Defining the role of the polyasparagine repeat domain of the *S. cerevisiae* transcription factor Azf1p

Taylor Stewart¹, Benjamin E. Wolfe¹, and Stephen M. Fuchs^{1,2,3*}

¹Department of Biology, Tufts University, Medford, MA 02155, USA

²Institute for Protein Innovation, Boston, MA 02115, USA

³Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA

* To whom correspondence should be addressed. Email: stephen.fuchs@proteininnovation.org

ABSTRACT

Across eukaryotes, homopolymeric repeats of amino acids are enriched in regulatory proteins such as transcription factors and chromatin remodelers. These domains play important roles in signaling, binding, prion formation, and functional phase separation. Azf1p is a prion-forming yeast transcription factor that contains two homorepeat domains, a polyglutamine and a polyasparagine domain. In this work, we report a new phenotype for Azf1p and identify a large set of genes that are regulated by Azf1p during growth in glucose. We show that the polyasparagine (polyN) domain plays a subtle role in transcription but is dispensable for Azf1p localization and prion formation. Genes upregulated upon deletion of the polyN domain are enriched in functions related to carbon metabolism and storage. This domain may therefore be a useful target for engineering yeast strains for fermentation applications and small molecule production. We also report that both the polyasparagine and polyglutamine domains vary in length across strains of *S. cerevisiae* and propose a model for how this variation may impact protein function.

INTRODUCTION

Repeat domains are prevalent in the eukaryotic proteome and are found in 13% of proteins in yeast, 17% in humans, and 21% in *Drosophila*. Repeats are commonly found within intrinsically disordered regions (IDRs), which are protein domains that do not have a single native fold corresponding to a lowest energy state. Instead, these domains dynamically sample a number of conformations with similar energies (Van Der Lee *et al.* 2014). Despite being historically dismissed as unimportant due to their lack of three-dimensional structure, repetitive protein domains are essential for binding, signaling, and functional phase separation. Across eukaryotes, repeats are enriched in cell-surface and regulatory proteins, such as transcription factors (TFs) and chromatin remodelers (Gemayel *et al.* 2010). Repeats enriched in a single amino acid are deemed homorepeats. Polyglutamine (polyQ) is the most common type of homorepeat in eukaryotic proteins, followed by polyasparagine (polyN). While polyQ domains are found across eukaryotes, polyN domains are rare in vertebrates and found almost exclusively in more primitive invertebrates (Faux *et al.* 2005). Synthetic polyN peptides have been shown to mediate aggregation (Peters and Huang 2007; Lu and Murphy 2014), but the role of natural polyN domains in protein function has only been explored in a few cases. For example, the polyN domain of Pdr1p has been shown to be important for transcriptional activation (Kolaczowska *et al.* 2002). Whether or not this function expands to other polyN-containing TFs, such as *CCR4*, *MIT1*, *ASG1*, *HAP4*, *ABF1*, *MOT3*, *AZF1*, and *RLM1*, has not been investigated.

Several polyQ and polyN-containing TFs in yeast (*HAP4*, *AZF1*, *MOT3*, *RLM1*, *CYC8*, *MGA1*, and *SWI1*) have been demonstrated to form prions. Prions are proteins that have multiple conformations, at least one of which is self-templating and epigenetically heritable. Prion-forming domains (PrDs) are characterized by low-complexity amino acid sequences that are often rich in N and Q, insensitive to scrambling, and intrinsically disordered in their native state (Cascarina and Ross 2014). Many prion-forming proteins contain repeats that are important to the function of both the prion and the native protein. Although there is only one bona fide prion in mammals, PrP^{Sc}, screens to identify yeast prions have revealed that they are

common in both lab and wild strains of yeast (Alberti *et al.* 2009; Halfmann *et al.* 2012; Chakrabortee *et al.* 2016; Itakura *et al.* 2020). While the mammalian PrP^{Sc} and prion-like proteins are well documented to cause both infectious diseases and age-related neurodegeneration in mammals (Prusiner 1982; Murakami *et al.* 2015; Patel *et al.* 2015; Collinge 2016; Harrison and Shorter 2017; Paul *et al.* 2017; Guo *et al.* 2018), there is mounting evidence that yeast prions act as adaptive strategies, especially during environmental stress. Yeast prions are well-suited to mediate adaptation due to their enrichment in TFs and other regulatory proteins (Tuite and Serio 2010; Halfmann and Lindquist 2010; Newby and Lindquist 2013; Holmes *et al.* 2013; Jarosz *et al.* 2014a, 2014b; Chakrabortee *et al.* 2016; March *et al.* 2016; Chernova *et al.* 2017; Harvey *et al.* 2018; Franzmann and Alberti 2019; Chakravarty *et al.* 2020; Itakura *et al.* 2020). A screen conducted by Chakrabortee and colleagues identified nearly 50 novel prion-like proteins in yeast that conferred epigenetically heritable phenotypes after transient overexpression. In many cases, the prion state improved yeast growth rate in different environmental stresses, including osmotic and acidic stress. Like previously identified prions, many of the proteins identified in this screen are TFs and RNA-binding proteins (Chakrabortee *et al.* 2016).

Prion-forming TFs have the capacity to impact yeast growth through multiple mechanisms. These proteins can alter gene expression in both their native and prion states, leading to changes cellular functions such as metabolism, surface display, and stress tolerance. These proteins are therefore important targets for yeast metabolic engineering. Yeast are used in a wide range of applications in research, food and drink production, and small molecule production. Because homorepeat domains are at the intersection of TFs and prions, these domains are particularly interesting targets for engineering. Target genes of Q-rich TFs in yeast have significantly higher levels of expression divergence, expression variability, mutational variance, and expression noise when compared to targets of non-repeat-containing TFs (Gemayel *et al.* 2015). These results highlight the importance of repeat domains in modulating gene expression.

