UNIVERSITY OF CALIFORNIA
RIVERSIDE

Alteration of Secretion Rates in Model Yeasts and Development of Assays for Genome-Wide Screening of Secretion

A Dissertation submitted for partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Bioengineering

by

Dominic Biondo

June 2026

Dissertation Committee:

Dr. Ian Wheeldon, Chairperson
Dr. Katherine Borkovich
Dr. Joshua Morgan

The Dissertation of Dominic Biondo is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

This journey through graduate school has been one of the most challenging yet rewarding times of my life. I would first like to thank Dr. Ian Wheeldon for his mentorship and guidance, especially during the difficult years. Joining his lab after working with my previous PI, Dr. Justin Chartron, was one of the smoothest transitions I could have hoped for, thanks to his quick acceptance and willingness to incorporate the projects I had begun exploring. This transition also happened at the start of the COVID-19 lockdown, which made his support even more meaningful. Under his tutelage, I not only grew as an experimentalist but also sharpened my skills in data analysis and figure preparation, which have become some of my strongest tools as a scientist.

I would also like to thank my committee members, Dr. Borkovich and Dr. Morgan, for their advice and encouragement. From them, I learned to approach experiments from multiple angles, to broaden my perspective, and to think more critically about how statistical analyses apply in a true laboratory setting—lessons that extended well beyond the classroom.

My lab mates have also been invaluable throughout this journey. Dr. Aida Tafrishi provided critical help and insight into my work on *K. marxianus* while others provided insight into how they would approach a problem. I am especially grateful to Dr. Clifford Morrison and Dr. Mario León-López, who have been not only amazing colleagues but also some of the closest friends I could

ask for. They supported me through both academic challenges and personal struggles, always being there when I needed them most.

I owe the deepest gratitude to my family, who have been my strongest support system both figuratively and literally. As the third of eight siblings, I have always had the privilege of being surrounded by a supportive group that has been there for me at every step of my life. My parents, in particular, have been my number one supporter from the very beginning, providing encouragement, strength, and unconditional love through every stage of this journey.

One person I especially want to acknowledge is my best friend, Brittany. We met through a mutual friend only a few years ago, but she quickly became one of the most supportive figures in my life. It did not take long for our friendship to grow, and she has since been a constant source of encouragement and joy. Her presence has made this journey not only easier but truly better.

Finally, I want to thank the many friends and students at UCR who made my time here so enjoyable outside of the lab. Their friendship and camaraderie provided balance to the intensity of graduate school and made this chapter of my life all the more meaningful.

ABSTRACT OF THE DISSERTATION

Alteration of Secretion Rates in Model Yeasts and Development of Assays for Genome-Wide Screening of Secretion

by

Dominic Biondo

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, June 2026
Dr. Ian Wheeldon, Chairperson

In this work, I investigated strategies to improve recombinant protein production in yeast by targeting both transcriptional regulation and protein secretion. First, I characterized how mutations within the promoter region of Saccharomyces cerevisiae influence gene expression. Targeted mutagenesis revealed that increasing GC content within the TATA box and pre-initiation complex consistently reduced expression, identifying these regions as highly sensitive to sequence composition. I then expanded this approach using CRISPR-enabled trackable genome engineering (CREATE) to introduce a broader range of mutation types across an extended promoter region. This analysis confirmed that GC-rich substitutions were strongly associated with decreased expression, while certain AT- or mixed-base substitutions in the scanning region and transcriptional start site contributed to increased promoter activity. Finally, I adapted a mid-western (colony blot) secretion assay for use in

Kluyveromyces marxianus. Using HiBiT-tagged constructs, I established a range of low, medium and high secretion outputs across multiple promoter–signal peptide–gene combinations. Together, these results demonstrate that combining promoter engineering with secretion screening provides a systematic framework for optimizing protein production in yeast.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background

Biotechnology has become one of the fastest growing industries in the world, with applications spanning medicine, agriculture, and industrial production. In the United States, the biotech sector has expanded at an average annual rate of nearly 10% over the past decade, while globally, biopharmaceuticals now generate over 160 billion USD, representing roughly 20% of the pharmaceutical market and continuing to rise [1, 2]. Recombinant proteins account for a major portion of this market, including monoclonal antibodies, vaccines, hormones, cytokines, blood factors, and industrial enzymes, all of which rely on optimized host systems for production.

The earliest recombinant proteins were produced in bacterial systems, most notably *Escherichia coli*. The FDA approval of recombinant human insulin in 1982 marked a milestone in biotechnology and demonstrated the potential of microbial systems for large-scale protein production [3]. While bacterial hosts remain widely used, they are limited by their inability to perform many post translational modifications and by challenges in secreting complex eukaryotic proteins. These shortcomings have accelerated the use of eukaryotic hosts, particularly yeasts, which combine the rapid growth and scalability of microbial systems with the ability to fold proteins and perform glycosylation.

Saccharomyces cerevisiae is the best-studied yeast and has long been employed for both industrial enzymes and therapeutic proteins. However, its secretory pathway can be inefficient, leading to problems such as misfolding, degradation, and hyperglycosylation of heterologous proteins [4, 5]. To address these limitations, non-conventional yeasts such as *Kluyveromyces marxianus*, *Yarrowia lipolytica*, and *Pichia pastoris* have gained increasing attention. Each species brings unique physiological traits and secretory capabilities: *K. marxianus* grows rapidly at elevated temperatures and can metabolize diverse carbon sources, *Y. lipolytica* is adapted for lipid metabolism and secretion of hydrophobic proteins, and *P. pastoris* possesses strong methanol-inducible promoters and robust secretion machinery [6, 7]. Together, these yeasts expand the biotechnological toolkit and provide alternative hosts for proteins that are challenging to produce in *S. cerevisiae*.

The secretory pathway is central to protein production in all yeasts. Proteins destined for secretion are translated into the endoplasmic reticulum (ER), where they undergo folding, disulfide bond formation, and glycosylation. Properly folded proteins are then trafficked through the Golgi apparatus and delivered to the cell surface, while misfolded proteins are targeted for ER associated degradation [8, 5]. Bottlenecks can occur at multiple stages of this pathway. Limited ER folding capacity, inefficient

trafficking, or proteolytic degradation often restrict secretion yields. Engineering strategies to improve secretion have therefore included overexpression of folding chaperones, optimization of signal peptides, modification of glycosylation pathways, and deletion of vacuolar proteases [9, 4]. Despite progress, secretion remains a common bottleneck, and high-throughput assays are needed to measure secretion capacity at scale. In Chapter 4 of this dissertation, I describe the development of a colony blot assay in K. marxianus that adapts a HiBiT/LgBiT luminescent complementation system to measure secretion directly from colonies, allowing hundreds of variants to be screened simultaneously.

While the secretory pathway determines how efficiently proteins exit the cell, the promoter region dictates how much protein is produced. The yeast promoter can be divided into distinct elements that coordinate transcription initiation. The TATA box recruits the TATA-binding protein (TBP), the Pre Initiation Complex (PIC) provides the scaffold for assembly of general transcription factors, the scanning region facilitates polymerase progression, and the transcriptional start site (TSS) defines the precise location of initiation [10, 11, 12]. These elements are not equally tolerant to mutation. The TATA box is particularly sensitive to sequence changes, with A/T-rich sequences favoring TBP binding and GC substitutions reducing binding affinity and transcriptional output [13, 14]. The PIC region is also critical for transcriptional regulation, while

the scanning region and TSS generally display greater flexibility. Promoter engineering strategies have exploited this sensitivity to tune expression levels, including random mutagenesis, saturation mutagenesis, and systematic design of synthetic promoters [15, 16, 17]. These studies show that disruptive mutations are more common than activating ones, reflecting the inherent vulnerability of transcription initiation.

1.2 Thesis Organization

Promoter regulation and the secretory pathway therefore represent complementary control points for recombinant protein production. Modifying promoter architecture provides a way to redistribute transcriptional resources, either by reducing expression of competing genes or tuning heterologous gene output, while engineering the secretory pathway ensures that proteins are folded and exported efficiently. By combining these approaches, it is possible to improve yields both at the level of transcription and at the level of secretion. This dissertation explores both strategies. Chapters 2 and 3 describe targeted promoter mutagenesis in *S. cerevisiae* using two approaches: a focused Inscripta-based library and a broader CREATE-based library. These experiments identified promoter elements most sensitive to mutation and established general principles of promoter sensitivity as well as

promoter-specific differences. Chapter 4 shifts to K. marxianus and describes the development of a high throughput colony blot assay for secretion, which enables systematic evaluation of secretion across many variants and lays the groundwork for genome-wide screening.

Recombinant protein production remains a cornerstone of biotechnology, and yeasts represent some of the most versatile hosts for this purpose. Achieving high yields, however, requires simultaneous control of transcriptional output and secretion capacity. The work presented in this dissertation develops tools that probe both levels of control. Promoter mutagenesis in *S. cerevisiae* establishes rules for tuning expression through the core promoter, while the colony blot assay in K. marxianus establishes a scalable method to evaluate and improve secretion. Together, these complementary approaches expand the toolkit for yeast engineering and provide new strategies for optimizing recombinant protein production.

1.3 References

[1] Walsh, G. (2018). Biopharmaceutical benchmarks 2018. Nature Biotechnology, 36(12), 1136–1145. https://doi.org/10.1038/nbt.4305

[2] Langer, E. S., & Rader, R. A. (2014). Biopharmaceutical manufacturing capacity and production. BioProcess International, 12(2), 40–47.

[3] Johnson, I. S. (1983). Human insulin from recombinant DNA technology. Science, 219(4585), 632–637. https://doi.org/10.1126/science.6337396

[4] Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M. L., Mattanovich, D., & Villaverde, A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. Microbial Cell Factories, 7, 11. https://doi.org/10.1186/1475-2859-7-11

[5] Delic, M., Mattanovich, D., & Gasser, B. (2013). Repression of gene expression in Pichia pastoris is one of the major limitations for recombinant protein production. Biotechnology Journal, 8(5), 564–576. https://doi.org/10.1002/biot.201200193

[6] Madzak, C. (2018). Engineering Yarrowia lipolytica for use in biotechnological applications: A review. Molecular Biotechnology, 60, 621–635. https://doi.org/10.1007/s12033-018-0093-4

[7] Spohner, S. C., Müller, H., Quitmann, H., & Czermak, P. (2015). Expression of enzymes for the usage in food and feed industry with Pichia pastoris. Journal of Biotechnology, 202, 118–134. https://doi.org/10.1016/j.jbiotec.2015.01.027

[8] Conesa, A., Punt, P. J., van Luijk, N., & van den Hondel, C. A. M. J. J. (2001). The secretion pathway in filamentous fungi: A biotechnological view. Fungal Genetics and Biology, 33(3), 155–171. https://doi.org/10.1006/fgbi.2001.1276

[9] Porro, D., Sauer, M., Branduardi, P., & Mattanovich, D. (2011). Recombinant protein production in yeasts. Molecular Biotechnology, 48(2), 191–203. https://doi.org/10.1007/s12033-010-9366-y

[10] Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell, 116(5), 699–709. https://doi.org/10.1016/S0092-8674(04)00205-3

[11] Hahn, S., & Young, E. T. (2011). Transcriptional regulation in *Saccharomyces cerevisiae*. Genetics, 189(3), 705–736. https://doi.org/10.1534/genetics.111.127019

[12] Rhee, H. S., & Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature, 483(7389), 295–301. https://doi.org/10.1038/nature10799

[13] Lubliner, S., Keren, L., & Segal, E. (2013). Sequence features of yeast and human core promoters. Nature Biotechnology, 31(6), 548–554. https://doi.org/10.1038/nbt.2590

[14] Mogno, I., Kwasnieski, J. C., & Cohen, B. A. (2013). Massively parallel promoter assays reveal effects of binding site variants. Genome Research, 23(11), 1908–1915. https://doi.org/10.1101/gr.157891.113

[15] Patwardhan, R. P., Lee, C., Litvin, O., Young, D. L., Pe'er, D., & Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nature Biotechnology, 27(12), 1173–1175. https://doi.org/10.1038/nbt.1589

[16] Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., & Segal, E. (2012). Inferring gene regulatory logic from high-throughput promoter measurements. Nature Biotechnology, 30(6), 521–530. https://doi.org/10.1038/nbt.2205

[17] Blazeck, J., Garg, R., Reed, B., & Alper, H. S. (2012). Controlling promoter strength in *Saccharomyces cerevisiae* using synthetic hybrid promoters. Biotechnology and Bioengineering, 109(11), 2884–2895. https://doi.org/10.1002/bit.24552

Chapter 2: Promoter Mutagenesis Using Inscripta Libraries in *S. cerevisiae*

2.1 Abstract

We generated random promoter mutations in *Saccharomyces cerevisiae* to identify sequence changes that reduce protein production without completely abolishing expression. Using CAN1::sfGFP and CCW22::DsRed reporter strains, we found that A/T → G/C substitutions in the TATA box and Pre-Initiation Complex (PIC) regions consistently reduced gene expression. In contrast, mutations in the scanning region and transcriptional start site (TSS) had weaker or inconsistent effects. These results demonstrate that specific classes of promoter mutations can fine-tune expression levels and provide a framework for targeted regulation of gene output in yeast.

