EPFL

Quantitative characterization of the inorganic phosphate gene regulatory networks in S. cerevisiae and bottom up engineering an orthogonal gene network

Présentée le 19 janvier 2024

Faculté des sciences et techniques de l'ingénieur
Laboratoire de caractérisation du réseau biologique
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Shiyu CHENG

Acceptée sur proposition du jury

Prof. M. Dal Peraro, président du jury
Prof. S. Maerkl, directeur de thèse
Prof. D. Shore, rapporteur
Prof. S. Martin, rapporteuse
Prof. J. Mckinney, rapporteur

École
polytechnique
fédérale
de Lausanne

2024

Title: Quantitative characterization of inorganic phosphate gene regulatory networks in *S. cerevisiae* and bottom up engineering an orthogonal gene network

Name: Shiyu Cheng
SCIPER: 293255

Date: 15th of May 2023
Place: Lausanne
Date of the oral thesis exam: 8th of August 2023

Thesis Committee
Thesis director: Prof. Sebastian Josef Maerkl, EPFL, Lausanne, Switzerland
President: Prof. Matteo Dal Peraro, EPFL, Lausanne, Switzerland

Panel experts
internal: Prof. John Mckinney, EPFL, Lausanne, Switzerland
external: Prof. David Shore, University of Geneva, Geneva, Switzerland
external: Prof. Sophie Martin, University of Geneva, Switzerland

Acknowledgements

First and foremost, I want to express my gratitude to my thesis advisor Prof. Sebastian Maerkl, who provided me with the opportunity to become a part of LBNC (Laboratory of Biological Network Characterization), a truly remarkable international team and family. Sebastian has fostered a pleasant and open working environment that has allowed us to grow both scientifically and personally. He is consistently direct, responsive, and supportive. Without him, I would not have had the chance to encounter so many incredible individuals here.

Secondly, I would like to express my appreciation for our Team Yeast. Evan, Ming, and I have been working together on the yeast project since I joined LBNC. Evan mentored me in synthetic biology, generously sharing his knowledge and helping me troubleshoot whenever I faced experimental challenges. We have had countless engaging discussions on both academic and personal topics, especially before the COVID lockdown. Without his assistance, my transition from a novice in biology to my PhD thesis work would not have been as smooth. Ming is not only a great friend but also a supportive and reliable teammate. He consistently maintains a calm and composed demeanor. Collaborating with him has been a truly enjoyable experience. In addition to Team Yeast, I have been fortunate to be surrounded by a group of amazing colleagues and friends. When I first arrived at EPFL for the hiring days, Ivan warmly welcomed me. He also helped me acclimate to life in Lausanne when I first moved here. Ivan is always there to support us when LBNC members face ups and downs in life. His enthusiasm for teaching has been inspiring. Participating in the iGEM team alongside Josh as teaching assistants was a fantastic experience. Thank you, Josh, for supporting one another during iGEM and occasionally troubleshooting together on our PhD projects. I would like to thank Simone for the friendship and joy he brought to our office and lab. I have always been inspired by his passion for work and cars. Thank you, Julia, for your constant encouragement. Thanks to Team

Acknowledgements

Adaptyv Biosystems and Team EirDrop as well – their strong motivation for starting their own companies has been inspiring. Moustafa, Danial, Federico, and Rafael, I truly enjoyed the time we spent conducting experiments together in the lab. I am grateful to Greg for all the interesting and thought-provoking conversations we've had over the years. Thanks to Maria for her assistance in the lab. Finally, I would like to thank all the other LBNC members, including Martine, Ekaterina, Nadanai, Zoe, Barbora, Fabien, Amir, Laura, Laura-R, Ragu, Fetameh, Jian, Amina, Pao-wan, Ludovica, Julian, Amogh, and Fanjun. Completing my PhD with all these incredible individuals has been an extraordinary experience.

Furthermore, I would like to express my gratitude to Prof. Roger G. Clerc, Prof. Didier Trono for their mentorship during my PhD and beyond. I am thankful to Prof. Xingyu Jiang, my bachelor's and master's advisor, who has supported my academic and personal development in recent years. I would also like to thank my candidacy president, Prof. John McKinney, and jury members, Prof. Bruno Correia and Prof. David Shore. Additionally, I appreciate my thesis president, Prof. Matteo Dal Peraro, and thesis jury members, Prof. David Shore, Prof. John McKinney, and Prof. Sophie Martin for taking the time to evaluate and refine my thesis.

Lastly, I wish to thank my parents, grandparents, sister, family, and loved ones for their unwavering support throughout these years. I am fortunate to have been able to come to Switzerland with my family. They have been understanding and patient with me, even when I have had to work late into many evenings, weekends, and holidays. I am grateful to Qian for taking care of our sons and to my elder sister Shizhu for looking after my parents and grandparents while I am away from home. With your love and support, I am confident that I can make a difference.

Lausanne, May 2023   Shiyu Cheng

Abstract

Gene regulatory networks (GRNs) play a crucial role in an organism's response to changing environmental conditions. Cellular behaviors typically result from the integration of multiple gene outputs, and their regulation often demands precise control of numerous genes. A comprehensive understanding of GRNs is essential for replicating their native function and engineering cellular behavior. In this work, we systematically characterized the yeast inorganic phosphate gene regulatory network at the single-cell level under well-controlled environmental conditions. Our analysis identified a robust, perfectly adapted Pho regulon state under intermediate Pi conditions and revealed a previously unknown intermediate nuclear localization state of the transcriptional master regulator Pho4. We also examined the robustness of the yeast phosphate GRN by adjusting the concentration of master regulator Pho4. We developed a Z3EV-mediated inducible phosphate regulatory network to effectively tune Pho4 expression. By adding different concentrations of β-estradiol, which yeast cells do not consume, we modulated Pho4 expression levels. We determined that Pho4 translocation robustness is limited to a specific range of Pho4 expression levels. At higher Pho4 levels, Pho4 can saturate promoter binding sites, resulting in saturated activation and demonstrating robustness to changes in Pho4 concentration. We constructed a synthetic orthogonal GRN using synthetic transcription factors and promoters on a yeast artificial chromosome. The synthetic transcription factor PhoZF was designed by replacing the basic domain of native Pho4 with a Zinc Finger binding domain, and the synthetic promoters were created by substituting consensus binding motifs in native Pho4-regulated promoters and a minimal CYC1 promoter with the Zinc Finger binding motifs. The synthetic transcription factor effectively activated the synthetic promoters, generating sufficient fold changes to fine-tune input-output levels. Our findings laid the groundwork for the bottom-up assembly of a complete synthetic gene regulatory network to replace native networks.

Keywords: yeast gene regulatory network, pho regulon, zinc finger, synthetic transcription factor, synthetic promoter, yeast artificial chromosome

Résumé

Les réseaux de régulation génique (GRN) jouent un rôle crucial dans la réponse d'un organisme aux changements des conditions environnementales. Les comportements cellulaires résultent généralement de l'intégration de plusieurs produits de gènes, et leur régulation nécessite souvent un contrôle précis de nombreux gènes. Une compréhension approfondie des GRN est essentielle pour reproduire leur fonction native et manipuler le comportement cellulaire. Dans ce travail, nous avons caractérisé systématiquement le réseau de régulation génique du phosphate inorganique de la levure au niveau des cellules individuelles dans des conditions environnementales bien contrôlées. Notre analyse a identifié un état de Pho regulon robuste et parfaitement adapté sous des conditions intermédiaires de Pi et a révélé un état de localisation nucléaire intermédiaire inconnu auparavant pour le régulateur transcriptionnel maître Pho4. Nous avons également examiné la robustesse du GRN phosphate de la levure en ajustant la concentration du régulateur maître Pho4. Nous avons développé un réseau de régulation du phosphate inductible médié par Z3EV pour réguler efficacement l'expression de Pho4. En ajoutant différentes concentrations de β-estradiol, que les cellules de levure ne consomment pas, nous avons modulé les niveaux d'expression de Pho4. Nous avons déterminé que la robustesse de la translocation de Pho4 est limitée à une plage spécifique de niveaux d'expression de Pho4. À des niveaux de Pho4 plus élevés, Pho4 peut saturer les sites de liaison des promoteurs, entraînant une activation saturée et démontrant une robustesse face aux changements de concentration de Pho4. Nous avons construit un GRN orthogonal synthétique en utilisant des facteurs de transcription et des promoteurs synthétiques sur un chromosome artificiel de levure. Le facteur de transcription synthétique PhoZF a été conçu en remplaçant le domaine de base du Pho4 natif par un domaine de liaison Zinc Finger, et les promoteurs synthétiques ont été créés en remplaçant les motifs de liaison consensus dans les promoteurs régulés par Pho4 natif et un promoteur minimal CYC1 par des motifs de liaison Zinc Finger. Le facteur de transcription synthétique a activé efficacement les promoteurs synthétiques, générant des changements de plis suffisants pour ajuster finement les niveaux d'entrée-sortie. Nos résultats ont jeté les bases de l'assemblage de fond en comble d'un réseau de régulation génique synthétique complet pour remplacer les réseaux natifs.

Résumé

Mots-clés : réseau de régulation génétique de levure, pho regulon, doigt de zinc, facteur de transcription synthétique, promoteur synthétique, chromosome artificiel de levure

Contents

Contents

List of Figures

List of Tables

1 Introduction

1.1 Gene regulatory networks

Biological organisms rely on complex gene regulatory networks (GRNs) to respond to changing environmental conditions. The cellular behaviors are usually an integration of multiple gene outputs and their regulation often requires the precise regulation of multiple genes. Systematic expression of thousands of genes can be regulated in the developmental process of higher order living organisms, such as animals or plants (1). The GRNs usually consist of genes encoding transcription factors and the cis-regulatory elements that control the gene expression. Each of the cis-regulatory elements contains specific target site sequences, and receives multiple inputs from other genes in the GRN. These inputs are transcription factors (activators and repressors) that recognize specific sequences within the regulatory elements (2). Understanding of the GRNs is important to help in solving different biological and biomedical problems (3). Due to very high number of interactions between about 20,000 genes in human and about 6000 genes in yeast, the GRNs are of tremendous help in narrowing down the number of interactions that can be correspondent to specific phenotypes.

1.1.1 Gene expression and regulation

The living cells need to organize and control the intake of nutrients, expulsion of waste, and arrangement of energy, to function properly or to survive under adverse circumstances. The gene expression patterns need to be arranged in place and time to make this possible. The most basic function of a gene is to control its expression levels, which in turn controls the interaction of cell with its environment. The process of transcription of DNA into RNA, and subsequent translation into protein, controlled by the central dogma of molecular biology (4), is the basis of gene expression and regulation. Gene expression requires the recruitment of a RNA polymerase to a promoter sequence to initiate the transcription, which then transcribe the gene into mRNA by forming RNA polymers. Transcription factors modulate the expression rate by inhibiting or enhancing the transcription rate of genes. Gene products can interact with other genes and proteins to form a network that allow self-regulation to emerge (1). Transcription factors, as the products of gene expression, can serve as a feedback loop for the regulation of the genetic networks. mRNA can also become other types of RNA (e.g. iRNA) to take their own regulatory role into protein production. One can identify specific cell types by fingerprinting the RNA expression levels through RNA sequencing techniques (5). Individual cells can exhibit a substantial variation in RNA expression between different cells, even with the same cell type, depending on circumstances and age (6). The difference between individual cell level can be resolved with the advent of single-cell transcriptomics(7).

The regulation of gene expression by GRNs is dynamic and occurs at multiple levels, including chromatin structure, DNA methylation, and post-transcriptional modifications. The expression of genes is dependent not only on their genetic sequences, but also on their accessibility. Epigenetics is the heritable features that are not encoded in the genome's sequences. Instead, it alters the sequence's physical structure by wrapping DNA around histones or methylating nucleotides. Chromatin is a collection of DNA, RNA, and protein that compacts the structure of these molecules, thereby increasing their stability, density, and order. DNA is wraped around histones to create nucleosomes within chromatin. The accessibility of DNA sequences within these nucleosomes significantly impacts the expression of genes situated within the nucleosome (8). Modification of histones can alter the gene expression in a predictive manner (9). DNA methylation is an additional form of epigenetic regulation. By adding one carbon and three hydrogens to an existing structure, DNA methylation could change the structure of nucleotides within the sequence, and effectively alter the binding affinity of the binding sites to prohibit interactions. Methylation can be induced by the environmental factors (10) and transmitted through multiple cell divisions and generations (11). Finally, post-transcriptional modifications (12), such as alternative splicing (13) and protein modification, can further influence the expression of genes.

1.1.2   Computational models for GRNs

In addition to various experimental methods developed for large-scale GRN studies, different computational models have also been developed for GRN analysis to reveal the mechanisms that underlie GRNs (14). These include i) logical models that describes GRN qualitatively, such as Bloolean network, probabilistic Boolean networks, MetaReg, Petri nets; ii) continuous models to understand behaviours that depend on timing and molecular concentrations; iii) Single molecule level models that are more detailed and can explain stochasticity (15). Since comprehensive models of GRNs would require a substantial degree of complexity, GRN models usually focus on particular aspects of genetic regulation. Hill kinetics, for example, is commonly used to model cooperative binding of two or more proteins to promoters, enhancers, or other region of genes, where the association/dissociation rate of the protein and gene and the cooperativity of multiple proteins are used to formulate these kinetics. The quality of a model can be accessed by the ability to predict the experimental data. The simpler model is more preferable if two models match the date equally well, because it requires less parameters. The stochasticity of gene expression could have big impact on the dynamics of GRNs and thus create the diversity of populations. The more predictive biological models of GRNs are of great benefit to elucidate mechanisms

or evolutionary process in the basic research, and replace some tedious and costly experimental work in bio-medicines.

1.2 Yeast gene regulatory networks

1.2.1 Yeast system biology

The baker's yeast, *Saccharomyces cerevisiae*, is a type of single-celled eukaryote with approximately 6,000 genes (16). It is one of the most well-studied and important model organisms in the biological sciences (17). As eukaryotes, yeast cells have a true nucleus and other complex cellular structures, and their genetic and metabolic structure is relatively simple compared to more complex organisms, such as mammals. This makes yeast a useful model organism for the study of cellular processes and the regulation of gene expression, metabolism, and signaling. In addition to its importance as a model organism, yeast also has many practical applications. For example, yeast is used in the production of bread, beer, and wine, and is also used as a tool for producing recombinant proteins and other biologically active molecules (18).

Yeast system biology (19) is an interdisciplinary field that combines the tools and concepts of genetics, biochemistry, molecular biology, and computational biology to study the biological systems of yeast. Research in yeast system biology has led to many important discoveries and insights into the regulation of gene expression, metabolic pathways, and cellular signaling, and has had important implications for our understanding of these processes in more complex organisms. From 2001, three Nobel Prizes have been awarded based on forward genetics using this organism (20). These include the prize awarded to Hartwell, Nurse, and Hunt in 2001 "for their discoveries of key regulators of the cell cycle", the prize awarded to Rothman, Scheckman, and Südhof in 2013 "for their ground-breaking work on cell membrane vesicle trafficking", and the prize awarded to Yoshinori Ohsumi in 2016 for "his discoveries of the mechanisms for autophagy", where all these ground-breaking works were accomplished using yeast as the model system. Additionally, the results of yeast system biology research have practical applications in areas such as biotechnology and medical research (21), as a better understanding of yeast biology can lead to the development of new drugs and therapies, and the improvement of yeast strains for bioproduction (18).

Over years, different collections and resources are developed and available for studying the genome of the yeast Saccharomyces cerevisiae. Thanks to the efforts of the yeast genomoe project, yeast became the first eukaryotic organism get fully sequenced (16). Concurrently, multiple academic (22; 23) and commercial (24) yeast research groups were exploring genome-wide gene expression analysis to uncover comprehensive transcriptional responses. Subsequently, a comprehensive analysis was conducted to uncover the localization of proteins in yeast (25). At the same time, the availability of genome-scale mutant collections allowed for systematic screening of

phenotypes. A transposon-tagged mutants library (26) with over 11,000 strains was generated for genome-wide analysis of disruption phenotypes, gene expression and protein localization, leading to the large-scale analysis of gene function in the budding yeast. The Yeast Gene Deletion Project is a large-scale effort to generate a comprehensive collection of yeast strains in which every gene in the yeast genome has been deleted. This collection provides an important resource for studying the function of individual genes in yeast, and it has been widely used to identify genes that are essential for specific cellular processes. A yeast deletion collection, also called yeast knockout collection (27; 28), containing over 21,000 mutant stains was constructed by precise deletion of the open reading frames. This collection is the first and the only complete deletion collection available for eukaryotic organism, which has been used in numerous genome-wide phenotypic screens to study fitness, mating, sporulation, membrane trafficking, stress (29). The discovery of green fluorescent protein (GFP) by Tsien et al. (30) and subsequent methods to incorporate it into yeast's genes led to the development of a yeast library with GFP tags (25). This library offers quantitative gene expression signals and subcellular localization and protein-protein interaction data in eukaryotic cells. The development of Saccharomyces Genome Database (SGD, www.yeastgenome.org) (31; 32) serves as the primary community resource for the budding yeast Saccharomyces cerevisiae. The SGD provides a powerful genome browser containing the highest-quality information that is manually curated from peer-reviewed literature about the yeast genome, its chromosomal features, and their functions. The SGD carefully manages and organizes data obtained primarily from focused studies to produce machine-readable Gene Ontology (GO) annotations (33) for yeast genes. Meanwhile, SGD database serves as a bridge between highly precise, yet limited, focused studies and broad global studies, allowing for comprehensive investigation of the roles of previously uncharacterized genes and the identification of novel functional connections between seemingly unrelated processes (19). The SGD database simplifies the retrieval and presentation of this information, streamlining experimental design and analysis. Users can effortlessly query information about a gene, whether it be based on a sequence, name, function, or position. These collections and resources represent valuable tools for investigating the yeast genome and play a pivotal role in expanding our comprehension of this essential model organism.

1.2.2   Yeast transcriptional gene regulatory network

Yeast GRNs (34) are composed of a variety of components, including genes, transcription factors (TFs), and regulatory elements such as enhancers and silencers. TFs play a critical role in regulating gene expression in yeast, and they can either activate gene expression by binding to enhancers or repress gene expression by binding to

silencers. The activity of TFs is controlled by various signals, such as nutrient availability and changes in the cellular environment. There are around 209 putative yeast TFs that contain a specific binding domain as suggested in the Yeast TF Specificity Compendium (35). The vast majority of yeast TFs can be classified into only a few categories based on their DNA-binding domain. The two most prevalent categories of yeast TFs are the GAL4/zinc cluster domain (comprising 57 proteins), and the zinc finger C2H2 domain (comprising 41 proteins), which is the most widespread among all eukaryotes (36). Other classes with a significant number of proteins include bZIP (15 proteins), Homeodomain (12 proteins), GATA (10 proteins), and basic helix-loop-helix (bHLH) (8 proteins). Only a small number of yeast TFs are critical for viability in rich or complete medium. The majority of TFs control transcription only under specific circumstances, such as under nutrient-deprived environments and various types of stress (37). Moreover, there exists genetic redundancy among TFs, as they can regulate an overlapping collection of genes, as previously identified by ChIP-chip (38). The identification of the sequence specificity of TFs represents a crucial element in the study of transcriptional regulatory mechanisms and networks (39).

In vitro methods

Various methods were developed for in vitro (40) characterization of the sequence specificity of TFs, including binding site enrichment (41; 42), DNA microarrys (43; 44), and microfuidic devices (45; 46). The systematic evolution of ligands by exponential enrichment (SELEX) involves incubating a purified TF with a pool of randomized DNA oligonucleotides. The oligonucleotides that the TF binds to are selected, amplified using PCR, and then re-incubated with the TF, allowing for successive rounds of selection to identify high-affinity binders or the TF's consensus TFBS. Protein binding microarrays (PBMs) (43) is a method for discovery protein – DNA binding events at a relatively high throughput. Approximately 150 yeast TFs were evaluated by using universal PBMs (44; 47; 48; 49). In a typical PBM experiment, a protein of interest is exposed to a microarray with a base priming site and approximately 40,000 double-stranded DNA probes, each containing a 25-base sequence that differs from one probe to the next by 10 bases. The probe set is carefully designed that each 10-base sequence occurs only once, while PBMs can produce longer motifs of at least 14 bases (44; 47). A high-throughput microfluidics platform, termed mechanically induced trapping of molecular interactions (MITOMI), was constructed by a process known as multilayer soft lithography (45). These devices are then aligned to an epoxy-coated glass slide that contains thousands of micro-arrayed DNA spots utilizing standard DNA microarray printing instruments. MITOMI is capable of detecting transient and low-affinity interactions that other techniques often miss due to the necessity of stringent wash steps. MITOMI has been used to evaluate the sequence specificity of 20 yeast TFs, where the off-rates of individual sequences were quantified (45; 50).

In vivo methods

Chromatin immunoprecipitation (ChIP) is the most frequently used *in vivo* method for probing genome-wide TF binding, which can be combined with either DNA microarry technology (ChIP-chip) (51; 52; 53) or with massive parallel sequencing (ChIP-seq) (54; 55). After extracting and fragmenting genomic DNA, antibodies specific to the DNA-binding protein are used to precipitate the resulting fragments. This process results in the formation of complexes between the DNA-binding protein and the DNA bound to it with the antibody. Finally, the DNA sequences of these complexes are analyzed (56; 57). In ChIP-seq (54), DNA fragments are sequenced using next-generation sequencing (NGS), which has enabled chromosome conformation capture (58), open chromatin detection, and ambitious projects such as the Encyclopedia of DNA Elements (ENCODE) (59) and modENCODE (60), made possible by the decreasing cost of sequencing. Compared to ChIP, the DNA adenine methyltransferase (Dam) identification (DamID) technique (61) does not require the use of specific antibodies, which can be subject to availability issues. However, DamID has lower resolution than ChIP (62). Another technique, DNase-seq (63), is based on identifying genomic sites accessible for DNaseI cleavage, enabling the discovery of open chromatin since DNaseI can only cut nucleosome-free chromatin in DNaseI hypersensitive sites. Formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq (64) does not require cell treatment prior to adding formaldehyde, which can quickly cross-link cellular machinery leading to 50% lethality in 100 s and 99% lethality in 360 s in yeast, making the analyses less dependent on time artifacts of the protocol. A more recent method, assay for transposase-accessible chromatin using sequencing (ATAC-seq) (65; 66), emerged as a faster and more sensitive alternative to previously used methods. Unlike DNase-seq and FAIRE-seq, ATAC-seq does not require ultrasound nuclei extraction or formaldehyde treatment, and the entire process takes under three hours. The hybrid system is another in vivo method. The first hybrid system, the yeast two-hybrid system (Y2H) (67; 68), was designed to discover protein-protein interactions and was later modified to study protein-DNA interactions, predominantly TF-promoter interactions as a Y1H system (69; 70), and ligand-protein interactions as Y3H (71). To measure the gene expression output in the living cells, protein of interest can be conjugated with different fluorescent proteins (72) and be quantified by measuring fluorescent signals. Various techniques are used to measure the fluorescent signal, including plate readers, fluorescent microscopy, and Flow cytometry (73). Plate reader could read the absorbance and fluorescent signal over time, with a medium to high throughput (in a 96-well or 384-well format). However, it can only perform the bulk measurement instead of single cell measurement. In addition, one could not control precisely the media conditions to a constant level due to the consumption during cell incubation. The fluorescence-activated cell sorting (FACS) has become a widely

used technique for studying gene regulatory networks and their regulation. FACS can enrich the concentration of a rare cell type by specifically fluorescently labeling it, such as by expressing cell biomarkers fused with fluorescent proteins. FACS technology has been developed to capture over 32 fluorescent channels (74). By using bar-coded promoter libraries to control fluorescence reporter genes, gene expression can be measured, potentially under inducible conditions, with the help of FACS. The enriched cell fractions resulting from the cell sorting can be linked with their genotype by single-cell sequencing. Models of gene regulatory network functioning can then be proposed based on the sequencing data analysis (75). However, it is not possible to generate temporal data to track single cell over time. In order to gain a better understanding of a biological system, it is crucial to address several factors that can act as bottlenecks, including the absence of high-throughput experiments, the lack of high-quality single-cell data, and the difficulty in maintaining stable media conditions.

Microfluidic platform for GRN analysis

Microfluidics is a technology that allows for the manipulation and control of fluids at the microscale, both temporally and spatially (76). With the invention of soft lithography (77), Polydimethylsiloxane (PDMS)-based microfluidic devices further expanded the capabilities of microfluidics, due to its bio-compatibility, transparency, and ability to be bonded to glass microscopy slides or coverslips (78). Microfluidics has been applied to many disciplines and has become a commonly used technology in different biological applications (79), including cell sorting (80), continuous cell culture (81), PCR (82), protein-DNA interactions (45; 46), encapsulation (83; 84), and organ-on-chips (85; 86). Microfluidics are also advantageous in studying system biology (87) and GRN dynamics (88) by generating many different environmental conditions in a high throughput and single cell manner (89). A chemostat device was developed by Hasty's lab to establish a continuous yeast cell culture on the chip (90), which allowed for monitoring of hundreds of yeast cells over time. Our lab developed a microfluidic chemostat array that contains over 1,000 cell culture chambers (81) and performed large-scale genotype-to-phenotype mapping by using this high throughput platform (91). To fabricate the microchemostat, yeast cells were deposited on a glass slide using a microarrayer and a PDMS microfluidic device was aligned on top of it. Yeast cells can grow continuously inside the chamber, and excess cells get removed via an exit channel. Single cell traps (92) were also developed by various labs (93) to immobilise a single cell for studying cell aging (94; 95) or cell decision-making processes (96; 97). Multiplexer (98), mixer (99), or gradient/dilution generator (100; 101), could be integrated into microfluidics to generate different media or chemical input for GRN dynamics studies. Our lab recently have developed a microfluidic platform, named MicroSTAR (102), that combines the following features: 1) continuous media supply,

2) high throughput for both strains and media conditions, 3) ability to measure at the single cell level, 4) observation of sub-cellular localization, and 5) long-term cell culturing. This new device was used in this study for the quantitative characterization of the yeast phosphate gene regulatory networks.

1.3   Yeast pho regulon: a phosphate-responsive network

Inorganic phosphate (P_i) is an essential nutrient for the growth and survival of all living organisms, including yeast (103). Yeast cells have developed a sophisticated system for sensing and responding to changes in P_i availability, which involves a complex network of phosphate-responsive regulatory genes (Figure 1.1). This network, known as the yeast phosphate gene regulatory network (PGRN), plays a critical role in the regulation of P_i uptake, transport, and utilization (104; 105; 106). The PGRN involves the interplay of various proteins (Table 1.1), including transcription factors (Pho4, Pho2), transporters (Pho84, Pho89, Pho87, Pho90, Pho91), phosphatases (Ppn1, Ppn2, Phm8, Pho5, Pho11, Pho12, Pho8), kinases (Pho80, Pho85, Pho81), and vaculor storage regulators (VTC1-4), and is tightly regulated to ensure that the cell has an adequate supply of phosphate for various cellular processes.
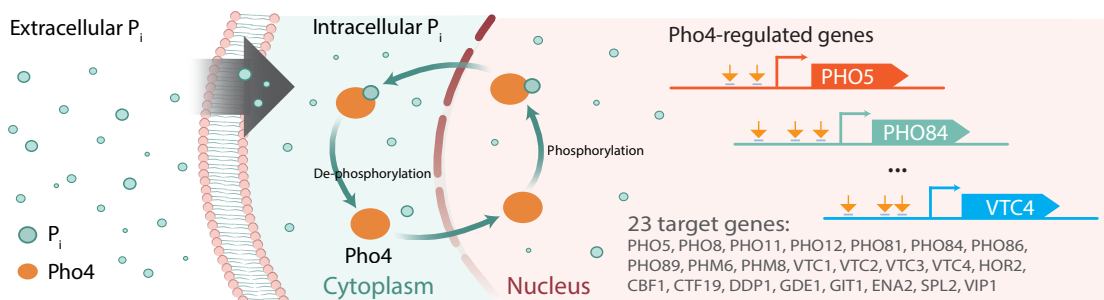
Figure 1.1: **Overview of the *S. cerevisiae* Pho regulon.** P_i is sensed internally by *S. cerevisiae*, leading to phsophorylation of the master regulator Pho4p, which in turn determines whether Pho4p is nuclear localized, or not. When nuclear localized, Pho4p can bind to binding sites (indicated by orange arrows) in its target promoters leading to expression of the Pho regulon genes.

The pho regulon is activated under P_i starvation (Figure 1.1). The master transcriptional regulator Pho4 is known to be expressed at a constitutive expression level. The localization of Pho4p is determined by its phosphorylation state, with full phosphorylation leading to cytoplasmic localization and inactivity. On the other hand, dephosphorylation leads to nuclear localization of Pho4p. Pho4p nuclear localization triggers the transcription of over 20 Pho regulon genes, which include phosphate transporters, phosphatases, and vacuolar storage regulators (103). The translocation of Pho4p is faciliated by the cyclin-dependent kinase complex Pho80p/Pho85p (107), which internally senses the P_i levels in yeast and transmits this information to Pho4p through phosphorylation (108; 109). Under P_i rich conditions, Pho4 is phosphorylated by this complex and localize to the cytoplasm. Under P_i depleted conditions, Pho81 inhibits the complex, and Pho4 is unphosphorylated and translocated into nucleus (107; 110), activating the phosphate-responsive genes. Under intermediate P_i

concentrations, the kinase complex is partially active, resulting in the accumulation of a nuclear form of Pho4 that is partially phosphorylated. This form of Pho4 is capable of inducing the transcription of a subset of phosphate-responsive genes, such as Pho84 (111).

The Pho system comprises two types of Pi transporters, including high-affinity (low Km) transporters (Pho84, Pho89) and low-affinity (high Km) transporters (Pho87, Pho90, Pho91), respectively (112). Pho87, Pho90, and Pho91 are expressed constitutively. During P_i rich condition, P_i can be imported by Pho87 and Pho90 (113; 114) from extracellular the environment, and be imported by Pho91 (115) from the vacuole. When P_i is depleted, Pho84 and Pho89 are transcriptionally upregulated (116). The low affinity transporter are replaced by high affinity transporters (117). P_i can be imported by high affinity (Pho84 and Pho89) from the extracellular environment. The Pho84 transporter plays a crucial role in the negative feedback loop by increasing P_i uptake. This helps to maintain the intracellular P_i levels and prevent excessive accumulation of P_i, which can be toxic to the cell. The upregulation of Spl2 could generate a positive feedback loop by suppressing the expression of Pho87 and Pho90 (118).

Under P_i rich conditions, P_i is stored in the vacuole for future use. VTC genes, including VTC1 to VTC4, play a crucial role in regulating inorganic phosphate storage in the vacuole as well as in the synthesis of polyphosphate (PolyP). PolyP is a polymer of up to a thousand P_i units linked through phosphoric anhydride bonds. PolyP accumulates in yeast cells to high concentrations in the lumen of the acidocalcisome, where it can be hydrolyzed to P_i by the polyphosphatases (Ppn1 and Ppn2) (119; 120; 121) and released back into the cytoplasm. When faced with P_i starvation, cells respond in a graded manner, with different genes being induced based on the severity of phosphate limitation. Under moderate P_i limitation, the cells induce the expression of high-affinity P_i importers (Pho84) and the polyP polymerase VTC. The VTC complex performs dual functions as both a polyP polymerase and a polyP translocase. It enables the cytosolic synthesis of polyP and simultaneously facilitates its translocation into the organelle lumen (122). However, the expression of proteins involved in P_i foraging is upregulated only under conditions of more severe P_i limitation (123). Phosphatases, such as Pho5, Pho11, and Pho12, can be secreted from cells to recover P_i from phosphorylated substrates in their environment (104). In addition, internal sources of P_i can also be recycled through the liberation of P_i from internal molecules (124), such as nucleotides or phospholipids via enzymes such as Phm8 (125; 126) and Gde1 (127).