The yeast TF Azf1p was recently shown to form a prion (Chakrabortee *et al.* 2016). Azf1p has a largely nonoverlapping set of target genes during growth in fermentable and non-

fermentable carbon sources, and the null mutant has been shown to have a severe growth defect in glycerol. *AZF1* is also important for cell wall integrity (Slattery *et al.* 2006). The prion conformer, designated [*AZF1*⁺], confers resistance to the drug radicicol in a gain-of-function manner but decreases the expression of Azf1p's target genes (Chakrabortee *et al.* 2016). Azf1p contains a polyN and polyQ domain, and the frequency of homorepeats in yeast TFs and PrDs predicts a role for these domains in transcription as well as prion formation. In this work, we investigated the function of the polyN domain of Azf1p and find that this domain plays a subtle role in transcription but is not required for Azf1p localization or prion formation.

METHODS

Yeast Strains and Plasmids—Plasmids used in this work are summarized in Table S1. The ΔN mutant plasmid was constructed by QuikChange site-directed mutagenesis using pBY011-*AZF1* (Harvard PlasmID) as a template with the primers described in Table S2. Yeast were grown on synthetic complete (SC) dropout medium, YPD (yeast extract, peptone, 2% dextrose), or YPG (yeast extract, peptone, 4% glycerol) at 30°C. Plasmids were freshly transformed into yeast and maintained on SC media lacking uracil (SC-Ura). Strains used in this study are summarized in Table S3 and were derived from BY4741 (*MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0*). The WT strain is from the yeast GFP collection in which the *AZF1* open reading frame is C-terminally tagged with GFP (Huh *et al.* 2003; Howson *et al.* 2005). *AZF1 Δ N* was constructed in this background using CRISPR (Laughery *et al.* 2015). A guide RNA sequence designed to target the polyQ domain of *AZF1* was cloned into the Cas9-expressing plasmid pML104. A ~1 kb fragment of *AZF1* containing the polyN deletion was amplified by PCR with 8 identical reactions using the *AZF1* F/R primers (Table S2) and pBY011- ΔN as a template. Ethanol-precipitated PCR products were co-transformed with pML104-polyQ, and cells were plated on SC-Ura. After 5 days, individual transformants were selected and grown overnight in 1 mL YPD in a 96 deep-well plate to allow for loss of the Cas9 plasmid. Overnight cultures were diluted ten-fold four times and spotted on FOA. Fast-growing FOA-resistant colonies were selected and characterized by

colony PCR with the *AZF1* F/R primers to identify yeast with the polyN deletion. This strain is referred to as ΔN in the text. *azf1- $\Delta 1$::URA3* was constructed in BY4741 using heterologous gene replacement with the *URA3* gene as a selectable marker. This strain was intended to be a full deletion of the *AZF1* open reading frame; however, *URA3* is instead integrated between *AZF1* and its promoter, abolishing *AZF1* expression (Table S4). In the text, this strain is referred to as *azf1 Δ* because it mimics the phenotypes associated with deletion of the *AZF1* open reading frame.

Spotting Assays—For phenotypic growth assays, yeast were grown overnight in YPD. Saturated overnight cultures were used to start fresh cultures in the same medium at an OD₆₀₀ of 0.2. The cells were allowed to double at least two times before approximately 1.0×10^7 cells were harvested and resuspended in sterile water in a 96-well plate. Cells were serially diluted five-fold five times and then spotted onto YPD or YPG plates using a 48-pin replicating tool. Plates were incubated at 30°C and imaged after three (YPD) to seven (YPG) days.

Growth Curve Assays—Each pBY011-*AZF1* plasmid (WT and ΔN) was transformed into the matched background and maintained on SC-Ura (*i.e.* the ΔN plasmid was transformed into *AZF1 ΔN*). Three biological replicates of each strain were grown overnight in SC_{Raff}-Ura (2% raffinose). 2 μ L of each culture were then transferred to 200 μ L of four types of fresh media in a 96-well plate: SC_{Gal}-Ura (2% galactose), SC_{Gal}-Ura+Radicicol (75 μ M), SC-URA, and SC-Ura+Radicicol. Growth was measured for 24 hours by OD₆₀₀ in a Molecular Devices SpectraMax M5 plate reader at 30°C with shaking prior to each reading. Readings were taken every 30 minutes. 10 μ L of each SC_{Gal}-Ura and SC-Ura culture were then transferred to 1 mL fresh SC-URA and allowed to grow to saturation overnight. The plate reader assay was then repeated in SC-URA media with and without radicicol. We compared the growth of cells whose ancestors had experienced *AZF1* overexpression to those that did not in order to determine the role of the polyN domain in [*AZF1*⁺] prion formation and activity.

A modified version of this plate reader assay was also performed with the WT, ΔN , and *azf1 Δ* strains without the galactose-inducible plasmids. Three biological replicates were grown

overnight in YPD. 2 μ L of each saturated culture were then transferred to 200 μ L of SC media, and growth was measured for 24 hours in the plate reader.

RNA-seq—Three biological replicates were grown overnight in either YPD or YPG. Saturated overnight cultures were diluted to OD₆₀₀ 0.2. Cells were grown to OD₆₀₀ 0.6-1.0 and collected by centrifugation. Total RNA was isolated using the GE Healthcare RNeasy Mini RNA Isolation Kit per the manufacturer's guidelines and was quantified by NanoDrop (Thermo Fisher Scientific). mRNA was isolated using a NEBNext poly(A) mRNA magnetic isolation module (New England Biolabs). mRNA was used to generate RNA-seq libraries using a NEBNext Ultra II RNA library prep kit for Illumina (New England Biolabs) following the manufacturer's recommended protocol with the following parameters: 15 minutes at 42°C for the RNA fragmentation time, a 1:25 dilution of the adapter, no size selection, and 15 cycles of denaturation / annealing / extension during the PCR enrichment step. After quantifying the library concentration with a Qubit HS DNA assay and determining the average peak size using a Fragment Analyzer, equimolar concentrations of each library were pooled together and sequenced on an Illumina NextSeq 550 at the Tufts University Genomics Core using a high output run with 150 cycles.

