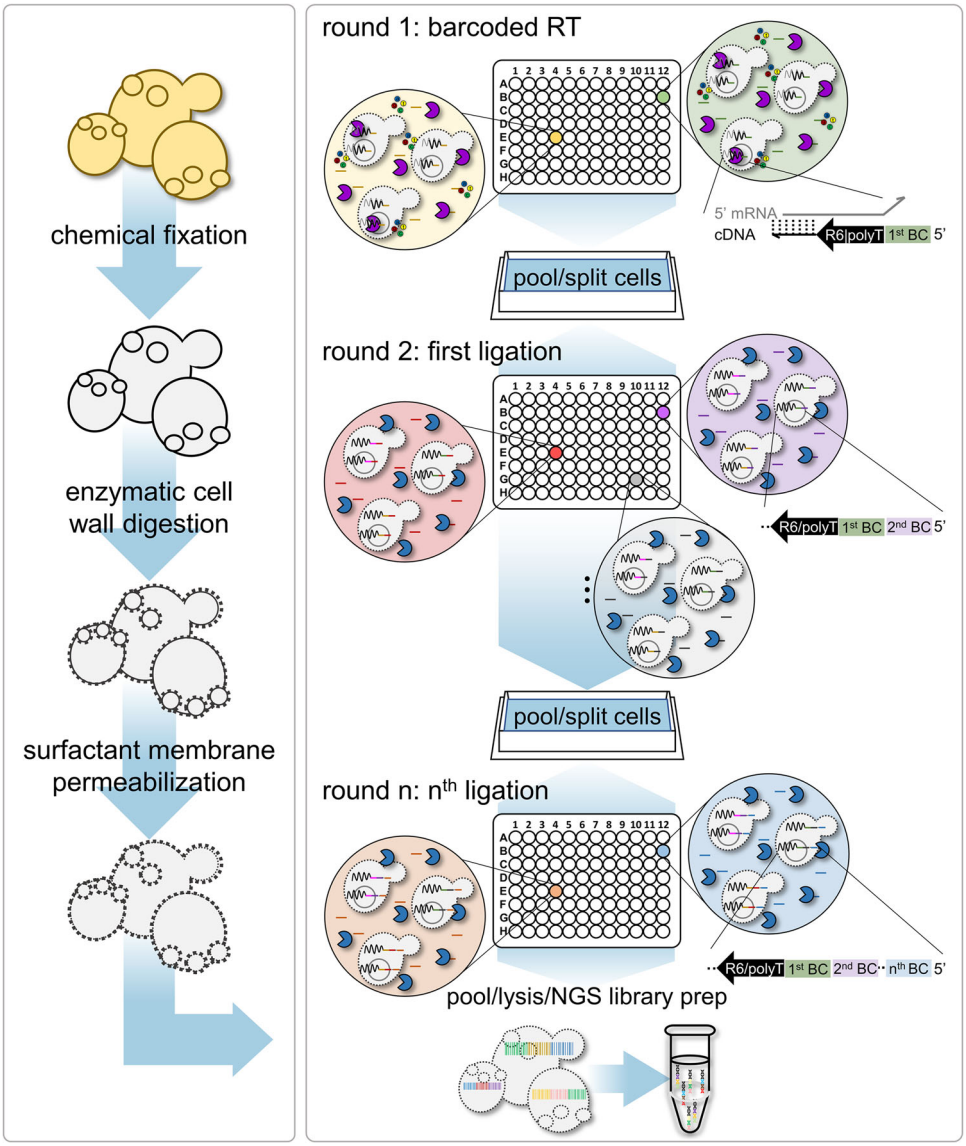

FIGURE 1 Yeast-optimized SPLiT-seq. (Left panel) Yeast-specific protocols were developed for the fixation, cell wall enzymatic digestion, and surfactant membrane permeabilization steps. These are the steps that are the most important to adapt to the working organism, as each creature is going to have different structural morphologies that will need customization, particularly for the permeabilization steps. For example, yeasts have cell walls, whereas mammalian cells do not. (Right panel) The split-pool reverse transcription and ligation steps are carried out in multi-well plates similar to the original protocols (Rosenberg et al., 2018; Kuchina et al., 2021). The resulting cNDA libraries are prepared and submitted for sequencing.

The intrinsic power of SPLiT-seq comes from its easy and inexpensive capabilities to scale up the number of cells processed. The number of cells that can be processed is a function of the number of wells in each barcoding step and the number of ligation reactions performed (Figure 2a). For example, one round of barcoded reverse transcription with 96 barcodes and two subsequent ligation steps, also with 96 barcodes, produces 96^3 (884,736) possible unique combinations. Adding one more ligation reaction yields almost 85 million unique combinations with only the trivial cost of adding another plate of 96 barcode oligos and reagents (Figure 2a; Kuchina et al., 2021; Rosenberg et al., 2018).

Another advantage of SPLIT-seq is the ability to multiplex many samples (i.e., genotypes or growth conditions) in the same run. Since each sample can be loaded intentionally into a specified well of a 96- or 384-well plate during the reverse transcription step, the first barcode in the combinatorial sequence can be used as a conditional signifier. This allows for the processing of as many unique samples as there are wells in the first step (Figure 2b). So, for example, if a 96-well scheme is used, 96 samples representing 96 genotypes or 96 environments can be processed in that run. In theory, using expressible barcodes (Figure 2c; Jackson et al., 2020) to label unique genotypes could increase the number of samples even further,