2.2 Introduction

Recombinant proteins are widely produced in both bacteria and yeast. Escherichia coli remains one of the most common production hosts and was used to produce the first U.S. Food and Drug Administration (FDA)-approved recombinant product, human insulin [1]. Since then, recombinant proteins have become essential for medicine and industry, including monoclonal antibodies, vaccines, cytokines, and enzymes [2, 3]. With the advent of CRISPR

technologies, high-throughput sequencing, and advances in upstream and bioprocessing workflows, recombinant protein production continues to expand in scale and scope [4, 5].

Among eukaryotic hosts, *S. cerevisiae* is particularly attractive due to its long history of industrial use, its robust folding and secretion machinery, and its ability to perform post-translational modifications [6, 7]. A key determinant of expression in yeast is the promoter region, which regulates transcription initiation. The core promoter can be divided into functional elements including the TATA box, PIC assembly site, scanning region, and the TSS. Each contributes differently to transcription: the TATA box positions the TATA-binding protein (TBP), the PIC recruits general transcription factors, the scanning region facilitates promoter clearance, and the TSS defines initiation [8, 9, 10]. Studies have shown that mutations in these elements can alter transcriptional output, either increasing or decreasing expression, depending on the sequence context [11, 12].

Based on this background, we hypothesized that targeted promoter mutations could tune expression by reducing activity without eliminating it entirely. This principle has practical value in biotechnology, where partial knockdown of genes can redistribute cellular resources toward heterologous protein secretion or pathway flux.

To test this hypothesis, we constructed fluorescent reporter strains in which CAN1, HIS3, and CCW22 were replaced with sfGFP, YFP, and DsRed, respectively. CAN1 encodes an arginine permease dispensable in rich media, HIS3 encodes an enzyme in histidine biosynthesis that is not required with supplementation, and CCW22 encodes a non-essential cell wall mannoprotein with an early stop codon. These loci were therefore suitable for neutral integration of reporters. Although HIS3::YFP was prepared, limitations in multicolor FACS prevented reliable separation of YFP from GFP. As a result, HIS3::YFP was excluded from downstream sorting and sequencing, though the construct remains available for future studies.

In this chapter, we describe the design and analysis of promoter mutagenesis libraries in CAN1::sfGFP and CCW22::DsRed. By combining fluorescence measurements, FACS, and sequencing, we identified promoter regions where mutations reproducibly reduce expression, providing insight into the sequence features that shape transcriptional sensitivity in yeast.

2.3 Results

To determine how specific mutations in promoter elements affect gene expression, we constructed reporter strains of *S. cerevisiae* containing CAN1::sfGFP, HIS3::YFP, and CCW22::DsRed integrations. Each gene was selected because it is non-essential under laboratory conditions: CAN1

encodes an arginine permease dispensable in rich media, HIS3 encodes an enzyme in histidine biosynthesis that is not required with histidine supplementation, and CCW22 encodes a non-essential cell wall mannoprotein that contains an early stop codon, making it dispensable for growth. These loci therefore served as neutral genomic sites for integrating fluorescent reporters without impairing host fitness.

We next designed promoter mutagenesis libraries targeting a 120 base pair (bp) window spanning the TATA box, PIC, scanning region, and TSS (Fig. 1a). Using a small library of mutations, we generated increased pockets of A/T rich and G/C rich regions, which generally follows the known trends of the promoter (Fig. 1b). Through the use of Inscripta's Black Box, the mutation library of roughly 600 mutations was made. The Black Box used plasmid containing the gRNA, the intended mutation, and a tracker sequence to make the mutations within the host strain as well as containing a trackable plasmid for sequencing (Fig. 1c).

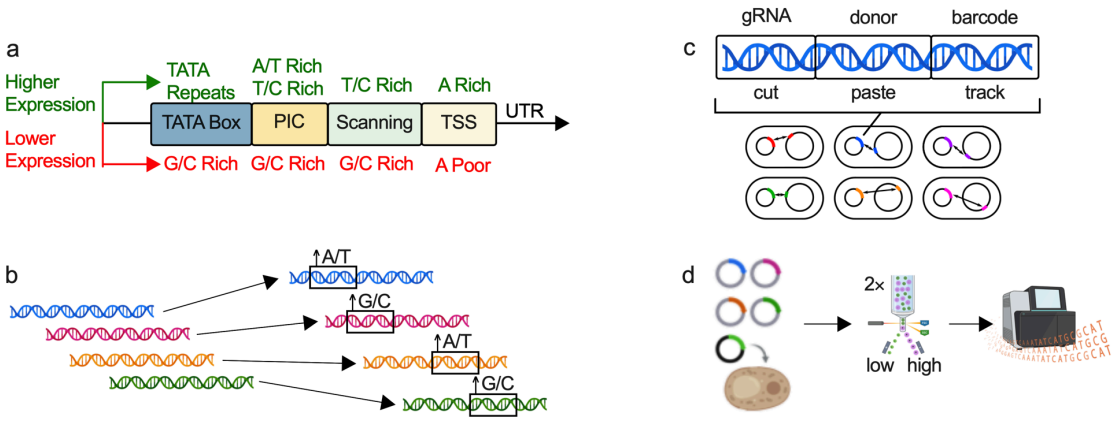

Fig. 1. Project design and work flow (a) General known promoter structure and mutations known to alter gene production. (b) The mutation strategy implimented was to use A/T to G/C and G/C to A/T random mutations. (c) Preperation of the library, done by Inscripta, for the mutation library. (d) Work flow of the project included mutation integration, double FACS sorting, and sequencing.

Prior to sorting, we confirmed that all reporter strains expressed detectable levels of their fluorescent proteins, with distributions clearly distinguishable from the parental BY4741 strain (Fig. 2a). Establishing these baselines was essential for defining gating strategies in fluorescence-activated cell sorting (FACS).

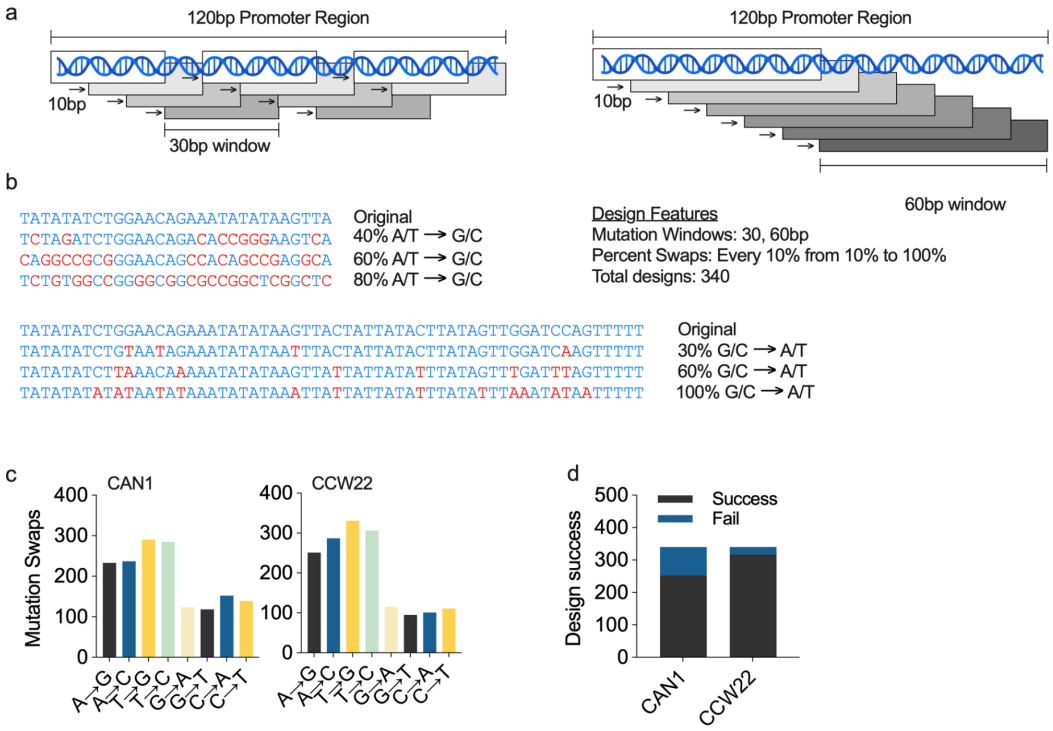

Fig. 2. Mutation library design. (a) The promoter was split up in two different ways to allow for a full analysis, while also being able to focus on specific motifs. Windows of 30bp and 60bp were used to shorten the mutation area, while these windows were sgifted over by 10bp in order to scan the entire promoter. (b) Example of the mutation design, where within the original sequence either A/T's or G/C's were randomly picked and changed, based on the percentage of those nucleotide. (c) The total number of mutations for each nucleotide change, showing that the mutations were not biased towards one mutation. (d) The designable number of mutations that were possible to make through Inscripta's Black Box.

In order to obtain true hits of both high and low secretors, a two-round fluorescence-activated cell sorting (FACS) was decided. In the first round, cells clustering near the control mean were removed from the population. In the second round, the most extreme 2% of cells were further enriched, capturing both high- and low-fluorescence populations (Fig. 1d). Sorted cells were regrown and re-analyzed, confirming that the distributions of the sorted

libraries diverged from controls and that the low-fluorescence pool contained the largest fraction of variants with altered expression.

To systematically probe this region, mutations were introduced using 30bp and 60bp sliding windows shifted in 10bp increments, ensuring that each nucleotide was covered multiple times across different sequence contexts. Mutations consisted of A/T ↔ G/C substitutions, applied at varying frequencies ranging from 10% to 100%. This design produced a diverse set of promoter variants for functional screening (Fig. 2b). Each nucleotide was examined to determine that whenever it was mutated, it was unbiased toward one specific nucleotide (Fig. 2c). For each A or T we would assume it would equally randomly mutate to G or C, while the inverse is also true. Each promoter region was composed of roughly two-thirds A/T and one-third G/C, accounting for the increased amount of mutations that occurred for A's and T's. Some of the mutations designed were not necessarily able to be made, with mutations that had 100% A/T or G/C commonly not being a feasible design to actually make. For CAN1::GFP only 253 of the 340 were possible to make, while for CCW22::DsRed only 317 of the possible 340 were possible to make (Fig. 2d). After the double FACS sort, samples from each pool and a control were run again on the flow cytometer to check the fluorescence spread (Fig. 3a).