Data analysis was performed using Geneious Prime v. 2020.2.4 (Biomatters Ltd.). Sequencing reads were aligned to the s288c genome using the Geneious RNA Mapper with Medium-Low Sensitivity / Fast settings. Multiple best matches were mapped randomly. Expression levels were then calculated, and ambiguously mapped reads were counted as multiple full matches. To identify differentially expressed genes, expression levels were compared between strains in each carbon source using DESeq2 with a parametric fit type. Expression levels were also compared between carbon sources for each strain. Genes that are statistically significantly differentially expressed ($p < 0.001$) in *azf1Δ* compared to WT were selected for GO Enrichment Analysis. The genes in this subset all have log₂ ratios with a magnitude of 0.30 or greater. Because very few genes are statistically significantly differentially expressed in Δ N, genes with log₂ fold ratios of ± 0.30 or more were selected for analysis. GO Enrichment Analysis was performed using PANTHER v. 14 (Ashburner *et al.* 2000; Carbon *et al.*

2019; Mi *et al.* 2019). The PANTHER Overrepresentation Test (Released 20200728) employs a Fisher's exact test and reports significant enrichment in biological processes ($p < 0.05$).

Target Gene Expression Measurements by RT-qPCR—RNA was isolated as described above, and cDNA synthesis was performed using the SuperScript™ First-Strand Synthesis System (Invitrogen) following the manufacturer's guidelines with an oligo-dT(20) primer and 300 ng of RNA per sample. cDNA was diluted 1:5 and stored at -20°C . Each 20 μL reaction was prepared using the Brilliant II SYBR Green QPCR system (Agilent) following the manufacturer's guidelines with the primers listed in Table S2. Samples were contained in MicroAmp 96-Well Reaction Plates (Applied Biosystems) sealed with MicroAmp Optical Adhesive Film (Applied Biosystems). Amplification was performed using an Applied Biosystems 7300 Real-Time PCR System with the following cycling conditions: 5 minutes at 95°C and 40 cycles of 10 seconds at 95°C followed by 30 seconds at 55°C . The reference gene ACT1 was used for relative quantification. Gene expression changes were calculated using the $2^{-\Delta\Delta\text{Ct}}$ method. Two independent experiments were performed with three technical replicates for each biological replicate.

Microscopy—Cells in which *AZF1* is tagged with GFP were imaged on a Leica Thunder Imager in the log and stationary phases of growth with either glucose or glycerol as the carbon source. WT and ΔN cells were grown overnight in YPD or YPG. 100 μL of saturated YPD overnight cultures were spun down and imaged to study cells in stationary phase with glucose as the carbon source. Saturated overnight YPD cultures were also diluted to an OD_{600} of 0.2 and grown to mid-log phase. 1.0 mL of these cultures was spun down and imaged to study cells growing in log phase with glucose as the carbon source. YPG cultures reached mid-log phase after about 24 hours. 1.0 mL of these cultures were spun down and imaged to study cells growing in log phase with glycerol as the carbon source. YPG cultures were also allowed to grow for 48 hours, at which time 100 μL were spun down and imaged to study cells growing in stationary phase with glycerol as the carbon source. Each condition was imaged two independent times, and at least 50 cells were imaged for each strain in each condition. Images were processed using Leica LAS X Instant Computational Clearing.

Evaluation of AZF1 repeat variation—Variation within *AZF1* was measured using the data available from 93 recently sequenced *S. cerevisiae* genomes (Strope *et al.* 2015) as described in Babokhov *et al.* 2018.

Data Availability—Strains and plasmids are available upon request. Supplemental files are available online.

RESULTS

Azf1p is a prion-forming yeast TF that has been shown to activate genes involved in carbon metabolism and energy production during growth in glucose (Slattery *et al.* 2006; Chakrabortee *et al.* 2016). Previous work on Azf1p has been limited, so we began by monitoring the activity of *azf1Δ* in different carbon sources. A null mutant has previously been shown by spotting assay to produce a severe growth defect on non-fermentable carbon sources (Slattery *et al.* 2006). Slattery and colleagues also investigated the impact of the null mutant on growth on glucose by spotting assay, which did not reveal any differences between *azf1Δ* and the WT. We sought to further explore the role of Azf1p in glucose by monitoring growth in liquid media by optical density in a plate reader. We find that the null mutant has a shortened lag phase and reaches the log phase of growth more quickly when compared to the parent strain (Figure 1A). To further investigate the role of Azf1p during growth in glucose, we performed RNA-seq with RNA isolated from *azf1Δ* and WT cells grown in glucose and found many genes that are differentially expressed in the null mutant (Figure 1B, Figure S1, Table S5).

Figure 1. *azf1Δ* shortens lag phase during growth in glucose. A) Growth of WT and *azf1Δ* with glucose as the carbon source. Growth was measured by OD₆₀₀ in a plate reader for 24 hours. The first 8 hours are highlighted to show the differences in growth more clearly and due to the reduced accuracy of the plate reader at high optical densities. *azf1Δ* has a dramatically shorter lag phase than WT. Growth curves are the average of three biological replicates, and error bars

indicate standard deviation. B) Volcano plot showing genes that are differentially expressed in *azf1Δ* compared to WT.

Genes downregulated by *azf1Δ* have not been previously reported, and we hypothesized that Azf1p may act as a co-repressor for these genes. The general co-repressor Cyc8p contains two Q-rich domains, one of which has been shown to mediate binding with cofactors (Gemayel *et al.* 2015). Azf1p contains a polyQ and polyN domain (Figure 2A), and we hypothesized that one or both of these domains may be important for transcriptional repression by Azf1p. We first deleted the polyN domain using CRISPR (Figure 2A). We attempted to delete the polyQ domain using the same approach; however, deletion of the polyQ domain using this strategy was never successful. In an effort to probe the role of the polyQ domain, we deleted this domain from diploid BY4743. While the heterozygous mutant was viable, the strain did not sporulate. Null mutants are known to be sporulation deficient (Slattery *et al.* 2006), and the sporulation deficiency observed in the heterozygous mutant suggests a fundamental role for the polyQ domain in Azf1p function.