This gene expression pattern results in elevated phosphate import and helps to sus-

tain intracellular Pi levels. The activation of the Pho regulon poses a challenge in preventing a cycle of futile oscillation, where activation of the Pho regulon leads to an increase in internal P_i levels, causing P_i levels to rise again above the activation threshold and turning off the system, only to be reactivated once Pho regulon gene products are diluted out. This has led to the proposition of one or more negative feedback loops that act to decrease internal P_i concentrations, which are necessary for the Pho regulon model. These feedback loops include Spl2p, which binds and inactivates low-affinity transporters (118), and Vtc1-4p, which sequesters and stores P_i in the vacuole (123), reducing internal P_i levels. According to the current model, the rate of P_i influx is influenced by the concentration of extracellular P_i, resulting in a range of internal P_i levels with varying concentrations (123; 117). Our understanding of Pho4 also suggests that, nuclear concentration of Pho4p should vary in response to internal P_i concentration in order to fine-tune gene expression (111). However, no differences in nuclear-localized Pho4p levels have been observed, and no intermediate states have been identified. The mechanistic details of how the Pho regulon functions are still not clear.

Table 1.1: Genes in the phosphate gene regulatory network.

Genes included in this study		
Gene	**Location**	**Function**
Pho4	Chr VI	Basic helix-loop-helix (bHLH) transcription factor of the myc-family; activates transcription cooperatively with Pho2p in response to phosphate starvation; binding to 'CACGTG' motif is regulated by chromatin restriction, competitive binding of Cbf1p to the same DNA binding motif and cooperation with Pho2p; function is regulated by phosphorylation at multiple sites and by phosphate availability.
PHO5	Chr II	Repressible acid phosphatase; 1 of 3 repressible acid phosphatases that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway derived cell surface glycoprotein; induced by phosphate starvation and coordinately regulated by Pho4p and PHO2.
Pho11	Chr I	One of three repressible acid phosphatases; glycoprotein that is transported to the cell surface by the secretory pathway; induced by phosphate starvation and coordinately regulated by Pho4p and PHO2; PHO11 has a paralog, PHO12, that arose from a segmental duplication.
Pho12	Chr VIII	One of three repressible acid phosphatases; glycoprotein that is transported to the cell surface by the secretory pathway; pregulated by phosphate starvation; PHO12 has a paralog, PHO11, that arose from a segmental duplication.
Pho84	Chr XIII	High-affinity inorganic phosphate (Pi) transporter; also low-affinity manganese transporter; regulated by Pho4pp and Spt7p; mutation confers resistance to arsenate; exit from the ER during maturation requires Pho86p; cells overexpressing Pho84p accumulate heavy metals but do not develop symptoms of metal toxicity.
Pho89	Chr II	Plasma membrane Na+/Pi cotransporter; active in early growth phase; similar to phosphate transporters of Neurospora crassa; transcription regulated by inorganic phosphate concentrations and Pho4pp; mutations in related human transporter genes hPit1 and hPit2 are associated with hyperphosphatemia-induced calcification of vascular tissue and familial idiopathic basal ganglia calcification.

PHM6	Chr IV	Phosphate metabolism protein 6. Expression is regulated by phosphate levels.
Pho8	Chr IV	Repressible vacuolar alkaline phosphatase; controls polyphosphate content; regulated by levels of Pi and by Pho4pp, Pho9p, Pho80p, Pho81p and Pho85p; dephosphorylates phosphotyrosyl peptides; contributes to NAD+ metabolism by producing nicotinamide riboside from NMN.
Pho86	Chr X	Endoplasmic reticulum (ER) resident protein; required for ER exit of the high-affinity phosphate transporter Pho84p, specifically required for packaging of Pho84p into COPII vesicles; protein abundance increases in response to DNA replication stress.
VTC1	Chr V	Regulatory subunit of the vacuolar transporter chaperone (VTC) complex; VTC complex is involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; also has mRNA binding activity; protein abundance increases in response to DNA replication stress.
VTC2	Chr VI	Regulatory subunit of the vacuolar transporter chaperone (VTC) complex; involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; VTC2 has a paralog, VTC3, that arose from the whole genome duplication.
VTC3	Chr XVI	Regulatory subunit of the vacuolar transporter chaperone (VTC) complex; involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; VTC3 has a paralog, VTC2, that arose from the whole genome duplication.
VTC4	Chr X	Vacuolar membrane polyphosphate polymerase; subunit of the vacuolar transporter chaperone (VTC) complex involved in synthesis and transfer of polyP to the vacuole; regulates membrane trafficking; role in non-autophagic vacuolar fusion; protein abundance increases in response to DNA replication stress.

VIP1	Chr XII	Bifunctional inositol pyrophosphate kinase and phosphatase; contains an N-terminal PP-InsP kinase domain that phosphorylates inositol hexakisphosphate and heptakisphosphate, and a C-terminal phosphatase domain that hydrolyzes both 1PP-InsP5 and 5PP-InsP5; IP7 levels decrease during starvation, suggesting a role for PP-InsP enzymes in Pi homeostasis; may regulate the dimorphic switch and the function of the cortical actin cytoskeleton.
PHM8	Chr V	Lysophosphatidic acid (LPA) phosphatase, nucleotidase; principle and physiological nucleotidase working on GMP, UMP and CMP; involved in LPA hydrolysis in response to phosphate starvation and ribose salvage pathway; phosphatase activity is soluble and Mg2+ dependent; expression is induced by low phosphate levels and by inactivation of Pho85p; repressed by Gcn4p under normal conditions; PHM8 has a paralog, SDT1, that arose from the whole genome duplication.
HOR2	Chr V	DL-glycerol-3-phosphate phosphatase involved in glycerol biosynthesis; also known as glycerol-1-phosphatase; induced in response to hyperosmotic or oxidative stress, and during diauxic shift; HOR2 has a paralog, GPP1, that arose from the whole genome duplication.
CTF19	Chr XVI	Outer kinetochore protein, needed for accurate Chr segregation; component of kinetochore sub-complex COMA (Ctf19p, Okp1p, Mcm21p, Ame1p) that functions as platform for kinetochore assembly; required for spindle assembly checkpoint; minimizes potentially deleterious centromere-proximal crossovers by preventing meiotic DNA break formation proximal to centromere; homolog of human centromere constitutive-associated network (CCAN) subunit CENP-P and fission yeast fta2.

CBF1	Chr X	Basic helix-loop-helix (bHLH) protein; forms homodimer to bind E-box consensus sequence CACGTG present at MET gene promoters and centromere DNA element I (CDEI); affects nucleosome positioning at this motif; associates with other transcription factors such as Met4p and Isw1p to mediate transcriptional activation or repression; associates with kinetochore proteins, required for Chr segregation; protein abundance increases in response to DNA replication stress.
ENA1	Chr IV	P-type ATPase sodium pump; involved in Na+ and Li+ efflux to allow salt tolerance.
ENA2	Chr IV	P-type ATPase sodium pump; involved in Na+ efflux to allow salt tolerance; likely not involved in Li+ efflux.
GDE1	Chr XVI	Glycerophosphocholine (GroPCho) phosphodiesterase; hydrolyzes GroPCho to choline and glycerolphosphate, for use as a phosphate source and as a precursor for phosphocholine synthesis; may interact with ribosomes.
DDP1	Chr XV	Diadenosine and diphosphoinositol polyphosphate phosphatase; hydrolyzes diphosphorylated inositol polyphosphates and diadenosine polyphosphates; high specificity for diadenosine hexa- and pentaphosphates; contains endopolyphosphatase activity with a high affinity for polyphosphates, an activity also observed for its human DIPP homologs; possesses mRNA decapping activity; nudix hydrolase family member; protein abundance increases in response to DNA replication stress.
GIT1	Chr III	Plasma membrane permease; mediates uptake of glycerophosphoinositol and glycerophosphocholine as sources of the nutrients inositol and phosphate; expression and transport rate are regulated by phosphate and inositol availability.
SPL2	Chr VIII	Protein with similarity to cyclin-dependent kinase inhibitors; downregulates low-affinity phosphate transport during phosphate limitation by targeting Pho87p to the vacuole; upstream region harbors putative hypoxia response element (HRE) cluster; overproduction suppresses a plc1 null mutation; promoter shows an increase in Snf2p occupancy after heat shock; GFP-fusion protein localizes to the cytoplasm.

YJL119C	Chr X	Dubious open reading frame; unlikely to encode a functional protein, based on available experimental and comparative sequence data.
YAR070C	Chr I	Dubious open reading frame; unlikely to encode a protein, based on available experimental and comparative sequence data; YAR070C has a paralog, YHR214C-B, that arose from a segmental duplication.
YNL217W	Chr XIV	Zn2+-dependent endopolyphosphatase; required with PPN1 to mobilize polyphosphate stores in response to phosphate starvation; member of the PPP-superfamily of metalloproteases; localizes to the vacuolar lumen via the MVB pathway; null mutant is highly sensitive to azaserine and resistant to sodium-O-vandate.
Reference: *Saccharomyces* Genome Database (SGD) (31; 32)		

1.4   Synthetic biology

Synthetic biology (128) is an interdisciplinary field that combines principles from biology, engineering, computer science, and other related disciplines to design and construct novel biological systems or modify existing ones. The goal of synthetic biology is to create biological systems with specific functions or properties that do not exist in nature, or to re-engineer existing systems to perform new tasks. Synthetic biology is a rapidly evolving field that plays a crucial role in addressing global challenges such as climate change, food security, and healthcare (129; 130).

Origin of synthetic biology

Synthetic biology has its roots in genetic engineering, which developed in the 1970s and 1980s (131). The development of molecular cloning (132) and polymerase chain reaction (PCR) (133) laid the foundation for genetic munipulation, which later on allowed researchers to manipulate DNA and engineer microbes to produce useful products such as insulin and human growth hormone (134). In the 1990s, DNA sequencing (135) and computational tools enabled the complete sequencing of the E. coli and yeast genomes (136; 16), thus equipping scientists with knowledge to understand and construct biological systems. The term "synthetic biology" was coined in the early 2000s to describe the use of engineering principles to design and construct new biological systems. This approach differed from traditional genetic engineering in its emphasis on the creation of new, synthetic biological parts and systems (137).

Synthetic circuits

The development of synthetic transcriptional networks has been a significant focus of the field of synthetic biology, with various circuit behaviors implemented. These include switches (138; 139; 140), oscillations (141; 142; 143; 144), logic operations (145), filtering (146), sensors (147), memory (148), and noise propagation (149; 150). In 2000, the first synthetic gene regulatory circuit, termed toggle switch, was constructed in *Escherichia coli*. With two repressible promoters arranged in a mutually inhibitory network, the toggle could flip between stable states using transient chemical or thermal induction with a nearly ideal switching threshold (138). At the same time, an oscillating network consisting of three transcriptional repressors, termed repressilator, was constructed in *Escherichia coli* (141). The network is capable of periodically triggering the production of GFP, and monitoring its status in individual cells. This was one of the first examples of synthetic biology, and demonstrated the potential for engineering cells with novel functions. Later on, autoregulatory gene networks with positive (151) and negative feedback loops (152) were engineered to gain stability of the gene networks, by combining theoretical and experimental approachs. As engi-

neering of gene regulatory networks was ongoing, the implementation of synthetic circuits proved to enhance our fundamental understanding of biological systems. This was initiated by studies that investigated the connection between gene expression and molecular noise (149; 150).

Synthetic genomes

In the past two decades, automated and miniaturized oligonucleotide synthesis has led to a considerable decrease in costs to meet the rising demand for DNA synthesis. This development has been significant as it allows researchers to synthesize custom-designed DNA sequences at a lower cost, making it more accessible to the wider scientific community (153). The technology for DNA synthesis has advanced to a stage where it is now feasible to synthesize complete genomes. In 2010, a complete bacterial genome was chemically synthesized and transplanted into a bacterial cell, Mycoplasma mycoides,(154) by J. Craig Venter Institute, demonstrating the potential for creating synthetic organisms with completely artificial DNA sequences. Later on, a minimal bacterial genome, containing only the essential genes required for bacterial life, was created and transplanted into a bacterial cell. The resulting organism, known as JCVI-syn3.0, was the smallest genome capable of supporting bacterial life (155). The development of genome editing technologies, such as CRISPR/Cas9, have revolutionized the field of synthetic biology, making it easier and more efficient to modify and engineer biological systems (156). In 2014, as a collaborative effort between several institutions, a 272-kilobase pair of the yeast genome was successfully synthesized, known as synthetic yeast chromosome III (157), demonstrating the ability to synthesize and assemble a complete chromosome of an eukaryotic organism. This also inspired the development of a much larger-scale synthetic biology project, the Synthetic Yeast 2.0 project (158; 159; 160; 161), which aims to synthesize the entire yeast genome, including all 16 chromosomes of the yeast genome (162; 163; 164; 165; 166; 167; 168; 169; 170; 171). Apart from the efforts of synthesis of individual genomes, a functional single-chromosome yeast was constructed by successive end-to-end chromosome fusions and centromere deletions, which exhibited nearly identical transcriptome and similar phenome profiles as the wild type strains (172).

1.5 Synthetic gene regulatory networks

While there has been progress in the development of engineered organisms with simple gene circuits, most applications do not fully utilize the potential of synthetic biology. There is a significant gap between the low-level genetic circuitry and the possibility of assembling these circuits into more complex gene networks that exhibit robust and predictable behaviors.

Building modular parts

To improve and accelerate the design of synthetic gene networks, one approach is the development of standard biological parts that can be inter-operable and combined to create new biological systems. The BioBricks Foundation, founded in 2003, played a major role in this effort, creating a repository of standardized DNA parts for use in synthetic biology (173; 174). The International Genetically Engineered Machine (iGEM) competition, launched in 2004, has also played a significant role in promoting this effort (175). In recent years, multiple toolkits containing modular parts were developed and rigorously characterized for complex genetic circuit engineering in yeast (176; 177) and mammalian cells (178).

Synthetic orthogonal networks

The goal of synthetic biology is to create predictable and precise control over the behavior of cells through user-designed interactions of biological components, which are often derived from natural sources. However, these components may inadvertently interact with host machinery, particularly within the host central dogma, which can compromise the reliability and predictability of gene circuits. To address this issue, it is necessary to reduce or eliminate these unintended interactions in order to ensure context-independent bioactivities and prevent negative effects on host fitness (179). Orthogonal gene regulatory networks are important tool designed to be independent of endogenous cellular regulatory pathways, and thus can be used to precisely control the expression of target genes without interfering with cellular processes (180).

Zinc-finger transcriptional network

Engineered zinc-finger nucleases (ZFNs) are a type of molecular tool used for genome editing, specifically for inducing targeted double-strand breaks in DNA (181). ZFNs consist of two parts: a zinc finger protein domain and a DNA cleavage domain, which together can be engineered to recognize and cleave specific DNA sequences (182). Zinc fingers (ZFs) are 30 amino acid domains that bind to 3 bps of DNA sequences (183), which was used in natural transcriptional gene networks to recognize DNA and bind to promoter sequences in tandem ways (184). Artificial transcription factors could created by fusing zinc-finger proteins with activation and repression domains

(185; 186). Advancements in the field have enabled intentional modification of the DNA-binding specificity of single ZFs to bind to various 3 bp sequences. These ZFs can then be linked together via covalent bonding to form synthetic, multi-finger arrays that can identify longer DNA sequences with a high level of precision. The OPEN (Oligomerized Pool ENgineering) (187; 188) and context-dependent assembly (coda) methods (189) were developed for rapid construction of multifinger arrays with high efficiency. Using the OPEN platform, Khalil et al. developed a library of orthogonal synthetic transcription factors (sTFs) in yeast. Theys utilized synthetic transcription factors (sTFs) with modular DNA-binding domains and artificial zinc fingers to construct advanced synthetic transcriptional circuits. By systematically adjusting key component properties, including DNA specificity, affinity, promoter design, and protein-protein interactions, the authors were able to create circuits with adjustable output strength and transcriptional cooperativity (190).

1.6   The objective of this thesis

The overall objective of thesis comprise four parts: 1) Quantitative characterization of the yeast inorganic phosphate gene regulatory network (GRN) at the single cell level; 2) Systematic perturbation of the phosphate GRN using a Z3EV-mediated inducible system to determine the robustness of the GRN; 3) Bottom up assembly of an orthogonal GRN with synthetic transcription factor and promoters on a yeast artificial chromosome; 4) Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo.

Chapter 2 discusses the efforts made to characterize of the yeast phosphate GRN. A library of 81 strains was built, each carrying a red fluorescent protein-tagged Pho4 and twenty-seven green fluorescent protein-tagged Pho4-regulated promoters. The input-output level of the Pho4 regulatory network, including the expression level of Pho4 and these Pho4-regulated promoters, was quantitatively characterized. In addition, the nuclear localization of the master regulator Pho4 as a function of the extracellular inorganic phosphate level was studied. To enable the high-throughput measurement, two state-of-the-art microfluidic chips were used, including a dilution-generator-integrated Microchemostat Array (MCA) device and a MICROfluidic S. cerevisiae Trap ARray (MICROSTAR) device.

Chapter 3 discusses the robustness characterization of the yeast phosphate GRN. A Z3EV-mediated inducible phosphate regulatory network was developed to better tune the expression level of Pho4. The expression level of Pho4 was tuned by adding different concentrations of beta-estradiol, which is not consumed by yeast cells. The output levels of regulated promoters were investigated with different Pho4 titrations. Additionally, a library of 12 Pho4 variants was created, where the basic binding domain of Pho4 was mutated with different binding affinities and then integrated with 23 phosphate-responsive promoter cassettes. The input-output levels of these two libraries were preliminarily studied using the plate reader and FACS.

Chapter 4 discusses the efforts of bottom-up assembly of an orthogonal GRN with synthetic transcription factors and promoters on a yeast artificial chromosome. A set of synthetic promoters regulated by a synthetic transcription factor (TF) was designed. The synthetic transcription factor consists of separate domains for different molecular functions: i) a synthetic Zn-finger domain enables binding to DNA at specific sequences embedded within the engineered promoters; ii) a protein-protein interaction domain enables cooperative interactions; and iii) the phosphorylation sites that encode regulation of nuclear import and export which will come from Pho4. The expression level of the synthetic TF is tuned by a Z3EV-mediated inducible system.

The synthetic promoters were engineered by replacing the binding motifs in the scaffold promoters (CYC1, Pho5, Pho84, Pho81, Pho8pr, etc.) with Zinc-Finger binding motifs. The synthetic TF and ZF promoters were then integrated into the native loci and the yeast artificial chromosome for functional tests.

Chapter 5 discusses the systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and *in vivo*. We developed an inverted MITOMI assay (iMITOMI), by adapting the original MITOMI platform (45) to generate quantitative measurements of transcription factors binding at equilibrium to longer DNA targets containing binding site clusters of up to 6 distinct binding sites (191). To assess whether the observed *in vitro* occupancies translate to transcriptional activation *in vivo*, we tested low-affinity binding site clusters in a synthetic and native gene regulatory network in *S. cerevisiae*.

Chapter 6 briefly summarizes and discusses the thesis work and the future work outlook.

2 Quantitative characterization of yeast inorganic phosphate gene regulatory network

Chapter 2. Quantitative characterization of yeast inorganic phosphate gene regulatory network

This chapter is adapted from an article published in *PNAS* 2023.

Authors: Hon Ming Yip*, Shiyu Cheng*, Evan J. Olson*, Michael Crone, Sebastian J. Maerkl (*equal contributions)

Contribution: S.C., M.C., and E.J.O. generated the strains used in this study. H.M.Y. designed the microfluidic devices and performed all on-chip experiments. E.J.O. wrote the image analysis pipeline. S.J.M. conceived and directed the study.

2.1   Abstract

Cells cope with and adapt to ever-changing environmental conditions. Sophisticated regulatory networks allow cells to adjust to these fluctuating environments. One such archetypal system is the *S. cerevisiae* Pho regulon. When external inorganic phosphate (P_i) concentration is low, the Pho regulon activates, expressing genes that scavenge external and internal P_i. However, the precise mechanism controlling this regulon remains elusive. We conducted a systems analysis of the Pho regulon on the single cell level under well-controlled environmental conditions. This analysis identified a robust, perfectly adapted Pho regulon state in intermediate P_i conditions, and we discovered a hitherto unknown intermediate nuclear localization state of the transcriptional master regulator Pho4p. The existence of an intermediate nuclear Pho4p state unifies and resolves outstanding incongruities associated with the Pho regulon, explains the observed programmatic states of the Pho regulon, and improves our general understanding of how nature evolves and controls sophisticated gene regulatory networks. We further propose that robustness and perfect adaptation are not achieved through complex network-centric control, but by simple transport biophysics. The ubiquity of multi-transporter systems suggests that similar mechanisms could govern the function of other regulatory networks as well.

2.2   Introduction

Cellular regulatory complexity evolved to allow cells to adapt to fluctuating environmental conditions. Extracellular nutrient availability changes over time and cells cope with these fluctuations by adapting their gene expression patterns. Inorganic phosphate (P_i) is one such essential nutrient for which organisms evolved a sophisticated regulatory network (192). In the budding yeast *S. cerevisiae*, P_i is sensed internally by the cyclin-dependent kinase complex Pho80p and Pho85p. Pho80p/Pho85p are located in the nucleus and transduce P_i levels to their target, the transcriptional master regulator Pho4p (107; 193; 194) through phosphorylation (108; 109). The phosphorylation state of Pho4p in turn determines whether Pho4p is localized to the nucleus or the cytoplasm. When Pho4p is fully phosphorylated it is localized to the cytoplasm and inactive, whereas it becomes nuclear localized when it is dephosphorylated. Nuclear localization of Pho4p leads to the expression of Pho regulon genes including phosphate transporters, phosphatases, and vacuolar storage regulators. Expression of these genes increases phosphate import and maintains intracellular P_i levels.

This raised the question of how the Pho regulon can activate without entering a futile oscillation cycle in which a P_i flux increase due to Pho-regulon activation leads to an increase in internal P_i, in turn raising P_i levels again above the critical activation threshold and turning the system off, just to be activated once again when the Pho-regulon gene products have been diluted out. The current Pho-regulon model therefore stipulates and necessitates the existence of one or more feedback loops that act to decrease internal P_i concentrations. These negative feedback loops are thought to be Spl2p which binds and inactivates low-affinity transporters (118), and Vtc1-4p which are thought to decrease internal P_i by sequestering and storing P_i in the vacuole (123).

The current model assumes that P_i influx rate depends on extracellular P_i concentration, which in turn means that internal P_i levels vary over a continuum of concentrations (123; 117). Varying internal P_i concentrations requires a sophisticated control mechanism that constantly adjusts the relative strength of the negative feedback loop (195), which is feasible and has been described in other systems such as osmoregulation (196) and tunable gene expression in the Gal system (197). Our current understanding of Pho4p also suggests that nuclear Pho4p concentration should track internal P_i concentration in order to achieve the required gene expression tuning (195; 111). But so far, no differences in nuclear-localized Pho4p levels have been observed and no intermediate states have been identified. It is also counter-intuitive that the model requires the existence of feedback loops that counteract the function of the Pho regulon, which is to increase P_i flux, not decrease it. And finally, one of the main

feedback loops suggested to be Vtc1-4p should become inactive under prolonged intermediate phosphate starvation, as P_i sequestration into the vacuole has to cease once vacuoles have reached capacity. Many questions therefore remain in regards to the mechanistic details of how the Pho regulon functions.

We observed that the Pho regulon can attain three clearly defined states as opposed to a continuum of states: an "off" state, a "plateau" state, and a fully activated "on" state. In the off state, when extracellular P_i concentrations are high, all Pho regulon genes are off. The plateau state is a previously observed but hitherto unexplained state in which the entire population of cells are partially activating the Pho regulon (118). We call this the plateau state because we found that this state is robust to differences in extracellular P_i concentration. The off and plateau states are linked by a bistable region (118). The third and final state is the fully activated Pho regulon and is observed in conditions of very low or depleted extracelullar P_i. We were also able to quantify nuclear Pho4p localization in these states. Pho4p is highly nuclear localized in the on state, as previously described. But we discovered that Pho4p is partially localized to the nucleus in the plateau state, and that the nuclear concentration of Pho4p in the plateau state is also robust and insensitive to differences in extracellular P_i concentration ($[eP_i]$). The existence of an intermediate Pho4 nuclear state resolves current incongruities and readily explains the observed Pho-regulon behavior. The observation of perfect adaptation and robustness in the plateau state led us to suggest that *S. cerevisiae* uses a different control mechanism than previously thought. Departing from the generally adopted regulatory network-centric dogma, we suggest that the Pho regulon is governed by simple transporter biophysics.

2.3 Results

2.3.1 Strain library construction

To conduct a systems-level analysis of the Pho regulon, we chose 23 genes known or presumed to be regulated by Pho4p (Figure 1.1, Table 1.1). The promoters of these 23 target genes were cloned upstream of yeast optimized GFP (yoGFP) and integrated into the *LYS2* locus (Figure 2.1). We created an "LF" library in which all cassettes use the same generic terminator ADH1te, and an "NN" combined readout / deletion library by replacing these genes in their native loci with yoGFP. To enable characterization of Pho4p localization, we generated a yomScarlet-i - Pho4p fusion in the endogenous locus to preserve native expression levels. These libraries were constructed by using the Cas9-assisted cloning by homologous recombination (Figure 2.2) (176). Under eP_i rich condition, Pho4 should localize to the cytoplasm and the promoter should express at their basal levels. Under eP_i free condition, Pho4 is expected to translocate to the nucleus and fully activate the promoters. However, the behaviours under intermediate eP_i conditions are unclear and are also an outstanding question in the field. We did not observe large differences in target gene expression between the native Pho4p and the N-terminally tagged version, indicating that the fusion protein did not considerably alter Pho4p function (Figure 2.3).
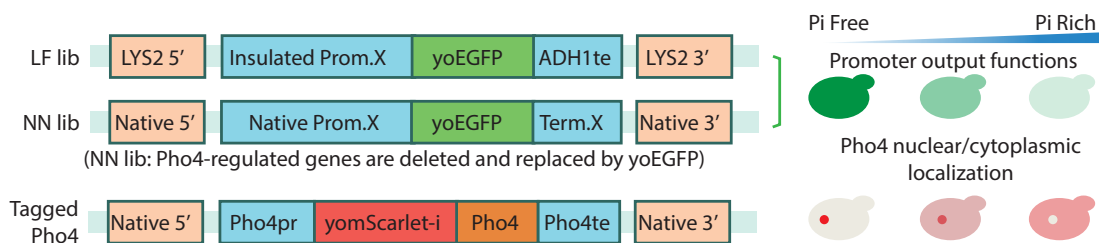
Figure 2.1: **Strain library design.** We generated two libraries: LF and NN. The LF library has 23 of the Pho regulon promoters cloned in front of yoGFP and all cassettes are terminated by a common terminator. The NN library has the native genes replaced with yoGFP leading to a combined deletion / reporter library. We also generated a yomScarlet-i-Pho4p strain to track Pho4p localization on the single cell level. Right hand side shows the expected response under three media conditions.
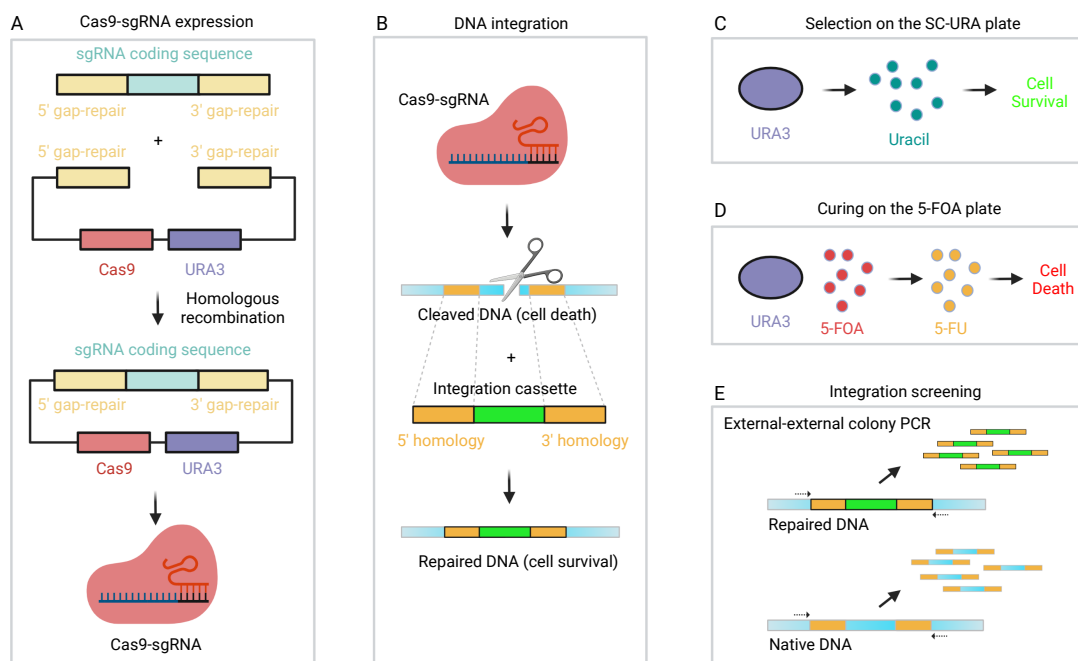
Figure 2.2: **Cas9-assisted cloning by homologous recombination. a,** The sgRNA coding sequence cassette and Cas9-URA3 cassette were digested with EcoRV and BsmBI, respectively, for yeast transformation. After transformation, they combined to form a plasmid and expressed the Cas9-sgRNA complex to target the integration loci. **b,** The DNA insertion cassette was digested with NotI for yeast transformation. The native genomic DNA was cleaved by the Cas9-sgRNA complex to create a double stranded break, which could lead to cell death. The integration cassette was inserted into the target loci via Cas9-assisted homologous recombination and the genomic DNA was repaired. **c,** The resultant transformants were selected on SC-URA plates. Yeast cells containing Cas9-URA3 plasmid could produce uracil and thus survive on the SC-URA plate. **d,** The colonies on the SC-URA plates were cured on 5-FOA plates to remove the Cas9-URA3 plasmid under selection pressure. 5-FOA can be converted into 5-FU when URA3 is expressed leading to cell death. **e,** Colonies on the 5-FOA plates were screened by external-external colony PCR and the PCR products were verified by Sanger sequencing.
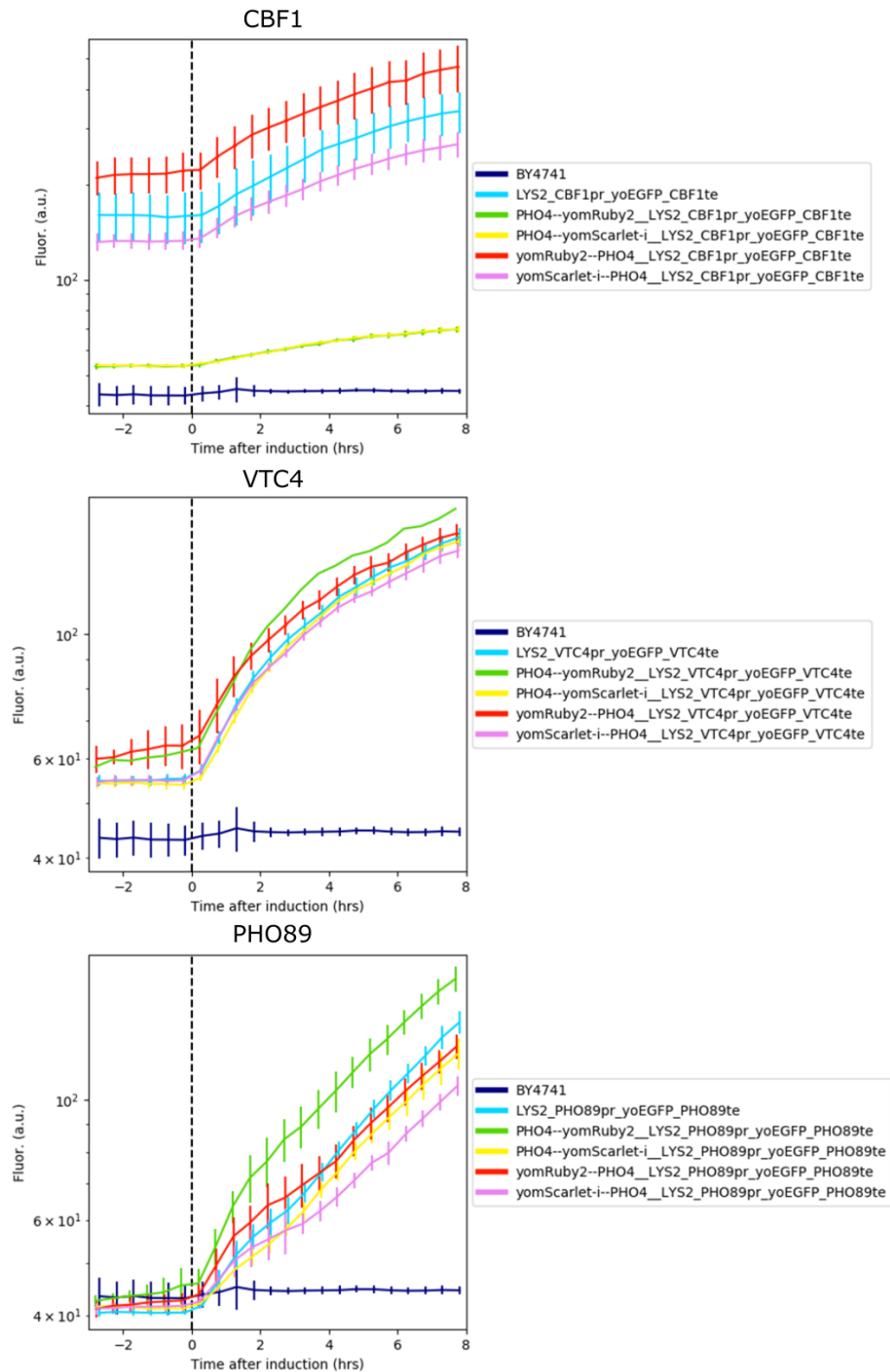
Figure 2.3: **Assessment of fluorescently tagged Pho4p variants.** Different fluorescent Pho4p fusion variants were tested for their ability to activate 3 target promoters: CBF1, VTC4, and PHO89 and compared to wild type Pho4p.

