Title
Engineering expanded spliceosome function in Saccharomyces cerevisiae

Author
DeNicola, Anthony

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Engineering expanded spliceosome function in *Saccharomyces cerevisiae*

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemical Engineering

by

Anthony DeNicola

2018

ABSTRACT OF THE DISSERTATION

Engineering expanded spliceosome function in *Saccharomyces cerevisiae*

by

Anthony DeNicola

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2018

Professor Yi Tang, Chair

Nature has long been a source of chemical diversity and its untapped potential is a major resource for solving our medical and energy problems. In particular, filamentous fungi have been prolific producers of medicinal compounds in the fight against human disease. Therefore, there has been interest in leveraging the advances in genomics to discover new fungal biosynthetic pathways that yield novel bioactive compounds. Heterologous expression of biosynthetic genes in model organisms will be increasingly necessary for high-throughput exploration of this genomic sequence space. Unfortunately, fungal genes contain many non-coding introns, which are difficult to manually annotate or predict *in silico*. Additionally, it is not possible to obtain intron-free cDNA from uncultivable species or transcriptionally silent gene clusters. This intron problem magnifies as the number of genomes increases and it risks derailing heterologous expression of this new genetic data. Unfortunately, the native spliceosome of the commonly used model eukaryote *Saccharomyces cerevisiae* cannot remove introns from distant fungi.

In this thesis, I will describe my efforts to engineer *S. cerevisiae* with expanded spliceosome functionality. I identified two failure modes that prevent splicing of an intron from *Aspergillus fumigatus*. This led to the generation of a chimeric yeast-fungal BranchBinding Protein that has enhanced specificity for an intron containing a fungal branchpoint site. Expression of this

mutant protein enabled a 2-fold improvement in splicing of an intron with a suboptimal branchpoint site. Additionally, we identified multiple synergistic splicing factor mutations with mBBP, *YHC1-D36A* and downregulation of *IFH1*, that enabled a 1.6-fold improvement of splicing of the *A. fumigatus* intron. Additional studies modifying the U2 small nuclear RNA as well splicing proofreaders *PRP5, PRP16,* and *PRP28* will be described, highlighting the drawbacks of these approaches. This study is the first demonstration of improved splicing of an *Aspergillus* intron through spliceosome engineering in *S. cerevisiae*. Using the tools, methodologies, and yeast strains provided by this work, the spliceosome can be engineered with new function, broadening the scope of how synthetic biology will be used to enhance heterologous expression in diverse research fields, such as in the elucidation of the splicing code and in natural products discovery.

The dissertation of Anthony DeNicola is approved.

Douglas L. Black

Guillaume Chanfreau

Yvonne Y. Chen

Yi Tang, Committee Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Ch. 1.1 – 1.6 contains material written by A.B. DeNicola from the publication J.M. Billingsley, ***A.B. DeNicola**, Y. Tang, Technology development for natural product biosynthesis in *Saccharomyces cerevisiae. Curr. Opin. Biotechnol*., 42 (2016), pp. 74-83

*Equal contributing author

The work described in this dissertation was supported by National Institutes of Health grant 1U01GM110706, awarded to the Fungal Team of the Genomes to Natural Products Network (GNPN), which is composed of Yi Tang and his lab as well as a team at the Stanford Genome Technology Center (SGTC). The SGTC team was initially headed by Maureen Hillenmeyer and received additional leadership from Ronald Davis, Robert P. St. Onge, Chaitan Khosla, and Lars Steinmetz.

I would like to begin by expressing my gratitude to Yi Tang. He has been my mentor for the five years of my PhD program and has guided me through multiple challenging projects. He always believed in my ability to get through some difficult science and encouraged me to try new techniques. He kept a strong focus on the major story and helped gently redirect me toward what was important. He's also opened a lot of doors for me in my career and really set the stage for me to begin my life post-PhD. I appreciate all he's done for me. I've learned a lot from him about leadership, academics, and especially about the power of Chinese liquor. We've been on many adventures together over the years! I distinctly recall barreling down the highway to the Bay Area in a minivan (Hsiao-Ching Lin and I clutching each other for dear life) as Yi was racing Wei Xu who was driving the other half of the lab in a second minivan. But we survived (pretty sure Yi won that race) and stopped to get fresh strawberries before piling back in the van and continuing north. The San Francisco trip was a blast, we saw so many sights, such as the Redwood forests, Lombard Street, and the insides of many, many bars. But of course, it was a highly professional

trip so we met with our Stanford collaborators for a bit before taking selfies on the Stanford lawn and making human pyramids.

I've seen many iterations of the Tang lab over the years of my PhD and seen so many exceptional scientists and incredible people come and go. Ralph Cacho was my mentor when I was a first-year student and my first roommate in Los Angeles! He set a high bar for scientific excellence and yet as hard as he worked, he was always down for an adventure such as when we went to Griffith Observatory and hiked down the mountain to get Korean BBQ. I'd also like to thank Jaclyn Winter who was a great friend and role model. We got lunches together at Narnia (a hidden food court in the labyrinth of CHS), spent many nights bonding at Liquid Kitty (great bar, may it RIP), and endured many trials together (such as the terrible UCLA flooding of 2014, may her car RIP). I'd also like to thank Mancheng Tang (TMC), who sat next to me in our office for nearly my entire PhD, who knows so much that he advises half the lab if not more, and who has the amazing ability to nap anywhere at anytime. As my main man also on the GNPN, we've traveled together a lot and I found that he can really hold his liquor.

I've also been fortunate enough to serve as a mentor to some truly exceptional scientists. John Billingsley may have been my first graduate student mentee, but he was the one who taught me a lot about what it means to be a mentor. His determination to see things through and his gifted artistic abilities in writing and figure design brought the iridoid project to new heights and scientifically inspired me. Also important, he taught me about Rupaul's Drag Race and just how many mimosas the human body can actually drink. My second graduate student mentee Danielle Yee has been truly wonderful and continues to impress me with her scientific prowess and unexpected comedic zingers.

I would also like to thank Jan Schlemmer who continues to support and encourage me as he continues in his own PhD program. Wai Man (Emily) Yu was my first undergraduate student and she demonstrated incredible loyalty and dedication to her projects, often working late into the night with me on chemical extractions and molecular cloning. Jenette Creso worked with me for

one summer during her undergraduate and has become a great friend as she continues in her PhD program at Yale. My undergraduate Vidya Subrahmanyam was an incredible help to me in my final two years and is a renowned expert in memology. Her humor brightens the office whenever she comes by and her magical experimental hands helped me with 52 simultaneous and perfectly successful yeast transformations.

I would also like to thank the team at Stanford for hosting me for several months during my fourth year. I received a lot of help from Kevin Roy, Justin Smith, and Bob St. Onge. They made a big impact on how I view experimental design and pushed me to think bigger with my experiments. TranSceInt was developed by Joe Horecka (12/1/1963 – 10/20/2017), who was an amazing scientist and always supportive of my work. He is missed! I would also like to thank Raeka Aiyar, a true friend, and my partner in adventures as two scientists out on the town.

I couldn't have an acknowledgment section without thanking Carly Bond. Our friendship really began when we did some volunteering for Yi's daughter's class by performing a science demonstration (also with the help of Muxun Zhao). We had a lot of fun coming up with experiments to do with the kids, such as making corn starch goop to teach about fluid dynamics. But most importantly we taught them how to have fun with liquid nitrogen! We had children dip roses into the liquid nitrogen and smash them on the table. There wasn't really a good lesson with this experiment, we just wanted to smash things. Carly and I also hosted the Carly-Anthony Show nearly every year as entertainment for departmental parties. We put in a ton of time and effort to host fun events with games, talent shows, caroling, and comedy skits lampooning the colorful characters in our department (there have been many).

I'd also like to give a shoutout to Karl Abad, a close friend from high school who was my roommate for most of my PhD as we lived together in West Hollywood. And finally, I'd like to thank my family, particularly my mother Laura DeNicola, father James DeNicola, and sister Dana Silber for endlessly supporting me and propping me up when I was struggling. I couldn't have completed this work without them.

VITA

2013	B.S., *magna cum laude* Chemical Engineering University of California, Irvine Irvine, CA
2014-2017	Chemistry-Biology Interface Training Fellow Department of Chemistry & Biochemistry University of California, Los Angeles
2013-2018	Research Assistant Department of Chemical and Biomolecular Engineering University of California, Los Angeles

PUBLICATIONS

J.M. Billingsley, **A.B. DeNicola**, J.S. Barber, M-C. Tang, J. Horecka, A. Chu, N.K. Garg, Y. Tang, Engineering the biocatalytic selectivity of iridoid production in *Saccharomyces cerevisiae*. *Metabolic Engineering*, 44 (2017), pp.117-125

J.M. Billingsley, *__A.B. DeNicola__, Y. Tang, Technology development for natural product biosynthesis in *Saccharomyces cerevisiae. Curr. Opin. Biotechnol*., 42 (2016), pp. 74-83
*Equal contributing author

PRESENTATIONS

Metabolic Engineering 11 (06/2016), The Western Awaji Island Conference Center, Japan
Poster presentation, short oral presentation: DeNicola, A.*; Tang, Yi.
Engineering the Spliceosome of Saccharomyces Cerevisiae to Splice Heterologous Introns from Distant Fungi

UCLA Molecular Biology Institute Research Conference (04/2016), Lake Arrowhead, CA
Oral presentation: DeNicola, A.*; Tang, Yi.
Engineering the spliceosome of Saccharomyces cerevisiae to splice heterologous introns from distant fungi

Chemistry-Biology Interface Symposium (09/2015) University of California, Los Angeles
Poster presentation: DeNicola, A.*; Tang, Yi.
Engineering reporters to assay splicing of heterologous introns in Saccharomyces cerevisiae

Chemistry-Biology Interface Symposium (08/2014), University of California, Los Angeles
Poster presentation: DeNicola, A.*; Tang, Yi.
Metabolic engineering of budding yeast for the high production of geraniol

1. INTRODUCTION

Natural products play an indispensable role as pharmaceuticals, specialty chemicals, and industrially relevant compounds. The recent surge of genomics data has revitalized the discovery and characterization of enzymes and small molecules from plants, fungi, and bacteria for use in medical and industrial applications as well as in bioenergy. Meanwhile, the development of new genetic tools for the yeast *Saccharomyces cerevisiae* has fueled efforts to refactor biosynthetic pathways and facilitate heterologous production of valuable compounds (Figure 1). In contrast to many native production hosts, yeast is fermentable, genetically tractable, and generally recognized as safe. Furthermore, many natural product biosynthesis enzymes and heterologous proteins have been successfully reconstituted in *S. cerevisiae* that show poor activity when expressed in the model prokaryote *Escherichia coli* (1). Strategies drawing from genomics, synthetic biology, and metabolic engineering have been applied to refactor biosynthetic pathways and optimize production of secondary metabolites. In the past several years, numerous pioneering accomplishments in the field have been realized, including reconstitution of complex pathways (2) and industrial scale-up of engineered systems (3). The development of tools and strategies employed to engineer yeast for natural product production will be discussed.

1.1 Background of technology development for gene expression in yeast

Nature has long been a source of chemical diversity and its untapped potential will increasingly become a powerful resource against antibiotic resistance (4) and in the development of a sustainable energy future (5). A plethora of chemical and biological products have been isolated from natural sources and adapted for use in medicine, industry, and bioenergy. Given the rapid improvements in molecular biotechnology and genome sequencing, we have access to more sequence content then we can successfully characterize. Therefore, it is imperative that new synthetic biology tools are generated to help in the discovery and characterization of Nature's diversity.

As roughly 70% of antibacterial and anticancer drugs are natural products or inspired by natural products (6), there has been renewed interest in leveraging the recent advances in genomics to discover biosynthetic gene clusters that yield new bioactive scaffolds (7). Recent results of genome-sequencing projects have revealed a wealth of uncharacterized secondary metabolites (8). In particular, filamentous fungi are prolific producers of valuable compounds such as the cholesterol lowering drug lovastatin (9), the antifungal compound griseofulvin (10), the immunosuppressant drug cyclosporine (11), and the penicillin and cephalosporin antibiotics (12). The rapid increase in genomic sequence data has revealed that as many as 5 million fungal species exist (13) and only 10% of natural products have been discovered in even well-characterized fungal species (14). Considering the list of putative biosynthetic gene clusters will soon number in the millions (15), it is essential that a high-throughput pipeline is developed to translate the genetic code into a characterized chemical readout. Traditional natural products research has utilized a “top-down” approach (16) that relies on expression of gene clusters in their natural host organism and screening of collected extracts. However, it can be costly and laborious to work with the native producer as they tend to synthesize low amounts of the product and often prove impervious to genetic manipulation (17). Also, gene clusters may be transcriptionally silent under laboratory growth conditions and potentially valuable products will not be expressed (18). Efforts to turn on these gene clusters have been successful in a variety of natural hosts (18-20), but these advancements will not be helpful for the 99% of microbial strains that are not readily cultivable (21). Considering the decreasing cost of gene synthesis (22) and improvements in metagenomic analysis (7,23,24), there is an increasing advantage to this “bottom-up” or genes-first approach to NP discovery (25).

Fungi are also a rich source of enzymes with uses in food, textile, and detergent industries (26). Additionally, fungi are the source of nearly all industrial enzymes used for lignocellulose digestion in bioenergy production (27). The recent discovery of the rich variety of fungal

cellulosomes for degradation of plant biomass (28) is an exciting prospect for further engineering efforts. However, due to the difficulties associated with establishing an industrial process, a high hit frequency covering extensive natural diversity is needed to improve industrial enzyme bioprospecting (29). Heterologous expression of fungal genes in a well characterized host will be increasingly necessary for the economical and efficient exploration of new sequence space (30,31), which will have a profound impact on academic and industrial fields as diverse as medicine and bioenergy.

Much of the recent development in the natural product discovery pipeline has focused on connecting natural products to the genes that encode them in bacterial species (8,32). Bacterial gene clusters identified through genome mining strategies can be refactored for expression in a model bacterial species such as *E. coli* or *Streptomyces coelicolor* (33). Through expression of individual enzymes or entire biosynthetic pathways, the gene clusters can be characterized and the structure and function of the products discovered. However, unlike prokaryotic gene clusters, pathway refactoring for eukaryotic gene clusters introduces additional challenges for heterologous expression.

High-throughput exploration of the diversity of fungal genomes requires correct expression of gene products in a heterologous host, which necessitates pathway refactoring. Refactoring involves restructuring the genetic elements of the pathway with characterized regulatory parts for expression in the native producer or in a heterologous host (34). I will emphasize the role of pathway refactoring for heterologous expression. Despite the improvements to molecular biotechnology methods for gene synthesis and assembly, there have been many problems associated with expression of refactored fungal heterologous pathways, such as poor protein expression, low precursor availability in the heterologous host, and intron splicing (35). Solutions have been identified for improving protein expression levels through codon optimization (36) and metabolic engineering has been used to increase supply of precursor pools (37), but no strategies exist to address the problem of intron splicing. Common fungal species contain an average of 2.5

introns per gene (38) and genes involved in NP biosynthesis can contain 5-9 introns (39) or more. Although intron free genes can be synthesized, the presence of an intron may be important to stimulate transcription (40) and even then, it is difficult to predict introns with *in silico* methods (41). These methods rely on accurate genome sequences and comparisons to homologous protein sequences; even one intron incorrectly predicted will render the protein nonfunctional when expressed (42). Annotated databases are frequently incorrect (25). Therefore, a universal host capable of rapid gene assembly, intron splicing, and sophisticated protein expression of minimally refactored heterologous genes would be ideal to explore the tremendous amount of genomic sequence diversity available.

1.2 Refactoring biosynthetic gene clusters

Refactoring biosynthetic gene clusters (BGCs) involves restructuring the genetic elements of the pathway with characterized regulatory parts for expression in the native producer or a heterologous host. The Voigt group established the methodology for biosynthetic pathway refactoring when they rebuilt the entire 23.5 kb nitrogen fixation gene cluster from *Klebsiella oxytoca* with synthetic parts for de-regulated expression in the native host (34). During pathway refactoring for heterologous expression in *S. cerevisiae*, each gene in a BGC is either synthesized or amplified from genomic DNA or cDNA. Promoters and terminators are cloned upstream and downstream of each coding region to generate expression cassettes, which are assembled and introduced into the desired platform strain (Figure 2).

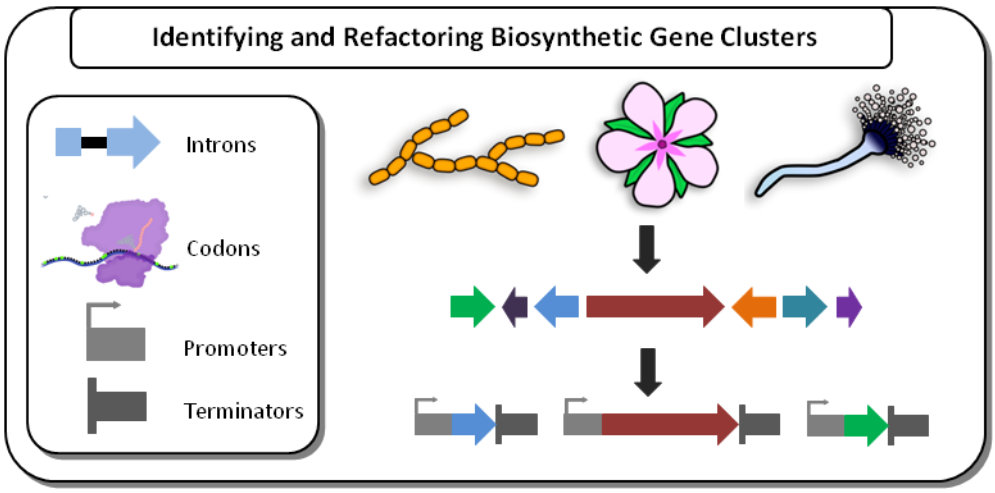


Figure 1. Refactoring biosynthetic genes. Overview of the process of refactoring biosynthetic gene clusters for heterologous expression in yeast.

1.3 Preparing pathway genes for heterologous expression

Pathway refactoring of gene clusters identified from eukaryotic genomic DNA often requires the removal of non-coding intron sequences. Fungal species, such as *Aspergillus*, contain many introns (43) and individual genes involved in natural products pathways frequently contain as many as 5-9 introns (39). Although there are several methods for successful *in silico* intron prediction from genome sequences (44), even one improperly annotated intron can derail pathway characterization efforts due to incorrect protein translation. Intron-less cDNA can be generated from active gene clusters in cultivable species, but silent gene clusters and uncultivable species cannot be interrogated by this approach. Further, the native spliceosome of *S. cerevisiae* does not have the capability to remove introns from distant fungi (45). Through the mutation of specific yeast splicing factors, we have enabled the recognition of intron sequence motifs from *Aspergillus fumigatus*. With the continued rise in genomic sequence data, it will become increasingly important to develop new computational and experimental tools to solve the intron problem.

Despite the difficulties caused by heterologous introns in the original pathway sequence, introns native to *S. cerevisiae* have recently been added to the synthetic biology toolbox as

predictable regulators of gene expression levels. Yofe *et al*. generated a gene expression library consisting of 240 unique yeast strains, each expressing a yellow fluorescent protein (YFP) gene interrupted by a natural *S. cerevisiae* intron (46). By quantifying the fluorescence, introns were shown to reliably reduce reporter gene expression. Incorporating *S. cerevisiae* introns into gene assembly strategies would be useful for tuning optimal expression ratios of the different pathway enzymes, especially if promoter choices are limited due to the need for specific pathway regulation.

Either through gene synthesis or site-directed mutagenesis, sequences can be codon-optimized for improved translational efficiency in *S. cerevisiae*. Codon-optimizing two genes in the carotenoid pathway (*crtI* and *crtYB*) improved beta-carotene production in yeast by 200% (47). However, as recent work by Lanza *et al*. has shown, genes designed by traditional algorithms do not always yield higher expression than the wild type gene (48). The authors proposed a condition-specific approach to determine optimal codon usage under a given growth condition instead of relying on the frequency of codon usage in the entire genome. The catechol 1,2-dioxygenase gene from *Acinetobacter baylyi* was optimized for production at stationary phase and had 2.6-fold higher catalytic activity than the wild type and 2.9-fold higher activity than a commercially optimized variant.

1.4 Regulatory DNA sequences control the strength of gene expression

Each coding region must be flanked by a promoter and a terminator, which are regulatory DNA sequences that influence the frequency of gene transcription and the stability of the transcripts, respectively. Promoters, classified as constitutive or regulated, are *cis*-acting regulatory elements upstream of a coding sequence where transcription factors bind and recruit the RNA polymerase in order to initiate transcription. The report by Sun *et al*. detailed the characterization of 14 constitutive promoters via expression of green fluorescent protein (GFP) under varied glucose and oxygen conditions (49). Promoters of high and medium strength were

used in the refactoring of xylan degradation and zeaxanthin biosynthesis pathways and the resultant yeast strain produced 0.74 mg/L of zeaxanthin.

To avoid the generation of toxic products and metabolic stress during heterologous expression, cell growth can be decoupled from the production phase using regulated promoters, which are inducible or repressible under different physiological conditions. For example, a modified GAL induction system, repressed by glucose and induced by galactose, was used successfully in the production of artemisinic acid at 25 g/L in *S. cerevisiae* (3). Another glucose repressible promoter, ADH2p, induces strong transcription once glucose is depleted and has been shown to provide higher expression levels than the GAL promoter (50). The *ADH2* promoter has proven highly useful for the heterologous expression of biosynthetic genes and recently enabled the discovery of new fungal indole diterpenes (51). A library of *ADH2*-like promoters would prove beneficial for pathway refactoring and characterization of new biosynthetic gene clusters.

Unfortunately, yeast promoters tend to be hundreds of base pairs in length, which compounds the difficulties for rapid pathway refactoring. Recent work by Redden and Alper has focused on the construction of short synthetic promoters (52). The authors combined core promoter elements with upstream activation sequences to generate minimal promoters. They achieved high levels of both inducible and constitutive expression with an up to 80% reduction in the size of the promoter.

The terminator sequences, downstream of the coding region, influence the mRNA half-life. Yamanishi *et al*. evaluated 5302 terminator regions by quantifying the level of GFP expressed under the control of a strong promoter (53). Work by Curran *et al*. focused on characterizing the activity of over 30 terminator regions for metabolic engineering applications in yeast (54). Terminators were coupled to promoters with a variety of strengths and a yellow fluorescent protein was expressed and quantified with fluorescence-activated cell sorting (FACS). The Alper group also systematically developed short (35-70 bp) synthetic terminators and proved they function as

effectively as native terminators for the expression of the heterologous proteins involved in the biosynthesis of itaconic acid (55).

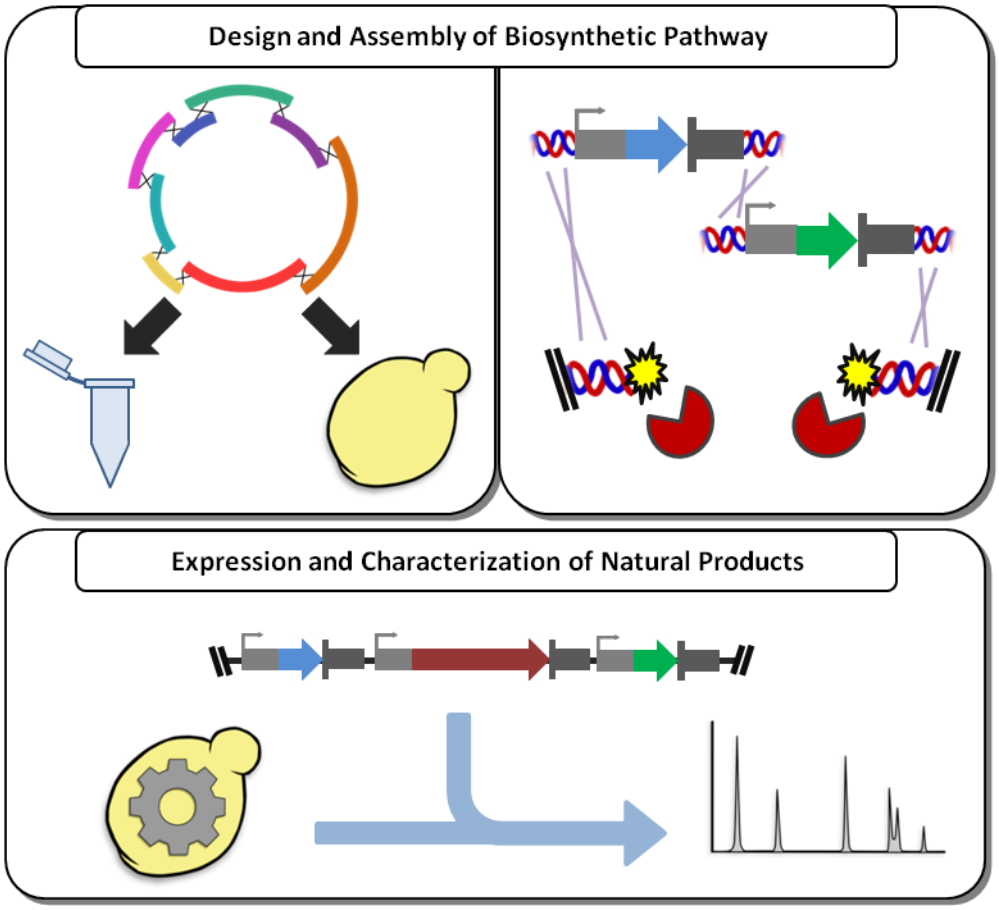


Figure 2. Assembly and expression of natural product pathways. (TOP) (Left) DNA Assembly in vitro and in vivo of plasmids for episomal expression. (Right) Multigene integration of refactored pathway genes facilitated by double-stranded break technology. (BOTTOM) Expressing the biosynthetic pathway in a pre-engineered platform strain for detection and characterization of natural products.

1.5 Episomal expression and plasmid assembly techniques

Refactored genes can be assembled into plasmids or integrated into the genome for expression in yeast. Episomal expression is an effective strategy as there are a variety of efficient methods for plasmid assembly and high copy numbers are accessible with few cloning steps. During plasmid construction, the identity of the selection marker and origin of replication must be decided upon depending on the genotype of the yeast strain used and desired copy number of

the plasmid. Prototrophic markers and markers conferring drug resistance are commonly used selection markers (56). However, the identity of the selection marker can negatively impact the growth rates of yeast and the metabolite production levels, particularly in haploid strains (57). Plasmid copy number is controlled by two frequently used origins: low copy CEN/ARS and high copy 2 µ. Low copy plasmids have been found to generate reliable expression patterns, while high copy plasmids can display more variable expression (58). The identity of the gene being expressed can also influence which origin to select. In a recent report by Trenchard and Smolke, it was found that low copy expression of plant P450s yielded higher production levels of the target compound cheilanthifoline (59). High copy expression of the endomembrane-localized P450s induced a stress response in the endoplasmic reticulum, causing membrane proliferation, and ultimately lowered cheilanthifoline biosynthesis.

There have been many recent advances for the *in vitro* and *in vivo* assembly of DNA parts. Gibson assembly is a sequence independent, one-pot method for assembling multiple overlapping DNA fragments through the combined activities of an exonuclease, polymerase, and ligase (60). Golden Gate method, which is sequence dependent, utilizes type II restriction enzymes to efficiently combine modular parts in a one-pot reaction (61) and an extension of this technology has recently been adapted specifically for yeast assemblies (58). The methodologies based on Golden Gate do not require downstream sequencing of assembled constructs as point mutations cannot arise once the individual components have been generated. The homologous recombination machinery of yeast can also be utilized for the assembly of plasmids *in vivo*, requiring only one transformation step and as few as 29 nucleotides of overlapping sequence between DNA fragments (62). Yeast homologous recombination has been used to refactor silent and orphan gene clusters, such as in the discovery of the antibiotic taromycin A from the marine actinomycete *Saccharomonospora* (19), and potent indolotryptoline antiproliferative agents, lazarimides A and B, from environmentally-derived DNA (18).

1.6 Genomic expression and gene integration technologies

Integration of heterologous genes has unique advantages over episomal expression, particularly for metabolic engineering applications. Genes integrated through homologous recombination are stably maintained in cell populations without the need for selection pressure, although simultaneous integration of a marker is necessary to select for integrants. If higher copy number is desired, multiple copies of pathway genes can be sequentially targeted to multiple locations in the genome or integrated using the repeating delta sites (63). For this purpose, the expression levels of many loci have been characterized using *lacZ* and luciferase reporters (64,65). In a recent report by Brown *et al.*, the plant natural product strictosidine was only detected after a rate-limiting enzyme in the pathway was integrated multiple times into the genome and expressed from a high copy plasmid (66).

Recent advancements in integration technology revolve around the induction of a double-strand break (DSB) in the DNA to greatly improve the efficiency of integration. The homing endonuclease I-SceI was used to introduce a DSB at a target locus and eight overlapping fragments, seven genes and one selection marker, were efficiently integrated simultaneously (67). Another robust method for the construction of multigene pathways is Reiterative Recombination (68). Two endonucleases, SceI and HO, are used iteratively in conjunction with recyclable markers to sequentially assembly DNA fragments into a single locus with high efficiencies. Despite these elegant advances, the development of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated systems (Cas) has ushered in a new era of efficient marker-free genome editing, promising greater flexibility and utility than previous DSB technologies. First adapted for utilization in yeast by DiCarlo *et al*. (69), CRISPR-Cas9 has proven to be a powerful tool for rapid pathway construction. Mans *et al.* demonstrated the simultaneous integration of 6 genes, composing the *Enterococcus faecalis* pyruvate dehydrogenase complex, at a single locus combined with a separate gene deletion to construct an acetyl-CoA overproducing strain (70). The potential of multigene pathway assembly using

CRISPR-Cas9 was further explored in a report by Jakočiūnas *et al.* (71). A total of 15 DNA parts (3 expression cassettes) were integrated into three separate loci without selection and with an efficiency of 31%. By programming CRISPR-Cas9 to cut at the repeating delta sites, 18 genomic copies of a combined xylose utilization and (*R,R*)-2,3-butanediol production pathway were integrated in one transformation step (72). The ability to assemble gene clusters at any genomic location through the combined use of DSB technologies and yeast homologous recombination will prove invaluable for the continued production of natural products in yeast.

1.7 *S. cerevisiae* and intron splicing

Saccharomyces cerevisiae has been a powerful tool in the characterization of many natural product gene clusters and industrial enzymes from both fungal sources. As a eukaryote, *S. cerevisiae* can successfully express membrane bound enzymes such as cytochromes P450 (59). Unparalleled homologous recombination abilities make yeast a prime choice for the refactoring of large multi-gene BGCs (62,73). High transformation efficiencies for yeast and its well-characterized mating capabilities allow generation of large libraries (74,75). Budding yeast has been extensively engineered with a suite of auxotrophic markers and many designer deletion strains available (56,76). Yeast is an excellent host for cellulosic bioethanol production (77) and can be engineered to utilize these alternative feedstocks (78). Additionally, *S. cerevisiae* has been the workhouse of the natural products community to express fungal gene clusters (79,80). Importantly, *S. cerevisiae* even has molecular machinery to remove introns, the spliceosome. Unfortunately, the budding yeast spliceosome usually fails to splice introns from other species (81-83), despite a few successful examples (84,85). Making introns more canonical has been shown to improve splicing in one case in budding yeast (81), but this is neither a universal solution nor feasible for rapid interrogation of new genomic space.

In order to improve *S.cerevisiae* as a host for the expression of fungal genes, the activity of its spliceosome needs to be enhanced. The spliceosome is the intron removal machinery in the eukaryotic cell (86), and splicing occurs co-transcriptionally (87). The spliceosome functions by two principles: 1) small nuclear RNAs (snRNAs) bind to intron splice sites to promote the two transesterification reactions needed for intron removal and exon ligation; and 2) proteins maneuver the snRNAs into correct position and offer additional roles in intron recognition, alternative splicing, splicing fidelity, and protecting reactive intermediates during spliceosome assembly (88). Intron splice sites, which are recognized by the snRNAs, are important for splicing catalysis and can vary in sequence identity across species. *S. cerevisiae* introns are close to the consensus sequence, as their splice sites bind perfectly to the snRNAs. In general, there is 1 intron per gene in yeast, although there are exceptions of 2 introns per gene (89). Limited alternative splicing exists in yeast with notable examples being *SUS1* (90), *PTC7* (91), *SCR1* (92,93), and *APE2* (94). Although there is more widespread use of non-productive alternative splice sites in budding yeast than previously thought, these isoforms are destroyed by the Nonsense-mediated Decay (NMD) RNA surveillance pathway and do not contribute to proteome diversity (95).

Distant fungal species have a greater number of introns per gene that *S. cerevisiae* (96), highly degenerate splice sites (38,97,98) and more frequent alternative splicing events (99-101). Although the core of the spliceosome is conserved from budding yeast to mammals (102), the degeneracy of splice sites coincides with the evolution of additional splicing factors (103). In particular, the SR protein family (104) aids in enhancing or repressing noncanonical splice site usage in higher eukaryotes (105-107). Specifically, the RS domain of the SR splicing factors contacts the splice signals to promote stable snRNA interaction (108) and to activate pre-mRNA splicing (109). *S. cerevisiae* lacks SR proteins, with the exception of its closest homolog *NPL3* (110-112). Despite the advancements in understanding fungal splicing, it is not known how many

SR proteins or splicing enhancers exist and, consequently, directly porting these heterologous splicing factors into budding yeast is not yet feasible.

2. THE SPLICEOSOME

2.1 Background of the spliceosome

The folding of pre-mRNA and excision of introns is completed by a multitude of *trans*-acting factors that comprise the spliceosome (113). Each of the U1, U2, U4, U5, and U6 small nuclear RNAs interact both with a variety of associated protein factors to form small nuclear ribonucleoproteins (snRNPs) that are the main building blocks of the spliceosome (114,115). The evolutionarily conserved core of the eukaryotic spliceosome consists of about ~80 proteins (116) and there are approximately 40 proteins involved in the maintenance of a catalytically active spliceosome (102,117). The life cycle of the spliceosome is highly ordered and dynamic as it forms stepwise on the pre-mRNA and none of the particles possess preformed active centers (118). The spliceosomal subunits undergo extensive remodeling during the course of splicing as the RNA active site components are initially sequestered as inactive conformations through a network of RNA-RNA interactions (119). These subunits become active through conformational changes in the spliceosome initiated by RNA helicases (120,121). During this process, the intron substrate is recognized multiple times by various splicing factors to ensure fidelity in splicing (116). The recent determination of the different stages of the spliceosome through cryo-electron microscopy is ushering in a new era in spliceosome research (122-133).

2.2 Role of the spliceosome in human disease

Alternative splicing of precursor messenger RNA is responsible for the precise regulation of gene expression and its dysregulation is the cause of various human genetic diseases (134). Multiple potential splice sites as well as the binding of accessory splicing factors influences the final fate of an RNA transcript and which specific isoform the spliceosome will generate (135).

There is also a complex relationship between splicing, transcription, and chromatin which further decides alternative splicing events (136). Aberrant splicing can occur when mutations arise in the splicing signals of a gene or in spliceosomal genes themselves. A splicing signal mutation can prevent proper association of the spliceosome at an intron/exon boundary or activate a cryptic splice site in another location of the gene to be used preferentially (137). Mutations in splicing factors change alternative splicing patterns by altering the binding affinities of the proteins for their targets (138). These can result in cancer-specific mis-splicing (139). For example, certain cancer cells contain mutations in splicing factor SF3b1 (associated with U2 snRNP) that promote alternate branchpoint site usage (140). Alternative splicing also contributes to mechanisms of cancer resistance by modulating drug targets (141). Multiple splicing factor mutations alter pre-mRNA splicing patterns associated with myelodysplastic syndrome (MDS) (142), particularly U2AF (143-147) and SF3b1 (148,149). As there has been much focus in the literature on U2AF and SF3b1 targets, and it will be crucial to continue to develop an understanding of how different splicing factor mutations contribute to alternative splicing regulation in the study of human disease.

2.3 Therapeutic approaches designed to alter splicing

Mutations involving the spliceosome can play a role in both therapy and disease, depending on where the splicing mutation occurs. First, if the disease-causing mutation is in a splicing signal of a gene (i.e. intron, exon, splicing enhancer/silencer sequence), then engineered spliceosome genes can be introduced to the host to correct the erroneous splicing at the level of the transcript. Second, mutations in spliceosome genes are related to the development of certain cancers. A cancer-causing mutation in a spliceosomal gene induces widespread changes to intron/exon recognition. Studies, described below, have shown cancer cells with spliceosome mutations become more susceptible to drugs that inhibit splicing. Therefore, studying spliceosome mutations will be beneficial for the design of therapies using these two approaches.

Precise engineering of spliceosome factors can rescue splicing impaired by mutations in the splicing signals, which has therapeutic potential for many RNA diseases. Spliceosome genes have been redesigned as therapies to correct splicing for different diseases using results from cell cultures or mouse models. For example, U1 small nuclear RNAs (a core component of the spliceosome) have been engineered to correct splicing defects in genes causing spinal muscular atrophy (SMA) (150-153), propionic acidema (154), and blood coagulation disorders (155-158), particularly when combined with antisense oligonucleotide molecules (159). These engineered U1 snRNAs provide more stable binding to a mutated splicing signal and promote correct splicing of transcripts, thus alleviating the disease phenotype. These engineered U1 snRNAs have also been designed to function as HIV-1 inhibitors. Targeting the U1 snRNP to the 3' end of HIV-1 mRNA substantially reduced viral protein expression in cell lines by blocking pre-mRNA polyadenylation and targeting the HIV-1 mRNA for degradation (160). To restore correct splicing of a gene causing retinitis pigmentosa, researchers have deployed both mutant U1 snRNAs (161,162) as well as another engineered splicing protein, U2AF (163) to promote the wild type splicing isoform. These studies highlight the wide variety of diseases that can be treated using this approach, but also show that current research has been primarily focused on U1 snRNA as a splicing modulation therapy.

Chemical compounds can also modulate the aberrant splicing in human RNA diseases (164). Their mechanisms of action are an area of active study, furthered by the advances in high-throughput RNA sequencing (165). Natural products have provided many leads for splicing-based therapeutic compounds (166,167), underscoring the importance of efficient exploration of new natural product sequence space (see Ch.1). Medicinal chemistry also plays an important role here and has been demonstrated to be useful to optimize tumor-selective spliceosome modulators (168). Some small molecules can be used to stabilize aberrant splicing factor associations as demonstrated with the stabilization of U1 snRNA interactions with *SMN2* in SMA mice (169).

These small molecule therapies have been proven to increase motor function and longevity in these mice models (170).

The spliceosome is also a therapeutic vulnerability for cancers (171). In cancer, malignancies from mutated spliceosomal genes are sensitive to pharmacological intervention (172), which has been exploited in the development of new therapeutic compounds. Splicing factor SF3b1 is a potential antitumor target as it is affected by the bacterial natural product pladienolide B (173) as well as spliceostatin A (174). Splicing inhibitor E7107 is well-studied in treating acute myeloid leukemia in mouse models (175). And sudemycins have been studied as treatments in haematological cancers, such as leukemia (176) due to their effect on alternative splicing (177) by altering interactions with the U2 snRNP and modifying chromatin (178).

Overall, learning how to control alternative splicing is crucial to designing new therapies to prevent splicing diseases. Studying how mutations in splicing factors influence their RNA-binding properties will have an impact on human health, both in understanding and treating human disease.

3. REPORTER DEVELOPMENT

In order to easily assay splicing in *S. cerevisiae*, I desired to have changes in splicing efficiency related to changes in cell fitness. This would allow for rapid testing of multiple different hypotheses without the need for RNA extraction and verification for every mutant or reporter. In this report, three different native yeast genes are utilized as splicing reporters. Introns are cloned into the open reading frames using different cloning techniques, described briefly in their appropriate section and in more detail in Materials and Methods. All reporter genes are tied to yeast growth and fitness phenotypes, allowing growth measurements on agar plates and in liquid media to serve as a proxy for splicing. In all cases, superior growth/fitness means superior splicing of the reporter gene transcript. An intron is defined as five modules: 5′ splice site (5′ss), In-Between 1 (IB1), BranchPoint Site (BPS), In-Between 2 (IB2), and 3′ splice site (3′ss). The

sequence identities of the 5′ss, BPS, and 3′ss are generally conserved across species due to their role in splicing catalysis. The two sequences in between the three conserved sites (IB1, IB2) vary in size and sequence even within the same organism. The nucleotides in the consensus modules for *S. cerevisiae* and *Aspergillus nidulans* vary greatly with BPS being the most divergent and the 5′ss and 3′ss being more similar between species (38). Two representative introns were chosen for the studies presented in this proposal: the second *MATa1* intron from *S. cerevisiae* and the fifth *PES1/NRPS1* intron (pes1-5) from *Aspergillus fumigatus*, with sizes of 52 bp and 54 bp respectively.

The intron MATa1 is unique in *S. cerevisiae.* The gene *MATa1* contains the two smallest introns in budding yeast, which are both spliced inefficiently (179) despite having consensus splice sites (180). For functional protein, both introns must be spliced (181) and efficient splicing requires both *IST3* (182) and *BUD13* (183). Considering the second *MATa1* intron has been previously placed in a synthetic context and shown to splice (184), this intron will be ideal for studying changes to a small intron, which is not frequently studied in yeast splicing literature.

We chose an intron from *A. fumigatus* as *Aspergillus* species are host to a plethora of valuable secondary metabolites and are a well-studied fungal species (24). *PES1/NRPS1* is one of the first non-ribosomal peptide synthetases identified for secondary metabolite production in *Aspergillus fumigatus* (185). The protein product improves fungal tolerance to oxidative stress during pathogenicity (186) and is involved in production of fumigaclavine C (187). Transcription of this biosynthesis gene is regulated extensively by a histone deacetylase HdaA (188) and a major regulator LaeA (189), although specifics are still under active investigation. Considering pes1-5 is on the smaller side for fungal introns and resides in a gene involved in a complex secondary metabolite pathway, we decided it would make an excellent test case as a difficult intron for *S. cerevisiae* to splice. It has a 5′ss that is used in *S. cerevisiae* (89), albeit at a low frequency, and an identifiable putative BPS.

3.1 *URA3* reporter

The first reporter utilized in these studies is based on the *URA3* marker. *URA3* encodes orotidine-5′-phosphate (OMP) decarboxylase, involved in uracil biosynthesis, and is used as a marker for selection in laboratory yeast strains, which have been designed to be auxotrophic for *URA3* (190). To determine the limitations of the yeast spliceosome when splicing the model intron pes1-5 from *Aspergillus fumigatus*, I constructed a library of intron variants by swapping modules between pes1-5 and the budding yeast MATa1. These hybrid introns were inserted into the coding sequence of the *URA3* gene using overlap-extension PCR and subsequently cloned into a low copy autonomous plasmid. This library was transformed into the common laboratory strain BY4741 (a *ura3*Δ0 strain) and plated on uracil dropout media to assay splicing; if the spliceosome assembled on the intron and performed catalytic activity, the intron would be removed from the *URA3* transcript and cells would be able to produce uracil and grow. If the intron was unable to be spliced or the splicing efficiency was too low to produce enough of the mature transcript, the cells would be unable to grow on the uracil dropout agar plates.

The wild type MATa1, pes1-5, and hybrid variants were tested using the *URA3* reporter. Yeast growth was observed, indicating active splicing, in strains containing the wild type MATa1, the hybrid MATa1 with the heterologous 5′ss (h5′ss), and the hybrid MATa1 with the heterologous IB2 (hIB2). Growth of the strain containing MATa1 served as a positive control and confirmed that functional OMP decarboxylase could be produced through correct excision of this intron from the open reading frame. It is not too surprising that the 5′ss spliced because the sequence from pes1-5 is used in wild type yeast introns (89). Substituting IB2 also allowed for efficient splicing, which is logical as yeast lacks factors that traditionally bind to the IB2 sequence (38).

Yeast growth was not observed, indicating abolished splicing, in strains containing the *URA3* reporter with the insertion of pes1-5, the hybrid MATa1 with the heterologous BPS (hBPS) and the hybrid MATa1 with both the heterologous 5′ss and heterologous IB1 (h5′ss_hIB1). Additionally, for an intron tested with the yeast BPS substituted into an otherwise pes1-5

sequence does not restore splicing, indicating that the BPS alone is not inducing splicing failure. An intron was also tested that had all consensus sequences from MATa1 (5′ss, BPS, 3′ss) but the hIB1 and hIB2 from pes1-5. This hybrid intron does not allow for any growth of the yeast on uracil dropout media indicating no or low splicing. These results offered a starting point for analysis of which modules prevent the heterologous splicing of pes1-5: the branchpoint site (hBPS) and the sequence in-between the 5′ splice site and the branchpoint site (hIB1).

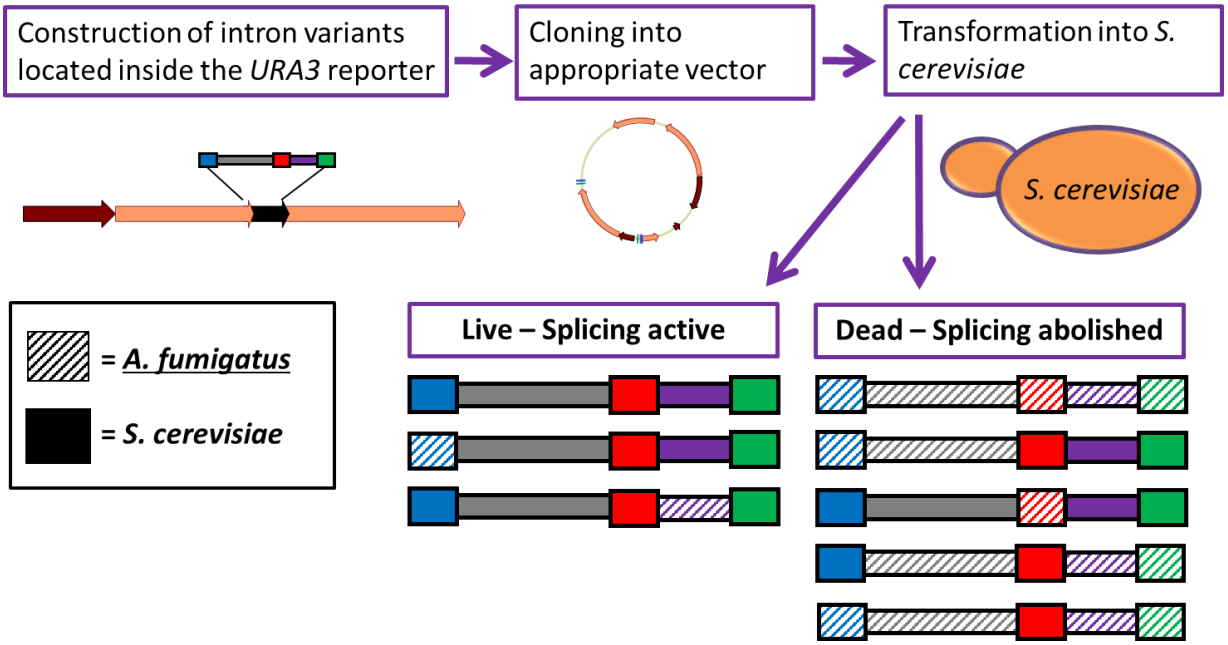


Figure 3. *URA3* is a useful splicing reporter. Schematic detailing the *URA3* reporter for assaying splicing in wild type yeast. Introns that were not efficiently spliced prevent yeast from growing on uracil dropout media.