Figure 2. The polyN domain of Azf1p impacts growth in different carbon sources. A) Schematic of WT and ΔN Azf1p. Azf1p contains a polyQ (teal), polyN (orange), and a zinc-finger DNA-binding domain (ZnF) (green). To probe the role of the polyN domain, we deleted this domain in *AZF1-GFP* (Huh *et al.* 2003; Howson *et al.* 2005). B) Spotting assay measuring the growth of WT, ΔN , and *azf1Δ* on glucose and glycerol. On glycerol, deletion of the polyN domain produces a mild growth defect, while *azf1Δ* produces a severe growth defect. Images are representative of 3 independent experiments. C) Growth of WT, ΔN , and *azf1Δ* with glucose as the carbon source. Growth was measured by OD₆₀₀ in a plate reader for 24 hours. The first 8 hours are highlighted to show the differences in growth more clearly and due to the reduced accuracy of the plate reader at high optical densities. Deletion of the polyN domain moderately shortens the lag phase, and *azf1Δ* dramatically shortens the lag phase compared to WT. Growth curves are the average of three biological replicates, and error bars indicate standard deviation.

Differential roles of the polyN domain during growth in different carbon sources

PolyQ domains have been extensively studied due to their involvement in a number of mammalian neurodegenerative diseases. Despite being the second most common type of homorepeat in the eukaryotic proteome, the function of polyN repeats has been

underexplored. We were therefore interested in characterizing the role of the polyN domain of Azf1p. The severe growth defect in glycerol media produced by the null mutant is one of Azf1p's most well-characterized phenotypes. We therefore measured the impact of deleting the polyN domain on growth in glycerol by spotting assay (Figure 2B). Deletion of the polyN domain produces a moderate growth defect on glycerol when compared to the parent strain. Consistent with previous reports, *azf1Δ* does not support growth on glycerol. WT, ΔN, and *azf1Δ* appear to grow similarly on glucose, which is consistent with the finding by Slattery and colleagues that *AZF1* is not essential for growth in fermentable carbon sources. In order to detect any more subtle differences in growth between these strains in glucose, we monitored growth for 24 hours by optical density in a plate reader (Figure 1C). We find that deletion of the polyN domain also reduces the length of the lag phase, although not to the same extent as the null mutant. This suggests that the polyN domain plays a fundamental role in Azf1p function, although deletion of this domain does not completely abolish Azf1p activity.

PolyN domains are prevalent in yeast TFs, but their function has only been explored in a few cases. In order to evaluate the role of the polyN domain in transcription, we performed RNA-seq using RNA isolated from WT and ΔN cells grown in both glucose and glycerol. We find that deletion of the polyN domain alters gene expression in both glucose (Figure 3A, Table S6) and glycerol (Figure 3B, Table S7) compared to WT. Notably, there are more statistically significantly differentially expressed genes in ΔN during growth in glucose compared to glycerol. The number of differentially expressed genes and the magnitudes of these changes, however, are much less than those observed in the null mutant, which indicates that the polyN domain plays a subtle Azf1p function.

In order to test the validity of the RNA-seq experiment, we measured relative gene expression of several of the most differentially expressed genes from each carbon source, as well as two known Azf1p target genes (*MDH2* and *GAS1*), by qPCR. Overall, the relative gene expression trends observed in ΔN and *azf1Δ* compared to WT are consistent with those measured by RNA-seq (Figure S2).

Figure 3. The polyN domain of Azf1p plays a subtle role in transcription. A) Volcano plot showing genes that are differentially expressed in ΔN compared to WT in glucose. B) Volcano

plot showing genes that are differentially expressed in ΔN compared to WT in glycerol. C) GO Enrichment Analysis of upregulated genes with a log2 ratio of 0.30 or more in ΔN in glucose compared to WT. Fold enrichment is indicated to the right of each bar. Processes related to environmental stress tolerance are shown in blue, and processes related to carbon metabolism and storage are shown in green. D) GO Enrichment Analysis of upregulated genes with a log2 ratio 0.30 or higher in ΔN in glycerol compared to WT. Fold enrichment is indicated to the right of each bar. Processes related to environmental stress tolerance are shown in blue, and processes related to carbon metabolism and storage are shown in green. D) GO Enrichment Analysis of upregulated genes with a log2 ratio -0.30 or lower in ΔN in glycerol compared to WT. rRNA processing is the only process for which this set of genes is statistically significantly enriched (4-fold enrichment).

In order to gain a better understanding of how the transcriptional changes in ΔN underlie the growth phenotypes observed in Figures 2B and 2C, we characterized differentially expressed genes with a log2 ratio of 0.30 or higher for upregulated genes and -0.30 or lower for downregulated genes by GO Enrichment Analysis (Ashburner *et al.* 2000; Carbon *et al.* 2019; Mi *et al.* 2019). Genes upregulated by ΔN during growth in glucose are enriched in two main biological processes: environmental stress tolerance and carbon metabolism and storage (Figure 3C). There are only 3 genes downregulated -0.30 -fold or more during growth in glucose. As a result, there is no statistically significant biological process enrichment. Out of 119 genes upregulated 0.30-fold or higher in ΔN during growth in glycerol, 5 are involved in gluconeogenesis (17-fold enrichment), 8 are involved in glycolysis (15-fold enrichment), and 13 are involved in cell wall organization (3-fold enrichment). Out of 56 genes downregulated -0.30 -fold or more in glycerol, 12 are involved in rRNA processing (4-fold enrichment) (Figure 3E).