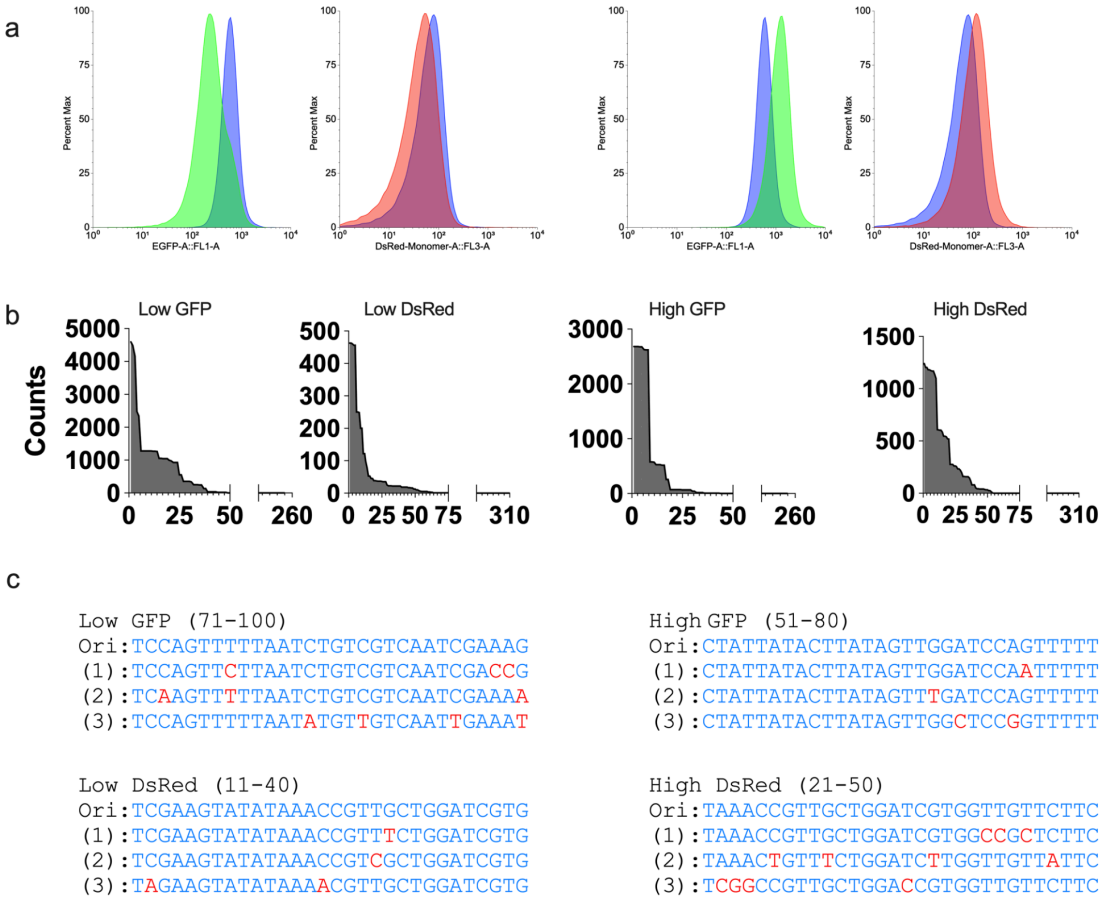

Fig. 3 FACS and NGS results. (a) FACS pools were rerun through flow cytometry to examine the spread of the double sorted pools. (b) Number of sequence hits and amount of times seen within the sequencing results. (c) Example sequences from each pool as compared to the original.

Samples were also prepared and sent out for next generation sequencing (NGS), to determine what the winning sequences are. Fig. 3b indicates the number of counts each individual mutation appeared post sequencing. Provided samples of individual mutations found within for each sorted pool as compared to the original sequence (Fig. 3c).

After sequencing, counts were pooled into bins from, where 1-100 equaled 1, 101-200 equaled 2, and so on. From there, sequences were separated into 10bp chunks so that specific motifs could be analysed easily. Frequency of mutations was established by taking the total number of mutated nucleotides per position and dividing by the total number of bin counts within the motif. From this, any overlapping nucleotides from both the high and low pools were subtracted to focus on the altering nucleotides (Fig. 4).

To further contextualize the effects of promoter mutagenesis, sequence variants were grouped according to three functionally distinct regions of the yeast core promoter: the TATA box, the pre-initiation complex (PIC) region, and the downstream scanning region (Fig. 4). The TATA box is an A/T-rich element located upstream of the transcription start site that serves as the primary binding site for the TATA-binding protein (TBP) and plays a central role in positioning the transcriptional machinery. Immediately downstream of the TATA box lies the PIC region, which encompasses binding sites for additional general transcription factors involved in assembling the RNA polymerase II pre-initiation complex. Further downstream, the scanning region spans the sequence traversed by the transcriptional machinery during start site selection and initiation. For each of these regions, mutations falling within the corresponding 10 bp windows were aggregated and analyzed separately to assess how perturbations to distinct promoter components influence transcriptional output.

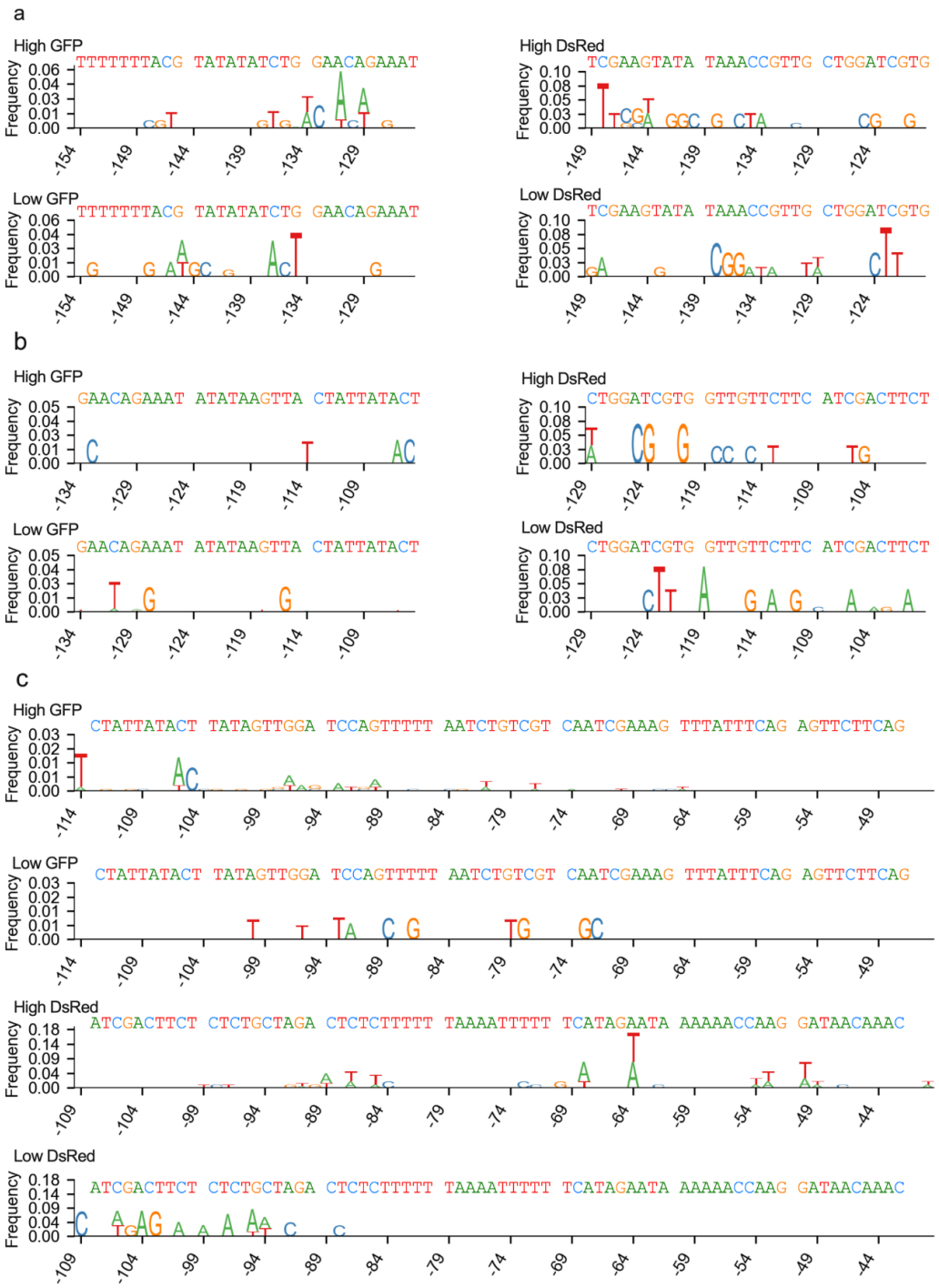

Fig. 4 Frequency plots of the TATA Box. Frequency calcultations calculated based on total number of counts within the motif and removal of overlapping common mutations between high and low pools. (a) Frequency of mutated nucleotides within the TATA Box. (b) Frequency of mutated nucleotides within the PIC. (c) Frequency of mutated nucleotides within the scanning region.

Taken together, these analyses demonstrate that promoter mutagenesis produces highly asymmetric effects on gene expression. While many sequence variants were phenotypically neutral or exhibited modest changes in fluorescence, GC-enriching substitutions within core promoter motifs—particularly the TATA box—were consistently associated with reduced expression. This effect was evident across multiple analytical scales, including individual nucleotide substitutions, 10 bp sliding windows, and aggregated promoter regions. In contrast, mutations that enhanced expression were rare, exhibited weaker reproducibility, and lacked consistent sequence features. These results indicate that promoter architecture is robust to a wide range of perturbations but contains specific sequence vulnerabilities that bias transcriptional output downward when disrupted. Collectively, this establishes GC enrichment in core promoter motifs as a predictable mechanism for tuning gene expression and motivates further exploration of alternative strategies for identifying expression-enhancing variants.

2.4 Discussion

This study sought to determine whether targeted promoter mutagenesis could be used to reproducibly reduce gene expression in *Saccharomyces cerevisiae* without abolishing transcription entirely. By combining sliding-window mutagenesis, fluorescence-activated cell sorting, and high-throughput

sequencing, we identified promoter regions and sequence features that are particularly sensitive to perturbation. Across two independent reporter systems, our results demonstrate that GC-enriching substitutions in core promoter motifs—most notably the TATA box and, to a lesser extent, the pre-initiation complex (PIC) region—consistently bias variants toward lower expression.

Across both CAN1::sfGFP and CCW22::DsRed reporters, mutations enriched in low-fluorescence populations were dominated by A/T → G/C substitutions within the core promoter. This trend was evident when mutations were analyzed at the level of individual nucleotides, 10 bp bins, and aggregated promoter regions (Fig. 4). In contrast, high-fluorescence populations showed comparatively weak and inconsistent enrichment for specific substitutions. This asymmetry indicates that promoter architecture is far more permissive to mutations that reduce transcription than to those that enhance it, consistent with prior saturation mutagenesis studies of yeast promoters [13, 14].

Importantly, these effects were not driven by overall mutation burden alone. Rather, mutations localized to specific motifs produced disproportionate effects on expression, underscoring the importance of promoter context rather than simple sequence disruption.

Among all promoter elements examined, the TATA box emerged as the most mutation-sensitive region. GC substitutions within the TATA box were strongly enriched in low-fluorescence pools and depleted from high-fluorescence

pools (Fig. 4a). Analysis of GC fraction within the 30 bp TATA region further revealed a striking pattern: low-expression variants tolerated a broad range of GC content, whereas high-expression variants were tightly constrained to a narrow GC range.

This distributional asymmetry suggests that there are many ways to disrupt TATA box function but relatively few ways to modify it without impairing transcription. This observation is consistent with the known requirement for A/T-rich sequences to support stable binding of the TATA-binding protein (TBP) [8, 10]. Increasing GC content likely alters DNA flexibility and minor groove geometry, weakening TBP–DNA interactions and reducing the efficiency of transcription initiation.

Notably, the effect was not driven by a single catastrophic mutation but rather by the cumulative impact of GC enrichment across the motif. This supports a model in which the TATA box function depends on distributed sequence properties rather than a single invariant consensus.

Mutations within the PIC region also biased variants toward lower expression, though the magnitude of this effect was smaller and more heterogeneous than that observed for the TATA box (Fig. 4b). This intermediate sensitivity is consistent with the role of the PIC as a platform for recruiting general transcription factors, where multiple protein–DNA and protein–protein interactions may provide partial redundancy [9, 10].

In contrast, mutations within the scanning region and transcription start site (TSS) had relatively weak or inconsistent effects on fluorescence (Fig. 4c). These regions appear to tolerate substantial sequence variation without strongly altering transcriptional output. Such robustness likely reflects flexibility in start site selection and promoter clearance mechanisms, as previously suggested by genome-wide analyses of yeast promoters [11].

Together, these results support a hierarchical model of promoter sensitivity in which the TATA box represents the primary vulnerability to sequence perturbation, followed by the PIC, while downstream regions are comparatively resilient.

From an applied perspective, these findings establish GC substitution within core promoter motifs as a predictable strategy for tuning gene expression downward without complete silencing. This approach is particularly relevant for metabolic engineering, where partial knockdown of endogenous genes can redirect cellular resources toward heterologous protein production or pathway flux optimization.

At the same time, this study has several limitations. The mutation libraries were restricted to two substitution classes and a 120 bp promoter window, and the Inscripta Black Box platform used for library construction is no longer available. Furthermore, the scarcity of reproducible high-expression variants highlights the inherent difficulty of enhancing transcription through random mutagenesis alone.

These limitations motivated the development of CREATE-based promoter libraries described in Chapter 3, which expand both mutational diversity and promoter coverage.

2.5 Materials & Methods

2.5.1 E. coli Heat Shock Transformation

For plasmid propagation, chemically competent NEB 10β (E. coli) cells were used. Aliquots of 50 µL were thawed on ice in sterile 1.5 mL microcentrifuge tubes. Plasmid DNA (0.2–1.0 µg) was added to each aliquot, and the mixture was incubated on ice for 5 minutes. The tubes were transferred to a 42 °C heat block for 45 seconds, then immediately returned to ice for 5 minutes. After heat shock, 950 µL of LB medium was added to each tube, and cells were recovered for 1 hour at 37 °C in a shaking incubator. Subsequently, 100 µL of the culture was plated onto LB agar plates containing the appropriate selective antibiotic, and plates were incubated at 37 °C for 24 hours. The remaining culture was supplemented with antibiotics and incubated at 37 °C in a shaking incubator for plasmid preparation.