3.2 *CUP1* reporter

Despite the utility of the *URA3* reporter, it suffers from several disadvantages as a splicing reporter. As there is no quantitative range between dead or alive, the reporter lacks sensitivity and is unable to assay events of low splicing efficiency. However, a sensitive splicing reporter based on the *CUP1* gene was previously developed in the Guthrie lab (191) and is used extensively by various groups (192,193). *CUP1* is a small nonessential yeast gene that encodes a copper-activated metallothionein. Expression of this small protein (62 amino acids) allows the cells to grow in the presence of copper in a dosage-dependent manner. The most famous

example of this technology is the *ACT1-CUP1* fusion (191). This construct consists of the several amino acids of the *ACT1* gene and the *ACT1* intron fused in frame with the *CUP1* ORF. It is frequently expressed from a high copy plasmid (2-micron origin) and under the regulation of a strong, constitutive promoter (GPDp). Using *CUP1* to assay splicing offers several advantages over fluorescent and prototrophic reporters: 1) intron splicing can be assayed with or without selection, depending on the goal of the experiment; 2) copper concentration can be modulated as a proxy for splicing efficiency; and 3) RNA transcripts can extracted and analyzed to quantify splicing efficiency.

In order to assay splicing in *S. cerevisiae*, I first developed a new reporter gene system. This is based on the landmark system described above, the *ACT1-CUP1* reporter gene, first developed by Guthrie lab and used widely since as a field standard (191). However, this system is most easily adapted to studying changes in splicing of the *ACT1* intron, which is much larger (309 bp) than the average intron from distant fungi (82% of *A. nidulans* introns are in the range of 0-79 bps (38)). Additionally, reports indicate that 2-micron plasmids can vary in copy number (57) and produce more variable expression profiles (58). In order to easily test a variety of introns, to minimize the variability between experiments, and to detect minute improvements in splicing, my collaborators and I designed a flexible integrated reporter system. However, before *CUP1* could be used as a splicing reporter, the endogenous *CUP1* activity in BY4741 had to be removed. Most laboratory strains of *S. cerevisiae* contain two copies of the *CUP1* gene in tandem. I deleted both copies from BY4741 by integration of the *LEU2* marker into yeast chromosome VIII. The genotype was confirmed by colony PCR and sequencing of upstream and downstream regions amplified from the genome. The phenotype was confirmed by plating the wild type strain and the *cup1*Δ strain on plates dosed with different concentration of copper. As expected, when the copper concentration increased, the *cup1*Δ strain was unable to grow.

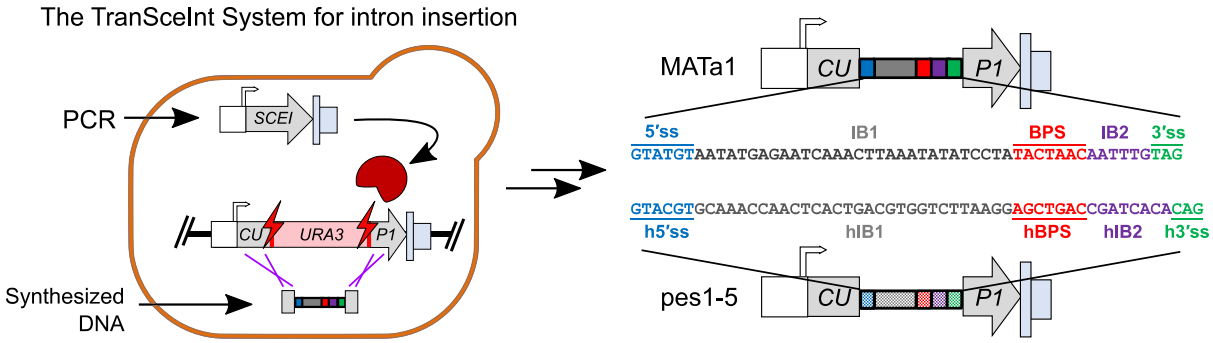


Figure 4. **Building integrated *CUP1* reporters without cloning.** System for rapid and cloning-free construction of yeast strains with integrated *CUP1*-intron reporters and copper assay development as a proxy for splicing. (Left) The TranSceInt system for intron insertion allows construction of a stably integrated reporter strain with a new intron using only two PCR reactions and one yeast transformation. (Right) Cartoon schematics of the MATa1 (yeast) and pes1-5 (*A. fumigatus*) *CUP1* reporters with the intron modules color-coded and labeled.

We designed our system to efficiently introduce new introns from synthesized DNA directly into the reporter locus in order to prevent excessive cloning steps (such as cloning each *CUP1*-intron construct *in vitro* and subsequent integration into the genome). For this purpose, a specialized *CUP1-URA3* landing pad was designed with the help of the late Dr. Joe Horecka. The *CUP1* open reading frame is interrupted 50 bps from the start codon by the *URA3* marker flanked by I-SceI sites for enzymatic cleavage. The meganuclease I-SceI is an effective inducer of double-stranded DNA breaks (194). This landing pad was integrated into the *cup1*Δ strain. To insert introns, the desired sequence is synthesized either as overlapping short oligonucleotides or as a gBlock (IDT) and amplified via PCR to make double-stranded DNA. Upon induction of *SCEI* in vivo (see Materials and Methods), the *URA3* marker is enzymatically cleaved by SceI and donor DNA containing homology to the *CUP1* gene can be inserted in its place. This system is called TranSceInt, due to the transient nature of SceI expression, and achieves high efficiency for genomic integrations (195). Expression of these *CUP1*-intron reporters is driven by the native *CUP1* promoter as it is induced by presence of copper and improves the number of transcripts from a single copy of the gene (50,196). The TranSceInt system allows rapid construction of *CUP1*-intron strains in a wild type background.

As demonstrated in Figure 5, insertion of the wild type intron MATa1 into the *CUP1* reporter allows for growth on copper media nearly to the level of intron-less *CUP1*. However, insertion of the pes1-5 intron into the reporter prevents growth on copper media indicating failure to splice. Spotting assay shows that MATa1 is still growing at 0.8 mM (stops showing growth ~1.0 mM, visible in Figure 6) while pes1-5 shows a growth defect after 0.1 mM and has no visible growth at 0.5mM.

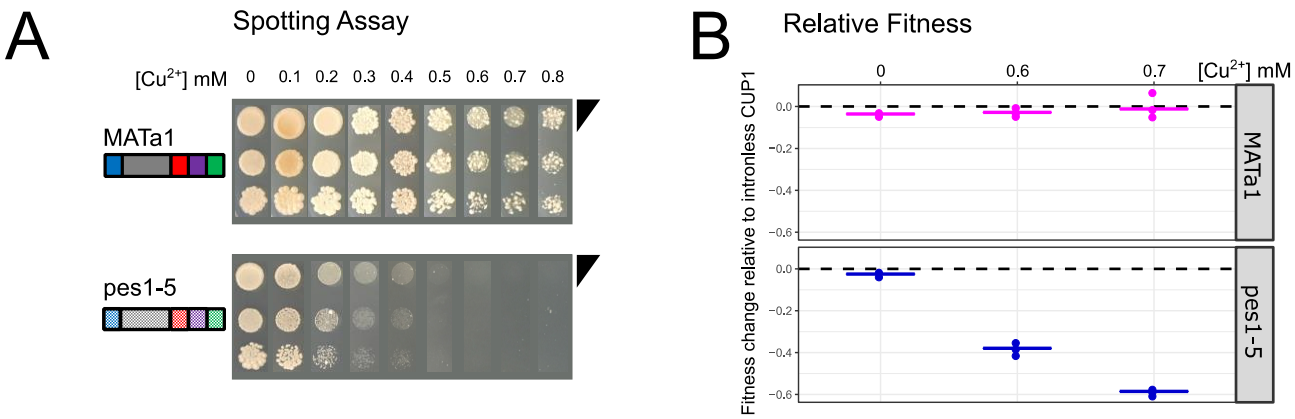


Figure 5. Data demonstrating MATa1 splices but pes1-5 does not splice. (**A**) Agar spotting assay on plates of different [Cu^{2+}] using two strains: one with the *CUP1*-MATA1 reporter and one with the *CUP1*-pes1-5 reporter. (**B**) The relative fitness metric is based on growth curve data collected in 96-well microtiter plates. The fold change of the area under the curve is calculated from a test strain versus a reference strain at each copper concentration, (n=3). Data is plotted with a bar through the median point. Here the reference strain contains the intron-less *CUP1* reporter in the same genetic context as the other strains (see Materials and Methods).

In order to detect minute changes in strain fitness as a proxy for splicing changes, I decided to use 96-well microtiter plates and define a metric based on the resulting growth curve data. Yeast microtiter plates have been used to study genetic interactions (197) and protein-protein interactions (198) in pooled screens. Some benefits of miniaturized liquid assays versus agar plate spotting are increased information regarding strain viability (e.g. growth rate, total OD), simplicity of the assay to set up, and ability to quantitatively resolve finer differences in growth using less of a toxic compound (199). The fitness change metric used in this report is defined as the fold change of the area under the curve (AUC) of each strain at each copper concentration relative to a reference strain's AUC at the same copper concentration (see Materials and

Methods). As the data shows (Figures 5B and 6B), this is a sensitive and complementary approach to the data obtained by agar spotting. Relative fitness experiments (Figure 5B) demonstrate the same growth pattern for reporter strains with MATa1 and pes1-5 introns as the agar spotting, with similar fitness at 0 mM Cu^{2+} (YPD) and a rapid drop in pes1-5 strain viability at each subsequent increase in copper concentration.

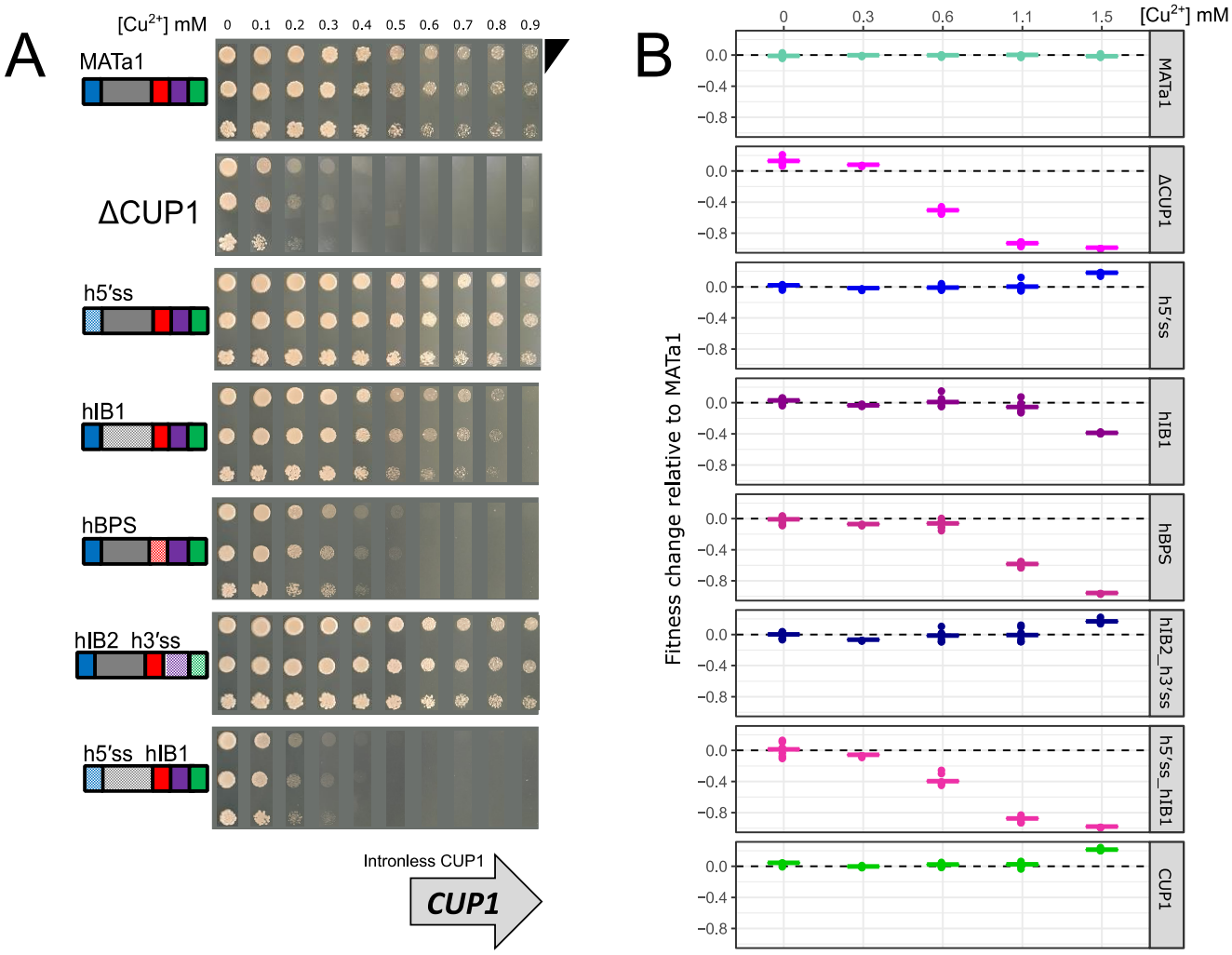


Figure 6. Splicing assay testing hybrid introns in wild type yeast. (**A**) Agar spotting assay on plates of different [Cu^{2+}] using one control intron (MATa1), a negative control strain *cup1*Δ for baseline copper tolerance, and five hybrid yeast-fungal introns (hatched module indicates pes1-5 sequence origin). (**B**) Relative fitness calculations from 96-well microtiter plate assays including the same strains from part D with the addition of intron-less *CUP1* as a positive control for highest possible copper tolerance from this reporter. Replicates: n=6 for 0, 0.6, and 1.1 mM; n=3 for 0.3 and 1.5 mM.

To determine the limitations of the yeast spliceosome when splicing the pes1-5 intron from *A. fumigatus*, I constructed another library of hybrid introns in *CUP1* by swapping modules between fungal pes1-5 and the yeast MATa1. Hybrid introns were created by replacing the sequences of the different modules of MATa1 with the corresponding pes1-5 sequence. For example, h5'ss indicates the heterologous 5'ss from pes1-5 (GTACGT) has been swapped in place of the MATa1 5'ss (GTATGT). Most yeast-fungal hybrid introns are the same size as MATa1 (52 bp) with the exception of hIB2_h3'ss (54 bp total), which has a longer IB2 region due to pes1-5 having a longer IB2 region. The strains containing the hybrid introns were tested using both a liquid spotting assay on copper media and liquid growth assays in a plate reader (Figures 6A and 6B). Figure 6A shows the different hybrid introns and strains constructed to the left and this annotation is mirrored in Figure 6B. A hatched color block indicates the module came from the *A. fumigatus* sequence while a solid color block indicates native *S. cerevisiae* MATa1 sequence. The liquid spotting assay results are aligned with the relative fitness results which are calculated with MATa1 as the reference strain, n=3-6 for each concentration (see Materials and Methods). The intron-less *CUP1* reporter and reporters containing the wild type MATa1, h5'ss, hIB1, hBPS, hIB2_h3'ss, or h5'ss_hIB1 were assayed on copper-containing media. The hIB2 and h3'ss were tested as a unit considering the 3'ss is used 50% of the time in native yeast introns (89) and wasn't considered likely to have a negative effect on splicing efficiency.

From these data, yeast growth was observed, indicating active splicing, for MATa1, h5′ss, and hIB2_h3′ss at all copper conditions tested, matching the *URA3* reporter results. The strain with the best fitness compared to MATa1 is the intron-less *CUP1* which is expected as introns inserted into non-native positions lowers gene expression in *S. cerevisiae* (46). Interestingly, h5′ss and hIB2_h3′ss actually have better growth metrics than wild type MATa1, as their fitness profiles (Figure 6B) are more similar to the strain containing the intron-less *CUP1*. This superior growth is correlated to a 1.25-fold increase of spliced RNA transcript relative to total *CUP1* transcript for each strain compared to the calibrator MATa1 strain (Figure 7). It is surprising that h5′ss has

superior growth to MATa1 given that the h5′ss sequence (GTACGT) is used sparingly in wild type yeast introns (89). Substitution of IB2 and 3′ss in hIB2_h3′ss also allowed for efficient splicing, which is unsurprising. Yeast has a single protein that binds to the IB2 region (MUD2) (38) compared to other fungi which have the U2AF65-U2AF35 heterodimer (38). The U2AF small subunit has two Zinc Fingers that specifically contact the 3′ss (200) in early spliceosome assembly. In contrast, *S. cerevisiae* 3′ss selection occurs during exon-ligation (201) and is also influenced by secondary structure (202). *CWC21* has been implicated in 3′ss selection by positioning the 3′ss during the transition to the second transesterification splicing step (203). There are critical spacing requirements between BPS and 3′ss (204) as 3′ss selection is based on a distance window from the intronic branch site (94) and increasing the length of the IB2 region may improve splicing efficiency (205). Considering that in *S. cerevisiae* alternative splicing occurs predominantly on the 3′ side of introns (206), it is possible that *S. cerevisiae* has a greater flexibility in accepting a variety of IB2 and 3′ss combinations provided they remain in its optimal effective distance. Given the average size of fungal introns, this is a likely possibility and bodes well for *S. cerevisiae* as an improved splicing host.

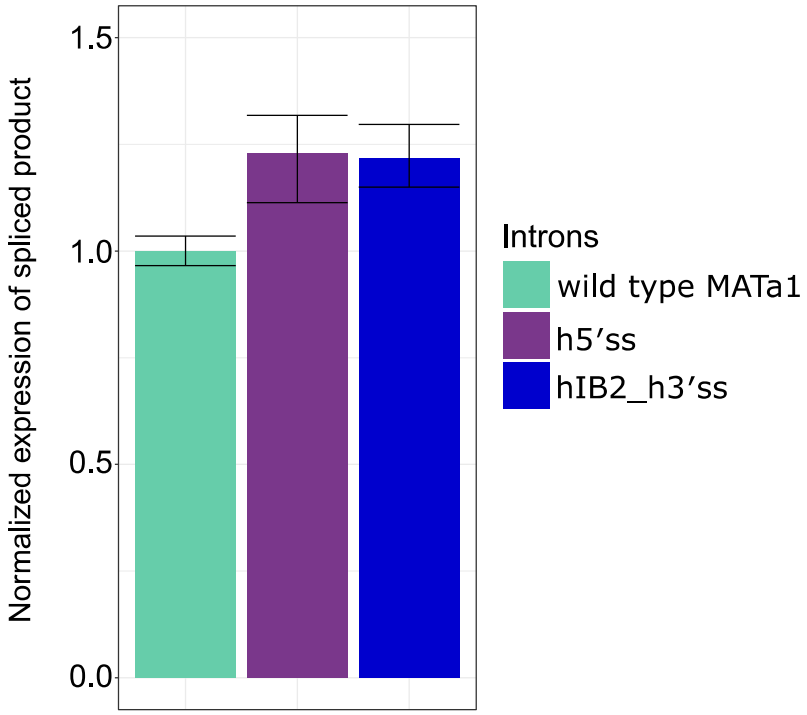


Figure 7. RT-qPCR results on hybrid introns that improve splicing. RT-qPCR results indicating that h5′ss and hIB2_h3′ss hybrid introns have higher splicing than the wild type MATa1 upon insertion into the *CUP1* reporter gene. Data is determined from three qPCR replicates using RNA extracted from a single biological replicate for each strain.

The strains with the weakest growth profiles are *cup1*Δ, h5′ss_hIB1, hBPS, and hIB1 in order from lowest copper resistance to highest. The h5′ss_hIB1 result is highly interesting as the h5′ss sequence improves splicing, and hIB1 still supports some splicing until the highest copper condition tested in this assay. It is noteworthy that the combination of h5′ss and hIB1 together cause the largest splicing defect. Replacing the h5′ss of pes1-5 with the MATa1 5′ss (pes1-5_C4T) does not restore splicing (Figure 8), likely due to the presence of hBPS. Previous work has shown that since MATa1 is the smallest intron in yeast, it does not tolerate deletions in the IB1 region (207), even when substituted with additional length in the IB2 region. This suggests a size requirement for efficient lariat formation during the first transesterification reaction. Taken together with these new findings suggests some secondary structure may be the cause of the splicing failure of h5′ss_hIB1. Particularly given the GC content of hIB1 is 50% to IB1's 20% and GC content influences splice site recognition (208). Higher GC content increases the stability of RNA secondary structures (209) and secondary structures influence alternative splicing (210,211) and

overall splicing efficiency (212). *Aspergillus* introns display stronger folding downstream of 5′ss (43), likely due to higher GC content in this region. In *S. cerevisiae*, *SUS1* alternative splicing regulation depends on secondary structure (213). I theorized an unfavorable secondary structure was forming due to the high GC content of the IB1 region, which ultimately changes intron size below the size limit imposed by MATa1. Perhaps the formation of a secondary structure in this intron plays a role in regulation of the native context of the gene. These results will be discussed more in Ch.5.

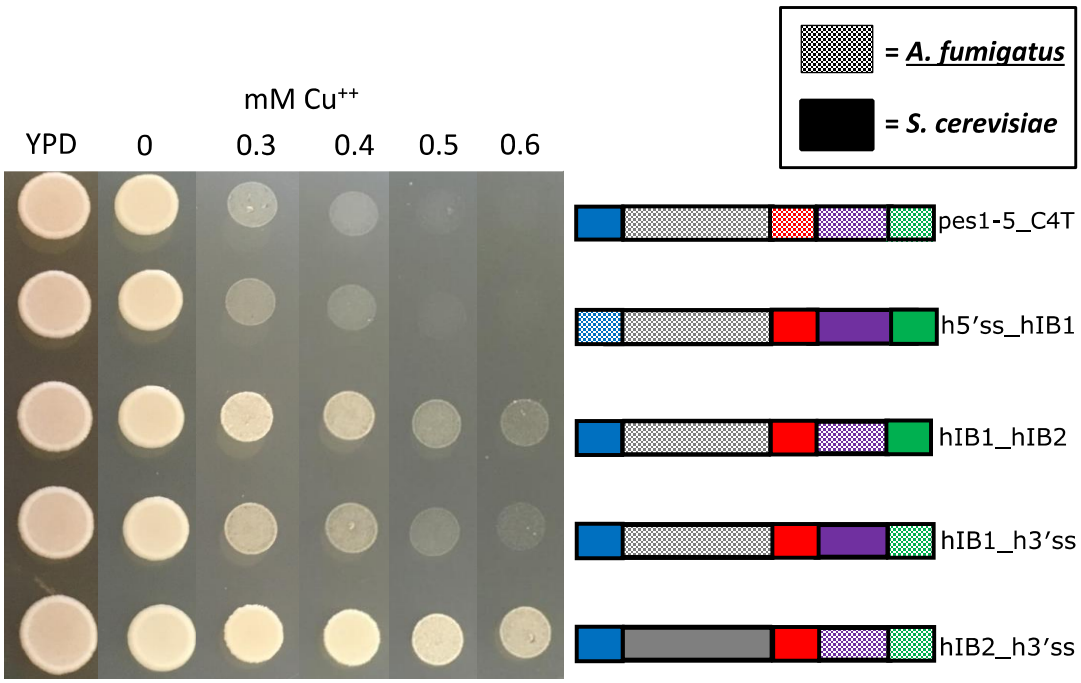


Figure 8. Liquid spotting assay of multiple intron module combinations. Additional liquid spotting assay of different intron variants on different concentration of copper agar plates. This further implicates hIB1 as a problematic module in combination with other pes1-5 modules.

It is clearly observed that hBPS and h5′ss_hIB1 are the two main failure modes that abolish splicing of pes1-5 in *S. cerevisiae*. This study fits in context of the literature as budding yeast introns are more information dense than other fungi such as *Aspergillus* (38), likely a result of their more rigid adherence to consensus sequence motifs and few accessory splicing factors. Therefore, splicing will be more sensitive to deviations from the consensus sequence and to the

formation of suboptimal secondary structures, which is observed using this set of yeast-fungal hybrid introns.

3.3 *HIS3* reporter

Now that the problematic modules of pes1-5 were identified by *URA3* and *CUP1* reporters, I had begun to engineer *S. cerevisiae* with enhanced functionality (see following chapters). However, it was desired to identify an additional splicing reporter with higher sensitivity than *URA3* to use in conjunction with *CUP1* for these engineering assays. Another sensitive genetic reporter that operated through a different mechanism than *CUP1* would strengthen any argument that the positive growth phenotypes observed were due to a positive enhancement of splicing and not due to a general increase in copper tolerance or cell fitness. For this purpose, I looked into *HIS3*. This gene encodes imidazoleglycerol-phosphate dehydratase, which is involved in histidine biosynthesis and is also partially deleted in all BY4741-derived designer deletion yeast strains (*his3*Δ1), leading to a non-functional gene product. The benefit of this as a reporter gene is that the enzyme His3 can be inhibited by addition of 3-Amino-1,2,4-triazole (3-AT) to the yeast media. Introns were cloned into the *HIS3* open reading frame. Strains that could tolerate higher concentrations of 3-AT were clearly producing higher amounts of His3 to overcome the 3-AT inhibition, indicating superior splicing of the *HIS3* transcript. Cloning the hBPS identified from previous assays was successful. Growth data is shown below at different concentrations of 3-AT. It is clear that a strain with wild type MATa1 cloned into the *HIS3* gene can tolerate greater amounts of 3-AT than *HIS3*-hBPS, matching the results of both the *URA3* and *CUP1* reporters. Unfortunately, the *HIS3* reporter is not as useful for studying the pes1-5 intron. Strains containing the *HIS3*-pes1-5 construct could not survive in histidine dropout media, even at 0 mM 3-AT (data not shown), emphasizing the severity of the splicing defect induced by pes1-5 and the great utility of the *CUP1* reporter.

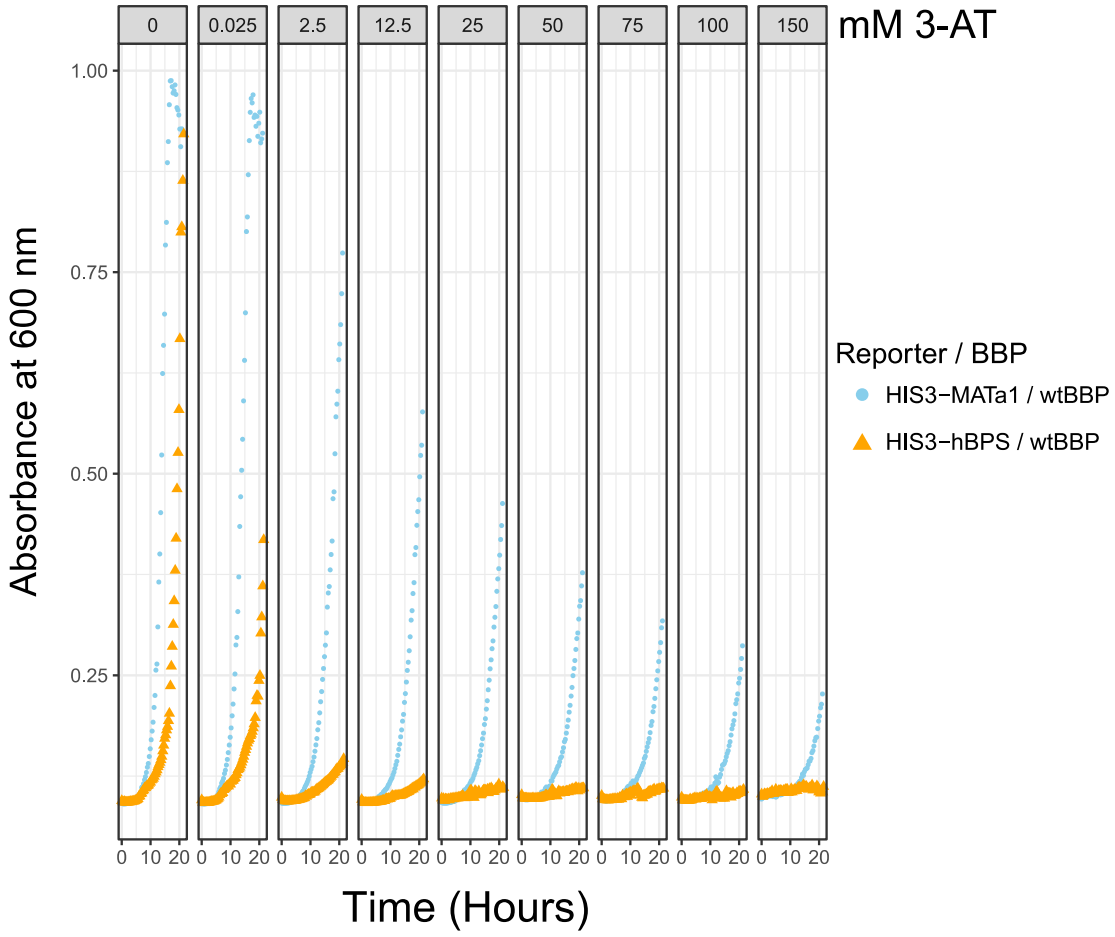


Figure 9. *HIS3* also serves as a sensitive splicing reporter. Yeast growth in microtiter 96-well plate. Raw absorbance is plotted against time. Each facet represents a different concentration of 3-AT. The two different reporter constructs are depicted by color and shape. For each growth curve an average of n=3 replicates is shown.

3.4 Materials and Methods

Strains

S. cerevisiae strains used in this study were derived from BY4741 (190). For use with the modified *CUP1* splicing reporter designed in this study, endogenous *CUP1* genes in BY4741 were removed. Both copies of the *CUP1* gene (*CUP1-1* and *CUP1-2*) on chromosome VIII were deleted simultaneously via targeted homologous recombination using a *LEU2* marker (strain S1, Appendix). This necessitated simultaneous deletion of the genes in-between the *CUP1* paralogs, *ARS810* and *YHR054C*, a putative replication origin and a putative protein of unknown function, respectively. However, the strain with these deletions displayed no fitness defects or changes to

growth rate when grown in YPD alongside BY4741 (data not shown). The genotype was confirmed by colony PCR. The phenotype change of S1 was confirmed by plating the wild type strain (BY4741) and the *cup1*Δ strain on plates dosed with different concentration of copper.

To generate the TranSceInt-receptor strain, a *cup1-ura3* landing pad was transformed into S1 to generate S2. The *cup1* gene is interrupted 50 bps from the start codon by the *URA3* marker flanked by I-SceI sites. Co-transformation of an *TDH3*promoter-*SCEI-PRM9*terminator expression cassette (as a PCR product) and double-stranded donor DNA intron with homology to the *cup1* ORF allows for efficient excision of the *URA3* marker and seamless replacement with any intron sequence. Post heat shock, strains were inoculated overnight in YPD and diluted in water before counter-selection of the *URA3* marker by plating on 5-Fluoroorotic Acid (5-FoA) containing media. To construct the strain with the intron-less single copy of *CUP1*, synthetic DNA was designed to restore the original ORF and was transformed using the TranSceInt system. This strain contains a single copy of *CUP1* in the same genetic context as the intron-containing reporter strains.

All yeast transformations in this study were carried out using the LiOAc/ssDNA method (214). Transformants were selected using leucine dropout media, uracil dropout media, and 5FoA media (0.8 g/L – 1 g/L). All genomic mutations were confirmed by colony PCR amplification of the region affected plus 50-500 bp of additional region upstream or downstream of the homology targets, to confirm specificity of the recombination event. PCR products were sequenced by Sanger sequencing (Laragen).

Plasmids

Standard cloning techniques were performed to maintain and propagate the plasmids in *E. coli* DH10b. pXP318 was a gift from Nancy Da Silva & Suzanne Sandmeyer (Addgene plasmid #26837) (65). pKR8 was a gift from Kevin Roy and Justin Smith (SGTC) that facilitated BspQI cloning. The hBPS variants plasmids were constructed using a BspQI cloning method (described

in more detail in Ch.4 Materials and Methods). A receptor plasmid was cloned containing two opposite BspQI cloning sites located 43 bp from the start codon of *CUP1*, a slight shift in intron position from the integrated reporter. Introns were ordered as 90-mers (IDT) with flanking BspQI sites. Upon digestion and ligation of the backbone vector and PCR-amplified intron oligos, the introns were seamlessly introduced into the *CUP1* ORF 43 bp downstream from the start codon.

Liquid Spotting Assays

Strains were inoculated into YPD or uracil dropout liquid media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. Optical Densities (ODs) were measured by spectrophotometer at 600nm for each strain and cells were normalized to each other by OD and then serially diluted into sterile water. 5 µL to 20 µL were spotted and plates were incubated at 30°C for 3-7 days before imaging.

Growth Curve Assays

Strains were inoculated into YPD, uracil dropout liquid media, or uracil and histidine double dropout media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. ODs were measured by spectrophotometer at 600nm and colonies were sub-cultured to OD 0.01 in 100 µL of culture per well in 96-well microtiter plates. Absorbance at 600nm was measured every 15 minutes over the course of 24 hours in a Tecan M200 Pro held at 30°C and shaking continuously alternating between linear shaking (before measurement) and orbital shaking (post-measurement) for each kinetic cycle.

Growth Curve Analysis

Raw data was exported to Excel from iControl Software for the Tecan M200 Pro and imported into R version 3.4.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Packages from the Tidyverse (Hadley Wickham (2017). tidyverse: Easily Install and

Load 'Tidyverse' Packages. R package version 1.1.1. https://CRAN.R-project.org/package=tidyverse) were used to manipulate the data for analysis and presentation with ggplot2 (Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009)). Fold change calculations were determined in an R script as follows: 1) the mean for the first 5 measurements of each strain replicate at each copper condition was calculated and subtracted from all absorbance readings to set the baseline of each growth curve to zero; 2) the area under the curve (AUC) was calculated as the sum of all adjusted absorbance readings using the trapezoidal rule; 3) the AUC of either the intron-less *CUP1* strain, *CUP1*-MATa1 strain, or *CUP1*-pes1-5 strain was used as the control value in the fold change equation: $(\mathrm{AUC}_{\mathrm{sample}} - \mathrm{AUC}_{\mathrm{reference}})/\mathrm{AUC}_{\mathrm{reference}}$.

RNA extraction

For hybrid intron strains, yeast cells were inoculated into YPD from agar plate colonies or frozen stocks and grown overnight in 5 mL of 5.0 mM Cu^{2+}, shaking at 30°C. The following day strains were then sub-cultured to an OD_{600} 0.1 in fresh 5 mL of YPD media containing 1.2mM Cu^{2+} and cultured for 7 hours. ODs were measured by spectrophotometer and cell count was normalized to 500 µL per OD_{600} 1.0. All samples were centrifuged, media was removed, and cell pellets were resuspended in 1 mL RNAprotect Cell Reagent before incubation at room temperature for 15 minutes. Cells for experiments were extracted fresh. Cell Reagent was removed by centrifugation. To digest the cell walls, cells were incubated for 1 hour at 37°C with gentle shaking in a zymolyase solution (100 µL RNase-free water, 10 µL zymolyase (Zymo Research), and 0.1 µL beta-Mercaptoethanol). All RNA was extracted using a slightly modified form of RNeasy Mini Plus Kit (Qiagen). On column DNase digestion (Qiagen) was extended to 1 hour. RNA was eluted from Qiagen columns using 100 µL 10 mM EDTA.

Quantitative RT-PCR experiments

Primers for RT-qPCR were designed using primer3 (215,216) or by hand when appropriate, such as for the splice junction primer. The exon2 product was amplified by primers that only bound in the second exon of the *CUP1*-intron reporter. This primer pair can amplify all intron-containing and spliced transcripts and represents the total *CUP1* transcript. For these strains, the spliced product was abundant enough to detect easily using total RNA. RNA samples were diluted to 1 ng/µL before adding 5 µL of these templates to each well on a 0.2 mL skirted 96-well PCR plate (Thermo Scientific, AB-0800/W). The template for each biological replicate was added to three separate wells to generate three qPCR replicates per biological replicate of a unique strain. A master mix for the Luna Universal One-Step RT-qPCR Kit (NEB, #E3005L) was generated for each primer pair used and 15 µL was aliquoted into each well and qPCR plates were sealed with Microseal 'B' seals (BioRad #MSB1001). Samples were run in a CFX96 Real-Time System with C1000 Thermal Cycler. The protocol for RT-qPCR was as follows: 55°C for 10 min (RT), 95°C for 1 min, [95°C for 10 sec, 60°C for 30 sec, plate read] x 45 cycles. Subsequent melt curve analysis protocol: 60°C incremented by 0.5°C to 95°C, holding temperature for 5 sec.

RT-qPCR analysis

Raw Cq values were exported to Excel from Bio-Rad CFX Manager 2.0 (determined by regression, algorithm similar to PCR Miner (217)) and then imported into R for analysis. ΔCq (spliced – reference) values for each replicate were calculated for each plate. Specifically, the Cq of the exon2 amplicon (representing total *CUP1* transcript) was subtracted from Cq of the spliced product. The mean ΔΔCq values were calculated from this pooled dataset of ΔCq values (218) using a two-sample magnitude bootstrap with replacement to generate the 95% confidence intervals (CIs) around the effect size. The mean of the resampled ΔCq values for the calibrator strain (containing the *CUP1*-MATa1 reporter) was subtracted from the mean of the resampled ΔCq values for the test strain (h5'ss or hIB2_h3'ss) to generate a simulated ΔΔCq. This process

was repeated for 10,000 iterations to generate the 95% confidence intervals. For each strain, the mean ΔΔCq value, upper CI value, and lower CI value were transformed into fold change values using $2^{-\Delta\Delta Cq}$ and then plotted in R.

4. ADDRESSING SPLICING FAILURE MODE #1

4.1 The Yeast-Fungal hybrid intron with heterologous BPS (hBPS) splices poorly

I first focused on rescuing the splicing defect observed in the heterologous/hybrid branchpoint site (hBPS) results from the *URA3, CUP1,* and *HIS3* assays (Ch.3). The BPS and its role in splicing are well-studied (219,220). Yeast introns predominantly contain the consensus module (UACUAAC) with few exceptions (89). The identified hBPS from pes1-5 (AGCUGAC) is not experimentally verified, which is a challenging feat in and of itself (221), and even computational BPS prediction is challenging, particularly in species with degenerate BPS (222). But the sequence fits the fungal consensus sequence RCURAY (38) and is present in many introns in fungal species, making it highly likely to be the BPS of the pes1-5 intron.

4.2 Mutation of the U2 snRNA for improved splicing

Therefore, I looked into strategies to recover splicing of this suboptimal BPS. In a compelling study, Charles Query's group has had success in restoring splicing of grossly-mutated BPSs by expressing an orthogonal U2 snRNA (223). The U2 snRNA binds to the BPS in order to facilitate the first transesterification by bulging out the catalytic adenosine (UACUAAC). The sequence of the U2 snRNA was examined and the binding site to the branchpoint site was determined to be the base pairs 34-39 in the 5′ end (Figure 10). When coupled to the hBPS sequence, the U2 snRNA binding motif does not match by 3 base pairs (positions 35, 38, and 39). Therefore, these base pairs were modified by incorporating the required changes into the sequence of a synthetic oligonucleotide used to amplify the U2 coding region by PCR before subsequent cloning of this mutant U2* snRNA into an expression vector.

I cloned both the wild type U2 RNA and U2* into a CEN/ARS and 2-micron plasmid for expression in yeast. I tried two methods of expressing the U2 sNRNA, one with a substituted terminator (PRM9t) for higher expression (54) and one with the native terminator. The substituted terminator shows no improvement in splicing as detected by copper assays. The native U2 promoter and terminator sequences were therefore used for expression to ensure that the transcript would undergo proper modification to produce a functional snRNA (224). These results also emphasized to me that I should use the native regulatory sequences to express splicing factors going forward. This expression cassette is denoted U2* and was cloned into a vector containing a low copy origin of replication (CEN/ARS) for replication in yeast.

After transforming the expression plasmid into the strain containing the hBPS and plating on copper media, it was observed that yeast containing U2* grew on concentrations of copper that strains with only a control plasmid did not. Therefore, the mutations in hBPS were suppressed by the addition of the orthogonal U2* and the splicing phenotype was restored.

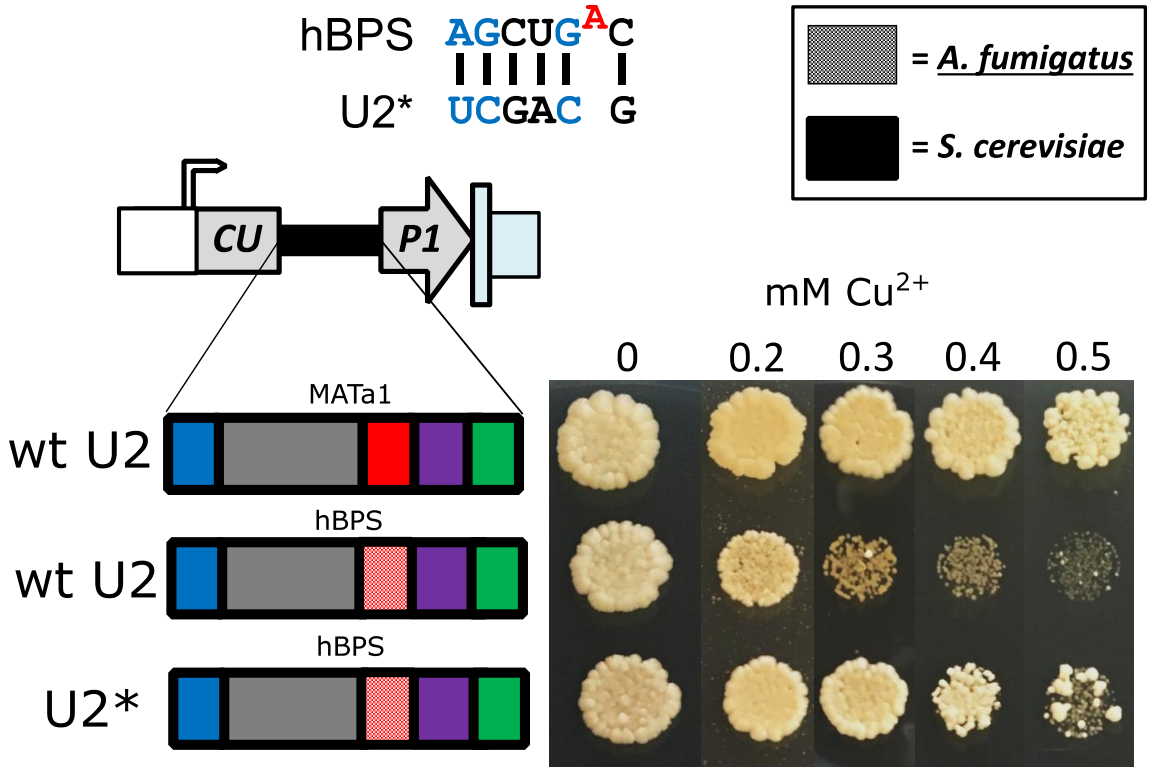


Figure 10. A mutant U2* snRNA can resuce splicing of hBPS. Demonstration of restored splicing of hBPS using an orthogonal mutant U2* snRNA that is able to bind perfectly to the suboptimal sequence. Liquid spotting assay results are shown with strains expressing wild type U2 snRNA or orthogonal U2* snRNA. The wild type MATa1 intron is represented by solid color blocks (each module is a different color). hBPS intron has a hatched block in BPS (red) and the sequence is shown at the top with blue nucleotides representing differences with the wild type sequences.

Despite the observed recovery of splicing, two problems were noted upon the addition of U2*: 1) splicing of the hBPS is still not as efficient as splicing of the native MATa1 as evidenced by the superior growth of MATa1 strains on higher concentrations of copper; and 2) the growth rate of the strain expressing U2* is reduced under all growth conditions compared to the strains expressing additional copies of U2 or the strains expressing the control plasmids. Although the specific reason for the reduction of growth rate is unclear, these observations are supported by Smith et al., who constructed a variety of U2 sequences to bind to grossly substituted branchpoint sites (223). This growth defect was exacerbated when the U2 snRNA was expressed on a 2-micron plasmid. Perhaps this is related to a defect in U2 remodeling (193) that prevents proper spliceosome activity on transcripts needed for growth. As multiple modified U2 snRNAs would be

required to rescue splicing of other potential heterologous BPS, these results indicate that strain fitness would be greatly compromised for a modest splicing improvement.

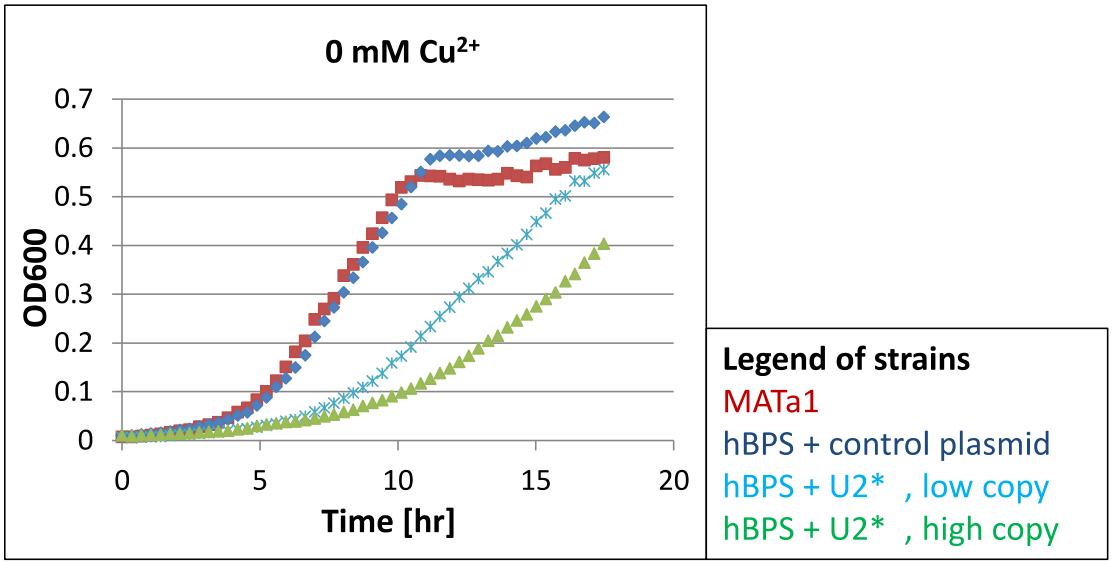


Figure 11. Overexpression of the mutant U2* snRNA causes a growth defect. Growth curves in uracil dropout media depicting the fitness defects caused by overexpression of mutant U2 snRNAs in yeast. Higher expression of U2* leads to a further decrease in fitness in uracil dropout media.