The polyN domain is dispensable for Azf1p localization

Azf1p has previously been shown to be localized to the nucleus during growth in glucose in both log and stationary phase, consistent with its function as a TF. Azf1p is also nuclear during log phase growth in glycerol, but relocalizes to cytoplasmic foci during stationary phase growth in glycerol (Slattery *et al.* 2006). Because polyN domains have previously been shown to promote protein aggregation (Peters and Huang 2007; Lu and Murphy 2014), we hypothesized

that the polyN domain is important for the formation of these foci. During growth in glucose, we find that both WT and ΔN localize to the nucleus (Figure 4A). In glycerol, both WT and ΔN localize to the nucleus during log phase, and both form cytoplasmic foci during stationary phase growth (Figure 4B). We therefore conclude that the polyN domain is not required for Azf1p relocation during stationary phase growth in glycerol.

Figure 4. Role of the polyN domain in localization of Azf1p. A) Representative images of localization of Azf1p in glucose during growth in log and stationary phase. During growth in glucose, Azf1p is localized to the nucleus. B) Representative images of localization of Azf1p during growth in glycerol in log and stationary phase. During log phase, Azf1p is localized to the nucleus. During stationary phase, both Azf1p and Azf1- ΔN p form cytoplasmic foci. At least 50 cells were imaged for each strain in each condition.

[AZF1⁺] prion formation does not require the polyN domain

Azf1p was recently shown to form a prion, designated [AZF1⁺], that confers resistance to the drug radicicol. Both the polyQ and polyN domains are predicted to be capable of promoting prion formation by the Prion-Like Amino Acid Composition (PLAAC) algorithm (Lancaster *et al.* 2014) (Figure 5A). Prion formation can be driven by overexpression of the constituent protein, and phenotypes conferred by the prion are stable over many generations after transient overexpression (Chakrabortee *et al.* 2016). In order to measure the impact of deleting the polyN domain on prion formation, we measured resistance to radicicol of strains that experienced previous overexpression of *AZF1*. Consistent with previous reports, overexpression of WT improves growth in radicicol compared to cells that have never experienced overexpression (Figure 5B). Unexpectedly, overexpression of ΔN also confers increased resistance to radicicol compared to naïve cells (Figure 5C). Yeast prions have been referred to as a “bet-hedging” strategy due to their ability to confer beneficial phenotypes under some conditions and be deleterious in others (Halfmann *et al.* 2010). We find that transient overexpression of both WT and ΔN do not impact growth in the absence of radicicol (Figures 5D and 5E), indicating that [AZF1⁺] is not toxic to cells under normal conditions.

Figure 5. Role of the polyN domain in prion formation and activity. A) Azf1p domains predicted to be capable of promoting prion formation by PLAAC. The polyQ and polyN domains of Azf1p are both predicted to have prion-like properties. B) Growth of WT cells that did (red) and did not (blue) experience previous overexpression (OE) of *AZF1* in SC-Ura + 75 μ M radicicol measured by OD₆₀₀ in a plate reader. C) Growth of Δ N cells that did (red) and did not (blue) experience previous overexpression of Δ N in SC-Ura + 75 μ M radicicol measured by OD₆₀₀ in a plate reader. Previous overexpression of both WT and Δ N confer resistance to radicicol. D) Growth of WT cells in SC-Ura with and without previous overexpression of *AZF1* measured by OD₆₀₀ in a plate reader. E) Growth Δ N cells in SC-Ura with and without previous overexpression of Δ N measured by OD₆₀₀ in a plate reader. Overexpression of *AZF1* does not impact growth under normal conditions. All growth curves are the average of three biological replicates. The first 10 hours are highlighted to show the differences in growth more clearly and due to the reduced accuracy of the plate reader at high optical densities.

Azf1p exhibits repeat-length variation across strains of yeast

Repeat domains are well known to be genetically unstable, and changes in repeat copy number can impact protein function. This variation can enable organisms to adapt to different environmental conditions and drive evolution (Sawyer *et al.* 1997; Ebisawa *et al.* 2001; Hoyer 2001; Johnsen *et al.* 2007; Fidalgo *et al.* 2008; Gemayel *et al.* 2015; Press and Queitsch 2017). We therefore examined the homorepeat domains of Azf1p for length variation. Through a bioinformatic analysis of the genomes of 93 lab and wild isolates of *S. cerevisiae* (Strope *et al.* 2015; Babokhov *et al.* 2018), we determined that the polyN and polyQ domains of Azf1p vary in repeat copy number across strains of yeast. The polyN domain has 8 alleles ranging from 17 to 25 N residues (Figure 6A), and the polyQ domain has 3 alleles, 9Q, 12Q, and 17Q (Figure 6B).

Figure 6. Repeat variation in the polyQ and polyN domains of Azf1p may impact protein function. A) The distribution of polyN domain variants across 93 strains of *S. cerevisiae*. The polyN domain has a range of alleles from 17 to 25 N residues. B) The distribution of polyQ domain variants across 93 strains of *S. cerevisiae*. The polyQ domain has 3 alleles: 9Q, 12Q, and 17Q. C) Model for ways that repeat copy number variation in Azf1p may impact protein function. Repeat variation may impact prion formation and/or stability, target gene expression, or protein aggregation and/or phase separation.

Azf1p is a prion-forming yeast TF that contains a polyQ and a polyN domain. To begin to elucidate how repetitive regions modulate Azf1p function, in this work we sought to investigate the role of the polyN domain in transcription and [AZF1⁺] prion formation. We find that the polyN domain plays a subtle role in transcription but is not required for prion formation. We also report that both the polyQ and polyN domains vary in copy number across strains of yeast.

DISCUSSION

Azf1p has previously been reported to be a transcriptional activator that is important for growth on glycerol and other non-fermentable carbon sources (Slattery *et al.* 2006). In this work, we report a new phenotype for *AZF1* during growth in glucose and identify a larger set of genes that are regulated by Azf1p. We monitored growth of WT and *azf1Δ* cells in glucose by optical density and find that the null mutant has a shortened lag phase compared to WT (Figure 1A). To further investigate the role of Azf1p during growth in glucose, we measured differential gene expression in the null mutant compared to WT by RNA-seq (Figure 1B and Table S5). To our knowledge, *azf1Δ* has not been previously shown to upregulate gene expression. Our data suggests that Azf1p has a repressive function.