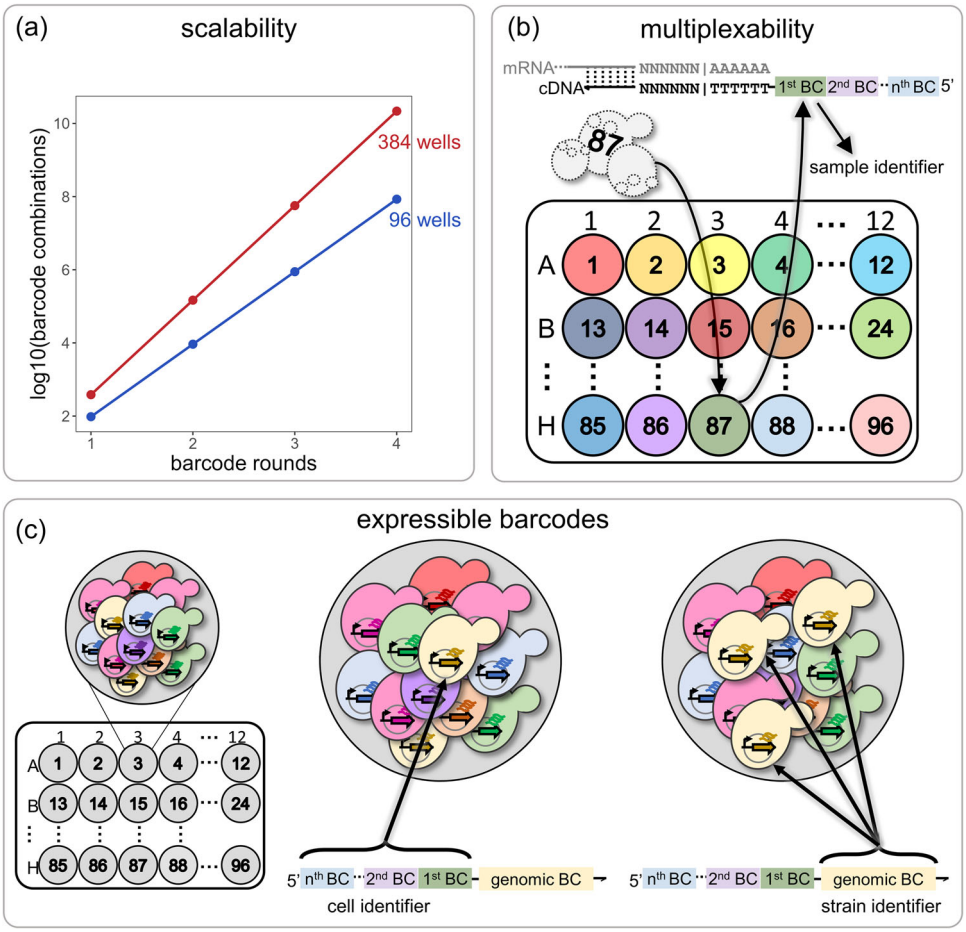

FIGURE 2 The high-throughput power of the SPLiT-seq method. (a) The number of available unique barcode combinations scales exponentially with addition of relatively few new barcodes (increments of 96 or 384). A similar scaling analysis can be found in the original SPLiT-seq publication (Rosenberg et al., 2018) (b) The number of samples that can be processed in a single run is equal to the number of barcoded primers in the reverse transcription step, which provides the first barcode in the combinatorial sequence. For example, 96 or 384 samples could be processed with standard multiwell plates. (c) Expressible barcodes, or expressed engineered sequences that identify genotype, can be combined with the SPLiT-seq method to provide large genotype and environment combinatorial power.

enabling massive GxE screens. Assuming each strain's transcriptome requires 500 cells per experiment to achieve adequate coverage, a single standard SPLiT-seq run (3 rounds of 96 barcodes) could process almost 900,000 cells. This is enough to cover as many as 1800 genotypes if expressible barcodes are used or 1800 different combinations of genotype and environment. As each SPLiT-seq run takes less than a week to perform, hundreds of conditions and thousands of strains could easily be sampled in only a few months. Additionally, as cells are fixed right after sampling, early samples can be stored in the freezer until all experiments are finished and ready for scRNAseq preparations.

In sum, SPLIT-seq is a high-throughput method for single-cell RNA sequencing that lends itself to the yeast system because there are so many engineered and natural yeast genotypes to explore, and so many environments in which yeast can be grown in the laboratory (Costanzo et al., 2016; Gasch et al., 2000; Kinsler et al., 2020). Here we optimize the method for the unique physiology of yeasts.

3 | RESULTS

3.1 | Chemical fixation, beta-glucan specific enzymes, and nonionic detergents successfully prepare yeast cells for in situ SPLiT-seq reactions

Yeasts are unique in that they possess a fungal cell wall. This wall must be permeabilized just enough to allow SPLiT-seq reagents to enter cells but not enough such that the cell can no longer act as a container for the reaction. Optimizing fixation and permeabilization methods in series is challenging as testing them one at a time would require going through the entire approximately $2000 protocol and sequencing every test. This is necessary because there are no good intermediate steps in the protocol to test library quality. Realistically, one must take the cells through the entire SPLiT-seq protocol, library preparation, and sequencing before one can discern the final data quality. cDNA gel traces or library concentrations are not a good indicator of barcoding success or of the number of reads or genes

that will be recovered per cell. Fortunately, the multiplexable nature of the SPLiT-seq method allowed us to test many cell fixation and permeabilization treatments in parallel, greatly reducing the time and cost of exploring the parameter space.

Fluorescent in situ hybridization cell preparation methods are an ideal place to start when adapting SPLiT-seq to any new organism, as FISH also relies on efficiently diffusing reagents into cells. Based on a method adapting FISH for flow-cytometry in yeast (Bertin et al., 1990), we initially tested (Figure 3, Experiment 1) 16 combinations of ethanol and/or formaldehyde fixation with an enzyme mixture, zymolyase, which degrades the beta-glucans in the cell wall, and nonionic detergents Tween-20 or Triton X-100 to permeabilize the cell membrane in both *Saccharomyces cerevisiae* and *Candida albicans* (Figure 3a). After identifying the best-performing initial condition by choosing the one that yielded the greatest median UMIs/cell (Figure 3a; red), we performed a second experiment (Figure 3, Experiment 2). In this experiment, we tested an additional 18 conditions representing more subtle perturbations of the same variables but found no further improvement (data not shown). The optimal protocol we have so far established by selecting the conditions yielding the greatest median UMIs/cell (Figure 3a; red), which works in multiple yeast species (Figure 3), is the one described