4.3 CRISPR-Cas9 strategy to modify essential splicing genes

Given the U2 snRNA results discussed above, I decided I needed to identify additional splicing targets that would better improve suboptimal intron splicing. This necessitated developing a technique to modify essential genes. Carly Bond (Yi Tang's lab) and I developed a CRISPR-Cas9 system for cloning-free and seamless genomic edits of essential splicing genes. Simultaneous transformation of the CRISPR-Cas9 plasmid components and donor DNA allows efficient essential gene modification through yeast homologous recombination of cut DNA (225) (see Materials and Methods for additional details). This technology was used to seamlessly modify multiple essential genes sequentially. The plasmid set (pCB30-G418, pCB32-hyg, and pCRCT-iCas9 from (226)) are amplified using a set of six 60-mer primers. 4 primers are universal and are used for every CRISPR-Cas9 assembly. Two 60-mer primers need to be designed for every genomic target (as they include the 20bp guide sequence for the sgRNA from CRISPy). Using

these six primers and the appropriate template for the desired antibiotic marker (pCB30 or pCB32), three PCR products are generated and co-transformed with the desired donor DNA. This technique allows for efficient replacement without direct selection for an integration event and antibiotic markers (from pCB30 and pCB32) can be alternated to allow for sequential edits.

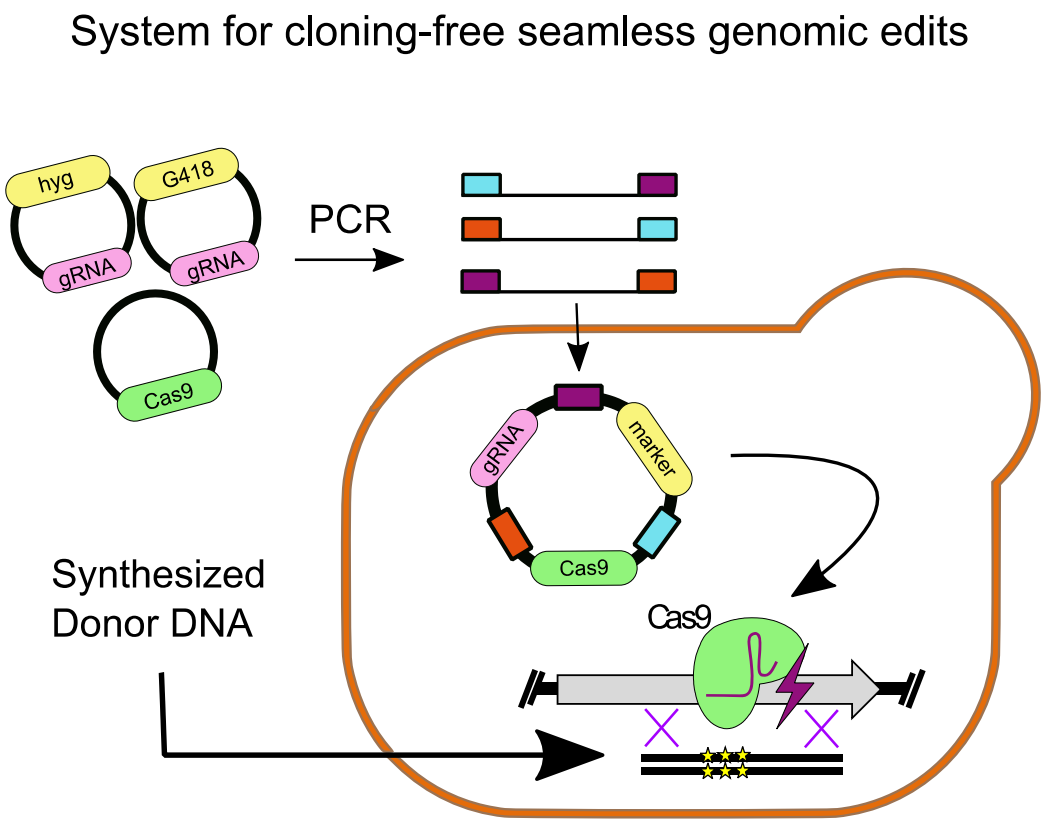


Figure 12. CRISPR-Cas9 system for editing essential yeast genes. Schematic detailing the flexible CRISPR-Cas9 strategy for seamless genomic edits. A template with the desired antibiotic resistance is amplified (2 amplicons) as well as the Cas9 sequence using specially designed primers with homology for assembly (third amplicon). These three amplicons are co-transformed with donor DNA of choice for gene editing using yeast homologous recombination stimulated by a Cas9-mediated double-stranded break.

4.4 Modification of RNA helicases to decrease splicing fidelity

The spliceosomal RNA helicases/ATPases act at multiple stages of spliceosome assembly and activity, enabling complex conformational rearrangements (120). Specific interactions between the spliceosome and its intron substrate are “proofread” by these enzymes, which, upon satisfaction, hydrolyze ATP to rearrange the spliceosome to the next conformation needed to splice. If a suboptimal interaction is detected, the RNA helicase sends the substrate to be

discarded and it does not proceed to the next step in splicing. These kinetic proofreading mechanisms are crucial to quality control of the spliceosome and a major reason for its incredible specificity in splicing (227). These enzymes are fairly well conserved across species, although higher order organisms have additional helicases not present in yeast (228). Interestingly, proofreading mechanisms are more stringent in yeast than high eukaryotes, which is likely due to the fact that there is less variability in RNA duplexes formed during yeast splicing due to the predominance of consensus sequences.

The DEAD-box RNA helicase *PRP5* remodels the U2 snRNP during pre-spliceosome formation through an ATP-dependent mechanism (229). Specifically, it disrupts the branchpoint-interacting stem loop of the U2 snRNA, thus promoting binding of U2 snRNA to the BPS (193). It has been demonstrated that pseudouridines in U2 snRNA alters the binding/ATPase activity of Prp5 (230). Prp5 mutations that destabilize the interaction with U2 snRNA have been shown to suppress BPS mutations by allowing more suboptimal substrates to continue splicing (229). *PRP5* is an integral component of the spliceosome and is thought to have additional ATP-dependent and ATP-independent functions throughout splicing (231). For example, Prp5 is regulated by Hub1-dependent negative feedback loop (232) and Prp5 physically interacts with SF3b1/Hsh155 during branchsite selection with implications for a mechanistic understanding for certain human diseases (233).

Considering that these RNA helicases discriminate against suboptimal introns substrates (234), they represent important targets for spliceosome engineering. Mutation of these splicing factors can decrease intron substrate flux to non-productive pathways, therefore allowing more intron-spliceosome complexes to continue through the productive splicing pathway. Prp5-N399D was identified as a suppressor of multiple splicing mutations at the BPS (235). This enzyme was shown to raise copper tolerance *in vivo* and had reduced ATP usage *in vitro* suggesting slowed kinetic proofreading which enabled more suboptimal BPS to be spliced (235).

As *PRP5* is an essential gene due to its critical roles in spliceosome function, we needed a powerful strategy to be able to quickly introduce the desired mutation. Thus, we utilized our CRISPR-Cas9 engineering (described previously) to introduce the N399D mutation into the genomic copy of *PRP5* in a strain containing the integrated *CUP1*-hBPS reporter. This new mutant strain was assayed against its parent strain and a control using a liquid spotting assay depicted in Figure 13. The results indicate a minor improvement in splicing, using improved yeast growth as a proxy for splicing.

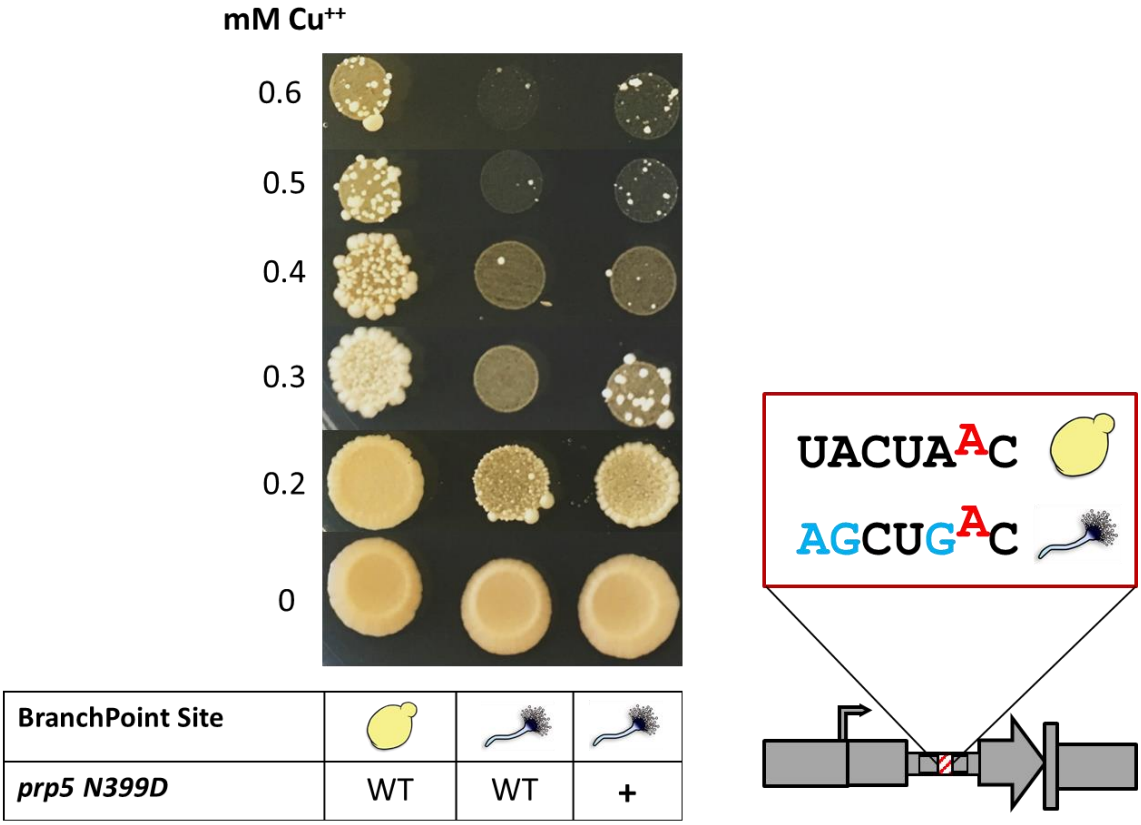


Figure 13. *prp5-N399D* enables minor rescue of hBPS splicing. Liquid spotting assay depicting wild type yeast containing the *CUP1*-MATa1 reporter (lane 1), wild type yeast with the *CUP1*-hBPS reporter (lane 2), and yeast with *prp5*-N399D mutation and *CUP1*-hBPS reporter. Improved growth indicates improved splicing of the hBPS, shown in the cartoon legend (right).

4.5 Enhance BPS recognition through mutation of BranchBinding Protein

Given the growth defects of U2* snRNA, and the mild success of *prp5*-N399D for rescuing splicing, we desired a more universal and effective solution. Therefore, we looked to the protein that U2 snRNP displaces during spliceosome assembly: BranchBinding Protein (BBP).

BBP, encoded by *MSL5*, is the first splicing factor to contact the BPS of the intron during spliceosome assembly. It is involved with the formation of Complex E in the *S. cerevisiae* spliceosome but is released immediately after pre-spliceosome formation (236) during incorporation of the U2 snRNP. BBP is essential for stable assembly of the spliceosome *in vivo* and forms a heterodimer with MUD2. The KH-QUA2 domain of BBP is responsible for binding to the BPS of an intron (237). The yeast BPS evolved to preferentially bind to the consensus BPS UACUAAC (catalytic adenosine underlined), which is the BPS of MATa1.

I performed a sequence alignment of the BBPs from different fungal species: *Saccharomyces cerevisiae* (Sce), *Schizosaccharomyces pombe* (Spo), *Aspergillus fumigatus* (Afu), *Penicillium oxalicum* (Pox), *Fusarium oxysporum* (Fox) (Figure 14) and the protein sequences of the KH-QUA2 domain of BBP proteins are displayed. Not surprisingly, the KH-QUA2 domain of *S. cerevisiae* is most dissimilar from every other sequence, with the largest differences in the RNA binding channels of the protein. These results are in line with the preference of the budding yeast spliceosome for the consensus BPS, which is logical as budding yeast lacks the SR proteins to aid the in the recognition of suboptimal splice site sequences.

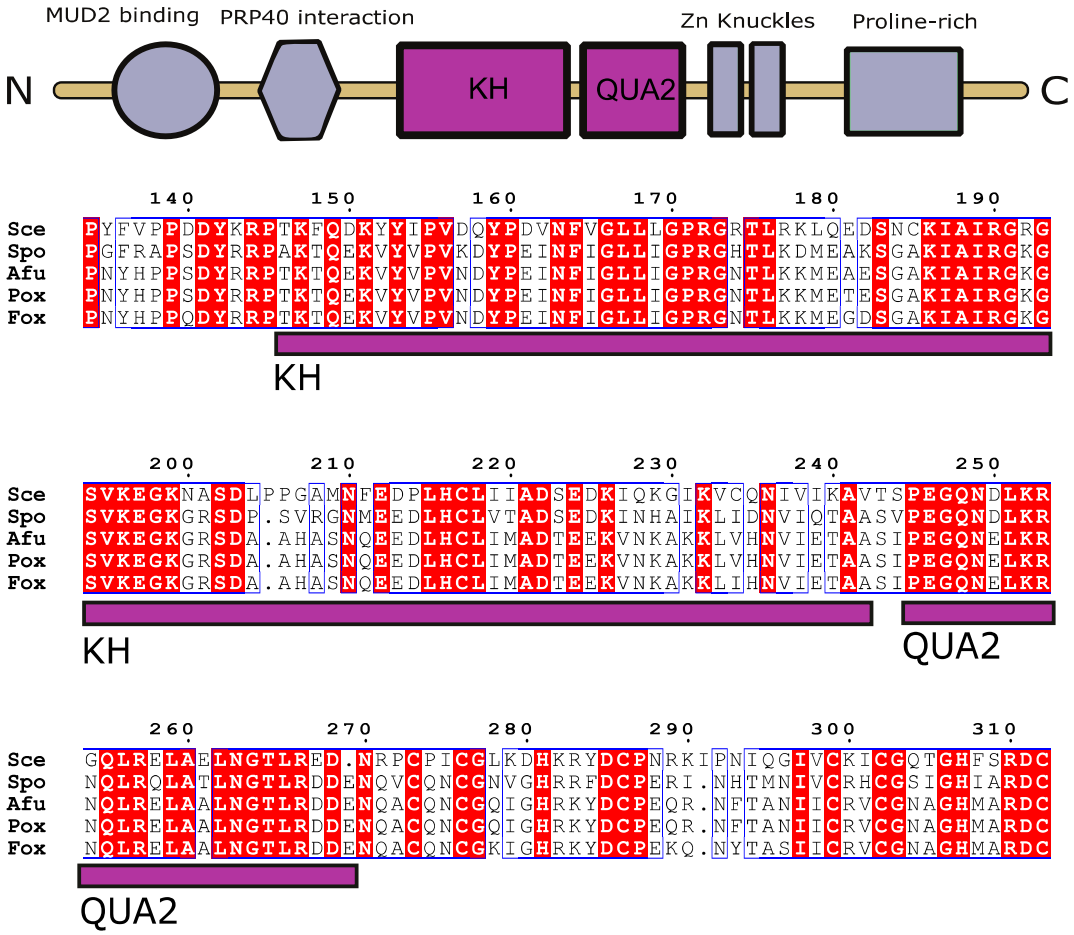


Figure 14. Protein domain alignment of fungal BBPs. Cartoon schematic detailing the layout of the different domains of BBP. (Bottom) Alignment of the KH-QUA2 domain responsible for binding to the branchpoint site. Protein sequences are from *Saccharomyces cerevisiae* (Sce), *Schizosaccharomyces pombe* (Spo), *Aspergillus fumigatus* (Afu), *Penicillium oxalicum* (Pox), and *Fusarium oxysporum* (Fox). The purple box underlines the specific sequences of KH and QUA2 sub-domains. Residues boxed in red demonstrate conserved sequence.

In order to study the BBP of different fungi, we considered direct heterologous expression of *A. fumigatus* BBP, but ultimately hypothesized that the protein would not function efficiently with the budding yeast spliceosome. In order to function in spliceosome assembly, BBP forms a heterodimer with the nonessential protein Mud2, and yeast Mud2 differs greatly compared to its homologs in other fungal species (the U2AF65-U2AF35 heterodimer) (238,239) (Appenidx), particularly because *S. cerevisiae* has no direct homolog to U2AF35 (103). Mutation in the *MSL5* domain responsible for its interaction with *MUD2* impairs spliceosome assembly, particularly for introns with BPS mutations (240). We theorized that domains that allow BBP to interact with the

other splicing factors would not be present in the fungal homologs as they evolved different points of contact for interaction with their own splicing machinery (the U2AF heterodimer). This is demonstrated by the difference in BBP's Mud2 contact compared to other species' BBP N-terminal regions (Appendix). This is also supported by the sequence alignment data of Mud2 to U2AF65 (Appendix) which has low homology in the BBP interaction domain. We hypothesized that hBPS doesn't splice because *S. cerevisiae* BBP does not recognize deviant sequences from the consensus and any spliceosome assembly is unstable on these transcripts. Therefore, we decided to replace the KH-QUA2 domain with the homologous domain from *A. fumigatus* and cloned the expression cassette containing this chimeric yeast-fungal BBP sequence into a pXP318 vector (see Materials and Methods). This chimeric protein is expressed using the native promoter and terminator of *MSL5*.

4.6 Combinatorial testing of mutant splicing factors

At this point, three splicing mutations had been constructed: U2* snRNA, *prp5*-N399D, and mBBP (*msl5*-AfKH-QUA2). We decided to test these mutations together in the search for positive synergistic mutations and to also test orthogonal expression of mBBP versus integrated replacement of the genomic copy.

The first figure shows orthogonal expression of U2*, wild type BBP, and mBBP versus an empty vector control. All strains are wild type *PRP5* and plated on uracil dropout agar media containing different concentrations of copper. The clearest improvement is with expression of orthogonal U2* snRNA, which matches the results described earlier. Expression of mBBP appears to offer a slight growth benefit, as colony density is thicker at 0.4 mM compared to the control. Overexpression of wild type BBP appears to have a slight negative effect, as colony density is decreased compared to empty vector at 0.4 mM.

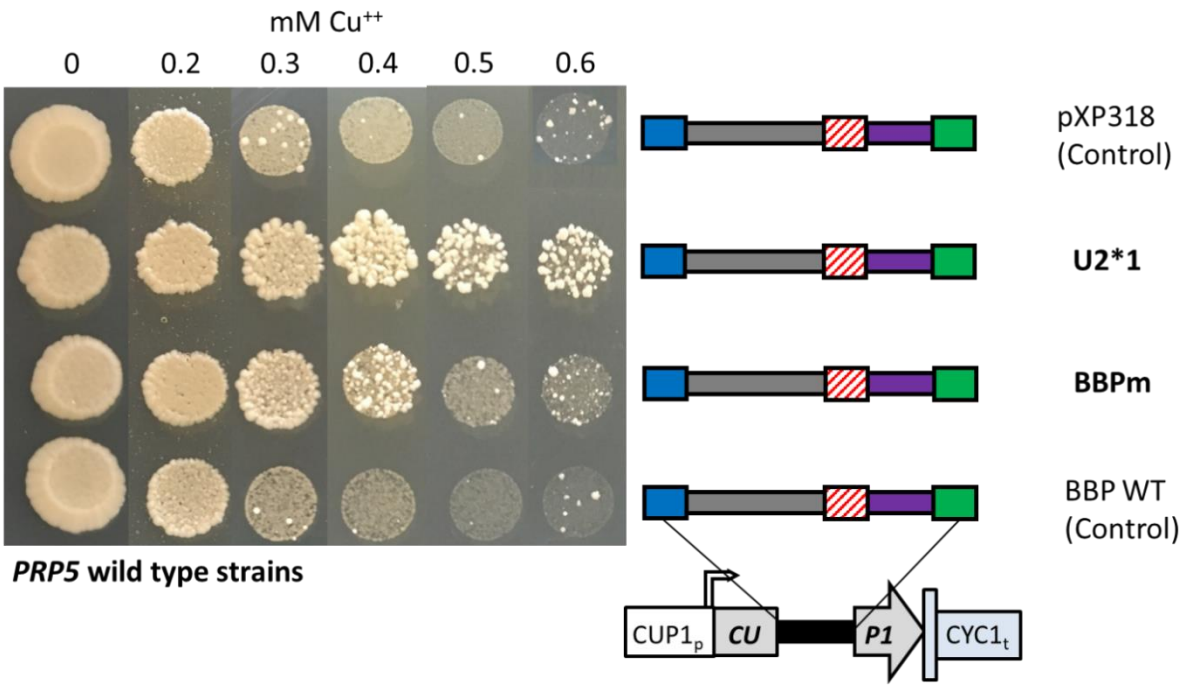


Figure 15. Testing U2* and mBBP in wild type *PRP5* background. Liquid spotting assay testing plasmid expression of U2* snRNA (U2*1, row 2) and mBBP mutation (BBPm, row 3) versus two control strains: (empty vector, row 1) and wild type BBP overexpression (row 4). All strains have only the genomic copy of wild type *PRP5.* Higher colony density and larger colonies indicate improved splicing of the integrated *CUP1*-hBPS reporter.

To test for synergy with these mutations in a *prp5-N399D* background, the same plasmids used above were transformed into this strain background (built in Ch. 4.4) and assayed on copper agar plates. The empty vector strain grows slightly better than in the first assay, reproducing the minor fitness improvement observed with *prp5*-N399D. However, expression of orthogonal U2* shows reduced colony growth compared to overexpression in a *PRP5* wild type background. This indicates that these two splicing mutations do not stack additively. Given the intimate relationship between *PRP5* and U2 snRNA, this is not a surprising result. Mutations that alter the function and stability of these splicing factors may negatively impact overall spliceosome activity. Perhaps the stabilized U2* snRNA cannot be properly remodeled by *prp5*-N399D prior to intron-lariat formation and stalled spliceosomes are discarded. Interestingly, there are appears to be a positive synergy between overexpression of mBBP and the genomic copy of *prp5*-N399D, visible starting at 0.4 mM. This effect is not witnessed upon overexpression of wild type BBP indicating that this

phenomenon is due to the *A. fumigatus* KH-QUA2 domain of mBBP. Overexpression of wild type BBP grows worse than the empty vector strain again, indicating that too high of expression of this splicing factor may induce a growth defect.

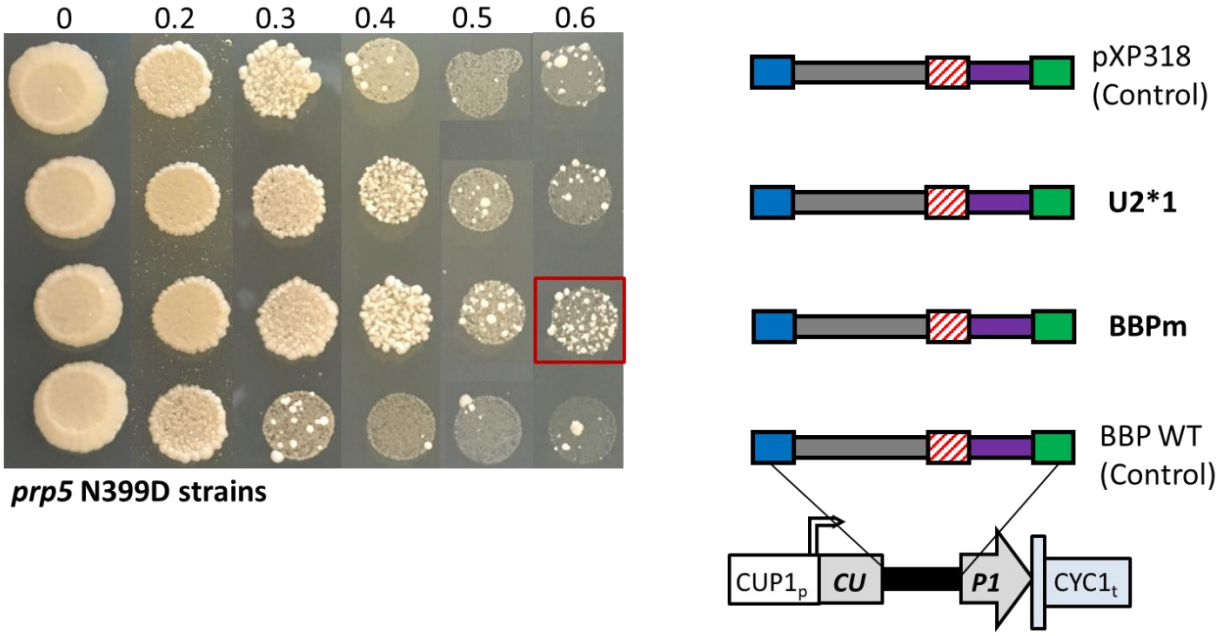
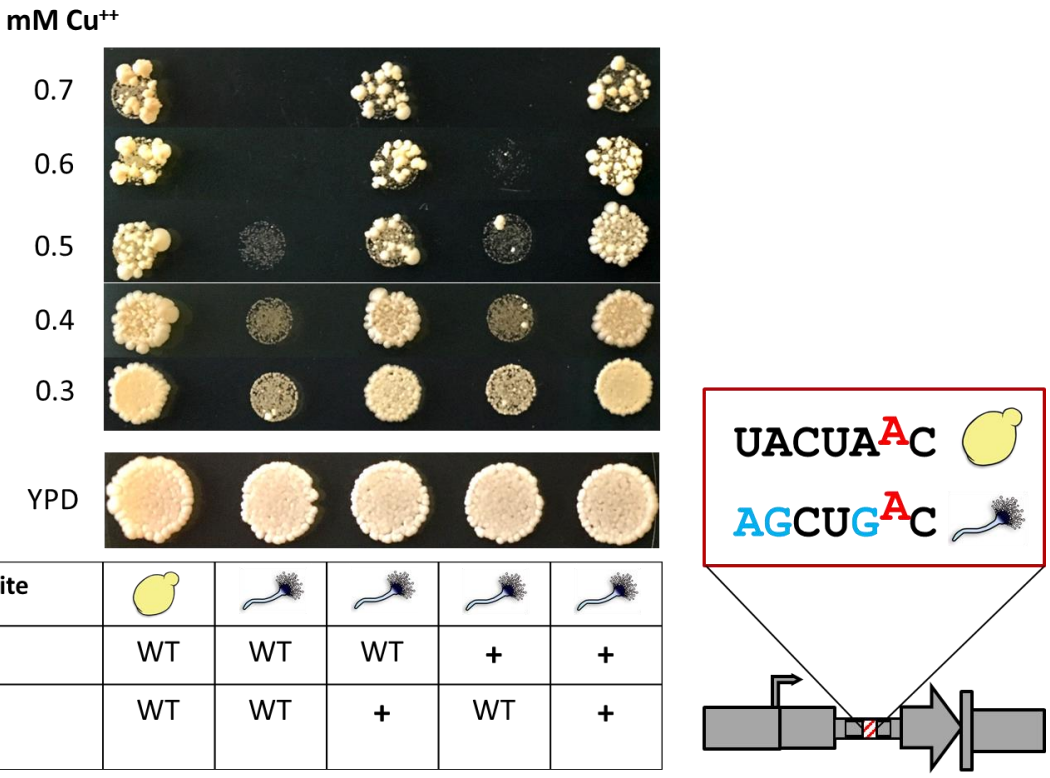


Figure 16. Testing U2* and mBBP in *prp5*-N399D background. Liquid spotting assay testing plasmid expression of U2* snRNA (U2*1, row 2) and mBBP mutation (BBPm, row 3) versus two control strains: (empty vector, row 1) and wild type BBP overexpression (row 4). All strains have the mutated genomic copy of *prp5*-N399D. Higher colony density and larger colonies indicate improved splicing of the integrated *CUP1*-hBPS reporter.

We were interested in further exploring the positive synergy between *prp5*-N399D and mBBP as well as different methods of expression of chimeric mBBP in greater detail. Therefore, mBBP was integrated (replacing the native BBP copy) into a wild type strain with the *CUP1*-hBPS reporter and also into the strain background containing *prp5*-N399D and the *CUP1*-hBPS reporter. All integrations were performed using the CRISPR-Cas9 strategy outlined earlier (Ch. 4.3). These integrated strains were tested against multiple control strains with integrated splicing factors. The wild type yeast background with *CUP1*-MATa1 was used as a positive control. As expected a wild type strain with the *CUP1*-MATa1 reporter grows well at the conditions of copper tested. The wild type strain containing *CUP1*-hBPS grows poorly at 0.3 mM and stops growing after 0.5 mM.

Surprisingly, the genomic copy of mBBP grows nearly identical to wild type *CUP1*-MATa1 at these copper conditions and shows much better growth results than orthogonal expression of mBBP in the assay performed previously. This is likely related to the growth defect caused by overexpression of BBP observed in the previous assays. The genomic replacement *prp5*-N399D again provides a very slight increase in phenotype compared to a wild type spliceosome. The dual mutant strain (*prp5*-N399D and mBBP) grows similarly to the mBBP integrated strain, and the wild type spliceosome strain with *CUP1*-MATa1 reporter indicating similar splicing levels.



BranchPoint Site					
prp5 N399D	WT	WT	WT	+	+
BBPm (AfKH-QUA2)	WT	WT	+	WT	+

Figure 17. Integrated mBBP restores nearly wild type growth. Liquid spotting assay comparing wild type spliceosome with individual genomic mutations *prp5*-N399D, mBBP and combined mutants. (Col. 1) The positive control with wild type spliceosome and *CUP1*-MATa1. (Col. 2) Negative control with wild type spliceosome and *CUP1*-hBPS. (Col. 3) *CUP1*-hBPS with genomic replacement of BBP with mBBP. (Col. 4) *CUP1*-hBPS strain with *prp5*-N399D mutation. (Col. 5) *CUP1*-hBPS strain with both genomic *prp5*-N399D and genomic mBBP. Higher colony density indicates superior splicing of the *CUP1* reporter.

This experiment was repeated with higher copper concentrations to see if any difference could eventually be detected between mBBP and *prp5*-N399D combinations. The results below

indicate that there is not a positive synergy when both mBBP and *prp5*-N399D are present as genomic replacements of their wild type counterparts. The lack of positive synergy across U2* snRNA, mBBP, and *prp5*-N399D encouraged me to pick mutation combinations carefully going forward as multiple splicing factor mutations may inhibit colony fitness. There is a careful balance between expanding spliceosome functionality and hampering its catalytic abilities. Given the multiple roles of *PRP5*, it is not surprising that mutation of this protein may work against spliceosome efficiency even if more suboptimal transcripts are kept in the productive splicing pathway. But the growth improvement offered by integrated mBBP (in place of wild type BBP) was impressive and unexpected. This result was pursued in greater detail and shows that the right mutation in a single protein has the capability of changing splicing dramatically.

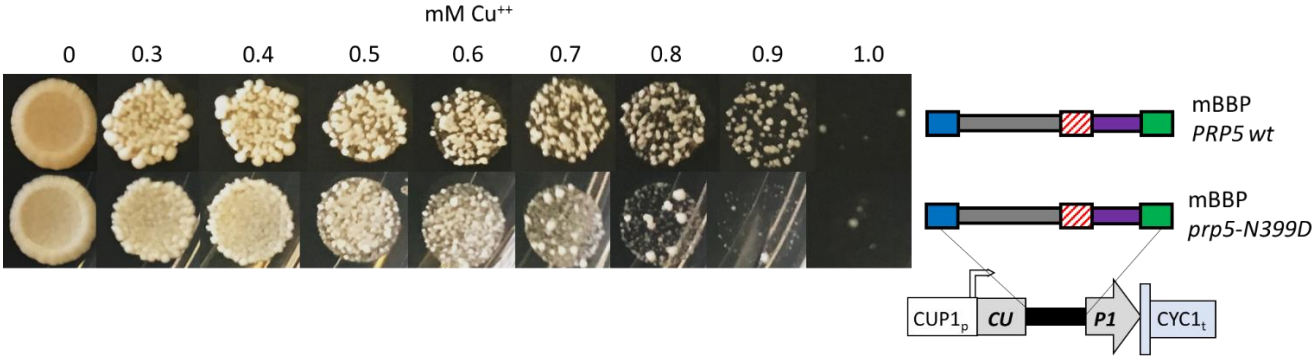


Figure 18. Integrated mBBP and *prp5*-N399D have negative synergy on cell fitness. Liquid spotting assay repeating previous strains tested, but at higher conditions of copper. (Row 1) mBBP replaces the genomic copy of BBP, but *PRP5* is wild type. (Row 2) Both mBBP and *prp5*-N399D replace their wild type alleles in the genome.

4.7 Further testing of genomic mBBP

Replacement of the wild type KH-QUA2 domain with the *A. fumigatus* homologous domain does not appear to have any negative effect on yeast growth and fitness. Yeast strains were grown at 30°C in the Tecan plate reader, but no noticeable growth defect was observed (Figure 19). To check for temperature sensitivity that may be caused by this mutation, cultures containing wild type BBP or mBBP were spotted onto YPD agar plates and incubated at 16°C, 25°C, 30°C,

and 37°C. No difference between the strains is visible at any of these temperatures (Appendix). It is logical considering functionally important residues in an essential protein will be conserved across related species and expanding BPS recognition still enables recognition of the consensus BPS. RNA-binding proteins have evolved to be modular (241) and a seamless swap of a single domain would be expected to alter function without altering protein stability.

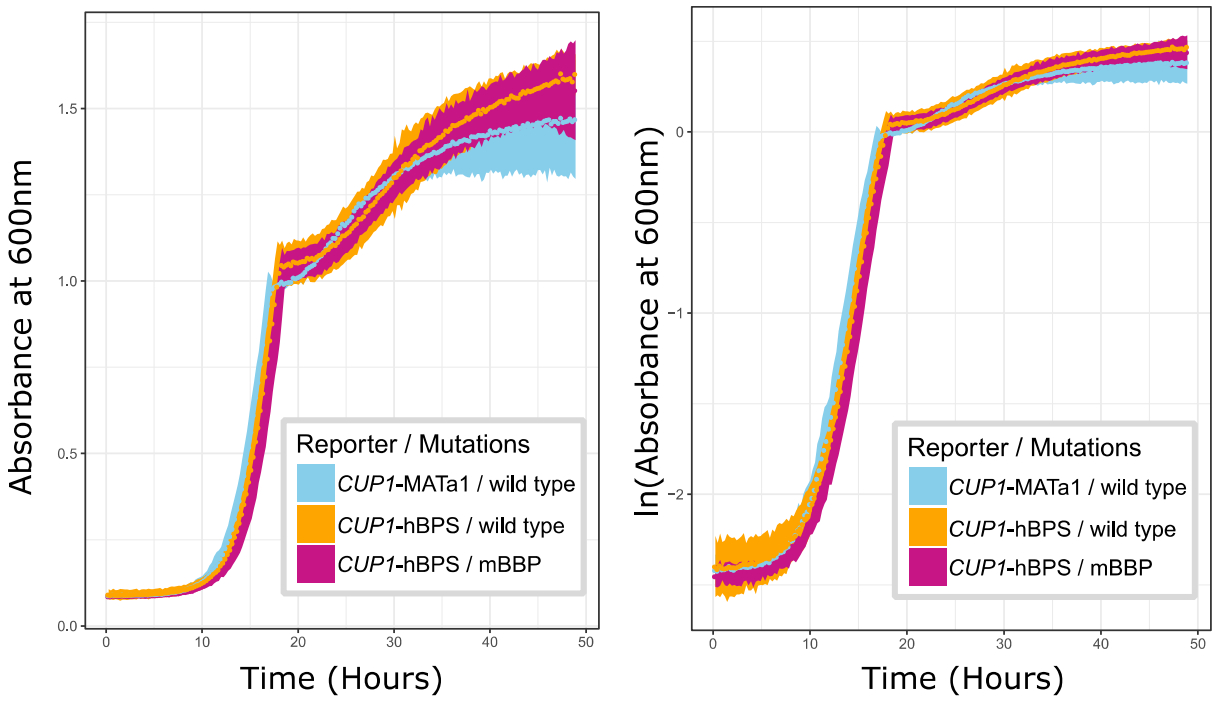


Figure 19. Integrated mBBP does not cause growth defects. Microtiter growth assay showing the growth curve of three yeast strains with wild type BBP or chimeric mBBP. No significant difference is observed measuring the absorbance (left) and no growth rate differences are seen based on the equivalent slopes in the natural log plot (right). The ribbons represent 95% confidence intervals of each measurement.

Despite not inducing any detectable effect on yeast fitness, the chimeric yeast-fungal BBP (mBBP) enabled strong splicing improvements of the hybrid intron with the heterologous BPS (hBPS) as demonstrated by yeast growth assays and RT-qPCR data. Figure 20 shows the liquid spotting assay and relative fitness results of strains containing the wild type MATa1 intron and wild type BBP (wtBBP), the hBPS hybrid intron and wt BBP, and the hBPS intron and chimeric mBBP. hBPS with wtBBP does not grow past 0.4mM and shows a steady decline in fitness relative to the MATa1 strain at every copper condition tested. However, once the mBBP is introduced into

the strain containing hBPS, growth is dramatically recovered to near wild type levels. However, the growth data indicates incomplete splicing compared to MATa1, as the mBBP strain only grows to 0.9mM (versus 1.0mM for MATa1 strain) in Figure 20A and more obviously shows a reduced relative fitness at 0.6mM in Figure 20B.

Considering that BBP is only involved in the first steps of spliceosome assembly emphasizes that Complex E formation is essential to proper splicing of suboptimal intron sequences in budding yeast. These results emphasize the importance of BBP for splicing of suboptimal transcripts as *msl5* mutants are particularly splicing impaired if the intron contains a suboptimal BPS (240,242). The remaining splicing defect detected in these results is likely due to the fact that the aberrant hBPS is not efficiently bound to U2 snRNA or otherwise destabilized leading up to the first splicing reaction.

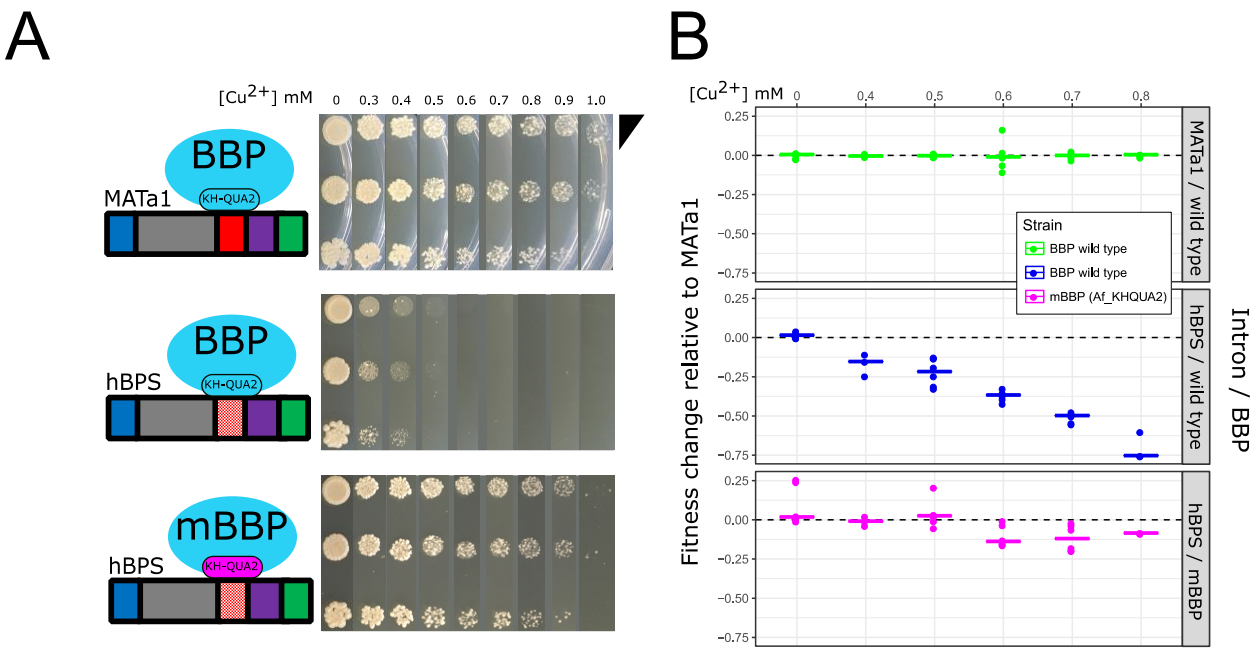


Figure 20. Integrated mBBP improves fitness on copper. The chimeric yeast-fungal mBBP has a significant positive impact of splicing the suboptimal hybrid intron hBPS. (**A**) Cartoon schematic (left) of wild type BBP (all blue) assembled on either the MATa1 intron (UACUAAC) or hBPS hybrid intron (AGCUGAC) and mBBP (magenta KH-QUA2 domain) assembled on hBPS hybrid intron. The liquid spotting assay (right) shows serial dilutions of these three strains spotted on agar plates with different copper concentrations. (**B**) Relative fitness calculations for the same three strains aligned with the cartoon schematic in A.

Orthogonal over-expression experiments reveal the importance of maintaining the correct expression levels of splicing factors because optimal growth is obtained using the strain containing mBBP at its genomic locus (Figure 21). Co-expression of mBBP and wtBBP still shows some growth improvement however, which is encouraging for studying orthogonal BBP systems. Overexpressing wtBBP does not offer any improvement in splicing, indicating that the function of mBBP itself is required to see splicing improvements. These results are in agreement with what was observed in previous assays and underscores the important of maintaining optimal expression of splicing factors.

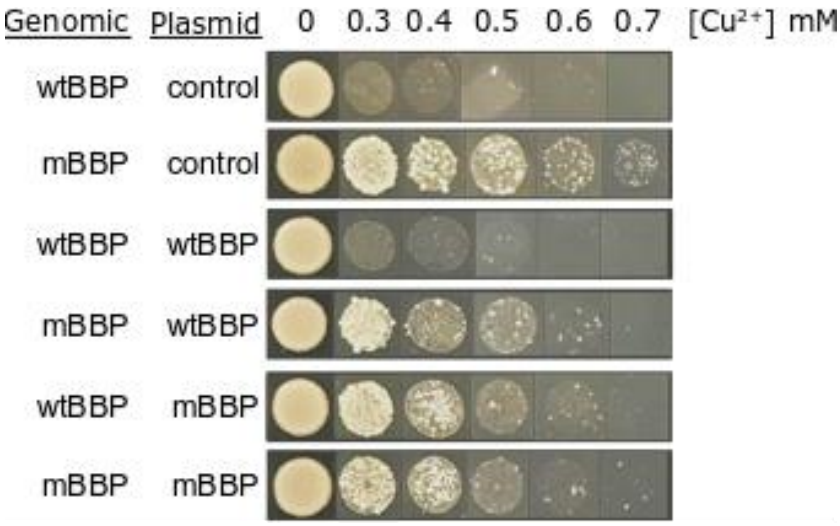


Figure 21. Overexpression of BBP and mBBP decrease cell fitness. Liquid spotting assay for overexpression of either wtBBP or mBBP on CEN/ARS low copy vectors. Strains were plated on -uracil agar plates containing different concentrations of copper.

The improvement is splicing is confirmed by RT-qPCR in Figure 22. To determine the percentage of each transcript, relative incidents were calculated and approximately add to 100% (see Materials and Methods). The leftmost facet shows the transcript levels of *CUP1* in a wild type BBP background. The percent of *CUP1* transcript containing intron is much higher than the percentage spliced (25%). However, upon introduction of mBBP, the percent spliced increases to 50%, a 2-fold increase. And as revealed by the fitness data, splicing of wild type MATa1 is still

much higher at 70%. It is interesting that such a strong phenotypic difference results from the detected change in relative amounts of transcript. But it is important to note that this transcript difference is reflected in the fitness results derived from liquid culture assays, reinforcing their sensitivity.

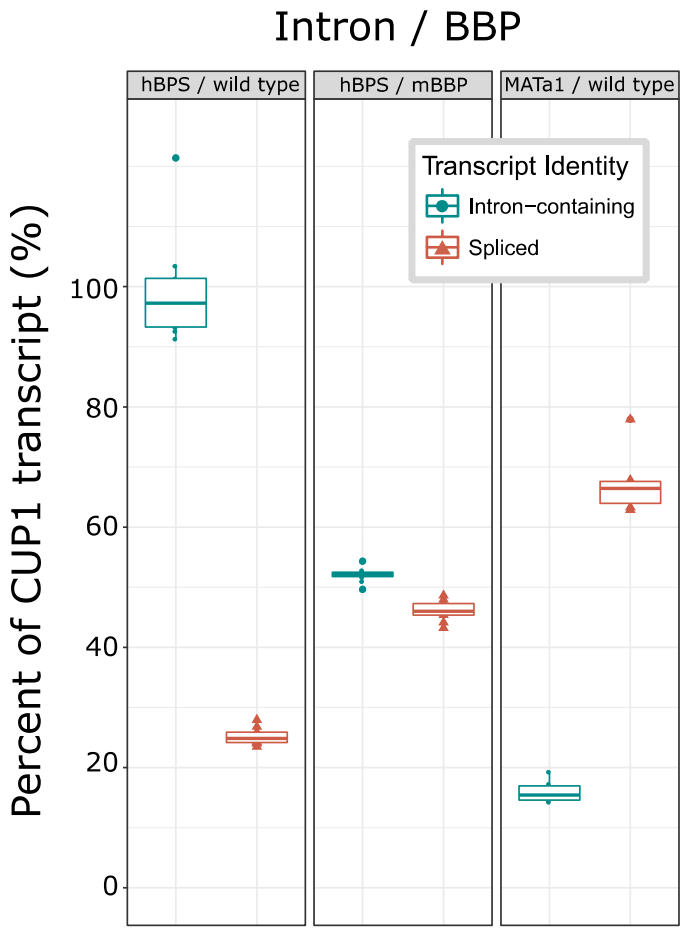


Figure 22. mBBP increases the amount of spliced *CUP1*. RT-qPCR data presented as the percentage of *CUP1* transcript that is either intron-containing or spliced.

For additional verification and testing of this exciting spliceosome mutation and to verify the utility of the *HIS3* reporter, we transformed strains containing wild type BBP and wild type mBBP with *HIS3*-hBPS plasmid reporters and assayed in histidine dropout liquid media with different concentrations of 3-AT in a microtiter 96-well plate. These results perfectly align with what was observed with the *CUP1* reporter system and the RT-qPCR results. The strain with the

highest growth at all concentrations of 3-AT is the wild type strain with *CUP1*-MATa1, followed by mBBP strain with *CUP1*-hBPS. Even at 0 mM 3-AT, the strain with wild type BBP and *CUP1*-hBPS displays a growth defect, indicating that enough histidine is being produced to support reduced growth. These results confirm the splicing defect of hBPS and the splicing rescue of mBBP are gene-independent.