2.5.2 Plasmid Mini Prep

Plasmid DNA was purified using the ZymoPURE Plasmid Miniprep Kit (Zymo Research), following the manufacturer's protocol with slight modifications. Briefly, 5 mL overnight LB cultures containing the appropriate antibiotic were

prepared from single colonies. From each culture, 800 µL was stored as glycerol stocks, and the remaining cells were harvested by centrifugation at 12,000 × g for 5 minutes. The cell pellet was resuspended in 200 µL P1 buffer, followed by 200 µL P2 buffer with gentle inversion. Then, 400 µL P3 buffer was added, mixed thoroughly, and centrifuged at 12,000 × g for 5 minutes. The supernatant was transferred to a spin column, washed with Endo wash buffer (200 µL) and plasmid wash buffer (400 µL), and centrifuged after each wash. DNA was eluted in 50 µL of nuclease-free water.

2.5.3 Sequencing

Plasmid constructs were submitted to Genewiz (Azenta Life Sciences) for Sanger sequencing to confirm integrity prior to use in experiments. Promoter mutagenesis libraries and pooled samples were sequenced at the UCR Institute for Integrative Genome Biology (IIGB) Genomics Core. Illumina sequencing was performed using [e.g., MiSeq or NextSeq] with 1 × 75 bp single-end reads. Barcoded adapters were used to separate samples prior to pooling. Custom in house scripts were used for demultiplexing, read trimming, and mapping to reference sequences.

2.5.4 S. cerevisiae Transformation and Genomic Integration

Saccharomyces cerevisiae transformations were performed using the PEG 3350/LiAc/salmon sperm carrier DNA method [15]. A single colony was inoculated into 5 mL of 2× YPD and grown overnight at 30 °C. The overnight

culture was diluted into 50 mL of fresh 2× YPD to a final density of ~2.5 × 10^8 cells and grown to mid-log phase (OD_{600} ~2.0, ~2 × 10^7 cells/mL). Cells were harvested by centrifugation at 3,000 × g for 5 minutes, washed twice with sterile water, and resuspended in 1 mL water. For each transformation, 100 µL of cell suspension was mixed with up to 3 µg of plasmid DNA, salmon sperm carrier DNA, and PEG 3350 solution. Tubes were incubated at 42 °C for 40–60 minutes for heat shock. After incubation, cells were pelleted at 13,000 × g for 30 seconds, resuspended in 1 mL sterile water, and plated on selective media. Plates were incubated at 30 °C for 1–3 days, depending on transformation type.

2.5.5 Promoter Mutagenesis Library Design

Random mutations were introduced into a 120 bp promoter region encompassing the TATA box, Pre-Initiation Complex (PIC), scanning region, and transcription start site (TSS). Mutagenesis was performed using sliding windows of 30 bp and 60 bp, shifted in 10 bp increments, to ensure each base was covered multiple times across different sequence contexts. Two substitution classes were applied: (1) A/T → G/C and (2) G/C → A/T. Within each window, all eligible bases were identified, and a subset was randomly selected for substitution according to the chosen mutation frequency. Frequencies ranged from 10% to 100% of available sites. At lower percentages, only a fraction of bases were mutated, whereas at 100%, all eligible sites were substituted.

Importantly, only nucleotides belonging to the targeted class were mutated, even at the maximum percentage.

2.5.6 Flow Cytometry and FACS Sorting

Library glycerol stocks were inoculated into YPD and grown in shaking culture at 30 °C for 4 hours. Cells were harvested, washed twice with PBS at 5,000 × g for 1 minute, and resuspended in sterile ultrapure water. Flow cytometry was performed on a Sony SH800 cell sorter to establish fluorescence distributions and define gating strategies. Fluorescence settings were as follows: sfGFP (ex: 488 nm, em: 510/20 nm), YFP (ex: 488 nm, em: 535/30 nm), and DsRed (ex: 561 nm, em: 617/30 nm).

For FACS sorting, cells were prepared as above and sorted on the Sony SH800. Populations were collected into three groups: (1) cells clustering near the control mean (“baseline”), (2) cells below baseline fluorescence, and (3) cells above baseline fluorescence. Sorted cells were regrown in selective medium for [2 hours] prior to further analysis.

2.5.7 NGS and Processing

DNA was extracted from sorted populations, and promoter regions were amplified for sequencing. Libraries were sequenced at the UCR Genomics Core using [Illumina platform, read length]. Mutation frequencies were determined by mapping reads to the reference promoter sequence. Custom Python scripts

were used to calculate enrichment between high- and low-fluorescence populations.

2.6 References

[1] Johnson, I. S. (1983). Human insulin from recombinant DNA technology. Science, 219(4585), 632–637. https://doi.org/10.1126/science.6337396

[2] Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances, 27(3), 297–306. https://doi.org/10.1016/j.biotechadv.2009.01.008

[3] Nielsen, J. (2013). Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. Bioengineered, 4(4), 207–211. https://doi.org/10.4161/bioe.22856

[4] Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213), 1258096. https://doi.org/10.1126/science.1258096

[5] Jakočiūnas, T., Jensen, M. K., & Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of yeast genomes. Metabolic Engineering, 34, 44–59. https://doi.org/10.1016/j.ymben.2015.12.004

[6] Nevoigt, E. (2008). Progress in metabolic engineering of Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews, 72(3), 379–412. https://doi.org/10.1128/MMBR.00025-07

[7] Porro, D., Sauer, M., Branduardi, P., & Mattanovich, D. (2011). Recombinant protein production in yeasts. Molecular Biotechnology, 48(2), 191–203. https://doi.org/10.1007/s12033-010-9366-y

[8] Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box–containing genes. Cell, 116(5), 699–709. https://doi.org/10.1016/S0092-8674(04)00205-3

[9] Hahn, S., & Young, E. T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: Transcription factor regulation and function, mechanisms of initiation, and promoter architecture. Cold Spring Harbor Perspectives in Biology, 3(12), a005132. https://doi.org/10.1101/cshperspect.a005132

[10] Rhee, H. S., & Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature, 483(7389), 295–301. https://doi.org/10.1038/nature10799

[11] Lubliner, S., Keren, L., & Segal, E. (2013). Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Genome Research, 23(5), 810–820. https://doi.org/10.1101/gr.150656.112
[12] Mogno, I., Kwasnieski, J. C., & Cohen, B. A. (2013). Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. Nucleic Acids Research, 41(9), 4845–4857. https://doi.org/10.1093/nar/gkt179

[13] Patwardhan, R. P., Lee, C., Litvin, O., Young, D. L., Pe'er, D., & Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nature Biotechnology, 27(12), 1173–1175. https://doi.org/10.1038/nbt.1589

[14] Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., & Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nature Biotechnology, 30(6), 521–530. https://doi.org/10.1038/nbt.2205

Chapter 3: Expanded Promoter Mutagenesis with CREATE Libraries in *S. cerevisiae*

3.1 Abstract

Promoter mutagenesis provides a strategy to dissect the sequence features that regulate transcription in yeast. In Chapter 2, we demonstrated that A/T → G/C substitutions in the TATA box and Pre-Initiation Complex (PIC) reduce gene expression, but our approach was constrained by the Inscripta platform, which permitted only two mutation classes within a 120 bp window. Here, we developed a CREATE-based mutagenesis strategy that expanded mutational diversity to 31 substitution types and extended coverage to 200 bp upstream of the transcriptional start site (TSS). Independent sfGFP reporter strains were constructed for CAN1, HIS3, and CCW22, avoiding the spectral overlap that limited analysis in Chapter 2. Flow cytometry and fluorescence activated cell sorting (FACS) revealed reproducible enrichment of loss-of-function variants, with the TATA box and PIC consistently the most mutation-sensitive regions. Interestingly, HIS3 displayed additional sensitivity in the scanning region, suggesting promoter-specific differences in mutational tolerance. These findings confirm general principles of promoter architecture while uncovering gene-specific effects, and they establish CREATE mutagenesis as a scalable platform for promoter engineering in yeast.

3.2 Introduction

In Chapter 2, targeted promoter mutagenesis was used to identify sequence changes that reproducibly reduce gene expression in *Saccharomyces cerevisiae*. By introducing A/T → G/C and G/C → A/T substitutions across a 120 bp region spanning the TATA box, pre-initiation complex (PIC), scanning region, and transcription start site (TSS), we showed that mutations within core promoter elements—particularly the TATA box and PIC—were strongly enriched in low-expression populations. These findings are consistent with prior studies demonstrating the sensitivity of core promoter architecture to sequence composition and GC content [3,4,5,6,7].

While this approach established general trends in promoter sensitivity, it was constrained by technical limitations of the Inscripta Black Box platform. Mutagenesis was restricted to two substitution classes within a fixed 120 bp window, limiting both mutational diversity and spatial coverage of promoter regions. As a result, it remained unclear whether additional base substitutions would produce similar effects or whether regions further upstream of the TATA box contribute to transcriptional regulation. In addition, multicolor fluorescence overlap prevented parallel analysis of all reporter constructs.

To overcome these limitations, we implemented a CREATE (CRISPR-enabled trackable genome engineering)–based promoter mutagenesis

strategy. CREATE was originally developed to enable high-resolution, trackable genome engineering through guide-encoded mutations and homology-directed repair [1,2]. This approach allows systematic introduction of diverse base substitutions across extended genomic regions. Using CREATE, we expanded promoter mutagenesis to include 31 distinct substitution types and extended coverage to 200 bp upstream of the transcription start site.

In parallel, independent sfGFP reporter strains were constructed for the CAN1, HIS3, and CCW22 promoters to allow promoter activity to be evaluated individually without spectral overlap. CAN1 encodes an arginine permease dispensable in rich media, HIS3 encodes an enzyme in histidine biosynthesis that is not required with supplementation, and CCW22 encodes a non-essential cell wall mannoprotein. These loci therefore serve as neutral genomic sites for reporter integration, as commonly employed in yeast promoter studies [8,9,10].

In this chapter, we describe the design, construction, and initial analysis of CREATE-based promoter mutagenesis libraries in *S. cerevisiae*. By combining expanded promoter mutagenesis with fluorescence-activated cell sorting (FACS) and next-generation sequencing, we examine how increased mutation diversity and promoter coverage influence enrichment patterns across promoter regions. Results are presented descriptively through Figure 3, establishing a framework for downstream quantitative and mechanistic analysis.

3.3 Results

Libraries were constructed for the *HIS3*, *CAN1*, and *CCW22* promoters. Each design included primers for amplification, homology arms flanking the mutation window, a guide sequence selected for the target region, and a Cas9 scaffold. The libraries were cloned into Cas9-containing plasmids and transformed into *S. cerevisiae*. Following library integration, cells were screened by flow cytometry and fluorescence-activated cell sorting (FACS) to enrich high and low-fluorescence populations. Sorted cells were regrown, and promoter variants were identified by next-generation sequencing (NGS) (Fig. 1).

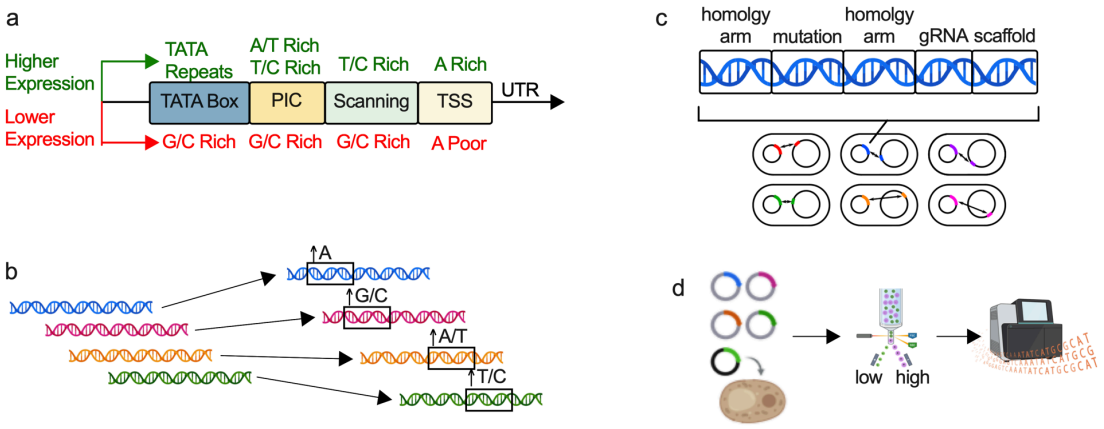

Fig. 1. Project design and work flow (a) General known promoter structure and mutations known to alter gene production. (b) The mutation strategy implimented expands the amount of mutation permutations. (c) Preperation of the library done using the CREATE-style mutation integration for the mutation library. (d) Work flow of the project included mutation integration, single FACS sorting, and sequencing.