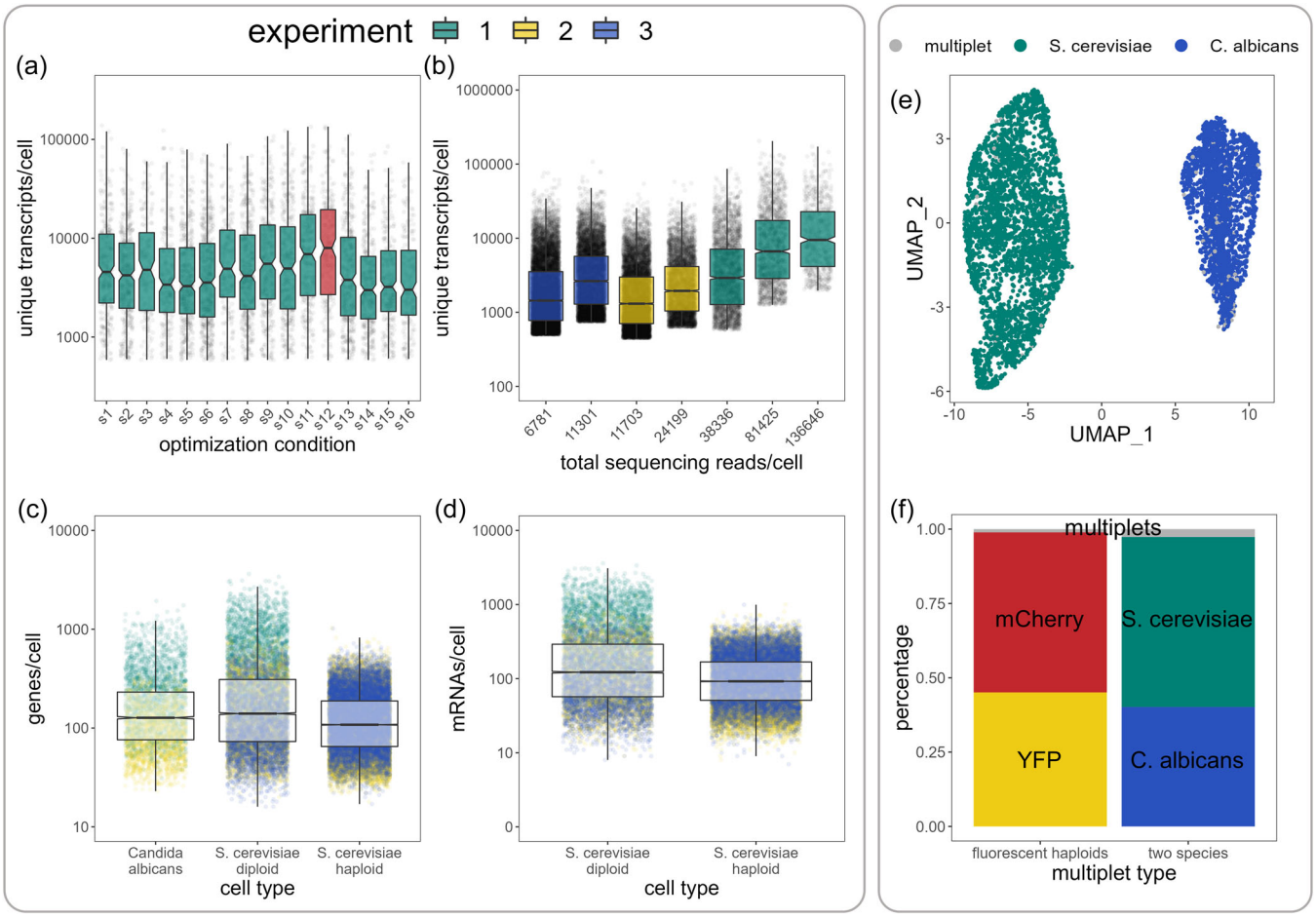

FIGURE 3 Yeast-optimized SPLiT-seq produces quality scRNAseq data. (a) Initial 16 optimization conditions from Experiment 1 comprising combinations of ethanol and/or formaldehyde fixations and zymolyase and/or detergent permeabilizations. While there were no strong differences between any of the conditions, we moved forward with the combination with the highest median unique transcripts per cell (highlighted in red), which is described in the methods. (b) Total unique transcripts detected per cell (including rRNAs) for multiple sublibraries across three experiments. Each point represents a single cell and boxes represent 1st–3rd quartiles. Each sublibrary contains 10,133, 5465, 13, 299, 9177, 2960, 1705, and 703 cells, respectively. Median transcripts per cell range from 1500 to 9500 per cell. When fewer cells are profiled (rightmost sublibraries), the sequencing depth per cell and the number of unique transcripts per cell increase. We use unique molecular identifiers to distinguish unique transcripts from duplicates created during PCR. (c) Total genes per cell detected in *C. albicans*, and diploid and haploid *S. cerevisiae*. Box represents 1st–3rd quartiles. (d) Total mRNAs per cell detected in diploid and haploid *S. cerevisiae*. Box represents 1st–3rd quartiles. (e) UMAP visual clustering of *S. cerevisiae* and *C. albicans* data. Each point represents a single barcode or cell, and its position relative to other points correlates with the similarity of their transcriptomes. The points are colored by their mapped or "true" identity. We see a distinct separation of the two species. (f) Percentages of barcodes that classify as YFP or mCherry expressing cells, *S. cerevisiae* or *C. albicans*, or multiplets (barcodes that have any reported expression of the opposing fluorescent reporter, or greater than 15% of the opposing species' genome).

below in the methods section. Briefly, it includes a 4% formaldehyde overnight fixation and a combination of 0.005 U/μL zymolyase at 37°C for 15 min and 0.4% Triton X-100 at 4°C for 3 min permeabilization steps. The additional 15 conditions tested in Figure 3a are also listed in the methods.

3.2 | Yeast-optimized SPLiT-seq detects thousands of unique RNAs per cell for 43,388 cells of two yeast ploidies and species

After modifying the protocol to work with yeasts, SPLiT-seq detects a median of ~1000 to 10,000 unique RNAs per cell depending on sequencing depth (Figure 3b). Higher sequencing coverage results in a deeper sampling of the transcriptome as can be seen in Figure 3b which shows median unique RNAs per cell for technical replicates of three experiments of different sequencing depths. In experiment 3, the barcoded cells were evenly divided into 10 sublibraries. The two sequenced sublibraries returned ~5,500 and 10,000 barcoded cells that passed computational filtering. Given we started the experiment with between 200,000 and 300,000 cells, this estimates a capturing efficiency between 25% and 50%.

SPLiT-seq is effective in both haploid and diploid cells of *S. cerevisiae*, and also works with *C. albicans*. We recover, on average, between 100 and 600 genes per cell (Figure 3c). Of the detected mRNAs, we see an average of ~120–700 mRNAs per cell depending on the experiment and sequencing depth. Across all cells, we see the range of detected mRNAs being between ~10 to 3600 per cell (Figure 3d). Given that there are approximately 30,000 mRNA molecules per cell for *Saccharomyces cereivisiae*, (Miura, 2008) and assuming haploid and diploid *S. cerevisiae* and *C. albicans* have transcriptome sizes in the same order of magnitude, this range represents 0.2-12% of the active transcriptome in a yeast cell. This level of single-cell transcriptome recovery is similar to that obtained using droplet-based methods in yeast, (Jackson et al., 2020; Jariani et al., 2020; Nadal-Ribelles et al., 2019; Urbonaite et al., 2021) and can likely be further optimized (see Discussion). On average, the transcript proportions we recover are 93.7% rRNA, 0.005% tRNA, 0.04% ncRNA, and 5.75% mRNA. The low rate of mRNA recovery is representative of the true fraction of a yeast cell's RNA belonging to mRNA (5%) (Warner, 1999). We recover the expected ratio of rRNA to mRNA because we use both polydT and random hexamer primers in the reverse transcription step. The addition of random hexamers allows for better coverage of the 5' ends of mRNA, but also increases rRNA recovery. Users who prefer to study mRNA should lower the percentage of rRNA recovered by decreasing the ratio of random hexamer to polydT primers, or by removing rRNA with commercially available kits post cell lysis, following previous work (Kuchina et al., 2021; O'Neil et al., 2013). Alternatively, a relatively low mRNA recovery rate is not necessarily a problem, as algorithms leverage the fact that the transcripts recovered differ for every cell, so they can construct a fuller transcriptome by clustering together many single-cell transcriptomes (Waltman & van Eck, 2013). Indeed, similar transcriptome recovery was adequate to illustrate known biological processes in bacteria, as well as discover new behaviors (Kuchina et al., 2021).

3.3 | Yeast-optimized SPLiT-seq is truly single-celled and experiences a negligible percentage of barcode collisions

A common concern in single-cell RNA sequencing experiments is barcode collisions or one barcode mapping to multiple cells. In physical isolation methods, this occurs when more than one cell is loaded into the same container. In the SPLiT-seq method, multiple cells can acquire the same barcode by physically aggregating or through cells traveling the same path through the barcode plates by chance. We performed two experiments to get an empirical estimate of the percentage of barcodes that map to multiple cells or multiplets. We first processed two disparate species, diploid *S. cerevisiae* and *C. albicans*, mixed together in the same SPLiT-seq run. We could then calculate the percentage of barcode collisions by calculating the number of barcodes that have a large percentage (> 15%) of uniquely mapping transcripts that align to both species' genomes. We removed rRNA reads before this calculation as there are considerable amounts of homology between these species in those genes and keeping them in falsely increases the number of calculated barcode collisions.

Since many engineered yeast strains are haploid *S. cerevisiae*, we also wanted to ensure that these cells did not have any odd properties that caused them to stick together. We looked at stickiness between haploid *S. cerevisiae* by using two haploid strains each expressing a different fluorescent protein. A total of 2209 cells were detected that expressed at least one of these markers. We looked for collisions by counting the number of cells (barcodes) that report expression of both fluorescent markers. Figure 3f shows that in both experiments, the percentage of detected barcode collisions is well under five percent. We also see good species resolution in UMAP-based clustering analyses (Figure 3e). This suggests that our method is truly single celled.

3.4 | Yeast-optimized SPLiT-seq identifies heterogeneous transcriptional states in clonal populations grown in similar conditions

A good test of any single-celled method is whether it captures enough of the diversity of the transcriptome to resolve transcriptionally distinct states. The cells in two of our experiments ("experiments 1 and 2" from Figure 3) were grown in rich media and sampled in early mid-log phase. Even in this carbon-rich environment, we see considerable diversity in gene expression (Figure 4a). We performed the Louvain clustering algorithm on these cells, and the cells from Experiment 3 sampled at mid-log to group cells based on the gene expression data we collected and used UMAP to visualize the cell clusters. For each of these clusters, we performed a differential gene