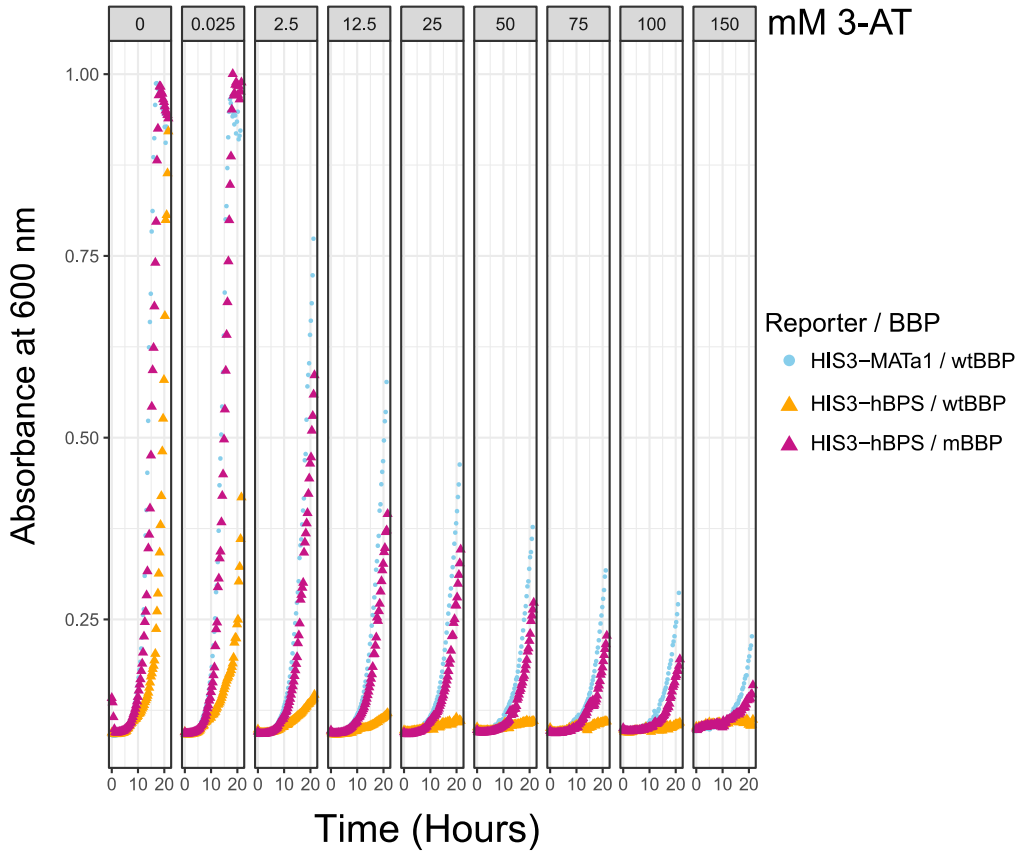


Figure 23. *HIS3*-hBPS assay shows mBBP improves splicing of hBPS. Yeast growth in a microtiter 96-well plate. Raw absorbance is plotted against time. Each facet represents a different concentration of 3-AT. The different reporter constructs are depicted by color and shape.

4.8 Assaying a randomized BPS library using high-throughput sequencing

In collaboration with the Stanford Genome Technology Center (SGTC), specifically Kevin Roy, Justin Smith, Sundari Suresh, and Bob St. Onge, we designed a high-throughput sequencing experiment to study the changes in intron-binding specificity of the mBBP spliceosome mutation. Not only would this yield interesting insights into the BPS sequence preference of mBBP, but this

would allow us to fine-tune the experimental details in order to scale up and test more fungal introns through high-throughput sequencing approaches. Our overall strategy is plasmid-based expression of a library in wild type and mutant yeast strains, culturing with selection to alter the reporter population, and targeted sequencing to identify the counts of each library member (Figure 24). First, we introduced an intron library with a randomly mutated BPS sequence into a plasmid reporter. This plasmid library was transformed into a strain containing the mBBP mutation in the genome and cultured at different concentrations of copper. After several growth periods (sub-culturing was performed by an automated liquid-handling robot), the genomic and plasmid DNA was harvested through extraction. The intron population in the reporter was amplified using PCR and this population of amplicons was sequenced using a MiSeq. Sequences with high abundance in the data would indicate well-spliced introns as they would have higher resistance to copper and thus a fitness advantage during the assay. We took the yeast intron (MATa1) used previously and had IDT synthesize it with a randomized BPS sequence. The BPS is 7 bps in total and all were randomized with equal probability for all 4 nucleotides (A, G, C, T). Then this was cloned into the *cup1* reporter on a plasmid for the selection assay and sequencing analysis. BspQI seamless cloning was performed to seamlessly insert the intron library into the backbone vector, as described in Materials and Methods. The library was passed through *E. coli* and transformed into two yeast strains: 1) S1 with a wild type spliceosome and *cup1*Δ, and 2) a child of S1, with mBBP integrated in place of wild type BBP.

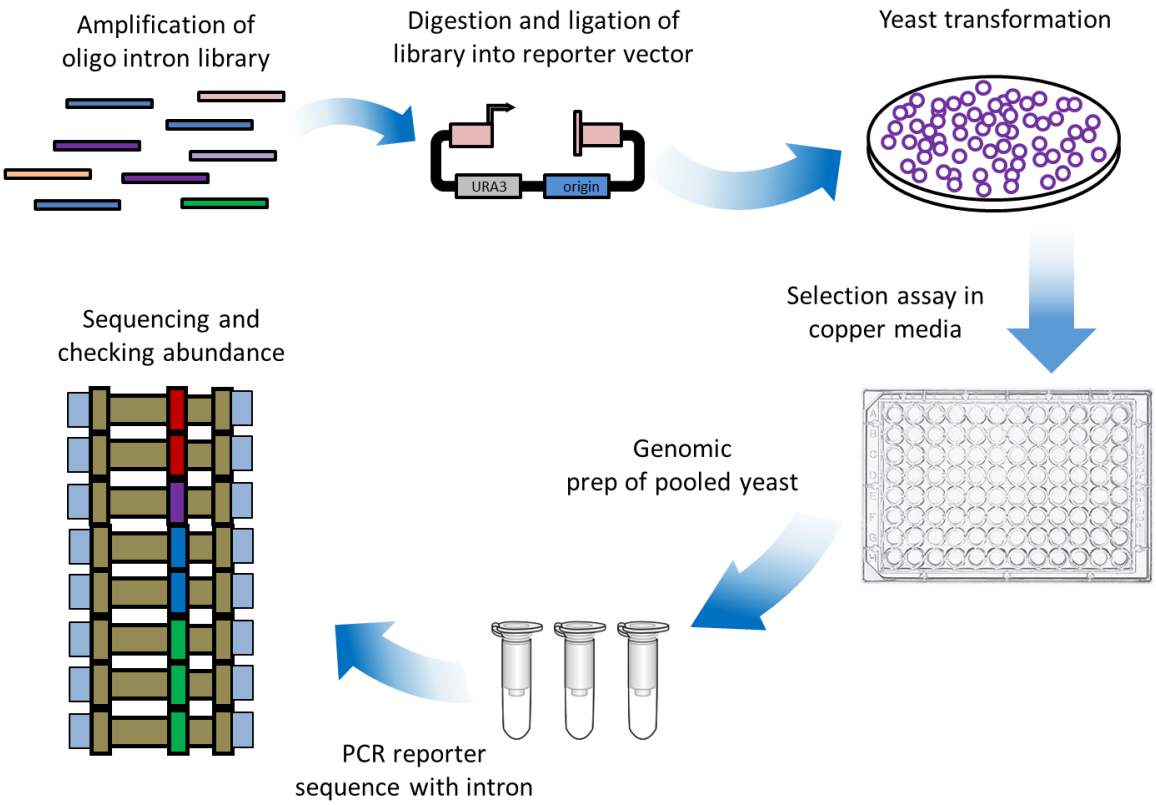


Figure 24. Procedure for building and assaying an intron. Overview of the procedure for high-throughput sequencing of an intron library.

The wild type and mBBP strains were cultured at multiple copper conditions for approximately 90 hours in a 48 well plate on a liquid handling robot (see photograph, Figure 25). Absorbance was measured periodically and once cultures reached a threshold absorbance (0.6), they were sub-cultured to a new well with fresh media at the same copper concentration as the initial media. This allowed the populations to remain in exponential growth phase for the duration of the experiment. Four growth periods of the strains came from three sub-cultures.

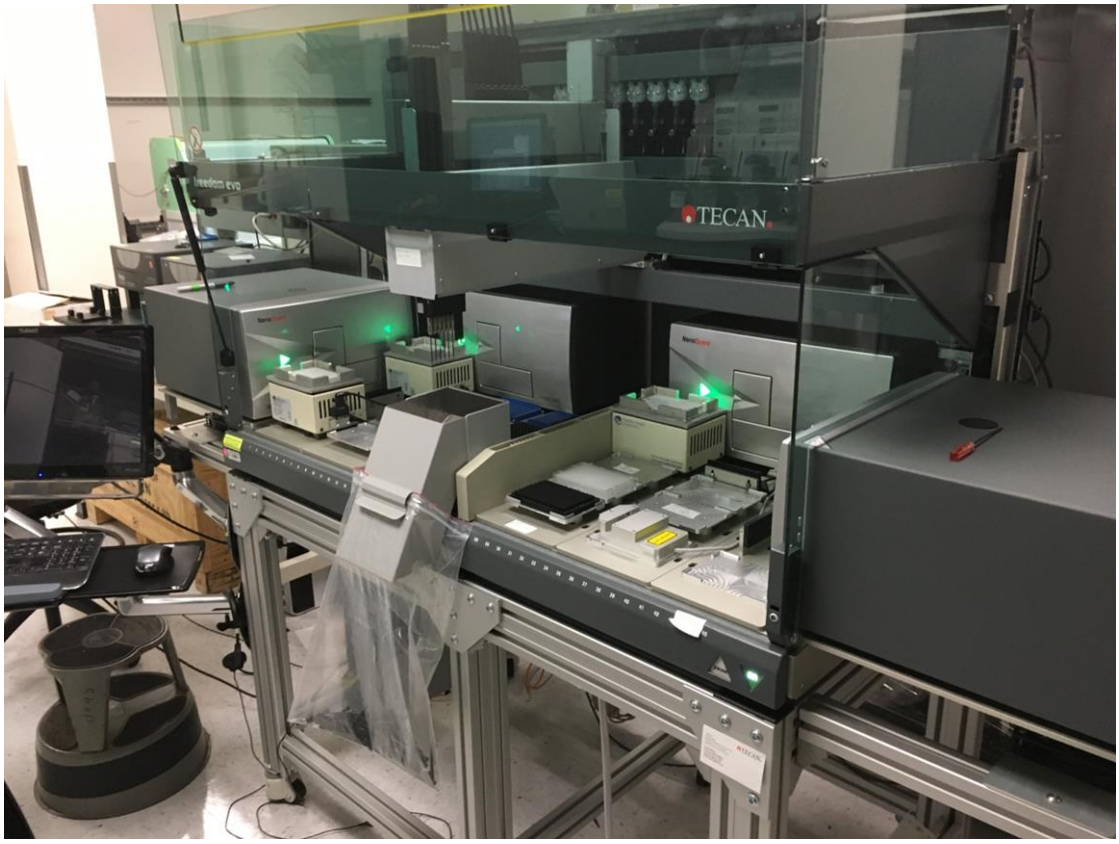


Figure 25. Photograph of the Tecan Liquid Handler. The liquid handling robot / plate reader was used for automated yeast culturing. Located at the Stanford Genome Technology Center in Palo Alto, CA.

At 0 mM, both strains grow nearly identical, which matches earlier experiments showing no difference between wild type spliceosome and strains with mBBP in place of BBP. It was not expected to see a difference in intron populations between the strains at 0 mM and the sequencing data reflected an unchanged library. There is no strong sequence bias in the strains without selection pressure. This also gave us insight into how well the library was constructed. S1 strains achieved 78% library coverage (12,746 variants detected out of 16,384 theoretical variants from 4^7 possibilities with a 7 nucleotide randomized sequence) while strains with mutant mBBP (BBPm) achieved 63% coverage (10,260 variants detected out of 16,384 total). While higher coverage is always desired, it was not critical for the design of this experiment considering we were more interested in identifying differences between a spliceosome containing mBBP and a wild type spliceosome.

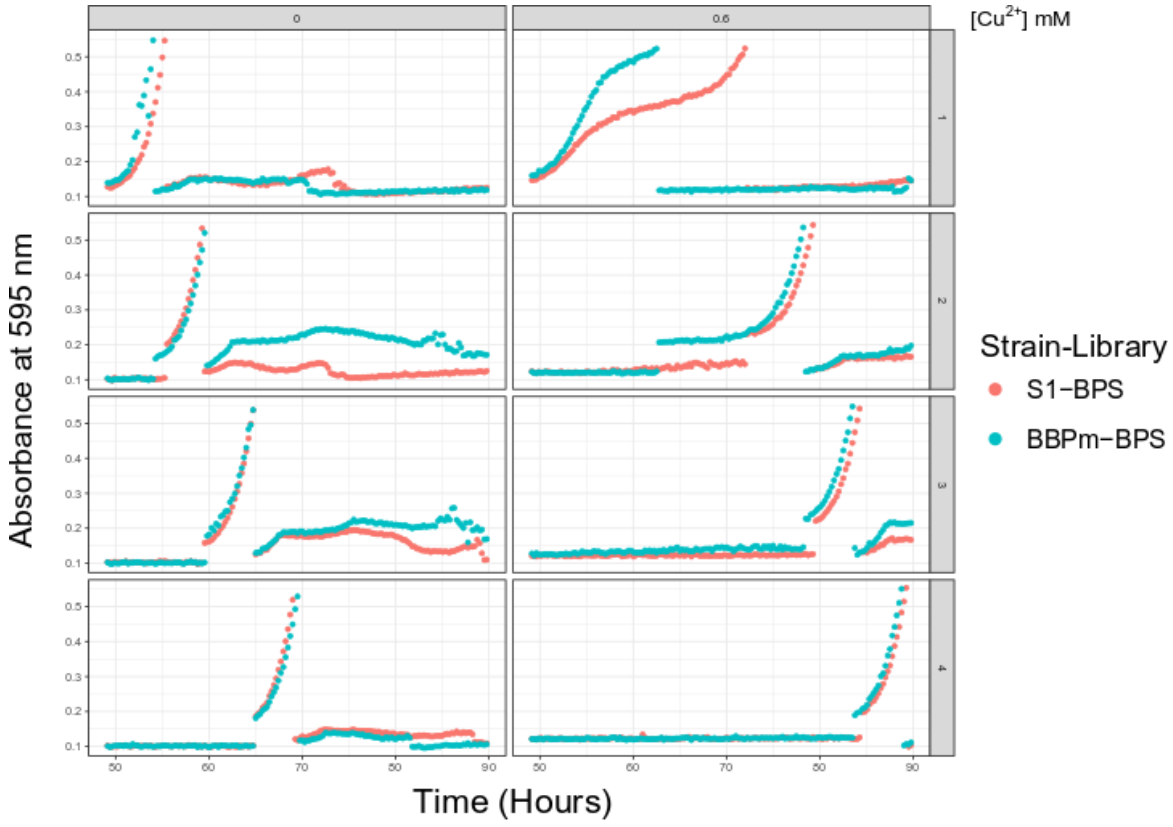


Figure 26. Growth curves of library-containing strains. Growth data from liquid handling pooled growth experiment. Each facet essentially shows all the absorbance readings from one of the wells of the 48-well plate. The pink strain (S1) is the wild type with the randomized BPS library. The blue strain (BBPm) has the mBBP genomic mutation and randomized BPS library. The left and right facets are 0 mM copper and 0.6 mM respectively. The top-bottom facets represent the different sub-culturing steps, starting at 1 and ending with 4. The residual growth in each well after sub-culturing is either due to some residual growth from the remaining culture or due to altered absorbance readings from the film being punctured and potential condensation collecting on the film.

Growth at 0.6 mM was a more interesting result. It is observed during the first growth period that the strains exhibit different growth profiles. The strain with mBBP (in blue) grows faster initially, then begins to level off and picks up again after sub-culturing, while the wild type strain (in red) grows more slowly before leveling off then picking up again. Starting in the second facet and continuing for the remainder of the experiment, the strains have essentially the same growth. We interpreted these results to mean that the mutant strain is capable of splicing a higher variety of sequences than wild type, allowing it to grow more quickly in the presence of copper. We extracted and sequenced both strains after the final sub-culturing period. Unfortunately, the results from this preliminary MiSeq experiment need to be repeated to remove sequencing artifacts. While the results from the wild type strain are logical (appearance of the consensus BPS in the data:

UACUAAC, particularly in the 3′ region of the BPS), the mutant spliceosome strain contains unusual motifs that could not be independently verified to splice. The top 5 sequences are hugely abundant relative to the remainder of the pool with counts 1-2 orders of magnitude more frequent than other sequences. These BPS sequences are TGTGGTC, GACTCAA, TAATGTC, GGGCGGG, TTGCACT. In fact, the first 12 most abundant sequences do not appear to be real BPS (no canonical A in position 6 of the BPS). We independently cloned the introns containing the most abundant BPS (TGTGGTC and GACTCAA) but these displayed no growth at any copper conditions tested (but grew well at 0 mM), suggesting these introns cannot be spliced. When the top 12 “non-splicers” are removed from the raw sequencing data and re-analyzed, a slightly more realistic BPS motif emerges. However, the total counts without the top 12 non-splicers are very low and are not likely to be a realistic picture of true mBBP preferences. Also, for S1 growth at 0.6 mM, there were 1671 unique variants remaining in the pool (from 12,746 at start). For mBBP growth at 0.6 mM, only 346 unique variants remained (out of 10,260).

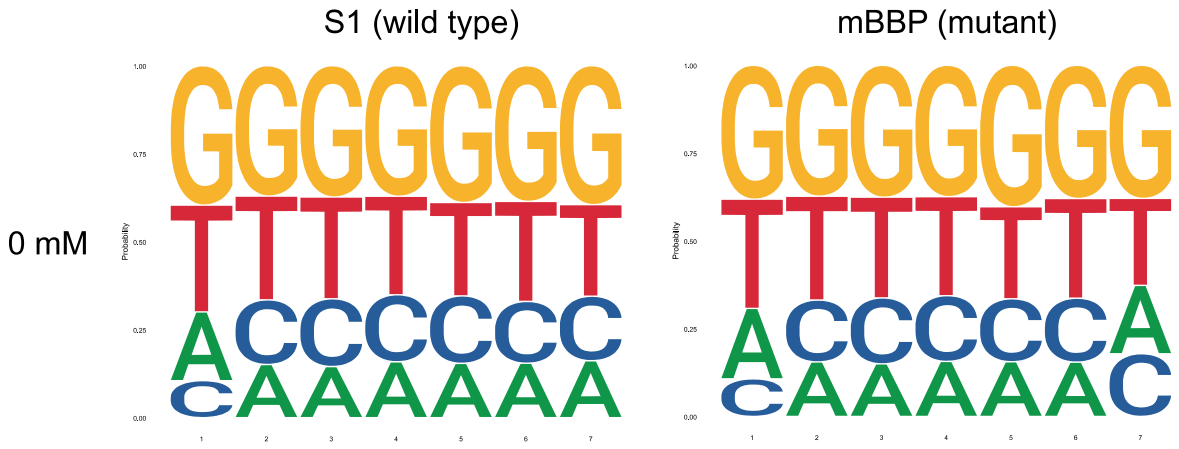


Figure 27. High-throughput sequencing results without selection pressure. Results from MiSeq data for wild type and mBBP strains cultured at 0 mM and extracted after the first growth period. There is no selection pressure for splicing and therefore there is no change in the intron population. Both strains are starting with essentially identical intron populations according to this probability sequence logo. The height of each letter represents its relative frequency at that position of the BPS.

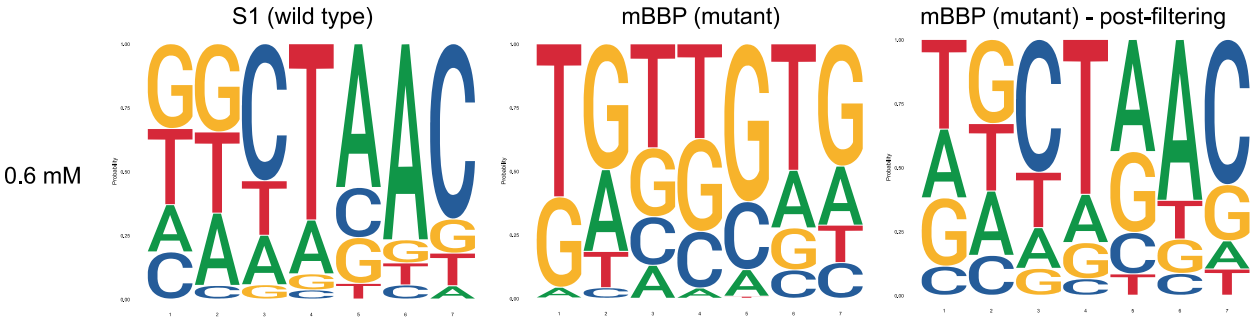


Figure 28. High-throughput sequencing results with selection pressure. Results from MiSeq data for wild type and mBBP strains cultured at 0.6 mM and extracted after the final growth period. The height of each letter represents its relative frequency at that position of the BPS. The middle image is from the raw data analyzed for the mutant strain, containing the top 12 "non-splicers" described in text. The final image is the raw data with the 12 "non-splicers" filtered out.

Overall, the success of the assay in a wild type strain shows the possibility of using high-throughput sequencing assays to study splicing of introns. The cloning strategy established here is simple and relatively inexpensive to perform. Further optimization of this assay with our collaborators will be performed. We aim to reduce the copper concentrations and extract the yeast pellets from earlier time points in order to capture the most sequence diversity.

4.9 Assaying hBPS variants to understand sequence specificity of mBBP

Given the difficulties with the high-throughput sequencing assay, we decided to study the BPS sequence in more detail using traditional cloning and constructed a small set of additional intron variants (Figure 29). We dissected the 3 bps that are mutated in hBPS (**AG**CU**G**AC) compared to wild type BPS (UACUAAC). For this assay, the mBBP mutation was re-built into strain S1 (*cup1*Δ) to generate S12 and introns were cloned into CEN/ARS vectors using a BspQI cloning method to seamlessly introduce introns into the *CUP1* ORF (see Materials and Methods). Reporters cloned into a CEN/ARS plasmid should produce expression nearly as stable as genomic integration (58). The set of 5 vectors was transformed either into strain S1 or S12 for assaying in a wild type BBP or mBBP background.

Importantly, mBBP causes no defect in splicing of wild-type MATa1 intron, demonstrated by the bottom panels (Figure 29 and 30) as no difference in strain fitness was detected compared to a wild type strain background. Splicing of hBPS is rescued in the strain with the mBBP mutation

compared to wtBBP, emphasizing the reproducibility of the results seen previously with the integrated reporters. The reduction in fitness effect observed compared to previous data is likely due to the fact that all strains displayed were observed to have a much higher sensitive to copper when assayed in uracil dropout liquid compared to YPD liquid. This underscores the utility of using integrated reporters, which can be assayed in YPD and display more robust growth in liquid assays.

These assays reveal that only the presence of both G2 and G5 in a BPS severely abolishes splicing in a wild type strain, but not in the mBBP strain. Substitution of either G2 or G5 singly into the native BPS sequence only demonstrates a slight reduction in splicing in a wild type strain. The mBBP mutation also gives a growth advantage to a G2 mutation (hBPS4) as revealed by fitness data, which is supported by the literature as the binding of BBP to a BPS with a G2 mutation is reduced (243). However, the wild type BBP gives a fitness advantage to splicing of an intron with a G5 mutation (hBPS3). This assay also shows that the initial bp A1 is not critical for yeast splicing as A1 is well-tolerated in wild type and mutant strains. These results are promising for the success of mBBP as high-impact splicing enhancer due to the fact that this “double-G” motif (G2+G5) is seen in other putative BPS of *PES1* introns, indicating it is fairly common at least in *A. fumigatus* introns.

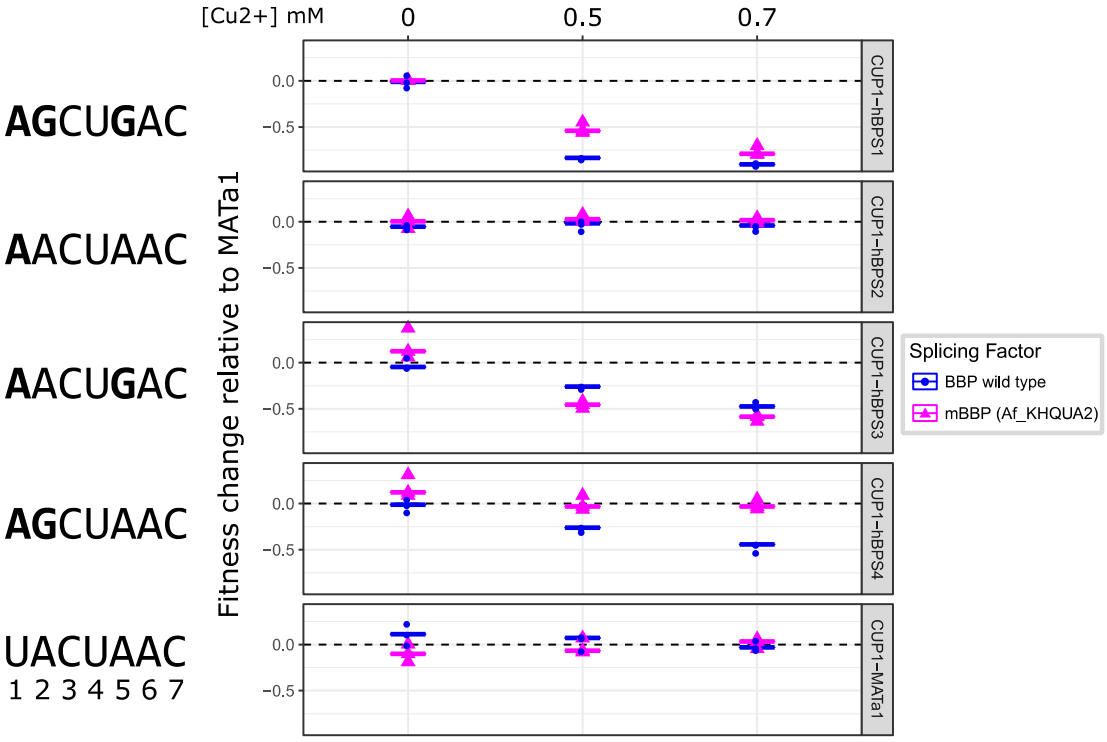


Figure 29. Fitness change of hBPS-variants. Relative fitness results are shown for variants of hBPS in order to test the effects of each mutated base pair of the BPS. hBPS is the top facet and consensus yeast BPS is the bottom.

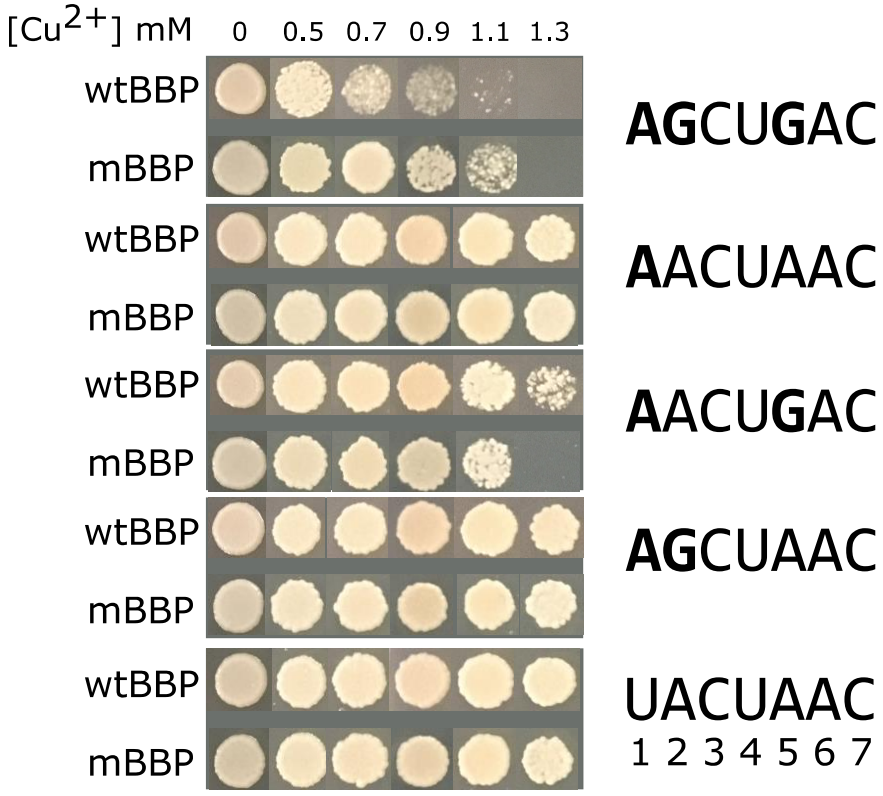


Figure 30. Liquid spotting assay of hBPS-variants. Liquid spotting assay of hBPS variants. Similar results are observed compared to the growth curve analysis. hBPS is the top facet and consensus yeast BPS is the bottom.

4.10 Materials and Methods

Strains

To modify the splicing factor genes, a flexible CRISPR-Cas9 system was deployed. Two plasmids are needed as templates to generate three PCR products needed for the CRISPR-Cas9 plasmid: 1) pCB30 (G418, Addgene #pending) or pCB32 (hyg, Addgene #pending) depending on the antibiotic resistance desired; and 2) pCRCT, a gift from Huimin Zhao (Cas9, Addgene #60621, (226)). All three PCR products were amplified by Accuprime Pfx Polymerase (Invitrogen) and primer sequences used are in Appendix. The primers used to amplify the iCas9 expression cassette have homology to the products amplified from the pCB plasmids to form a complete plasmid *in vivo* through yeast homologous recombination. The primers used to amplify the pCB plasmid are designed to have homologous overlap at the 20bp gRNA sequence site, allowing any desired guide sequence to be designed into the oligo (IDT). Guide sequences were chosen using CRISPy Cas9 target finder (244). All 3 PCR products were co-transformed with the donor DNA into the recipient yeast strain. After the heat shock, strains were inoculated into 3 mL YPD and shaken overnight at 30°C. Cultures were diluted and plated for single colonies on the appropriate selection for the CRISPR-Cas9 plasmid. Efficiencies varied per genomic loci targeted. For mBBP: out of 10 colonies screened, 60-80% would have mutation. For *YHC1-D36A*: 1/6 would have the mutation. Approximately the same efficiency was observed for *prp5*-N399D.

All yeast transformations in this study were carried out using the LiOAc/ssDNA method (214). Transformants were selected using leucine dropout media, uracil dropout media, 5FoA media (0.8g/L – 1g/L), or 2% YPD supplemented with 200 µg/mL hygromycin or 200 µg/mL G418. All genomic mutations were confirmed by colony PCR amplification of the region affected plus 50-500 bp of additional region upstream or downstream of the homology targets, to confirm specificity of the recombination event. PCR products were sequenced by Sanger sequencing (Laragen).

Plasmids

Standard cloning techniques were performed to maintain and propagate the plasmids in *E. coli* DH10b. pXP318 was a gift from Nancy Da Silva & Suzanne Sandmeyer (Addgene plasmid #26837) (65). To clone the U2 snRNA cassette, *LSR1* and its promoter and terminator were amplified using colony PCR from BY4741 and assembled into pXP318 (replacing the TEF1 promoter and CYC1 terminator). To create U2* snRNA with 3 mutated bps, the mutations were designed into oligos that overlap at the site of the mutation. *LSR1* was amplified in two pieces (promoter in piece 1 and terminator in piece 2) and then assembled into one product using overlap-extension PCR before assembly into pXP318. A similar procedure was used to clone *MSL5* with its native promoter and terminator into pXP318. Once the wild type BBP expression vector was created, PCR was used to create a linear product of the entire vector minus the sequence of the KH-QUA2 domain. The *A. fumigatus* domain was synthesized as a gBlock (IDT) with 60 bp homology to wild type BBP upstream and downstream of the native KH-QUA2. This gBlock was amplified using PCR and assembled with the linear backbone product to create pXP318-mBBP. Plasmids with the *CUP1*-hBPS variants (hBPS1, hBPS2, hBPS3, hBPS4) were designed and cloned using the BspQI library cloning procedure (described below).

Protein Alignments

Protein sequences of *MSL5* and *MUD2* homologs were downloaded from Uniprot, NCBI, and *Saccharomyces* Genome Database FUNGI BLASTP, aligned using T-Coffee (245-248), and the image was generated using ESPript (249).

Liquid spotting assays

Strains were inoculated into YPD or uracil dropout liquid media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. Optical Densities (ODs) were measured by spectrophotometer at 600nm for each strain and cells were normalized to each other by OD and then serially diluted into sterile water. 5 µL to 20 µL were spotted and plates were incubated

at 30°C for 3-7 days before imaging. Temperature growth assays were performed by spotting serial dilutions on YPD or uracil dropout media at 16°C (imaged after 12 days), 25°C, 30°C, and 37°C (imaged after 4 days).

Growth Curve Analysis

Raw data was exported to Excel from iControl Software for the Tecan M200 Pro and imported into R version 3.4.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Packages from the Tidyverse (Hadley Wickham (2017). tidyverse: Easily Install and Load 'Tidyverse' Packages. R package version 1.1.1. https://CRAN.R-project.org/package=tidyverse) were used to manipulate the data for analysis and presentation with ggplot2 (Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009)). Fold change calculations were determined in an R script as follows: 1) the mean for the first 5 measurements of each strain replicate at each copper condition was calculated and subtracted from all absorbance readings to set the baseline of each growth curve to zero; 2) the area under the curve (AUC) was calculated as the sum of all adjusted absorbance readings using the trapezoidal rule; 3) the AUC of either the intron-less *CUP1* strain, *CUP1*-MATa1 strain, or *CUP1*-pes1-5 strain was used as the control value in the fold change equation: $(\mathrm{AUC}_{\mathrm{sample}} - \mathrm{AUC}_{\mathrm{reference}})/\mathrm{AUC}_{\mathrm{reference}}$.

RNA extraction

For hBPS strains, yeast cells were inoculated into YPD from agar plate colonies or frozen stocks and grown overnight in 5 mL of 5.0 mM Cu^{2+}, shaking at 30°C. The following day strains were then sub-cultured to an OD_{600} 0.1 in fresh 5 mL of YPD media containing 1.2mM Cu^{2+} and cultured for 7 hours, each strain with three biological replicates. ODs were measured by spectrophotometer and cell count was normalized to 500 µL per OD_{600} 1.0. All samples were

centrifuged, media was removed, and cell pellets were resuspended in 1 mL RNAprotect Cell Reagent before incubation at room temperature for 15 minutes. Cells for hBPS experiments were extracted fresh. RNAprotect Cell Reagent was removed using centrifugation. To digest the cell walls, cells were incubated for 1 hour at 37°C with gentle shaking in a zymolyase solution (100 µL RNase-free water, 10 µL zymolyase (Zymo Research), and 0.1 µL beta-Mercaptoethanol). All RNA was extracted using a slightly modified form of RNeasy Mini Plus Kit (Qiagen). On column DNase digestion (Qiagen) was extended to 1 hour. RNA was eluted from Qiagen columns using 100 µL 10 mM EDTA.

Quantitative RT-PCR experiments

Primers for RT-qPCR were designed using primer3 (215,216) or by hand when appropriate, such as for the splice junction primer. For these strains, the spliced product was abundant enough to detect easily using total RNA. RNA samples were diluted to 1 ng/µL before adding 5 µL of these templates to each well on a 0.2 mL skirted 96-well PCR plate (Thermo Scientific, AB-0800/W). The template for each biological replicate was added to three separate wells to generate three PCR replicates per biological replicate. A master mix for the Luna Universal One-Step RT-qPCR Kit (NEB, #E3005L) was generated for each primer pair used and 15 µL was aliquoted into each well and plates were sealed with Microseal 'B' seals (BioRad #MSB1001). Samples were run in a CFX96 Real-Time System with C1000 Thermal Cycler. The protocol for RT-qPCR was as follows: 55°C for 10 min (RT), 95°C for 1 min, [95°C for 10 sec, 60°C for 30 sec, plate read] x 45 cycles. Subsequent melt curve analysis protocol: 60°C incremented by 0.5°C to 95°C, holding temperature for 5 sec.

RT-qPCR analysis

Raw Cq values were exported to Excel from Bio-Rad CFX Manager 2.0 (determined by regression, algorithm similar to PCR Miner (217) and then imported into R for analysis. RT-qPCR

experiments were performed using an all genes approach where samples were split across multiple plates but all genes were tested for each plate. Therefore, ΔCq (spliced – reference) values for each strain were calculated for each plate before combination of all replicate values. For hBPS strains, the ΔCq values were used to determine % transcript using the following equations: $\%\,\text{spliced} = 100 * 2^{(\text{Cq}_{\text{exon2}} - \text{Cq}_{\text{spliced}})}$ and $\%\,\text{intron} = 100 * 2^{(\text{Cq}_{\text{exon2}} - \text{Cq}_{\text{intron}})}$ (and these transformed ΔCq values are plotted directly.

Library cloning

The oligo containing the MATa1 intron with a randomized 7 nucleotide BPS (90-mer from IDT) was amplified using Q5 DNA Polymerase (NEB) (Tm = 69°C, 5 sec extension) for 13 cycles. The product was purified using Binding Buffer column purification and digested with BspQI to open unique sticky ends that match the backbone vectors. The backbone vectors were derived from pKR8 (gift from Kevin Roy and Justin Smith at SGTC) which lacks any BspQI sites on the backbone. The *CUP1*promoter- *CUP1* ORF-*CYC1*terminator expression cassette was cloned into this backbone. The promoter and ORF 43 bp downstream from the start codon were amplified as one piece and with two BspQIs introduced as part of the oligo to allow for scar-less insertion of introns with matching BspQI sites. A homologous oligo was used to amplify the remainder of the *CUP1* ORF-*CYC1*terminator and these two PCR amplicons were assembled with a linear pKR8 backbone using Gibson Assembly (60). The backbone was also digested with BspQI and ligated to the intron library for 2 hours at room temperature using T4 ligase. The samples were then ethanol precipitated and re-suspended in 5 µL prior to electroporation of 1 µL into electrocompetent *E. coli.* Cells were recovered in pre-warmed SOC liquid media (1 mL) for 1 hour at 37°C and plated on LB-carb agar plates. The cells from the agar plate were scraped and pooled before miniprep. Transformation of the plasmid libraries into two yeast strains was carried out using the LiOAc/ssDNA method (214).

Pooled Screen Assay

The two libraries (S1-BPS and BBPm-BPS) were expanded in uracil droput liquid media before sub-culturing into a 48-well plate, containing different copper concentrations. Absorbance at 595 nm was taken periodically in a Tecan plate reader over the duration of the 90 hour experiment. Cultures were maintained in log phase growth by automatic sub-culturing once ODs reached 0.6 using a Freedom EVO Liquid Handling System (Tecan), controlled by a custom Pegasus software (Tecan).

Preparing DNA templates for High-Throughput Sequencing on MiSeq

DNA from the cell pellets harvested from the pooled screening assay were extracted using the following procedure: 1) Add 400 µL of extraction buffer to pellet (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris HCl pH 8.8, 1 mM EDTA pH 8.0); 2) Mix solution and add 100-200 µL glass beads; 3) Add 400 µL phenol:chloroform:isoamyl alcohol (24:25:1) pH 8.0 to each tube; 4) Vortex briefly; 5) Incubate at 65°C for 10 min, vortexing at 5min point; 6) Spin down at top speed for 5 min; 7) Transfer 200 µL of supernatant to a new tube containing 0.5 mL ethanol and 20 µL NaOAc pH 5.2; 8) Add 1 µL glycoblue for DNA visualization; 9) Spin tubes down for 1 min at top speed and remove supernatant; 10) Wash with 500 µL 70% ethanol; 11) centrifuge and remove supernatant; 12) Resuspend DNA pellet in 50 µL 1X Tris-EDTA (TE) buffer. These samples were digested using RNase and the DNA was further cleaned using Binding Buffer column purification and resuspended in a smaller volume.

Adaptor sequences were designed as part of the primers for the reporter. Four primers were used to amplify each of the four gDNA templates to introduce unique adaptors for multiplexing during the MiSeq run. A 10 fold enrichment of outer:inner primers was used. For a 50 µL Q5 Polymerase (NEB) reaction, 2.5 µL of each 10 µM outer primer and 0.8 µL of each 5 µM inner primer were added to each gDNA template. The remainder of the reaction was set to manufacturer’s conditions. Templates for S1 at 0 mM, S1 at 0.6 mM, and BBPm at 0 mM all were

amplified for 30 PCR cycles, (annealing temperature was 61°C for 20 sec, and extension was 72°C for 4 sec). The template for BBPm was run at the same thermal cycler conditions but only amplified for 24 cycles and was noted to amplify much earlier than the other templates consistently during PCR condition optimization/troubleshooting.

Analysis of High-Throughput Sequencing Data

We thank Nathan Lubock (Sriram Kosuri lab, UCLA) for helpful discussion on MiSeq data analysis and for generation of a bash script to run a BBtools pipeline (JGI) to trim the paired end reads by overlap and remove contaminating sequence information. Cleaned fastq files were imported into R using the ShortRead package (250), BPS sequences were extracted by locating the intron position in the string and then removing the 7 bps following the IB1 region. BPS were plotted as a sequence logo using ggseqlogo (251).

5. ADDRESSING SPLICING FAILURE MODE #2

5.1 Introduction to 5′ end problem

Originally, the heterologous IB1 region was presumed to the main cause of the splicing failure of the h5′ss_hIB1 hybrid intron tested using the *URA3* reporter. I reasoned this from the positive splicing results from the h5′ss hybrid intron. If the heterologous 5′ss (GTACGT) from pes1-5, which is still used natively in some *S. cerevisiae* introns, splices alone, then it seemed likely that the reason h5′ss_hIB1 didn't spliced was simply the hIB1 module. It has high GC content (50%) compared to *S. cerevisiae* MATa1 (20% GC) indicated a potential issue with secondary structure formation.

5.2 Assaying hIB1 variants

Based on the *URA3* reporter assay described in Ch.3, I reasoned the yeast spliceosome cannot splice a hybrid intron containing the heterologous IB1 sequence from pes1-5. First, I explored the mechanism through which the heterologous IB1 prohibits splicing using the more

sensitive *CUP1*-intron reporter. The sequences of the *S. cerevisiae* IB1 (from MATa1) and the *A. fumigatus* hIB1 (from pes1-5) differ significantly in GC content despite being the same length (30 nt); the IB1 has a GC content of 20% while hIB1 is 50%. As GC content has been shown to correlate with the stability of secondary structures (252), it is possible that the formation of a secondary structure is abolishing splicing activity. As yeast lack auxiliary factors to influence the recognition of various introns, splicing is more dependent on *cis*-acting elements such as the secondary structure in intron modules (253).

Secondary structures formed from IB1 and IB2, the two sequences in-between the consensus sequences, play important regulatory roles during intron splicing (94,254). Work by Rogic et al. focused on elucidating how secondary structure interactions in IB1 sequences influenced splicing efficiency by altering the distance between the 5′ splice site and the branchpoint site (254). The Eyras group isolated IB2 sequence data for the introns in yeast, computationally predicted the minimum free energy structures, and determined that predicted secondary structures influenced which potential 3′ splice site would be chosen to complete splicing (205). However, Rogic et al. focused on introns with long IB1 sequences. The intron size distribution in yeast is bimodal, with long introns (~400 bp) and short introns (~100 bp), where the drastic difference in length is primarily due to long IB1 sequences (89). Fungal species such as *Aspergillus* tend to have short introns (mean length ~73 bp) (38) and therefore studying structures formed by short IB1 sequences will provide insight into heterologous splicing. Therefore, I dissected the role of the secondary structure interactions in a short heterologous IB1 sequence and to improve knowledge of how the IB1 sequence can influence intron splicing.

To examine possible secondary structures adopted by the IB1 from the yeast MATa1 intron and the heterologous IB1 from the pes1-5 intron, mfold (255), RNAfold (256), and forna (257) were used to calculate the minimum free energy structures associated with each sequence. All prediction software agreed on the set of structures presented using forna in Figure 31. The IB1

from MATa1 forms no secondary structure and hIB1 from pes1-5 forms a stable stem loop containing 19 base pairs.

In previous work done by Köhrer and Domdey, it was found that deleting as few as 6 nucleotides from the IB1 of MATa1 was sufficient to abolish splicing of the intron (207). They proposed that there is a minimum distance requirement between the 5′ splice site and the branchpoint site that must be satisfied in order for splicing to occur. As MATa1 is among the smallest introns in *S. cerevisiae* (179), it is logical that causing further reduction in size will prevent this intron from being spliced. The Eyras group (253) defined the 'effective distance' between the branchpoint site and the 3′ splice site as the total number of nucleotides less the number of nucleotides involved with secondary structure formation (94). The effective 5′ss-BPS distance of the hIB1 from pes1-5 was determined to be 13 nucleotides while the effective distance of the IB1 from MATa1 is 30 nucleotides.

To determine if the reduction of the effective distance is abolishing splicing, five different intron sequences were constructed using the *CUP1-URA3* landing pad as described in Ch.3. These will consist of the wild type MATa1 and four MATa1 variants containing specific nucleotide changes to the IB1 sequence.

First, the wild type IB1 in the intron was replaced by the hIB1 from pes1-5 (as for the *URA3* reporter). Second, the entire 30 nucleotides of hIB1 will be replaced by the 13 nucleotides uninvolved with the formation of the stem loop, generating hIB1-1. This will mimic the effective distance allowed by hIB1 according to computational predictions. If the intron can still be spliced out of the *CUP1* transcript, the cells will grow on copper-dosed media, and effective 5′ss-BPS distance can be ruled out as the cause of abolished splicing. This result is unlikely given the results of the work by Köhrer and Domdey (207). It is expected that the intron will not splice efficiently, the cells will not grow on copper-dosed media, and the shortening of the effective distance of the intron will be identified as the cause of abolished splicing.