Despite being historically dismissed as functionally unimportant, homorepeats, particularly polyQ and polyN, are conserved across eukaryotic TFs and have been shown to play important roles in signaling, binding, and protein solubility. N-rich and Q-rich sequences are also characteristic of the prion-forming domains of yeast prions, and many repeat-containing TFs form prions. Azf1p contains a polyQ and a polyN repeat domain, and we sought to investigate the roles of these domains in transcription and prion activity. We attempted to delete each of these domains in the *AZF1-GFP* strain from the yeast GFP collection (Huh *et al.* 2003; Howson *et al.* 2005). While deletion of the polyN domain was successful (Figure 2A), the polyQ domain could not be deleted using multiple strategies. This predicts a fundamental role for the polyQ domain of Azf1p in yeast but requires significant additional exploration that is beyond the scope of this manuscript. Here we continued the interrogation of the polyN domain

as they are thought to serve similar functions to polyQ domains but are not nearly as well studied.

Due to the known involvement of a homorepeat domain in the repressive activity of Cyc8p, we hypothesized that the polyN domain may be important for this function. In order to address this hypothesis, we first examined the role of the polyN domain in growth in different carbon sources. We find that deletion of the polyN domain produces a mild growth defect on glycerol and moderately reduces the length of lag phase compared to WT (Figure 2B and 2C). Because each of these phenotypes is less pronounced than the corresponding phenotype observed in the null mutant, we conclude that the polyN domain has a minor but not essential role in Azf1p function. In order to better understand these phenotypes, we compared gene expression levels in ΔN to WT as measured by RNA-seq. Consistent with our phenotypic growth data, we find that deletion of the polyN domain impacts gene expression in both glucose (Figure 3A) and glycerol (Figure 3B), suggesting that this domain is involved in Azf1p's TF activity. During growth in both carbon sources, more genes are upregulated than downregulated upon deletion of the polyN domain. This result is consistent with our hypothesis that the polyN domain plays a role in Azf1p's repressive function.

In order to explore how the changes in gene expression measured by RNA-seq give rise to the observed growth patterns in each carbon source, we performed GO Enrichment Analysis on genes differentially expressed between ΔN and WT. In glycerol, genes upregulated by ΔN are enriched in gluconeogenesis (17-fold enrichment), glycolysis (15-fold enrichment), and cell wall organization (3-fold enrichment) (Figure 3D). These processes are consistent with the those previously reported to be regulated by Azf1p (Slattery *et al.* 2006). Genes upregulated by ΔN during growth in glucose fit into two main categories: environmental stress tolerance and carbon metabolism and storage (Figure 3C). Upregulation of carbon metabolism genes may underlie the shortened lag phase observed in ΔN , as cells that can import and utilize glucose more rapidly or efficiently may be able to enter the log phase of growth more quickly. This result may have implications for fermentation applications and small molecule production. Yeast are used in the production of a variety of foods and drinks, as well as natural and non-natural small molecules for pharmaceuticals, fuels, and even perfumes (Lee *et al.* 2012; Zhang

et al. 2017; Chen *et al.* 2020). Metabolic engineering is a crucial technology geared towards improving the efficiency and output of these processes, and the new phenotypes we report for a null mutant and a ΔN mutant indicate that Azf1p may be a useful target.

We hypothesized that the transcriptional changes observed in ΔN may result from altered Azf1p localization. During both log and stationary phase growth, we find that ΔN exhibits the same localization patterns as WT regardless of the carbon source (Figure 4). Deletion of the polyN domain therefore does not appear to impact Azf1p localization. This observation is consistent with the result that deletion of the polyN domain impacts the expression of only a subset of Azf1p's target genes, as altered localization would likely affect all of Azf1p's targets. The polyN domain may instead act as a binding domain with transcriptional cofactors. This hypothesis is supported by the different growth and gene expression patterns observed between ΔN and *azf1 Δ* . Different target genes have different combinations of transcriptional regulators. The polyN domain may be important for binding with only a subset of Azf1p's cofactors, causing deletion of this domain to impact expression of only a subset of Azf1p's target genes.

In a large-scale screen for yeast prions conducted by Chakrabortee and colleagues, Azf1p was found to form a prion, [AZF1⁺], that confers resistance to the drug radicicol. Overexpression of the constituent protein drives prion formation, and phenotypes conferred by the prion are stable over many generations after transient overexpression (Chakrabortee *et al.* 2016). The PrD of Azf1p has not been characterized experimentally, but PLAAC predicts that the polyN domain is one of several sequences capable of promoting prion formation (Lancaster *et al.* 2014) (Figure 5A). In order to determine if the polyN domain plays a role in [AZF1⁺] prion formation or activity, we monitored growth in radicicol of WT and ΔN cells that previously overexpressed WT and ΔN , respectively. Consistent with Chakrabortee and colleagues' results, transient overexpression of WT improves growth in radicicol compared to cells that have never experienced overexpression (Figure 5B). Surprisingly, overexpression of ΔN also confers resistance to radicicol compared to naïve cells (Figure 5C), suggesting that the polyN domain is dispensable for prion formation. The polyQ domain is within a large N-terminal IDR predicted

by PLAAC to be PrD-like (Lancaster *et al.* 2014) (Figure 5A). This domain may therefore be a more relevant candidate for future studies aimed at determining the PrD of Azf1p.

Finally, we report that the polyN and polyQ domains of Azf1p vary in repeat copy number across strains of yeast. The polyN domain has 8 alleles ranging from 17 to 25 N residues (Figure 6A), and the polyQ domain has 3 alleles, 9Q, 12Q, and 17Q (Figure 6B). We propose a model outlining several ways that repeat-length variation may impact Azf1p function (Figure 6C). First, repeat variation could alter prion formation and/or stability. In yeast, expanded repeats in PrDs have been shown to increase prion formation (Liu and Lindquist 1999; Parham *et al.* 2001; Tank *et al.* 2007). While we find that the polyN domain does not play a significant role in prion formation, PLAAC predicts that the polyQ domain is PrD-like (Figure 5A). If the polyQ domain is in fact important for prionogenesis, short polyQ alleles may reduce prion formation and/or stability of the prion conformer, while long alleles or abnormally expanded repeats could result in the formation of an irreversible amyloid fiber.