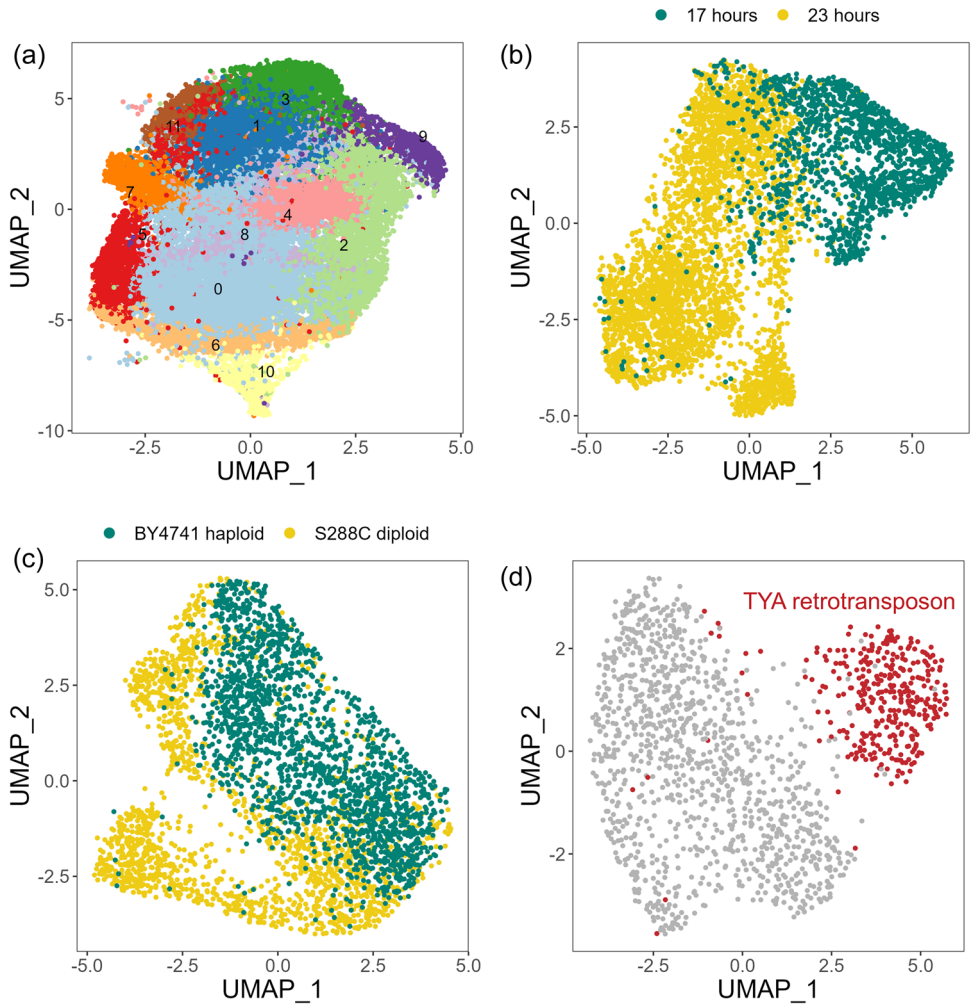

FIGURE 4 Yeast-optimized SPLiT-seq can discern biologically relevant cell states. (a) UMAP visual clustering of combined haploid and diploid *S. cerevisiae* cells from Experiments 1 and 2. The number-labeled, colored clusters represent clusters detected using a standard Louvain algorithm in Seurat (Kinsler et al., 2020). (b) UMAP visual clustering of *S. cerevisiae* cells from the first and last timepoints of the Experiment 3 growth curve corresponding to 17 and 23 h after inoculation. (c) UMAP visual clustering of haploid by 4741 and diploid s288c cells from the 17-h time point of Experiment 3. (d) UMAP visual clustering of the 23-h time point of Experiment 3, identifying a subset of cells expressing the TYA retrotransposon.

expression analysis to investigate what biological processes were distinguishing cells based on their gene ontology as assessed using Metascape and the Saccharomyces Genome Database (Cherry et al., 2012; Zhou et al., 2019). For example, cluster 2 appears to be cells that have very recently divided, as the top differentially expressed genes are related to septum digestion after cytokinesis. In this same cluster, we also observe several daughter-specific upregulated genes such as DSE1, DSE2, and DSE4, thus, this cluster may specifically tease out the newly divided daughter cells. Cells in other clusters have different unique transcriptional profiles. For example, cells in cluster 4 appear to be in late G1 or early S phase as we see upregulated expression of the G1 cyclin CLN2, genes involved in the formation of the bud neck such as HSL1 and GIN4, and genes related to DNA replication such as POL1 and RNR1. A full list of each cluster's statistically differentially expressed genes after multiple comparison correction is available in Supporting Information S1: Table 1. The observation that our single-cell transcriptomes cluster by cell-cycle state, combined with previous work demonstrating that single-cell RNA-seq often clusters cells by cell-cycle state, provides strong evidence that our yeast-optimized SPLiT-seq method is effective at capturing single-cell transcriptomes.

To provide further validation that our protocol adapting SPLiT-seq to yeasts captures biologically relevant details in single cells, we also endeavored to discern known information about cells via their transcriptomes. Cells in one of our experiments ("experiment 3" from Figure 3) consist of both haploid and diploid cells from samples across a standard growth curve in synthetic complete media. Cells sampled early have access to more glucose relative to those sampled later which are subsequently also exposed to more metabolic waste. Thus, there are likely differences in the transcriptomes of these populations. Previous work using SPLiT-seq demonstrates that cells cluster based on the time and cell density they were sampled across a

standard growth curve (Kuchina et al., 2021). Consistent with this expectation, using UMAP visualization, we can see separation of our yeast cells sampled from the earliest (17 h after inoculation) and latest timepoints (23 h after inoculation) in our growth curve in Figure 4b. Within the 17-h time point, we can also distinguish between the haploid and diploid cells (Figure 4c). This would not be possible unless our protocol optimizing SPLiT-seq for yeasts preserves the integrity of the transcriptomes sampled and the single-cell nature of the method.

Finally, we are able to discern something previously unknown about our yeast cells. Cells from the later timepoint in Experiment 3 (23 h after inoculation) contain a distinct group of cells that express the TYA retrotransposon (Figure 4d) with significantly upregulated YAR009C, YAR010C, YBR012W-A&B, YDR050C, YDR261C-D, YDR365W-A&B, YER138C, YJR029W, YKL096W-A, YOL103W-A&B, YOR142W-B, and YPR158C-D genes. Understanding the significance of this retrotransposon-expressing population is outside the scope of this work. Instead, we highlight this finding, and the others listed in this section, to validate our protocol for optimizing SPLiT-seq for the unique biology of yeasts. The clustering approaches performed in Figure 4 all provide evidence that our method is sampling the transcriptomes of individual cells with enough depth to cluster cells based on their transcriptionally distinct states.

4 | DISCUSSION

We demonstrate that SPLiT-seq can be used to investigate the transcriptomes of two important yeast species: *S. cerevisiae* and *C. albicans*. Though other methods exist for performing single-cell RNA sequencing in yeasts (Dohn, 2021; Jariani et al., 2020; Nadal-Ribelles et al., 2019; Urbonaite et al., 2021; Zheng et al., 2017) SPLiT-seq has several advantages. One major advantage is that SPLiT-seq enables multiplexing hundreds of genotypes and environments in a single experiment. It is thus better able to leverage the "awesome power of yeast genetics" to explore how genetic changes affect the transcriptome (Macreadie & Dhakal, 2022; Scannell et al., 2011). Another advantage of SPLiT-seq is that it does not require any expensive or specialized equipment to perform. Thus, SPLiT-seq brings the possibility of transcriptomics to a wider group of yeast researchers. Even researchers who have access to the specialized equipment necessary for droplet-based single-cell RNA sequencing may benefit from the yeast-optimized SPLiT-seq protocol described here. The in situ barcoding reactions that comprise SPLiT-seq have recently been leveraged to extend the scalability of droplet based RNA sequencing methods (Ma et al., 2023). Thus, we hope the yeast-optimized approach for SPLiT-seq described in this manuscript will be broadly useful to the community.