Sequencing of high- and low-fluorescence pools revealed reproducible enrichment of promoter variants. In *CAN1*, mutations clustered around the TATA

box and PIC, while *CCW22* displayed additional sensitivity in the scanning region. *HIS3* showed strong enrichment in the PIC and scanning region closer to the transcriptional start site. Across all three promoters, A/T → G/C substitutions were enriched in low-fluorescence populations, indicating that increased GC content reduces promoter activity (Fig. 2). High- and low-fluorescence pools were compared across the promoter regions of *CAN1*, *CCW22*, and *HIS3*. Peaks in significance aligned with regions of mutation enrichment, confirming that the TATA box and PIC were consistently the most sensitive promoter elements. *HIS3* showed additional enrichment extending into the scanning region, highlighting promoter-specific differences in sensitivity (Fig. 3).

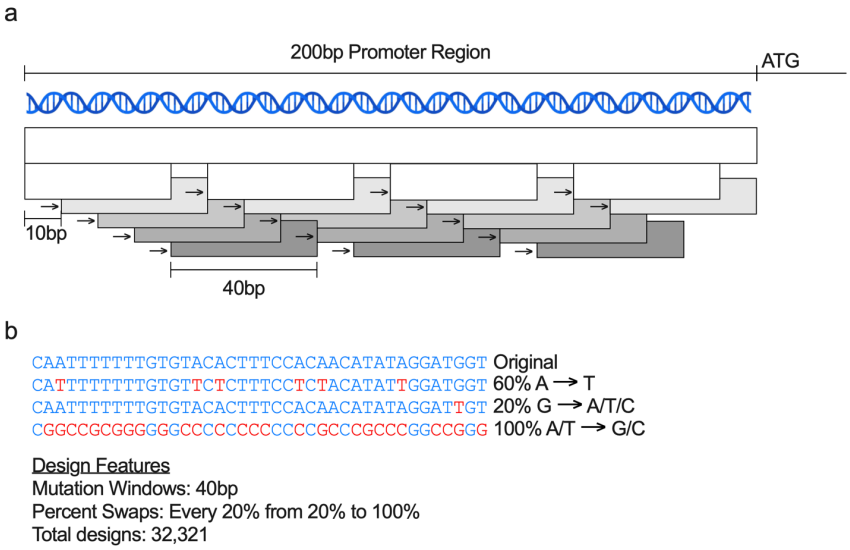

Fig. 2. Mutation library design. (a) The promoter was split up in two different ways to allow for a full analysis, while also being able to focus on specific motifs. Windows of 40bp were and were shifted over by 10bp in order to scan the entire promoter. (b) Example of some of the varying designs.

To move beyond the limitations of Chapter 2, we applied CREATE mutagenesis to expand both the diversity and coverage of promoter variants in *S. cerevisiae*. This strategy generated libraries with 31 substitution types across 200 bp windows upstream of the TSS, representing a substantial increase compared to the two mutation classes and 120 bp coverage previously available with Inscripta. In total, the CREATE design yielded 46,981 unique promoter variants, with each mutation type repeated five times across the target windows to improve sampling. Although some windows were inaccessible due to guide sequence constraints, the libraries provided broad and systematic coverage of promoter elements.

To avoid fluorescence overlap that complicated analysis in Chapter 2, independent sfGFP reporter strains were constructed for *CAN1*, *HIS3*, and *CCW22*. This allowed promoter activity to be evaluated separately in each background without spectral interference.

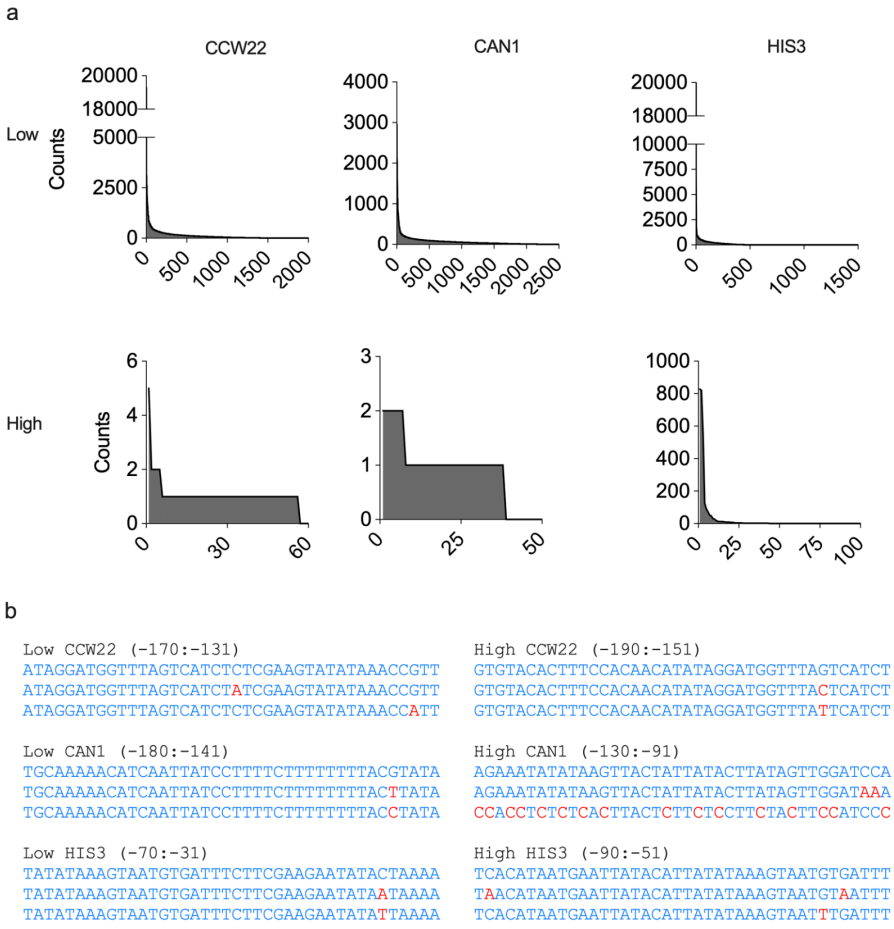

Fig. 3 NGS results. (a) Number of sequence hits and amount of times seen within the sequencing results. (b) Example sequences from each pool as compared to the original.

Control strains were first analyzed to establish baseline fluorescence distributions, which served as gating references for sorting. The CREATE libraries were then subjected to two rounds of FACS, isolating high- and low fluorescence populations relative to controls. As with the Inscripta libraries, the low-fluorescence pools showed the strongest divergence from baseline, suggesting that disruptive mutations were more common than activating ones.

Sorted populations were regrown and re-analyzed by flow cytometry, confirming the stability of the high- and low-fluorescence distributions. DNA from these populations was sequenced to map mutation frequencies across promoter regions.

Sequencing revealed reproducible patterns of enrichment in all three promoters. In *CAN1*, low-fluorescence variants clustered around the TATA box (– 180 to –160) and extended into the PIC region (–150 to –120). A/T → G/C substitutions were especially enriched in these windows. In *CCW22*, enrichment also concentrated at the TATA box and PIC, but additional sensitivity was observed in the scanning region (–100 to –70). In *HIS3*, enrichment shifted further downstream, spanning the PIC and scanning regions (–90 to –50). This distribution suggested promoter-specific sensitivity to sequence variation in regions beyond the TATA box (Fig. 2). The most significant peaks aligned with regions of high mutation enrichment, confirming that the TATA box and PIC were consistently the most mutation-sensitive elements across promoters. In *HIS3*, statistical enrichment extended toward the TSS, further highlighting promoter-specific differences (Fig. 3).

3.4 Discussion

The CREATE-based mutagenesis libraries allowed us to move beyond the limitations of Chapter 2 and examine promoter sensitivity with much greater depth. By expanding the mutational diversity to thirty-one substitution types and extending the target window to 200 bp upstream of the TSS, we were able to capture a more complete picture of promoter architecture in *S. cerevisiae*.

The results from *CAN1*, *CCW22*, and *HIS3* confirm the central importance of the TATA box and PIC, while also revealing promoter-specific nuances that were not visible in the smaller Inscripta libraries. Across all three promoters, A/T → G/C substitutions were consistently associated with reduced expression, underscoring the destabilizing effect of increased GC content within promoter elements. This trend reinforces the observations made in Chapter 2 and is supported by earlier studies showing that higher GC content decreases transcriptional activity in yeast promoters [8, 9].

The TATA box and PIC emerged as the most mutation-sensitive regions, as expected from their roles in transcriptional initiation. Sequence changes in the TATA box are known to disrupt TBP binding [5], reducing the ability of the transcriptional machinery to assemble. Similarly, substitutions within the PIC region likely interfere with the recruitment of general transcription factors [6, 7], diminishing promoter strength.

At the same time, the CREATE libraries revealed promoter-specific differences in scanning region sensitivity. In *CAN1* and *CCW22*, mutations upstream of the TSS had relatively minor effects, suggesting these promoters can tolerate sequence variation without major disruption. By contrast, *HIS3* showed clear enrichment of low-fluorescence variants in the scanning region, extending toward the TSS. This difference suggests that the functional importance of scanning sequences may vary depending on promoter context.

Another key observation is the imbalance between loss-of-function and gain-of-function variants. Consistent with the Inscripta libraries in Chapter 2, down-regulating mutations were far more common than those that enhanced expression. This asymmetry reflects the general principle that it is easier to disrupt promoter function than to strengthen it, a conclusion echoed in large scale promoter saturation mutagenesis studies [3, 4].

Together, these results highlight both the shared features of promoter sensitivity—the consistent impact of GC substitutions in the TATA box and PIC— and the gene-specific differences that emerge when broader regions are interrogated. The CREATE mutagenesis strategy therefore not only confirmed the trends observed in Chapter 2 but also provided new insights into promoter architecture that were inaccessible with smaller libraries allowing for expanded use for future studies.

3.5 Materials and Methods

3.5.1 Plasmid Mini Prep

Plasmid DNA was purified using the ZymoPURE Plasmid Miniprep Kit (Zymo Research) according to the manufacturer's instructions with slight modifications. Briefly, 5 mL overnight LB cultures containing the appropriate antibiotic were prepared from single colonies. From each culture, 800 µL was stored as glycerol stocks, and the remaining cells were harvested by centrifugation at 12,000 × g for 5 minutes. The pellet was resuspended in 200 µL P1 buffer, followed by 200 µL P2 buffer with gentle inversion. Then, 400 µL P3 buffer was added, mixed thoroughly, and centrifuged at 12,000 × g for 5 minutes. The supernatant was transferred to a spin column, washed with Endo wash buffer (200 µL) and plasmid wash buffer (400 µL), and centrifuged after each wash. DNA was eluted in 50 µL of nuclease-free water.

3.5.2 Sequencing

Plasmid constructs were submitted to Genewiz (Azenta Life Sciences) for Sanger sequencing to confirm integrity prior to use in experiments. Promoter mutagenesis libraries and pooled samples were sequenced at the UCR Institute for Integrative Genome Biology (IIGB) Genomics Core. Illumina sequencing was performed using [e.g., MiSeq or NextSeq] with 1 × 75 bp single-end reads. Barcoded adapters were used to separate samples prior to pooling. Custom in-

house scripts were used for demultiplexing, read trimming, and mapping to reference sequences.

3.5.3 *S. cerevisiae* Transformation and Integrations

S. cerevisiae transformations were performed using the PEG 3350/LiAc/salmon sperm DNA method [8]. A single colony was inoculated into 5 mL of 2× YPD and grown overnight at 30 °C. The overnight culture was diluted into 50 mL of fresh 2× YPD to a final density of ~2.5 × 10^8 cells and grown to mid-log phase (OD_{600} ~2.0, ~2 × 10^7 cells/mL). Cells were harvested by centrifugation at 3,000 × g for 5 minutes, washed twice with sterile water, and resuspended in 1 mL water.

For each transformation, 100 µL of cells was mixed with up to 3 µg of plasmid DNA, salmon sperm carrier DNA, and PEG 3350. Tubes were incubated at 42 °C for 40–60 minutes for heat shock. After incubation, cells were pelleted at 13,000 × g for 30 seconds, resuspended in 1 mL sterile water, and plated on selective media. Plates were incubated at 30 °C for 1–3 days, depending on transformation type.