2.3.2 Microfluidic platforms for single cell characterization

Two state-of-art microfluidic devices were developed by Hon Ming Yip to enable the single cell characterization: 1) the microchemostat array with a dilution generator (MCA-DG) and 2) MICROfluidic S. cerevisiae Trap ARray (MICROSTAR) device. These two platforms support testing of a large number of strains and media conditions. Also, a 256-chamber microchemostat array (256-MCA) was designed based on a previously developed high-throughput MCA (81; 91), and used for smaller scale, proof-of-concept experiments.

MCA device

The 256-MCA device was used to validate the strains from the promoter characterisation library and the strain for transcription factor localisation measurement. There are two sets of experiments that were done using 256-MCA. The first is a characterisation of the LF library under full starvation. The second is an evaluation of the influence of the red fluorescent protein yomScarlet-i and yomRuby2 fusion on the transcription factor Pho4. For the strain library validation, the goal is to see if the strains are induced under phosphate starvation and to confirm we are able to extract single cell data from the images using the image analysis pipeline (see section 2.3.3). Although the strain throughput is lower than the original MCA, it is sufficient for testing all of the strains in the libraries with multiple replicates in a single experiment. The experiment phase is six hours of pre-starvation, and 16 hours of starvation. Only full starvation was performed in these experiments. For large-scale characterization of all strains, we employed a previously developed high-throughput MCA, that allows characterization of 1,152 yeast strains in parallel (81; 91). We added a microfluidic dilution generator (DG) to automatically generate different extracellular P_i concentrations (Figure 2.4, Figure 2.5) (100). The original microchemostat array was divided into 12 zones and each zone is connected to an outlet of a dilution generator in order to measure 96 strains under 12 Pi concentrations. The device contains 1) 12 microchemostat sub-arrays, 2) a ten-output dilution generator, and 3) two independent media zones for positive and negative controls. To generate different concentrations, two media formulations are introduced from two different inlets, which are combined in the dilution generator (DG) to generate 10 unique output formulations. The dilution generator can also be used with a single medium inlet, in which case all DG output channel media conditions are identical and equal to the media input.

Figure 2.4: **Microchemostat array (MCA) chip design.**

The figure legend content:

Flow layer
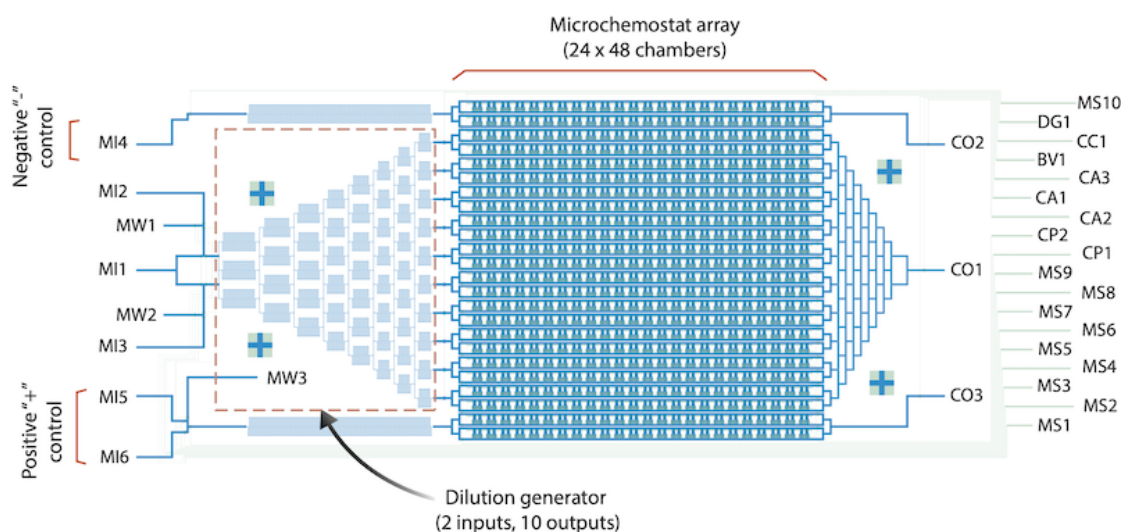
MI1 A media inlet of DG and typically feeding Pi-rich media.
MI2 A media inlet of DG and typically feeding Pi-high media.
MI3 A media inlet of DG and typically feeding Pi-low/Pi-free media.
MW1 A media outlet for "MI1" or "MI2" and is used during media switch to clear dead volume and air bubble.
MW2 A media outlet for "MI1" or "MI3" and is used during media switch to clear dead volume and air bubble.
MI4 A media inlet of "-" control and typically feeding Pi-rich media.
MI5 A media inlet of "+" control and typically feeding Pi-free media.
MI6 A media inlet of "+" control and typically feeding Pi-rich media.
MW3 A media outlet for "MI5" or "MI6" and is used during media switch to clear dead volume and air bubble.
CO1 The exit for cells and used media during on-chip cell culture in DG zone.
CO2 The exit for cells and used media during on-chip cell culture in "-" zone.
CO3 The exit for cells and used media during on-chip cell culture in "+" zone.

Control layer

MS1 Control "MI1" and is typically used in combination with "MS 1-5" during media switch.
MS2 Control the inlet of DG zone and is typically used in combination with "MS 1-5" during media switch.
MS3 Control all "MW" and is typically used in combination with "MS 1-5" during media switch.
MS4 Control "MI2" and "MI3", and is typically used in combination with "MS 1-5" during media switch.
MS5 Control "MI2", "MI3" and all "MW", and is typically used in combination with "MS 1-5" during media switch.
MS6 Control "MW3" and is typically used in combination with "MS 6-9" during media switch.
MS7 Control "MI5" and is typically used in combination with "MS 6-9" during media switch.
MS8 Control "MI6" and is typically used in combination with "MS 6-9" during media switch.
MS9 Control the inlet of "+" control and is typically used in combination with "MS 6-9" during media switch.
MS10 Control "MI4" and hence the inlet of "-" control.
CP1 Control the lower flow channel in the chamber array.
CP2 Control the upper flow channel in the chamber array.
DG1 Control the channels connecting DG and MCA, and typically closed to clear air bubbles in the DG during chip priming.
CA1 Control the media and cell exit (CO1), and typically closed to clear air bubbles in the cell trap array during chip priming.
CA2 Control the media and cell exit (CO2), and typically closed to clear air bubbles in the cell trap array during chip priming.
CA3 Control the media and cell exit (CO3), and typically closed to clear air bubbles in the cell trap array during chip priming.
BV1 Gently press the cell culture chamber to guide cell grow in monolayer.
CC1 Control the outlet of chambers to prevent cross-contamination during chip priming and initial growth.
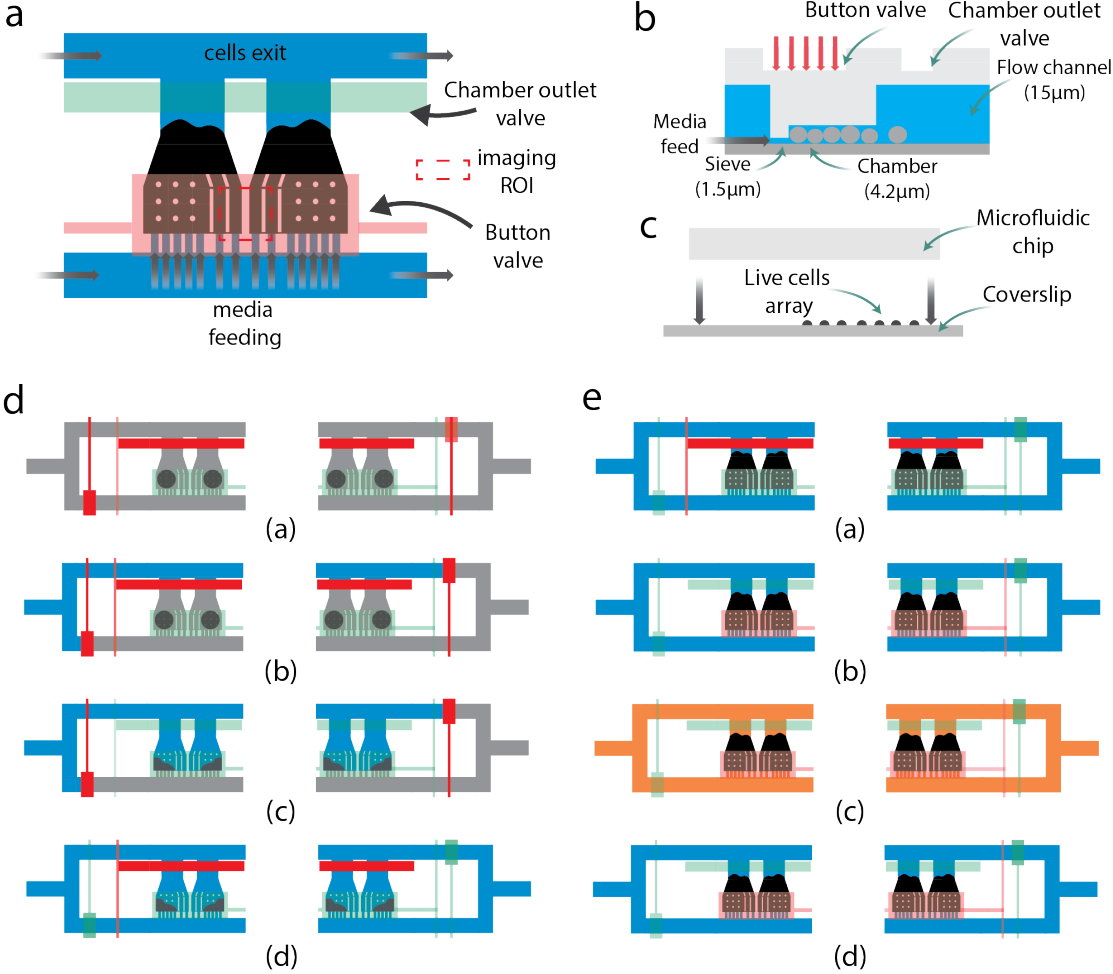
Figure 2.5: **MCA experimental workflow.**

MICROSTAR device

In order to achieve more precise control over growth and environmental conditions we engineered a MICROSTAR (MICROfluidic *S. cerevisiae* Trapping ARray) device, on which yeast cultures are confined to narrow, independent culturing grooves roughly two cell diameters in width (Figure 2.6, Figure 2.7, Figure 2.26). This design is similar to the "mother-machine" for bacterial cultures (198), but with the added feature that cells are perfused from both ends of the groove, while permitting cell exit only from the top. The MICROSTAR device allows 16 strains to be analyzed under eight media conditions for 128 experiments per device. The MICROSTAR device consists of 256 cell culture areas, each area in turn contains 8 parallel culture grooves 8 μm in width or roughly 2 *S. cerevisiae* cell diameters (Figure 2.6). The device contains 2,048 cell grooves in total. Each groove is 3.8 μm in height to force cells to grow in monolayers. Each culture groove is connected to two parallel media perfusion channels. On the bottom of each groove, we created a narrow sieve channel that is 1.2 μm in height, 11 μm wide, and 5.5 μm long. The cell-traps array contains 16 columns and 16 rows. Each column can be individually loaded with different yeast strains. The 16 rows in turn are divided into 8 individually addressable rows to provide different media conditions, generated by an upstream dilution generator.
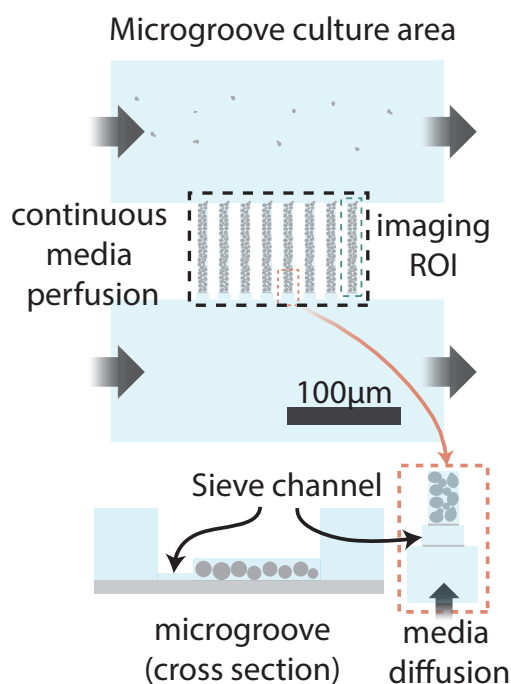
Figure 2.6: **MICROSTAR device.** The MICROSTAR cultures cells in narrow grooves, with the bottom of each groove connected to a media supply channel via a narrow groove.
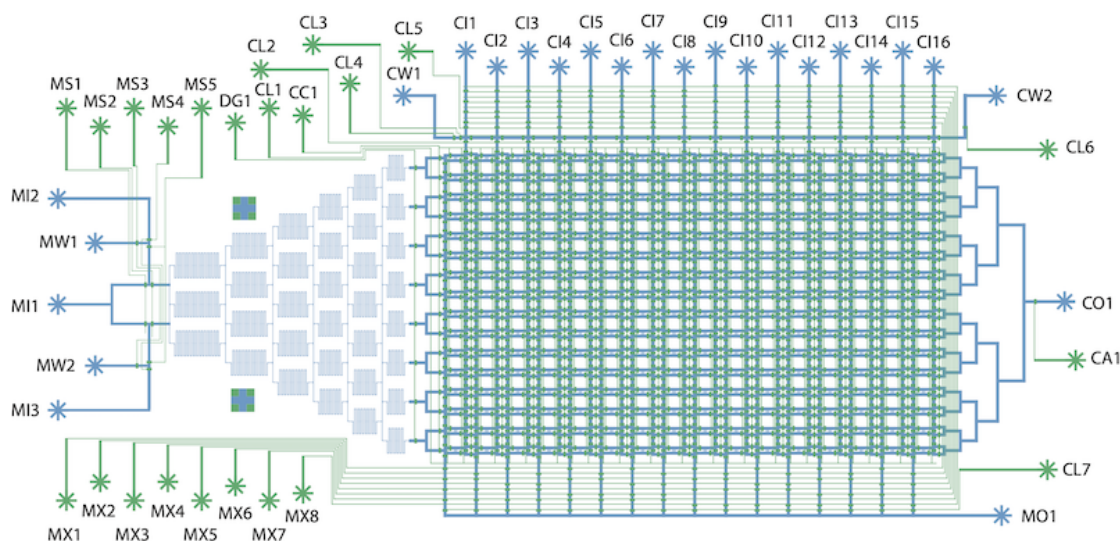
Flow layer

MI1	A media inlet and typically feeding Pi-rich media.
MI2	A media inlet and typically feeding Pi-high media.
MI3	A media inlet and typically feeding Pi-low/Pi-free media.
MW1	A media outlet and is used during media switch to clear dead volume and air bubble.
MW2	A media outlet and is used during media switch to clear dead volume and air bubble.
CO1	An exit for cells and used media during on-chip cell culture.
MO1	An exit for used media during cell loading process.
CI1-16	16 cell loading inlets.
CW1	A media inlet/outlet for optional washing operation
CW2	A media inlet/outlet for optional washing operation

Control layer

MS1	Control "MI1" and is typically used in combination with other "MS" during media switch.
MS2	Control all "MI" and is typically used in combination with other "MS" during media switch.
MS3	Control all "MW" and is typically used in combination with other "MS" during media switch.
MS4	Control "MI2"and "MI3", and is typically used in combination with other "MS" during media switch.
MS5	Control "MI2", "MI3" and all "MW", and is typically used in combination with other "MS" during media switch.
DG1	Control the accessibility of media into the cell trap array and typically closed to clear air bubbles in the DG during chip priming.
CL1	Disconnect the columns of cell trap array and is typically used during cell loading.
CL2	Disconnect the cell trap array and "CI1-16", and is typically used during cell loading.
CC1	Disconnect the rows of cell trap array and is typically used during on-chip cell culture.
CL3	Disconnect "CI1-16" and is typically used during cell loading.
CL4	Control "CW1" and is typically used during media switch.
CL5	Control "CI1-16" and is typically used during cell loading.
CL6	Control "CW2" and is typically used during media switch.
CA1	Control the media and cell exit (CO1), and typically closed to clear air bubbles in the cell trap array during chip priming.
CL7	Disconnect the cell trap array and "MO1", and is typically used during chip priming.
MX1-8	Control the accessibility of 16 cell loading paths using a combinatorial operation.

Figure 2.7: **MICROSTAR chip design.**

The DG on MICROSTAR generates eight media conditions including the high and low media inputs (Figure 2.8, 2.27, 2.31, 2.32).

Diluation generator output:
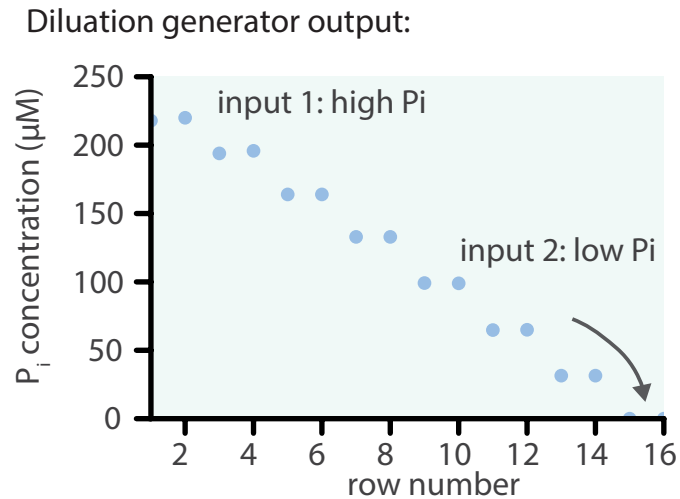
Figure 2.8: **Dilution generator output.** Both the MCA and MICROSTAR devices are equipped with an on-chip dilution generator to automatically generate 10 and 8 P_i media conditions, respectively.

Media conditions can be changed on the fly so that a standard experiment initially cultures cells under nominal, rich P_i conditions for six hours, followed by a switch to eight different P_i concentrations for 16 hours, followed by a 16 hour recovery period in rich conditions (Figure 2.9).
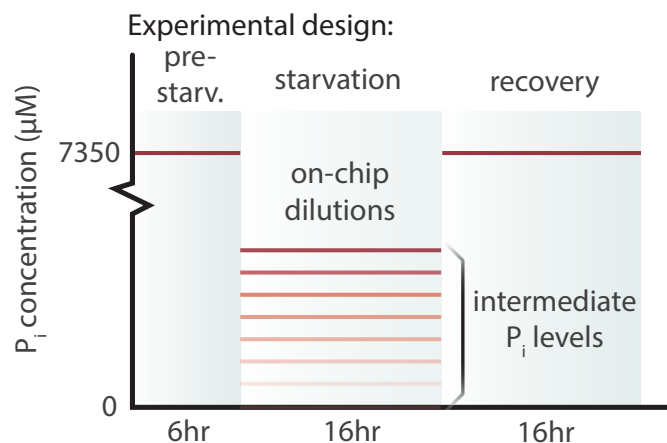
Figure 2.9: **Experimental design.** Automated media switching enables complex experiments: growing strains under rich media conditions, followed by a switch to different P_i starvation conditions, and a switch back to rich conditions.

2.3.3   Single cell analysis pipeline

The devices are imaged on an automated microscope with a time resolution of 20-30 minutes (Figure 2.10A-F). After the acquisition of the fluorescence and phase contrast images, single cell data were then extracted from them. A single cell analysis pipeline (detailed in Section 2.5.6) was developed by Evan J. Olson to extract the data and perform the quantitative analysis. We can therefore follow promoter activity (Fig. 2.10G), or assess Pho4p nuclear localization (Fig. 2.10H).
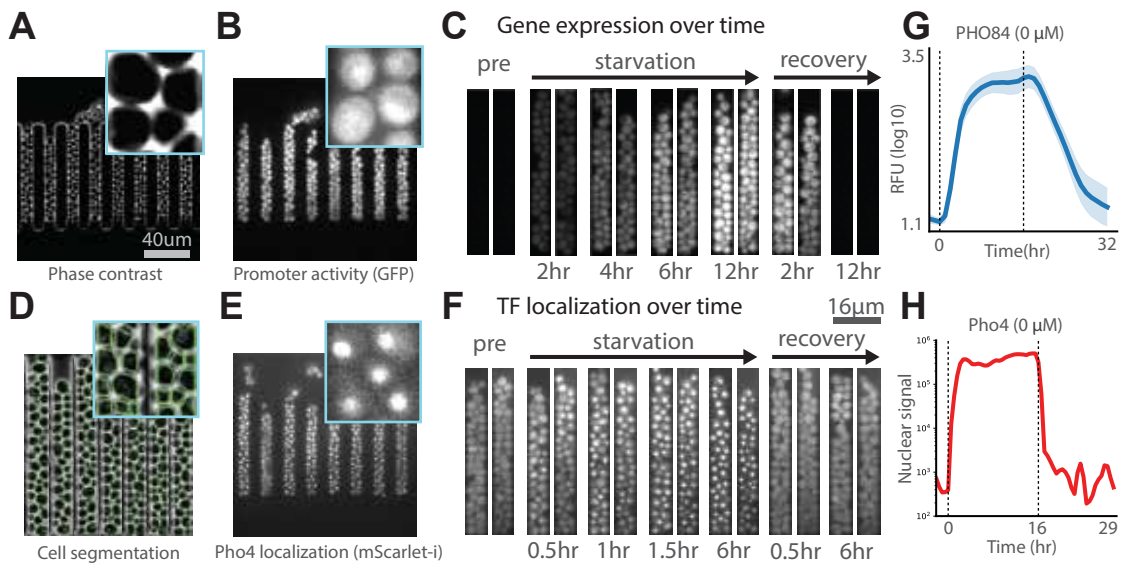
Figure 2.10: **Single cell analysis pipeline.** Using time-lapse imaging with high spatio-temporal resolutions returns information on promoter activation and Pho4p nuclear localization.