To confirm if the presence of the predicted stem loop or a variety of similar, but suboptimal structures (253), are responsible for the reduction of the effective 5′ss-BPS distance, an additional two hybrid introns will be created with various structural mutations in the hIB1 stem loop. First, the stem will be destabilized by introducing mutations C6G and G21C (hIB1-2). The MFE structure predicted has no secondary structure, similar to the wild type IB1 sequence in MATa1 (Figure 31). To reintroduce the stem, the compensatory mutations G22C and C7G will be inserted into hIB1-2 to generate hIB1-3. It is expected that intron splicing will be improved by destabilizing the stem and the yeast containing hIB1-2 will survive on higher concentrations of copper compared to strains containing hIB1. When the stem is reinstated in strains containing, hIB1-3, it is expected that splicing will again be abolished similar to in strains containing hIB1.

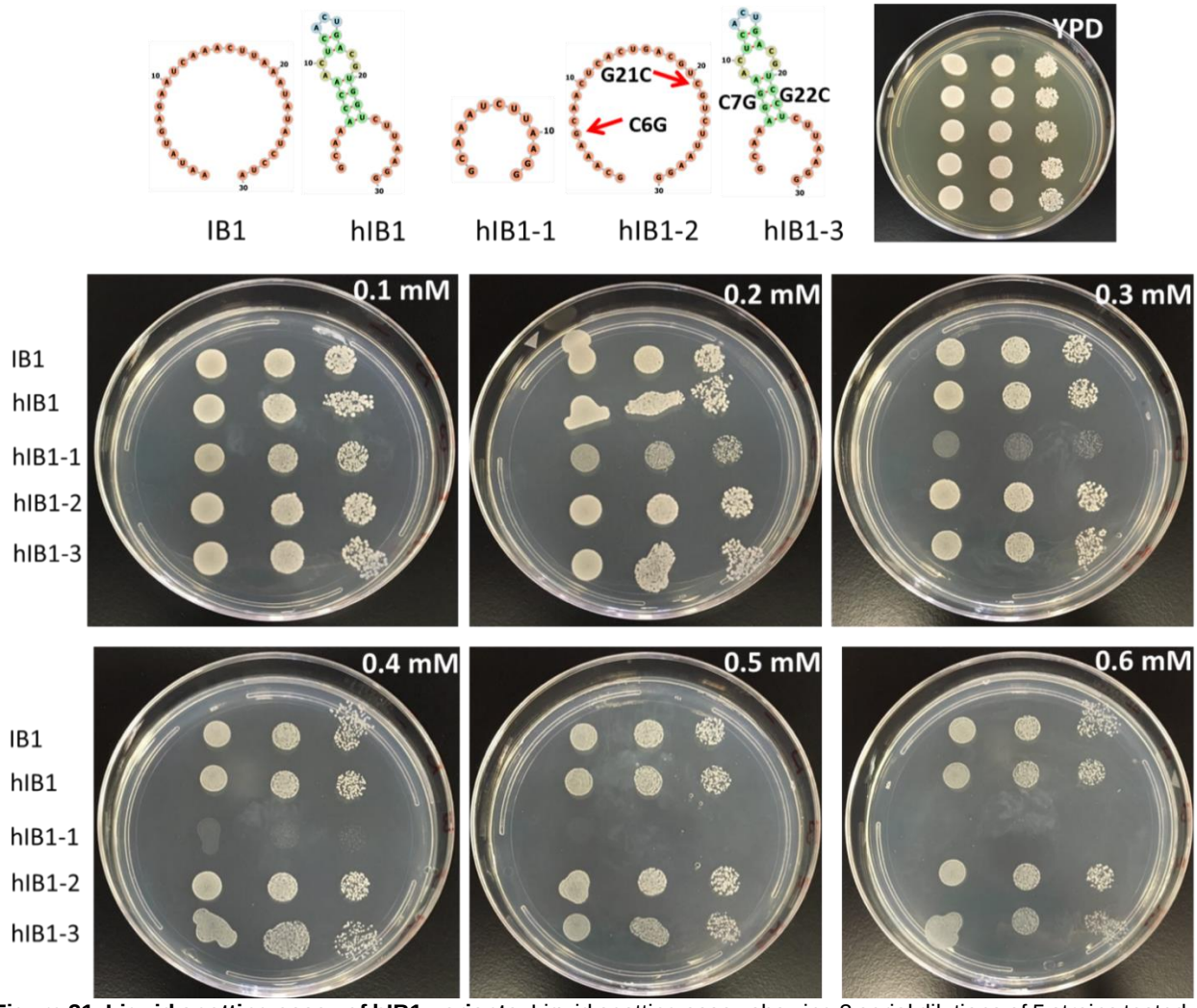


Figure 31. Liquid spotting assay of hIB1-variants. Liquid spotting assay showing 3 serial dilutions of 5 strains tested, each containing a different intron variant inserted into the genomic *CUP1* reporter. The top strain (IB1) is the wild type yeast intron MATa1. The remaining introns are all derived from MATa1 with changes made to the IB1 sequence. (hIB1) is the heterologous IB1 sequence from pes1-5 swapped into MATa1. (hIB1-1) The predicted stem loop deleted and is shorter than all other introns tested here. (hIB1-2) Two point mutations are made to the hIB1 from pes1-5 to remove the stem loop. (hIB1-3) Two additional point mutations are made to hIB1-2 sequence to re-introduce the stem loop.

From these results, all strains grow equally well on the YPD control plate, indicating any changes in growth are due to splicing of the reporter under copper selective conditions. It is apparent from the copper agar plates imaged that hIB1-1 is the worst performing strain and begins to show a growth defect as early as 0.1 mM. This is expected because any deletions to the IB1 region of this intron have already been shown to be deleterious to its splicing. The remaining introns are not so dramatically affected. As expected the MATa1 strain grows fairly well, particularly at earlier concentrations of copper. Unexpectedly at the time, it was observed that

hIB1 still supports growth even up to 0.6 mM, although it is not growing as densely as the MATa1 strain. It is worth noting that hIB1-2 appears to be growing more densely than hIB1 and hIB1-3, which provides support to the hypothesis that the stem loop structure identified computationally is negatively impacting splicing by reducing the effective distance of the 5′ region of the intron.

Finer differences between each of these regions could be resolved at higher concentrations of copper and with RT-qPCR studies. However, this was not pursued for two reasons. First, it became clear with this assay that the hIB1sequence is not as a severe as a problem by itself than previously believed. These results led me to reconsider the assumption that the reason h5′ss_hIB1 does not splice was due solely to the hIB1 portion. Second, this experiment provided support that secondary structure due to high GC content was promoting inefficient splicing of this intron.

5.3 Identification of dual module problem: h5′ss_hIB1

Following up with the results from the previous section, we decided to revisit the h5′ss_hIB1 results from the *URA3* assay (described in Ch.3). Therefore, h5′ss, hIB1 and h5′ss_hIB1 hybrid introns were reconstructed in the *CUP1* reporter and assayed as previously described. Some data is shown below and additional data is displayed in Ch. 3 where the other results from the hybrid intron assays were studied and discussed. Essentially, h5′ss and MATa1 splice very well. hIB1 alone introduces a slight growth defect but still supports some yeast growth on copper media, particularly at lower concentrations. However, the combination of h5′ss and hIB1 in the same intron severely abolish splicing relative to either module alone. This is a striking result because the heterologous 5′ss only differences from the yeast 5′ss by 1 base pair (the 4th position is a C instead of a T).

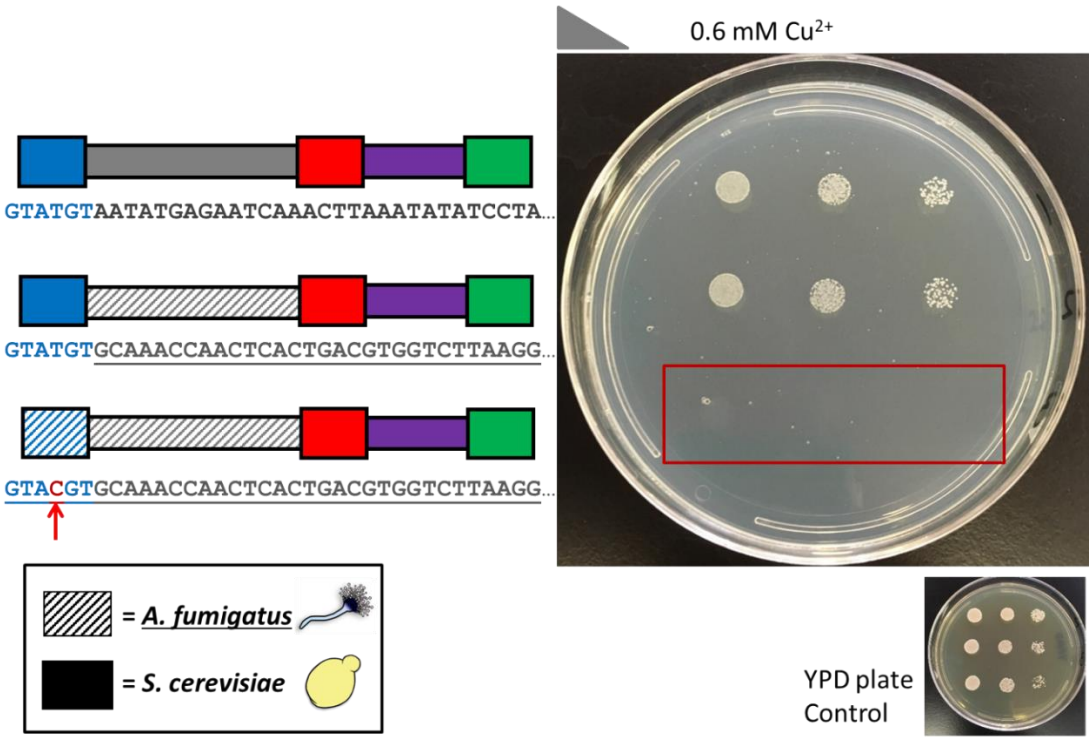


Figure 32. Liquid spotting assay identifying h5′ss_hIB1 dual module problem. Liquid spotting assay looking specifically at the wild type MATa1 intron (top), hIB1 hybrid intron (middle), and h5′ss_hIB1 intron (bottom). All introns grow on the YPD control but show dramatically different growth properties on the 0.6 mM copper sulfate plate. The only real difference between hIB1 and h5′ss_hIB1 introns is highlight: the 4th T is a C nucleotide.

We had considered the possibility that this specific problem might be arising from introduction of this hybrid intron into *URA3* and may be an unfortunate accident of a suboptimal structure happening to form from the combination of these sequences. However, this was disproven for multiple reasons. First, if this was a fluke then this problem should have disappeared once the intron is introduced into a different genetic context, such as another reporter like *CUP1* or *HIS3*. Since the problem was accurately reproduced, the secondary structure is clearly a result of the intron itself. Additionally, if the secondary structure was an accident, the only problem with pes1-5 should have been the BPS. But mBBP provides little improvement in the splicing phenotype of pes1-5 (Ch. 6), indicating other issues with the intron that the BPS. Overall h5′ss_hIB1 is an unexpected issue arising when trying to splice heterologous introns and indicates that GC content can be used to gauge how likely it is that structures are forming that inhibit yeast

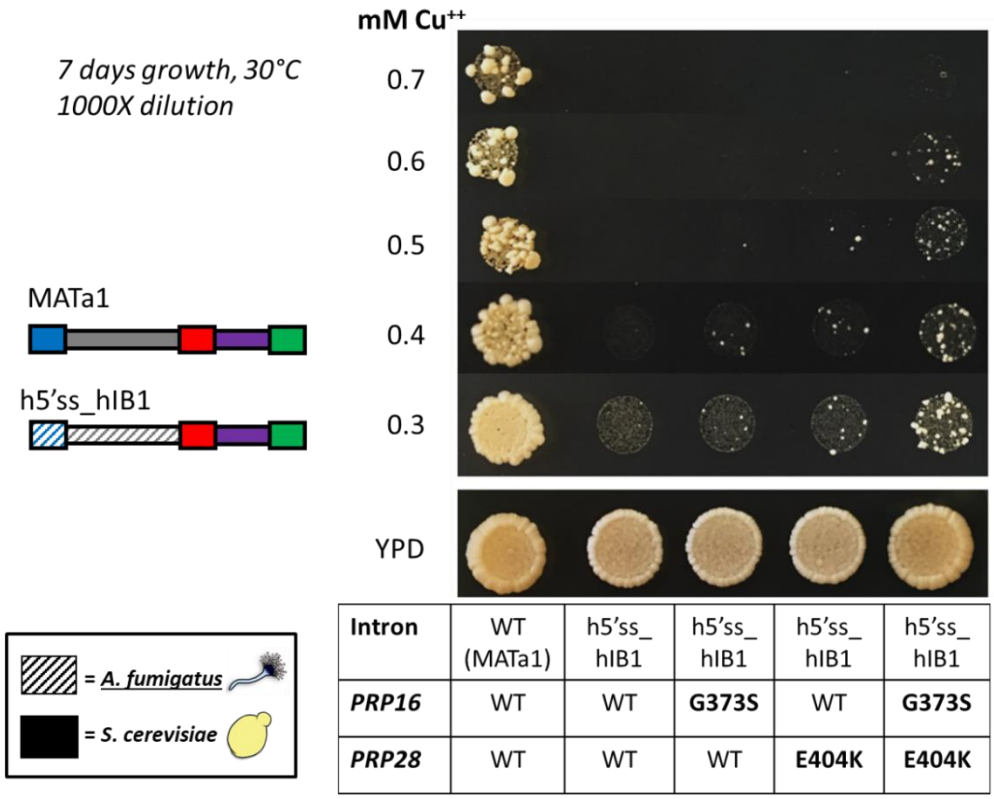
splicing. Considering that splicing can be used to regulate gene expression, and fungi has sophisticated secondary metabolism that is well-regulated, many fungal introns may form structures involved in this regulation. Therefore, new techniques need to be established to engineer the *S. cerevisiae* spliceosome to splice these types of introns.

5.4 Addressing dual module problem using RNA helicase mutations

Previous splicing assays revealed that the heterologous BPS (hBPS) and the combination of heterologous 5′ss and IB1 sequences from pes1-5 abolish splicing of their respective hybrid introns. We improved splicing of a hybrid intron (hBPS) previously using mBBP. We decided to again look into splicing proofreading RNA helicases to look for solutions into h5′ss_hIB1.

We improved splicing of hybrid introns through mutation of protein splicing factors involved in intron recognition and proofreading. We were originally interested in studying variants of U1 small nuclear RNA to rescue splicing. It has been demonstrated that U1 snRNA can tolerate extension near its binding sequence up to 30 bps without causing growth defects in a wild type strain background (258). Extended base-pairing between the U1 snRNA and the 5′ss increases 5′ss recognition in higher eukaryotes (259). However, in yeast, the increased binding can induce hyperstabilization of the U1 snRNA on a transcript, blocking activity of Prp28, and preventing splicing. Given the moderate success of *prp5*-N399D (Ch.3), we decided to focus on other RNA helicases to restore splicing of this suboptimal intron: *PRP28*, and *PRP16*. We reasoned that these splicing factors should also provide a more general solution than the U1 snRNA which operates through sequence-specific binding and would not work for introns with different sequences. Also, considering that the snRNAs have highly conserved sequences due to the many roles for their secondary structures in splicing catalysis (260), evolution has favored the development of accessory spliceosome proteins to expand intron splicing capabilities. Therefore, we focused on mutating the protein splicing factors involved in splicing sequence fidelity (*PRP28* and *PRP16* proofread splicing catalysis).

The failure of the yeast spliceosome to splice the h5′ss_hIB1 hybrid intron is likely due to the formation of a suboptimal intron secondary structure, as discussed earlier. We chose a generalizable approach to solving this problem and studied the RNA helicases (*PRP16* and *PRP28*) which are responsible for discarding spliceosomes stalled on suboptimal transcripts. The Guthrie and Query labs have identified mutants of these enzymes with slower kinetic proofreading. Prp28 proofreads the U6:5′ss duplex during incorporation of the tri-snRNP, when U6 snRNA replaces U1 snRNA at the 5′ss. The mutant E404K has slower proofreading activities (261), thus allowing more time for binding of the U6 snRNA before discarding the spliceosome. Prp16 proofreads the 5′ss and BPS during 5′ss cleavage. The mutant G373S (prp16-101) is able to suppress mutations in the BPS without inducing a yeast growth defect (262). These enzymes act more slowly on transcripts and allow more suboptimal transcripts to pass through the splicing catalysis pathway. The wild type genes were replaced with these mutants using CRISPR-Cas9 and their combination improved splicing of the h5′ss_hIB1 hybrid intron, Figure 33.



Intron	WT (MATa1)	h5'ss_ hIB1	h5'ss_ hIB1	h5'ss_ hIB1	h5'ss_ hIB1
PRP16	WT	WT	**G373S**	WT	**G373S**
PRP28	WT	WT	WT	**E404K**	**E404K**

Figure 33. Multiple mutant RNA helicases improve growth of h5′ss_hIB1 strain. Modifying multiple RNA helicases using CRISPR-Cas9 enables minor splicing improvements to a hybrid intron.

5.5 Materials and Methods

Strains

The hIB1 variants were constructed using the TranSceInt system to replace the genomic *URA3* marker interrupting the *CUP1* ORF with each individual variant so that a unique yeast strain only had one reporter variant integrated into its genome. The genomic copies of *PRP16* and *PRP28* were modified using the CRISPR-Cas9 system previously described (Ch. 4 Materials and Methods). Each mutation and appropriate homology was ordered as a gBlock (IDT) and gRNA sites were designed using CRISPy. The three pieces of the CRISPR-Cas9 plasmid were amplified and co-transformed with the appropriate donor DNA before selection for plasmid assembly on the antibiotic plates. The donor DNA included the point mutation (G373S or E404K for *PRP16* or *PRP28* respectively) as well as a silent mutation at the PAM site of the CRISPR guide RNA to prevent cutting a repaired locus. There was no great difference between the efficiencies for either

genomic locus and approximately 1/4 to 1/6 colonies screened via sequencing contained the point mutation.

Liquid Spotting Assay

Strains were inoculated into YPD liquid media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. Optical Densities (ODs) were measured by spectrophotometer at 600nm for each strain and cells were normalized to each other by OD and then serially diluted into sterile water. 5 µL to 20 µL were spotted and plates were incubated at 30°C for 3-7 days before imaging.

6. ENGINEERING THE SPLICEOSOME WITH EXPANDED FUNCTIONALITY

6.1 Overview of methodology for spliceosome engineering

Two failure modes were identified for the splicing of the *A. fumigatus* intron pes1-5: hBPS and h5′ss_hIB1. The problem of hBPS recognition was solved using the genomic replacement of the KH-QUA2 domain of yeast BranchBinding Protein (BBP) with the homologous domain from *A. fumigatus* (described in Ch.3). Unfortunately, mBBP is not sufficient to rescue splicing of pes1-5 alone due to the additional failure mode identified by h5′ss_hIB1. Therefore, additional mutations in RNA helicases *PRP16* and *PRP28* were studied to solve the h5′ss_hIB1 problem. For further study we decided to focus on how these and additional mutations would synergistically combine with mBBP to best ameliorate the splicing defect of this fungal intron.

Below is a figure depicting the spliceosome engineering methodology used in this chapter. Specifically, three aims are followed: 1) Improve splicing factor recognition of fungal intron modules; 2) Improve competition of fungal intron-containing transcripts by reducing fidelity of the spliceosome; and 3) Improve splicing factor accessibility by downregulating pools of competing transcripts. Aim 1 will focus on mutation of the splicing factor associated with each problematic module (BBP and *YHC1*). Aim 2 will discuss the results of introducing mutations that decrease

kinetic proofreading abilities of *PRP16* and *PRP28*. And Aim 3 will focus on strategies to downregulate the pool of ribosomal protein gene transcripts.

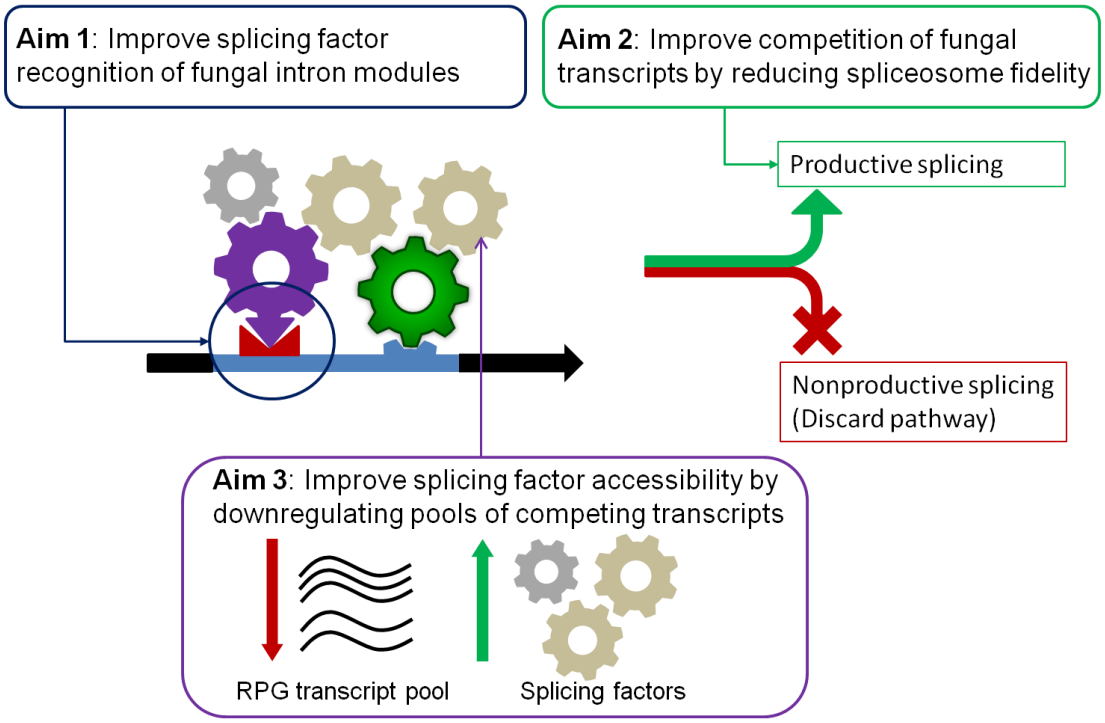


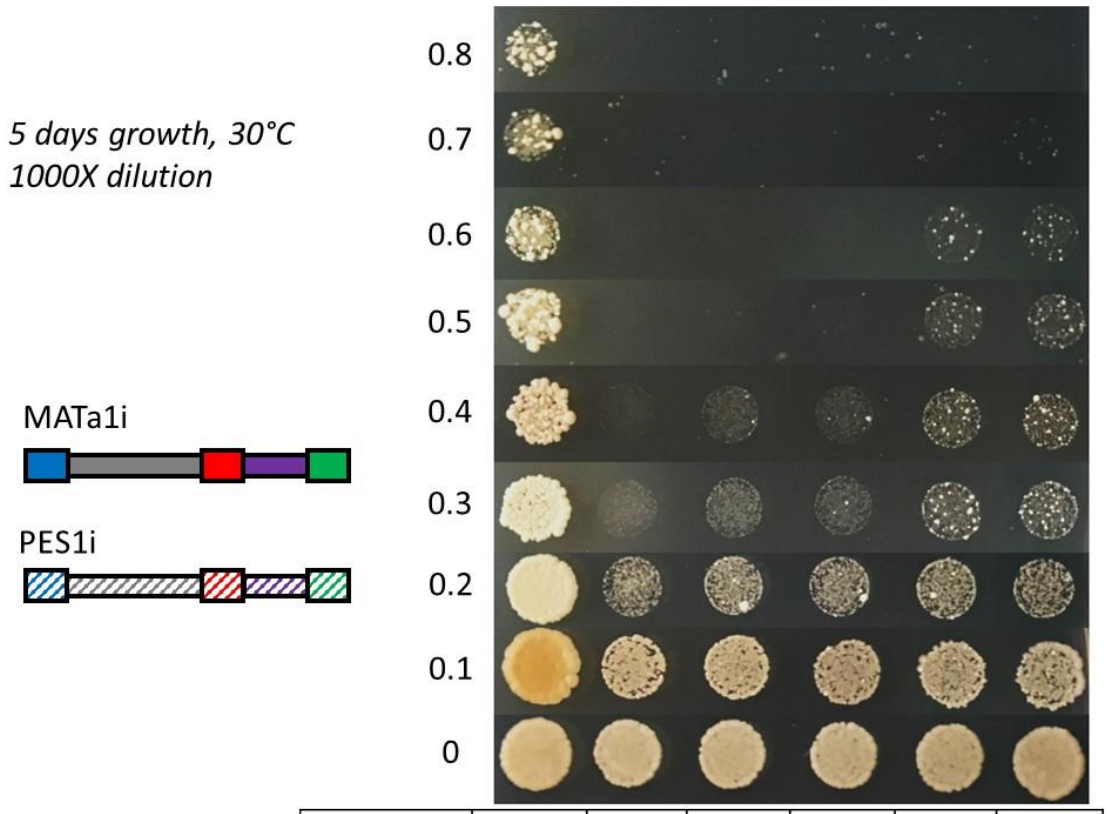
Figure 34. Three main spliceosome engineering principles. Overview of spliceosome engineering principles used throughout this work. Aim 1 emphasizes work done with mBBP and *YHC1-D36A.* Aim 2 involves modification of RNA helicases to decrease spliceosome fidelity (*PRP16* and *PRP28*). Aim 3 induces the "hungry spliceosome" phenotype by decreasing *IFH1* expression.

6.2 Improving splicing of the pes1-5 intron through mutation of RNA helicases

The Aim 1 strategy was deployed against the hBPS failure mode through mutation of BBP recognition of suboptimal BPS (Ch. 3). The chimeric mBBP showed strong improvement of incorporating the hybrid intron with hBPS (AGCUGAC) into a productive splicing pathway. However, this mutation is not sufficient to improve splicing of the pes1-5 intron due to the problem associated with the combination of its h5′ss_hIB1 modules, which I reasoned induced an inhibitory secondary structure that pushed effective intron size below the requirements for intron splicing in *S. cerevisiae.* In the preceding chapter, I described how introduction of mutant alleles of *PRP16* and *PRP28*, G373S and E404K respectively, positively synergized and allowed some recovered

growth of the hybrid intron h5′ss_hIB1 on copper plates. Therefore, I was interested in introducing these mutant alleles into the strain with the integrated *CUP1*-pes1-5 reporter as well as mBBP. Theoretically, splicing of pes1-5 should improve because both failure modes have now been addressed. I sequentially built the *PRP16* and *PRP28* mutations were built into strain containing the *CUP1*-pes1-5 reporter using CRISPR-Cas9 (see Materials and Methods).

The liquid spotting assay testing synergy of mBBP, *prp16*-G373S, and *prp28*-E404K on pes1-5 splicing shows a similar trend as the results with the h5′ss_hIB1 hybrid intron studied in isolation (Figure 35). The strain with the *CUP1*-pes1-5 reporter and a wild type spliceosome does show some slightly increased growth relative to S1 (*cup1*Δ) which indicates minor but detectable splicing of the reporter. Upon introduction of mBBP into this strain, no change in splicing is observed on these plates, indicating that mBBP alone cannot solve all failure modes of the pes1-5 intron. When *prp16*-G373S is introduced into this strain, however, a slight growth improvement is detected and the strains exhibit small colony formation up to 0.6 mM. Interestingly, introduction of *prp28*-E404K doesn't introduce a further increase, which contradicts what was observed with h5′ss_hIB1. This is possibly due to three reasons: 1) the entirety of pes1-5 may alter the secondary structure problems observed due to h5′ss_hIB1, which means we should be careful in drawing too broad of conclusions regarding the failure modes of introns from hybrid analysis; or 2) spliceosome efficiency is impaired by introduction of three splicing mutations and too many mutations will not stack additively; or 3) there is growth defect from the combination of all three splicing mutations that begins to hamper cell fitness even if there is superior splicing of the pes1-5 intron.



CUP1-**Intron**	MATa1i	N/A	PES1i	PES1i	PES1i	PES1i
MSL5	WT	WT	WT	**BBPm**	**BBPm**	**BBPm**
PRP16	WT	WT	WT	WT	**G373S**	**G373S**
PRP28	WT	WT	WT	WT	WT	**E404K**

Figure 35. RNA helicase mutations have positive synergy with mBBP. Liquid spotting assay showing changes in splicing of the *CUP1*-pes1-5 reporter. Columns 1 and 2 are controls, depicting the wild type yeast with *CUP1*-MATa1 reporter and *cup1*Δ. Column 3 shows growth of the strain with the *CUP1*-pes1-5 reporter and no splicing mutations. Columns 4 – 6 introduces a new splicing factor mutation into the strain preceding it and all are assayed on copper media together.

This assay revealed that solutions that specifically addressed each failure mode identified from hybrid intron studies could combine and increase splicing of the full *A. fumigatus* intron. However, given the results only introduced a minor improvement and we stopped seeing additive growth improvements, we decided to pursue alternative mutations to the RNA helicases. Lowering spliceosome fidelity is not an ideal goal in the process of engineering enhanced splicing of fungal introns.

6.3 Improving splicing of pes1-5 through mutation of the U1 snRNP assembly

To improve splicing of pes1-5, we decided to search for mutations that could act synergistically with mBBP and improve spliceosome assembly on the pre-mRNA containing the pes1-5 intron. We reasoned that stabilization of Complex E formation might help alleviate the secondary structure formation in the h5′ss_hIB1 region. Proteins stabilize the interactions of the U1 snRNP with the 5′ss when U1 snRNA-pre-mRNA base pairing is unstable (263). Therefore, we searched the splicing literature and identified the D36A mutation in *YHC1*, a gene involved in U1 snRNP formation for 5′ss recognition (264). This mutation hyperstabilizes U1 snRNP formation on introns with aberrant 5′ connections to the spliceosome (265,266) and we theorized it could attenuate the failure mode introduced by h5′ss_hIB1.

Figure 36 shows the combinations of two splicing factor mutations mBBP and *YHC1-D36A*. We hypothesized *YHC1-D36A* might serve to stabilize interactions of the spliceosome with the suboptimal pes1-5 due to its suboptimal 5′ region (and likely secondary structure). For these assays, fitness was calculated using pes1-5 wild type strain as the reference, so an increase in fitness will be seen data above the horizontal y = 0 line (no change). Interestingly, neither *YHC1-D36A* nor mBBP alone are sufficient to provide much improvement in splicing, although at higher copper concentration, an increase in fitness is seen for both mutations. However, the combination of both mutations (third bar, purple in each facet) demonstrates consistently higher growth than either mutation alone. These results show mBBP and *YHC1-D36A* both stack synergistically, while either mutation alone is not sufficient to rescue any splicing of pes1-5. This implies that the h5′ss_hIB1 splicing defect can be rescued by *YHC1-D36A* mutation, which may indicate improved spliceosome assembly through a bridging interaction between mBBP and the U1 snRNP during Complex E formation.

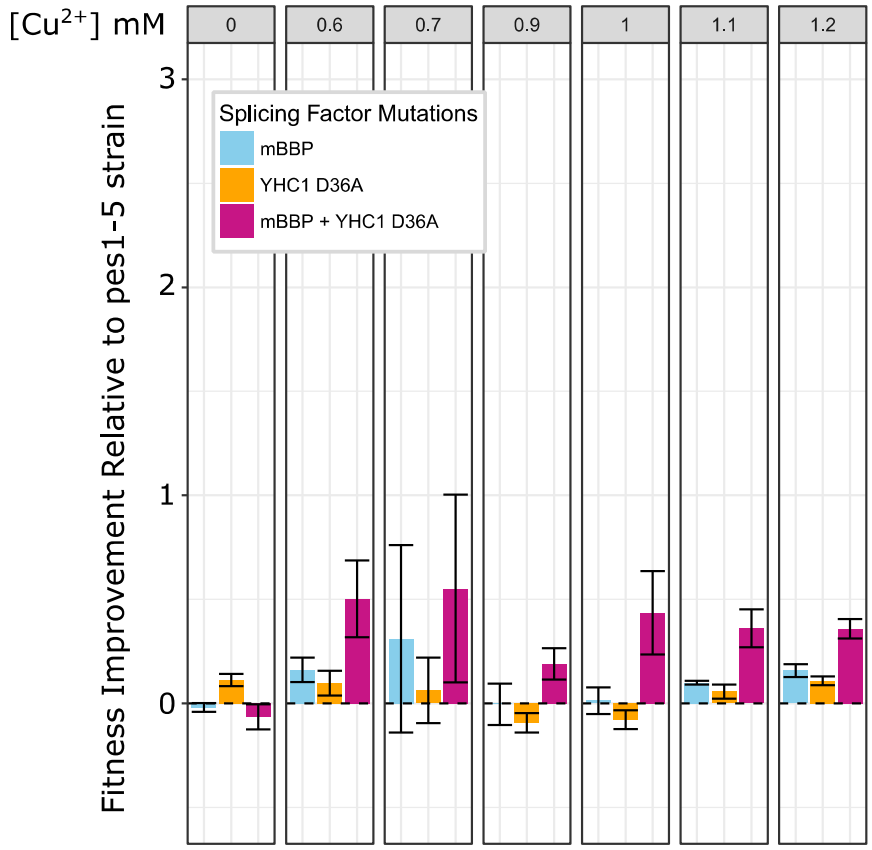


Figure 36. Fitness changes due to synergy between mBBP and *YHC1-D36A*. Fitness improvements are shown for each mutation relative to growth of the strain with the *CUP1*-pes1-5 reporter and a wild type spliceosome. This graph emphasizes the splicing factors mutations introduced into the genome using CRISPR-Cas9 in single or in combination. The scale is set to emphasize comparisons with Figure 41.

6.4 Improve splicing of pes1-5 by increasing accessibility of splicing factors

Influential work by the Ares group has shown that splicing of suboptimal transcripts improves when there is reduced competition for splicing machinery (267). In yeast, 90% of intron-containing transcripts in vegetative cells are from ribosomal protein genes (RPGs) (268-270) and 101 of 293 intron-containing genes are RPGs (267). As demonstrated by Munding et al., expression of RPGs requires extensive use of the splicing machinery and reducing the total amount of RPG transcripts significantly decreases the pool of transcripts competitive for the spliceosome. The limiting factors become available to act on challenging transcripts, global splicing is improved, and certain growth defects caused by mutations in the spliceosome can be ameliorated. Downregulating this gene decreases the number of ribosomal protein gene transcripts, which compete for precious splicing machinery.

To prove the concept of *IFH1* downregulation in their work, the Ares group utilized strains with the GAL promoter integrated in front of the *IFH1* ORF (267). This repressed gene expression when yeast were grown on glucose and allowed for excellent testing of the “hungry spliceosome” hypothesis. For our engineering purposes, we needed to be able to quickly introduce *IFH1* downregulation into multiple yeast strains and genomic reporter replacement can be challenging. Therefore, to flexibly introduce *IFH1* downregulation into different strain backgrounds we developed an inducible CRISPR-interference strategy to repress transcription of *IFH1.* A plasmid containing an expression cassette for deactivated/dead Cas9 (dCas9) fused to a human transcriptional repressor domain (Mxi1) is induced by the addition of anhydrotetracycline (ATc) (271). The introduction of this plasmid and repression system necessitated using only liquid culture media for these experiments.

The results of liquid growth assays are shown below (Figure 37). No growth defect is apparent for strains at 0 mM copper, indicating that leaky repression of *IFH1* does not have a significant negative impact of cell fitness. However, addition of the CRISPRi plasmid offers a growth benefit to each strain tested after copper is added to the media. Particularly, growth of the strain with *CUP1*-hBPS in a wild type splicing background is increased dramatically by repression of *IFH1*. However, mBBP still represents a more powerful and specific solution for suboptimal BPS, but it is truly impressive how RPG downregulation globally improves splicing of difficult introns. Downregulation of *IFH1* in the strain with *CUP1*-hBPS and mBBP (in place of genomic wild type BBP) does not provide much added benefit, indicating the mBBP is capable of strong splicing recovery of hBPS. Splicing of pes1-5 in a wild type spliceosome background shows growth improvement upon induction of the CRISPRi repression system. This indicates that splicing of pes1-5 can be improved upon repression of RPGs. Additionally, a similar growth improvement can be observed in strains with *CUP1*-pes1-5 and mBBP, *prp16*-G373S, and *prp28*-E404K mutations, indicating that repression of *IFH1* stacks additively with these three mutations.

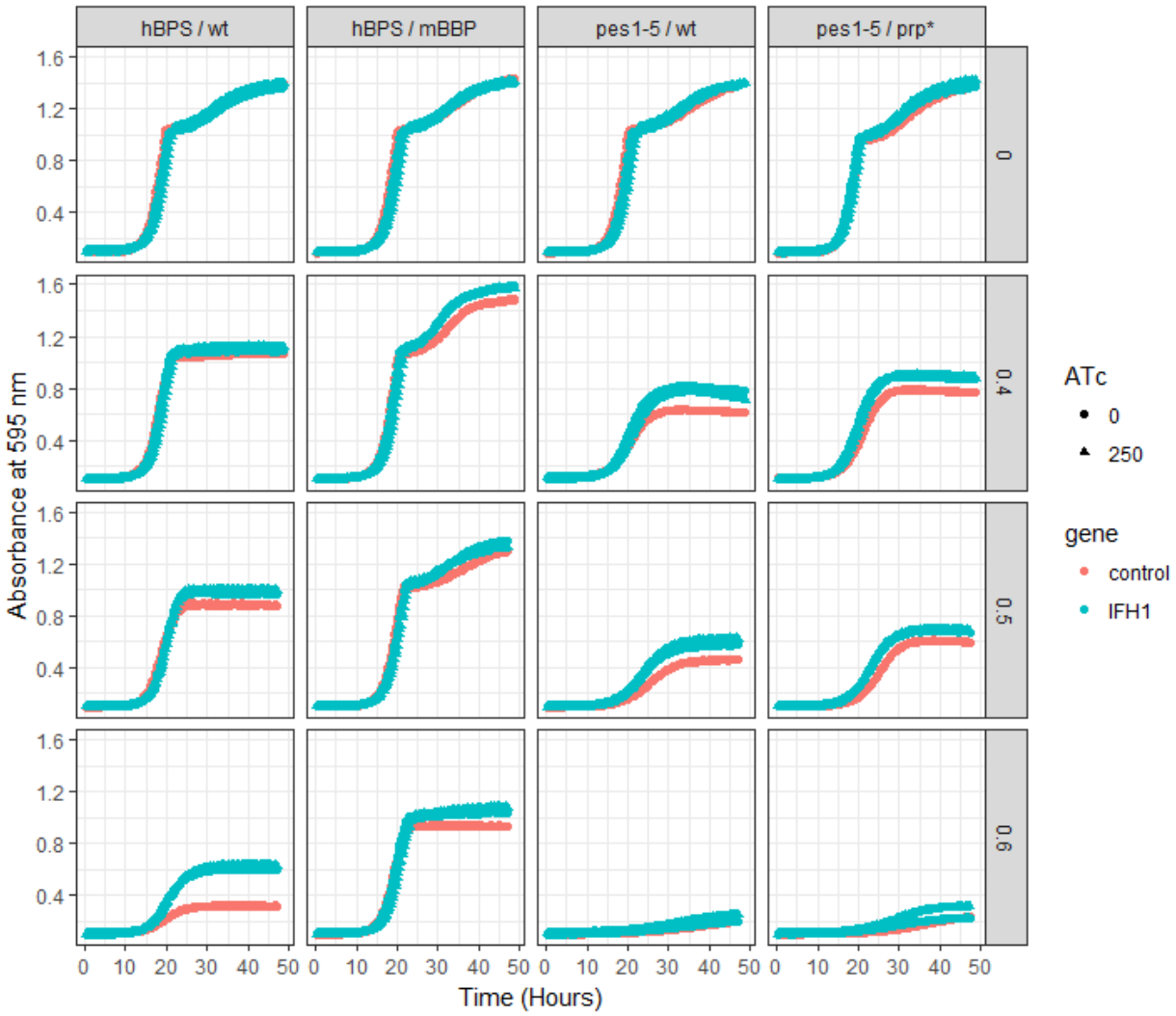


Figure 37. Growth curves for CRISPRi repression of *IFH1*. Liquid growth curves examining the impact of CRISPRi downregulation of *IFH1* on different introns in strains with different splicing factor mutations. The left to right faceting indicates a different yeast strain ("intron / splicing factor mutations" is shown above each facet); wt = wild type, prp* = mBBP + *prp16*-G373S + *prp28*-E404K (condensed to save space). The top to bottom faceting is a different copper concentration (mM). The blue curves are strains containing the *IFH1* CRISPRi plasmid (triangle is addition of ATc and circle is no ATc). There is significant overlap between lack of inducer or presence of inducer for this assay, indicating that the repression system is leaky. The red curves are the same strain with an empty control vector and represents a true negative control.

6.5 Testing combinatorial strain improvements

Following the spliceosome engineering methodology introduced in Ch. 6.1, we identified five splicing factor mutations that operate through distinct mechanisms to improve splicing of the *A. fumigatus* intron pes1-5. We wanted to identify the most effective combination that required fewest number of changes to the spliceosome but still resulted in fitness improvement as detected by

copper assays. Specifically, we wanted to examine mBBP, *YHC1-D36A*, and *IFH1* repression compared to *prp16*-G373S and *prp28*-E404K. All strains grow approximately the same at 0 mM copper indicating no strong growth inhibition is occurring. At 0.5 mM, the strain growths begin to separate and interestingly mBBP + *YHC1-D36A* performs as well as mBBP + *prp16*-G373S + *prp28*-E404K. Important to note was the phenotype exerted by mBBP + *dIFH1*. At lower copper concentrations (0.5 mM), it provides a lower growth phenotype than mBBP + *YHC1-D36A* or mBBP + *prp16*-G373S + *prp28*-E404K. However, at 0.7 mM and above, mBBP + *dIFH1* has the highest growth on copper media.

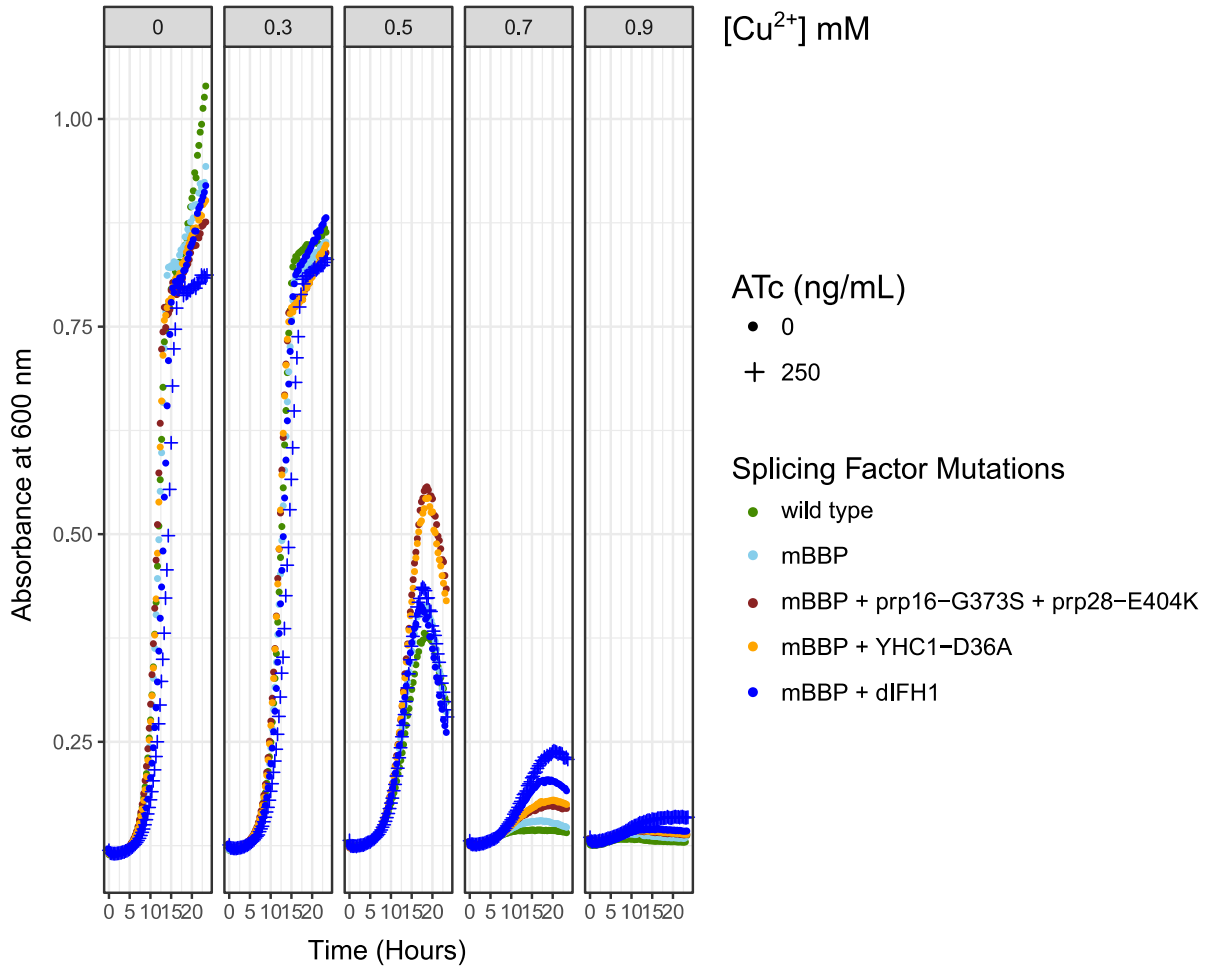


Figure 38. Comparative analysis between RNA helicase mutations and *YHC1-D36A*. Five strains are tested in a liquid growth assay in 96-well plates. The left to right facets each represent a different copper concentration (mM, shown at the top of the facet). ATc was only added to mBBP + dIFH1 in order to activate the CRISPRi repression system (repress *IFH1* expression). All strains contain the *CUP1*-pes1-5 integrated reporter.

These results indicated that *YHC1*-D36A stacked synergistically with mBBP and performed equivalently to the strain with mBBP + *prp16*-G373S + *prp28*-E404K. This was a positive result because stabilization of Complex E formation provides as much improvement to splicing as slowing the proofreading capabilities of the spliceosome. Therefore, these mutations should prevent less accurate spliced isoforms from being created by a spliceosome with reduced fidelity. Given the positive synergy simultaneously observed with mBBP + *dIFH1*, we decided to introduce the CRISPRi plasmid into the strain background of mBBP + *YCH1-D36A.*

Results from growth assays in 96-well plates are shown below (Figure 39). Additional controls were tested and the results are displayed. These results show the fold change of the area under the curve values (see Materials and Methods). Data above the line indicate growth improvement while data below the line is decreased fitness. The negative control S1 (*cup1*Δ) shows growth below that of the reference strain: wild type spliceosome with *CUP1*-pes1-5. This result is in agreement with results from the liquid spotting assay displayed previously in this chapter. As a positive control, the strain with integrated *CUP1*-MATa1 and a wild type spliceosome is also grown and easily displays the highest fitness improvement over pes1-5 splicing, which is to be expected. This represents the upper limits of obtainable splicing of this intron. The remaining strains all have the reporter *CUP1*-pes1-5 integrated into the genome and different splicing factor mutations. The best performing mutant combinations at the copper concentrations tested are mBBP + *dIFH1* and mBBP + *YHC1-D36A* + *dIFH1*, with the triple mutant showing the highest fitness improvement in this experiment.