3.5.4 Flow Cytometry and FACS Sorting

Library glycerol stocks were inoculated into YPD and grown in shaking culture at 30 °C for 4 hours. Cells were harvested, washed twice with PBS at 5,000 × g for 1 minute, and resuspended in sterile ultrapure water.

Flow cytometry was performed using a Sony SH800 cell sorter to define fluorescence distributions and establish gating strategies. Fluorescence settings were: sfGFP (excitation 488 nm, emission 510/20 nm), YFP (excitation 488 nm, emission 535/30 nm), and DsRed (excitation 561 nm, emission 617/30 nm).

For FACS sorting, cells were prepared as above and sorted on the same instrument. Gates were set to collect three groups: (1) cells clustering near the control mean (“baseline”), (2) cells below the baseline (low fluorescence), and (3) cells above the baseline (high fluorescence). Sorted populations were regrown in selective medium for [X hours] before downstream analysis.

3.7 References

[1] Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., ... & Gill, R. T. (2017). Genome-wide mapping of mutations at single nucleotide resolution for protein, metabolic, and genome engineering in Escherichia coli. Nature Biotechnology, 35(1), 48–55. https://doi.org/10.1038/nbt.3718

[2] Liang, L., Liu, R., Garst, A. D., Lee, T., Sánchez i Nogué, V., Beckham, G. T., & Gill, R. T. (2017). CRISPR EnAbled Trackable genome Engineering (CREATE) for isopropanol production in Escherichia coli. Metabolic Engineering, 41, 1–10. https://doi.org/10.1016/j.ymben.2017.02.008

[3] Patwardhan, R. P., Lee, C., Litvin, O., Young, D. L., Pe'er, D., & Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nature Biotechnology, 27(12), 1173–1175. https://doi.org/10.1038/nbt.1589

[4] Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., & Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nature Biotechnology, 30(6), 521–530. https://doi.org/10.1038/nbt.2205

[5] Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box–containing genes. Cell, 116(5), 699–709. https://doi.org/10.1016/S0092-8674(04)00205-3

[6] Hahn, S., & Young, E. T. (2011). Transcriptional regulation in *Saccharomyces cerevisiae*: Transcription factor regulation and function, mechanisms of initiation, and promoter architecture. Cold Spring Harbor Perspectives in Biology, 3(12), a005132. https://doi.org/10.1101/cshperspect.a005132

[7] Rhee, H. S., & Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature, 483(7389), 295–301. https://doi.org/10.1038/nature10799

[8] Lubliner, S., Keren, L., & Segal, E. (2013). Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Genome Research, 23(5), 810–820. https://doi.org/10.1101/gr.150656.112

[9] Mogno, I., Kwasnieski, J. C., & Cohen, B. A. (2013). Massively parallel

synthetic promoter assays reveal the in vivo effects of binding site variants. Nucleic Acids Research, 41(9), 4845–4857. https://doi.org/10.1093/nar/gkt179

[10] Blazeck, J., Garg, R., Reed, B., & Alper, H. S. (2012). Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. ACS Synthetic Biology, 1(7), 326–331. https://doi.org/10.1021/sb300034z

Chapter 4: Development of a Mid-western Blot Assay for Protein Secretion in *K. marxianus*

4.1 Abstract

Kluyveromyces marxianus is a promising host for heterologous protein production due to its rapid growth, thermotolerance, and metabolic versatility, but tools for high-throughput secretion screening are limited. Here, we adapted a colony-based “mid-western blot” assay from *S. cerevisiae* to *K. marxianus*. The assay employs HiBiT/LgBiT complementation with NanoLuc luciferase to generate a chemiluminescent signal proportional to secreted protein, enabling secretion measurements across hundreds of colonies simultaneously. To evaluate assay performance, we tested twelve plasmid-based cassettes combining two promoters (GPD1, NC1), two signal peptides (OST ProMFα, synthetic), and three model proteins (CelA, HSA, PTE). Secretion outputs spanned nearly three orders of magnitude, with PTE constructs near background, HSA constructs at intermediate levels, and CelA constructs producing the highest titers. These results establish the mid-western blot as a robust and scalable assay for secretion screening in *K. marxianus*, providing a direct, quantitative alternative to growth-based methods and a platform for genome-wide library applications.

4.2 Introduction

Protein secretion is a central challenge in the development of yeast hosts for recombinant protein production. While *S. cerevisiae* has long been the dominant eukaryotic host, its limitations in protein folding, glycosylation, and secretion capacity have motivated increasing interest in non-conventional yeasts [4]. Among these, *K. marxianus* has emerged as a particularly attractive host due to its rapid growth rate, thermotolerance, and ability to metabolize a broad range of carbon sources, including both hexoses and pentoses [4, 7]. These traits make *K. marxianus* a promising chassis for industrial bioproduction, provided that robust tools are available to evaluate and optimize protein secretion.

To date, most high-throughput screens for improved phenotypes in *K. marxianus* have been based on growth or resistance markers. For example, Robertson et al. recently developed biosensor-driven CRISPR libraries in which production of specific metabolites was coupled to antibiotic resistance, enabling genome-wide identification of beneficial mutations [9]. While powerful, such approaches provide only indirect measurements of secretion and may miss mutations that enhance protein export without affecting growth [8]. Direct assays of secretion are therefore needed to complement these tools.

Our lab previously developed a colony-based secretion assay, termed the “mid-western blot,” in *S. cerevisiae*. This method relies on tagging proteins of

interest with the small HiBiT peptide, which forms an active NanoLuc luciferase when complemented with the LgBiT fragment in the presence of Furimazine [2]. Colonies are grown on nitrocellulose membranes, and secreted HiBiT-tagged proteins remain bound to the membrane. Upon addition of LgBiT and substrate, chemiluminescence provides a quantitative readout of secretion normalized by colony size. This approach enables the simultaneous screening of hundreds of colonies in a high-throughput and scalable manner.

To adapt the assay to *K. marxianus*, we constructed secretion cassettes containing different combinations of promoters, signal peptides, and model genes. The promoters GPD1 (strong) and NC1 (medium) were chosen based on previous characterization in *K. marxianus* [4, 5]. Signal peptides included OST ProMFα, previously used for secretion in *K. marxianus*, and a synthetic signal peptide designed in *S. cerevisiae*, which has been shown to enhance heterologous protein export [10, 3]. Three genes with distinct biological relevance were selected as secretion targets: CelA (a cellulase from *Clostridium thermocellum*), HSA (human serum albumin), and PTE (phosphotriesterase from *Pseudomonas diminuta*). Together, these combinations provided twelve unique cassettes for testing the capacity and sensitivity of the mid-western blot in *K. marxianus*.

The goal of this chapter is to demonstrate that the mid-western blot can be successfully ported to *K. marxianus*, to benchmark secretion across multiple

promoter–signal peptide–gene combinations, and to establish the assay as a platform for genome-wide screening of knockouts improving protein production.

4.3 Results

The goal of this work was to establish both an assay and a protein secretion cassette suitable for genome-wide knockout library screening. Our lab has already made an assay system, called a mid-western blot, in *S. cerevisiae* that is able to determine secretion rates of tagged genes based on the chemiluminescence and size of the colonies that remain on a nitrocellulose membrane. Porting this system into *K. marxianus* is simple, plasmids systems have already been utilized widely in our and other labs. The protein secretion cassettes were based on differing genes, promoters, and signal peptides previously tested in *K. marxianus* by both our lab and others. Promoters include NC1 and GPD1, secretion signals include synthetic signal peptide and OST ProMFα, and genes include HSA, CelA, and PTE. I utilized these parts in order to establish a system in *K. marxianus* for examining genome wide libraries for knockouts that are able to increase secretion rates.

The mid-western blot was designed for screening for changes in secretion rates of libraries using the Nano Lucifierase protein (NanoLuc). First, proteins of interest are tagged with a short 33 nucleotide sequence called HiBiT. These proteins are then secreted by the host organism, either through

integration or on a plasmid. The organism is then plated on the desired media, with a nitrocellulose membrane. The colonies that form secrete the POI-HiBiT and remain on the membrane. The membranes with the cells are removed and copied to another empty plate. The membranes are then cleaned with TBST, and then LgBiT is added with binds and completes NanoLuc. When Furimazine is added to the membrane, it reacts with NanoLuc and generates chemiluminescence. Imaging both the Chemiluminescence and Alexa 647 gives both the amount of secreted POI, based on intensity, and the size of each colony. With a single blot, we are able to look at hundreds of colonies at a time, focusing on the secretion of the POI.

The cassettes used are in plasmid form for these experiments, but I set up for integration and replacement of the HIS3 gene for use as a selectable marker. A low copy plasmid was used under the use of both the GPD1 and NC1 promoters, a medium and strong promoter respectively, which have been previously characterized in our lab previously. The signal peptides used were OST-ProMFα, previously used in our collaborators lab, and synthetic signal peptide (SS), which was used in the *S. cerevisiae* version of this assay. To expand on the types of genes for this assay, three genes with varying important products. CelA, an endoglucanase from *C. thermocellum*, is an example of an enzyme that degrades cellulose, HSA, human serum albumin, is an abundant protein in human blood serum, and PTE, phosphotriesterase from P. diminuta, is

an esterase that can degrade organophosphates. This leads to the 12 permutations of cassettes for determining the capabilities of the mid-western blot (FIG.1b).

To determine if this assay works in *K. marxianus* one of the cassettes was chosen, pKM014, to compare to a negative control, pIW1991, which was the backbone for these experiments. Cells were grown for 24 hours at 30°C, and cultures were diluted to an OD600 of ~700 cells each were plated together on the mid-western blot [Fig. 2]. Due to the close variance between the negative control and the positive control, I used Otsu thresholding of the two-population system to separate the low secreted positive hits with the negative control. This takes the two populations and finds the cutoff that maximizes separation between the populations while minimizing variance within each group. The samples had a difference from 13.58 ± 7.710 for the negative control, to 44.20 ± 16.45 for the positive control [1].