2.3.4   Systems analysis of the Pho regulon

To obtain a systems-level overview of the possible states the Pho regulon can attain as a function of $[eP_i]$, we characterized the activity of all 23 promoter strains across both libraries, LF and NN, on the MCA device (Figure 2.11, 2.12) and analyzed all LF strains on the MICROSTAR device (Figure 2.13, 2.14).
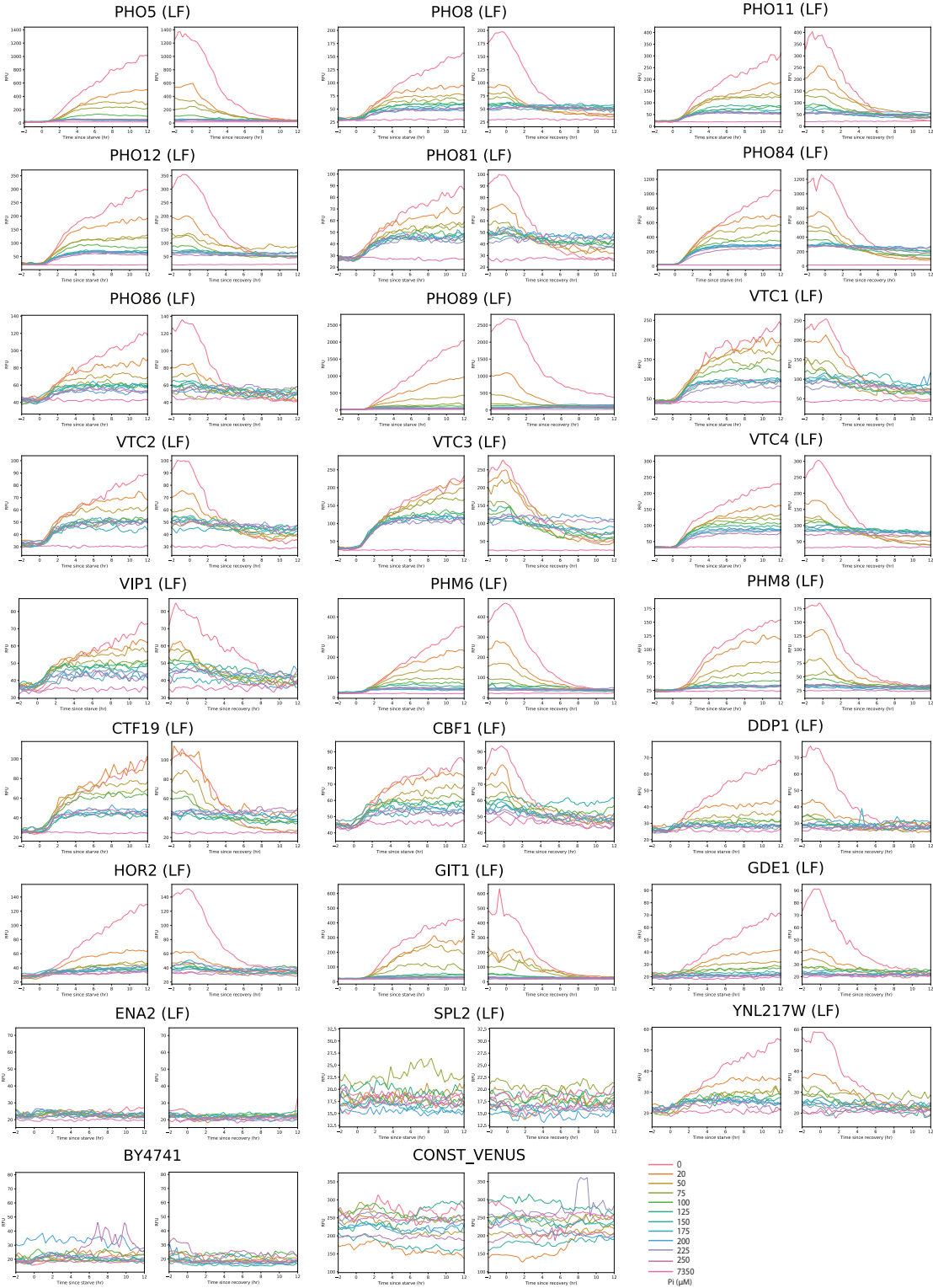
Figure 2.11: **Promoter expression over time at different P_i concentrations for LF library strains measured on MCA.**
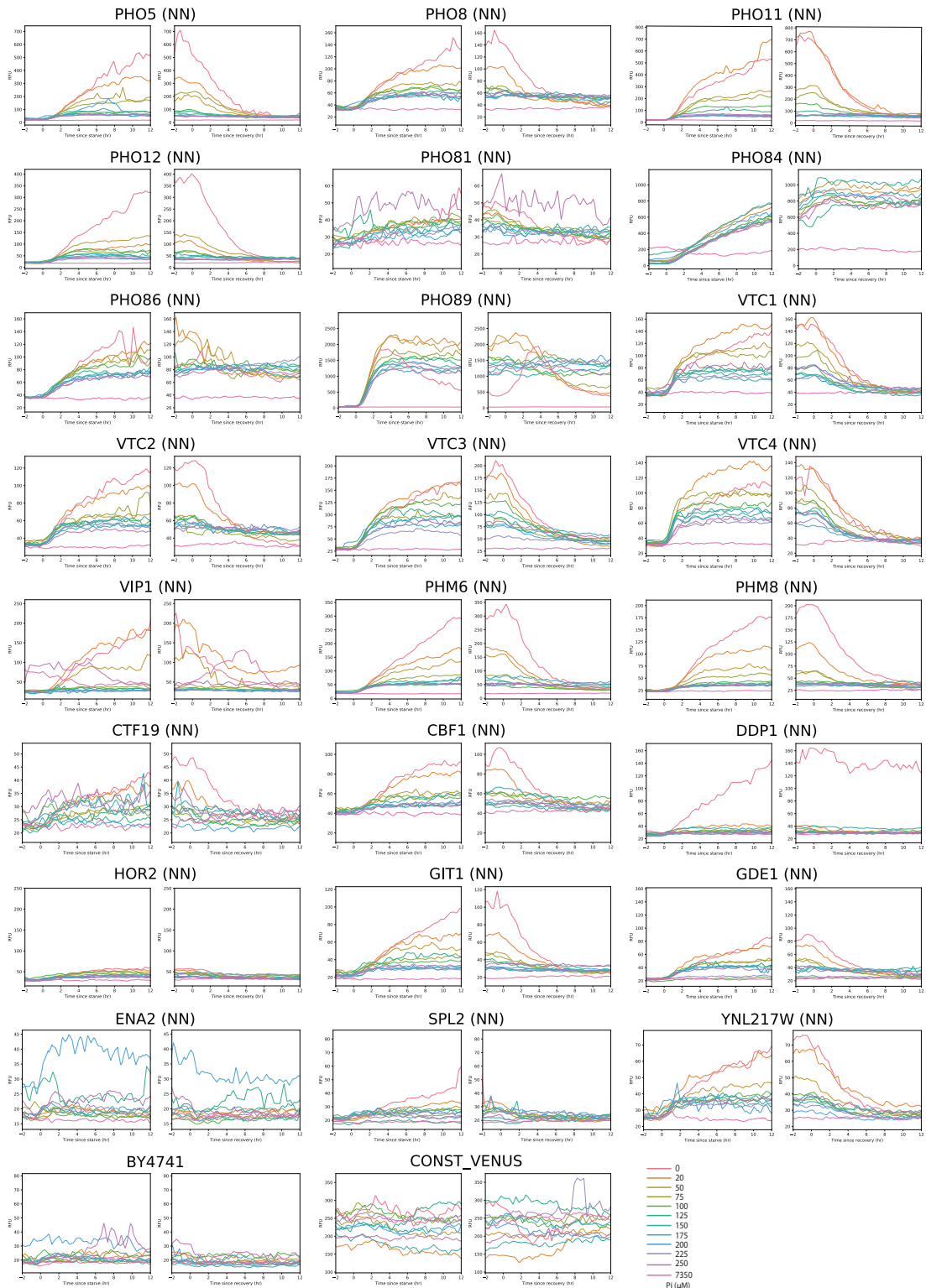
Figure 2.12: **Promoter expression over time at different P_i concentrations for NN library strains measured on MCA.**
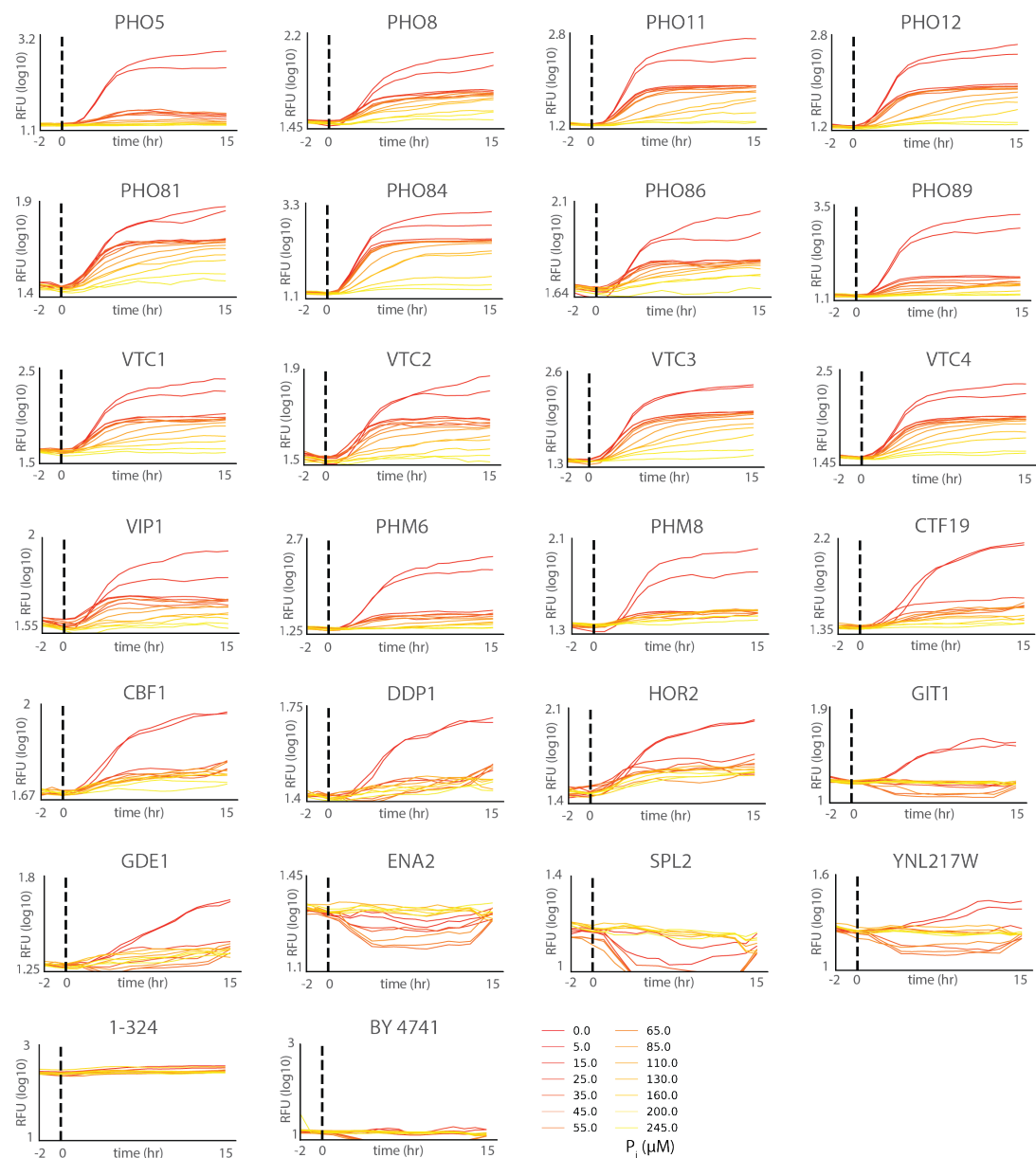
Figure 2.13: **Promoter expression over time at different P_i concentrations for all strains measured on MICROSTAR.**
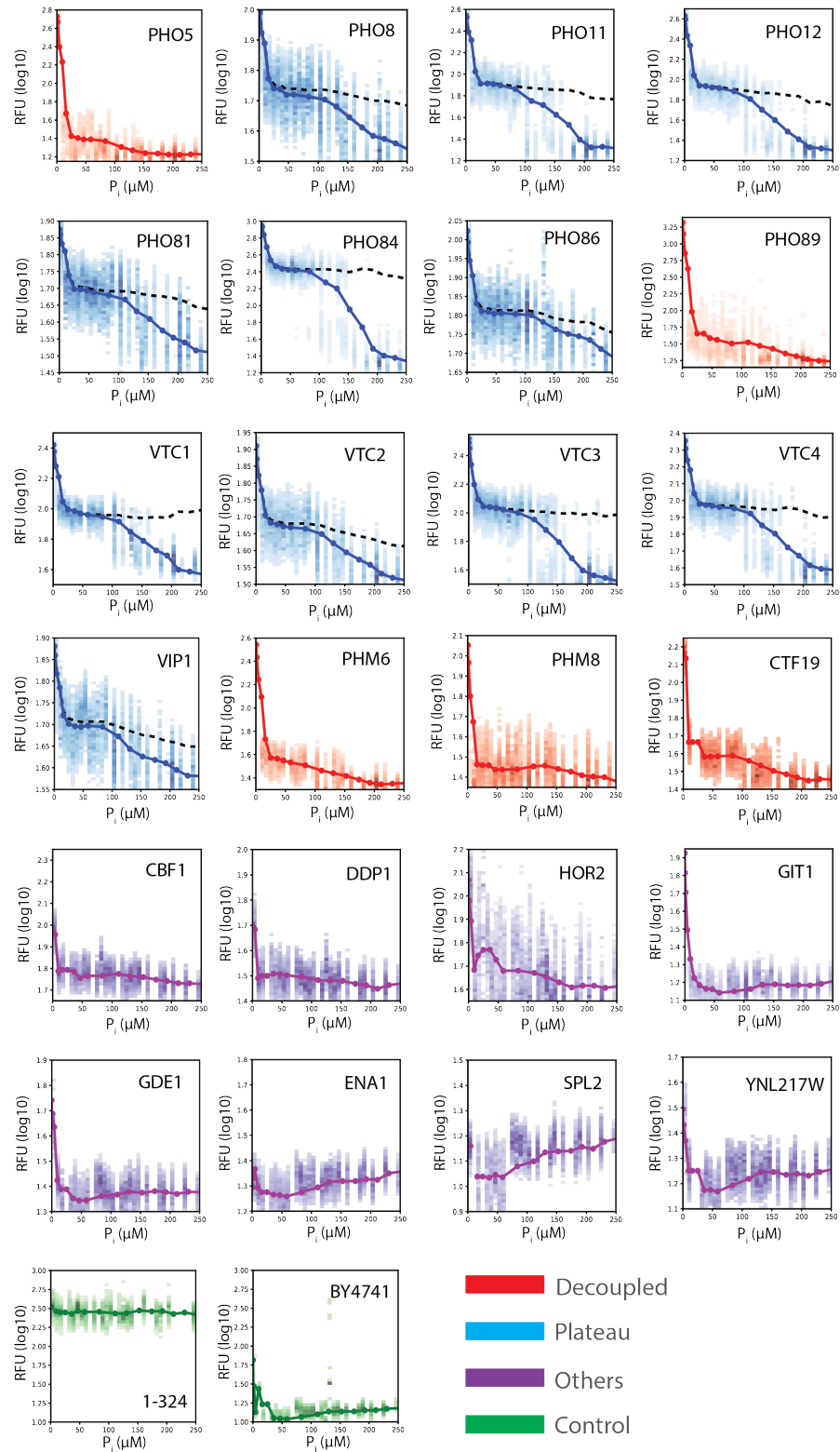
Figure 2.14: **Promoter expression versus P_i for all strains measured on the MI-CROSTAR device.**

We were able to map the precise output function of the Pho regulon under defined steady-state conditions. This showed that, rather than a plethora of possible promoter classes and output states, Pho-regulated promoters fall into two classes (Figure 2.15).
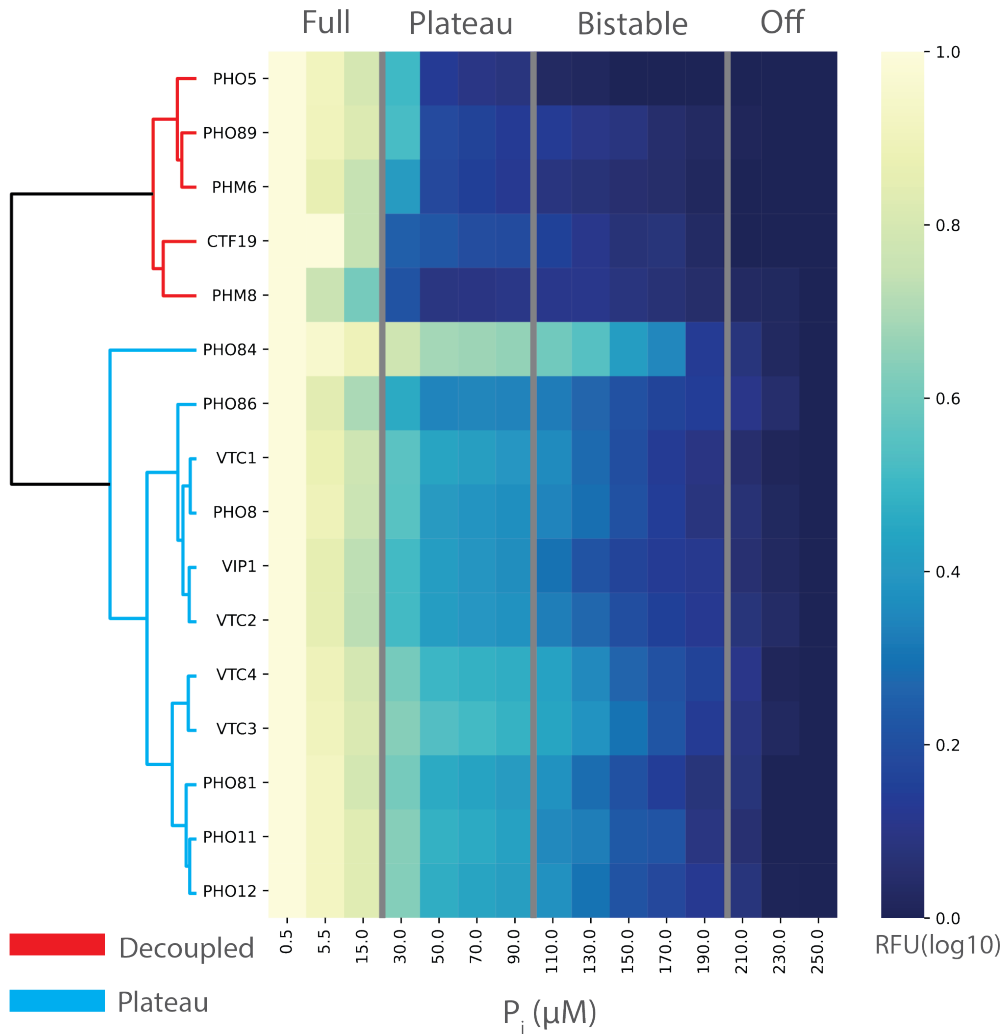
Figure 2.15: **Normalised promoter expression across P_i levels and hierarchical clustering.** Normalised promoter activity levels of PHO promoters are plotted in a heatmap. The activation level is normalised by its maximum and minimum levels. The hierarchical tree on the left shows the clustering result. The promoters in red are clustered as decoupled promoters, and the promoters in blue are clustered as plateau promoters.

One class are the previously described decoupled promoters including *PHO5* (Figure 2.16a-b) (199). Decoupled promoters show no activity in conditions above 200 μM eP_i, very slight activation in intermediate eP_i ranges of 20-200 μM, and very high activation below 20 μM eP_i. The second, and major, class of promoters to which *PHO84* belongs, we call "plateau" promoters (Figure 2.16c-h). Plateau promoters remain inactive above 200 μM eP_i, but enter an intermediate activation state in the range of 20-200 μM eP_i. Strikingly, these promoters exhibit a surprisingly invariable expression level over an order of magnitude range of $[eP_i]$. In other words, these promoters exhibit perfect adaptation and robust control. When $[eP_i]$ fell below 20 μM, expression of these promoters increased further. It should be noted that even decoupled promoters display a low-level plateau behaviour.
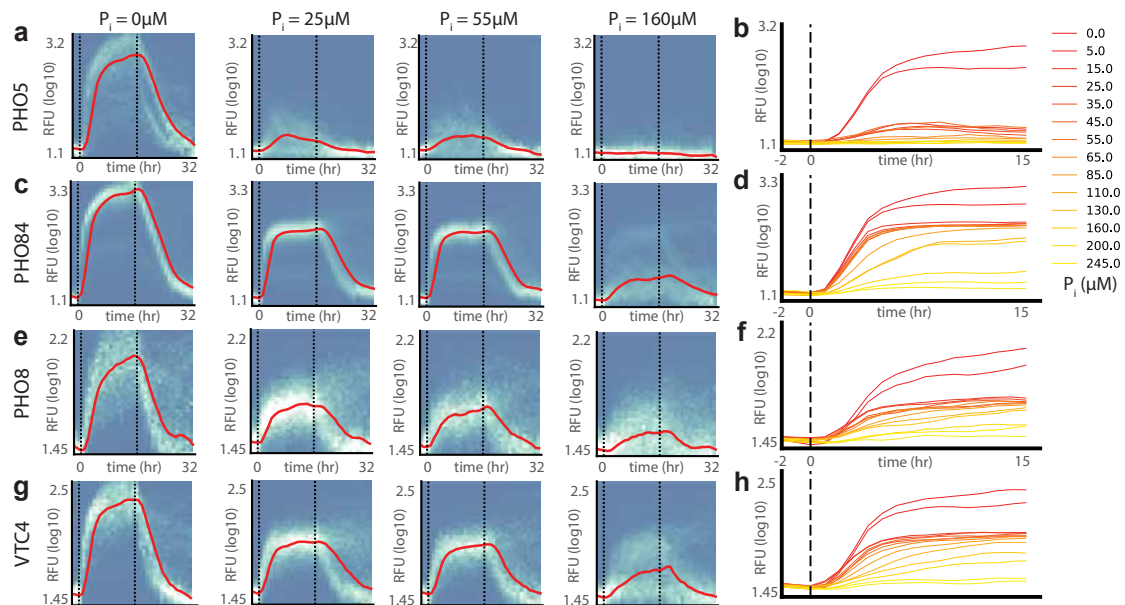
Figure 2.16: **Promoter activity dynamics as a function of [eP_i].** **a-b,** *PHO5,* **c-d,** *PHO84,* **e-f,** *PHO8,* and **g-h,** *VTC4* promoter activity measured on a MICROSTAR device over time and different $[eP_i]$. **a, c, e, g,** Vertical dashed lines indicate when media switches occurred. Graphs show single cell promoter activity distributions as a heat map, and moving median intensities are indicated by red lines. **b, d, f, h,** Median promoter activity levels for all P_i concentrations measured.

The Pho regulon can therefore attain three distinct states or programs (Figure 2.17). When $[eP_i]$ is above 200 μM the entire Pho regulon remains in the off state. In a relatively narrow range of 100-200 μM eP_i the Pho regulon exhibits bistability, where cells can be either in the off or plateau state (118). In the range of 20 - 100 μM eP_i the Pho regulon exhibits a high degree of robustness in that promoter activity is largely independent of $[eP_i]$. This plateau region functionally extends all the way to the upper end of the bistable region at 200 μM eP_i. At $[eP_i]$ below 20 μM the Pho regulon is fully activated, with decoupled promoters turning on and plateau promoters further increasing in activity above their plateau levels.
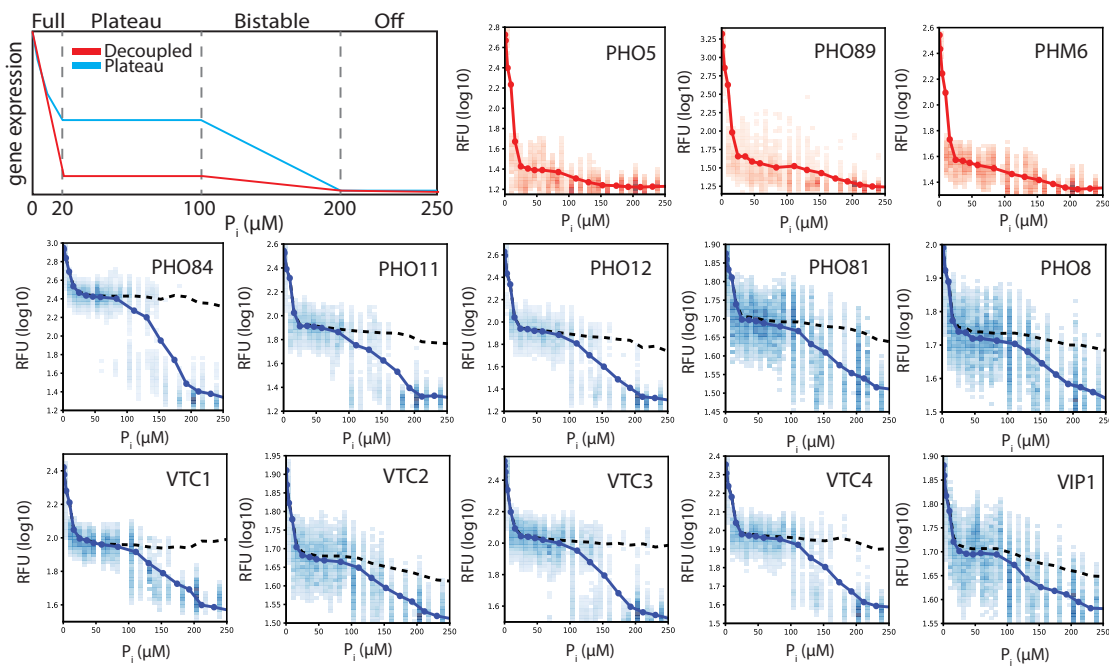
Figure 2.17: **Classification of the Pho regulon promoters.** Pho regulon promoters can be classified into "decoupled" and "plateau" promoters. **Red,** Decoupled promoters: *PHO5, PHO89,* and *PHM6*. **Blue,** Plateau promoters: *PHO8, PHO11, PHO12, PHO81, PHO84, VTC1-4,* and *VIP1*. Single cell intensity distributions are plotted as heat maps, and solid lines show the median value. Dotted lines indicate median values above a certain threshold to capture only activated cells.

The relative activation levels of the PHO promoters in the four programs are also analyzed and shown as pie charts (Figure 2.18). The four pie charts show the activated contribution of individual promoters and clusters in the four programs. The bar chart shows the total activation level of the PHO system and the two clusters. The mean activation output of individual promoters in each program (ePi level range) is used to calculate the individual relative to the total activation output level. Here the contribution merely refers the promoters' output and is not relevant to gene function and importance of the gene. In the "full" program, all the promoters are fully activated. Decoupled promoters share similar activation levels as plateau promoters do even with fewer members. Then, their amount dramatically drops in the "plateau" program. In the meantime, the total PHO system output drops around four-fold. A similar ratio happens in the "bistable" and the "off" program and the total activation decreases further.
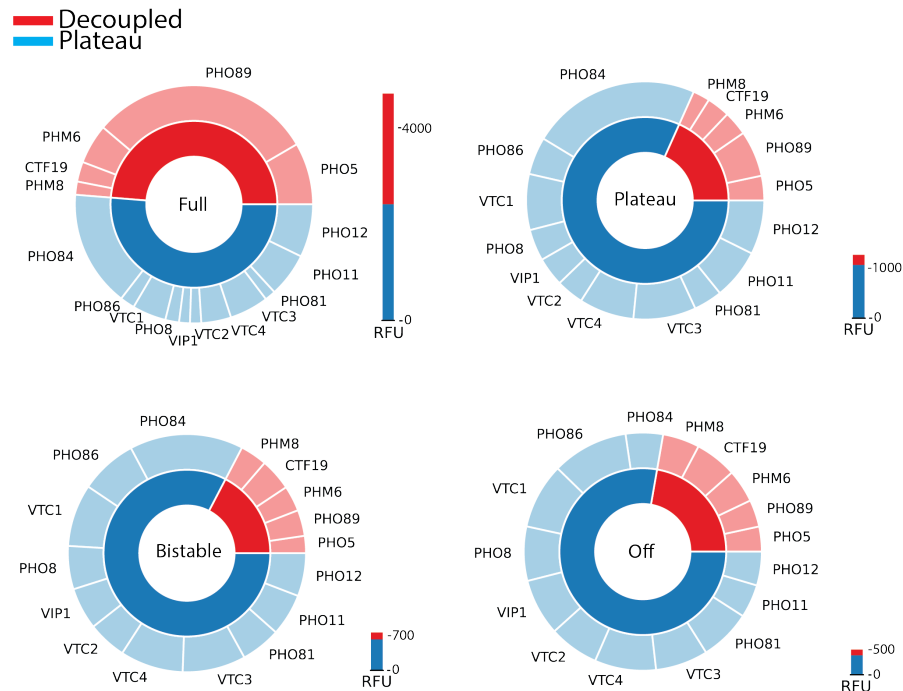
Figure 2.18: **Promoter activation levels during the different expression programs of the pho-regulon.** The relative activation levels of the PHO promoters in the four programs are shown as pie charts. The bar charts indicate the absolute level of activation. Promoters are grouped in either "decoupled" or "plateau" represented as red and blue, respectively.

2.3.5 Pho4p nuclear localization

The existence of a plateau region is puzzling. Pho4p is known to be localized to either the cytoplasm or nucleus and no intermediate states of Pho4p nuclear localization have been described thus far. It is therefore not clear how promoter decoupling can be achieved at intermediate phosphate conditions, nor is it at all understood how the system achieves a robust, perfectly adapted state.

We attempted to quantify Pho4p localization under steady-state conditions over the entire functionally relevant eP_i range (Fig. 2.2). Nuclear localization of Pho4p was clearly visible under full starvation conditions. In intermediate eP_i conditions, nuclear localization of Pho4p was not as obvious, but we were able to observe a detectable level of Pho4p nuclear localization (Fig. 2.2A).
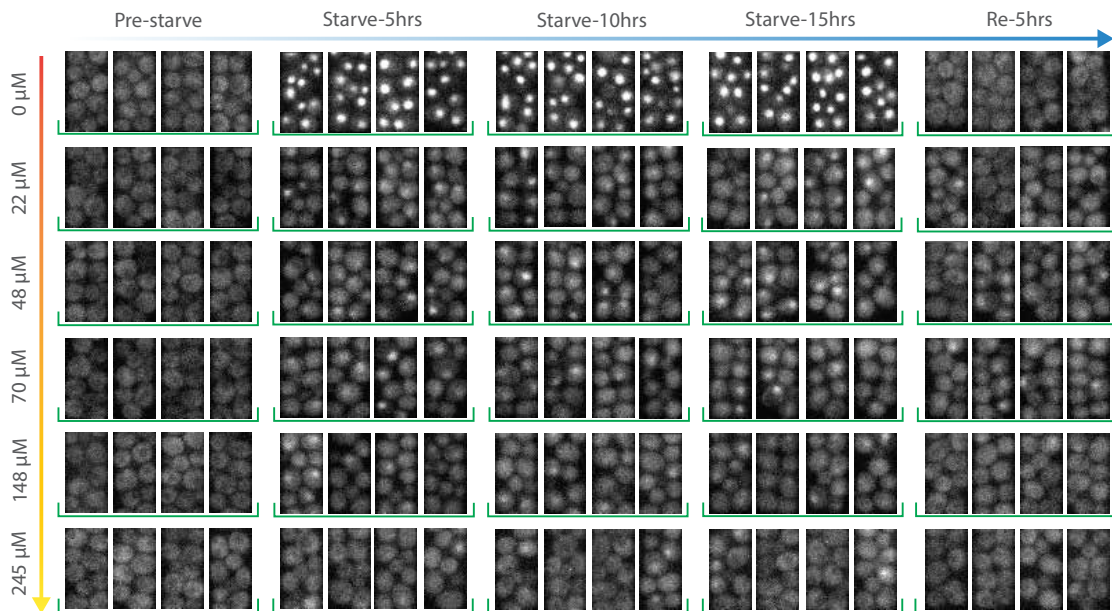
Figure 2.19: **Pho4p nuclear localization as a function of [eP_i].** Time-lapse image series showing single-cell Pho4p nuclear localization. No Pho4p is visibly nuclear localized pre-starvation. Upon shift to lower [eP_i], nuclear localization is clearly visible for 0 μM eP_i. In the range of 22 - 245 μM eP_i low levels of nuclear Pho4p are visible.

Although nuclear Pho4p signal was rather low at intermediate conditions, we attempted to quantify Pho4p localization. We used a simple, but effective analysis method based on the fact that when Pho4p is localized to the nucleus it also becomes locally more concentrated (Figure 2.20).
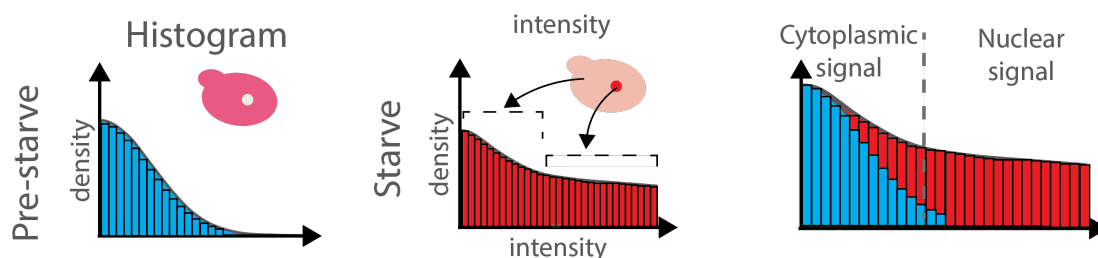
Figure 2.20: **Schematic explanation of the image histogram based analysis for Pho4p nuclear localization.** When Pho4p becomes nuclear localized it also becomes locally more concentrated, giving rise to higher intensity pixels.

Localization of Pho4p to the nucleus should thus be reflected in an increased number of higher intensity pixels observable in image histograms. For low $[eP_i]$ the transition of Pho4p to the nucleus is visually obvious and results in very pronounced histogram tails (Figure 2.21). This histogram method is sufficiently sensitive to detect nuclear Pho4p at intermediate and high $[eP_i]$ (Figure 2.21, Supplementary Figure 2.33). For example, at 70 μM eP_i, a clear difference in Pho4p localization can be seen between the pre-starvation state and the starvation state (Figure 2.21).
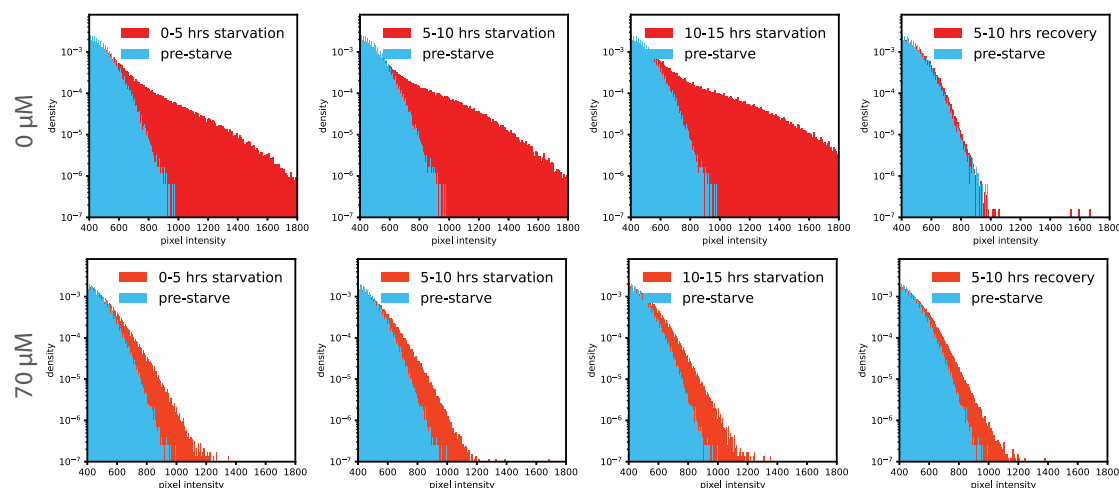
Figure 2.21: **Image histograms for Pho4 nuclear localization.** Image histograms showing pixel intensity distributions for 0 μM and 77 μM eP_i. In every plot the blue histogram shows the pixel intensity distribution prior to starvation for reference. The red histograms show either the intensity distribution during starvation or during recovery as indicated.

In full starvation conditions Pho4p localized to the nucleus to a large degree (Figure 2.22, Supplementary Figure 2.34). Intermediate conditions gave rise to low but measurable Pho4p nuclear localization, while rich P_i conditions showed very slight or no Pho4p nuclear localization. Strikingly, Pho4p nuclear localization in intermediate conditions also appeared to be robust and invariant to $[eP_i]$.
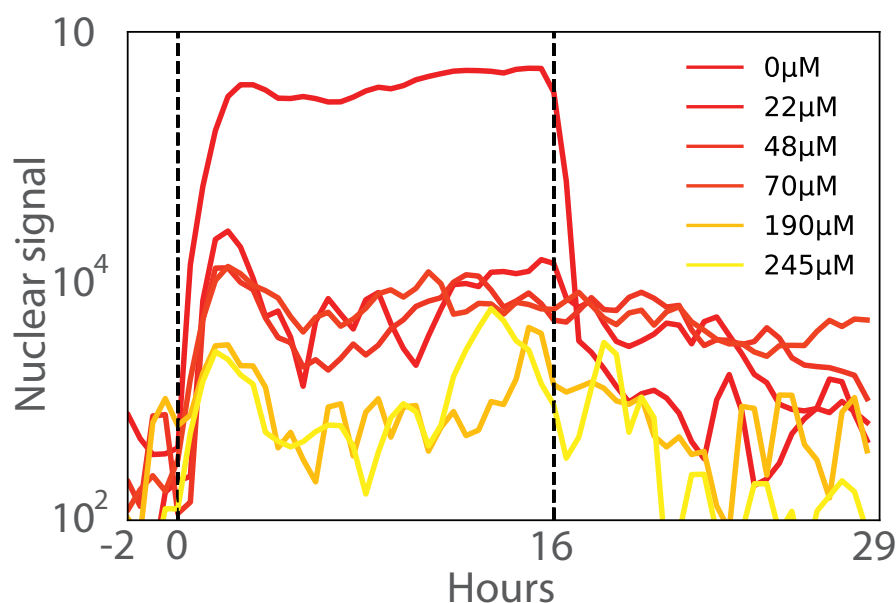
Figure 2.22: **Pho4p nuclear localization plotted over time for different $[eP_i]$s.** In full starvation conditions Pho4p localized to the nucleus to a large degree. Intermediate conditions gave rise to low but measurable Pho4p nuclear localization, while rich Pi conditions showed very slight or no Pho4p nuclear localization. Strikingly, Pho4p nuclear localization in intermediate conditions also appeared to be robust and invariant to $[eP_i]$s.