This method should be fairly straightforward to adapt for researchers studying yeasts other than those studied here, such as *S. pombe*, or flocculating natural isolates of *S. cerevisiae*. Zymolyase is not very effective against the *S. pombe* cell wall, but industrial enzymes can be used to permeabilize these cells (Molines, 2022). Additionally, flocculation is a common problem in beer and wine fermentation, and techniques exist to gently break the noncovalent bonds between cells immediately before fixation (Stahl et al., 1983; Verstrepen et al., 2003).

The number of mRNAs we recover using our yeast-optimized SPLiT-seq method (120–700 on average per cell), and the number of genes (100–600 on average per cell) is in the same ballpark but less than reported recovery for yeast experiments performed with the 10X genomics microfluidics-based platform (~2000 mRNAs per cell and ~700 genes per cell; Jackson et al., 2020). Nonetheless, there are several reasons our method is likely to be useful to the yeast community. First, an average of hundreds of unique reads per cell is enough for many applications. For example, we recover enough data to delineate biologically relevant cell states (Figure 4). Additionally, published studies detect less than 5% of each cell's transcriptome, but by clustering similar cells, can comprehensively map different transcriptional states (Kuchina et al., 2021) Second, we observe up to thousands of mRNAs in some cells, suggesting further optimization may yet yield further improvements. For example, in one experiment, where we used a reverse transcriptase with H minus activity and sequenced more deeply (Experiment 1 in Figure 3), we seemed to improve our reads per cell to approximately 10,000 (including rRNA) and our genes per cell to approximately 500 (Figure 3b,c). One obvious area for further optimization is changing the ratio of polydT to random hexamer primers, which would likely increase the yield of mRNA over rRNA (Kuchina et al., 2021). This might also be achieved by using a commercially available reagent to destroy rRNA before sequencing (O'Neil et al., 2013). And finally, we believe the highly scalable and massively multiplexable nature of the technique makes it an important addition to the yeast community.

5 | METHODS

5.1 | Experimental methods

5.1.1 | Yeast cell culture

Diploid s288c, genetically naive haploid BY4741, *MATa ura3A0 his3Δ1 met17Δ0 P ACT1-GAL3::SpHIS5 gal1Δgal10Δ::LEU2 leu2Δ0::P PGK1 -mCherry-KanMX6 ybr209wΔ::B103-HphMX6*, and *MATa ura3A0 his3Δ1 met17Δ0 P ACT1-GAL3::SpHIS5 gal1Δgal10Δ::LEU2 leu2Δ0::P GAL1-YFP-KanMX6 ybr209wΔ::BC3-HphMX6*, and ATCC 3147 *C. albicans* strains were used. Cells were streaked on YP plus 2% dextrose agar plates from frozen −80C glycerol stocks and grown at 30 C for 48 h. In Experiments 1 and 2, single colonies of s288c and ATCC 3147 were picked into YP plus 2% dextrose liquid media (YPD). The YFP and mCherry-engineered cells were picked into Synthetic Complete minus glucose (Sunrise) plus 2.5% sucrose, 1.25% raffinose, and 0.625% galactose media to induce YFP expression in Experiment 2. In Experiment 3, diploid s288c cells and haploid BY4741 cells were picked into Synthetic Complete (Sunrise) plus 2% glucose media (SCD). All cultures were grown with shaking at 30 C for 24 h. For

Experiments 1 and 2, cells were then transferred into fresh YPD media at a 1:250 dilution and grown until cultures reached approximately 2–4 × 10^7 cells/mL, or early mid-log phase. For Experiment 3, approximately 15,000 total cells were transferred into 50 mL of fresh SCD and sampled at 17, 19, 21, and 23 h after inoculation, which correspond to approximately 0.9 × 10^7, 2.7 × 10^7 × 10^7, and 8.6 × 10^7 cells/mL.

5.1.2 | Fixation and permeabilization

At the time of sampling, 3 mL of yeast cultures were immediately spun down in a room-temperature centrifuge at 3000*g* for 5 min. The media supernatant was removed, and the cell pellet was resuspended in 2 mL of cold 4% formaldehyde in molecular-grade phosphate-buffered saline (PBS). The cells were then fixed cold in a 4°C refrigerator for approximately 18 h. After fixation, the samples were spun down in a 4°C centrifuge at 3000*g* for 5 min. The formaldehyde supernatant was removed, and the pellet was resuspended in 1 mL cold molecular grade 100 mM Tris HCl pH 7 plus 0.1 U/μL SUPERase-In RNase inhibitor (Invitrogen) to quench the formaldehyde. Cells were then centrifuged, resuspended in 250 μL of a zymolyase enzymatic solution of 0.1 M Na_2EDTA, 1 M sorbitol, and 0.005 U/μL zymolyase (Zymo Research) (pH ~7.5), and incubated at 37°C for 15 min. At the end of the incubation, 1 mL of cold PBS plus 0.1 U/μL SUPERase-In RNase inhibitor and 0.1 U/μL Enzymatics RNase inhibitor (Enzymatics) was immediately added, and the cells were centrifuged at 4°C, 3000*g* for 5 min. The cells were then resuspended in 250 μL of cold 0.4% Triton X-100 in PBS plus RNase inhibitors and incubated on ice for 3 min. Again, 1 mL of cold PBS plus RNase inhibitors was added at the end of the 3 min and the cells were centrifuged. They were then resuspended in 500 μL of cold PBS plus RNase inhibitors, vortexed on high for approximately 30 s, and lightly filtered through a 40 μm pluriStrainer (pluriSelect). The cells were then counted using a Beckman Coulter Cell Counter and diluted to 1 million cells/mL into fresh cold PBS plus RNase inhibitors.

The additional conditions tried (Figure 3a, minus s12 described above) are outlined below:

s1: 2% formaldehyde overnight fixation with 0.01 U/μL zymolyase (37°C for 15 min).

s2: 2% formaldehyde overnight fixation with 0.01 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

s3: 2% formaldehyde overnight fixation with 0.005 U/μL zymolyase (37°C for 15 min).

s4: 2% formaldehyde overnight fixation with 0.005 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

s5: 2% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C overnight with 0.01 U/μL zymolyase (37°C for 15 min).

s6: 2% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C with 0.01 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

s7: 2% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C overnight with 0.005 U/μL zymolyase (37°C for 15 min).

s8: 2% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C with 0.005 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min),

s9: 4% formaldehyde overnight fixation with 0.01 U/μL zymolyase (37°C for 15 min).

s10: 4% formaldehyde overnight fixation with 0.01 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

s11: 4% formaldehyde overnight fixation with 0.005 U/μL zymolyase (37°C for 15 min).

s13: 2% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C overnight with 0.01 U/μL zymolyase (37°C for 15 min).

s14: 4% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C with 0.01 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

s15: 4% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C overnight with 0.005 U/μL zymolyase (37°C for 15 min).

s16: 4% formaldehyde for 15 min fixation stored in 70% EtOH at −20°C with 0.005 U/μL zymolyase (37°C for 15 min) and 0.4% Triton X-100 (4°C for 3 min).

In situ reverse transcription and ligation, post lysis processing, and library preparation were performed as described in previous work (Kuchina et al., 2021; Rosenberg et al., 2018) and are described in detail below.

5.1.3 | Reverse transcription

Reverse transcription (in situ) was performed in 25 μL reactions in 48 wells of a 100 μL 96-well plate with each well containing well-specific, barcoded primers. For each reaction, the final concentration of cells and reagents equated to 3 μM barcoded random hexamer primers (Supporting Information S2: Table 2), 3 μM barcoded 15-dT primers (Supporting Information S2: Tables 2), 0.25 U/μL Enzymatics RNase inhibitor, 0.25 U/μL SUPERase-In RNase inhibitor, 1 mM dNTPs (per base), 20 U/μL Maxima H minus Reverse Transcriptase (ThermoFisher), 1X RT buffer (ThermoFisher), 7.5% PEG6000 (formerly PEG8000), and 200,000 cells/mL (5 μL of the final preparation of 1 million cells/mL). These reactions were covered with an adhesive seal and placed in a standard thermocycler, and set to 23°C for 10 min followed by 50°C for 50 min. At the end of the thermal incubation, all 48 reactions were pooled with 9.6 μL of 10% Triton-X100 and centrifuged at 4°C, 3000*g* for 5 min and the supernatant was carefully removed from the cell pellet. It was then resuspended in 2 mL of cold PBS plus RNase inhibitors and vortexed on high for approximately 30 s. The 2 mL of cells were then filtered through a 15 μM pluriStrainer (pluriSelect).