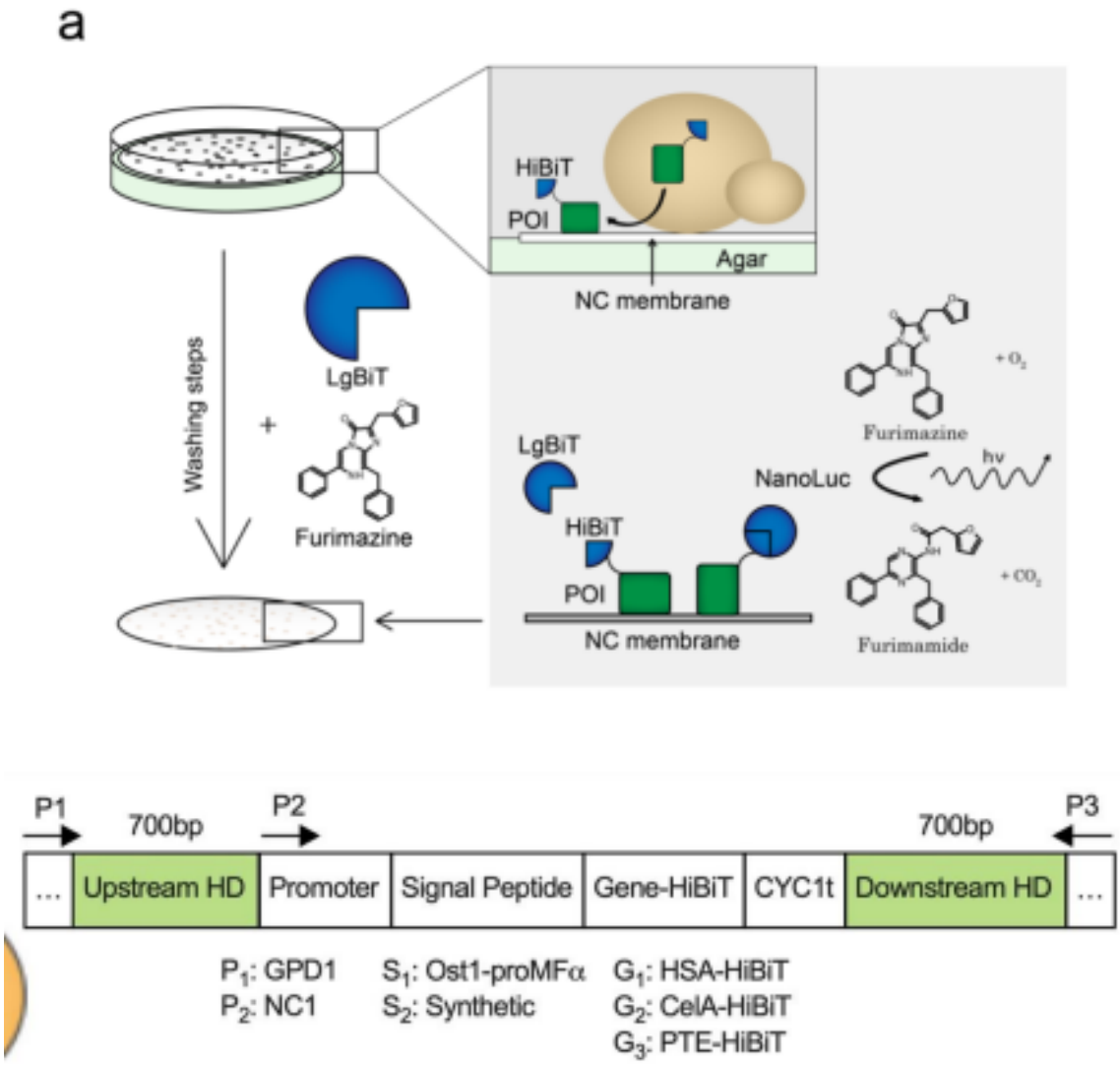

Fig. 1. Mid-western blot workflow and secretion cassette design. (a) Schematic of the mid-western blot assay. Colonies expressing HiBiT-tagged proteins are grown on nitrocellulose membranes, and secreted proteins remain bound to the surface. After washing, LgBiT and Furimazine are added, generating a chemiluminescent signal proportional to secretion, which is normalized by colony size. (b) Design of the 12 secretion cassettes. Two promoters (GPD1, NC1), two signal peptides (OST-ProMFα, synthetic SP), and three model genes (CelA, HSA, PTE) were combined in all permutations to evaluate secretion capacity.

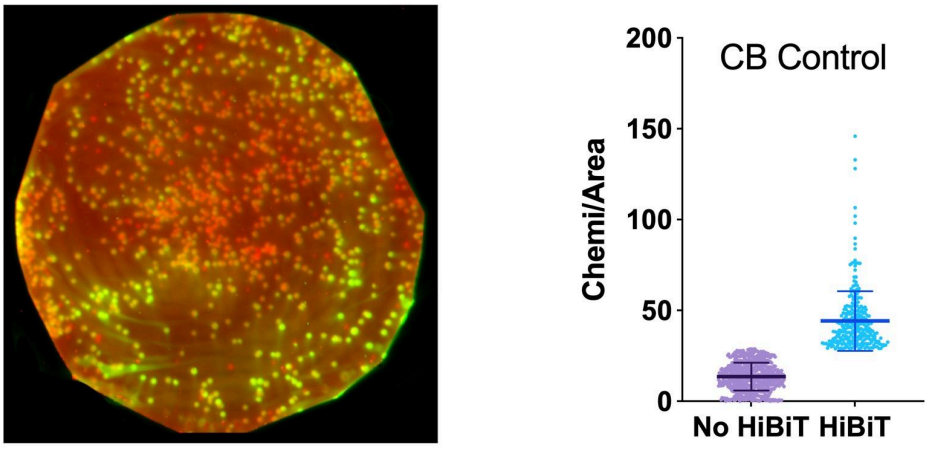

Figure 2. Validation of the mid-western blot assay in *K. marxianus*. Comparison of a representative secretion cassette with a negative-control backbone. Colonies were plated on nitrocellulose membranes and analyzed by chemiluminescence normalized to colony area. Otsu thresholding was applied to separate positive secretors from background. Positive controls produced significantly higher secretion values than the negative control (44.2 ± 16.5 vs 13.6 ± 7.7 chemi/area, mean ± SD).

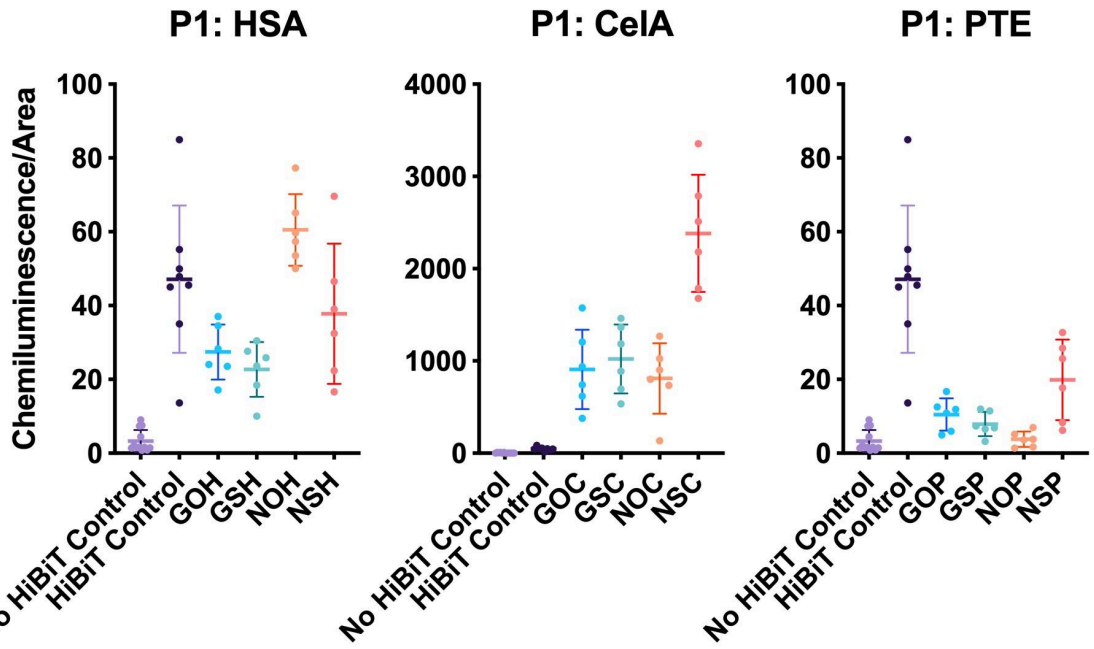

Fig. 3. Benchmarking secretion across 12 promoter–signal peptide–gene combinations. Quantitative secretion values measured by the mid-western blot assay for CelA, HSA, and PTE constructs under GPD1 or NC1 promoters and OST-ProMFα or synthetic signal peptides. Data represent mean ± SD of six biological replicates (controls, n = 12). PTE constructs clustered near background, HSA constructs produced intermediate secretion, and CelA constructs exhibited the highest secretion.

Table 1. Secretion levels of 12 promoter–signal peptide–gene cassettes in *K. marxianus* measured by the mid-western blot assay. Values represent mean ± SD (chemi/area) from six biological replicates (controls, n = 12). Gene Promoter Signal Peptide Mean ± SD (Chemi/Area)

PTE	GPD1	OST-ProMFα	10.49 ± 4.39
		Synthetic (SS)	7.92 ± 3.29
	NC1	OST-ProMFα	3.78 ± 2.10
		Synthetic (SS)	19.85 ± 10.92
HSA	GPD1	OST-ProMFα	27.43 ± 7.45
		Synthetic (SS)	22.68 ± 7.41
	NC1	OST-ProMFα	60.52 ± 9.72
		Synthetic (SS)	37.78 ± 19.00
CelA	GPD1	OST-ProMFα	909.3 ± 429.7
		Synthetic (SS)	1022 ± 373.7
	NC1	OST-ProMFα	811.3 ± 381.6
		Synthetic (SS)	2383 ± 259.3

By utilizing these cassettes in *K. marxianus* we were able to determine differing secretion responses, allowing for low, medium, and high secretion sets. A 96-well plate style mid-western blot was done in order to ascertain the secretion levels of the 12 cassettes. The values represent the mean ± SD from six biological replicates while the controls had twelve. The positive control, a *S. cerevisia*e sample from the initial mid-western blot experiments (unpublished),

while the negative control was the FS_expr strain.Samples were grown in the 96- well plate for 24hours at 30°C, then plated on selective media using a 96-well plate pin replicator and grown for 24hours at 30°C. The control samples had a chemi/area value of 47.15 ± 19.95, while the negative control was 3.309 ± 3.002. Generally the secretion levels of each gene were close, with PTE having the lowest secretion rates, HSA being in the middle, and CelA having a high secretion rate. Secretion levels varied across the 12 promoter–signal peptide– gene combinations (Table 1). PTE constructs showed the lowest secretion (3.8– 19.9 chemi/area), HSA constructs were intermediate (22.7–60.5), and CelA constructs were substantially higher [2]. Within each gene set, promoter and signal peptide choice shifted secretion output, but the overall rank order of PTE < HSA < CelA remained consistent.

These results demonstrate that the mid-western blot can be successfully applied in *K. marxianus* to measure secretion of heterologous proteins across a wide dynamic range. Using a single assay, we distinguished constructs with secretion levels close to background (PTE, 3.8–19.9 chemi/area) from intermediate constructs (HSA, 22.7–60.5) and highly secreted proteins [3] (Table 1).

Importantly, the assay was sensitive enough to resolve differences among promoter–signal peptide combinations within each gene, while maintaining clear separation between low, medium, and high secretion groups.

This establishes the assay as a scalable and quantitative platform for evaluating secretion in *K. marxianus* and provides a foundation for integrating genome-wide libraries.

4.4 Discussion

In this chapter, we established a mid-western blot assay in *K. marxianus* and demonstrated its ability to detect secretion differences across a diverse set of promoter–signal peptide–gene combinations. The assay was able to clearly resolve low, medium, and high secretion constructs, with PTE cassettes clustering near background, HSA cassettes providing an intermediate and tunable range, and CelA cassettes producing secretion levels an order of

magnitude higher ([Table 1]). This dynamic range is critical for downstream screening, as it ensures that both increases and decreases in secretion can be detected depending on the genetic perturbation. By validating the assay with multiple constructs and controls, we show that the system is robust, scalable, and directly applicable to genome-wide knockout libraries for strain optimization [4, 5].

Previous work from our lab has relied on growth-based selection methods for library screening in *K. marxianus*. For example, Robertson et al. developed biosensor-driven genome-wide CRISPR screens that coupled

metabolite production to antibiotic resistance, enabling enrichment of improved variants through selective growth [9]. More broadly, CRISPR-enabled genome-wide screens in non-conventional yeasts have primarily been limited to indirect phenotypes such as survival or growth under selective pressure [8]. While effective, these approaches risk biasing results toward mutations that primarily favor fitness rather than those that directly enhance secretion.

In contrast, the mid-western blot assay described here quantifies secretion itself, independent of cellular growth rate. By leveraging the HiBiT/NanoLuc system [2], secretion differences can be measured across hundreds of colonies simultaneously, producing a quantitative readout that spans nearly three orders of magnitude. This dynamic range ensures that both subtle decreases and substantial increases in secretion can be reliably detected, making the assay ideally suited for integration with genome-wide knockout libraries in *K. marxianus*.

Our results highlight the different ranges provided by the three gene sets. PTE constructs produced secretion values near the negative control, limiting their usefulness for detecting further decreases. CelA constructs secreted at very high levels, often saturating the assay, which may mask incremental improvements. In contrast, HSA constructs consistently displayed intermediate secretion values that span both directions of measurable change. These findings indicate that the HSA set represents the most practical choice for

integration into genome-wide knockout screens, as it provides a balance between sensitivity to both increases and decreases in secretion.

Promoter and signal peptide choice also influenced secretion output within each gene set, consistent with prior studies showing that promoter strength and signal peptide efficiency significantly shape heterologous protein secretion in *K. marxianus* [4, 10]. For instance, the NC1 promoter yielded higher secretion in several constructs compared to GPD1, while the synthetic signal peptide derived from work in *S. cerevisiae* [3] performed comparably to the OST-ProMFα sequence used in previous *K. marxianus* studies.

One limitation of this work is that the constructs tested here were expressed from low-copy plasmids. Variation in plasmid copy number may contribute to colony-to-colony variability and could confound precise quantification. Integration of secretion cassettes into the genome will help stabilize expression levels and improve reproducibility, as has been demonstrated in previous *K. marxianus* engineering studies [6]. Additionally, the dynamic range of the assay, while broad, still has limitations: PTE secretion is too close to background for reliable detection of decreases, and CelA secretion may be too high to resolve incremental improvements.