Repeat variation could also impact expression of Azf1p's target genes. Another prion-forming yeast TF, Cyc8p, has a naturally variable polyQ domain. Variation in this domain impacts Cyc8p solubility and interaction with its binding partners, which in turn alters target gene expression and the associated phenotypes (Gemayel *et al.* 2015). We find that the polyN domain of Azf1p plays a subtle role in transcription, possibly by mediating binding to transcriptional cofactors. The impact of repeat variation in the polyN domain of Azf1p, as well as the variable repeat domains of many other TFs across eukaryotes, on target gene expression warrants further study. The gene expression and resulting metabolic changes we observed in ΔN may have implications in fermentation applications and small molecule production. Deletion of the polyN domain moderately shortens the lag phase in glucose (Figure 1C), a phenotype that can potentially be explained by the upregulation of genes involved in carbon metabolism and storage (Figure 3C). Repeat variation may serve as a mechanism to further tune this phenotype by altering gene expression.

One possible mechanism by which repeat variation could impact Azf1p's function in transcription is phase separation. Several eukaryotic TFs have been shown to phase separate with Mediator via their IDRs, and the formation of these droplets was demonstrated to be

required for transcription (Boija *et al.* 2018; Sabari *et al.* 2018). The polyQ and polyN domains of Azf1p are both intrinsically disordered (Babokhov *et al.* 2018), and length variation within these domains could impact the extent to which Azf1p can be incorporated into phase separated droplets.

Repetitive domains were historically dismissed as functionally unimportant due to their lack of three-dimensional structure. These domains are now known to be important for signaling, binding, prion formation, and phase separation. Homorepeats are particularly common in eukaryotic TFs and often exhibit length variation across organisms. The function of these domains and the impact of repeat variation, however, has only been minimally explored. We report a new role for the variable polyN domain of the yeast TF Azf1p in transcription and highlight repeat copy number variation as a possible driver of target gene expression variance in this and other eukaryotic TFs. We also report metabolic changes that may be useful for fermentation applications and small molecule production.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Mackenzie Parmenter and Mike Babokhov for early contributions to this project. We thank M. McVey, K.A. McElroy, and members of the McVey Lab and Tufts Biology department for their helpful discussions during the preparation of this manuscript.

FUNDING

This work was supported by the Army Research Office [W911NF-16-1-0175, W911NF-19-1-0299, and W911NF-20-1-0083 to S.M.F.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Alberti, S., R. Halfmann, O. King, A. Kapila, and S. Lindquist, 2009 A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* 137: 146–158.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25: 25–29.
- Babokhov, M., B. I. Reinfeld, K. Hackbarth, Y. Bentov, and S. M. Fuchs, 2018 Tandem repeats drive variation of intrinsically disordered regions in budding yeast. *bioRxiv*.
- Boija, A., I. A. Klein, B. R. Sabari, A. Dall’Agnese, E. L. Coffey *et al.*, 2018 Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175: 1842–1855.
- Carbon, S., E. Douglass, N. Dunn, B. Good, N. L. Harris *et al.*, 2019 The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47: D330–D338.
- Cascarina, S. M., and E. D. Ross, 2014 Yeast prions and human prion-like proteins: Sequence features and prediction methods. *Cell. Mol. Life Sci.* 71: 2047–2063.
- Chakrabortee, S., J. S. Byers, S. Jones, D. M. Garcia, B. Bhullar *et al.*, 2016 Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* 167: 369–381.e12.
- Chakravarty, A. K., T. Smejkal, A. K. Itakura, D. M. Garcia, and D. F. Jarosz, 2020 A Non-amyloid Prion Particle that Activates a Heritable Gene Expression Program. *Mol. Cell* 77: 251–265.
- Chen, R., S. Yang, L. Zhang, and Y. J. Zhou, 2020 Advanced Strategies for Production of Natural Products in Yeast. *iScience* 23: 100879.
- Chernova, T. A., Y. O. Chernoff, and K. D. Wilkinson, 2017 Prion-based memory of heat stress in yeast. *Prion* 11: 151–161.
- Collinge, J., 2016 Mammalian prions and their wider relevance in neurodegenerative diseases. *Nature* 539: 217–226.
- Ebisawa, T., M. Uchiyama, N. Kajimura, K. Mishima, Y. Kamei *et al.*, 2001 Association of

- structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep.* 2: 342–346.
- Faux, N. G., S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison *et al.*, 2005 Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* 15: 537–551.
- Fidalgo, M., R. R. Barrales, and J. Jimenez, 2008 Coding repeat instability in the FLO11 gene of *Saccharomyces* yeasts. *Yeast* 25: 879–889.
- Franzmann, T. M., and S. Alberti, 2019 Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* 294: 7128–7136.
- Gemayel, R., S. Chavali, K. Pougach, M. Legendre, B. Zhu *et al.*, 2015 Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol. Cell* 59: 615–627.
- Gemayel, R., M. D. Vines, M. Legendre, and K. J. Verstrepen, 2010 Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu. Rev. Genet.* 44: 445–477.
- Guo, L., H. J. Kim, H. Wang, J. Monaghan, F. Freyermuth *et al.*, 2018 Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. *Cell* 173: 677–692.
- Halfmann, R., S. Alberti, and S. Lindquist, 2010 Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol.* 20: 125–133.
- Halfmann, R., D. F. Jarosz, S. K. Jones, A. Chang, A. K. Lancaster *et al.*, 2012 Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature* 482: 363–368.
- Halfmann, R., and S. Lindquist, 2010 Epigenetics in the extreme: Prions and the inheritance of environmentally acquired traits. *Science* (80-.). 330: 629–632.
- Harrison, A. F., and J. Shorter, 2017 RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474: 1417–1438.
- Harvey, Z. H., Y. Chen, and D. F. Jarosz, 2018 Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol. Cell* 69: 195–202.
- Holmes, D. L., A. K. Lancaster, S. Lindquist, and R. Halfmann, 2013 Heritable remodeling of yeast multicellularity by an environmentally responsive prion. *Cell* 153: 153–165.
- Howson, R., W. K. Huh, S. Ghaemmamghami, J. V. Falvo, K. Bower *et al.*, 2005 Construction,