We thus plotted Pho4p nuclear localization at steady state as a function of $[eP_i]$ (Figure 2.23), showing that Pho4p is only partly nuclear localized in the plateau region. Complete nuclear localization only occurs in the on state of the system under low $[eP_i]$. Partial nuclear Pho4p localization in the plateau state is furthermore largely constant and insensitive to $[eP_i]$.
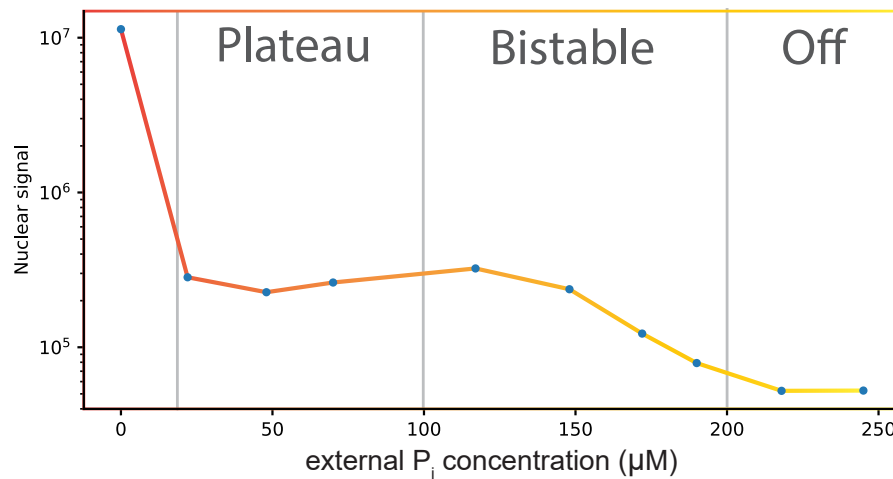
Figure 2.23: Pho4p nuclear localization at steady state under different $[eP_i]$.

The output of the Pho regulon therefore appears to be correlated with, and likely determined by, Pho4p nuclear concentration. In the off state, no, or very small quantities of Pho4p are nuclear localized, leading to a nuclear concentration that is insufficient to activate any Pho regulon promoters. In the plateau state, Pho4p is partially nuclear localized, leading to a nuclear Pho4p concentration that is sufficient to activate plateau promoters but remains below the threshold to fully activate decoupled promoters. Under low eP_i conditions, Pho4p is largely nuclear, giving rise to a high nuclear concentration that is sufficient to activate all promoter classes.

2.3.6 Transport-centric model

Given these insights we set out to derive a model that could explain the observed phenotypes. Current models assume that P_i influx rate varies with $[eP_i]$. At the same time, we and others have not observed measurable growth rate decreases at intermediate $[eP_i]$. We posit that P_i consumption or assimilation rate has to be proportional to growth rate, which means that at constant growth rate, P_i consumption rate is also constant. Under these conditions: a P_i influx rate that varies with $[eP_i]$, and a constant P_i consumption rate; internal P_i concentrations ($[iP_i]$) should track $[eP_i]$. A continuum of $[iP_i]$ makes it difficult to explain the observed nuclear localization behaviour of Pho4p and general behaviour of the Pho regulon, aside from appearing to be a suboptimal control strategy.

We therefore asked whether, instead of a control mechanism established by a finely tuned network architecture, the Pho regulon could be governed by an entirely different mechanism. This led us to look more closely at the transporter system, which consists of constitutively expressed low-affinity transporters (Pho90p and Pho87p), and high-affinity transporters (Pho84p and Pho89p) that are expressed at intermediate and low $[eP_i]$. The Pho regulon expresses Spl2p at intermediate and low $[eP_i]$, which is thought to bind to low-affinity transporters turning them off. The existence of the dual transporter system has been suggested to provide a competitive advantage, albeit with the same supposition that P_i influx varies with $[eP_i]$ (117). More recently Bosdriesz and colleagues put forth a transport carrier model as opposed to the standard Michaelis-Menten model (200). They argue that the carrier model explains the existence of different affinity transporters, as transporters of different affinities will achieve optimal transport rates at different external substrate concentrations.

Another important consequence of the carrier model that has not been described thus far is the fact that it predicts that $[iP_i]$ are robust and insensitive to $[eP_i]$. This can be easily seen when plotting the import rate (J) as a function of both internal and external substrate concentration (Fig. 2.24A). The import rate function J collapses on a single curve when external substrate concentrations are sufficiently high. The system is also stable, as the P_i import rate J will always tend towards matching the cellular P_i consumption rate. Choosing a constant import rate J, one can then plot internal versus external P_i concentration which shows that internal substrate concentration is stable and robust to $[eP_i]$, and below a critical threshold, $[iP_i]$ will tend to zero (Fig. 2.24B). A Michaelis-Menten model on the other hand stipulates that above the critical threshold $[iP_i]$ will always track $[eP_i]$ (Fig. 2.24C-D).
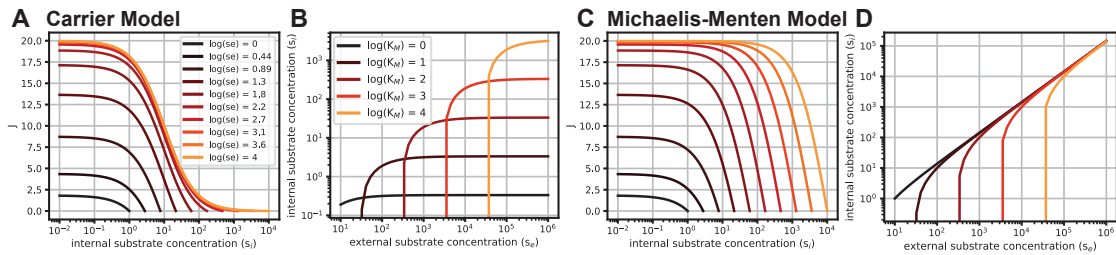
Figure 2.24: **Carrier model *versus* Michaelis-Menten model. A,** Carrier model steady-state import rate (J) as a function of external and internal substrate concentration (201). Parameters were: $k_2 = 4$; $k_4 = 4$; $K_d = 10$; $e_t = 10$. **B,** Carrier model: internal substrate concentration as a function of external substrate concentration at an import rate $J = 15$. **C,** Michaelis-Menten model for comparison. Parameters were the same as for the carrier model. **D,** Michaelis-Menten model: internal substrate concentration as a function of external substrate concentration at an import rate $J = 15$.

A simple carrier model consisting of two transporters of different affinities is consistent with experimental observations of the Pho regulon, including the apparent robustness to $[eP_i]$ (Figure 2.25). In the off-state, when $[eP_i]$ is high, low-affinity transporters are active and set an $[iP_i]$ that lies above the activation threshold (Figure 2.25E,I,L). As $[eP_i]$ approaches the critical threshold for the low-affinity transporters, a bistable region exists in which some cells remain in the off state, whereas others may enter the plateau state (Figure 2.25D,H). At intermediate $[eP_i]$, $[iP_i]$ levels fall below the activation threshold, leading to partial phosphorylation and partial nuclear localization of Pho4p. This in turn activates Spl2p which inactivates existing low-affinity transporters. Concomitantly, Pho4p expresses the high-affinity transporters, which establish a new stable state of constant $[iP_i]$, that lies below the activation threshold, but above the threshold at which growth limitations occur (Figure 2.25C,G,K). The cell therefore has to cope and adapt to only two $[iP_i]$, which can be relatively similar (only limited by the transfer function of the Pho80p/85p kinase system), and are constant and robust to changes in $[eP_i]$. When $[eP_i]$ falls below the critical level for the high-affinity transporter, $[iP_i]$ will tend to zero leading to complete dephosphorylation of Pho4p and complete nuclear localization, which activates all Pho regulon genes (Figure 2.25B,F,J). The model also predicts hysteresis when returning from intermediate $[eP_i]$ to high $[eP_i]$ since the high-affinity transporters are turned off and the high-affinity transporters will keep $[iP_i]$ constant. This hysteresis has been previously seen and described as commitment (123), and we also observed hysteresis in promoter activity (Figure 2.11) and Pho4p nuclear localization (Figure 2.21, 2.22, 2.33).
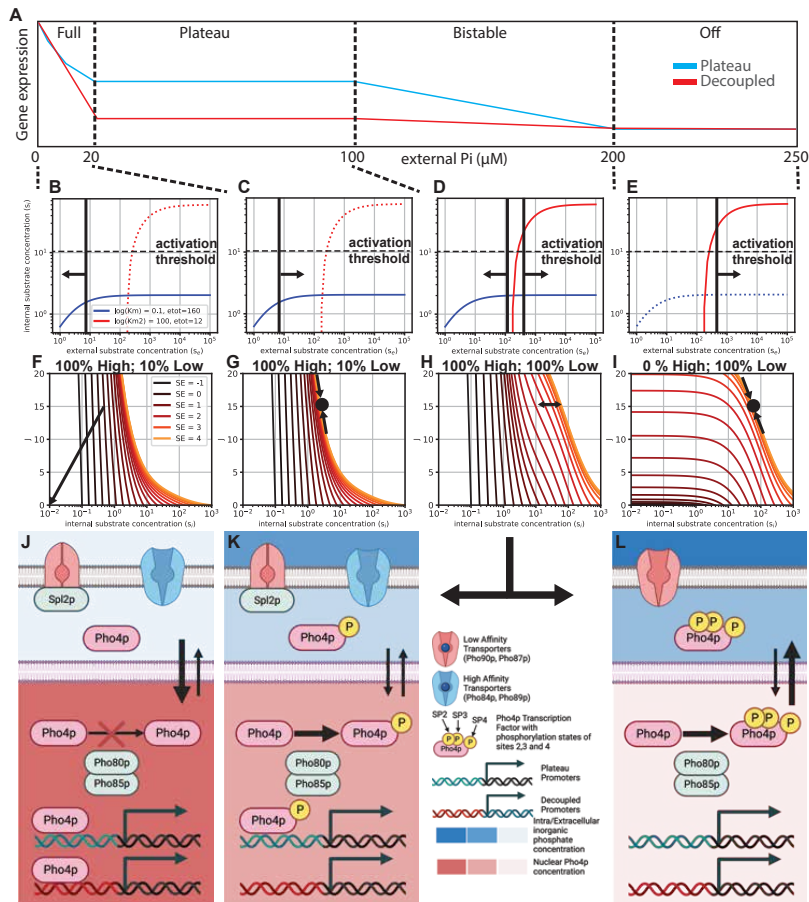
Figure 2.25: **Transport-centric Pho regulon model. A,** Schematic representation of the Pho regulon output function with its three major states (fully active, plateau, off) and the transition (bistable) state demarcated by dashed vertical lines. **B-L,** Each of the states can be mapped to a specific cellular state. **E, I, L,** At high $[eP_i]$ the Pho regulon is turned off. Low-affinity transporters are active, setting a robust $[iP_i]$ that is above the activation threshold. Pho4p is fully phosphorylated and localized to the cytoplasm. No Pho regulon genes are expressed. **D, H,** When $[eP_i]$ approach the critical value set by the low-affinity K_D (or K_M) value a bistable state exists in which some cells remain in the off state while others enter the plateau state. **C, G, K,** When $[eP_i]$ is below the critical import threshold of the low-affinity importers, $[iP_i]$ falls below the Pho regulon activation threshold and Pho4p becomes partially dephosphorylated, leading to low levels of nuclear localization. This nuclear Pho4p concentration is high enough to partially activate the Pho regulon (plateau-promoters). High-affinity transporters are expressed, which establish a new, robust $[iP_i]$ below the activation threshold. Spl2p is also expressed which turns off the low-affinity transporters. **B, F, J,** When $[eP_i]$ falls below the critical import threshold defined by the high-affinity transporter K_D (or K_M) value, $[iP_i]$ falls, and tend towards zero, leading to full dephosphorylation and full nuclear localization of Pho4p, which in turn activates the entire Pho regulon (plateau and decoupled promoters).

2.4 Discussion

The *S. cerevisiae* Pho regulon has been studied extensively but the precise mechanism governing the regulon remained unclear. The current Pho regulon model stipulates that it is regulated by feedback loops. It furthermore is thought that P_i import rates vary with $[eP_i]$ (123; 202), leading to changing $[iP_i]$ that are transduced to the master regulator Pho4p. Nuclear Pho4p concentration is therefore also thought to change as a function of external or internal P_i concentration following a standard transfer function (111), although only two Pho4p localization levels had thus far been described: a fully nuclear state and a fully cytoplasmic state. It was therefore argued that affinity or kinetic differences due to phosphorylation of site 6 must play a role in the differential expression observed between *PHO5* and *PHO84* at intermediate P_i levels leading to an activity rather than a concentration model, but it was also acknowledged that an unknown mechanism must exist that contributes to this differential expression (111).

By conducting a systems-level analysis of the Pho regulon we found that the Pho regulon can attain three well-defined states: a fully activated state, a plateau state, and an off state, with a bistable transition state between the off and the plateau state. The plateau state is perfectly adapted and robust to external P_i concentration changes over a range of 20 - 100 μM, or 20 - 200 μM if the bistable region is included. We characterized nuclear localization of Pho4p over a wide range of external P_i concentrations and discovered that, rather than attaining two nuclear localization states, Pho4p can attain three distinct nuclear localization states. In the plateau region Pho4p is partially localized to the nucleus, and this localization level is robust to external P_i concentration.

This intermediate nuclear localization state of Pho4p readily explains the observed states of the Pho regulon and provides an appealing solution of how the Pho regulon is designed and tuned on the promoter level. The original model of Pho4p with only two nuclear localization levels (zero and full) required an "activity" model in order to explain observed differential gene expression. Alternatively, a continuum of nuclear Pho4p concentration could have explained the differential expression, but failed to predict the observed perfectly adapted and robust plateau state. Three nuclear Pho4p concentration states on the other hand readily account for the three programmatic output states of the Pho regulon and the three nuclear Pho4 concentration states can be explained via the phosphorylation states and phosphorylation preferences of sites 2,3 and 4 (109). The effect of site 6 is less clear but it is possible that although not necessary for differential expression of plateau and decoupled promoters in intermediate phosphate conditions (111), that it either plays a role in preventing aberrant expression of plateau promoters under non-starving conditions, or provides

increased dynamic range between starved and intermediate expression levels by tuning the interaction strength with Pho2p.

Having three defined nuclear Pho4p concentrations of zero, intermediate, and high, also simplifies Pho regulon promoter evolution. Decoupling is used to prevent promoter activation in the plateau region, which is easily accomplished because nuclear Pho4p concentrations are sufficiently low to prevent outcompeting nucleosomes. Only when nuclear Pho4p concentrations are high can Pho4p compete with nucleosomes leading to full promoter activation. Furthermore, expression levels can be easily tuned as promoter binding site affinity can evolve to two specific Pho4p concentrations as opposed to coping with a continuum of nuclear Pho4p concentrations. This is an effective and simple approach to gene regulatory network design and is instructive for the design of complex synthetic gene regulatory networks.

The observation of three nuclear Pho4p states raises the question of how these states are established in the first place, and how a robust and perfectly adapted plateau state can exist. We excluded the possibility that the Pho regulon uses a network feedback approach to achieve perfect adaptation because we did not identify any components that act as a negative feedback nor found any negative feedback components that track $[iP_i]$ or $[eP_i]$. It is possible that the signaling cascade leading to Pho4p phosphorylation establishes three distinct Pho4p phosphorylation states as a function of continually varying $[iP_i]$. But we consider varying $[iP_i]$ to be inherently flawed and explored whether it might be physically possible to establish robust $[iP_i]$ instead.

A standard Michaelis-Menten transporter model predicts that internal substrate concentration tracks external substrate concentration under steady-state conditions. This seems to be a sub-optimal strategy as cells would need to function optimally over a wide range of $[iP_i]$. A continuum of iP_i levels would also necessitate complex regulatory control to achieve the observed Pho regulon program. A carrier transport model recently explained the existence of different affinity transporters (200). We realized that the carrier model also predicts that internal substrate concentrations are constant and robust to external substrate concentrations. Therefore a simple system of two transporters with different affinities could establish two robust $[iP_i]$, which then simply need to establish two Pho4p phosphorylation states through the Pho80p/85p kinase system. These two phosphorylation states in turn determine whether Pho4p is sequestered to the cytoplasm, or nuclear localized to intermediate levels. When iP_i levels fall further, Pho4p is completely dephosphorylated and fully nuclear localized. This model is compatible with all known Pho regulon phenotypes, including the observation of commitment or hysteresis that occurs when cells are returned to rich eP_i conditions. It eliminates the inconsistencies of the current model,

particularly in regards to the existence of the feedback loops, and predicts a stable plateau region, which has hitherto not been described. The model also predicts the establishment of homeostatic conditions. Therefore we suggest that complex gene regulatory networks such as the Pho regulon could be governed by simple control mechanisms based on transporter biophysics.

This work provides insight into how nature evolved simple, yet robust solutions to controlling sophisticated gene regulatory networks. The Pho regulon exhibits a complex response consisting of three distinct states: off, intermediate activation, and full activation. The differences between the intermediate and full activation state are complex and involve not only differences in activation levels, but also differences in the set of promoters being activated in each state. These states are in turn determined by 3 corresponding nuclear concentration levels of Pho4p. Therefore, a combination of Pho4 nuclear concentration states and promoter decoupling through nucleosome occupation of TF binding sites can fully account for this complexity. This example demonstrates that distinct and well-defined nuclear TF concentrations can be established and give rise to complex programs. It is thus likely that TF concentration is instrumental in other regulatory networks as well. Further, we posit that overall system control is not achieved by complex regulatory network architectures, but rather by simple transporter biophysics. It will be interesting to determine whether nature uses similarly simple, yet elegant solutions, to regulate other complex systems such as the glucose regulatory network, which also uses a combination of high and low affinity transporters.

2.5 Methods

2.5.1 Media

To prepare Lysogeny Broth (LB) medium, we dissolved 25 g LB powder (MP, 3002022) in deionized water to a final volume of 1 L. To prepare LB agar plates, we dissolved 25 g LB powder, 20 g Agar (Alfa, 10752-36) in deionized water to a final volume of 1 L. The medium was autoclaved before adding antibiotics. To prepare YPD medium, we dissolved 50 g YPD broth (Sigma, Y1375) in deionized water to a final volume of 1 L. To prepare YPD agar plates, we dissolved 50 g YPD powder, 20 g Agar (Alfa, 10752-36) in deionized water to a final volume of 1 L. The YPD medium was autoclaved before use. To prepare SC-URA plates, we added 1.92 g yeast synthetic dropout medium supplement (without uracil) (Y1501-20G), 6.7 g Yeast Nitrogen Base without amino acids (Y0626-1KG), and 20 g Agar in deionized water to 960 mL and after autoclaving added 40 mL 50% filter sterilized glucose solution (Corning 500 mL Vacuum Filter/Storage Bottle System, 0.22 μm Pore, Cat#431097) to a final volume of 1 L. To prepare the 5-FOA plates, we added 1.92 g yeast synthetic dropout medium supplement (without uracil) (Y1501-20G), 6.7 g Yeast Nitrogen Base without amino acids (Y0626-1KG), and 20 g Agar in deionized water to 880 mL and after autoclaving added 40 mL 50% filter sterilized glucose solution, 66 mL 760 μg mL^{-1} uracil solution, 10 mL 10% 5-FOA solution (EZSolutionTM, 00386260) to a final volume of 1 L.

2.5.2 Plasmid Library Construction

We used NEB 10-beta Competent *E. coli* (High Efficiency) (New England Biolabs, Cat #C3019H) for all plasmid cloning following the transformation protocol from the manufacturer. We conducted bacterial selection and growth in LB plates or LB medium at 37°C supplemented with appropriate antibiotics (chloramphenicol 34 μg mL^{-1}, ampicillin 100 μg mL^{-1}, or kanamycin 50 μg mL^{-1}). For the NN library (native locus, native terminator), the integration cassettes consisted of the native promoter sequence (1500 bp), yeast codon optimized enhanced green fluorescent protein (yoEGFP), and native terminator sequence (500 bp). The native ORFs were replaced by a yeast codon optimized enhanced green fluorescent protein (yoEGFP) using homologous recombination. For the LF library (LYS2 locus, ADH1 terminator), the integration cassettes consisted of two LYS2 homology arms (each 500 bp), a neutral insulation sequence (200 bp), the native promoter sequence (1500 bp), yoEGFP, and ScADH1 terminator (225 bp). We designed different sgRNAs (Table 2.5) targeting the native loci and LYS2 Locus and constructed corresponding sgRNA integration plasmids. These DNA sequences were generated from PCR or ordered as DNA fragments (IDT DNA).

All integration cassettes were constructed via BsaI or BsmBI Goldengate/TypeIIs assembly of the DNA fragments. Goldengate reactions are prepared as following: 50 ng of backbone plasmid, 150 ng of insertion plasmid, 2 μL of 10x T4 ligase buffer (NEB), 2 μL of T4 ligase (NEB), 1 μL of BsaI or BsmBI (NEB), and adding water to a final volume of 20 μL. Thermocycler setup: (BsmBI assembly: 45°C for 2 min; BsaI assembly: 42°C for 2 min, 16°C for 5 min) x 25 cycles, 60°C for 10 min, 80°C for 20 min, hold at 4°C forever.

2.5.3   Yeast Strain Construction

For yeast cloning, we used CRISPR/Cas9 assisted integration by homologous recombination (Figure 2.2) using the lithium acetate/polyethylene glycol (PEG) method (203). The GFP-TAG strains, including, Pho87 (tag), Pho90 (tag), and SPL2 (tag), were obtained from a commercially available yeast GFP Clone Collection (Thermo Fisher Scientific, US) (204). Two yeast strain libraries, including NN, and LF, were constructed for this study. The sgRNA plasmids, CRISPR/Cas9 plasmid, and integration cassettes were sequence-verified and digested by EcoRV, BsmBI, and NotI for yeast integration, respectively. EcoRV digestion was set up as follows: DNA 1 μg, 10X NEBuffer r3.1 5 μL, EcoRV 1 μL, nuclease-free water to 50 μL, incubate at 37°C for 1 h and 65°C for 20 min. BsmBI digestion was set up as follows: DNA 1 μg, 10X NEBuffer r3.1 5 μL, BsmBI-v2 1 μL, nuclease-free water to 50 μL, incubate at 55°C for 1 h and 80°C for 20 min. NotI digestion was set up as follows: DNA 1 μg, 10X CutSmart buffer 5 μL, NotI-HF 1 μL, nuclease-free water to 50 μL, incubate at 37°C for 1 h and 65°C for 20 min. After DNA purification (ZYMO DNA clean and concentrator-25 kit), they were transformed and integrated into the parental strain (BY4741) by homologous recombination using the lithium acetate/polyethylene glycol (PEG) method. Briefly, the parental strain BY4741 was grown in the yeast extract peptone dextrose (YPD) medium over night at 30°C and shaking at 250 rpm. After overnight culture, the yeast cells were back diluted to OD 0.175 in 5 mL fresh YPD medium. The OD600 reached around 0.75 after incubation for 4-5 hours. We pelleted the cells by centrifuging for 10 min at 2000 rpm. While harvesting cells, we prepared the lithium acetate transformation solution and DNA mixture solution. For each transformation, we prepared 306 μL chemical solution, including 260 μL PEG3350 (50% w/v), 36 μL Lithium acetate solution, and 10 μL ssDNA (10 mg mL^{-1}). In the meantime, we prepared 54 μL DNA mixture solution for each transformation, including 1 μg integration cassette, 100 μg Ca9/URA3 cassette, and 200 μg sgRNA cassette. After centrifuging, we washed the yeast cells with 2.5 mL water once and with a 100 mM lithium acetate solution twice. Afterwards, the yeast cells were suspended in the DNA mixture solution (54 μL) first and then mixed with the lithium acetate transformation solution (306 μL). After

incubation at 30°C for 30 min, we incubated the yeast cell transformation mixture at 42°C for a further 15 min in a water bath. We centrifuged the cells at 8000 rpm for 2 min and removed the supernatant. We re-suspended the cells in 200 μL water and plate 50 μL on the selection plates. We conducted transformant selection on synthetic complete uracil dropout (SC-URA) agar plates and cured the selected colonies on 5-FOA supplemented synthetic complete (SC 5-FOA) plates. We screened transformants for correct integration by colony PCR and verified the PCR product by sanger sequencing.

To create the mScarleti-Pho4p strain, we created a mScarleti-Pho4p plasmid with the Pho4p promoter, N-terminus tagged mScarleti-Pho4p, and the Pho4p terminator using a BsmBI assembly. The mScarleti-Pho4p plasmid, CRISPR/Cas9 plasmid, and two sgRNA plasmids (targeting the Pho4p chromosomal site), were sequence verified and digested by Not-I, BsmBI, and EcoRV for integration, respectively. After DNA purification, they were transformed into the parental strain (BY4741) using the lithium acetate/polyethylene glycol (PEG) method. We conducted the transformant selection and curing on SC-URA and SC 5-FOA plate, respectively. We screened transformants for correct integration by colony PCR and verified the PCR product by sanger sequencing.

2.5.4   MICROSTAR Device

This device was developed and fabricated by Hon Ming Yip in this study.

Device Design

The MICROSTAR (MICROfluidic *S. cerevisiae* Trapping ARray) device was designed in Autocad (Autodesk, U.S.A.) and is fabricated by standard multi-layer soft lithography (205) to produce a two layer device consisting of a control and a flow layer.

The MICROSTAR device consists of 256 culture areas, each area in turn contains 8 parallel culture grooves 8 μm in width or roughly 2 *S. cerevisiae* cell diameters. The device contains 2,048 cell grooves in total. Each groove is 3.8 μm in height to force cells to grow in monolayers. Each culture groove is connected to two parallel media perfusion channels. On the bottom of each groove we created a narrow sieve channel that is 1.2 μm in height, 11 μm wide, and 5.5 μm long. This allows media to perfuse the bottom as well as the top of each culture groove, but allows cell escape only from the top. This design creates a well controlled culture area proximal to the sieve channel. The MICROSTAR device contains 16 columns and 16 rows. Each column can be

individually loaded with different yeast strains. The 16 rows in turn are divided into 8 individually addressable rows to provide different media conditions, generated by an upstream dilution generator.

The dilution generator (206) is designed to perform on-chip dilutions from two inlets to 8 outputs. The channel width of the dilution generator was designed to be $20\mu m$, which greatly reduces the required dilution time. The dilution generator can either be used with a single medium inlet, in which case all DG output channel media conditions are identical and equal the media input. To generate different concentrations, two media formulations are introduced from two different inlets, which are combined in the DG to generate 8 unique output formulations. The media switching unit enables changing the inorganic phosphate concentration during the experiment.

Mold Fabrication

The control and flow molds for the MICROSTAR chip were fabricated using conventional photolithography techniques in the center of micronanotechnology at EPFL.

The CAD file of the device was converted to CIF format for mask production. Due to PDMS shrinkage, the control and flow layer were re-scaled to 103.5 percent and 100.5 percent, respectively. Chrome masks with photoresist AZ351 coating were used for mask production. They were written on a Heidelberg Instruments VPG200 (Heidelberg Instruments Mikrotechnik GmbH, Germany). Masks were then processed in a mask developer (Hamatech HMR900) using standard development, chrome etching, resist stripping, and final rinse dry parameters. The completed masks were then ready to use after overnight drying.

For flow layer mold fabrication, a 4-inch silicon wafer was cleaned at 500W for 7 minutes in an oxygen plasma cleaner Tepla 300 (PVA TePla, Germany). The wafer was then ready to be processed using standard SU-8 fabrication process consisting of (1) photoresist coating, (2) soft bake (SB), (3) exposure, (4) post-exposure bake (PEB), (5) development, and (6) hard bake (HB). The parameters of the process, materials, and the machines are listed in Tables 2.7, 2.8, and 2.9. No development after post-exposure baking was performed for the sieve layer. The photoresist for the chamber layer was directly coated on top of the sieve layer, and both layers were developed together after the chamber layer PEB. This measure prevented the sieve layer structure from negatively affecting coating of the chamber layer which otherwise would cause uneven coating. However, this measure was not applied to the dilution generator layer because: (1) photoresist became more resistant to developer after 2/3 rounds of long soft bake and post-exposure bake that made some parts undeveloped, and (2)

DG layer has no overlap with the sieve and chamber layer and so the influence was minimal. After finishing all three SU-8 layers, the wafer was then patterned using a positive photoresist AZ 10XT-60 (MicroChemicals, U.S.A.). The reason for using this type of photoresist is that it is able to generate rounded structures through annealing which can be completely closed by "Quake style valves". The fabrication procedure includes (1) HMDS treatment, (2) photoresist coating, (3) baking, (3) exposure, (4) relaxation, (5) development, and (6) annealing. The details are also listed on Tables 2.7, 2.8, and 2.9. To round the AZ flow channels, the wafer was annealed on a hotplate at 115°C for 1 hour.

The control mold was fabricated using a standard SU-8 workflow as it consists of only a single layer. The parameters of the process, materials, and the machines are listed in Tables 2.7, 2.8, and 2.9. After completion of the molds, they were inspected under a microscope to ensure features were fully developed and the height of each layer was measured using a surface profiler Bruker Dektak XT (Bruker, Germany). The feature heights of the flow mold are 1.2μm (sieve), 3.8μm (chamber), 15μm (flow channel) and 40μm (dg), and the feature height on the control mold is 20μm. Both newly fabricated control and flow molds were treated with TMCS (trimethylchlorosilane) vapor in a wafer box inside a fume hood overnight before the first usage.

Device Fabrication

Device fabrication includes (1) PDMS preparation, (2) spin coating (for flow layer), (3) 1st baking, (4) hole punching, (5) alignment, (6) final bake, and (7) bonding to glass. Before chip fabrication, both control and flow wafers were treated with TMCS vapor in a wafer box inside a fume hood for half an hour. Then, two PDMS ratios were prepared. For the control layer, 48 g of PDMS (with a base to curing agent ratio of 5:1) was prepared and mixed in a Thinky mixer ARE-250 CE (Thinky, U.S.A) for 1 min at 2,000 rpm and then de-foamed for 2 min at 2,200 rpm. The control mold was placed in a 6-inch glass petri dish coated with aluminium foil. The mixed PDMS was poured on the control mold and then degassed in a vacuum chamber for half an hour. After that, the PDMS was inspected and all air bubbles and dust particles were removed using a plastic pipette. For the flow layer, 21g PDMS (with base to curing agent ratio of 20:1) was prepared using the same procedure. The flow wafer was placed in a spin-coater G3P (Specialty Coating Systems SCS, U.S.A.) and about 8g of PDMS was poured onto it. Likewise, all air bubbles and dust particles were removed. The PDMS was then spin coated at 1,800 r.p.m. for 1 minute. The flow mold with spin-coated PDMS was then placed on a flat surface for 50 min to allow the coating to flatten. Both PDMS coated wafers were then cured for 30 minutes at 80°C. After that, the three control layers per mold were cut and peeled off the mold and the inlets of each control line

were punched (OD = 889 mm) using a precision manual-punching machine (Syneo, USA). Each control layer was then aligned by hand on top of a flow layer using a Nikon stereo microscope. The aligned devices were then baked at 80°C for 90 minutes, allowing the two layers to bond together. The assembled devices were then cut and peeled off the flow mold. The flow layer inlets were then punched. Finally, the device and a No. 1 coverslip (24x60mm, BRAND 470820) were treated in an oxygen plasma machine FEMTO Model 1A (Diener, Germany) for 20s, and then bonded together. The assembled device was then baked for 2 mins at 90°C.

Device Operation

The MICROSTAR device contains ~2,800 microvalves, which allow fluid and cells inside the device to be precisely manipulated. Chip priming (Figure 2.26a) is described in more detail in section 2.5.5. The 16 cell loading inlets are used to transport liquid cultures into the 16 cell culture areas of each column. Figure 2.26b shows the control line setting during the cell loading process. A multiplexer (205) is used to control and select a specific cell loading inlet (Figure 2.7). Another control line "CL1" isolates columns during cell loading to avoid cross-contamination (Figure 2.26c). In the same control line, two valves are placed at the top-left and bottom-right positions of each cell culture area to direct flow into the microgrooves for cell-loading (Figure 2.26c). Only media is able to pass through the sieve structure at the bottom of each microgroove, effectively trapping and retaining cells since their diameter is more than 3 times larger than the sieve channel height. In each column, there are 128 possible flow paths (the number of microgrooves). The length of each path is designed to be identical in order to match flow resistance and thus flow velocities. Hence, the trapping efficiency of each microgroove should be identical in the beginning of the loading process. Once cells are beginning to be trapped, the trapped cells in a microgroove increase the fluid resistance of its path causing other microgrooves that do not yet contain cells to be loaded. Once the cell loading process of a column is completed, the next column is selected and the process repeated until all columns are loaded.