The next step for this project is to integrate the HSA-based secretion cassettes into the genome at the *HIS3* locus of the FS_exp strain background, which will reduce variability and simplify maintenance compared to plasmid

systems. Subsequent experiments should compare plasmid-based and integrated constructs using the colony blot assay alongside SDS-PAGE or Western blot to verify secretion levels. Once validated, the integrated strains will serve as the foundation for applying our genome-wide knockout libraries in *K. marxianus*. Such libraries have already provided insight into essential and non essential genes in other non-conventional yeasts, including *P. pastoris* and *Y. lipolytica* [4, 1]. By coupling these libraries with the mid-western blot assay, we will be able to directly identify knockouts that enhance protein secretion. This strategy complements earlier findings in *S. cerevisiae*, where disruption of the cell wall protein CWP2 improved secretion [6], suggesting that non-obvious genetic perturbations can substantially affect protein export.

Ultimately, this assay provides a powerful new tool for exploring the genetic landscape of secretion in *K. marxianus*. By enabling direct, high throughput measurement of protein export across hundreds of colonies, it addresses a key gap left by growth-based screens. The integration of this platform with genome-wide libraries will facilitate discovery of beneficial knockouts, potentially uncovering combinations of genetic changes that maximize secretion titers without compromising cellular health. As such, this work expands the synthetic biology toolkit for *K. marxianus* and strengthens its role as a next-generation host for industrial protein production [7].

4.5 Materials and Methods

4.5.1 *K. marxianus* Transformation

K. marxianus FS_expm cells were transformed using a modified PEG 3350/LiAc method. Overnight cultures were grown in 50 mL YPD at 30 °C in a shaking incubator. The following day, cells were harvested by centrifugation at 3,000 × g for 5 minutes, washed once with sterile water, and resuspended in 1 mL of sterile water. For each transformation, 100 µL of cells were transferred to a sterile 1.5 mL microcentrifuge tube. Plasmid DNA (1–2 µg) and denatured salmon sperm DNA (10 µg) were added, followed by 400 µL of PEG 3350/LiAc solution. Tubes were vortexed briefly and incubated at 30 °C for 30 minutes with gentle shaking. After incubation, cells were heat shocked at 42 °C for 40 minutes. Cells were pelleted at 13,000 × g for 30 seconds, resuspended in 1 mL sterile water, and plated on selective media. Plates were incubated at 30 °C for 2–3 days, after which colonies were picked for plasmid confirmation and downstream assays.

4.5.2 Mid-Western Blot Assay

Mid-Western blot assays were performed using a HiBiT/LgBiT NanoLuc system (Promega). Approximately 700 cells from overnight cultures were diluted and plated on 0.45 µm nitrocellulose membranes (placed on selective agar plates). Plates were incubated at 30 °C until visible colonies formed.

Once grown, membranes were carefully lifted and replica-plated onto fresh agar plates for backup. The original membranes were washed twice in TBST buffer (20 mM Tris, 150 mM NaCl, 0.1% Tween-20) for 10 minutes each to remove intact cells, leaving secreted proteins bound to the nitrocellulose surface.

Membranes were then incubated with LgBiT protein solution (Promega) according to the manufacturer's instructions to allow complementation with the HiBiT tag. After incubation, Furimazine substrate (Promega) was added, and chemiluminescence was detected using a ChemiDoc imaging system. For colony size normalization, membranes were also imaged using Alexa 647 fluorescence. Secreted protein levels were quantified as chemiluminescence intensity divided by colony area (chemi/area).

4.5.3 Data Analysis

Colony images were processed using CellProfiler (Broad Institute). Fluorescence images were used to define colony boundaries, and integrated chemiluminescence intensity was divided by colony area to generate normalized secretion values. Otsu thresholding was applied to separate negative controls from weak secretors, providing an unbiased cutoff between populations. All coding and statistical analyses were performed in Python (Spyder IDE). Data are reported as mean ± standard deviation from at least

three biological replicates. Statistical significance was determined using Student's t-test, with $P < 0.05$ considered significant.

4.5.4 E. coli Heat Shock Transformation

Chemically competent NEB 10β (E. coli) cells were used for plasmid transformation. Aliquots of 50 µL were thawed on ice in sterile 1.5 mL microcentrifuge tubes. Plasmid DNA (0.2–1.0 µg) was added, and the mixture was incubated on ice for 5 minutes. Tubes were then heat shocked at 42 °C for 45 seconds, followed by immediate return to ice for 5 minutes. After heat shock, 950 µL of LB medium was added to each tube, and cells were recovered at 37 °C for 1 hour in a shaking incubator. A 100 µL aliquot was plated onto LB agar plates containing the appropriate antibiotic and incubated at 37 °C for 24 hours. The remaining culture was supplemented with antibiotic and incubated at 37 °C in a shaking incubator for plasmid preparation.

4.5.5 Plasmid Mini Prep

Plasmid DNA was purified using the ZymoPURE Plasmid Miniprep Kit (Zymo Research). A 5 mL LB culture containing the selective antibiotic was grown overnight at 37 °C. From each culture, 800 µL was stored as glycerol stocks, and the remaining cells were harvested by centrifugation at 12,000 × g for 5 minutes. The cell pellet was resuspended in 200 µL P1 buffer, followed by 200 µL P2 buffer with gentle inversion. Then, 400 µL P3 buffer was added, mixed

thoroughly, and centrifuged at 12,000 × g for 5 minutes. The supernatant was transferred to a spin column, washed sequentially with 200 µL Endo wash buffer and 400 µL plasmid wash buffer, and centrifuged after each wash. DNA was eluted in 50 µL of nuclease-free water.

4.6 Supporting Information

Supplemental Table 1. Strains and Plasmids

Supplemental Table 1. Strains and Plasmids

Strain	Genometype	Source
Escherichia coli		
NEB10b	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	New England Biolabs
Kluyveromyces marxianus		
CBS712		
FS_ctrl	*ura3Δ kat1Δ alpha3Δ*	Wheeldon Lab
FS_expm	*ura3Δ kat1Δ alpha3Δ URA3::Cas9*	Wheeldon Lab
CS_ctrl	*ura3Δ his3Δ kat1Δ alpha3Δ ku70Δ*	Wheeldon Lab
CS_expm	*ura3Δ his3Δ kat1Δ alpha3Δ ku70Δ URA3::Cas9*	Wheeldon Lab
kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-$SP_{OST\text{-}pro\text{-}MFa}$-HSA-HiBiT-$T_{CYC1}$*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-$SP_{OST\text{-}pro\text{-}MFa}$-CelA-HiBiT-$T_{CYC1}$*	This Study

kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-$SP_{OST\text{-}pro\text{-}MFa}$-PTE-HiBiT-$T_{CYC1}$*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-SP_{Syn}-HSA-HiBiT-T_{CYC1}*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-SP_{Syn}-CelA-HiBiT-T_{CYC1}*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{GPD1}-SP_{Syn}-PTE-HiBiT-T_{CYC1}*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-$SP_{OST\text{-}pro\text{-}MFa}$-HSA-HiBiT-$T_{CYC1}$*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-$SP_{OST\text{-}pro\text{-}MFa}$-CelA-HiBiT-$T_{CYC1}$*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-$SP_{OST\text{-}pro\text{-}MFa}$-PTE-HiBiT-$T_{CYC1}$*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-SP_{Syn}-HSA-HiBiT-T_{CYC1}*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-SP_{Syn}-CelA-HiBiT-T_{CYC1}*	This Study
kmCRG0XX	*FS_expm_HIS3::P_{NC1}-SP_{Syn}-PTE-HiBiT-T_{CYC1}*	This Study

Plasmid	Description	Source
pKD-ost1proMFa-CelAt-C	*OST-pro-MFa; CelA*	Da Silva Lab
pCRG200	*HSA_HiBiT stuff*	Chartron Lab
pCRG176	*Synthetic Signal*	Chartron Lab
pIW1191	*PTE; CYC1 Term*	Wheeldon Lab
pIW601	*Cas9*	Wheeldon Lab
pCRG252	*pIW601_HIS3$_{guide}$*	Wheeldon Lab
pIW1198	*HIS3_HD$_{URA3}$-P$_{GPD1}$-GFP-T$_{CYC1}$-HD$_{URA3}$*	Wheeldon Lab
pCRG244	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{OST-pro-MFa}$-CelA-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study
pCRG245	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{OST-pro-MFa}$-HSA-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study
pCRG246	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{OST-pro-MFa}$-PTE-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study
pCRG247	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{Syn}$-CelA-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study
pCRG248	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{Syn}$-HSA-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study
pCRG249	*pIW1198_URA3_HD$_{HIS3}$-P$_{GPD1}$-SP$_{Syn}$-PTE-HiBiT-T$_{CYC1}$-HD$_{HIS3}$*	This Study

pCRG250	*pIW1198_URA3_HD$_{HIS3}$-* *P_{NC1}-$SP_{OST\text{-}pro\text{-}MFa}$-CelA-HiBiT-$T_{CYC1}$-$HD_{HIS3}$*	This Study
pCRG251	*pIW1198_URA3_HD$_{HIS3}$-* *P_{NC1}-SP_{Syn}-PTE-HiBiT-T_{CYC1}-HD_{HIS3}*	This Study
pCRG253	*pIW1198_URA3_HD$_{HIS3}$-*	This Study
pCRG254	*pIW1198_URA3_HD$_{HIS3}$-*	This Study
pCRG255	*pIW1198_URA3_HD$_{HIS3}$-*	This Study
pCRG256	*pIW1198_URA3_HD$_{HIS3}$-*	This Study

4.7: References

[1] Bever, J. N., et al. (2022). RNA polymerase II–driven CRISPR–Cas9 system for efficient non-growth-biased metabolic engineering of Kluyveromyces marxianus.

[2] Dixon, A. S., Schwinn, M. K., Hall, M. P., Zimmerman, K., Otto, P., Lubben, T. H., ... & Hurst, R. (2016). NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. ACS Chemical Biology, 11(2), 400–408. https://doi.org/10.1021/acschembio.5b00753

[3] Kishkevich, A., et al. (2024). YTK Display-and-Secrete: Screening for optimal protein secretion.

[4] Lang, C., et al. (2020). Developing a broad-range promoter set for metabolic engineering in Kluyveromyces marxianus. [APA details/DOI needed]

[5] Kumar, A., et al. (2021). Identification of novel promoters and terminators in Kluyveromyces marxianus. [APA details/DOI needed]

[6] Li, X., et al. (2020). CRISPR-mediated multigene integration in Kluyveromyces marxianus; also CWP2 disruption in *Saccharomyces cerevisiae*.

[7] McTaggart, T. L., et al. (2019). Synthesis of polyketides from low-cost substrates by Kluyveromyces marxianus.

[8] Robertson, A. W., et al. (2022). Advances in CRISPR-enabled genome-wide screens in yeast.

[9] Robertson, A. W., et al. (2025). PYR1 biosensor-driven genome-wide CRISPR screens for improved monoterpene production in Kluyveromyces marxianus.

[10] Yarimizu, T., et al. (2015). Synthetic signal sequences enabling efficient secretory protein production in Kluyveromyces marxianus.

Chapter 5: Conclusions

In this thesis, I examined promoter mutagenesis as a strategy for tuning gene expression in *Saccharomyces cerevisiae* and developed tools to expand this approach to new contexts. Using Inscripta-based libraries, I showed that mutations in core promoter elements—especially A/T → G/C substitutions in the TATA box and PIC—consistently reduced transcriptional output. These results demonstrated that promoter activity can be predictably decreased without being eliminated, offering a practical method for partial gene knockdowns in metabolic engineering.

To overcome the limitations of the Inscripta platform, I applied CREATE based promoter libraries. This approach expanded both the diversity of substitutions and the coverage of promoter regions, generating tens of thousands of variants across *CAN1*, *HIS3*, and *CCW22*. The CREATE data confirmed the sensitivity of the TATA box and PIC but also revealed promoter-specific differences, such as scanning region effects in *HIS3*. Together, these results show that while some principles of promoter architecture are general, others depend on the sequence context of individual genes.

Finally, I developed and validated a mid-western blot assay for protein secretion in *Kluyveromyces marxianus*. This system enabled direct, quantitative measurement of secretion across hundreds of colonies and distinguished constructs with low, medium, and high levels of protein export. The assay is

well suited for integration with genome-wide knockout libraries, creating an opportunity to identify novel genetic perturbations that improve secretion in non conventional yeasts.

Overall, this work advances promoter mutagenesis as a scalable tool for fine-tuning gene expression and contributes a new assay for studying protein secretion. The findings highlight both general rules of promoter sensitivity and the importance of gene-specific differences. Going forward, combining promoter engineering with genome-wide screening will expand the design space for optimizing yeast hosts, strengthening their role as next-generation platforms for recombinant protein production.