5.1.4 | Ligations

Ligations (in situ) were performed in 50 and 60 μL reactions in a 100 μL 96-well plate with each well containing well-specific, barcoded ligation oligos. Before the experimental protocol, the round

2 and round 3 ligation plates were prepared by annealing the well-specific oligos with overhanging linker strands that guide the barcode oligos to the 5′ end of the singly barcoded cDNA (Supporting Information S2: Table 2, BC_0335 and BC_0284). In brief, for round 2, 12 µM barcode oligos and 11 µM linking strands per well were heated to 95 C in a 96-well plate and cooled to 20°C at a rate of 0.1°C/s. The same was performed for round 3, except 14 µM barcode oligos and 13 µM linking strands were used. 10 µL of these prepared annealed barcode plates were transferred to new 96-well plates for each experiment. A ligation mix was prepared containing the 2 mL of cells, and final concentrations of 8 U/µL T4 DNA Ligase (New England Biosciences), 1X Ligase Buffer (NEB), 0.32 U/µL Enzymatics RNase inhibitor, 0.32 U/µL SUPERase-In RNase inhibitor, and 7.5% PEG6000 in a total of 4.04 mL. 40 µL of this ligation mixture was added to the prepared round 2 barcode plate, covered with an adhesive seal, and incubated at 37°C for 30 min. A blocking strand solution was then prepared which preferentially binds to the linker strand, displacing it, and preventing future ligation reactions with the round two barcode oligos (Supporting Information S2: Table 2 BC_0340). This solution contained 26.4 µM blocking strand (BC_0340), 2.5X Ligase Buffer, and molecular grade water in 1200 µL. At the end of the round 2 ligation incubation, 10 µL of this blocking solution is added to each well, the plate was covered with an adhesive seal and then incubated at 37°C for a further 30 min. The ligation reactions were then pooled into a basin through a 40 µM pluriStrainer. An additional 20 µL of 2 M U/µL T4 DNA Ligase was also added to the basin and thoroughly mixed. 50 µL of this ligation mixture was then added to the round 3 barcode plate, covered with an adhesive seal, and incubated at 37°C for 30 min. A blocking solution of 11.5 µM BC_0066 (Supporting Information S2: Table 2), 125 mM EDTA, and molecular grade water in a total of 3200 µL was prepared, and 20 µL was added to each well of the round 3 plate at the end of the incubation. The EDTA immediately terminated the ligation reaction, so no further incubation was needed.

5.1.5 | Sublibrary generation and cell lysis

The final ligation mixture was pooled into a new basin and filtered through a 40 µM pluriStrainer into two 5 mL centrifuge tubes with 70 µL of 10% Triton-X100. The mixtures were then centrifuged at 4°C, 3000 G for 5 min. The supernatants of each tube were carefully aspirated, leaving a small volume in each since a pellet was not always visible. The pellets were then washed with a solution of 4 mL PBS, 40 µL 10% Triton-X100, and 10 µL of SUPERase-In RNase inhibitor, and consolidated into one 5 mL tube. The washed cell mixture was then centrifuged at 4 C, 3000 G for 5 min, and the supernatant carefully aspirated from the cell pellet. The pellet was then resuspended in a final 50 µL of PBS plus RNase inhibitors, and a small aliquot was counted using a Beckman Coulter Culture Counter. An appropriate volume of the cell suspension and PBS plus RNase inhibitors were mixed to make sublibraries ranging from approximately 5000–20,000 cells in 50 µL. These sublibraries were either then stored at −80°C or moved directly into the lysis step.

To lyse the cells, 60 µL of a solution of 20 mM Tris-HCl pH 8.0, 400 mM NaCl, 100 mM EDTA pH 8.0, 4.4% sodium dodecyl sulfateSDS, and 0.02 mg/mL proteinase K was added to each 50 µL sublibrary and incubated with shaking at 55°C for 2 h.

5.1.6 | Bead-bound template switch reaction

After lysis, 5 µL of either 100 µM PMSF or AEBSF was added to each sublibrary and incubated for 10 min at room temperature to inactivate the proteinase K. For each sublibrary, 44 µL of Dynabeads MyOne Streptavidin C1 (Invitrogen) were washed three times with 800 µL of a 1X wash solution of 5 mM Tris-HCl pH 8.0, 1 M NaCl, 500 µM EDTA, 0.05% Tween-20, and 0.1 U/µL SUPERase-In RNase inhibitor using a magnetic 1.5 mL tube rack. The beads were then resuspended in 100 µL per sample of a 2X wash solution containing 10 mM Tris-HCl pH 8.0, 2 M NaCl, 1 mM EDTA, and 0.2 U/µL SUPERase-In RNase inhibitor. 100 µL of this bead solution was added to each lysed sublibrary and agitated at room temperature for 1 h to bind the biotinylated, barcoded cDNA.

The samples were then placed in the magnetic rack and resuspended in 250 µL of the 1X wash solution and agitated at room temperature for 5 min, and repeated. After the two 1X wash steps, the samples were resuspended in 250 µL of 10 mM Tris-HCl pH 8.0, 0.1% Tween-20, and 0.1 U/µL SUPERase In RNase inhibitor. Following these three wash steps, the samples were rinsed with 250 µL of molecular-grade water while the beads were still bound to the magnetic rack.

For each sublibrary, the beads were then resuspended in 200 µL of a template switch reaction mix containing 10 U/µL Maxima H minus Reverse Transcriptase, 1X RT buffer, 1 mM dNTPs (per base), 0.5 U/µL SUPERase-In RNase inhibitor, 2.5 µM template switch oligo BC_0127 (Supporting Information S2: Tables 2), and 7.5% PEG6000. The samples were then agitated at room temperature for 30 min and then 42°C for 90 min. The tubes were then placed in a magnetic rack, and the supernatant removed. At this point, the beads could be rinsed with molecular-grade water and moved on to subsequent steps or resuspended in 250 µL of the Tris-HCL Tween-20 buffer and stored at 4°C overnight.

5.1.7 | cDNA amplification

If the samples were stored overnight in the Tris-Tween buffer, the tubes were placed in a magnetic rack and the beads were rinsed with 250 µL of molecular-grade water. The bead-bound cDNA were then amplified in 220 µL reactions with 2X high fidelity polymerase mix (KAPA HiFi, Q5) and 0.4 µM BC_0062 and BC_0108 primers (Supporting Information S2: Table 2) for 3 min at 95°C, and then five cycles of 98°C for 20 s, 65°C for 45 s, and 72 for 3 min. The bead PCR product was then placed against a magnetic rack and the

supernatant was transferred to new optical-grade PCR tubes with qPCR dye (EvaGreen 20X). The samples were then amplified on a qPCR machine for a further 10–20 cycles until the amplification curves exited log-linear phase. The PCR products were then cleaned using a 0.8X SPRI size selection and eluted in 20 µL of molecular-grade water. 5 µL were then run on a 1% agarose gel at 120 V for 15 min. A properly amplified cDNA library should appear as a smear starting at approximately 5–7 kB and ending at approximately 300 bp on a gel or bioanalyzer. 1 µL was also used to test the concentration on a qubit. A successful library will need a concentration of at least 1.5 ng/µL to have enough starting material for library preparation as described in these methods.