Row isolation (Figure 2.26d, e) and media switching (Figure 2.26F) are the two main operations required for on-chip cell culturing. In the starvation phase, each pair of rows is perfused with a different P_i concentration. In order to maintain stable P_i levels supplied by the dilution generator, it is necessary to physically isolate these rows. The control line "CC1" is designed to disconnect adjacent rows to ensure that there is no mixture/dilution of media occurring in the cell trap array during on-chip cell culture (Figure 2.26e). Other than that, "CC1" also disconnects the two parallel channels in the chamber, cell wash channel (upper channel), and media feeding channel (lower

channel), to keep the media feeding channels clean. The three media inlets and two wash channels form a media switching unit which enables changes in P_i level (Figure 2.26f). During on-chip cell culture, media is supplied from the media switching unit, to the cell trap arrays via the dilution generator, and eventually exits at the chip outlet "CO1". The detailed microvalve operation of cell loading and on-chip cell culturing will be discussed in detail in section 2.5.5.

Dilution Generator Characterization

To validate P_i concentrations, a tracer signal was quantitated throughout the duration of a typical experiment. Measurements were obtained at the cell-trap inlets, the cell-trap outlets, the first and the last culture chamber in each row. The media used in this experiment was the same as the media used in promoter characterisation experiments (see section 2.5.5). Furthermore, another experiment was performed to study the stability of the tracer used in by tracking the signal in the "MI2" channel, "MI3" channel, DG output supplying row 1-2, and DG output supplying row 15-16 over time.

Figure 2.27a and b show the tracer signals across DG outputs over time. The data in Figure 2.27a and b are the mean signal of the cell-trap inlets and outlets, and the chambers, respectively. Tracer signals rose when the dilution generator generated 8 different levels after the first switch and back to the basal level after the second switch. The tracer signal in each zone was stable over time. Figure 2.27c shows the mean normalised signal of inlet/outlet and chambers in each DG output zone.

Figure 2.27d shows that the signal at "MI2" and DG output 1 were identical over time. DG output 1 followed the trend of "MI2". The same result occurred in "MI3" and DG output 8. It explains that the initial fluctuation in Figure 2.27a and b was due to the changes in the "MI2" source but was not a systematic change indicating that on-chip dilution was stable and robust over time.

Switching rate was also measured. The last chamber in a row was captured in the first 10 min of the two switches with 10 second time resolution. Figure 2.27e shows that media reached the culture areas in 40 s and steady state in 1 min.

2.5.5   MICROSTAR Experiments

Cell and Media Preparation

Glycerol stocks (in 96-well plates or in tubes) were taken from storage at -80°C. For 96-well plate stocks, 2μL culture from the 96-well plate stock were inoculated by a 96-pin sampling replicator VP 408FS2AS (V&P Scientific, U.S.A) and then cultured in a 96-well plate with 100 μL YPD in each well overnight at 30°C in a plate shaker (Jencons Millennium 2000, Jencons Scientific, UK) at 900 rpm. Three rounds of flame sterilisation were done using ethanol to sterilise the sampling replicator before each copying process. For freezer stocks in tubes, they were sampled with a loop and transferred into 3 mL YPD and then cultured overnight at 30°C in an incubator (GYROMAX 737R, Amerex Instruments, U.S.A) at 250 rpm. The next day, 2 μL of these cultures were inoculated into tubes containing 1 mL SC media and cultured in the incubator for 5 hours, and then loaded into the microstar device.

Typically, three types of media were prepared for the on-chip experiment, (1) P_i-rich SC medium, (2) P_i-high SC medium, and (3) P_i-low or P_i-free SC medium. Inorganic phosphate from a pre-diluted stock (8.75 μM for P_i-high media and 1.75 μM for P_i-low media) and a tracer (6.3 μg/mL sulforhodamine for promoter characterisation experiments and 5.3 μg/mL fluorescein for transcription factor localization experiments) were added to the P_i-high medium, whereas P_i-low medium received only inorganic phosphate. Media were stored at 4°C and kept in an incubator for at least 2 hours at 30°C before on-chip experiments. All aforementioned medium contained chloramphenicol at a final concentration of 34 μg/mL.

Chip Priming

First, all control inlets were connected to Tygon tubing pre-loaded with oil (Fluorinert FC 40, CAS# 51142-49-5). The lines were then pressurised at 41 kPa to out-gas prime the on-chip control lines. Tygon tubing and therefore the on-chip control lines were regulated by computer controlled solenoid valves. After that, pressure was increased to 179 kPa and all valves were inspected visually to ensure they were closed under this pressure. The flow layer was then ready for perfusion. The flow layer was then completely primed with SC medium. A tubing connected to a bottle with P_i-rich SC media was plugged into the media inlet "MI1"and two intermediate P_i SC media were plugged into the media inlet "MI2" and "MI3" as shown in Figure 2.7 and 2.26a. All media bottles were pressurised at 31 kPa. The media wash channels "MW1" and "MW2" were then flushed by the two intermediate P_i SC media from "MI2" and "MI3", respectively, by releasing the valve "MS3" and "MS4" as shown in Figure 2.26a(b).

Then, "MS4" was closed and "MS1" was opened to let "MI1" wash both "MW1" and "MW2". All air bubbles and dead volumes of the two P_i-intermediate media inputs should be cleared at this step. After that, the main body of the device was ready to be perfused. "DG1" remains closed until all air in the dilution generator was cleared. Then, "DG1", all the valves controlling the cell trap array and cell loading channels were opened to perfuse the remaining parts of the device. Visual inspection under microscope with a 4x objective was performed to ascertain that no air bubbles were present in the chip. Normally, the entire process took around 20 minutes.

Cell Loading

Before loading cells into the cell traps, "CL7" and "CA1" were closed and opens the "MX1-8" to entirely fill all the cell loading channels "CI1-16" until a droplet was formed on each "CI" to ensure there was no air bubble in the inlets. "MX1-8" were closed again afterwards. Liquid cultures of each strain (see above) were taken from the incubator and loaded into individual Tygon tubing. Tubing with cells were then connected to the 16 cell loading inlets CI1-16. After that, the device was placed into cell loading mode by closing "CL3" and "CL1". These two valves were used to physically separate columns and create actively directed flow towards the cell traps. Typically, the cell loading process starts from the first column. Tubing connecting to "CI1' was connected to a pressure source (24.1 kPa) and then the multiplexer (MX1-8) switched to open the path for the first column (both upstream and downstream). The cell loading process was tracked manually using a low magnification objective (4x). It took ~30 seconds to trap sufficient cells into the microgrooves. The loading process stopped once the traps were filled. The same procedure was repeated sequentially for the other columns. Nominally all 16 columns were used in promoter characterisation experiments and 5 columns were used in transcription factor localisation experiments.

On-chip Cell Culture

After cell loading, the chip was placed into cell culture mode. "CL1" was opened and then "CC1" was closed allowing media to enter the cell trap array via the dilution generator to wash away cells not located in the culture grooves and to continuously supply the trapped cells with medium. Typically, the on-chip experiments consisted of three phases, pre-starvation, starvation and recovery. This was accomplished by controlling the media inlet control unit (Figure 2.26f). The duration of phase 1, 2 and 3 were 6, 16 and 16 hours, respectively. Media bottle was pressurised at 31 kPa when it was used, and unpressurised otherwise. The microscope housed in an incubator chamber (ice-Cube&Box, Life Imaging Service, Switzerland) and

temperature control unit which maintained the temperature at 30°C throughout on-chip operations including chip priming, cell loading, and cell-culture.

Time-lapse Acquisition

Time-lapse imaging was fully automated using a custom written visual basic program controlling a fully automated epi-fluorescent microscope (Nikon Ti-E, Nikon Japan). The MICROSTAR device contains 256 cell trap areas. Images of each area were acquired every 20 minutes (promoter characterisation experiments) or 30 minutes (transcription factor localization experiments). The location of each area was calculated by the software and then manually adjusted. Phase contrast and fluorescent images were acquired at every position. The hardware for imaging were (1) Objective CFI Plan Apochromat Lambda D 60X Oil (Nikon, Japan), (2) LED illumination source CoolLED pE-2 (Custom interconnected Ltd., UK), (3) shutter Lambda SC (Sutter Instrument, U.S.A.), and an EMCCD camera with a 1024x1024 pixel sensor array (Ixon DU-888, Andor Technology, UK) was used to acquire images. The camera sensor temperature was set to -70°C and cooled by a liquid-based cooling system (Oasis 160, Solid State Cooling Systems, U.S.A.) at 20°C . The microscope built-in intermediate magnification switch was set to 1.5x to obtain a 90x magnification. With this magnification, the entire cell culture chamber and a portion of upper and lower flow channel were visible. Typically, an experiment lasted 38 hours. The Nikon hardware-based perfect focus system (PFS) was used to maintain a stable focus throughout the experiment. Objective oil (Nikon type F, Nikon, Japan) was replaced and refilled every 12 hours to compensate for evaporation and property changes due to incubation at 30°C. For promoter characterisation experiments, 50 ms and 400 ms exposure times were used for phase contrast and fluorescent images, respectively. For transcription factor localization experiments, 3-layer z stacks (+0.5 μm, center, and -0.5 μm) were taken, and the exposure times for phase contrast and fluorescent images were 50 ms and 4 s, respectively. FITC filter and 470 nm light excitation was used to capture yoEGFP (promoter characterisation), whereas a TxRed filter and 535 nm light excitation was used to capture mScarlet-i (transcription factor) and red tracer signal (5 s exposure time). A GFP filter was used to capture green tracer signal (5 s exposure time).

Calibration Image Acquisition

Darkfield and flatfield calibration images were taken right after completion of the experiment. To capture darkfield images, the light path into the camera was blocked. Then, 500 images were taken using the same camera settings and exposure times as used for taking both phase contrast and fluorescent image. To capture flatfield images,

81 positions (a 9x9 array) of a blue auto-fluorescent plastic slide (92001, Chroma, U.S.A) were sampled (both phase contrast and fluorescent channel) with the same imaging setting as used in time-lapse imaging.

Strains analysed in MICROSTAR experiments and sample size

Table 2.13 shows the strains analyzed in each MICROSTAR promoter activity experiment (exp) and transcription factor localisation experiment (TF). To ensure high data quality, data filtering was performed. For exp experiments, two rounds of exclusions were done based on (1) image quality control (see section 2.5.6), and (2) exclusion of chambers/rows with defects during the experiment (e.g. channel clogging and fabrication defects). For TF experiments, sample grooves were manually screened (see section 2.5.6).

Figure 2.28a-c and Figure 2.30a-c show the number of single cells, grooves and chambers measured in all reported MICROSTAR experiments. The number of chambers and grooves measured are the number of chambers/grooves containing cells by the time point of two hours before starvation. They were detected and registered by the automated image analysis pipeline (see section 2.5.6). For promoter characterisation experiments, the measured number of single cells is the number of cells detected and analyzed by the image analysis pipeline. Figure 2.29 shows the number of single cells measured across P_i concentrations for each strain. For TF experiments, the number of grooves and chambers were registered manually as mentioned. As there was no cell segmentation applied to TF data, the number of cells was estimated by multiplying the number of frames, the number of grooves and 8 (typical number of cell at the bottom 4 layers).

Dilution Generator Performance

The quality of on-chip dilution was quantified by monitoring the fluorescent tracer that was added to the media (see section 2.5.5). During time lapse imaging, tracer signals were measured at the time point of (1) 10 min before switching to starvation, (2) 10 min after switching to starvation, (3) 10 min before switching to recovery, and (4) 10 min after switching to recovery. Measurements were taken at the cell-trap inlets and cell-trap outlets. Figure 2.27c shows that these positions reflect the concentration level in chambers. At the end of experiment, media with tracer (originally in "MI2") was supplied to the entire chip from "MI1" to acquire a reference signal of undiluted tracer.

Figure 2.31 shows the mean tracer signals across 8 DG output zones (two inlets and

two outlets) at the four time point of each MICROSTAR experiment. Two switches were successfully performed in each experiment. The concentration profile was maintained throughout starvation phase. The small difference between the start and the end point of starvation is explained in section 2.5.4 and Figure 2.27d. The on-chip diluted P_i concentrations were calculated using these tracer signals given the user-defined intermediate concentrations in high and low P_i media sources. Figure 2.32 shows the calibrated P_i levels of each experiment. The mean values of the start and end point were used in this paper.

2.5.6   MICROSTAR Image Processing and Data Analysis

The image analysis pipeline was developed by Evan J. Olson.

Image Processing

Processing of raw microscope imagery into summarized, per-cell, per-frame features was performed using a custom, parallelized analysis pipeline written in Python. Furthermore, the pipeline depends on both CellProfiler (207), a tool for automated image analysis, and Ilastik (208), an interactive learning and segmentation toolkit.

The pipeline consists of the following files:

1. *runAnalysis.py* - contains analysis configuration and control flow. The pipeline is initiated by executing this file.

2. *rawprocessing.py* - a module containing methods related to sorting and evaluating the raw timelapse and calibration images.

3. *preprocessing.py* - a module containing methods related to processing the raw images into calibrated, aligned, single-trap timelapses.

4. *singlecellprocessing.py* - a module containing methods related to cell region of interest (ROI) segmentation and single-cell feature extraction

5. *qualityassessment.py* - a module containing methods related to assessing the quality of the imagery and analysis of each timelapse (e.g. focus, alignment, cell growth, etc.)

Before analysis commences, each experiment is given a unique name for reference, [EXPTNAME]. The nomenclature for our experiments is "TY_[DATE]," with the date

given in YYMMDD format (e.g. "TY_200320"). All data and analysis results are archived within a directory given this name. After analysis is completed, the directory structure is organized as follows:

1. *[EXPTNAME]/code* - contains a complete copy of the pipeline at the moment the analysis was completed.

2. *[EXPTNAME]/raw* - contains all raw microscopy imagery as well as four spreadsheets which contain specifications of the positions of strains, timings, media conditions, any traps that should be excluded due to defects or failures, as well as any other notes associated with the execution of the experiment.

3. *[EXPTNAME]/preprocessed* - contains analysis results relating to the production of calibrated, aligned, single-trap timelapses. The resulting ImageJ hyperstacks are contained in the 'stacks' folder. The 'calibration' folder contains the flatfield and darkfield images used to calibrate each raw image. The 'metadata' folder contains metadata associated with each microscopy image, such as the camera, filter, and illumination settings. Finally, the 'movies' folder contains a movie of each single-trap timelapse.

4. *[EXPTNAME]/single_cell* - contains analysis results relating to the segmentation and feature extraction of the single cells within the trap timelapses. The 'scData' folder contains a spreadsheet for each trap listing the measured features of each cell identified in each frame of the trap timelapse. The 'quality' folder contains the results of the quality assessment in a compact, 'pickled' Python format, which can be unpacked using the 'load_stack_quality' method within 'qualityassessment.py.' The 'plots' folder contains a variety of plots which exhibit the behavior of the cells within each trap, including cell fluorescence and area over time as well as diagnostic plots arising from the quality assessment. The 'masks,' 'overlays,' and 'multilayer_masks' folders contain the results of the cell segmentation process, with 'masks' and 'multilayer_masks' containing the regions of each trap which have been identified to contain cells and multilayer cells respectively, and 'overlays' containing diagnostic images of each brighfield image with the identified cell ROI contours overlayed. Finally, the 'single_cell' folder contains the batch processing file 'Batch_data.h5' used by CellProfiler during the segmentation process.

Analysis Preparation

To prepare for the analysis, the user must decide the [EXPTNAME] and generate an initial directory structure consisting of '[EXPTNAME]/raw/' that contains the 'strains,'

'conditions,' and 'manualexclusion' experiment specification spreadsheets, as well as the 'microscopy' folder which contains the raw microscopy imagery. Then, the analysis is initialized via the command 'python runAnalysis.py [EXPTNAME]' which will commence the analysis on all trap positions identified in the raw data. Optionally, the user may specify additional arguments to this command to isolate the analysis to specific traps, which can be identified via their numerical position (1-256) and alphabetic trap designation (A-H). For example, 'pos105_D' for the fourth trap among the eight traps at position 105. The user can add additional arguments as desired.

Pipeline Configuration
The pipeline configuration process is run once at the commencement of the analysis. During this process, the user configuration supplied within the 'runAnalysis.py' file as well as the experiment specification spreadsheets contained within the 'raw' folder are read and the run-time configuration to be used by all following steps is generated. If the microscopy images within the 'raw' folder have not yet been sorted into subfolders the user will be presented with a listing of identified timelapses, positions, and channels and can then choose whether or not the imagery should be sorted based on these characteristics. This sorting step is required to continue with the remainder of the pipeline. Finally, the complete configuration is printed to the user within the terminal and the user can evaluate and choose whether or not to proceed.

Raw Image Processing
The raw image processing is run once during the analysis, immediately following the user authorization to proceed. During this step, the following actions are taken:

1. Establish directory structure

2. Sort raw images

3. Extract metadata from the raw and calibration

4. Ensure that the calibration images match the exposure time and camera settings of the raw images

5. Prepare darkfield (DF) and flatfield (FF) calibration images

6. Prepare image list and batch processing file for later CellProfiler-based image segmentation

7. Prepare multiprocessing workers and initiate batch processing

The directory structure is established as previously described. Raw timelapse images are sorted according to their phase of the timelapse, position, and channel, whereas

the raw calibration images are sorted according to the type of calibration (darkfield or flatfield) as well as the channel and shutter exposure settings. Metadata, such as camera and optical filter settings and measurements (e.g. exposure time and sensor temperature) for each acquired frame are extracted and stored. A verification is performed to confirm that the measured calibration image conditions match those of the acquired timelapses.

Darkfield and flatfield calibration images are computed from the raw DF and FF images acquired following the acquisition of the trap timelapses. Single DF images are highly noisy, so the mean of 500 frames is computed to establish a single DF frame. Furthermore, it was observed that there were rare outlier events in which single pixels would receive a signal far outside the typical range. Even when averaged among 500 frames, these outliers had a substantial effect on the mean of the affected pixels. In order to compensate for this behavior, before the DF mean was computed, first the median and interquartile range (IQR, i.e. 75th percentile minus the 25th percentile) were computed for each pixel in the frame. Any pixels within any frames which fell outside the range of $median \pm 10 \times IQR$ were removed from the computation of the mean. The FF calibration images are calculated for each channel by first subtracting the mean DF image with the same exposure time, and then calculating the median across the 81 acquired FF images. By using the median, across a large number of images acquired at different positions of the fluorescent calibration slide, we can isolate within the FF calibration image only the nonuniformities which are present across all frames, rather than those that are simply the result of dust or defects within a single frame.

Before completion of the raw image processing step, a one-time preparation step for the later single-cell segmentation process is performed in which a CellProfiler batch processing file is generated along with an indexed spreadsheet containing a list of all acquired dia images within the timelapses.

Finally, a list of jobs to be performed by the multiprocessing workers is prepared, and distributed to the worker threads which will run them in the following steps. Each worker thread runs on one core of the analysis computer. Given our hardware configuration, we utilized 38 out of the available 40 cores for workers with the remaining two being available for any overhead operations.

Timelapse Processing

The timelapse processing step, which prepares the timelapse images for subsequent single-cell analysis, is executed once per each of the 256 cell trap areas. During this

step, the following actions are taken:

1. Calibrate raw images using DF and FF images

2. Align subsequent frames to compensate for drift

3. Split images into eight separate cell trap images

4. Produce a single ImageJ hyperstack timelapse including all channels and z-levels for each of the eight traps

5. Generate a movie of the timelapse for each of the eight traps

Each raw timelapse image is corrected for both darkfield and flatfield artifacts which are inherent to the optical system. The DF corrects for the background counts observed on any image, even within the presence of any light incident on the sensor, whereas the FF corrects for the nonuniformity of the light intensity that arises as light travels through the various apertures of the optical path of the microscope system. Furthermore, given that the FF calibration is collected in a reproducible manner, the FF can also serve as a method for standardization of fluorescence across experiments, even in the presence of variations within the excitation intensity as well as adjustments to the optical path.

Each calibrated image is calculated as $IM_{cal} = \text{int}\left(1000 \times \frac{IM_{raw}-DF}{FF}\right)$. Thus the background DF counts are subtracted, and the resulting image is then normalized by the FF. However, because the intensity of the FF image is approximately on the same order as the DF-corrected image, the resulting floating-point image is comprised of values which are on the order of unity. Many image processing operations require that the pixel values are stored as integers, including those operations used in the downstream steps of this pipeline. Thus, we multiply this floating-point image by a fixed value which is established in the configuration of the pipeline, herein set to 1000. Following this re-scaling, the result is recast as an integer.

Following calibration, an image alignment step is performed to compensate for slight drifts over time in the position of the trap area within the frame. This alignment is performed using the images at the center of the z-stack in the brightfield channel. These registration corrections are purely translational - no rotation or other affine transforms are performed. Alignments are all performed relative to the first frame of the sequence. Each frame is aligned in a two-step process, the first being a coarse alignment and the second being a greater precision alignment.

First, the coarse alignment is performed using a binary 'background mask' of the frame, wherein the cell-containing regions are excluded and only regions corresponding to the static background portions remain. This background mask is itself calculated by evaluating the signal generated by the mean of a horizontal slice of the image containing the cell traps. This signal exhibits a periodicity with bright regions corresponding to walls and cell-containing areas, and dark regions corresponding to the background between them. Given this periodicity and the known geometry of the spacing and width of the traps, the horizontal extent of each trap can be determined. To determine the vertical positioning of the traps, the mean of vertical slices of the previously identified trap-containing horizontal regions is evaluated, and a peak corresponding to the transition from the trap to the sieve at the bottom of the trap is then determined for each of the eight traps. These eight values should be in agreement, as the traps should all be at the same vertical position within the frame. Thus, we calculate the median and IQR of the eight values, remove any that exceed a distance of two IQR, and then calculate the mean of the remainders to establish the bottom of the traps. Finally, the top of the trap is established as a fixed distance above the bottom of the trap given the known geometry of the traps. After this process we have determined coordinates for the top, bottom, left, and right sides of each cell-containing region of the frame, hence the regions outside of these areas can be considered as a background which should remain static during the timelapse. The coarse alignment can be performed simply using the position of the top-left corner of the left-most trap in the background mask.

The fine alignment follow the coarse alignment, and again uses the background mask - however for this alignment the masked brightfield image is used for alignment rather than just the mask itself. Before alignment, a histogram equalization is performed on these background regions to compensate for frame-to-frame variations using the 'createCLAHE' method of the OpenCV library. Fine alignment is then performed on the resulting images using the 'findTransformECC' method of the same library. The resulting translational warps are applied to the original frames (across all channels and z-stacks) and are also recorded within the metadata of the final timelapse.

Following the timelapse alignment, the regions containing each of the eight individual traps are isolated and the resulting single-trap timelapse stacks are recorded. All metadata is included within the stack file itself, including the raw image metadata, experiment metadata (such as strain, media conditions, and experiment timings), and preprocessing metadata (calibration file locations and alignment warps). Finally, a '.avi' movie is automatically generated for each timelapse to facilitate manual experiment evaluation using an ImageJ macro.

Cell Region of Interest (ROI) Segmentation

The cell ROI segmentation process, in which the brightfield images are used to determine the locations of cell objects within each frame, is performed once per each of the 256 cell trap areas. During this step, the following actions are taken:

1. Collect and join into a single images the brightfield (dia) images from the eight cell traps for each frame

2. Execute the CellProfiler/Ilastik batch processing routine to produce per-cell, per-frame ROIs

3. Separate the segmented ROI results for each of the eight traps

To prepare for segmentation, the single-trap image stacks for each of the 256 trap regions are joined back together. While the segmentation process can indeed be performed on the single-trap stacks themselves, it was identified that performance is improved by segmenting 256 larger stacks each containing eight traps versus 2048 single-trap stacks.

The brightfield images at the center of the z-stack used for segmentation. Each image passes through a CellProfiler pipeline with the following steps:

1. Histogram equalization

2. Classification of pixels into probabilities for cell, membrane, background, and multilayer classes via Ilastik

3. Determine initial cell objects based on a filled watershed using the predicted membrane image

4. Measure the size, shape, and intensity of the initial cell objects

5. Filter the initial objects to remove non-cell and multilayer objects

6. Expand the identified cell objects slightly

7. Produce and record an indexed mask of the resulting cell objects

Histogram equalization is used to minimize variations from frame-to-frame or experiment-to-experiment in the brightfield intensity. The pixel classification executed using a previously-trained classifier using the 'Autocontext' workflow within Ilastik. A set of 13 histogram-equalized images from various experiments, trap areas, and experiment timings were used for training. Training consisted of manually labeling the

classes of representative pixels within these images as cell, membrane, background, or multilayer. The resulting classifier can then process novel images and generate the probabilities of for each pixel to be within each of the four classes.

Following the pixel classification, a watershed operation is performed using the predicted membrane image to identify an initial set of objects. This set of objects contains the cell objects, but also contains objects which contain no cells or, occasionally, regions in which the cells have grown in a manner in which the trap is no longer maintaining a monolayer of cells. These erroneous objects are identified and filtered using several criteria as follows using the CellProfiler 'FilterObjects' module:

1. LowerQuartileIntensity of the predicted cell minus the predicted background image within the object must be greater than 0.0

2. LowerQuartileIntensity of the predicted cell minus the predicted multilayer image within the object must be greater than 0.0

3. Eccentricty of the object is less than 1.0

4. Solidity of the object is greater than 0.8

5. Compactness of the object is less than 1.5

6. Area of the object is between 100 and 10000 pixels

7. FormFactor of the object is greater than 0.6

The resulting cell objects are then expanded by two pixels to better represent the cell areas. This was necessary as a result of conservative classification of the membrane regions between cell areas which substantially improved the earlier initial object detection. These final cell objects are then indexed and recorded, along with a diagnostic image of the identified contours overlayed upon the initial brightfield image. Finally, the resulting cell ROIs for the full eight-trap image are separated into results for their constituent individual traps.

Cell ROI Feature Extraction
During ROI feature extraction, the characteristics of all cell ROIs are summarized for all cells in all frames. This process is run once per each of the 2048 trap timelapses.

For each cell ROI, the following statistics are measured and recorded:

1. Frame index

2. ROI index

3. Area

4. Position of the center of the ROI

5. Fluorescence mean

6. Fluorescence background

7. Fluorescence mean minus background

The fluorescence background is calculated by measuring the 5th percentile of pixel intensities within a annulus around the ROI. The ring mask is produced by the difference of a slightly dilated ROI and the original ROI.

The results of the feature extraction for each trap are recorded in a .csv. In addition to the individual ROI measurements, this file also contains the timing and media condition data corresponding to the observed frames of the trap. The file name also contains the position index of the trap area as well as the strain identifier for the contained strain. In this manner, the resulting .csv files are self-contained for downstream analysis.

Timelapse Quality Assessment and Plotting
During the automated quality assessment and plotting process, the timelapses are evaluated and scored on a variety of quality metrics. Additionally, diagnostics plots are created for each timelapse. This process is run ones per each of the 2048 trap timelapses.

During this process the following steps are performed:

1. Calculate focus quality

2. Calculate timelapse alignment quality

3. Generate a motion map (y-vs-t correlation kymograph of the dia channel)

4. Calculate a score to assess lack of motion (i.e. growth quality) using the motion map

5. Measure cell counts within the different regions of the trap

6. Calculate epi map (y-vs-t kymograph of the epi channel)

7. Measure the extent of any multilayer growth within the trap

8. Generate quality assessment plots

9. Compare quality scores to pre-determined quality criteria and annotate the quality performance of each trap

Focus quality is scored by evaluating the spatial sharpness of the image. To do this, the image is first normalized using the 1st and 99th percentile values in order to remove any disruptive outliers. Then, the Laplacian of the image is calculated, which produces large values in areas with high spatial gradients (i.e. sharp regions). Finally, the focus score is calculated as the difference between the 99th and 1st percentile of the Laplacian image. High scores correspond to a large difference between the tails of the distribution of the Laplacian image, meaning that there were regions of high sharpness within the image, and conversely, low values correspond to images that are blurred.

The alignment quality of the timelapse is evaluated via two scores - one which compares the correlation of each frame to it's previous frame, and another which compares to the initial frame. These 2d correlation coefficients are calculated only for background regions of the frame in order to avoid regions with frame-to-frame motion due to cell growth. Four background regions are used, corresponding to the areas above, below, to the left, and to the right of the cell trap. For each set of two frames being compared, the highest of the four region's correlation coefficients is used as the score in order to avoid possible disturbances due to cells passing through the region.

A y-vs-t correlation kymograph of the dia channel, herein termed a 'motion map,' is calculated for each timelapse. This produces a single image that facilitates evaluation of the motion of cells within the trap throughout the timelapse. To produce this image, first a mask is used so that only the interior of the trap is evaluated. Then the Pearson correlation coefficient is calculated on pixels corresponding to each horizontal slice of the trap between each frame and it's subsequent. When the correlation is high, there is a strong correspondence between the pixels in one frame to the next for this particular horizontal slice. Performing this calculation across all horizontal slices for a set of two frames produces a column of pixels which summarizes the similarity between the slices for all slices. Subsequent columns can be calculated for subsequent frames, and these columns can be assembled into a y-vs-t image. In the motion map, regions of high correlation imply that no motion is occurring between frames within a portion of the trap. On the other hand, regions of low correlation mean that substantial differences are present from one frame to the next, implying motion within that part of the trap.

The motion map can be used to produce summary statistics of the motion within the trap. A 'no-motion score' is calculated by simply calculating the proportion of the trap which exceeds a threshold correlation value of 0.8 for each frame. This per-frame score is then smoothed with using a window of five frames.

A count of cells identified within various regions of the trap is performed. These regions include the trap itself, the sieve below the trap, the area above the trap, and the area below the trap.

A y-vs-t kymograph of the epi channel is calculated, termed the 'epi map'. This follows a simplified process compared to the motion map. For the epi map, for each frame the mean is calculated over the x-axis, producing a single column of values. These columns are then assembled into a y-vs-t image. The epi map facilitates evaluation of the fluorescence both over time and as a function of vertical position within the groove by expressing this information within a single image.

The extent of any multilayer growth within the trap is evaluated using the multilayer results from the Ilastik pixel classification process which was evaluated during image segmentation. The score is simply the area proportion of the trap which was classified as multilayer.

Next, a variety of plots are generated in order to summarize these quality characteristics as well as the cell behavior measured within each timelapse. This includes density plots of the distributions of the measured cell fluorescence statistics, cell area, and a summary of the noise characteristics of the population over time. The assessed quality scores are also summarized to facilitate quality assessment.

Finally, the automatically-generated quality scores can be compared with pre-determined exclusion and warning criteria, and a summary of the pass-fail results per-trap is recorded. By incorporating these automated results alongside direct evaluation, we performed a manual screening the trap timelapses.

Image Analysis Computer Hardware

Image analysis was performed on a custom-built workstation with dual ten-core Intel Xeon E5-2680 v2 CPUs, a total of 128 GB of RAM, and a 2 TB SSD for storage of the active dataset. With this CPU configuration, hyperthreading enables the utilization of up to 40 logical cores. The image analysis pipeline was run with 38 worker threads, with two cores remaining for system overhead. This allows for approximately 3 GB of RAM per thread. Typical MICROSTAR experiments will have around 120 dual-channel frames captured per timelapse at each of the 256 positions. Given our hardware configuration, these approximately 30,000 frames can be fully analysed within 24

hours.

Promoter Characterisation

The single cell data generated by the image processing pipeline was used to characterise the promoters of the 23 strains. Only the bottom four layers of cells were analysed. In Figure 2A to H, single cell data of each strain were binned according to time and concentration level. The time bin size was 1 hour. Concentrations were divided into 15 bins and the concentration level shown in the label is the mean of the interval. The line plot shows the median of each concentration bin with moving average (window size = 3). The same method applied to 2.13. The heatmaps in Figure 2A, C, E and G show the distribution of single cell values in each concentration bin. 50 equally sized GFP signal bins were used given the maximum and minimum value.