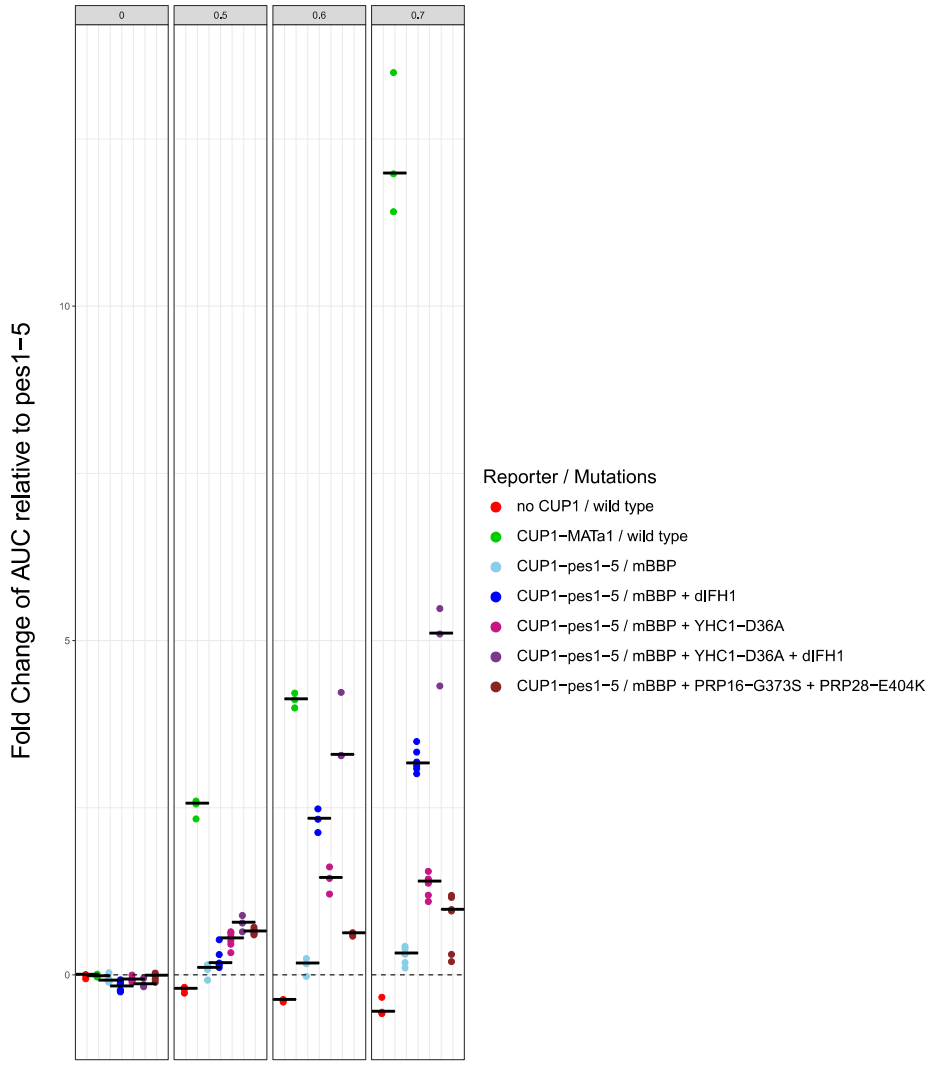


Figure 39. Splicing factor mutations cause additive improvements in yeast growth. Liquid growth assay performed in a 96-well plate. Growth curves were converted into an area under the curve value by the trapezoidal rule and then further converted into fold-change values relative to the AUC of pes1-5 in a wild type background. Data points above the y = 0 horizontal dashed line indicate improved growth relative to a wild type yeast with pes1-5. Data below the dashed line indicate reduced growth. Each different color is a different yeast strain and the legend displays the reporter / mutation combinations of each strain. The order of the legend mirrors the order of the strains along the x-axis. Each facet represents a different copper condition.

These results conclusively show that the most optimal combination of splicing factor mutations of the mutations studied here is mBBP + *YHC1-D36A* + *dIFH1*. Considering the potential negative ramifications of modification of *PRP16* and *PRP28,* and the lower copper resistance provided, we decided not to pursue these strains further. Instead we were interested in exploring additional combinations and conditions of mBBP, *YHC1-D36A*, and *dIFH1*.

6.6 Further characterization of *YHC1-D36A* and *IFH1*

To specifically look at the combinations of mBBP, *YHC1-D36A*, and *dIFH1*, we repeated the 96-well plate growth experiment. For each strain with a transformed CRISPRi plasmid (*dIFH1*), ATc was added to 250 ng/mL or not added. The results from the growth curve show that the presence of the CRISPRi plasmid does confer some growth improvement even without inducer ATc, indicating the CRISPRi system is leaky. However, the growth improvement increases upon addition of ATc, indicating the CRISPRi repression system is still being induced by ATc. This growth assay shows the same trends of these three splicing mutations stacking additively to offer greater growth improvements over a wild type spliceosome.

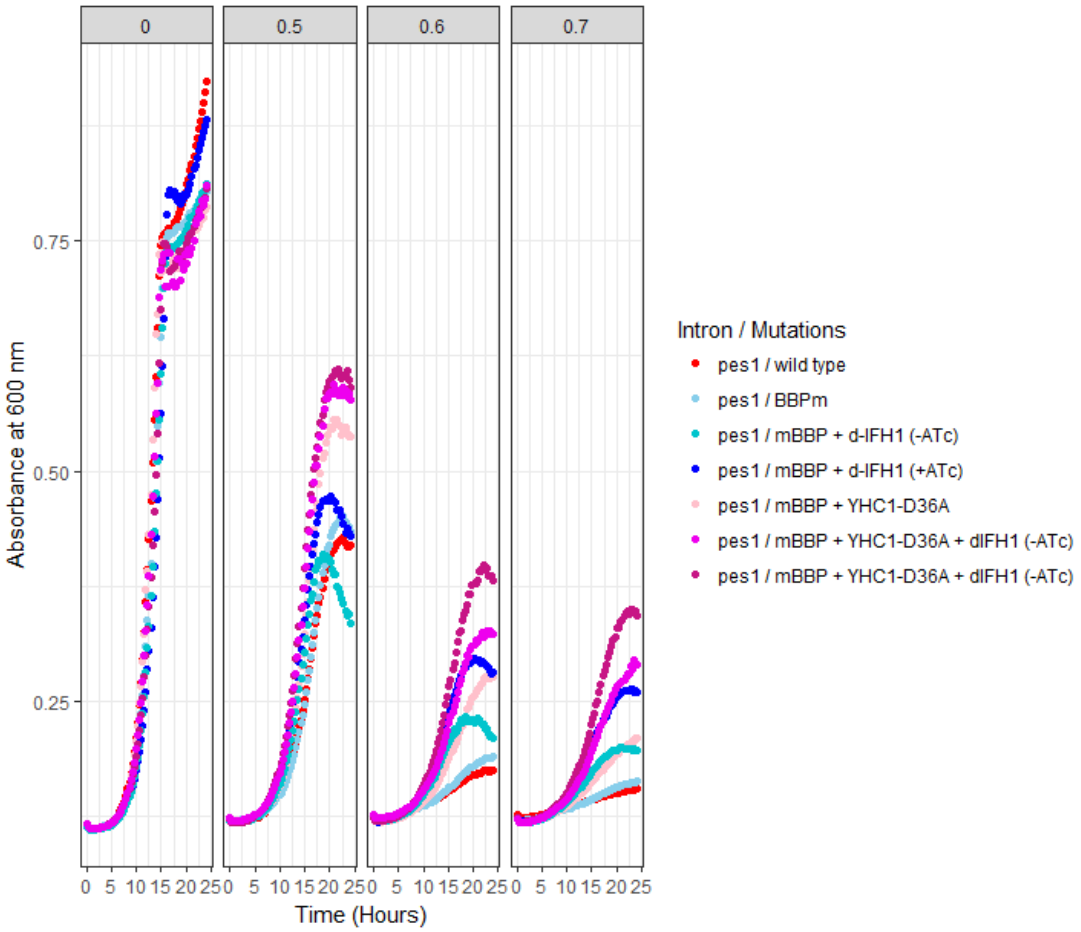


Figure 40. Growth assay reveals splicing synergy between mBBP, *YHC1-D36A*, and *dIFH1.* Liquid growth curve experiment from a 96-well plate. The left to right facets are different copper concentrations (mM). Each color represents a strain with different splicing factor mutations (or the absence/presence) of CRISPRi inducer ATc. All strains have the integrated *CUP1*-pes1-5 reporter.

Figure 41 shows that the mBBP and *YHC1-D36A* mutations both stack additively with the knockdown of *IFH1*. Results from Figure 41 indicate that the single most beneficial mutation for pes1-5 splicing is repression of *IFH1*, which indicates that depletion of ribosomal gene transcripts still allows enough Cup1 protein to be produced to offer an improved fitness phenotype in growth experiments. Additionally, this mutation still stacks well with other splicing mutations. The combination of all three mutations (mBBP, *YHC1-D36A*, and *dIFH1*) shows the highest fitness at copper concentrations at 0.9mM or below. Interestingly, the fitness advantage from the combination of all three mutations stops increasing after 1.0 mM, perhaps indicating that a balance is reached between growth advantage from superior pes1-5 splicing and growth defects caused by multiple spliceosome mutations and RPG depletion. The combination of mBBP and *dIFH1* begins to show the highest growth phenotype at 1.0mM and above, followed by the combination of *YHC1-D36A* and *dIFH1*. Overall, the growth defect from the combination of all splicing mutations becomes exacerbating at high copper conditions points toward a potential limitation of using a selection assay to screen for splicing improvements.

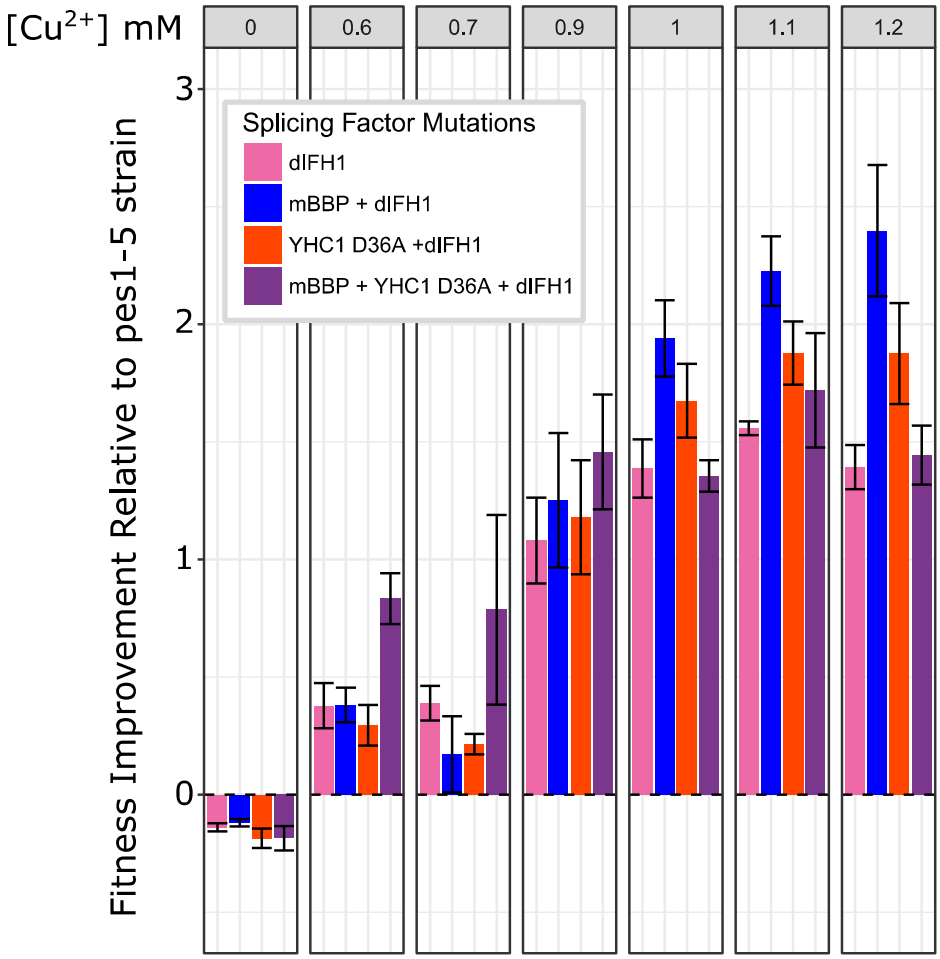


Figure 41. Fitness improvements across a range of copper concentrations. The fitness improvements over a wild type strain using combinations of mBBP, *YHC1-D36A,* and *IFH1* downregulation (through CRISPRi plasmid).

To conclude with the most effective general methodology to address the failure modes identified by this study, we developed a comprehensive approach to spliceosome engineering (Figure 42). Three strategies were deployed: 1) improve splicing factor recognition of substrate intron (mBBP); 2) improve stability of spliceosome assembly by addressing the two failure modes of pes1-5 (using mBBP and *YHC1-D36A*); and 3) increase availability of limiting splicing factors (through *IFH1* repression). The combination of these three general strategies demonstrated splicing improvements that "stack" additively and allowed for increased splicing of pes1-5 as demonstrated by fitness assays and RT-qPCR.

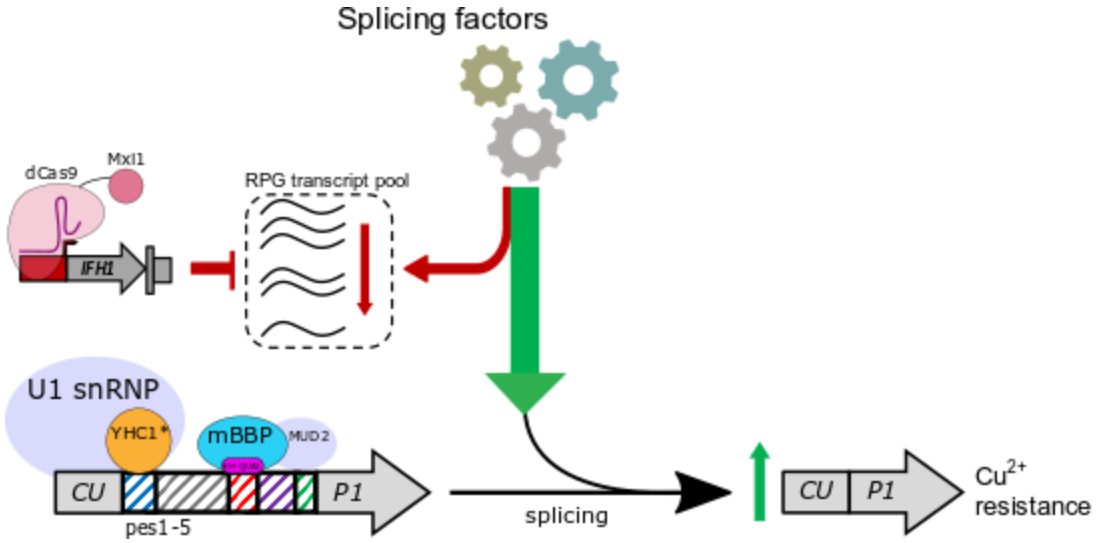


Figure 42. Overview of the most effective mutations for improving splicing of a fungal intron. A combinatorial approach to spliceosome engineering of *S. cerevisiae* for enhanced fungal intron splicing. Cartoon schematic detailing the three splicing factor mutations that enable improved splicing. Complex E formation is stabilized by introduction of mBBP and *YHC1-D36A* mutations. Accessibility of splicing factors is increased by transcriptional repression of *IFH1* using CRISPRi.

Figure 43 shows a representative growth curve used to calculate the fitness improvements of mutant strains. The wild type background with MATa1 intron is shown by green triangles and grows approximately the same at both 0 mM and 0.7 mM (facet 1 and 2, respectively). The *cup1*Δ strain is shown by red circles and has the weakest growth at 0.7mM. The strain containing pes1-5 and a wild type spliceosome displays slightly better growth, indicating very limited splicing. With three mutations introduced into this strain, splicing is improved and the strain achieves higher copper tolerance (purple plus signs). However, it is important to note that its growth is slightly reduced at 0mM copper compared to wild type as multiple splicing mutations begin to cause growth defects. Particularly *YHC1-D36A* and *IFH1* repression, as mBBP alone was shown not to inhibit growth. However, temperature sensitives were assayed at 16°C, 25°C, 30°C, and 37°C on agar plates (YPD or -ura agar plates where appropriate) and no temperature-specific growth defects were detected in any of the mutants (Appendix).

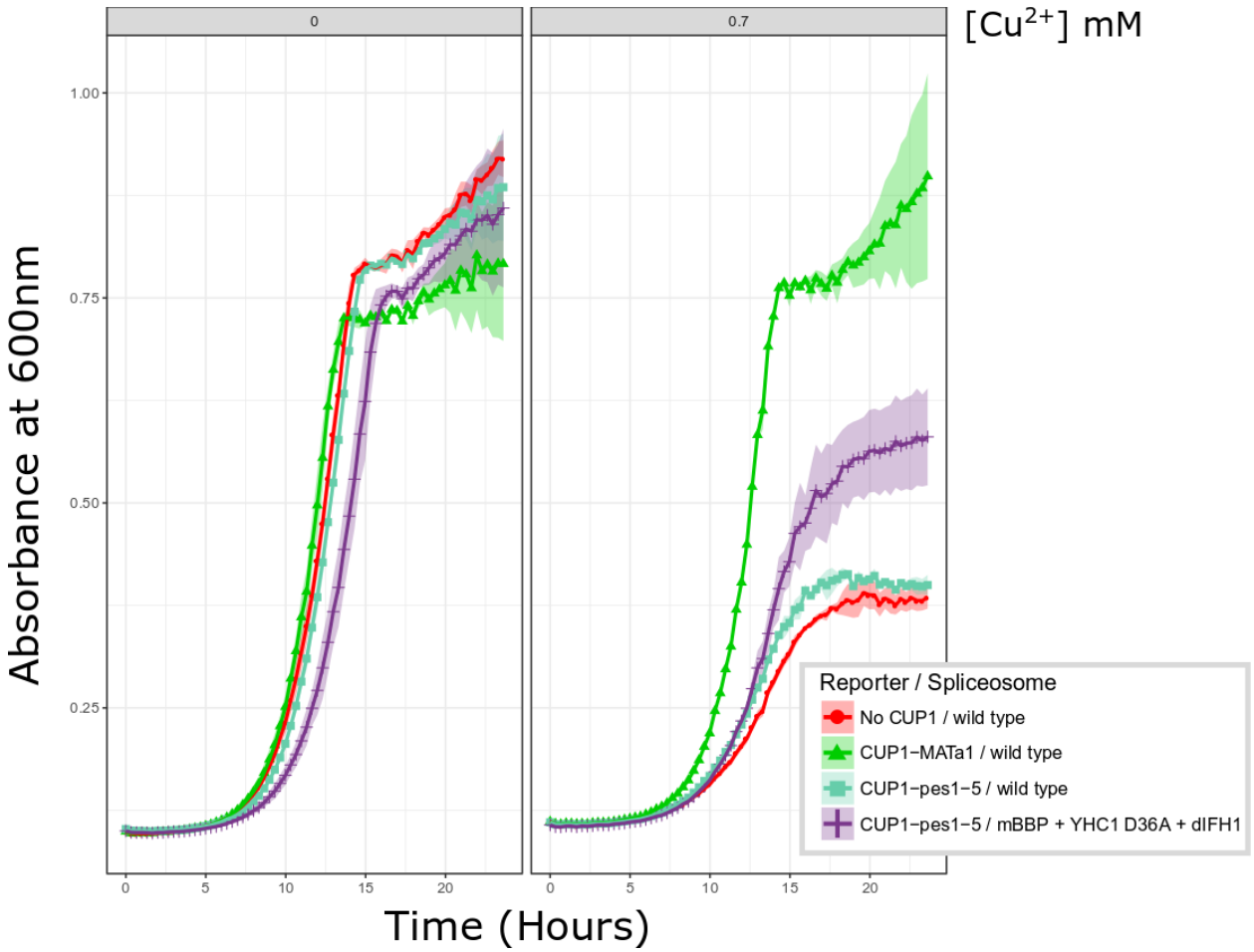


Figure 43. Growth curve detailing growth improvements from three splicing mutations. Liquid growth curves are shown for four different yeast strains with the standard deviation of replicates represented by the ribbon: 1) S1, the *cup1*Δ strain, as a negative control in red circles; 2) S4, the *CUP1*-MATa1 strain with wild type spliceosome; 3) S12, the *CUP1*-pes1-5 strain with wild type spliceosome; and 4) S15, the *CUP1*-pes1-5 strain with all three splicing factor mutations. Two growth conditions are shown, YPD with 0 mM [Cu^{2+}] (left facet) and YPD with 0.7 mM [Cu^{2+}] (right facet).

We next quantified splicing improvement via RT-qPCR (Figure 44). As overall strain health is poor at higher copper conditions despite improvements in relative fitness, the three mutant strain was extracted. The amount of spliced product relative to geometric mean of the reference genes *ACT1* and *SEC65* is 1.6-fold higher than in a wild type strain. We observed a mean effect magnitude of 1.55 with a 95% confidence interval of [1.20, 2.03]. This is a significant improvement in splicing (p value = 0.0046, two-sample permutation test). We achieved the first demonstration of improving splicing of a poorly spliced *Aspergillus* intron through spliceosome engineering in *S. cerevisiae*.

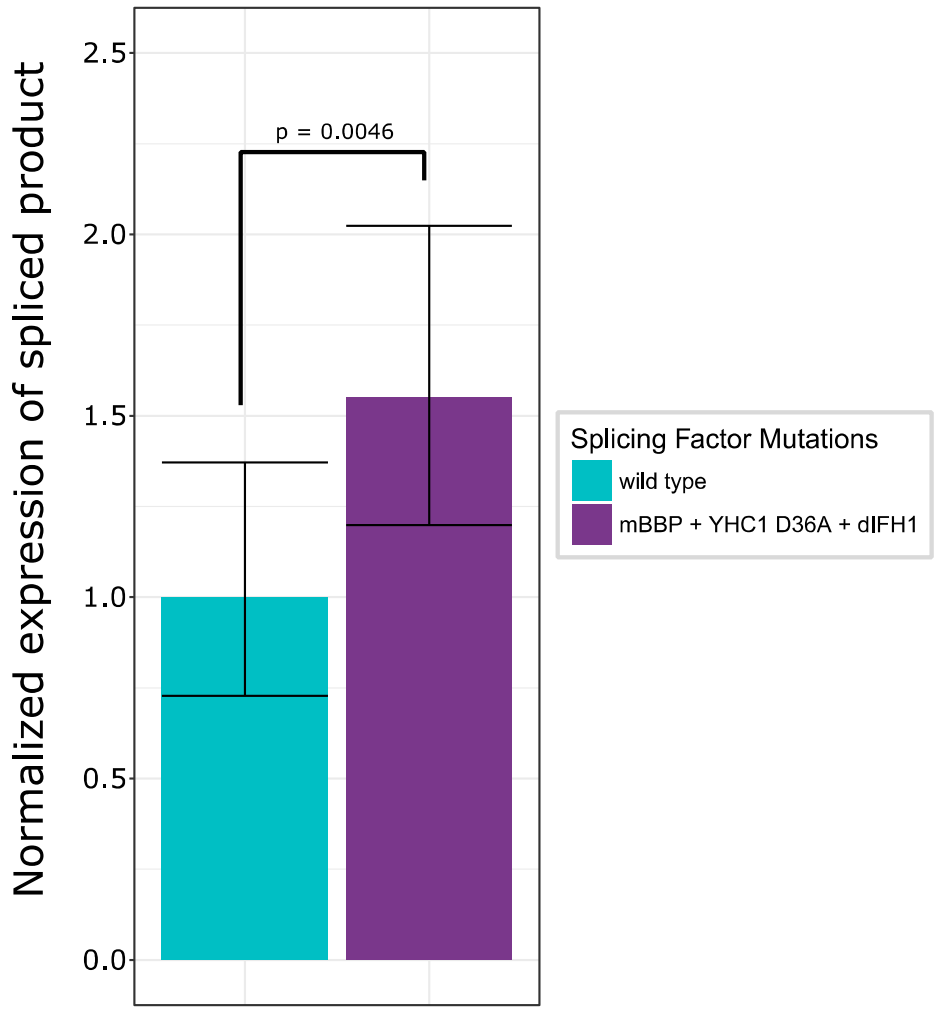


Figure 44. Splicing of the pes1-5 intron is significantly improved in the engineered strain. RT-qPCR results showing a significant increase of the spliced product in the engineered strain. Cq values of spliced product were normalized by the geometric mean of the Cq values of *ACT1* and *SEC65* reference genes. The wild type strain served as the calibrator for the engineered strain. Error bars are transformed 95% confidence intervals (see Materials and Methods). The data is derived from 12 data points for each strain (4 biological replicates with 3 qPCR replicates each).

6.7 Materials and Methods

Strains

To modify the splicing factor genes, a flexible CRISPR-Cas9 system was deployed. Two plasmids are needed as templates to generate three PCR products needed for the CRISPR-Cas9 plasmid: 1) pCB30 (G418, Addgene #pending) or pCB32 (hyg, Addgene #pending) depending on the antibiotic resistance desired; and 2) pCRCT, a gift from Huimin Zhao (Cas9, Addgene #60621,

(226)). All three PCR products were amplified by Accuprime Pfx Polymerase (Invitrogen) and primer sequences used are in Supplementary Table. The primers used to amplify the iCas9 expression cassette have homology to the products amplified from the pCB plasmids to form a complete plasmid *in vivo* through yeast homologous recombination. The primers used to amplify the pCB plasmid are designed to have homologous overlap at the 20bp gRNA sequence site, allowing any desired guide sequence to be designed into the oligo (IDT). Guide sequences were chosen using CRISPy Cas9 target finder, (244). All 3 PCR products were co-transformed with the donor DNA into the recipient yeast strain. Donor DNA for *PRP16* and *PRP28* mutations also included a silent mutation at the PAM site of the CRISPR guide RNA to prevent cutting a repaired locus. After the heat shock, strains were inoculated into 3 mL YPD and shaken overnight at 30°C. Cultures were diluted and plated for single colonies on the appropriate selection for the CRISPR-Cas9 plasmid. Efficiencies varied per genomic loci targeted. mBBP: out of 10 colonies screened, 60-80% would have mutation. For *YHC1-D36A*: 1/6 would have the mutation. There was no great difference between the efficiencies for either *PRP16* or *PRP28* and approximately 1/4 to 1/6 colonies screened via sequencing contained the point mutation.

All yeast transformations in this study were carried out using the LiOAc/ssDNA method (214). Transformants were selected using leucine dropout media, uracil dropout media, 5FoA media (0.8g/L – 1g/L), or 2% YPD supplemented with 200 µg/mL hygromycin and 200 µg/mL G418. All genomic mutations were confirmed by colony PCR amplification of the region affected plus 50-500 bp of additional region upstream or downstream of the homology targets, to confirm specificity of the recombination event. PCR products were sequenced by Sanger sequencing (Laragen).

Plasmids

Standard cloning techniques were performed to maintain and propagate the plasmids in *E. coli* DH10b. pRS416gT-Mxi1 was a kind gift from Kevin Roy, Justin Smith, and Bob St. Onge

(271). The guide region to repress IFH1 (5′-TGTTAGAATTGTGGAAAGGG) was synthesized as part of two overlapping oligos (each a 60-mer). These oligos primed each other to form a short PCR amplicon that was assembled with NotI-digested pRS416gT-Mxi1 using Gibson Assembly.

Liquid Spotting Assay

Strains were inoculated into YPD or uracil dropout liquid media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. Optical Densities (ODs) were measured by spectrophotometer at 600nm for each strain and cells were normalized to each other by OD and then serially diluted into sterile water. 5 µL to 20 µL were spotted and plates were incubated at 30°C for 3-7 days before imaging. Temperature growth assays were performed by spotting serial dilutions on YPD or uracil dropout media at 16°C (imaged after 12 days), 25°C, 30°C, and 37°C (imaged after 4 days).

Growth Curves

Strains were inoculated into YPD or uracil dropout liquid media from agar plate colonies or frozen stocks and grown overnight, shaking at 30°C. ODs were measured by spectrophotometer at 600nm and colonies were sub-cultured to OD 0.01 in 100 µL of culture per well in 96-well microtiter plates. Absorbance at 600nm was measured every 15 minutes over the course of 24 hours in a Tecan M200 Pro held at 30°C and shaking continuously alternating between linear shaking (before measurement) and orbital shaking (post-measurement) for each kinetic cycle.

Growth Curve Analysis

Raw data was exported to Excel from iControl Software for the Tecan M200 Pro and imported into R versions 3.4.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Packages from the Tidyverse (Hadley Wickham (2017). tidyverse: Easily Install and

Load 'Tidyverse' Packages. R package version 1.1.1. https://CRAN.R-project.org/package=tidyverse) were used to manipulate the data for analysis and presentation with ggplot2 (Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009)). Fold change calculations were determined in an R script as follows: 1) the mean for the first 5 measurements of each strain replicate at each copper condition was calculated and subtracted from all absorbance readings to set the baseline of each growth curve to zero; 2) the area under the curve (AUC) was calculated as the sum of all adjusted absorbance readings using the trapezoidal rule; 3) the AUC of *CUP1*-pes1-5 strain was used as the control value in the fold change equation: $(\mathrm{AUC}_{\mathrm{sample}} - \mathrm{AUC}_{\mathrm{reference}})/\mathrm{AUC}_{\mathrm{reference}}$.

RNA extraction

For pes1-5 strains, strains were inoculated into YPD or uracil dropout media without Cu^{2+} overnight. The following day, strains were sub-cultured to an OD_{600} of 0.5 in fresh 5 mL of YPD media containing 1.2mM Cu^{2+} and cultured for 7 hours; each strain with four biological replicates. 250 ng/mL anhydrotetracycline (ATc) was added to the strain containing the CRISPRi plasmid before addition of Cu^{2+}. OD was measured by spectrophotometer and cell count was normalized to 1 mL per OD_{600} 1.0 prior to extraction.

All samples were centrifuged, media was removed, and cell pellets were resuspended in 1 mL RNAprotect Cell Reagent before incubation at room temperature for 15 minutes. RNAprotect Cell Reagent was removed via centrifugation prior to freezing. Cell pellets for pes1-5 experiments were frozen prior to extraction at -80°C. To digest the cell walls, cells were incubated for 1 hour at 37°C with gentle shaking in a zymolyase solution (100 µL RNase-free water, 10 µL zymolyase (Zymo Research), and 0.1 µL beta-Mercaptoethanol). All RNA was extracted using a slightly modified form of RNeasy Mini Plus Kit (Qiagen). On column DNase digestion (Qiagen) was extended to 1 hour. RNA was eluted from Qiagen columns using 100 µL 10 mM EDTA.

cDNA synthesis

For pes1-5 strains, the spliced product was difficult to reliably detect from total RNA as Cq values were consistently above 35. Therefore, RNA extracts were further cleaned and concentrated using RNA precipitation. For volumes of 100 µL, 1 µL glycoblue, 10 µL 3M NaOAc, and 250 µL 100% EtOH (ice-cold) were added and samples were stored in -20°C for overnight. The following day, samples were spun at top speed for 30 min at 4°C. Supernatant was removed carefully with pipette. 200 µL of ice cold 70% ethanol was added to wash and samples were spun for 10 minutes at top speed at 4°C. Ethanol was removed via pipette and the wash step was repeated. Samples were spun down and all ethanol was removed and samples were resuspended in 21 µL RNase-free water for Nanoquant plate quantification (Tecan M200 Pro). Approximately 1 µg of each RNA template was used to prepare cDNA. For single-stranded cDNA synthesis, SuperScript III Reverse Transcriptase (ThermoFisher Scientific) was used according to the manufacturer's protocol, including an RNase H digestion to remove the RNA template. cDNA synthesis was achieved using a mixture of primers amplifying the 3′ end of *CUP1*, *ACT1*, and *SEC65*.

Quantitative RT-PCR experiments

Primers for RT-qPCR were designed using primer3 (215,216) or by hand when appropriate, such as for the splice junction primer. cDNA samples were diluted 1:6 and 5 µL of these diluted templates were added to each well on a 0.2 mL skirted 96-well PCR plate (Thermo Scientific, AB-0800/W). The template for each biological replicate was added to three separate wells to generate three qPCR replicates per biological replicate. A master mix for the Luna Universal One-Step RT-qPCR Kit (NEB, #E3005L) was generated for each primer pair used and 15 µL was aliquoted into each well and plates were sealed with Microseal 'B' seals (BioRad #MSB1001). Samples were run in a CFX96 Real-Time System with C1000 Thermal Cycler. The protocol for RT-qPCR was as follows: 55°C for 10 min (RT), 95°C for 1 min, [95°C for 10 sec, 60°C for 30 sec, plate read] x

45 cycles. Subsequent melt curve analysis protocol: 60°C incremented by 0.5°C to 95°C, holding temperature for 5 sec.

RT-qPCR analysis

Raw Cq values were exported to Excel from Bio-Rad CFX Manager 2.0 (determined by regression, algorithm similar to PCR Miner (217) and then imported into R for analysis. RT-qPCR experiments were performed using an all genes approach where samples were split across multiple plates but all genes were tested for each plate. Therefore, ΔCq (spliced – reference) values for each strain were calculated for each plate before combination of all replicate values. The reference Cq value was the geometric mean of the Cq values from *ACT1* and *SEC65* amplicons. The mean ΔΔCq values were calculated from this pooled dataset (218) using a two-sample magnitude bootstrap with replacement to generate the 95% confidence intervals (CIs) around the effect size. The mean of the resampled ΔCq values for the calibrator strain (containing the *CUP1*-pes1-5 reporter in a wild type strain background) was subtracted from the mean of the resampled ΔCq values for the test strain (*CUP1*-pes1-5 with 3 splicing factor alterations) to generate a simulated ΔΔCq. This process was repeated for 10,000 iterations to generate the 95% confidence intervals. For each strain, the mean ΔΔCq value from the experimental data, upper CI value, and lower CI value were transformed into fold change values using $2^{-\Delta\Delta Cq}$ and then plotted in R. A null hypothesis was assumed that the ΔCq values are from the same underlying population, which means that the mean of ΔΔCq should be approximately 0 because there would be no true difference in the population. Statistical analysis was performed using a two-sample permutation significance test comparing the engineered strain to the calibrator strain (pes1-5 wild type). Simulated chance values that were more extreme than the positive and negative absolute value of the observed test statistic (ΔΔCq generated from the ΔCq data) were summed and used to determine the “two-sided” p-value.

7. CONCLUSION

The challenges we are facing to prevent antibiotic resistance and find sustainable energy sources are plentiful. Therefore, there is interest in using all biotechnological advances to mine Nature's diversity in order to find solutions to our medical and energy problems. Advances in high-throughput sequencing technology have already led to immense scientific breakthroughs, but even these results have only offered a taste of the raw potential hidden in Nature. To derive meaning from this avalanche of sequencing information, we will need new computational and genetic tools for physical expression of this genetic data. Heterologous expression of gene sequences in model organisms such as *S. cerevisiae* has already proven to be a foundation of biotechnology and will be increasingly necessary for high-throughput exploration of this new genomic sequence space. Given the high value targets that have come from fungi, there is considerable interest in further exploration of these eukaryotic genomes. Therefore, the intron problem will continue to become a larger and larger issue facing successful heterologous expression of this overwhelming amount of new genetic data.

In this study, we identify the failure modes of intron splicing using an *Aspergillus* intron expressed in *S. cerevisiae*. By breaking the fungal intron into individual "modules" and swapping these with the corresponding modules in a native *S. cerevisiae* intron, we were able to determine which regions of the *A. fumigatus* intron prevented its splicing in yeast. The two failure modes were identified as the heterologous branchpoint site (hBPS) and combined modules at the 5′ end of the intron, heterologous 5′ss and heterologous In-Between 1 (h5′ss_hIB1). We reasoned these caused splicing failure due to lack of recognition by splicing factors (hBPS) and formation of a suboptimal secondary structure (h5′ss_hIB1).

Tools developed in this work were used for building multiple splicing reporters that could be tied to growth phenotypes. Overall, three genes were assayed as splicing reporters: *URA3, HIS3*, and *CUP1,* in order of least to greatest utility. *URA3* provided the initial results, which were consistent with the data from subsequent reporters but it lacked any sensitivity as any inefficiency

in splicing prevented cell growth on uracil dropout media. *HIS3* provided a similar sensitivity to *CUP1*, but ultimately was not as useful because it requires a minimal level of splicing for strains to grow in histidine dropout media. This threshold was not reached when the *A. fumigatus* pes1-5 intron was inserted into the *HIS3* ORF. However, *HIS3* proved a reliable reporter for hBPS studies as the splicing defect of this single module was not severe enough to abolish all splicing. *CUP1* is a splicing field standard and provided sensitive and consistent results across growth and RT-qPCR experiments.

To build and test *CUP1*-intron reporters, we adapted a powerful technology, TranSceInt, to insert introns directly into a *CUP1* reporter integrated into the genome. TranSceInt only required two PCRs (1 for the SceI expression cassette and 1 for the donor DNA intron) and a yeast transformation to easily incorporate an intron seamlessly into the open reading frame of *CUP1*. An intron can be designed as synthetic oligos or a gBlock and then transformed into this recipient strain for further splicing studies. We believe this will save on time spent cloning and also increase result reproducibility because a genomic reporter will produce more stable and consistent results. Integrated reporters using TranSceInt will allow for screening plasmid libraries in high-throughput assays as no additional plasmids are required to maintain the reporter gene. TranSceInt also facilitates efficient insertion of intron libraries at a single locus if a genomically integrated intron library is preferred. Additionally, we have demonstrated the efficacy of using low-copy plasmid reporters to assay for splicing differences in yeast strains with different splicing factor mutations. The cloning procedure is greatly simplified using the BspQI method described in the text and this could allow for screening of intron libraries in high-throughput assays.

The reporters tied successful splicing to improved growth under selective conditions. Liquid growth curves were shown to detect minute changes in splicing efficiency that proved more difficult to see or quantify on an agar plate. Overall, results from growth data collected on solid versus liquid media correlate very well, although the copper concentrations that induce toxicity do vary between media types. However, liquid growth has the advantage of being able to contrast

minute changes in copper resistance at fewer concentrations of copper and also requires less copper per experiment.

As the spliceosome is such a dynamic and complex ribonucleoprotein machine, it presents a highly challenging target to interrogate its mechanisms of action and to engineer it for new functions. We report a methodology to engineer the spliceosome for enhanced function on suboptimal introns, primarily through assembly of stable early-stage spliceosomes to maximize intron retention. CRISPR-Cas9 proved to be a flexible and efficient system for introducing mutations in essential splicing factor genes using only a few PCRs. With this system we introduced a chimeric yeast-fungal BranchBinding Protein (mBBP), which contains the RNA-binding KH-QUA2 domain from *A. fumigatus.* This mutation rescued growth on copper media to nearly wild type levels and enabled a 2-fold improvement in splicing of an intron with the fungal BPS. This is the first study to look at BBP domain replacement with a related homolog for alterations to splicing specificity. The rescue of the mutated yeast intron (hBPS) using mBBP is an unexpectedly powerful result. Previous works to focus on the splicing of grossly substituted BPS have focused on U2 snRNA (223), *PRP5* (235), and *HSH155* (272) mutations. Making the yeast BBP more *Aspergillus*-like should be helpful in studying additional introns considering that *Aspergilli* have demonstrated the capability to splice foreign introns (273), although this is not always the case (274). This study also points toward potential successes of humanizing BBP to study understand disease models of splicing. We dissected the 3 bps in hBPS that differ from the yeast consensus and found that the substitutions A2G and A5G (UACUAAC) were the reason hBPS failed to splice in *S. cerevisiae* unless mBBP was expressed. The BspQI cloning method described was used to build the hBPS-variants as well as to screen a library of BPS mutants in a selection assay to determine differences between wild type yeast and yeast with mBBP.

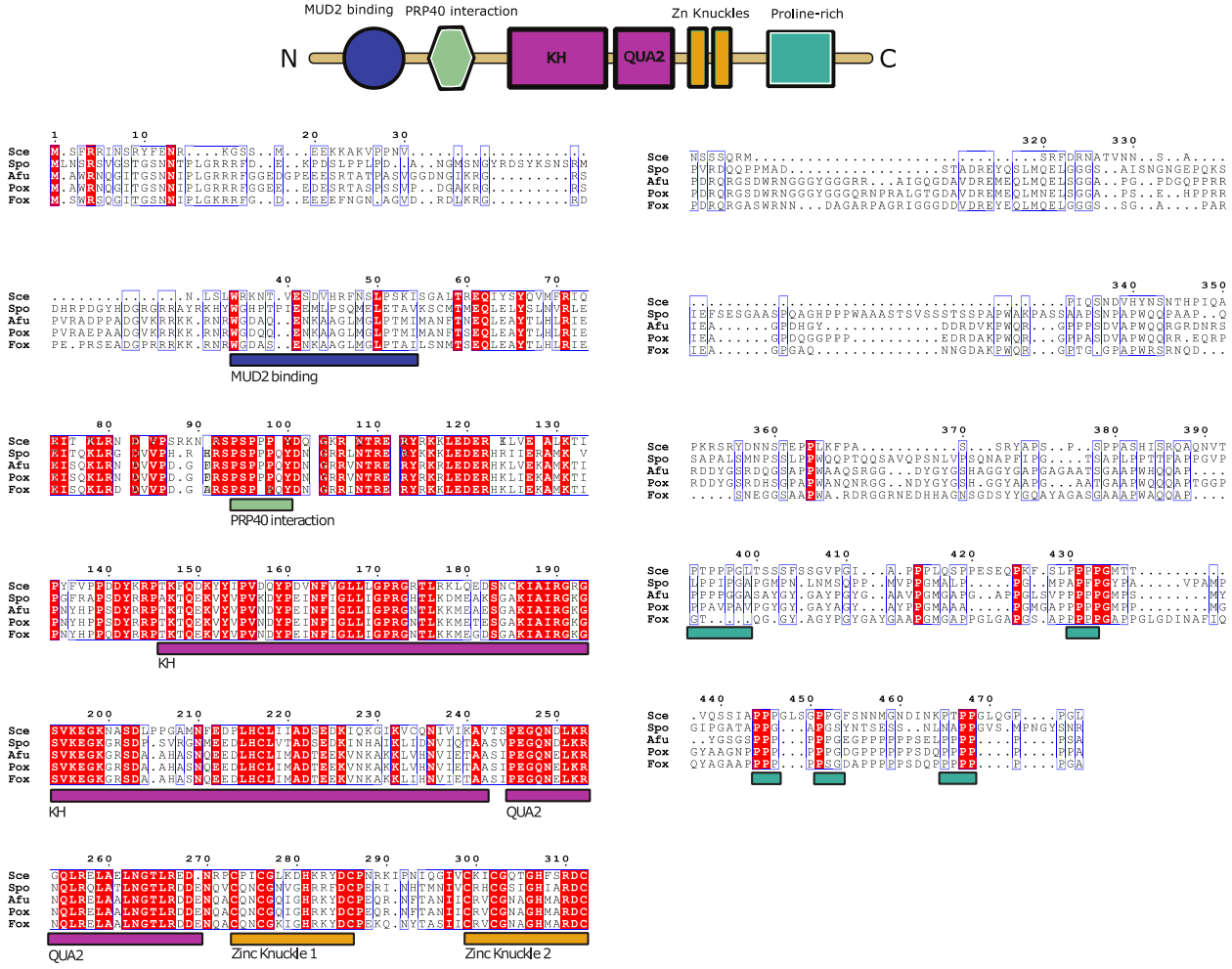
Additionally, we identified multiple synergistic splicing mutations with mBBP that enabled a 1.6-fold improvement of splicing of a fungal intron. mBBP likely stacks with the U1 snRNP mutation (*YHC1-D36A*) by promoting stable association of the U1 snRNP on the pes1-5 transcript.

Formation of a stable Complex E spliceosome prevents leakage of pre-mRNA containing pes1-5 (which has premature stop codon) into the cytoplasm where it will be targeted for degradation by NMD surveillance. Repression of *IFH1* was also shown to stack synergistically with mBBP and *YHC1-D36A* mutations. The logic is that most splicing resources are used by ribosomal protein gene (RPG) transcripts as they are the largest population of intron-containing transcripts in the cell. When the number of these RPG transcripts is reduced, more splicing factor components are available for the splicing of suboptimal transcripts and as a result, these transcripts see an increase in splicing efficiency. This effect is confirmed to be useful for improved pes1-5 splicing. *IFH1* downregulation (*dIFH1*) was achieved using a CRISPRi system for inducible repression. This is the first study to show this “hungry spliceosome” phenotype can be recreated using an easily portable CRISPRi system. Even with modifications to RPG production, the yeast strains would still display an increase in copper resistance, indicating that cell fitness was not negatively compromised by RPG downregulation.