5.1.8 | WGS fragmentation and ligation library preparation

For each sublibrary, between 20 and 110 ng of amplified and cleaned cDNA products were combined with an appropriate amount of molecular-grade water to make 35 µL. This was then added to 5 µL of WGS Fragmentation buffer (Enzymatics) and 10 µL of WGS Fragmentation Enzyme mix (Enzymatics) on ice and pipette mixed. The fragmentation mix was placed into a chilled 4 C thermocycler and then incubated at 32°C for 10 min and 65°C for 30 min. The samples were then transferred back on ice, cleaned with a double-sided SPRI 0.6X-0.8X size selection, and then eluted in 50.5 µL of molecular grade water.

An adapter ligation reaction mix was made containing 17.5 µL of molecular grade water, 20 µL of WGS Ligation buffer (Enzymatics), 10 µL of WGS DNA Ligase (Enzymatics), and 2.5 µL of annealed adapter mix with BC_0243 and BC_0244 at 50 µM each (Supporting Information S2: Table 2) per sublibrary. 50 µL of the eluted fragmentation product was added to this mix and incubated at 20 C for 15 min. The ligated adapter product was then cleaned using a 0.8X SPRI size selection and eluted in 20 µL of molecular grade water.

To generate a final product ready for Illumina sequencing, 18.5 µL of the eluted adapter ligation product was PCR amplified using a 2X high fidelity polymerase mix (KAPA HiFi, Q5) and Illumina indexing adapter primers with either single 6 bp or dual 8 bp indices (Supporting Information S2: Table 2). The PCR was cycled for 8–11 cycles depending on the amount of starting cDNA added to the initial fragmentation step. The final product was cleaned using a double-sided 0.5X-0.7X SPRI size selection and eluted in 20 µL of molecular-grade water.

5.1.9 | Detailed cost breakdown

300 µL 100 µM prediluted plates for the round 1 reverse transcription and rounds 2 and 3 ligations were ordered from Integrated DNA Technologies (IDT) for $7699.40. These plates contain enough volume to complete 300 RT, 250 r2, and 215 r3 ligations with careful pipetting. Using the lowest capacity of the round 3 plate, this comes to $36 in barcoding oligos per protocol. The remaining stand-alone oligos were purchased as lyophilized DNA for $270.60 (Supporting Information S2: Table 2). The most expensive consumables in the protocol are the reverse transcription and ligation enzymes. Thermo Fisher Maxima H minus Reverse Transcriptase (EP0753) now sells for between $835 and $845 depending on the vendor. The New England Biolabs T4 DNA Ligase (M0202M) is currently listed at $270. We recognize that cheaper brand alternatives do exist at similar enzyme concentrations, we just have no personal verification that they work with our protocol. These reagents are enough for just over 1 experiment, but not enough for 2, so we calculate them to be repurchased for every instance of the protocol. The other large cost items are the RNase inhibitors and Dynabeads. Both RNAse inhibitors together cost approximately $630 and last for about 5 repetitions of the protocol, totaling $127 per run. The Dynabeads cost $638 per 2 mL, and 44 µL are needed per sublibrary sample, giving a cost-per-sample of $14. The amplified cDNA libraries can be prepared for sequencing however the user desires, but using our method with the WGS fragmentation and ligation kits from Enzymatics, it comes to $39 per sample. The other reagents and plastics are common to most molecular and microbiology labs, and inexpensive enough to be considered negligible to the total cost of the protocol. This gives a start-up cost of approximately $10,000, a per-protocol cost of approximately $1300–1700 including added leeway for the unpriced consumables, and an additional $55 per sublibrary.

Alternatively, kits containing all relevant regents for the steps from in situ reverse transcription to WGS library preparation are available through the company, Parse. However, their fixation and permeabilization steps are optimized for mammalian cells, so independent preparations like those described here will be needed to more successfully apply the kits to yeast.

Illumina sequencing libraries from each of the three experiments were submitted to Psomagen. Experiment 1 was paired end sequenced on a full lane of the HiSeq X platform with a six basepair index on read 2. Experiments 2 and 3 were also sequenced on full lanes of the HiSeq X platform but with dual 8 bp indices.

Oligonucleotides: All oligonucleotide sequences used are available in Supporting Information S2: Table 2.

5.2 | Computational methods

5.2.1 | Transcriptome alignment

Sequencing reads were processed by barcode and aligned to the R64 s288c genome or a combined R64 s288c and sc5314 V4 genome from NCBI using STARsolo (Kaminow et al., 2021). This new version of STAR parses out cell barcodes and allows for the multi-mapping of reads, meaning it keeps reads that map to multiple places in the genome. This is important for yeast as *S. cerevisiae* recently underwent a full genome duplication, and there is significant

homology between many paralogs (Wolfe, 2015). We used the most basic uniform multi-mapping algorithm. This produced a highly sparse gene-by-cell matrix with cell columns for every possible barcode combination, regardless of gene detection.

5.2.2 | Data quality thresholding

To remove empty barcodes and low gene detection barcodes, we applied a "knee" detection filtering to the STARsolo-generated gene-by-cell matrix (Kaminow et al., 2021). Briefly, the barcodes were negatively ordered by log(total read counts). We remove any barcodes after the curve begins to drastically decrease, or any barcodes past the bend or the "knee." Depending on the library quality, this keeps between 60% and 80% of the non-zero barcodes.

5.2.3 | Single-cell data analysis

All data analysis was performed with the R based package, Seurat (Hao et al., 2021). We performed normalization, scaling, nearest neighbor calculations, clustering, and differential expression analyses similar to the tutorial provided by the Satija Lab (Getting Started with Seurat, 2024). Unique transcripts/cell and genes/cell were calculated with all genes left in the data. However, mRNAs/cell, Louvain clustering, and UMAP visualization were done on data with the ribosomal RNA removed. rRNA genes were identified using the Saccharomyces Genome Database (SGD; Cherry et al., 2012). For the clustering analyses, we first normalized and scaled the data, and found the variable features on which to perform principal component analyses using the default Seurat settings. For the PCA, we computed an initial 20 components and then used the JackStraw function to calculate how many components significantly contributed to the variance for each data set. The downstream Louvain clustering and UMAP analyses were performed using this number of dimensions. All other parameters were kept at the Seurat defaults.

5.2.4 | Gene ontology analysis

Simple gene ontology analyses were done by manually entering genes into both Metascape and SGD (Cherry et al., 2012; Zhou et al., 2019).

AUTHOR CONTRIBUTIONS

Leandra Brettner and Kerry Geiler-Samerotte conceived the study and contributed to writing the manuscript. Leandra Brettner, Rachel Eder, and Kara Schmidlin performed the experiments.

ACKNOWLEDGMENTS

We would like to thank Yuichi Eguchi and Pak Poon for their helpful insights into yeast morphology, physiology, and culturing methods. We would also like to thank Georg Seelig and Anna Kuchina for feedback and troubleshooting on the SPLiT-seq method. Funding for this work was provided by a National Institutes of Health grant R35GM133674 (to KGS), an Alfred P Sloan Research Fellowship in Computational and Molecular Evolutionary Biology grant FG-2021-15705 (to KGS), and a National Science Foundation Biological Integration Institution grant 2119963 (to KGS). We would also like to thank the members of the Geiler-Samerotte Lab for their support and feedback.