In Figure 2J and K, the heatmaps show the distributions of single cell data in each concentration bin for each strain. Only steady state data (10 to 15 hours of starvation) was used for this analysis. The concentration bin size is 7 μM, and 50 equally sized GFP signal bins were used given the maximum and minimum value on the y axis. For the line plot, it shows the median of each concentration bin with moving average (window size = 3). Concentrations were divided into 20 bins. The markers on the line plot is the median of corresponding bin range. The same method applied to Figure 2.14.

Figure 2.15 summarises normalised promoter activity across P_i levels and plots it as a heatmap. It is divided into four programs as shown in Figure 2I. An unsupervised hierarchical clustering was applied to the data shown in the heatmap and cluster 2 groups of promoters, named decoupled and plateau. The clustering was done by a clustering function AgglomerativeClustering in a machine learning package (scikit-learn) written in Python. The metric used to compute the distance is euclidean.

Pho4p Nuclear Signal Analysis

To measure nuclear Pho4p signal in intermediate P_i conditions proved to be challenging due the overall low abundance of Pho4p and the limited degree to which Pho4p localized to the nucleus in intermediate P_i conditions. We developed a simple, effective, and intuitive method that merely quantifies pixel intensity changes in image histograms and used a simple cut-off value to classify cytoplasmic and nuclear signal. Although this method is straightforward, it required a clean dataset as the method is sensitive to the occurrence of dead or dying cells, which become brightly fluorescent. Therefore, a quality control screening step was performed. All time-lapse videos were

manually annotated and only perfect microgrooves were included in the analysis. The screening criteria were: (1) the microgroove maintained good growth, (2) no dead cells formed within the ROI, and (3) no dying cells or unhealthy cells were present in the ROI. The above criteria were applied to a time window ranging from 3 hours before starvation to 3 hours before the end of experiment. Videos were cropped to 115x115 pixels (ROI) which covered the bottom ~4 cell layers (approximately 8 cells per frame) and a portion of background. Videos were binned into 10 P_i concentrations (1 in full, 3 in plateau, 4 in bi-stable, and 2 in off state). 8 of the 10 concentration bins have 2-3 chip replicates and all of them have chamber and groove replicates. In total, 230 microgrooves were analyzed. In the analysis, the maximum intensity of each pixel position in the z stack was used in the analysis. Any pixel intensity higher than 3000 was filtered. Figure 2.33 shows the histograms of Pho4p signal across the 10 concentration bins. To plot Pho4p nuclear signal as a function of P_i concentration, images from 2 hours into starvation to 2 hours before the end of starvation were analyzed. Any pixel intensity higher than 850 was defined as nuclear signal and below 850 considered to be cytoplasmic. The nuclear signal of each concentration bin is the mean nuclear signal of microgrooves in each bin (sum of nuclear signal divided by no. of sample grooves in the bin). The result is reported in Figure 2F with a moving average (window size =2).

2.5.7 Micro-Chemostat-Array (MCA)

Device Design

The microchemostat array (MCA) device used in this study is based on a previous work (81) and includes design modifications to adapt to the requirements of this study. The device design is shown in Figure 2.4. The MCA contains 1,152 cell culture chambers (48 columns x 24 rows) which support on-chip cell culture with continuous media supply. Unlike the MICROSTAR chip, cells are deposited directly onto the coverslip by live-cell spotting using a microarrayer (Arrayit, U.S.A.). Cell-arraying greatly increases the number of strains that can be analyzed on a single chip. The original MCA was able to measure 384 strains under three different media conditions. In this study, we required higher throughput in terms of conditions (multiple P_i concentrations). We therefore divided the MCA into 12 zones and each zone was connected to an outlet of a dilution generator to measure 96 strains under 12 P_i concentrations (normally 10 intermediate P_i levels, one 0 P_i, and one P_i-rich).

We also modified the cell culture chamber design. The original chamber design is a wide compartment which allows cells to grow freely in any direction. To better define

and constrict the direction in which cells will be pushed due to growth we included two vertical walls in the imaging ROI in each chamber (see Figure 2.5A). These walls guide the cells and minimize lateral movement.

Mold Fabrication

The mold fabrication workflow is identical to the procedure described above for the MICROSTAR device. However, different types of SU-8 (GM series, Gersteltec, Switzerland) were used and therefore the fabrication parameters differed (see Table 2.10, 2.11 and 2.12). The major difference is that these types of photoresist require ramp heating and cooling in baking steps instead of a step temperature.

Device Fabrication

The chip fabrication method is identical to the procedure described above for the MICROSTAR device.However, the device was bonded onto epoxy coated coverslip instead of plasma bonding.

Cell Culturing

The glycerol stocks stored at -80°C were used to inoculate a 96-well plate with 100 μL YPD in each well. The plates were then placed in a plate shaker for two days at 30°C. This long incubation period rendered the cells more robust to physical arraying.

Cell Arraying

We use a commercial microarrayer (Qarray2, Genetix, UK) equipped with 4 arraying pins (SMP3, ArrayIt Corp, Sunnyvale, CA) to transfer cells from the 96-well plate onto an epoxy coated coverslip. The duration of spotting was around 50 min. The spotting pins were cleaned and dried after each spot deposition using water and compressed air. The humidity in the spotting chamber was maintained at 75%. After spotting, the MCA device was aligned to the spotted coverslip under a stereo microscope.

2.5.8 MCA Experiments

Chip Priming

All control lines were connected to tubing pre-loaded with oil (Fluorinert FC 40, CAS# 51142-49-5) and pressurised at 41.3 kPa to out-gas prime the control lines. Pressure

was then increased to 151 kPa and each on-chip microvalve was inspected to ensure that they were fully closed. The above procedure does not apply to the pressure element located in the roof of each microchemostat. This element was primed before device-to-coverslip alignment. If this element was primed with cells underneath, the pressure caused cells to enter the sieve channel or break through the chamber outlet valve. The button valve remained un-pressurised until just before 90x time lapse imaging started.

MCA has the same media switch unit as MICROSTAR. The priming step is also the same as shown in Figure 2.26A. After that, media was allowed to enter the dilution generator and prime it while valve "DG1" remained closed. Then, "DG1" was opened allowing media to fill the upper flow channels. Upper "CP2" and lower channel control "CP1" remained closed to direct flow towards the cell culture chambers in order to ensure that cells remained in the chamber and avoid cross-contamination (Figure 2.5d(a) and (b)). After that, the pressure of chamber outlet valve "CC1" was gradually decreased from 151 kPa to 20.6 kPa. Media then started to perfuse the cell culture chamber and cells (Figure 2.5d(c)). Once the cell culture chamber was filled, media would enter the sieve channels and then enter the lower flow channel. At that moment, the lower flow channel control "CP1" was opened to prime the lower flow channel and the chamber outlet valve was closed to prevent cell exit (Figure 2.5d(d)). The chip priming procedure was completed when all air bubbles were successfully removed by out-gas priming.

On-chip Cell Culture

The next step is initial growth of the cell spot to fill the cell culture chamber. Normally, live cell spots take 15-20 hours to recover from spotting. During this period, time lapse imaging was performed with a 4x objective. These images are used for manual quality annotation. Once the cells filled up the chambers, the chamber outlet valve was opened to allow excessive cells to exit via the upper flow channel, and the button valve was pressurised at 13.7 kPa to keep cells growing in a mono-layer (Figure 2.5e(a) and (b)). Similar to MICROSTAR experiments, a typical experiment consists of three phases, (1) pre-starvation, (2) starvation, and (3) recovery (Figure 2.5e(b), (c) and (d)). The duration of these phases were 6, 16, and 24 hours. The media in negative control zone remained the same throughout the experiment.

Time-lapse Acquisition

The hardware, software and imaging parameters/conditions are identical to the procedure described above for the MICROSTAR device.

2.5.9 MCA Image Processing and Data Analysis

Image Processing

The image analysis process for the MCA experiments largely follows the procedure described above for the MICROSTAR devices.

During the image alignment portion of the preprocessing phase, the background identification process cannot be performed in the same manner, as the MICROSTAR approach relies on identifying the periodic spacing of the cell traps. Instead, for the MCA images, the background region determination is performed by first calculating an adaptive binary threshold on the dia image followed by identification of the large region of background along the bottom of the frame and along a vertical column in the middle of the frame between the two chambers. The performance of this mask generation is susceptible to disruption by cells which have made their way into the channel below the chambers, as well as occasional dust or debris which makes identification of the central column difficult. Thus, if the mask was not able to be identified as desired (that is, the central column was identified with a width of less than 25 pixels), then the background mask will be recalculated with a new bias term included on the binary threshold. This process will repeat with new bias values until either a successful mask is generated or until a maximum number of bias values have been attempted, in which case the previous frame's warp will be used and a warning will be logged. In the rare event that the background extraction fails on the reference frame (i.e. first frame), then the alignment of the of chambers in that position will fail to occur.

During segmentation, a new Ilastik autoclassification pixel classifier was trained on and for MCA images.

The full suite of quality assessment analyses were not adapted for MCA image analysis. Only plots summarizing the population fluorescence and area characteristics were generated.

2.5.10   Transport Models

The carrier model was as previously described (209). The high affinity transporters are expressed under conditions of low substrate availability and the low affinity transporters when substrate is plentiful. They suggested that an increased affinity not only increased the rate of the substrate influx into the cell, but also that of substrate efflux out of cell. The main difference between high and low affinity facilitated-diffusion transporters is that for the high-affinity transporters often both sides are saturated with substrate, while for the low-affinity transporter only the extracellular side is. A high affinity of transporter will cause both the inward facing and the outward facing binding sites of the transporter to be saturated with substrate. As a result, the efflux rate will be nearly as high as the influx rate, and the net uptake is very low. A low affinity of the transporter reduces the saturation of the transporter at the intracellular side. When the extracellular substrate concentration is high enough, the transporter will still be saturated at the extracellular side. The efflux will be reduced and the net uptake rate increases. The transporter switches between conformation with an inward-facing and outward-facing substrate binding site. The conformation change results in translocation of the substrate over the membrane. Transport of substrate s over a membrane by means of facilitated diffusion can generally be described by a four-step process: (i) extracellular substrate s_e to carrier binding, (ii) transport of s over the membrane, (iii) release of s in the cytoplasm, (iv) return of the substrate-binding site to the periplasm-facing position. The step (iv) distinguishes this model from reversible Michaelis-Menten kinetics.

The carrier model by facilitated diffusion underlying Equation:

$$v = e_t \frac{\frac{k_{cat}}{K_M}(s_e - s_i)}{1 + \frac{s_e}{K_M} + \frac{s_i}{K_M} + \alpha \frac{s_e s_i}{K_M^2}} \tag{2.1}$$

with

$$k_{cat} = \frac{k_2 k_4}{k_2 + k_4} \tag{2.2}$$

$$K_M = 2 K_D \frac{k_4}{k_2 + k_4} \tag{2.3}$$

$$\alpha = 4 \frac{k_2 k_4}{(k_2 + k_4)^2} \tag{2.4}$$

87

Solving for s_i:

$$s_i = \frac{s_e - \frac{vK_M}{e_t k_{cat}} - \frac{v s_e}{e_t k_{cat}}}{1 + \frac{v}{e_t k_{cat}} + \alpha \frac{v s_e}{e_t k_{cat} K_M}} \tag{2.5}$$

The Michaelis-Menten model was also as described previously (209):

$$v_{MM} = e_t \frac{\frac{k_{cat}}{K_M}(s_e - s_i)}{1 + \frac{s_e}{K_M} + \frac{s_i}{K_M}} \tag{2.6}$$

and solving for s_i:

$$s_{i_{MM}} = \frac{\frac{v}{e_t} + \frac{v s_e}{e_t K_M} - \frac{k_{cat} s_e}{K_M}}{\frac{-k_{cat}}{K_M} + \frac{-v}{e_t K_M}} \tag{2.7}$$

We used the following parameters for all plots: $k_2 = 4$, $k_4 = 4$ and otherwise only adjusted K_D or e_t values as indicated in the figure captions.

2.6   Supplemental Information

2.6.1   Supplemental Figures
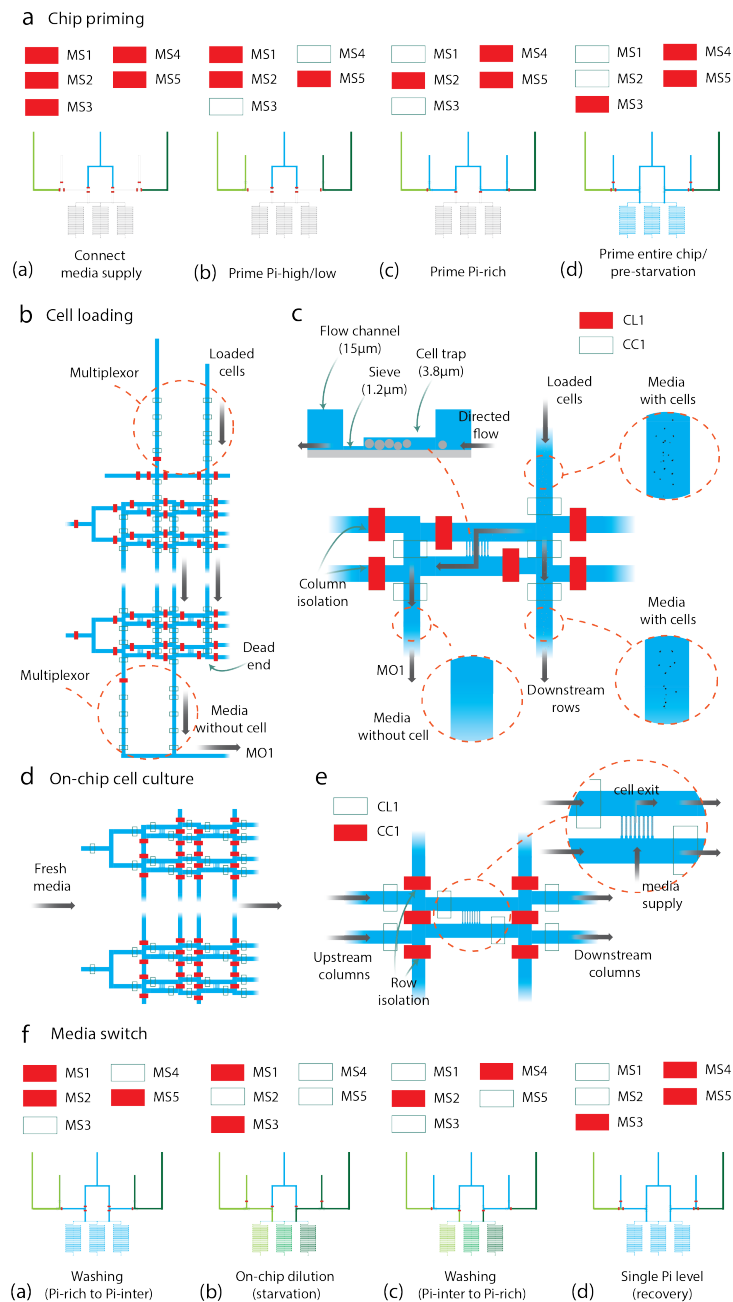
Figure 2.26: **MICROSTAR experimental workflow. a.** Microvalve settings for the chip priming process. These settings apply to both the MCA and MICROSTAR. **b.** Microvalve settings for the cell loading process. **c.** A close-up schematic of a MICROSTAR chamber during the cell loading process. A cross-sectional view is also shown. **d.** Microvalve settings for the cell culture process. **e.** A close-up schematic of a MICROSTAR chamber during the cell culture process. **f.** Microvalve settings for media switching (applies to both the MCA and MICROSTAR devices).
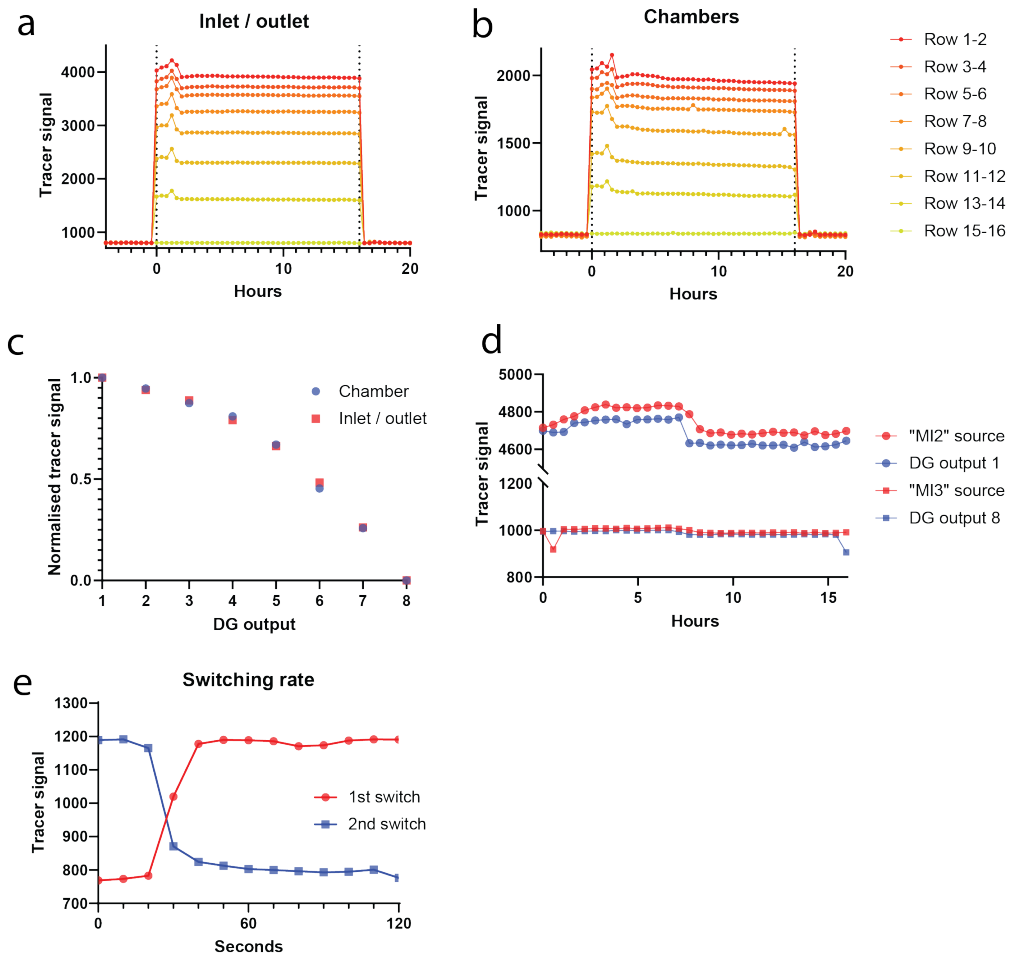
Figure 2.27: **Dilution generation validation and characterisation. a.** Tracer signals measured at the inlets and outlets of the eight media zones in the MICROSTAR array over time. **b.** Tracer signals measured at the chambers in the eight media zones in the MICROSTAR array over time. **c.** Normalised mean tracer signal of the inlets and outlets and chamber measured at the eight media zones. **d.** Tracer signals measured at the two media inlets for the dilution generator, the inlets of the first media zone (DG output1), and the inlets of the eighth media zone (DG output8). **e.** Media switching rate measured in a MICROSTAR chamber. The red line indicates the switch for "starvation", and the blue line indicate the switch for "recovery".
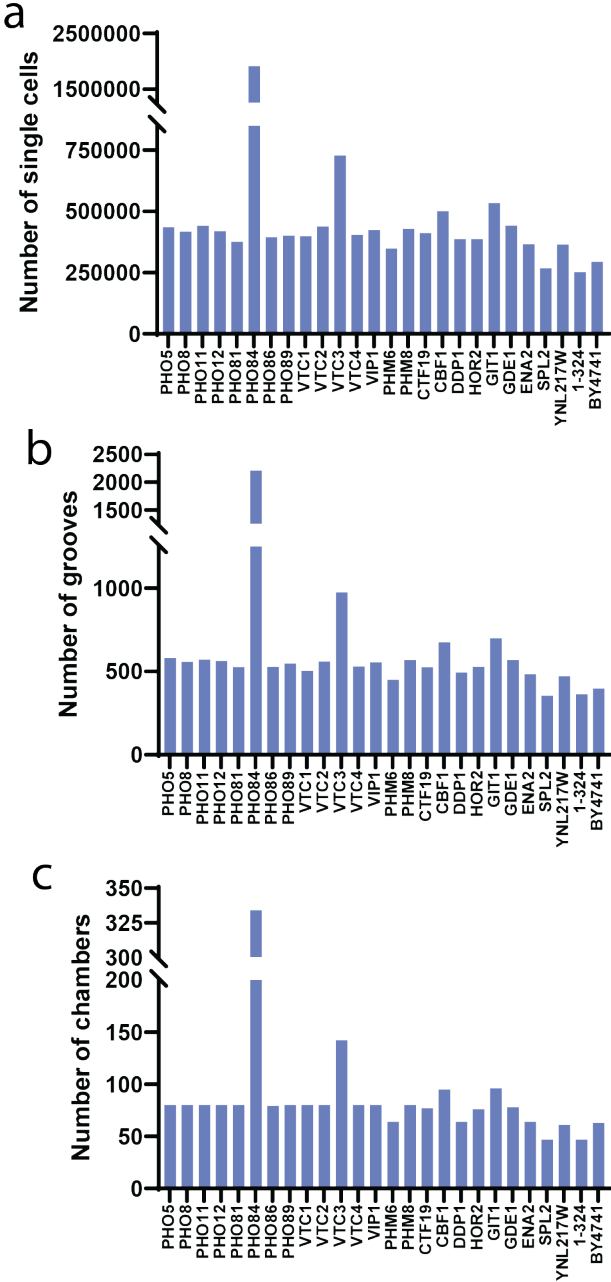
Figure 2.28: **MICROSTAR promoter expression sample sizes. A,** Number of single cells, **B,** number of grooves, and **C,** number of chambers measured for each strain in MICROSTAR promoter expression experiments.
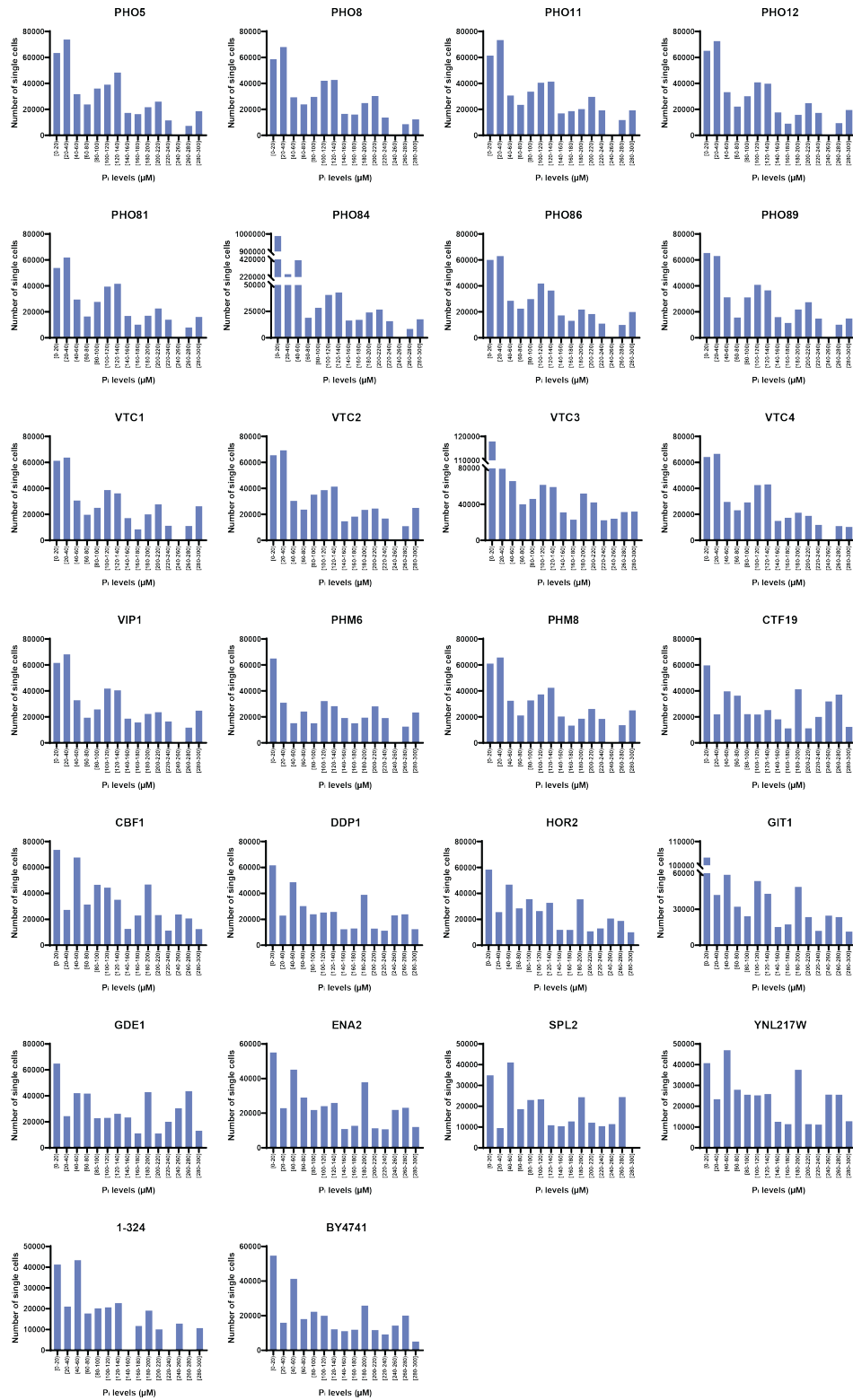
Figure 2.29: **Number of single cells measured across P_i concentrations for each strain in MICROSTAR promoter expression experiments.**
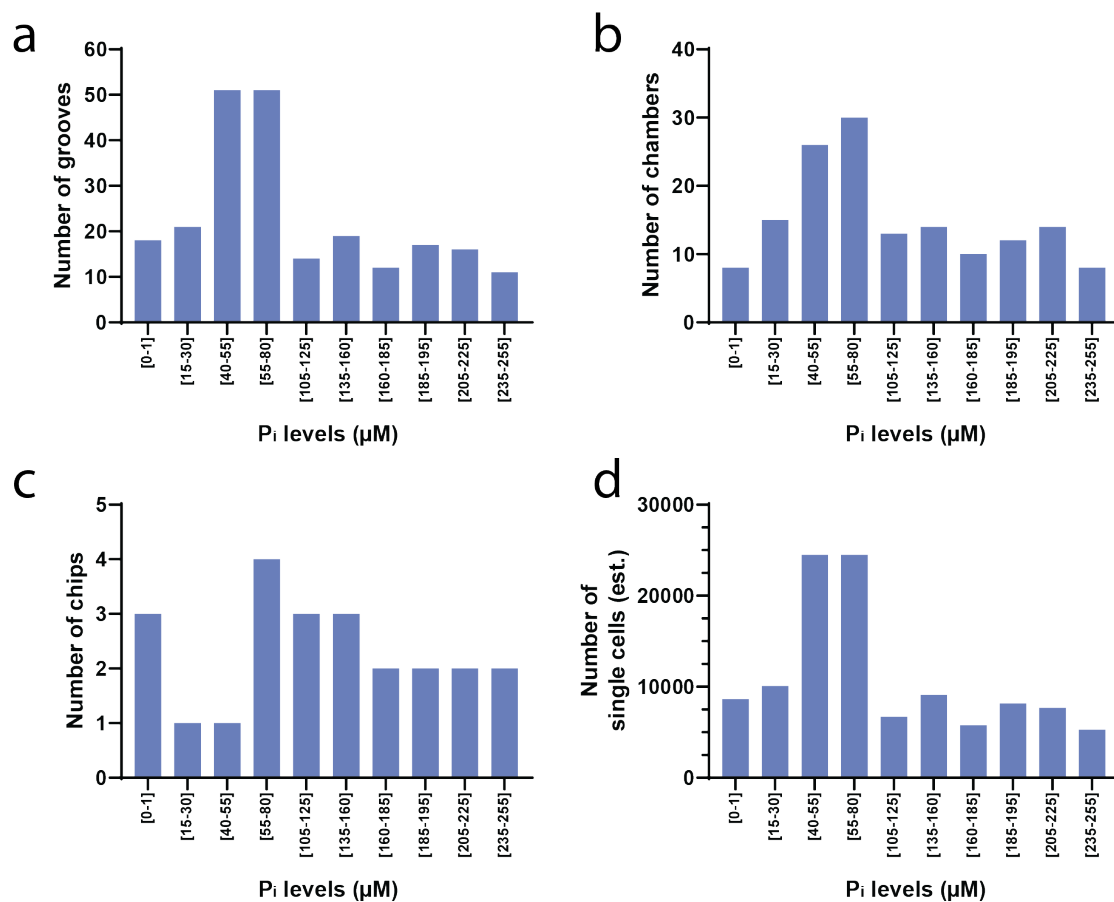
Figure 2.30: **Pho4p nuclear localization analysis sample size. A,** Number of grooves, **B,** number of chambers, **C,** number of chips, and **D,** estimated number of single cells measured in each P_i concentration bin for Pho4p nuclear localization analysis.
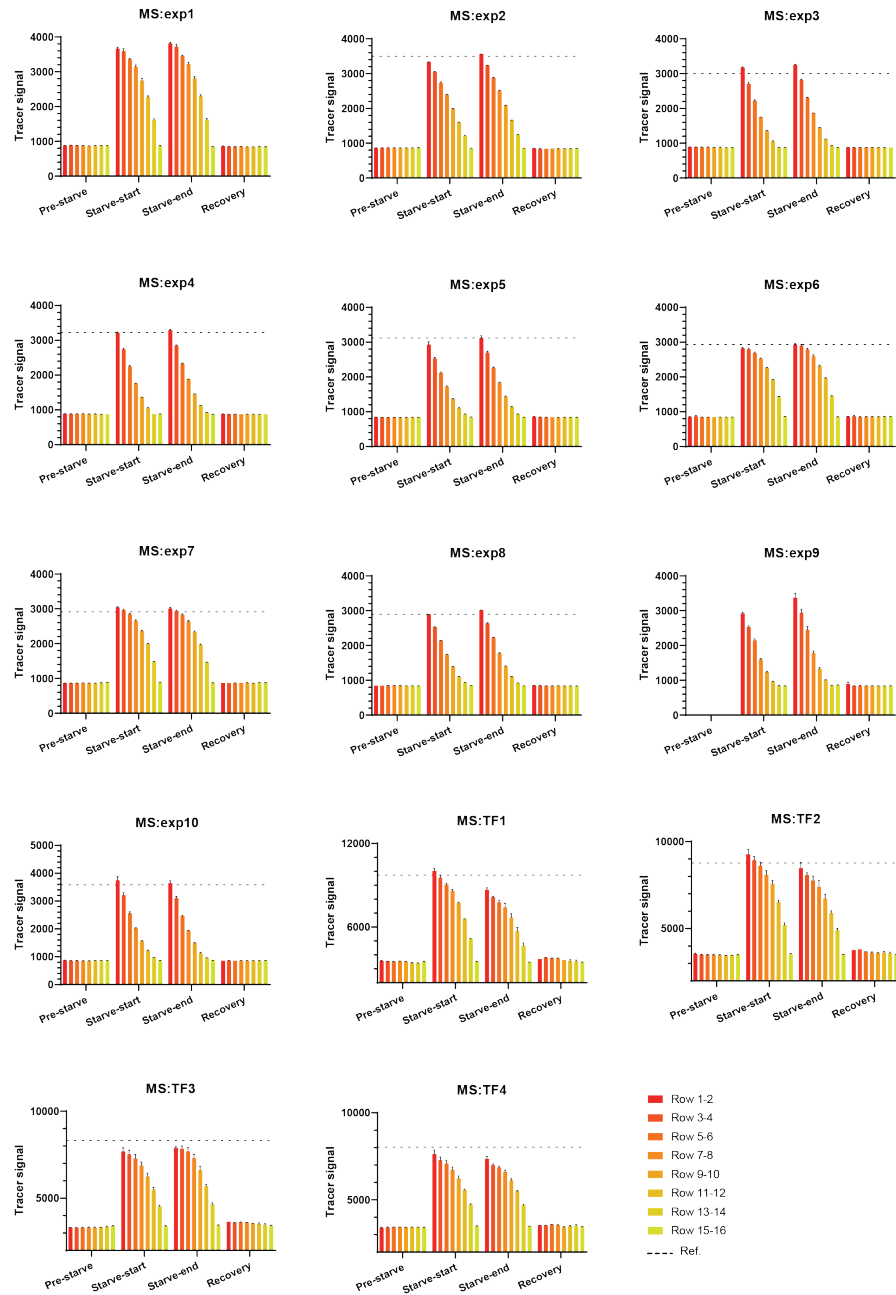
Figure 2.31: **Tracer signals across rows at four time points in MICROSTAR experiments.**
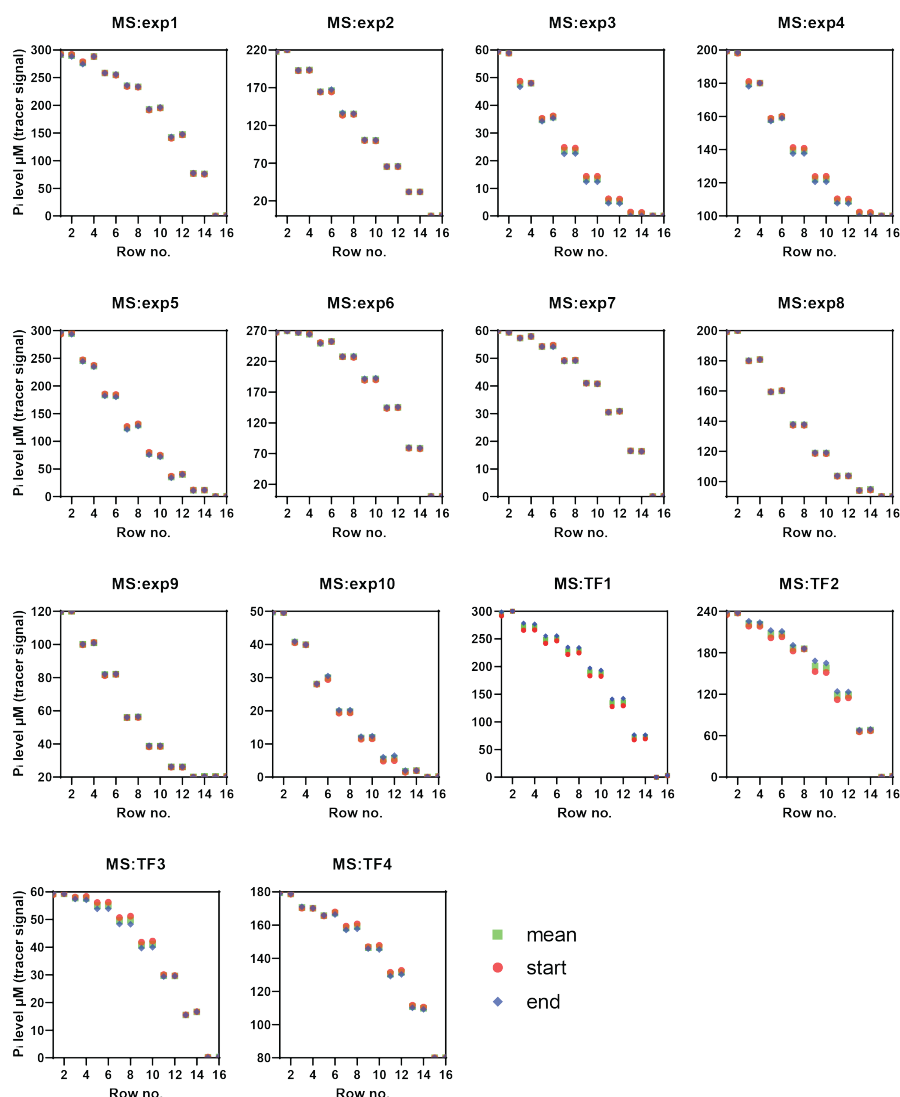
Figure 2.32: **Dilution generator calibrations for MICROSTAR experiments.** "Exp" indicates promoter characterisation experiments and "TF" denotes transcription factor localization experiments.
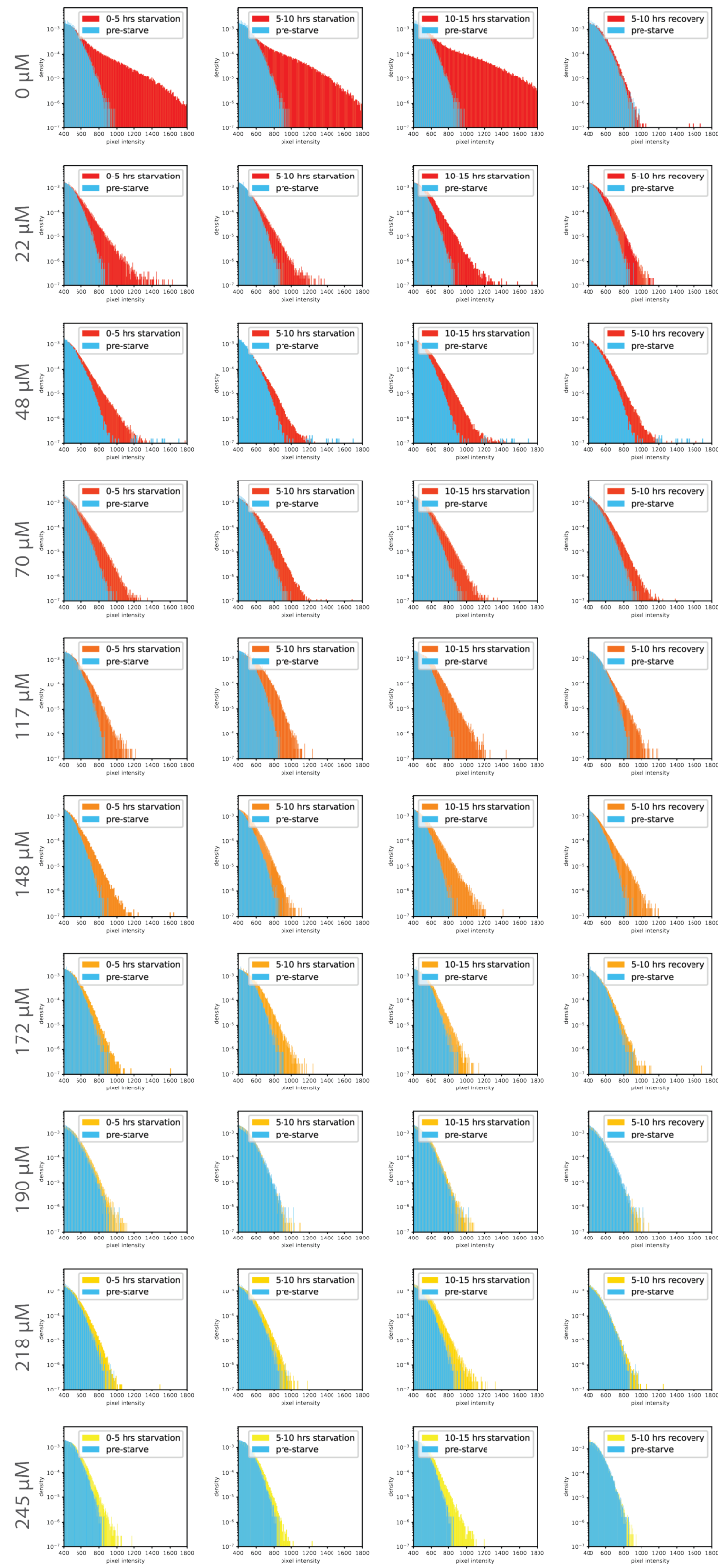
Figure 2.33: **Histograms of Pho4p signal at different P_i levels.**
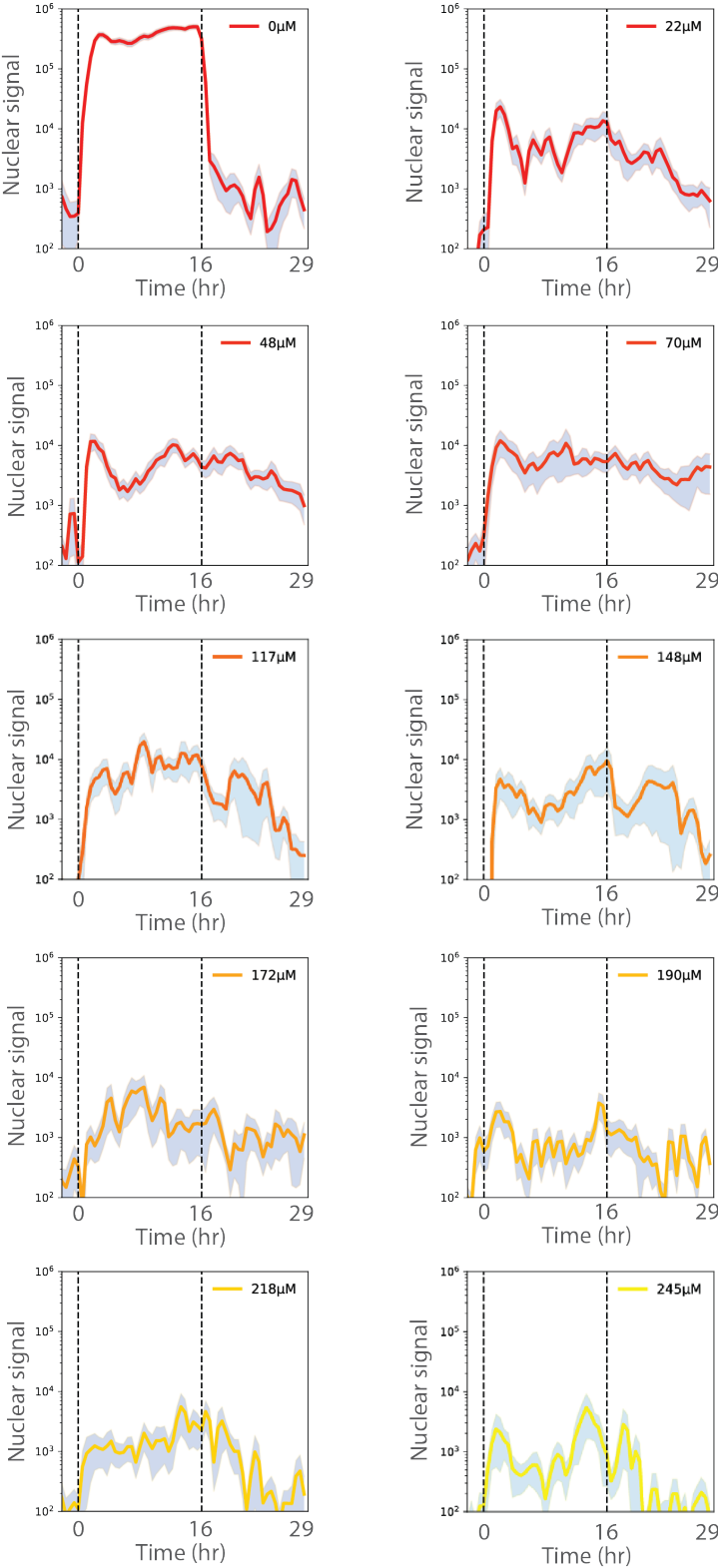
Figure 2.34: **Nuclear signal showing standard error (shaded region).**