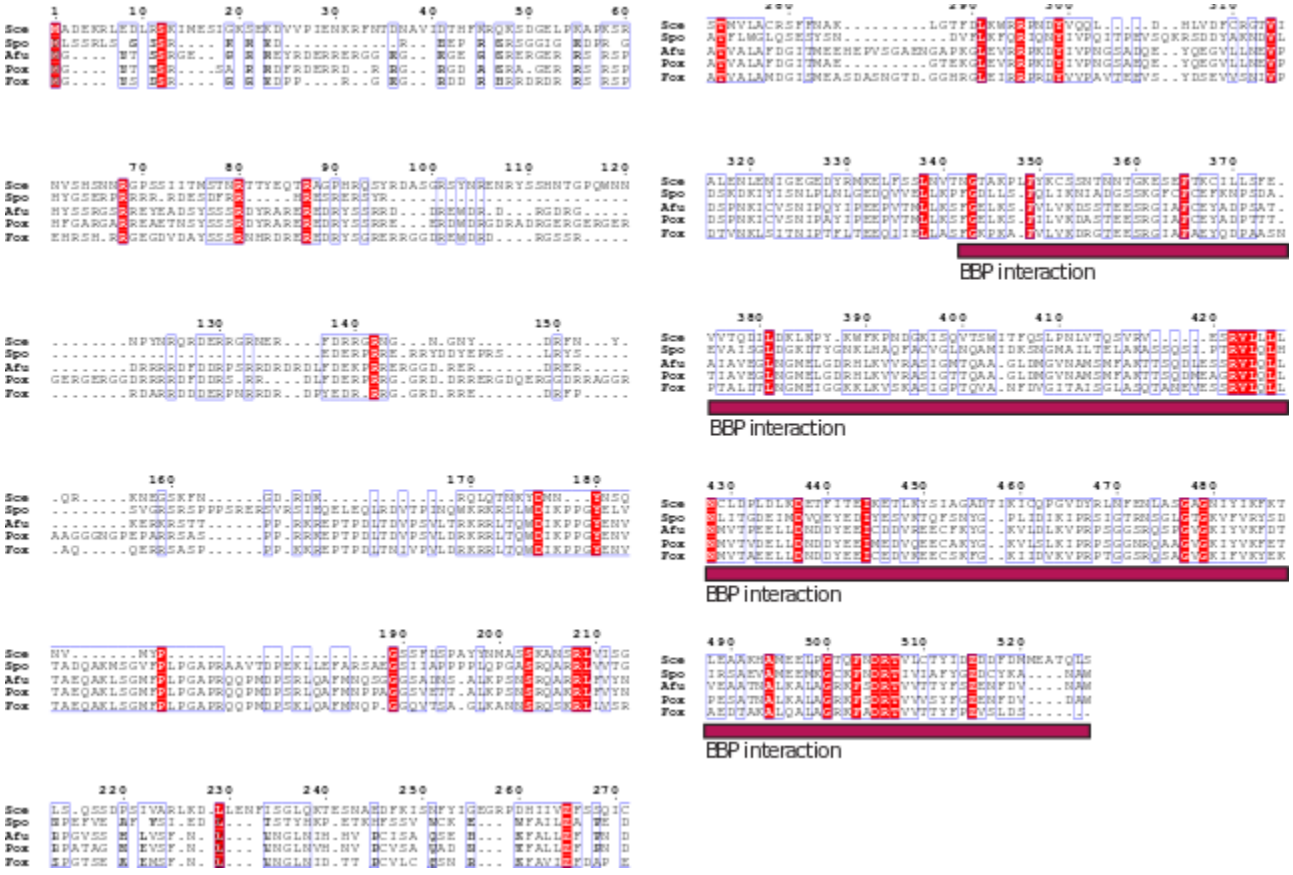
This is the first study to develop a methodology for spliceosome engineering and show that several mutant components of the spliceosome have been additively combined to enhance splicing of a heterologous intron from *A. fumigatus*. This work represents a major step forward in combining the results from research on the spliceosome with engineering techniques to elucidate how changes in intron sequence and mutations in splicing factors map to altered spliceosome function. Applying an engineering methodology to spliceosome studies will have important impacts on heterologous expression of fungal genes as well as providing techniques to study mutant splicing factors with greater throughput. Using the tools, methodologies, and yeast strains provided by this work, the spliceosome will be engineered with new function, enabling rapid discovery of natural products and broadening the scope of how synthetic biology will be used to enhance heterologous expression in diverse research fields, such as in the elucidation of the splicing code and in natural products discovery.

8. APPENDICES

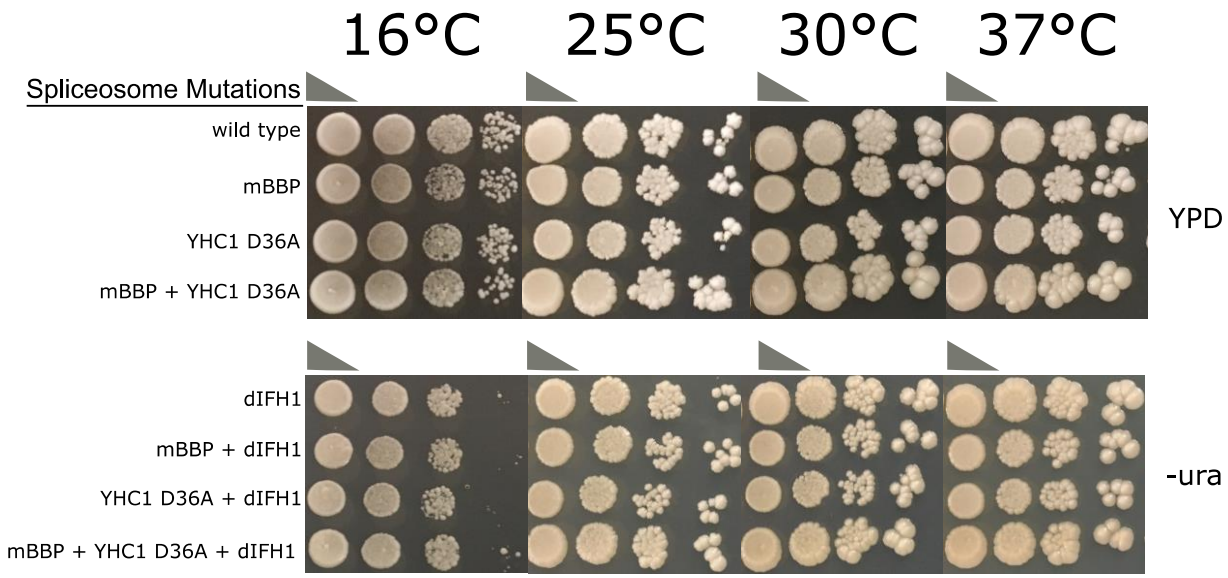
8.1 Full protein alignment of BBPs including *S. cerevisiae* (Sce), *Schizosaccharomyces pombe* (Spe), *Aspergillus fumigatus* (Afu), *Penicillium oxalicum* (Pox), and *Fusarium oxysporum* (Fox).



8.2 Full protein alignment of MUD2 and fungal homologs including *S. cerevisiae* (Sce), *Schizosaccharomyces pombe* (Spe), *Aspergillus fumigatus* (Afu), *Penicillium oxalicum* (Pox), and *Fusarium oxysporum* (Fox).



8.3 Temperature sensitivity assay of multiple splicing factor mutations.



8.4 Plasmids used in this study

Plasmid	Backbone	Description	Reference
	pXP320	CEN/ARS; *HIS3;* AmpR	2011 Fang
pAD01	pXP320	CEN/ARS; *URA3;* AmpR	*This study*
pAD02	pAD01	CEN/ARS; *URA3-MATa1;* AmpR	*This study*
pAD03	pAD01	CEN/ARS; *URA3-h5ss;* AmpR	*This study*
pAD04	pAD01	CEN/ARS; *URA3-hIB2;* AmpR	*This study*
pAD05	pAD01	CEN/ARS; *URA3-pes1-5;* AmpR	*This study*
pAD06	pAD01	CEN/ARS; *URA3-h5ss_hIB1;* AmpR	*This study*
pAD07	pAD01	CEN/ARS; *URA3-hBPS;* AmpR	*This study*
pAD08	pAD01	CEN/ARS; *URA3-hIB1_hIB2_h3ss;* AmpR	*This study*
pAD09	pAD01	CEN/ARS; *URA3-h5ss_hIB1_hIB2_h3ss;* AmpR	*This study*
pAD10	pCR-blunt	pCR-blunt - GPDp.*SCEI*.CYC1t	*This study*
pAD11	pKR8	CEN/ARS; *URA3;* AmpR; *HIS3*-BspQI	*This study*
pAD12	pKR8	CEN/ARS; *URA3;* AmpR; *CUP1*-BspQI	*This study*
pAD13	pAD11	CEN/ARS; *URA3;* AmpR; *HIS3*-MATa1	*This study*
pAD14	pAD11	CEN/ARS; *URA3;* AmpR; *HIS3*-hBPS	*This study*
pAD15	pXP318	CEN/ARS; *URA3;* AmpR; LSR1p-*LSR1*-LSR1t	*This study*
pAD16	pXP318	CEN/ARS; *URA3;* AmpR; LSR1p-*mLSR1*-LSR1t	*This study*
pAD17	pXP318	CEN/ARS; *URA3;* AmpR; MSL5p-*MSL5*-MSL5t	*This study*
pAD18	pXP318	CEN/ARS; *URA3;* AmpR; MSL5p-*msl5::AfKH-QUA2*-MSL5t	*This study*
pAD19	pAD12	CEN/ARS; *URA3;* AmpR; CUP1p-*CUP1*-MATa1-CYC1t	*This study*
pAD20	pAD12	CEN/ARS; *URA3;* AmpR; CUP1p-*CUP1*-hBPS1-CYC1t	*This study*
pAD21	pAD12	CEN/ARS; *URA3;* AmpR; CUP1p-*CUP1*-hBPS2-CYC1t	*This study*
pAD22	pAD12	CEN/ARS; *URA3;* AmpR; CUP1p-*CUP1*-hBPS3-CYC1t	*This study*
pAD23	pAD12	CEN/ARS; *URA3;* AmpR; CUP1p-*CUP1*-hBPS4-CYC1t	*This study*
pAD24	pAD12	CEN/ARS; *URA3;* AmpR; *CUP1*-hBPS_library	*This study*
pAD25	pRS416gT-Mxi1	CEN/ARS; *URA3;* AmpR; TEFp-dCas9-Mxi1-CYC1t	*This study*

8.5 Strains developed in this study

Strain	Parent	Genotype	Reference
BY4741	S288c	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Brachmann *et al.*, Yeast, 1998
S1	BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*	*This study*
S2	S1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-ura3*	*This study*
S3	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1*	*This study*
S4	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-(MATa1)*	*This study*
S5	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-(pes1-5)*	*This study*
S6	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-(h5ss)*	*This study*
S7	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-(hIB1)*	*This study*
S8	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ*::*cup1-(hBPS)*	*This study*

S9	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hIB2_h3ss)*	*This study*
S10	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(h5ss_hIB1)*	*This study*
S11	S8	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hBPS) prp5-N399D*	*This study*
S12	S8	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hBPS) msl5::AfKH-QUA2*	*This study*
S13	S11	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hBPS) prp5-N399D msl5::AfKH-QUA2*	*This study*
S14	S1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ msl5::AfKH-QUA2*	*This study*
S15	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hIB1-1)*	*This study*
S16	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hIB1-2)*	*This study*
S17	S2	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(hIB1-3)*	*This study*
S18	S10	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(h5ss_hIB1) prp16-G373S*	*This study*
S19	S10	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(h5ss_hIB1) prp28-E404K*	*This study*
S20	S19	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(h5ss_hIB1) prp28-E404K prp16-G373S*	*This study*
S21	S5	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(pes1-5) msl5::AfKH-QUA2*	*This study*
S22	S21	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(pes1-5) msl5::AfKH-QUA2 prp16-G373S*	*This study*
S23	S22	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(pes1-5) msl5::AfKH-QUA2 prp16-G373S prp28-E404K*	*This study*
S24	S5	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(pes1-5) yhc1-D36A*	*This study*
S25	S21	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cup1Δ::cup1-(pes1-5) msl5::AfKH-QUA2 yhc1-D36A*	*This study*

8.6 Synthetic DNA used in this study

gBlocks		
Name	**Sequence**	**Category**
prp5-N399D	CATCCATTAGATCAGTATGCTGTACAGGAGGTTCTGAAATGAAAAAGCAGATT ACTGATCTTAAAAGAGGCACTGAGATTGTTGTTGCCACACCGGGACGATTTAT TGATATATTAACACTAGACGATGGGAAATTACTTAGTACTAAAAGAATAACGTT CGTAGTAATGGATGAGGCAGACAGGCTGTTCGATTTAGGTTTTGAACCTCAAA TAACGCAAATCATGAAAACTGTTCGACCGGATAAACA	Splicing Factor
Af KH-QUA2	GTTGAGATAGCCCTAAAGACCATCCCTTACTTCGTTCCTCCGGATGATTACAA GAGACCTACCAAGACCCAGGAGAAGGTCTATGTTCCAGTGAATGACTATCCA GAGATTAACTTCATTGGCTTACTCATAGGTCCTCGTGGAAATACCTTGAAGAA GATGGAGGCCGAATCTGGTGCCAAGATTGCCATTCGAGGCAAAGGCTCCGTC AAAGAAGGAAAAGGCCGATCTGACGCCGCTCACGCTAGTAACCAGGAAGAAG ACCTCCATTGTCTGATCATGGCAGATACCGAAGAGAAGGTTAACAAGGCCAA GAAGCTTGTGCACAATGTTATTGAAACAGCTGCCTCGATTCCCGAAGGCCAG AACGAACTCAAGAGAAACCAGTTGCGAGAGTTGGCTGCTCTCAACGGTACCC TCCGTGATGATGAGAACAGGCCCTGTCCAATCTGTGGTTTAAAAGATCATAAA AGGTACGATTGTCCAAACAGA	Splicing Factor
prp16-G373S	AAGAGCAACCAAAAGATTTGCGACGATACAGCTCTTTTCACGCCATCAAAAGA TGACATTAAACATACTAAAGAGCAACTGCCTGTTTTCCGCTGCAGATCTCAATT GTTATCATTGATAAGAGAAAATCAAGTAGTAGTGATAATTTCTGAAACGGGCTC AGGTAAAACCACGCAACTTGCACAGTATTTATATGAAGAAGGATATGCCAACG ATAGGGGGAAATCTATTGTTGTCACACAGCCGAGAAGAGTAGCAGC	Splicing Factor
prp28-E404K	GGACCAAGTAACAAACATTTTAACTAAAGTCGATATAAATGCTGACTCTGCTGT GAATAGACAAACCCTGATGTTTACCGCTACAATGACACCCGTTATAGAAAAAA TTGCAGCAGGATACATGCAAAAGCCTGTTTATGCAACAATTGGGGTTAAAACG GGTTCTGAACCTTTGATTCAACAGGTTGTGGAATATGCAGATAATGACGAAGA CAAATTCAAAAAGTTGAAGCCAATTGTCGCTAAATATGATCCACCAA	Splicing Factor

hIB1_hIB2	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTAT GTGCAAACCAACTCACTGACGTGGTCTTAAGGTACTAACCGATCACATAGCCA ATGTGGTAGCTGCAAAAATAATGAACAATGCCAAAAATCATGTAGCTGCCCAA CGGGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAAAC CAAGAAGTCATGCTGCTCTGGGAAATGA	Reporter-Intron
h5ss_hIB1	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTA CGTGCAAACCAACTCACTGACGTGGTCTTAAGGTACTAACAATTTGTAGCCAA TGTGGTAGCTGCAAAAATAATGAACAATGCCAAAAATCATGTAGCTGCCCAAC GGGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAAACC AAGAAGTCATGCTGCTCTGGGAAATGA	Reporter-Intron
pes1-5	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTA CGTGCAAACCAACTCACTGACGTGGTCTTAAGGAGCTGACCGATCACACAGC CAATGTGGTAGCTGCAAAAATAATGAACAATGCCAAAAATCATGTAGCTGCCC AACGGGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAA ACCAAGAAGTCATGCTGCTCTGGGAAATGA	Reporter-Intron
hIB2_h3ss	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTAT GTAATATGAGAATCAAACTTAAATATATCCTATACTAACCGATCACACAGCCAA TGTGGTAGCTGCAAAAATAATGAACAATGCCAAAAATCATGTAGCTGCCCAAC GGGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAAACC AAGAAGTCATGCTGCTCTGGGAAATG	Reporter-Intron
hIB1_h3ss	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTAT GTGCAAACCAACTCACTGACGTGGTCTTAAGGTACTAACAATTTGCAGCCAAT GTGGTAGCTGCAAAAATAATGAACAATGCCAAAAATCATGTAGCTGCCCAACG GGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAAACCA AGAAGTCATGCTGCTCTGGGAAATGA	Reporter-Intron

Primers		
Name	**Sequence**	**Category**
4bpLSR1 F	ACGAATCTCTTTGCCTTTTGGCTTAGATCAAGTGCAGTATCTGTTC	Splicing Factor
BBP seq1 F	TGAGCAGAGGTATAGAAAGAAACTAGAGG	Splicing Factor
BBP seq1 R	CTCTACTGAAATGTCCAGTTTGTCCAC	Splicing Factor
BPS lib F	TCGATCGCTCTTCGGAGG	Splicing Factor
BPS lib R	CTACGAGCTCTTCCGCAC	Splicing Factor
homCYC1t LSR1 R	AATGTAAGCGTGACATAACTAATTACATGAAAGAGCGAACGGGAAGAC	Splicing Factor
homLSR1p - LSR1 F	TGTTTCTACTTGTTTTTTTTTTAAATCCCCACGAATCTCTTTGCCTTTTGGC	Splicing Factor
homPRM9t - LSR1 R	AGTTGTGTGCTAGTGTCTCCCGTCTTCTGTAAGAGCGAACGGGAAGACG	Splicing Factor
homPXP - LSR1p F	GTTGAATACTCATACTCTTCCTTTTTCAATGCAGTTGCAGCAGGATAGC	Splicing Factor
hom-TEF1p LSR1 F	TAGCAATCTAATCTAAGTTTTAATTACAAAACGAATCTCTTTGCCTTTTGG	Splicing Factor
homU2p-linear 1	CTATGAAGCTTCGCTATCCTGCTGCAAC	Splicing Factor
homU2t-linear 2	ACATAACATCTACCTCCAGCATCTCATAATGCAACGCTTCGGAAAATACG	Splicing Factor
LSR1 ORF R	AAGAGCGAACGGGAAGAC	Splicing Factor
LSR1p F	GCAGTTGCAGCAGGATAGC	Splicing Factor
LSR1p R	CCAAAAGGCAAAGAGATTCGTGG	Splicing Factor
LSR1t F	CTTTTACTTTGGTCGCTTGATG	Splicing Factor
LSR1t R	ATTATGAGATGCTGGAGGTAGATG	Splicing Factor
MSL5d R	GAAACTTACGGTACTGAACTCGATCAAG	Splicing Factor
MSL5u F	CCAATTGAAATGCTCCGTAGTAATCC	Splicing Factor

prp16 seq1 F	GATCCAGCAGTGTGGAACGATC	Splicing Factor
prp16 seq1 R	TTTGTACACTCAGAATCCGTCACATC	Splicing Factor
PRP16 seq2 R	CAGCTTCTACGTAGTCTTGAACAGG	Splicing Factor
PRP16u seq2 F	CATTCCTGGAGCATCTCTGTGC	Splicing Factor
prp28 seq1 F	GTTGGTGGACACTCCTTGGAGG	Splicing Factor
prp28 seq1 R	CTTGAAAGTTGAGATCTTGGGTCCATC	Splicing Factor
U2 ER 1 R	CCAAAAAATGTGTATTGTAACAAATTAAAAGG	Splicing Factor
U2 ER 2 F	ATTTGTTACAATACACATTTTTTGGGACGCCTGTTTTTAAAGTTAG	Splicing Factor
U2p back F	GTTTCTACTTGTTTTTTTTTTAAATCCCC	Splicing Factor
U2p back REV	GGGGATTTAAAAAAAAAACAAGTAGAAAC	Splicing Factor
U2t front F	CTTCCCGTTCGCTCTTTTAT	Splicing Factor
U2t front R	ATAAAAGAGCGAACGGGAAG	Splicing Factor
YCH1 D36A F	TCGTATTTGACCCATGACACGTTGAGCGTTCGTAAATCGCACTTGGTCGGTAA GAATCACCTTCGTATAACAGCTGCTTATTATAGGAAC	Splicing Factor
YHC1 D36A R	GCCTCTTTTTCCAATGTGGCGTCTTTTATGATTATGTTTATTAATAATGTCTCTT GCTTTGTTCCTATAATAAGCAGCTGTTATACGAAG	Splicing Factor
YHC1 ORF seq1	GTAACTTTTAGCGTCTCATTTTGACTACTG	Splicing Factor
YHC1 ORF seq2	GGAATATCCTGGTGAGCCATCG	Splicing Factor
YHC1u seq1	GATCAACAACAGGCACAGCC	Splicing Factor
YHC1u seq2	GATAAATGCCGTTATGTTCCCTCAGAG	Splicing Factor
5-BPS-3 pes1 U F	TGACGTGGTCTTAAGGATACTAACCGATCACACAGGGCCCAGGTATTGTTAG C	Reporter-Intron
5-BPS-3 pes1 U R	CTGTGTGATCGGTTAGTATCCTTAAGACCACGTCAGTGAGTTGGTTTGCACAT AC CACCACACCGTGTGC	Reporter-Intron
5ssMATa1-URA F	CAAACTTAAATATATCCTATACTAACAATTTGTAGGGCCCAGGTATTGTTAGC	Reporter-Intron
5ssMATa1-URA R	CTACAAATTGTTAGTATAGGATATATTTAAGTTTGATTCTCATATTACGTACCAC CACACCGTGTGC	Reporter-Intron
BPS-3ss-pes1 U F	CACTGACGTGGTCTTAAGGATACTAACAATTTGTAGGGCCCAGGTATTGTTAG C	Reporter-Intron
BPS-3ss-pes1 U R	CTACAAATTGTTAGTATCCTTAAGACCACGTCAGTGAGTTGGTTTGCACGTAC CACCACACCGTGTGC	Reporter-Intron
BPS-lib fwd	TAGTACGTCAGTCGACTGCGAGC	Reporter-Intron
BPS-lib rev	TCTGCTCCTGTCAGAGCTACTGC	Reporter-Intron
BPSmata1 – URA-mid F	CTTAAATATATCCTAAGCTGACAATTTGTAGGGCCCAGGTATTGTTAGC	Reporter-Intron
BPSmata1 – URA-mid R	CTACAAATTGTCAGCTTAGGATATATTTAAGTTTGATTCTCATATTACATACCAC CACACCGTGTGC	Reporter-Intron
BPSpes1-URA mid F	GTGGTCTTAAGGATACTAACCGATCACACAGGGCCCAGGTATTGTTAGC	Reporter-Intron
BPSpes1-URA mid R	CTGTGTGATCGGTTAGTATCCTTAAGACCACGTCAGTGAGTTGGTTTGCACGT ACCACCACACCGTGTGC	Reporter-Intron
BPS-var F	GACTGCGAGCTCTTCGGAG	Reporter-Intron
BPS-var R	GAGCTACTGCTCTTCCGCAC	Reporter-Intron
BPS-variant	gactgcgaGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAN NNNNNNAATTTGTAGTGCGGAAGAGCagtagctc	Reporter-Intron
CU	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTAT GTAATATGAGAATCAAAC	Reporter-Intron

CU-MATa1_IB1 F	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTATGTGCAAACCAACTCACTG	Reporter-Intron
CUP1 amp1	CCTAAGCTCTTCTCTCATGACCTTC	Reporter-Intron
CUP1 amp2	CCTATAGCTCTTCTTGCCAATGCC	Reporter-Intron
CUP1 c1 FOR	GACTGATCTGTTGTACTATCCGCTTC	Reporter-Intron
CUP1 FOR	ATGTTCAGCGAATTAATTAACTTCC	Reporter-Intron
CUP1 hom 2B	CGCTGTTACACCCCGTTGGGCAGCTACATGATTTTTGGCATTGTTCATTATTTTTGCAGCTACCACATTGGATTACCC	Reporter-Intron
CUP1 homology 3	AGCTGCCCAACGGGGTGTAACAGCGACGACAAATGCCCCTGCGGTAACAAGTCTGAAGAAACCAAGAAGTCATGCTGCTCTGGGAAATGA	Reporter-Intron
CUP1 restore	CTTCCAAAATGAAGGTCATGAGTGCCAATGCCAATGTGGTAGCTGCAAAAATAATGAACA	Reporter-Intron
CUP1 REV	TCATTTCCCAGAGCAGCATG	Reporter-Intron
CUP1-BQ F	AGAAGAGCATTGGCGCGCCTAAGCTCTTCTTGCCAATGCCAATGTGG	Reporter-Intron
CUP1-BQ R	AGAAGAGCTTAGGCGCGCCAATGCTCTTCTCTCATGACCTTCATTTTGGAAG	Reporter-Intron
CUP1-BQ v1.1 F	CCAAAATGAAGGTCATGAGAGAAGAGCTTAGGCGCGCCTATAG	Reporter-Intron
CUP1-BQ v1.1 R	ACCACATTGGCATTGGCAAGAAGAGCTATAGGCGCGCCTAAGCTC	Reporter-Intron
CUP1-BQv2 F	AATAGAAGAGCTTAGGCCATAGTGGCCTAGTAGCTCTTCTGAAGGTCATGAGTGCCAATG	Reporter-Intron
CUP1-BQv2 R	AAGAGCTACTAGGCCACTATGGCCTAAGCTCTTCTATTTTGGAAGTTAATTAATTCGCTG	Reporter-Intron
CUP1d-CYC1t R	AATTATTACTTCACCACCCTTTATTTCAGGCTGATATCTTAGCCTTGTTAGGCCGCAAATTAAAGCCTTCG	Reporter-Intron
CUP1d-dwn R	CACCACCCTTTATTTCAGGCTGATATCTTAGCCTTGTTAACGAATTCGAGCTCGGTACCC	Reporter-Intron
CUP1d-extend R	TTAAAACACTTTTGTATTATTTTTCCTCATATATGTGTATAGGTTTATACGGATGATTTAATTATTACTTCACCACCCTTTATTTCAGGC	Reporter-Intron
CUP1d-KO-pXP R	AATTATTACTTCACCACCCTTTATTTCAGGCTGATATCTTAGCCTTGTTAACGAATTCGAGCTCGGTACC	Reporter-Intron
CUP1-ext F	AAACTGTACAATCAATCAATCAATCATCACATAAAATGTTCAGCGAATTAATTAACTTCC	Reporter-Intron
CUP1-ext R	CGTGAATGTAAGCGTGACATAACTAATTACATGACTCGAGTCATTTCCCAGAGCAGCATG	Reporter-Intron
CUP1-hpRS416 F	AAAAAAGGATCTCAAGAAGATCCTTTGATCGTACCGAGCTCGAATTCCTAGTTAG	Reporter-Intron
CUP1-hpRS416 R	ACTGATTAAGCATTGGTAACTGTCAGACCAGATATCTTAGCCTTGTTAGGCCGC	Reporter-Intron
CUP1-KO-pXP F	TTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTGCCAAGCTTGCATGCC	Reporter-Intron
CUP1-KO-pXP R	AATTATTACTTCACCACCCTTTATTTCAGGCTGATATCTTAGCCTTGTTACGGTACCCgggATAACTTCG	Reporter-Intron
CUP1-MATa1	CTTCCAAAATGAAGGTCATGAGTGCCAATGGTATGTAATATGAGAATCAAACTTAAATAT	Reporter-Intron
CUP1-MATa1 (full)	CTTCCAAAATGAAGGTCATGAGTGCCAATGGTATGTAATATGAGAATCAAACTTAAATATATCCTATACTAACAATTTGTAG	Reporter-Intron
CUP1-mid R	CATTGGCACTCATGACCTTC	Reporter-Intron
CUP1p F	CTAGTTAGAAAAAGACATTTTTGCTGTC	Reporter-Intron
CUP1-pes1 (full)	CTTCCAAAATGAAGGTCATGAGTGCCAATGGTACGTGCAAACCAACTCACTGACGTGGTCTTAAGGAGCTGACCGATCACACAG	Reporter-Intron
CUP1p-front F	CTAGTTAGAAAAAGACATTTTTGCTGTCAG	Reporter-Intron
CUP1p-pXP R	AATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAAATGTCTTTTTCTAACTAGGAATTCGAGCTCGGTACCC	Reporter-Intron
CUP1-prim1 F	ATCACATAAAATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATG	Reporter-Intron

CUP1-prim2 R	CCCGTTGGGCAGCTACATGATTTTTGGCATTGTTCATTATTTTTGCAGCTACCACATTGG	Reporter-Intron
CUP1-repair F	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGCC	Reporter-Intron
CUP1-repair R	ATTGTTCATTATTTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTC	Reporter-Intron
CUP1u seq1	GGTTTCTCGGTCTAAGAGCTTATACG	Reporter-Intron
CUP1u-extend F	GTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTACCTTTAAAAGACGTTCTCATAATACATTTTAGG	Reporter-Intron
CUP1u-M F	TTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCGCGTTTCTGGGTGAGC	Reporter-Intron
CUP1u-ups F	TTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTGCCAAGCTTGCATGC	Reporter-Intron
CU-pes_IB1 F	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTACGTAATATGAGAATCAAACTT	Reporter-Intron
CU-pes1 F	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGGTACGTGCAAACCAACTCACTGACG	Reporter-Intron
CUP-MATa1 F	GAAGGTCATGAGGTATGTAATATGAGAATC	Reporter-Intron
CUP-NGS F	ACTCTTTCCCTACACGACGCTCTTCCGATCTTCCAAAATGAAGGTCATGAG	Reporter-Intron
CUP-NGS R	GGAGTTCAGACGTGTGCTCTTCCGATCTCACATTGGCATTGGCA	Reporter-Intron
h5ss	AGGTCATGAGTGCCAATGGTACGTAATATGAGAATCAAACTTAAATATATCCTATACTAACAATTTGTAGCCAATGTGGTAGCTGCAA	Reporter-Intron
h5ss screen F	GAGTGCCAATGGTACGTAATATGAGAATC	Reporter-Intron
h5ss-hIB1 check R	CACATTGGCTACAAATTGTTAGTACC	Reporter-Intron
hBPS	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAGCTGACAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS v2	TCGATCGCTCTTCGAATGTATGTAATATGAGAATCAAACTTAAATATATCCTAAGCTGACAATTTGTAGGAAGGAAGAGCTCGTAG	Reporter-Intron
hBPS10	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAGACTCAAAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS11	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAGTTAGCAAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS12	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAGGGGGGGAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS2	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAACTAACAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS3	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAACTAATAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS4	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAACTGACAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS5	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAACTGATAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS6	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAGCTAACAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS7	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAGCTAATAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS8	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAAGCTGATAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hBPS9	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTATGTGGTGAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
hIB2_h3ss screen R	ACATTGGCTGTGTGATCGGTTAG	Reporter-Intron
HIS3 amp1	CCTAAGCTCTTCTCTCAACGATTAGC	Reporter-Intron
HIS3 amp2	CTATAGCTCTTCTTGCATTGGTGACT	Reporter-Intron
HIS3-BQ F	AGAAGAGCATTGGCGCGCCTAAGCTCTTCTTGCATTGGTGACTTACACATAGAC	Reporter-Intron

HIS3-BQ R	AGAAGAGCTTAGGCGCGCCAATGCTCTTCTCTCAACGATTAGCGACCAGC	Reporter-Intron
HIS3-repair F	CATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTG ACT	Reporter-Intron
HIS3-repair R	TTCAGTGGTGTGATGGTCGTCTATGTGTAAGTCACCAATGCACTCAACGATTA G	Reporter-Intron
HIS-BQ v1.1 F	CTGGTCGCTAATCGTTGAGAGAAGAGCTTAGGCGCGCCTATAG	Reporter-Intron
HIS-BQ v1.1 R	GTGTAAGTCACCAATGCAAGAAGAGCTATAGGCGCGCCTAAGCTC	Reporter-Intron
HIS-MATa1 F	TCGTTGAGGTATGTAATATGAGAATCAAAC	Reporter-Intron
HIS-NGS F	ACTCTTTCCCTACACGACGCTCTTCCGATCTCGGCTGGTCGCTAATCGTTGAG	Reporter-Intron
HIS-NGS R	GGAGTTCAGACGTGTGCTCTTCCGATCTGTCGTCTATGTGTAAGTCACCAATG CA	Reporter-Intron
hom1-CUP1-pXP F	ATGTTCAGCGAATTAATTAACTTCCAAAATGAAGGTCATGAGTGCCAATGCCA AGCTTGCATGCCTGC	Reporter-Intron
hom2-CUP1-pXP R	CGCTGTTACACCCCGTTGGGCAGCTACATGATTTTTGGCATTGTTCATTATTTT TGCAGCTACCACATTGGcGAATTCGAGCTCGGTACC	Reporter-Intron
homCUP1-2 (rcomp)	TGTTCATTATTTTTGCAGCTAC	Reporter-Intron
homPXP - PRM9t R	GACAATAACCCTGATAAATGCTTCAATAATCATTTTCAACATCGTATTTTCCGA AG	Reporter-Intron
homSceI-PRM9t F	CTCCTCCGAAACTTTCCTGAAATAACTCGAGACAGAAGACGGGAGACACTAG C	Reporter-Intron
hom-SpeI-GPD R	CCTGGTTTTTTTTGATGTTTTTCATACTAGTCTTTGTTTGTTTATGTGTG	Reporter-Intron
hUra3-CUP1p F	CTTTTTTGCGAGGCATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGCTAG TTAGAAAAAGACATTTTTGCTGTCAGTC	Reporter-Intron
hUra3-CYC1t R	ATTTTTTTTTTTTCGTCATTATAGAAATCATTACGACCGAGATTCCCGGGGCAA ATTAAAGCCTTCGAGC	Reporter-Intron
Hybrid-BQ	AGCTAGGCTCTTCCGAGGTACGTGCAAACCAACTCACTGACGTGGTCTTAAG GAGCTGACAATTTGTAGTGCAGAAGAGCAGCATC	Reporter-Intron
iBQ v2 F	TCGATCGCTCTTCGAATG	Reporter-Intron
iBQ v2 R	CTACGAGCTCTTCCTTCC	Reporter-Intron
MATa1	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTTAGTATAGGATAT ATTTAAGTTTGATTCTCATATTACATACCATTGGCA	Reporter-Intron
MATa1 – URA-mid F	CTTAAATATATCCTATACTAACAATTTGTAGGGCCCAGGTATTGTTAGC	Reporter-Intron
MATa1 - URA-mid R	CTACAAATTGTTAGTATAGGATATATTTAAGTTTGATTCTCATATTACATACCAC CACACCGTGTGC	Reporter-Intron
MATa1 (partial) - CUP1	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTTAGTATAGGATAT ATTTAA	Reporter-Intron
MATa1 intron (rcomp)	CTACAAATTGTTAGTATAGGATATATTTAAGTTTGATTCTCATATTACATAC	Reporter-Intron
MATa1_IB2 R	TGTTCATTATTTTTGCAGCTACCACATTGGCTATGTGATCGGTTAGTATAGGAT ATATTTAAGTTTGATTCTCATATTACATACC	Reporter-Intron
MATa1-BPS	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTCAGCTTAGGATAT ATTTAAGTTTGATTCTCATATTACATACCATTGGCA	Reporter-Intron
MATa1-BPS2	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTCAGTATAGGATAT ATTTAAGTTTGATTCTCATATTACATACCATTGGCA	Reporter-Intron
MATa1-BPS3	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTTAGCATAGGATAT ATTTAAGTTTGATTCTCATATTACATACCATTGGCA	Reporter-Intron
MATa1-BPS4	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTTAGTTTAGGATAT ATTTAAGTTTGATTCTCATATTACATACCATTGGCA	Reporter-Intron
MATa1-BQ	TCGATCGCTCTTCGGAGGTATGTAATATGAGAATCAAACTTAAATATATCCTAT ACTAACAATTTGTAGTGCGGAAGAGCTCGTAG	Reporter-Intron
MATa1-CUP1	TTAAATATATCCTATACTAACAATTTGTAGCCAATGTGGTAGCTGCAAAAATAA TGAACA	Reporter-Intron

MATa1-IB1 R	TGTTCATTATTTTTGCAGCTACCACATTGGCTACAAATTGTTAGTACCTTAAGA CCACGTCAGTGAGTTGGTTTGCACATACCATTG	Reporter-Intron
MATa1-IVS2 URA-mid F	TAACCGATCACATAGGGCCCAGGTATTGTTAGCGG	Reporter-Intron
MATa1-IVS2 URA-mid R	CCGCTAACAATACCTGGGCCCTATGTGATCGGTTAGTATAGGATATATTTAAG TTTGATTCTCATATTACATACCACCACACCGTGTGC	Reporter-Intron
nest CUP1 F	CGAATTAATTAACTTCCAAAATGAAGG	Reporter-Intron
nest CUP1 R	GAGCAGCATGACTTCTTGG	Reporter-Intron
nested URA F	CAAACTTGTGTGCTTCATTGGATG	Reporter-Intron
nested URA R	CCTCTTCCAACAATAATAATGTCAGATCC	Reporter-Intron
oKR444	AATGATACGGCGACCACCGAGATCTACACTGAACCTTACACTCTTTCCCTACA CGACGCT	Reporter-Intron
oKR445	AATGATACGGCGACCACCGAGATCTACACTGCTAAGTACACTCTTTCCCTACA CGACGCT	Reporter-Intron
oKR446	CAAGCAGAAGACGGCATACGAGATATCACGACGTGACTGGAGTTCAGACGTG TGCTCTTC	Reporter-Intron
oKR447	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTGACTGGAGTTCAGACGTG TGCTCTTC	Reporter-Intron
oKR458	AATGATACGGCGACCACCGAGATCTACACACTGTGCAACACTCTTTCCCTACA CGACGCT	Reporter-Intron
oKR459	CAAGCAGAAGACGGCATACGAGATTAGTGAGTGTGACTGGAGTTCAGACGTG TGCTCTTC	Reporter-Intron
pes1-3b	AGCTAGGCTCTTCCGAGGTGCGTCTCCACCAATTCCAATCCGGTCACCTTGG TGCTAACCCCAGAACAGTGCAGAAGAGCAGCATC	Reporter-Intron
pes1-5a	TCGATCGCTCTTCGGAGGTACGTGCAAACCAACTCACTGACGTGGTCTTAAG GAGCTGACCGATCACACAGTGCGGAAGAGCTCGTAG	Reporter-Intron
pes1-5a v2	TCGATCGCTCTTCGAATGTACGTGCAAACCAACTCACTGACGTGGTCTTAAGG AGCTGACCGATCACACAGGAAGGAAGAGCTCGTAG	Reporter-Intron
pes1-CUP1 (partial)	TGTTCATTATTTTTGCAGCTACCACATTGGCTGTGTGATCGGTCAGCTCCTTAA GACCAC	Reporter-Intron
PES1i check R	GCTGTGTGATCGGTCAGC	Reporter-Intron
pes1-IB1 P1 R	TGTTCATTATTTTTGCAGCTACCACATTGGCTGTGTGATCGGTCAGCTTAGGA TATATTTAAGTTTGATTCTCATATTACGTACCATTGG	Reporter-Intron
pes1-P1 R	TGTTCATTATTTTTGCAGCTACCACATTGGCTGTGTGATCGGTCAGCTCCTTAA GACCACGTCAGTGAGTTGGTTTGCAC	Reporter-Intron
pes1-URA mid F	CGTGGTCTTAAGGAGCTGACCGATCACACAGGGCCCAGGTATTGTTAGC	Reporter-Intron
pes1-URA mid R	CTGTGTGATCGGTCAGCTCCTTAAGACCACGTCAGTGAGTTGGTTTGCACGTA CCACCACACCGTGTGC	Reporter-Intron
pes-IB1-BPS P1 R	TGTTCATTATTTTTGCAGCTACCACATTGGCTGTGTGATCGGTTAGTATAGGAT ATATTTAAGTTTGATTCTCATATTACGTACCATTGG	Reporter-Intron
pKR8-bb 1	CGTATCACGAGGCCCTTTCG	Reporter-Intron
pKR8-bb 2	GGCACTTTTCGGGGAAATGTG	Reporter-Intron
RSC30 check 2 R	GGTATTACTACGGCAAACTTCAACG	Reporter-Intron
RSC30 check REV	CTCAAGACATTCGCTTCTAGGTCAG	Reporter-Intron
RSC30 seq1	CTCAAGACATTCGCTTCTAGGTCAG	Reporter-Intron
RSC30 seq2	GATTCAATGACCCTATTCAATAAGCAGG	Reporter-Intron
S1 LEU2 F	CTGTGGAGGAAACCATCAAGAAC	Reporter-Intron
S1 RSC30 R	CGAGATGAAATGAATAGCAACGGAAG	Reporter-Intron
SceI.URA3 F	TAGGGATAACAGGGTAATCCAAGC	Reporter-Intron
SceI.URA3 R	ATTACCCTGTTATCCCTACGAATTCG	Reporter-Intron
ScHIS3-hpRS416 F	AAAAAAGGATCTCAAGAAGATCCTTTGATCCGTACGCTGCAGGTCGAC	Reporter-Intron

ScHIS3-hpRS416 R	ACTGATTAAGCATTGGTAACTGTCAGACCACCATCGATGAATTCGAGCTCG	Reporter-Intron
qPCR ACT1 F	GCCTTGGACTTCGAACAAGAAATG	qPCR
qPCR ACT1 R	GGGCTCTGAATCTTTCGTTACCA	qPCR
qPCR BBP F	CGCCAATACAGAGCAATGACG	qPCR
qPCR BBP R	GAGGTGAAGGTGATGGTGC	qPCR
qPCR BBPv2 R	GGTGAAGGTGATGGTGCATAAC	qPCR
qPCR CUP1e R	CAGCATGACTTCTTGGTTTCTTCAGAC	qPCR
qPCR CUP1e1 F	CTTCCAAAATGAAGGTCATGAGTGCC	qPCR
qPCR CUP1e2 F	TCATGTAGCTGCCCAACGG	qPCR
qPCR CUP1ej F	GAGTGCCAATGCCAATGTGG	qPCR
qPCR CUP1ejv2 F	ATGAGTGCCAATGCCAATGTG	qPCR
qPCR CUP1-hij F	CCTAAGCTGACAATTTGTAGCCAATGTG	qPCR
qPCR CUP1-Mij F	GTCATGAGTGCCAATGGTATGTAATATGAG	qPCR
qPCR H-e1e2 F	GCTAATCGTTGAGTGCATTGGTGAC	qPCR
qPCR H-e2 F	CATCACACCACTGAAGACTGC	qPCR
qPCR H-e2 R	CTCTGGAAAGTGCCTCATCC	qPCR
qPCR H-hij F	CCTAAGCTGACAATTTGTAGTGCATTGG	qPCR
qPCR H-Mij F	CCTATACTAACAATTTGTAGTGCATTGGTGAC	qPCR
qPCR H-p F	GAGCTGACCGATCACACAG	qPCR
qPCR IFH1 F	ACTGACGCCGATATCCTAGC	qPCR
qPCR IFH1 R	GCTGCTTCCACCATACTTTGC	qPCR
qPCR IFH1v2 F	ACCGCTCCTGTGCAATTCG	qPCR
qPCR pes1i F	CAACTCACTGACGTGGTCTTAAGG	qPCR
qPCR pes1iv2 F	CGTGCAAACCAACTCACTGAC	qPCR
qPCR sCR1 F	TTTCTGGTGGGATGGGATAC	qPCR
qPCR sCR1 R	TTTACGACGGAGGAAAGACG	qPCR
qPCR SEC65 F	GATTCCATATGGCCCTGATTTCGAC	qPCR
qPCR SEC65 R	GTGGCTTGAACGACTTTTCTGC	qPCR
qPCR SEC65v2 F	GATTGTTCCGCTGCATAGTCC	qPCR
qPCR YHC1 F	AGAGATGAGACGTGCGAGAG	qPCR
qPCR YHC1 R	CTCCTTGATCGTGGTTTGCAG	qPCR
qPCR YHC1v2 F	TACTGCAAACCACGATCAAGGAG	qPCR
qPCR YHC1v2 R	GCACTCGACGGTTTCCGTATG	qPCR
qPCR.2 Ce2 F	TGAACAATGCCAAAAATCATGT	qPCR

qPCR.2 Ce2 R	ATTTCCCAGAGCAGCATGAC	qPCR
qPCR.2 He1 F	GCTGGTCGCTAATCGTTGAG	qPCR
qPCR.2 He2 F	ACATAGACGACCATCACACCAC	qPCR
qPCR.2 He2 R	GCGCAAATCCTGATCCAAACC	qPCR
qPCR.2 Hej F	CTAATCGTTGAGTGCATTGGTGACTTAC	qPCR
qPCR.2 HM F	GTCGCTAATCGTTGAGGTATGTAATATGAG	qPCR
dCas9 amp F	GCCTACCTGAATGCAGTGGTAG	CRISPRi
dCas9 amp R	CCTATTTCCTGCTCAGACTTTGCG	CRISPRi
dCas9 seq2	CCTCGAAGTTCCAGGGAGTG	CRISPRi
dCas9 seq3	AACTGCACGCTATCCTCAGG	CRISPRi
gIFH1 seq1	ACAATTCTAACACTGCCAATCGC	CRISPRi
pRPR1 F	TCGCGGCTGGGAACG	CRISPRi
gYHC1-pCB 2	CTAGCTCTAAAACTCGTATAACAGCTGACTATTGATCATTTATCTTTCACTGCGGAGAAG	CRISPR-Cas9
gYHC1-pCB 4	ATGATCAATAGTCAGCTGTTATACGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	CRISPR-Cas9
iCas9 assem F	ctCACCTTTCGAGAGGACGATGCCCGTGTCTACTGCATAGCTTCAAAATGTTTCTACTCC	CRISPR-Cas9
iCas9 assem R	GCCCGTGCCATAGCCATGCCTTCACATATAGTTCTAGAACTAGTATGAGAAATATCGAGG	CRISPR-Cas9
iCas9 seq 1a	gattgtaggtttgtaccaactgg	CRISPR-Cas9
iCas9 seq 2a	ggatggtactgaggaattattggtg	CRISPR-Cas9
iCas9 seq 2b	ggcgtttgcgaatctctcc	CRISPR-Cas9
iCas9 seq 2c	gcttcattaggtacctaccatg	CRISPR-Cas9
iCas9 seq 2d	cttggtccacatacatgtctc	CRISPR-Cas9
iCas9 seq 3a	ggctctgccaagcaaatatg	CRISPR-Cas9
iCas9 seq1 R	cgttgtcaaaggtccgttgc	CRISPR-Cas9
iCas9 seq3 F	gaaataggcaaagcaaccgc	CRISPR-Cas9
iCas9 seqX1 R	atgcaaaacgactattgccacg	CRISPR-Cas9
iCas9 seqX3 F	cgtatctaaatgccgtcgttgg	CRISPR-Cas9
iCas9-back F	gcattgatttgagtcagctaggagg	CRISPR-Cas9
iCas9ec F	CTGCATAGCTTCAAAATGTTTCTACTCC	CRISPR-Cas9
iCas9ec R	tctagaactagtatgagaaatatcgagggactc	CRISPR-Cas9
iCas9-pCB301	tgaatcgagtccctcgatatttctcatactagttctagagcgcgccaacaaatatattgc	CRISPR-Cas9
iCas9-pCB303	AAGAGTAAAAAAGGAGTAGAAACATTTTGAAGCTATGCAGacgaaagggcctcgtg	CRISPR-Cas9

9. REFERENCES

1. Siddiqui, M.S., Thodey, K., Trenchard, I. and Smolke, C.D. (2012) Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *Fems Yeast Res*, **12**, 144-170.
2. Galanie, S., Thodey, K., Trenchard, I.J., Interrante, M.F. and Smolke, C.D. (2015) SYNTHETIC BIOLOGY Complete biosynthesis of opioids in yeast. *Science*, **349**, 1095-1100.
3. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**, 528-+.
4. Martens, E. and Demain, A.L. (2017) The antibiotic resistance crisis, with a focus on the United States. *J Antibiot*, **70**, 520-526.
5. Fatma, S., Hameed, A., Noman, M., Ahmed, T., Sohail, I., Shahid, M., Tariq, M. and Tabassum, R. (2018) Lignocellulosic Biomass: A Sustainable Bioenergy Source for Future. *Protein Pept Lett*.
6. Cragg, G.M. and Newman, D.J. (2013) Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta*, **1830**, 3670-3695.
7. Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H. and Fedorova, N.D. (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol*, **47**, 736-741.
8. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C. *et al.* (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol*, **11**, 625-631.
9. Tobert, J.A. (2003) Lovastatin and beyond: The history of the HMG-CoA reductase inhibitors. *Nat Rev Drug Discov*, **2**, 517-526.
10. Oxford, A.E., Raistrick, H. and Simonart, P. (1939) Studies in the biochemistry of micro-organisms. LX. Griseofulvin, C17H17O6Cl, a metabolic product of Penicillium griseo-fulvum Dierckx. *Biochem J*, **33**, 240-248.
11. Faulds, D., Goa, K.L. and Benfield, P. (1993) Cyclosporin. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic use in immunoregulatory disorders. *Drugs*, **45**, 953-1040.
12. Aharonowitz, Y., Cohen, G. and Martin, J.F. (1992) Penicillin and Cephalosporin Biosynthetic Genes - Structure, Organization, Regulation, and Evolution. *Annu Rev Microbiol*, **46**, 461-495.
13. Blackwell, M. (2011) The fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot*, **98**, 426-438.
14. Takeda, I., Umemura, M., Koike, H., Asai, K. and Machida, M. (2014) Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: application to sequenced genomes of Aspergillus and ten other filamentous fungal species. *DNA research : an international journal for rapid publication of reports on genes and genomes*, **21**, 447-457.
15. Medema, M.H. and Fischbach, M.A. (2015) Computational approaches to natural product discovery. *Nat Chem Biol*, **11**, 639-648.
16. Luo, Y.Z., Cobb, R.E. and Zhao, H.M. (2014) Recent advances in natural product discovery. *Curr Opin Biotech*, **30**, 230-237.
17. Zotchev, S.B., Sekurova, O.N. and Katz, L. (2012) Genome-based bioprospecting of microbes for new therapeutics. *Curr Opin Biotechnol*, **23**, 941-947.
18. Montiel, D., Kang, H.S., Chang, F.Y., Charlop-Powers, Z. and Brady, S.F. (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *P Natl Acad Sci USA*, **112**, 8953-8958.

19. Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C. and Moore, B.S. (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A*, **111**, 1957-1962.
20. Yeh, H.H., Ahuja, M., Chiang, Y.M., Oakley, C.E., Moore, S., Yoon, O., Hajovsky, H., Bok, J.W., Keller, N.P., Wang, C.C. *et al.* (2016) Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in Aspergillus nidulans Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor. *ACS Chem Biol*, **11**, 2275-2284.
21. Streit, W.R. and Schmitz, R.A. (2004) Metagenomics - the key to the uncultured microbes. *Curr Opin Microbiol*, **7**, 492-498.
22. Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods*, **11**, 499-507.
23. Roumpeka, D.D., Wallace, R.J., Escalettes, F., Fotheringham, I. and Watson, M. (2017) A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Front Genet*, **8**.
24. Inglis, D.O., Binkley, J., Skrzypek, M.S., Arnaud, M.B., Cerqueira, G.C., Shah, P., Wymore, F., Wortman, J.R. and Sherlock, G. (2013) Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae. *BMC Microbiol*, **13**, 91.
25. Harvey, C.J.B., Tang, M., Schlecht, U., Horecka, J., Fischer, C.R., Lin, H.C., Li, J., Naughton, B., Cherry, J., Miranda, M. *et al.* (2018) HEx: A heterologous expression platform for the discovery of fungal natural products. *Sci Adv*, **4**, eaar5459.
26. McKelvey, S.M. and Murphy, R.A. (2011) Biotechnological Use of Fungal Enzymes. *Fungi: Biology and Applications, 2nd Edition*, 179-204.
27. Solomon, K.V., Haitjema, C.H., Henske, J.K., Gilmore, S.P., Borges-Rivera, D., Lipzen, A., Brewer, H.M., Purvine, S.O., Wright, A.T., Theodorou, M.K. *et al.* (2016) Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science*, **351**, 1192-1195.
28. Haitjema, C.H., Gilmore, S.P., Henske, J.K., Solomon, K.V., de Groot, R., Kuo, A., Mondo, S.J., Salamov, A.A., LaButti, K., Zhao, Z. *et al.* (2017) A parts list for fungal cellulosomes revealed by comparative genomics. *Nat Microbiol*, **2**, 17087.
29. Ferrer, M., Martinez-Martinez, M., Bargiela, R., Streit, W.R., Golyshina, O.V. and Golyshin, P.N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol*, **9**, 22-34.
30. Ongley, S.E., Bian, X.Y., Neilan, B.A. and Muller, R. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep*, **30**, 1121-1138.
31. Harvey, A.L., Edrada-Ebel, R. and Quinn, R.J. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov*, **14**, 111-129.
32. Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Brown, L.C.W., Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J. *et al.* (2014) Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell*, **158**, 412-421.
33. Medema, M.H., Breitling, R., Bovenberg, R. and Takano, E. (2011) Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev Microbiol*, **9**, 131-137.
34. Temme, K., Zhao, D. and Voigt, C.A. (2012) Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc Natl Acad Sci U S A*, **109**, 7085-7090.
35. Billingsley, J.M., DeNicola, A.B. and Tang, Y. (2016) Technology development for natural product biosynthesis in Saccharomyces cerevisiae. *Curr Opin Biotechnol*, **42**, 74-83.

36. Al-Hawash, A.B., Zhang, X.Y. and Ma, F.Y. (2017) Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems. *Gene Rep*, **9**, 46-53.
37. Nielsen, J. (2015) BIOENGINEERING. Yeast cell factories on the horizon. *Science*, **349**, 1050-1051.
38. Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H.S., Zhu, H., Dyer, D.W., Roe, B.A. and Murphy, J.W. (2004) Introns and splicing elements of five diverse fungi. *Eukaryot Cell*, **3**, 1088-1100.
39. Nicholson, M.J., Eaton, C.J., Starkel, C., Tapper, B.A., Cox, M.P. and Scott, B. (2015) Molecular Cloning and Functional Analysis of Gene Clusters for the Biosynthesis of Indole-Diterpenes in Penicillium crustosum and P. janthinellum. *Toxins (Basel)*, **7**, 2701-2722.
40. Moabbi, A.M., Agarwal, N., El Kaderi, B. and Ansari, A. (2012) Role for gene looping in intron-mediated enhancement of transcription. *P Natl Acad Sci USA*, **109**, 8505-8510.
41. Davis, C.A., Grate, L., Spingola, M. and Ares, M. (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res*, **28**, 1700-1706.
42. Xu, W., Cai, X.L., Jung, M.E. and Tang, Y. (2010) Analysis of Intact and Dissected Fungal Polyketide Synthase-Nonribosomal Peptide Synthetase in Vitro and in Saccharomyces cerevisiae. *J Am Chem Soc*, **132**, 13604-13607.
43. Zafrir, Z. and Tuller, T. (2015) Nucleotide sequence composition adjacent to intronic splice sites improves splicing efficiency via its effect on pre-mRNA local folding in fungi. *RNA (New York, N.Y.)*, **21**, 1704-1718.
44. Cacho, R.A., Tang, Y. and Chooi, Y.H. (2015) Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Frontiers in Microbiology*, **5**.
45. Kakule, T.B., Jadulco, R.C., Koch, M., Janso, J.E., Barrows, L.R. and Schmidt, E.W. (2015) Native Promoter Strategy for High-Yielding Synthesis and Engineering of Fungal Secondary Metabolites. *Acs Synth Biol*, **4**, 625-633.
46. Yofe, I., Zafrir, Z., Blau, R., Schuldiner, M., Tuller, T., Shapiro, E. and Ben-Yehezkel, T. (2014) Accurate, Model-Based Tuning of Synthetic Gene Expression Using Introns in S. cerevisiae. *Plos Genet*, **10**.
47. Li, Q., Sun, Z.Q., Li, J. and Zhang, Y.S. (2013) Enhancing beta-carotene production in Saccharomyces cerevisiae by metabolic engineering. *Fems Microbiol Lett*, **345**, 94-101.
48. Lanza, A.M., Curran, K.A., Rey, L.G. and Alper, H.S. (2014) A condition-specific codon optimization approach for improved heterologous gene expression in Saccharomyces cerevisiae. *Bmc Syst Biol*, **8**.
49. Sun, J., Shao, Z.Y., Zhao, H., Nair, N., Wen, F., Xu, J.H. and Zhao, H.M. (2012) Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. *Biotechnol Bioeng*, **109**, 2082-2092.
50. Shen, M.W., Fang, F., Sandmeyer, S. and Da Silva, N.A. (2012) Development and characterization of a vector set with regulated promoters for systematic metabolic engineering in Saccharomyces cerevisiae. *Yeast*, **29**, 495-503.
51. Tang, M.C., Lin, H.C., Li, D.H., Zou, Y., Li, J., Xu, W., Cacho, R.A., Hillenmeyer, M.E., Garg, N.K. and Tang, Y. (2015) Discovery of Unclustered Fungal Indole Diterpene Biosynthetic Pathways through Combinatorial Pathway Reassembly in Engineered Yeast. *J Am Chem Soc*, **137**, 13724-13727.
52. Redden, H. and Alper, H.S. (2015) The development and characterization of synthetic minimal yeast promoters. *Nat Commun*, **6**.
53. Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., Moriya, H. and Matsuyama, T. (2013) A Genome-Wide Activity Assessment of Terminator Regions in

Saccharomyces cerevisiae Provides a "Terminatome" Toolbox. *Acs Synth Biol*, **2**, 337-347.
54. Curran, K.A., Karim, A.S., Gupta, A. and Alper, H.S. (2013) Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng*, **19**, 88-97.
55. Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A. and Alper, H.S. (2015) Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *Acs Synth Biol*, **4**, 824-832.
56. Siewers, V. (2014) An Overview on Selection Marker Genes for Transformation of Saccharomyces cerevisiae. *Yeast Metabolic Engineering: Methods and Protocols*, **1152**, 3-15.
57. Karim, A.S., Curran, K.A. and Alper, H.S. (2013) Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *Fems Yeast Res*, **13**, 107-116.
58. Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *Acs Synth Biol*, **4**, 975-986.
59. Trenchard, I.J. and Smolke, C.D. (2015) Engineering strategies for the fermentative production of plant alkaloids in yeast. *Metab Eng*, **30**, 96-104.
60. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343-U341.
61. Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *Plos One*, **3**, e3647.
62. Joska, T.M., Mashruwala, A., Boyd, J.M. and Belden, W.J. (2014) A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. *J Microbiol Meth*, **100**, 46-51.
63. Shao, Z.Y., Zhao, H. and Zhao, H.M. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*, **37**.
64. Flagfeldt, D.B., Siewers, V., Huang, L. and Nielsen, J. (2009) Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast*, **26**, 545-551.
65. Fang, F., Salmon, K., Shen, M.W., Aeling, K.A., Ito, E., Irwin, B., Tran, U.P., Hatfield, G.W., Da Silva, N.A. and Sandmeyer, S. (2011) A vector set for systematic metabolic engineering in Saccharomyces cerevisiae. *Yeast*, **28**, 123-136.
66. Brown, S., Clastre, M., Courdavault, V. and O'Connor, S.E. (2015) De novo production of the plant-derived alkaloid strictosidine in yeast. *P Natl Acad Sci USA*, **112**, 3205-3210.
67. Kuijpers, N.G.A., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2013) One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in Saccharomyces cerevisiae. *Fems Yeast Res*, **13**, 769-781.
68. Wingler, L.M. and Cornish, V.W. (2011) Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. *Proc Natl Acad Sci U S A*, **108**, 15135-15140.
69. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res*, **41**, 4336-4343.
70. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *Fems Yeast Res*, **15**.

71. Jakociunas, T., Rajkumar, A.S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C.B., Skjodt, M.L., Nielsen, A.T., Borodina, I., Jensen, M.K. *et al.* (2015) CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. *Acs Synth Biol*, **4**, 1226-1234.
72. Shi, S., Liang, Y., Zhang, M.M., Ang, E.L. and Zhao, H. (2015) A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. *Metab Eng*, **33**, 19-27.
73. Yuan, Y., Andersen, E. and Zhao, H. (2016) Flexible and Versatile Strategy for the Construction of Large Biochemical Pathways. *Acs Synth Biol*, **5**, 46-52.
74. Benatuil, L., Perez, J.M., Belk, J. and Hsieh, C.M. (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel*, **23**, 155-159.
75. Meehl, M.A. and Stadheim, T.A. (2014) Biopharmaceutical discovery and production in yeast. *Curr Opin Biotech*, **30**, 120-127.
76. Giaever, G. and Nislow, C. (2014) The yeast deletion collection: a decade of functional genomics. *Genetics*, **197**, 451-465.
77. Ko, J.K. and Lee, S.M. (2018) Advances in cellulosic conversion to fuels: engineering yeasts for cellulosic bioethanol and biodiesel production. *Curr Opin Biotechnol*, **50**, 72-80.
78. Yaguchi, A., Spagnuolo, M. and Blenner, M. (2018) Engineering yeast for utilization of alternative feedstocks. *Curr Opin Biotechnol*, **53**, 122-129.
79. Cacho, R.A. and Tang, Y. (2016) Reconstitution of Fungal Nonribosomal Peptide Synthetases in Yeast and In Vitro. *Methods Mol Biol*, **1401**, 103-119.
80. Bond, C., Tang, Y. and Li, L. (2016) Saccharomyces cerevisiae as a tool for mining, studying and engineering fungal polyketide synthases. *Fungal Genetics and Biology*, **89**, 52-61.
81. Ashikari, T., Amachi, T., Yoshizumi, H., Horiuchi, H., Takagi, M. and Yano, K. (1990) Correct splicing of modified introns of a Rhizopus proteinase gene in Saccharomyces cerevisiae. *Mol Gen Genet*, **223**, 11-16.
82. Langford, C., Nellen, W., Niessing, J. and Gallwitz, D. (1983) Yeast Is Unable to Excise Foreign Intervening Sequences from Hybrid Gene Transcripts. *P Natl Acad Sci-Biol*, **80**, 1496-1500.
83. Moreau, A., Durand, S. and Morosoli, R. (1992) Secretion of a Cryptococcus-Albidus Xylanase in Saccharomyces-Cerevisiae. *Gene*, **116**, 109-113.
84. Woudt, L.P., van den Heuvel, J.J., van Raamsdonk-Duin, M.M., Mager, W.H. and Planta, R.J. (1985) Correct removal by splicing of a Neurospora intron in yeast. *Nucleic Acids Res*, **13**, 7729-7739.
85. Chavez, R., Navarro, C., Calderon, I., Peirano, A., Bull, P. and Eyzaguirre, J. (2002) Secretion of endoxylanase A from Penicillium purpurogenum by Saccharomyces cerevisiae transformed with genomic fungal DNA. *Fems Microbiol Lett*, **212**, 237-241.
86. Shi, Y.G. (2017) Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Bio*, **18**, 655-670.
87. Herzel, L., Ottoz, D.S.M., Alpert, T. and Neugebauer, K.M. (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Bio*, **18**, 637-650.
88. Papasaikas, P. and Valcarcel, J. (2016) The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci*, **41**, 33-45.
89. Spingola, M., Grate, L., Haussler, D. and Ares, M., Jr. (1999) Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, **5**, 221-234.

90. Hossain, M.A., Rodriguez, C.M. and Johnson, T.L. (2011) Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing. *Nucleic Acids Res*, **39**, 8612-8627.
91. Juneau, K., Nislow, C. and Davis, R.W. (2009) Alternative Splicing of PTC7 in Saccharomyces cerevisiae Determines Protein Localization. *Genetics*, **183**, 185-194.
92. Grund, S.E., Fischer, T., Cabal, G.G., Antunez, O., Perez-Ortin, J.E. and Hurt, E. (2008) The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. *J Cell Biol*, **182**, 897-910.
93. Mishra, S.K., Ammon, T., Popowicz, G.M., Krajewski, M., Nagel, R.J., Ares, M., Jr., Holak, T.A. and Jentsch, S. (2011) Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature*, **474**, 173-178.
94. Meyer, M., Plass, M., Perez-Valle, J., Eyras, E. and Vilardell, J. (2011) Deciphering 3 ' ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Mol Cell*, **43**, 1033-1039.
95. Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M. and Chanfreau, G.F. (2014) Widespread Use of Non-productive Alternative Splice Sites in Saccharomyces cerevisiae. *Plos Genet*, **10**.
96. Phasha, M.M., Wingfield, B.D., Coetzee, M.P.A., Santana, Q.C., Fourie, G. and Steenkamp, E.T. (2017) Architecture and Distribution of Introns in Core Genes of Four Fusarium Species. *G3-Genes Genom Genet*, **7**, 3809-3820.
97. Wang, K., Ussery, D.W. and Brunak, S. (2009) Analysis and prediction of gene splice sites in four Aspergillus genomes. *Fungal Genetics and Biology*, **46**, S14-S18.
98. Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyras, E. and Ast, G. (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res*, **18**, 88-103.
99. Sibthorp, C., Wu, H.H., Cowley, G., Wong, P.W.H., Palaima, P., Morozov, I.Y., Weedall, G.D. and Caddick, M.X. (2013) Transcriptome analysis of the filamentous fungus Aspergillus nidulans directed to the global identification of promoters. *Bmc Genomics*, **14**.
100. Xie, B.B., Li, D., Shi, W.L., Qin, Q.L., Wang, X.W., Rong, J.C., Sun, C.Y., Huang, F., Zhang, X.Y., Dong, X.W. *et al.* (2015) Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus Trichoderma longibrachiatum. *Bmc Genomics*, **16**.
101. Kempken, F. (2013) Alternative splicing in ascomycetes. *Appl Microbiol Biot*, **97**, 4235-4241.
102. Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell*, **36**, 593-608.
103. Neuveglise, C., Marck, C. and Gaillardin, C. (2011) The intronome of budding yeasts. *Cr Biol*, **334**, 662-670.
104. Busch, A. and Hertel, K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wires Rna*, **3**, 1-12.
105. Plass, M., Agirre, E., Reyes, D., Camara, F. and Eyras, E. (2008) Co-evolution of the branch site and SR proteins in eukaryotes. *Trends Genet*, **24**, 590-594.
106. Zhou, Z.H. and Fu, X.D. (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, **122**, 191-207.
107. Lipp, J.J., Marvin, M.C., Shokat, K.M. and Guthrie, C. (2015) SR protein kinases promote splicing of nonconsensus introns. *Nat Struct Mol Biol*, **22**, 611-617.
108. Shen, H.H. and Green, M.R. (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Gene Dev*, **20**, 1755-1765.

109. Graveley, B.R. and Maniatis, T. (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell*, **1**, 765-771.
110. Kress, T.L., Krogan, N.J. and Guthrie, C. (2008) A Single SR-like Protein, Npl3, Promotes Pre-mRNA Splicing in Budding Yeast. *Mol Cell*, **32**, 727-734.
111. Holmes, R.K., Tuck, A.C., Zhu, C.C., Dunn-Davies, H.R., Kudla, G., Clauder-Munster, S., Granneman, S., Steinmetz, L.M., Guthrie, C. and Tollervey, D. (2015) Loss of the Yeast SR Protein Npl3 Alters Gene Expression Due to Transcription Readthrough. *Plos Genet*, **11**.
112. Muddukrishna, B., Jackson, C.A. and Yu, M.C. (2017) Protein arginine methylation of Npl3 promotes splicing of the SUS1 intron harboring non-consensus 5 ' splice site and branch site. *Bba-Gene Regul Mech*, **1860**, 730-739.
113. Matera, A.G. and Wang, Z. (2014) A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*, **15**, 108-121.
114. Chen, H.C. and Cheng, S.C. (2012) Functional roles of protein splicing factors. *Bioscience Rep*, **32**, 345-359.
115. Maeder, C. and Guthrie, C. (2008) Modifications target spliceosome dynamics. *Nat Struct Mol Biol*, **15**, 426-428.
116. Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701-718.
117. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. and Luhrmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846-850.
118. Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J. and Moore, M.J. (2011) Ordered and Dynamic Assembly of Single Spliceosomes. *Science*, **331**, 1289-1295.
119. Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, **3**.
120. Cordin, O., Hahn, D. and Beggs, J.D. (2012) Structure, function and regulation of spliceosomal RNA helicases. *Curr Opin Cell Biol*, **24**, 431-438.
121. Cordin, O. and Beggs, J.D. (2013) RNA helicases in splicing. *Rna Biol*, **10**, 83-95.
122. Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B. *et al.* (2016) Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science*, **353**, 1399-1405.
123. Liu, S., Li, X., Zhang, L., Jiang, J., Hill, R.C., Cui, Y., Hansen, K.C., Zhou, Z.H. and Zhao, R. (2017) Structure of the yeast spliceosomal postcatalytic P complex. *Science*, **358**, 1278-1283.
124. Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J. and Nagai, K. (2016) Cryo-EM structure of the spliceosome immediately after branching. *Nature*, **537**, 197-+.
125. Plaschka, C., Lin, P.C. and Nagai, K. (2017) Structure of a pre-catalytic spliceosome. *Nature*, **546**, 617-+.
126. Yan, C.Y., Wan, R.X., Bai, R., Huang, G.X.Y. and Shi, Y.G. (2016) Structure of a yeast activated spliceosome at 3.5 angstrom resolution. *Science*, **353**, 904-911.
127. Wan, R.X., Yan, C.Y., Bai, R., Huang, G.X.Y. and Shi, Y.G. (2016) Structure of a yeast catalytic step I spliceosome at 3.4 angstrom resolution. *Science*, **353**, 895-904.
128. Yan, C.Y., Wan, R.X., Bai, R., Huang, G.X.Y. and Shi, Y.G. (2017) Structure of a yeast step II catalytically activated spliceosome. *Science*, **355**, 149-155.
129. Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.C., Newman, A.J. and Nagai, K. (2017) Structure of a spliceosome remodelled for exon ligation. *Nature*, **542**, 377-380.

130. Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W. and Nagai, K. (2016) Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 angstrom resolution. *Nature*, **530**, 298-+.
131. Wan, R.X., Yan, C.Y., Bai, R., Lei, J.L. and Shi, Y.G. (2017) Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. *Cell*, **171**, 120-+.
132. Hang, J., Wan, R.X., Yan, C.Y. and Shi, Y.G. (2015) Structural basis of pre-mRNA splicing. *Science*, **349**, 1191-1198.
133. Yan, C.Y., Hang, J., Wan, R.X., Huang, M., Wong, C.C.L. and Shi, Y.G. (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*, **349**, 1182-1191.
134. Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nat Rev Genet*, **17**, 19-32.
135. Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291-336.
136. Allemand, E., Batsche, E. and Muchardt, C. (2008) Splicing, transcription, and chromatin: a menage a trois. *Curr Opin Genet Dev*, **18**, 145-151.
137. Pagani, F. and Baralle, F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet*, **5**, 389-U382.
138. Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S. and Manley, J.L. (2015) Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *P Natl Acad Sci USA*, **112**, E4726-E4734.
139. Dvinge, H., Kim, E., Abdel-Wahab, O. and Bradley, R.K. (2016) RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer*, **16**, 413-430.
140. Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S. *et al.* (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun*, **7**.
141. Siegfried, Z. and Karni, R. (2018) The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev*, **48**, 16-21.
142. Thol, F., Kade, S., Schlarmann, C., Loffeld, P., Morgan, M., Krauter, J., Wlodarski, M.W., Kolking, B., Wichmann, M., Gorlich, K. *et al.* (2012) Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*, **119**, 3578-3584.
143. Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T. *et al.* (2015) Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*, **27**, 631-643.
144. Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K.M. *et al.* (2015) U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia*, **29**, 909-917.
145. Jenkins, J.L. and Kielkopf, C.L. (2017) Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. *Trends Genet*, **33**, 336-348.
146. Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E. *et al.* (2012) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet*, **44**, 53-U77.
147. Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P. and Bradley, R.K. (2015) U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res*, **25**, 14-26.
148. Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L.L., Gambe, R.G. *et al.* (2016) Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell*, **30**, 404-417.

149. Kesarwani, A.K., Ramirez, O., Gupta, A.K., Yang, X., Murthy, T., Minella, A.C. and Pillai, M.M. (2017) Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3 ' splice sites within RNA secondary structures. *Oncogene*, **36**, 1123-1133.
150. Alanis, E.F., Pinotti, M., Dal Mas, A., Balestra, D., Cavallari, N., Rogalska, M.E., Bernardi, F. and Pagani, F. (2012) An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum Mol Genet*, **21**, 2389-2398.
151. Dal Mas, A., Fortugno, P., Donadon, I., Levati, L., Castiglia, D. and Pagani, F. (2015) Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. *Hum Mutat*, **36**, 504-512.
152. Rogalska, M.E., Tajnik, M., Licastro, D., Bussani, E., Camparini, L., Mattioli, C. and Pagani, F. (2016) Therapeutic activity of modified U1 core spliceosomal particles. *Nat Commun*, **7**, 11168.
153. Singh, N.N., del Rio-Malewski, J.B., Luo, D., Ottesen, E.W., Howell, M.D. and Singh, R.N. (2017) Activation of a cryptic 5 ' splice site reverses the impact of pathogenic splice site mutations in the spinal muscular atrophy gene. *Nucleic Acids Res*, **45**, 12214-12240.
154. Sanchez-Alcudia, R., Perez, B., Perez-Cerda, C., Ugarte, M. and Desviat, L.R. (2011) Overexpression of adapted U1snRNA in patients' cells to correct a 5 ' splice site mutation in propionic acidemia. *Mol Genet Metab*, **102**, 134-138.
155. Pinotti, M., Rizzotto, L., Balestra, D., Lewandowska, M.A., Cavallari, N., Marchetti, G., Bernardi, F. and Pagani, F. (2008) U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood*, **111**, 2681-2684.
156. Pinotti, M., Balestra, D., Rizzotto, L., Maestri, I., Pagani, F. and Bernardi, F. (2009) Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood*, **113**, 6461-6464.
157. Balestra, D., Faella, A., Margaritis, P., Cavallari, N., Pagani, F., Bernardi, F., Arruda, V.R. and Pinotti, M. (2014) An engineered U1 small nuclear RNA rescues splicing-defective coagulation F7 gene expression in mice. *J Thromb Haemost*, **12**, 177-185.
158. Balestra, D., Scalet, D., Pagani, F., Rogalska, M.E., Mari, R., Bernardi, F. and Pinotti, M. (2016) An Exon-Specific U1snRNA Induces a Robust Factor IX Activity in Mice Expressing Multiple Human FIX Splicing Mutants. *Mol Ther-Nucl Acids*, **5**.
159. Balestra, D., Barbon, E., Scalet, D., Cavallari, N., Perrone, D., Zanibellato, S., Bernardi, F. and Pinotti, M. (2015) Regulation of a strong F9 cryptic 5 ' ss by intrinsic elements and by combination of tailored U1snRNAs with antisense oligonucleotides. *Hum Mol Genet*, **24**, 4809-4816.
160. Knoepfel, S.A., Abad, A., Abad, X., Fortes, P. and Berkhout, B. (2012) Design of modified U1i molecules against HIV-1 RNA. *Antivir Res*, **94**, 208-216.
161. Glaus, E., Schmid, F., Da Costa, R., Berger, W. and Neidhardt, J. (2011) Gene Therapeutic Approach Using Mutation-adapted U1 snRNA to Correct a RPGR Splice Defect in Patient-derived Cells. *Mol Ther*, **19**, 936-941.
162. Schmid, F., Glaus, E., Barthelmes, D., Fliegauf, M., Gaspar, H., Nurnberg, G., Nurnberg, P., Omran, H., Berger, W. and Neidhardt, J. (2011) U1 snRNA-Mediated Gene Therapeutic Correction of Splice Defects Caused by an Exceptionally Mild BBS Mutation. *Hum Mutat*, **32**, 815-824.
163. Agrawal, A.A., McLaughlin, K.J., Jenkins, J.L. and Kielkopf, C.L. (2014) Structure-guided U2AF65 variant improves recognition and splicing of a defective pre-mRNA. *Proc Natl Acad Sci U S A*, **111**, 17420-17425.
164. Kataoka, N. (2017) Modulation of aberrant splicing in human RNA diseases by chemical compounds. *Hum Genet*, **136**, 1237-1245.
165. Ohe, K. and Hagiwara, M. (2015) Modulation of Alternative Splicing with Chemical Compounds in New Therapeutics for Human Diseases. *Acs Chemical Biology*, **10**, 914-924.

166. Martinez-Montiel, N., Rosas-Murrieta, N.H., Martinez-Montiel, M., Gaspariano-Cholula, M.P. and Martinez-Contreras, R.D. (2016) Microbial and Natural Metabolites That Inhibit Splicing: A Powerful Alternative for Cancer Treatment. *Biomed Res Int*.
167. Leon, B., Kashyap, M.K., Chan, W.C., Krug, K.A., Castro, J.E., La Clair, J.J. and Burkart, M.D. (2017) A Challenging Pie to Splice: Drugging the Spliceosome. *Angew Chem Int Edit*, **56**, 12052-12063.
168. Lagisetti, C., Palacios, G., Goronga, T., Freeman, B., Caufield, W. and Webb, T.R. (2013) Optimization of Antitumor Modulators of Pre-mRNA Splicing. *J Med Chem*, **56**, 10033-10044.
169. Palacino, J., Swalley, S.E., Song, C., Cheung, A.K., Shu, L., Zhang, X.L., Van Hoosear, M., Shin, Y., Chin, D.N., Keller, C.G. *et al.* (2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat Chem Biol*, **11**, 511-+.
170. Naryshkin, N.A., Weetall, M., Dakka, A., Narasimhan, J., Zhao, X., Feng, Z.H., Ling, K.K.Y., Karp, G.M., Qi, H.Y., Woll, M.G. *et al.* (2014) SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science*, **345**, 688-693.
171. Hsu, T.Y.T., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Echeverria, G.V., Sun, T.T., Kurley, S.J., Tyagi, S. *et al.* (2015) The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature*, **525**, 384-+.
172. Bonnal, S., Vigevani, L. and Valcarcel, J. (2012) The spliceosome as a target of novel antitumour drugs. *Nat Rev Drug Discov*, **11**, 847-859.
173. Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M. and Mizui, Y. (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol*, **3**, 570-575.
174. Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H. *et al.* (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol*, **3**, 576-583.
175. Lee, S.C.W., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., Durham, B.H., Yoshimi, A., Kim, Y.J., Thomas, M. *et al.* (2016) Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins (vol 22, pg 672, 2016). *Nat Med*, **22**, 692-692.
176. Xargay-Torrent, S., Lopez-Guerra, M., Rosich, L., Montraveta, A., Roldan, J., Rodriguez, V., Villamor, N., Aymerich, M., Lagisetti, C., Webb, T.R. *et al.* (2015) The splicing modulator sudemycin induces a specific antitumor response and cooperates with ibrutinib in chronic lymphocytic leukemia. *Oncotarget*, **6**, 22734-22749.
177. Fan, L.Y., Lagisetti, C., Edwards, C.C., Webb, T.R. and Potter, P.M. (2011) Sudemycins, Novel Small Molecule Analogues of FR901464, Induce Alternative Gene Splicing. *Acs Chemical Biology*, **6**, 582-589.
178. Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagisetti, C., de la Grange, P., Horbinski, C., Fondufe-Mittendorf, Y.N., Webb, T.R. and Stamm, S. (2014) Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic Acids Res*, **42**, 4947-4961.
179. Egecioglu, D.E., Kawashima, T.R. and Chanfreau, G.F. (2012) Quality control of MATa1 splicing and exon skipping by nuclear RNA degradation. *Nucleic Acids Res*, **40**, 1787-1796.
180. Miller, A.M. (1984) The yeast MATa1 gene contains two introns. *Embo J*, **3**, 1061-1065.
181. Ner, S.S. and Smith, M. (1989) Role of Intron Splicing in the Function of the Mata1 Gene of Saccharomyces-Cerevisiae. *Molecular and Cellular Biology*, **9**, 4613-4620.
182. Schmidlin, T., Kaeberlein, M., Kudlow, B.A., MacKay, V., Lockshon, D. and Kennedy, B.K. (2008) Single-gene deletions that restore mating competence to diploid yeast. *Fems Yeast Res*, **8**, 276-286.

183. Tuo, S., Nakashima, K. and Pringle, J.R. (2012) Apparent Defect in Yeast Bud-Site Selection Due to a Specific Failure to Splice the Pre-mRNA of a Regulator of Cell-Type-Specific Transcription. *Plos One*, **7**.
184. Kohrer, K., Vogel, K. and Domdey, H. (1990) A yeast tRNA precursor containing a pre-mRNA intron is spliced via the pre-mRNA splicing mechanism. *Embo J*, **9**, 705-709.
185. Cramer, R.A., Jr., Stajich, J.E., Yamanaka, Y., Dietrich, F.S., Steinbach, W.J. and Perfect, J.R. (2006) Phylogenomic analysis of non-ribosomal peptide synthetases in the genus Aspergillus. *Gene*, **383**, 24-32.
186. Reeves, E.P., Reiber, K., Neville, C., Scheibner, O., Kavanagh, K. and Doyle, S. (2006) A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in Aspergillus fumigatus. *FEBS J*, **273**, 3038-3053.
187. O'Hanlon, K.A., Gallagher, L., Schrettl, M., Jochl, C., Kavanagh, K., Larsen, T.O. and Doyle, S. (2012) Nonribosomal Peptide Synthetase Genes pesL and pes1 Are Essential for Fumigaclavine C Production in Aspergillus fumigatus. *Appl Environ Microb*, **78**, 3166-3176.
188. Lee, I., Oh, J.H., Shwab, E.K., Dagenais, T.R., Andes, D. and Keller, N.P. (2009) HdaA, a class 2 histone deacetylase of Aspergillus fumigatus, affects germination and secondary metabolite production. *Fungal Genet Biol*, **46**, 782-790.
189. Perrin, R.M., Fedorova, N.D., Bok, J.W., Cramer, R.A., Wortman, J.R., Kim, H.S., Nierman, W.C. and Keller, N.P. (2007) Transcriptional regulation of chemical diversity in Aspergillus fumigatus by LaeA. *PLoS Pathog*, **3**, e50.
190. Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P. and Boeke, J.D. (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115-132.
191. Lesser, C.F. and Guthrie, C. (1993) Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics*, **133**, 851-863.
192. Smith, D.J., Query, C.C. and Konarska, M.M. trans-Splicing to Spliceosomal U2 snRNA Suggests Disruption of Branch Site-U2 Pairing during Pre-mRNA Splicing. *Mol Cell*, **26**, 883-890.
193. Perriman, R. and Ares, M., Jr. (2010) Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol Cell*, **38**, 416-427.
194. Kuijpers, N.G., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2013) One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in Saccharomyces cerevisiae. *Fems Yeast Res*, **13**, 769-781.
195. Smith, J.D., Schlecht, U., Xu, W.H., Suresh, S., Horecka, J., Proctor, M.J., Aiyar, R.S., Bennett, R.A.O., Chu, A., Li, Y.F.G. *et al.* (2017) A method for high-throughput production of sequence-verified DNA libraries and strain collections. *Mol Syst Biol*, **13**.
196. Peng, B., Williams, T.C., Henry, M., Nielsen, L.K. and Vickers, C.E. (2015) Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb Cell Fact*, **14**, 91.
197. St Onge, R.P., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R.W., Nislow, C., Roth, F.P. and Giaever, G. (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet*, **39**, 199-206.
198. Schlecht, U., Miranda, M., Suresh, S., Davis, R.W. and St Onge, R.P. (2012) Multiplex assay for condition-dependent changes in protein-protein interactions. *P Natl Acad Sci USA*, **109**, 9213-9218.

199. Smith, A.M., Ammar, R., Nislow, C. and Giaever, G. (2010) A survey of yeast genomic assays for drug and target discovery. *Pharmacol Ther*, **127**, 156-164.
200. Yoshida, H., Park, S.Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y. *et al.* (2015) A novel 3 ' splice site recognition by the two zinc fingers in the U2AF small subunit. *Gene Dev*, **29**, 1649-1660.
201. Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., Newman, A.J. and Nagai, K. (2017) Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science*, **358**, 1283-1288.
202. Gahura, O., Hammann, C., Valentova, A., Puta, F. and Folk, P. (2011) Secondary structure is required for 3 ' splice site recognition in yeast. *Nucleic Acids Res*, **39**, 9759-9767.
203. Gautam, A., Grainger, R.J., Vilardell, J., Barrass, J.D. and Beggs, J.D. (2015) Cwc21p promotes the second step conformation of the spliceosome and modulates 3' splice site selection. *Nucleic Acids Res*, **43**, 3309-3317.
204. Cellini, A., Felder, E. and Rossi, J.J. (1986) Yeast Pre-Messenger-Rna Splicing Efficiency Depends on Critical Spacing Requirements between the Branch Point and 3-' Splice Site. *Embo J*, **5**, 1023-1030.
205. Plass, M., Codony-Servat, C., Ferreira, P.G., Vilardell, J. and Eyras, E. (2012) RNA secondary structure mediates alternative 3 ' ss selection in Saccharomyces cerevisiae. *Rna-a Publication of the Rna Society*, **18**, 1103-1115.
206. Schreiber, K., Csaba, G., Haslbeck, M. and Zimmer, R. (2015) Alternative Splicing in Next Generation Sequencing Data of Saccharomyces cerevisiae. *Plos One*, **10**, e0140487.
207. Kohrer, K. and Domdey, H. (1988) Splicing and spliceosome formation of the yeast MATa1 transcript require a minimum distance from the 5' splice site to the internal branch acceptor site. *Nucleic Acids Res*, **16**, 9457-9475.
208. Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B. *et al.* (2012) Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Reports*, **1**, 543-556.
209. Zhang, J., Kuo, C.C. and Chen, L. (2011) GC content around splice sites affects splicing through pre-mRNA secondary structures. *Bmc Genomics*, **12**, 90.
210. Warf, M.B. and Berglund, J.A. (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci*, **35**, 169-178.
211. Jin, Y.F., Yang, Y. and Zhang, P. (2011) New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *Rna Biol*, **8**, 450-457.
212. Buratti, E. and Baralle, F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol*, **24**, 10505-10514.
213. Abuqattam, A., Gallego, J. and Rodriguez-Navarro, S. (2016) An intronic RNA structure modulates expression of the mRNA biogenesis factor Sus1. *RNA (New York, N.Y.)*, **22**, 75-86.
214. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, **2**, 31-34.
215. Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289-1291.
216. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res*, **40**, e115.
217. Zhao, S. and Fernald, R.D. (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology*, **12**, 1047-1064.

218. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*, **25**, 402-408.
219. Rymond, B.C. (2010) THE BRANCHPOINT BINDING PROTEIN In and Out of the Spliceosome Cycle. *Post-Transcriptional Regulation by Star Proteins: Control of Rna Metabolism in Development and Disease*, **693**, 123-141.
220. Parker, R., Siliciano, P.G. and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, **49**, 229-239.
221. Gould, G.M., Paggi, J.M., Guo, Y., Phizicky, D.V., Zinshteyn, B., Wang, E.T., Gilbert, W.V., Gifford, D.K. and Burge, C.B. (2016) Identification of new branch points and unconventional introns in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, **22**, 1522-1534.
222. Zhang, Q., Fan, X.D., Wang, Y.J., Sun, M.A., Shao, J.L. and Guo, D.J. (2017) BPP: a sequence-based algorithm for branch point prediction. *Bioinformatics*, **33**, 3166-3172.
223. Smith, D.J., Konarska, M.M. and Query, C.C. (2009) Insights into branch nucleophile positioning and activation from an orthogonal pre-mRNA splicing system in yeast. *Mol Cell*, **34**, 333-343.
224. Perumal, K. and Reddy, R. (2002) The 3' end formation in small RNAs. *Gene expression*, **10**, 59-78.
225. Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S. *et al.* (2015) Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Systems*, **1**, 88-96.
226. Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T. and Zhao, H. (2015) Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. *Acs Synth Biol*, **4**, 585-594.
227. Egecioglu, D.E. and Chanfreau, G. (2011) Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. *RNA (New York, N.Y.)*, **17**, 383-389.
228. De, I., Schmitzova, J. and Pena, V. (2016) The organization and contribution of helicases to RNA splicing. *Wires Rna*, **7**, 259-274.
229. Liang, W.W. and Cheng, S.C. (2015) A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev*, **29**, 81-93.
230. Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C.C. and Yu, Y.T. (2016) Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *Embo J*, **35**, 654-667.
231. Kosowski, T.R., Keys, H.R., Quan, T.K. and Ruby, S.W. (2009) DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *RNA (New York, N.Y.)*, **15**, 1345-1362.
232. Karaduman, R., Chanarat, S., Pfander, B. and Jentsch, S. (2017) Error-Prone Splicing Controlled by the Ubiquitin Relative Hub1. *Mol Cell*, **67**, 423-+.
233. Tang, Q., Rodriguez-Santiago, S., Wang, J., Pu, J., Yuste, A., Gupta, V., Moldon, A., Xu, Y.Z. and Query, C.C. (2016) SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. *Genes Dev*, **30**, 2710-2723.
234. Koodathingal, P. and Staley, J.P. (2013) Splicing fidelity DEAD/H-box ATPases as molecular clocks. *Rna Biol*, **10**, 1073-1079.
235. Xu, Y.Z. and Query, C.C. (2007) Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol Cell*, **28**, 838-849.

236. Rutz, B. and Seraphin, B. (1999) Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *Rna-a Publication of the Rna Society*, **5**, 819-831.
237. Garrey, S.M., Voelker, R. and Berglund, J.A. (2006) An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *J Biol Chem*, **281**, 27443-27453.
238. Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *Rna-a Publication of the Rna Society*, **4**, 551-565.
239. Chang, J., Schwer, B. and Shuman, S. (2012) Structure-function analysis and genetic interactions of the yeast branchpoint binding protein Msl5. *Nucleic Acids Res*, **40**, 4539-4552.
240. Wang, Q., Zhang, L., Lynn, B. and Rymond, B.C. (2008) A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Res*, **36**, 2787-2798.
241. Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, **8**, 479-490.
242. Rutz, B. and Seraphin, B. (2000) A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *Embo J*, **19**, 1873-1886.
243. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89**, 781-787.
244. Jakounas, T., Sonde, I., Herrgard, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K. and Keasling, J.D. (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metab Eng*, **28**, 213-222.
245. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, **302**, 205-217.
246. Wallace, I.M., O'Sullivan, O., Higgins, D.G. and Notredame, C. (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*, **34**, 1692-1699.
247. Moretti, S., Armougom, F., Wallace, I.M., Higgins, D.G., Jongeneel, C.V. and Notredame, C. (2007) The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res*, **35**, W645-648.
248. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F. and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*, **39**, W13-17.
249. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, **42**, W320-324.
250. Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H. and Gentleman, R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, **25**, 2607-2608.
251. Wagih, O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645-3647.
252. Zhang, J., Kuo, C.C.J. and Chen, L.A. (2011) GC content around splice sites affects splicing through pre-mRNA secondary structures. *Bmc Genomics*, **12**.
253. Plass, M. and Eyras, E. (2014) Approaches to Link RNA Secondary Structures with Splicing Regulation. *Methods Mol Biol*, **1126**, 341-356.

254. Rogic, S., Montpetit, B., Hoos, H.H., Mackworth, A.K., Ouellette, B.F.F. and Hieter, P. (2008) Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *Bmc Genomics*, **9**.
255. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*, **288**, 911-940.
256. Hofacker, I.L. (2004) RNA secondary structure analysis using the Vienna RNA package. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, **Chapter 12**, Unit 12.12.
257. Kerpedjiev, P., Hammer, S. and Hofacker, I.L. (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377-3379.
258. Schwer, B., Chang, J. and Shuman, S. (2013) Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5 center dot Mud2 branchpoint-binding complex and other spliceosome assembly factors. *Nucleic Acids Res*, **41**, 7485-7500.
259. Freund, M., Hicks, M.J., Otte, M., Hertel, K.J. and Schaal, H. (2005) Extended base pair complementarity between U1 snRNA and the 5 ' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5 ' splice site recognition. *Nucleic Acids Res*, **33**, 5112-5119.
260. Mitrovich, Q.M. and Guthrie, C. (2007) Evolution of small nuclear RNAs in S. cerevisiae, C. albicans, and other hemiascomycetous yeasts. *RNA (New York, N.Y.)*, **13**, 2066-2080.
261. Yang, F., Wang, X.Y., Zhang, Z.M., Pu, J., Fan, Y.J., Zhou, J.H., Query, C.C. and Xu, Y.Z. (2013) Splicing proofreading at 5' splice sites by ATPase Prp28p. *Nucleic Acids Res*, **41**, 4660-4670.
262. Burgess, S.M. and Guthrie, C. (1993) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell*, **73**, 1377-1391.
263. Du, H. and Rosbash, M. (2001) Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. *Rna-a Publication of the Rna Society*, **7**, 133-142.
264. Kondo, Y., Oubridge, C., van Roon, A.M.M. and Nagai, K. (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5 ' splice site recognition. *Elife*, **4**.
265. Schwer, B. and Shuman, S. (2014) Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. *Nucleic Acids Res*, **42**, 4697-4711.
266. Larson, J.D. and Hoskins, A.A. (2017) Dynamics and consequences of spliceosome E complex formation. *Elife*, **6**.
267. Munding, E.M., Shiue, L., Katzman, S., Donohue, J.P. and Ares, M. (2013) Competition between Pre-mRNAs for the Splicing Machinery Drives Global Regulation of Splicing. *Mol Cell*, **51**, 338-348.
268. Ares, M., Grate, L. and Pauling, M.H. (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *Rna-a Publication of the Rna Society*, **5**, 1138-1139.
269. Lopez, P.J. and Seraphin, B. (1999) Genomic-scale quantitative analysis of yeast pre-mRNA splicing: Implications for splice-site recognition. *Rna-a Publication of the Rna Society*, **5**, 1135-1137.
270. Warner, J.R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci*, **24**, 437-440.

271. Smith, J.D., Suresh, S., Schlecht, U., Wu, M., Wagih, O., Peltz, G., Davis, R.W., Steinmetz, L.M., Parts, L. and St Onge, R.P. (2016) Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol*, **17**, 45.
272. Carrocci, T.J., Zoerner, D.M., Paulson, J.C. and Hoskins, A.A. (2017) SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. *Nucleic Acids Res*, **45**, 4837-4852.
273. Yaegashi, J., Oakley, B.R. and Wang, C.C. (2014) Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in Aspergillus nidulans. *J Ind Microbiol Biotechnol*, **41**, 433-442.
274. Song, Z., Bakeer, W., Marshall, J.W., Yakasai, A.A., Khalid, R.M., Collemare, J., Skellam, E., Tharreau, D., Lebrun, M.H., Lazarus, C.M. *et al.* (2015) Heterologous expression of the avirulence gene ACE1 from the fungal rice pathogen Magnaporthe oryzae. *Chem Sci*, **6**, 4837-4845.