- p>verification and experimental use of two epitope-tagged collections of budding yeast strains.
- Comp. Funct. Genomics*
- 6: 2–16.
- Hoyer, L. L., 2001 The ALS gene family of *Candida albicans*. *Trends Microbiol.* 9: 176–180.
- Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson *et al.*, 2003 Global analysis of protein localization in budding yeast. *Nature*.
- Itakura, A. K., A. K. Chakravarty, C. M. Jakobson, and D. F. Jarosz, 2020 Widespread Prion-Based Control of Growth and Differentiation Strategies in *Saccharomyces cerevisiae*. *Mol. Cell* 77: 266–278.
- Jarosz, D. F., J. C. S. Brown, G. A. Walker, M. S. Datta, W. L. Ung *et al.*, 2014a Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism. *Cell* 158: 1083–1093.
- Jarosz, D. F., A. K. Lancaster, J. C. S. Brown, and S. Lindquist, 2014b An evolutionarily conserved prion-like element converts wild fungi from metabolic specialists to generalists. *Cell* 158: 1072–1082.
- Johnsen, A., A. E. Fidler, S. Kuhn, K. L. Carter, A. Hoffmann *et al.*, 2007 Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Mol. Ecol.* 16: 4867–4880.
- Kolaczowska, A., M. Kolaczowski, A. Delahodde, and A. Goffeau, 2002 Functional dissection of Pdr1p, a regulator of multidrug resistance in *Saccharomyces cerevisiae*. *Mol. Genet. Genomics* 267: 96–106.
- Lancaster, A. K., A. Nutter-Upham, S. Lindquist, and O. D. King, 2014 PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 30: 2501–2502.
- Laughery, M. F., T. Hunter, A. Brown, J. Hoopes, T. Ostbye *et al.*, 2015 New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* 32: 711–720.
- Lee, J. W., D. Na, J. M. Park, J. Lee, S. Choi *et al.*, 2012 Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8: 536–546.
- Van Der Lee, R., M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill *et al.*, 2014 Classification

- of intrinsically disordered regions and proteins. *Chem. Rev.* 114: 6589–6631.
- Liu, J. J., and S. Lindquist, 1999 Oligopeptide-repeat expansions modulate “protein-only” inheritance in yeast. *Nature* 400: 576.
- Lu, X., and R. M. Murphy, 2014 Synthesis and disaggregation of asparagine repeat-containing peptides. *J. Pept. Sci.* 20: 860–867.
- March, Z. M., O. D. King, and J. Shorter, 2016 Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* 1647: 9–18.
- Mi, H., A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas, 2019 PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47: D419–D426.
- Murakami, T., S. Qamar, J. Q. Lin, G. S. K. Schierle, E. Rees *et al.*, 2015 ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* 88: 670–690.
- Newby, G. A., and S. Lindquist, 2013 Blessings in disguise: Biological benefits of prion-like mechanisms. *Trends Cell Biol.* 283: 251–259.
- Parham, S. N., C. G. Resende, and M. F. Tuite, 2001 Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. *EMBO J.* 20: 2111–2119.
- Patel, A., H. O. Lee, L. Jawerth, S. Maharana, M. Jahn *et al.*, 2015 A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162: 1066–1077.
- Paul, K. R., A. Molliex, S. Cascarina, A. E. Boncella, J. P. Taylor *et al.*, 2017 Effects of Mutations on the Aggregation Propensity of the Human Prion-Like Protein hnRNPA2B1. *Mol. Cell. Biol.* 37: e00652-16.
- Peters, T. W., and M. Huang, 2007 Protein aggregation and polyasparagine-mediated cellular toxicity in *Saccharomyces cerevisiae*. *Prion* 1: 144–153.
- Press, M. O., and C. Queitsch, 2017 Variability in a Short Tandem Repeat Mediates Complex Epistatic Interactions in *Arabidopsis thaliana*. *Genetics* 205: 455–464.
- Prusiner, S. B., 1982 Novel proteinaceous infectious particles cause scrapie. *Science* (80-.). 216: 136–144.

- Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey *et al.*, 2018 Coactivator condensation at super-enhancers links phase separation and gene control. *Science* (80-.). 361: eaar3958.
- Sawyer, L. A., J. M. Hennessy, A. A. Peixoto, E. Rosato, H. Parkinson *et al.*, 1997 Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* (80-.). 278: 2117–2120.
- Slattery, M. G., D. Liko, and W. Heideman, 2006 The Function and Properties of the Azf1 Transcriptional Regulator Change with Growth Conditions in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 5: 313–320.
- Strope, P. K., D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone *et al.*, 2015 The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25: 762–774.
- Tank, E. M. H., D. A. Harris, A. A. Desai, and H. L. True, 2007 Prion Protein Repeat Expansion Results in Increased Aggregation and Reveals Phenotypic Variability. *Mol. Cell. Biol.* 27: 5445–5455.
- Tuite, M. F., and T. R. Serio, 2010 The prion hypothesis: From biological anomaly to basic regulatory mechanism. *Nat. Rev. Mol. Cell Biol.* 11: 823–833.
- Zhang, Y., J. Nielsen, and Z. Liu, 2017 Engineering yeast metabolism for production of terpenoids for use as perfume ingredients, pharmaceuticals and biofuels. *FEMS Yeast Res.* 17:.

A.

B.

Figure 1

A.

B.

C.

Figure2

A.

B.

Figure 4

Figure 5

A.

B.

C.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.05.22.991010>; this version posted February 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Natural repeat length variants from ~100 yeast genomes:

Figure 6

A.

B.

C.

D.

E.

Figure 3