DATA AVAILABILITY STATEMENT

Data processing and analysis scripts, as well as processed alignment data can be found on the Open Science Framework: https://osf.io/xh8s3/files/osfstorage. Both raw fastq files and alignment data can also be accessed through NCBI Geo GSE251966. The alignment data consists of barcode (cell)-by-gene count matrices that were created after first aligning all sequenced reads to the appropriate genome(s) using STARsolo, (Kaminow et al., 2021) and then assigning each read according to its barcode. These matrices are in their raw form without any filtering. All subsequent analyses performed in this publication, including quantifying the number of reads and genes per cell, and all clustering analyses are repeatable with these data. They are the basic data format that feed into single-cell RNA sequencing pipelines such as Seurat (Getting Started with Seurat) or Scanpy (Wolf et al., 2018). We also provide raw sequencing reads in the form of fastq files. There are multiple fast files pertaining to each of the 3 experiments because we sequenced multiple redundant sub-libraries per each experiment. Due to an unforeseen data loss, 3/14 original fastq files were truncated, resulting in a minor data loss (5.7% of total reads).

REFERENCES

10X Genomics single-cell libraries. (2024). | Cornell Institute of Biotechnology | Cornell University. https://www.biotech.cornell.edu/core-facilities-brc/services/10x-genomics-single-cell-libraries

10X Pricing. (2024). https://www.kumc.edu/research/genome-sequencing-facility/services/10x-services/10x-pricing.html

Bertin, B., Broux, O., & Van Hoegaerden, M. (1990). Flow cytometric detection of yeast by in situ hybridization with a fluorescent ribosomal RNA probe. *Journal of Microbiological Methods*, *12*, 1–12.

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S. N., Steemers, F. J., Adey, A., Waterston, R. H., Trapnell, C., & Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, *357*, 661–667.

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., … Wong, E. D. (2012). Saccharomyces genome database: The genomics resource of budding yeast. *Nucleic Acids Research*, *40*, D700–D705.

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z. Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., … Boone, C. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, *353*(6306), aaf1420.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S.,

Friedman, N., & Regev, A. (2016). Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, *167*, 1853–1866.

Dohn, R., Xie, B., Back, R., Selewa, A., Eckart, H., Rao, R. P., & Basu, A. (2021). mDrop-seq: Massively parallel single-cell RNA-seq of Saccharomyces cerevisiae and Candida albicans. *Vaccines*, *10*(1), 30. https://www.biorxiv.org/content/10.1101/2021.07.06.451338v1

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., & Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental. *Molecular Biology of the Cell*, *11*, 4241–4257.

Getting Started with Seurat. (2024). https://satijalab.org/seurat/articles/get_started.html

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., … Johnston, M. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, *418*, 387–391.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*, 3573–3587.

Jackson, C. A., Castro, D. M., Saldi, G.-A., Bonneau, R., & Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *eLife*, *9*, e51254.

Jariani, A., Vermeersch, L., Cerulus, B., Perez-Samper, G., Voordeckers, K., Van Brussel, T., Thienpont, B., Lambrechts, D., & Verstrepen, K. J. (2020). A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. *eLife*, *9*, e55320.

Kaminow, B., Yunusov, D., & Dobin, A. (2021). STARsolo: Accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. bioRXiv. Advance published online. https://doi.org/10.1101/2021.05.05.442755

Kinsler, G., Geiler-Samerotte, K., & Petrov, D. A. (2020). Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*, *9*, e61271.

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Molecular Cell*, *58*, 610–620.

Kuchina, A., Brettner, L. M., Paleologu, L., Roco, C. M., Rosenberg, A. B., Carignano, A., Kibler, R., Hirano, M., DePaolo, R. W., & Seelig, G. (2021). Microbial single-cell RNA sequencing by split-pool barcoding. *Science*, *371*(6531), eaba5257.

Levy, S. F., Blundell, J. R., Venkataram, S., Petrov, D. A., Fisher, D. S., & Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, *519*, 181–186.

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen, A. N., Jones, M., … Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*, 337–341.

Liu, S., & Trapnell, C. (2016). Single-cell transcriptome sequencing: Recent advances and remaining challenges. *F1000Research*, *5*, 182.

Ma, P., Amemiya, H. M., He, L. L., Gandhi, S. J., Nicol, R., Bhattacharyya, R. P., Smillie, C. S., & Hung, D. T. (2023). Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell*, *186*, 877–891.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual. *Cell*, *161*, 1202–1214.

Macreadie, I., & Dhakal, S. (2022). 'The awesome power of yeast'. *Microbiology Australia*, *43*, 19–21.

McNulty, R., Sritharan, D., Liu, S., Hormoz, S., & Rosenthal, A. Z. (2021). Droplet-based single cell RNA sequencing of bacteria identifies known and previously unseen cellular states. *BioRxiv*. Advance published online. https://doi.org/10.1101/2021.03.10.434868

Miura, F., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y., & Ito, T. (2008). Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics*, *9*, 574.

Molines, A. T., Lemière, J., Gazzola, M., Steinmark, I. E., Edrington, C. H., Hsu, C. T., Real-Calderon, P., Suhling, K., Goshima, G., Holt, L. J., Thery, M., Brouhard, G. J., & Chang, F. (2022). Physical properties of the cytoplasm modulate the rates of microtubule polymerization and depolymerization. *Developmental Cell*, *57*, 466–479.

Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P., Nguyen, M., de Nadal, E., Posas, F., & Steinmetz, L. M. (2019). Sensitive high-throughput single-cell RNA-Seq reveals within-clonal transcript-correlations in yeast populations. *Nature Microbiology*, *4*, 683–692.

O'Neil, D., Glowatz, H., & Schlumpberger, M. (2013). Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Current Protocol Molecular Biology*, *103*, 4.19.1–4.19.8.

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., & Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, *556*, 339–344.

Pricing & Quoting – 10x Genomics. (2024). https://store.10xgenomics.com/build?stytch_token_type=magic_links&token=PQm56V8faR9gSVsaFGCR4YeoD_HWvdeb0cavKNJkoNW4

Pricing. (2024). | Single Cell Sequencing Core. https://www.bumc.bu.edu/singlecell/pricing/

Rodriguez-Fraticelli, A. E., Weinreb, C., Wang, S. W., Migueles, R. P., Jankovic, M., Usart, M., Klein, A. M., Lowell, S., & Camargo, F. D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature*, *583*, 585–589.

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L. T., Peeler, D. J., Mukherjee, S., Chen, W., Pun, S. H., Sellers, D. L., Tasic, B., & Seelig, G. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, *360*, 176–182. https://doi.org/10.1126/science.aam8999

Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., Rine, J., Johnston, M., & Hittinger, C. T. (2011). The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces* sensu stricto. *G3&58; Genes|Genomes|Genetics*, *1*, 11–25.

Sharon, E., Chen, S. A. A., Khosla, N. M., Smith, J. D., Pritchard, J. K., & Fraser, H. B. (2018). Functional genetic variants revealed by massively parallel precise genome editing. *Cell*, *175*, 544–557.

Stahl, U., Kues, U., & Esser, K. (1983). Flocculation in yeast, an assay on the inhibition of cell aggregation. *European Journal of Applied Microbiology and Biotechnology*, *17*, 199–202.

Urbonaite, G., Lee, J. T. H., Liu, P., Parada, G. E., Hemberg, M., & Acar, M. (2021). A yeast-optimized single-cell transcriptomics platform elucidates how mycophenolic acid and guanine alter global mRNA levels. *Communications Biology*, *4*, 822.

Verstrepen, K. J., Derdelinckx, G., Verachtert, H., & Delvaux, F. R. (2003). Yeast flocculation: what brewers should know. *Applied Microbiology and Biotechnology*, *61*, 197–205.

Waltman, L., & van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B: Condensed Matter and Complex Systems*, *86*, 471.

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, *24*, 437–440.

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, *19*, 15.

Wolfe, K. H. (2015). Origin of the yeast whole-genome duplication. *PLoS Biology*, *13*, e1002221.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*, 1523.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Brettner, L., Eder, R., Schmidlin, K., & Geiler-Samerotte, K. (2024). An ultra high-throughput, massively multiplexable, single-cell RNA-seq platform in yeasts. *Yeast*, *41*, 242–255. https://doi.org/10.1002/yea.3927