2.6.2   Supplemental Tables

Table 2.1: Media, chemicals, enzymes, etc. used in this study.

Reagent or Kit	Source	Identifier
YPD broth	Sigma Aldrich	Cat# Y1375-1KG
Glucose	Sigma Aldrich	Cat# G8270-1KG
LB powder	AppliChem GmbH	Cat# A0954
Agar	Sigma Aldrich	Cat# 05039-500G
Chloramphenicol	Sigma Aldrich	Cat# C0378-5G
Ampicillin	Roche	Cat# 10835242001
Kanamycin Sulfate	Life Technologies	Cat# 11815024
Yeast Synthetic Drop-out Medium Supplements	Sigma Aldrich	Cat# Y1501-20G
Yeast Nitrogen Base Without Amino Acids	Sigma Aldrich	Cat# Y0626-1KG
Uracil	Sigma Aldrich	Cat# U1128-25G
EZSolution 5-Fluoroorotic acid monohydrate (5-FOA)	Biovision	Cat# 2797-10
YNB with ammonium sulfate, powder	MP Biomedicals	Cat# MP4027-812
Complete Supplement Mixture (CSM), powder	MP Biomedicals	Cat# MPB-114500022
Adenine hemisulfate salt	Sigma Aldrich	Cat# A3159-25G
L-Tryptophan	Sigma Aldrich	Cat# 1083740100
Monobasic potassium phosphate	Sigma Aldrich	Cat# P5655-100G
Sodium chloride	Sigma Aldrich	Cat# S9888-25G
Potassium chloride	Sigma Aldrich	Cat# P5405-250G
Phire Plant Direct PCR Master Mix	Molecular Dimensions	Cat# 133083
NeXtalStock PEG 3,350 (200) - 200 ml	Sigma Aldrich	Cat# 05039-500G
QIAprep Spin Miniprep Kit (250)	QIAGEN	Cat# 27106
Wizard SV Gel and PCR Clean-Up System	Promega Corp	Cat# A9281
QIAquick Gel Extraction Kit (50)	QIAGEN	Cat# 28704
BsaI-HF	New England Biolabs Inc.	Cat# R3733L
BsmBI-v2	New England Biolabs Inc.	Cat# R0739L
EcoRV	New England Biolabs Inc.	Cat# R0195L
T4 DNA ligase	New England Biolabs Inc.	Cat# M0202L
NotI-HF (High Fidelity)	New England Biolabs Inc.	Cat# R3189L

NEB 10-beta Competent E. coli (High Efficiency)	New England Biolabs Inc.	Cat# C3019H
Midori Green Advance (1 ml)	Nippon Genetics Europe	Cat# MG04
UltraPure Salmon Sperm DNA Solution	Life Technologies	Cat# 15632011
500 mL Vacuum Filter/Storage Bottle System	Corning	Cat# 431097

Table 2.2: Plasmids used or generated in this study.

Plasmid ID	Strain Name	Source	Identifier
pYTK001-096	Yeast MoClo Toolkit	John Dueber lab	Addgene# 1000000061
pTY001	LF_PHO5_yoEGFP_tADH1	This work	N/A
pTY002	LF_PHO89_yoEGFP_tADH1	This work	N/A
pTY003	LF_PHM6_yoEGFP_tADH1	This work	N/A
pTY004	LF_PHO84_yoEGFP_tADH1	This work	N/A
pTY005	LF_PHO11_yoEGFP_tADH1	This work	N/A
pTY006	LF_PHO12_yoEGFP_tADH1	This work	N/A
pTY007	LF_VTC1_yoEGFP_tADH1	This work	N/A
pTY008	LF_VTC3_yoEGFP_tADH1	This work	N/A
pTY009	LF_VTC4_yoEGFP_tADH1	This work	N/A
pTY010	LF_VTC2_yoEGFP_tADH1	This work	N/A
pTY011	LF_PHO8_yoEGFP_tADH1	This work	N/A
pTY012	LF_PHO86_yoEGFP_tADH1	This work	N/A
pTY013	LF_VIP1_yoEGFP_tADH1	This work	N/A
pTY014	LF_PHO81_yoEGFP_tADH1	This work	N/A
pTY015	LF_HOR2_yoEGFP_tADH1	This work	N/A
pTY016	LF_PHM8_yoEGFP_tADH1	This work	N/A
pTY017	LF_CTF19_yoEGFP_tADH1	This work	N/A
pTY018	LF_SPL2_yoEGFP_tADH1	This work	N/A
pTY019	LF_CBF1_yoEGFP_tADH1	This work	N/A
pTY020	LF_DDP1_yoEGFP_tADH1	This work	N/A
pTY021	LF_ENA1_yoEGFP_tADH1	This work	N/A
pTY022	LF_ENA2_yoEGFP_tADH1	This work	N/A
pTY023	LF_GDE1_yoEGFP_tADH1	This work	N/A
pTY024	LF_GIT1_yoEGFP_tADH1	This work	N/A
pTY025	LF_YAR070C_yoEGFP_tADH1	This work	N/A
pTY026	LF_YJL119C_yoEGFP_tADH1	This work	N/A
pTY027	LF_YNL217W_yoEGFP_tADH1	This work	N/A
pTY028	NN_PHO5_yoEGFP_tPHO5	This work	N/A
pTY029	NN_PHO89_yoEGFP_tPHO89	This work	N/A
pTY030	NN_PHM6_yoEGFP_tPHM6	This work	N/A
pTY031	NN_PHO84_yoEGFP_tPHO84	This work	N/A
pTY032	NN_PHO11_yoEGFP_tPHO11	This work	N/A
pTY033	NN_PHO12_yoEGFP_tPHO12	This work	N/A
pTY034	NN_VTC1_yoEGFP_tVTC1	This work	N/A
pTY035	NN_VTC3_yoEGFP_tVTC3	This work	N/A
pTY036	NN_VTC4_yoEGFP_tVTC4	This work	N/A

pTY037	NN_VTC2_yoEGFP_tVTC2	This work	N/A
pTY038	NN_PHO8_yoEGFP_tPHO8	This work	N/A
pTY039	NN_PHO86_yoEGFP_tPHO86	This work	N/A
pTY040	NN_VIP1_yoEGFP_tVIP1	This work	N/A
pTY041	NN_PHO81_yoEGFP_tPHO81	This work	N/A
pTY042	NN_HOR2_yoEGFP_tHOR2	This work	N/A
pTY043	NN_PHM8_yoEGFP_tPHM8	This work	N/A
pTY044	NN_CTF19_yoEGFP_tCTF19	This work	N/A
pTY045	NN_SPL2_yoEGFP_tSPL2	This work	N/A
pTY046	NN_CBF1_yoEGFP_tCBF1	This work	N/A
pTY047	NN_DDP1_yoEGFP_tDDP1	This work	N/A
pTY048	NN_ENA1_yoEGFP_tENA1	This work	N/A
pTY049	NN_ENA2_yoEGFP_tENA2	This work	N/A
pTY050	NN_GDE1_yoEGFP_tGDE1	This work	N/A
pTY051	NN_GIT1_yoEGFP_tGIT1	This work	N/A
pTY052	NN_YAR070C_yoEGFP_tYAR070C	This work	N/A
pTY053	NN_YJL119C_yoEGFP_tYJL119C	This work	N/A
pTY054	NN_YNL217W_yoEGFP_tYNL217W	This work	N/A
pTY055	LYS2 sgRNA1	This work	N/A
pTY056	LYS2 sgRNA2	This work	N/A
pTY057	pWS158 [o]	Ellis lab	Addgene# 90517
pTY058	pWS082 - sgRNA Entry Vector	Ellis lab	Addgene# 90516
pTY059	VTC3 (YPL019C) sgRNA1	This work	N/A
pTY060	VTC3 (YPL019C) sgRNA2	This work	N/A
pTY061	CTF19 (YPL018W) sgRNA1	This work	N/A
pTY062	CTF19 (YPL018W) sgRNA2	This work	N/A
pTY063	PHO84 (YML123C) sgRNA2	This work	N/A
pTY064	PHO84 (YML123C) sgRNA1	This work	N/A
pTY065	VIP1 (YLR410W) sgRNA1	This work	N/A
pTY066	VIP1 (YLR410W) sgRNA2	This work	N/A
pTY067	CBF1 (YJR060W) sgRNA2	This work	N/A
pTY068	CBF1 (YJR060W) sgRNA1	This work	N/A
pTY069	YJL119C (YJL119C) sgRNA1	This work	N/A
pTY070	YJL119C (YJL119C) sgRNA2	This work	N/A
pTY071	VTC4 (YJL012C) sgRNA2	This work	N/A
pTY072	VTC4 (YJL012C) sgRNA1	This work	N/A
pTY073	SPL2 (YHR136C) sgRNA1	This work	N/A
pTY074	SPL2 (YHR136C) sgRNA2	This work	N/A

pTY075	PHO5 (YBR093C) sgRNA1	This work	N/A
pTY076	PHO5 (YBR093C) sgRNA2	This work	N/A
pTY077	PHO81 (YGR233C) sgRNA1	This work	N/A
pTY078	PHO81 (YGR233C) sgRNA2	This work	N/A
pTY079	VTC2 (YFL004W) sgRNA1	This work	N/A
pTY080	VTC2 (YFL004W) sgRNA2	This work	N/A
pTY081	PHO8 (YDR481C) sgRNA1	This work	N/A
pTY082	PHO8 (YDR481C) sgRNA2	This work	N/A
pTY083	PHM8 (YER037W) sgRNA1	This work	N/A
pTY084	PHM8 (YER037W) sgRNA2	This work	N/A
pTY085	GIT1 (YCR098C) sgRNA2	This work	N/A
pTY086	GIT1 (YCR098C) sgRNA1	This work	N/A
pTY087	PHO89 (YBR296C) sgRNA1	This work	N/A
pTY088	PHO89 (YBR296C) sgRNA2	This work	N/A
pTY089	YAR071W (PHO11) sgRNA (OT)	This work	N/A
pTY090	YDR039C (ENA2) sgRNA (OT)	This work	N/A
pTY091	YAR070C sgRNA (OT)	This work	N/A
pTY092	YHR215W (PHO12) sgRNA	This work	N/A
pTY093	YDR040C (ENA1) sgRNA	This work	N/A
pTY094	YDR281C (PHM6) sgRNA	This work	N/A
pTY095	YER062C (HOR2) sgRNA	This work	N/A
pTY096	YER072W (VTC1) sgRNA	This work	N/A
pTY097	YJL117W (PHO86) sgRNA	This work	N/A
pTY098	YOR163W (DDP1) sgRNA	This work	N/A
pTY099	YNL217W sgRNA	This work	N/A
pTY100	YPL110C (GDE1) sgRNA	This work	N/A
pTY101	Pho4p knockout sgRNA1	This work	N/A
pTY102	Pho4p knockout sgRNA2	This work	N/A
pTY103	mScarlet-i Pho4p N-tagged	This work	N/A
pTY104	Pho4p mScarlet-i C-tagged	This work	N/A

Table 2.3: Yeast strains used or generated in this study.

Strain ID	Strain Name	Cassette	Source	Identifier
sTY001	BY4741	N/A	HorizonDiscovery	Cat# YSC1048
sTY002	mScarleti-Pho4p	N/A	This work	N/A
sTY003	Yeast GFP library, Pho87-GFP	N/A	ThermoFisher	Cat# 95700
sTY004	Yeast GFP library, Pho90-GFP	N/A	ThermoFisher	Cat# 95700
sTY005	Yeast GFP library, SPL2-GFP	N/A	ThermoFisher	Cat# 95700
sTY006	LF library, PHO5_yoEGFP_tADH1	pTY001	This work	N/A
sTY007	LF library, PHO89_yoEGFP_tADH1	pTY002	This work	N/A
sTY008	LF library, PHM6_yoEGFP_tADH1	pTY003	This work	N/A
sTY009	LF library, PHO84_yoEGFP_tADH1	pTY004	This work	N/A
sTY010	LF library, PHO11_yoEGFP_tADH1	pTY005	This work	N/A
sTY011	LF library, PHO12_yoEGFP_tADH1	pTY006	This work	N/A
sTY012	LF library, VTC1_yoEGFP_tADH1	pTY007	This work	N/A
sTY013	LF library, VTC3_yoEGFP_tADH1	pTY008	This work	N/A
sTY014	LF library, VTC4_yoEGFP_tADH1	pTY009	This work	N/A
sTY015	LF library, VTC2_yoEGFP_tADH1	pTY010	This work	N/A
sTY016	LF library, PHO8_yoEGFP_tADH1	pTY011	This work	N/A
sTY017	LF library, PHO86_yoEGFP_tADH1	pTY012	This work	N/A
sTY018	LF library, VIP1_yoEGFP_tADH1	pTY013	This work	N/A
sTY019	LF library, PHO81_yoEGFP_tADH1	pTY014	This work	N/A
sTY020	LF library, HOR2_yoEGFP_tADH1	pTY015	This work	N/A
sTY021	LF library, PHM8_yoEGFP_tADH1	pTY016	This work	N/A

sTY022	LF library, CTF19_yoEGFP_tADH1	pTY017	This work	N/A
sTY023	LF library, SPL2_yoEGFP_tADH1	pTY018	This work	N/A
sTY024	LF library, CBF1_yoEGFP_tADH1	pTY019	This work	N/A
sTY025	LF library, DDP1_yoEGFP_tADH1	pTY020	This work	N/A
sTY026	LF library, ENA1_yoEGFP_tADH1	pTY021	This work	N/A
sTY027	LF library, ENA2_yoEGFP_tADH1	pTY022	This work	N/A
sTY028	LF library, GDE1_yoEGFP_tADH1	pTY023	This work	N/A
sTY029	LF library, GIT1_yoEGFP_tADH1	pTY024	This work	N/A
sTY030	LF library, YAR070C_yoEGFP_tADH1	pTY025	This work	N/A
sTY031	LF library, YJL119C_yoEGFP_tADH1	pTY026	This work	N/A
sTY032	LF library, YNL217W_yoEGFP_tADH1	pTY027	This work	N/A
sTY033	NN library, PHO5_yoEGFP_tPHO5	pTY028	This work	N/A
sTY034	NN library, PHO89_yoEGFP_tPHO89	pTY029	This work	N/A
sTY035	NN library, PHM6_yoEGFP_tPHM6	pTY030	This work	N/A
sTY036	NN library, PHO84_yoEGFP_tPHO84	pTY031	This work	N/A
sTY037	NN library, PHO11_yoEGFP_tPHO11	pTY032	This work	N/A
sTY038	NN library, PHO12_yoEGFP_tPHO12	pTY033	This work	N/A
sTY039	NN library, VTC1_yoEGFP_tVTC1	pTY034	This work	N/A
sTY040	NN library, VTC3_yoEGFP_tVTC3	pTY035	This work	N/A
sTY041	NN library, VTC4_yoEGFP_tVTC4	pTY036	This work	N/A

sTY042	NN library, VTC2_yoEGFP_tVTC2	pTY037	This work	N/A
sTY043	NN library, PHO8_yoEGFP_tPHO8	pTY038	This work	N/A
sTY044	NN library, PHO86_yoEGFP_tPHO86	pTY039	This work	N/A
sTY045	NN library, VIP1_yoEGFP_tVIP1	pTY040	This work	N/A
sTY046	NN library, PHO81_yoEGFP_tPHO81	pTY041	This work	N/A
sTY047	NN library, HOR2_yoEGFP_tHOR2	pTY042	This work	N/A
sTY048	NN library, PHM8_yoEGFP_tPHM8	pTY043	This work	N/A
sTY049	NN library, CTF19_yoEGFP_tCTF19	pTY044	This work	N/A
sTY050	NN library, SPL2_yoEGFP_tSPL2	pTY045	This work	N/A
sTY051	NN library, CBF1_yoEGFP_tCBF1	pTY046	This work	N/A
sTY052	NN library, DDP1_yoEGFP_tDDP1	pTY047	This work	N/A
sTY053	NN library, ENA1_yoEGFP_tENA1	pTY048	This work	N/A
sTY054	NN library, ENA2_yoEGFP_tENA2	pTY049	This work	N/A
sTY055	NN library, GDE1_yoEGFP_tGDE1	pTY050	This work	N/A
sTY056	NN library, GIT1_yoEGFP_tGIT1	pTY051	This work	N/A
sTY057	NN library, YAR070C_yoEGFP_tYAR070C	pTY052	This work	N/A
sTY058	NN library, YJL119C_yoEGFP_tYJL119C	pTY053	This work	N/A
sTY059	NN library, YNL217W_yoEGFP_tYNL217W	pTY054	This work	N/A

Table 2.4: Oligonucleotides used for PCR or sequencing validation.

Oligonucleotides	Name	Source
LF lib (LYS2 locus) validation		
5'-TATGCTGGATCTGGTAGAGTGGAG-3'	LYS2 Seq Fw	This work
5'-TGCGTCAAGGGCTGAAAAGAC-3'	LYS2 Seq Rev	This work
NN lib (Native locus) validation		
5'-TGCCCACACATCTAATCAAACG-3'	CBF1 Native Fw	This work
5'-TAATTCCCTCTTTTATGCTTTAGTATCG-3'	CBF1 Native Rev	This work
5'-AATGGGCCCTCTCGATATTG-3'	CTF19 Native Fw	This work
5'-CTCATATTAAACCCATAGGAGACG-3'	CTF19 Native Rev	This work
5'-CCATTAGTAATATGGCATGGAACC-3'	DDP1 Native Fw	This work
5'-TAATTCCCTCTTTTATGCTTTAGTATCG-3'	DDP1 Native Rev	This work
5'-GTTTGTTAGGGCAGGGATGTAG-3'	ENA1 Native Fw	This work
5'-GAAGGAAAATAAAATTCTTTCCTCTTTC-3'	ENA1 Native Rev	This work
5'-CGAACTTTTCACAGGTTCTGAAC-3'	ENA2 Native Fw	This work
5'-CTTTGCCTCGAGAGATAATATGGAG-3'	ENA2 Native Rev	This work
5'-AACATGGTGTTCCAAAGCAC-3'	GDE1 Native Fw	This work
5'-GAAAATCTTGAAAGATCTGGGTATCG-3'	GDE1 Native Rev	This work
5'-GAAAGCAGAGAATCAAAAGAAGC-3'	GIT1 Native Fw	This work
5'-AACATGGTGTTCCAAAGCAC-3'	GIT1 Native Rev	This work
5'-GTTTGGCCTTCTCAAGTATTTTGG-3'	HOR2 Native Fw	This work
5'-CTCTGGAATCTTCCCCACTG-3'	HOR2 Native Rev	This work
5'-CAAATGTGCATTAGCGTGTAATG-3'	PHM6 Native Fw	This work
5'-GGATGCGTTGACGCTAATG-3'	PHM6 Native Rev	This work
5'-AGGTGAACGAAGAAAAAAAAAAG-3'	PHM8 Native Fw	This work
5'-ATGATGATGTTAGAGAATTCTTTCTG-3'	PHM8 Native Rev	This work
5'-CATTCATCGTGGGTGTCTAATAAAG-3'	PHO11 Native Fw	This work
5'-CTTGAATAAAAAGCAATGTAGAGAACAG-3'	PHO11 Native Rev	This work
5'-GTGGGTGTCTAATAAAGTTTTAATGAC-3'	PHO12 Native Fw	This work
5'-CCAATAATTTTGAATAAAAGAAGCAC-3'	PHO12 Native Rev	This work
5'-ACGTATTTGGAAGTCATCTTATGTG-3'	Pho5 Native Fw	This work
5'-ACCTGCATTAACTAAAGCAGCG-3'	Pho5 Native Rev	This work
5'-GCAGACCACAGGGTAGTCAAC-3'	PHO8 Native Fw	This work
5'-AGGAAGAAGTTGGCTGGTAGATC-3'	PHO8 Native Rev	This work
5'-AAACGCGCTAATTGCATCAG-3'	PHO81 Native Fw	This work
5'-GATGGACACCTTTCCATATTG-3'	PHO81 Native Rev	This work
5'-TGCCCACACATCTAATCAAACG-3'	PHO84 Native Fw	This work
5'-TAATTCCCTCTTTTATGCTTTAGTATCG-3'	PHO84 Native Rev	This work
5'-GAGGAGAGACAACCCTGTTCTC-3'	PHO86 Native Fw	This work
5'-AGATGGTTCCGATGGTTGC-3'	PHO86 Native Rev	This work

5'-AGACCTTTTTTTTCTTTTTCTGC-3'	Pho89 Native Fw	This work
5'-GATAACTGTAAGTCAAAAGGCCATG-3'	Pho89 Native Rev	This work
5'-ATGTTTCCTACCCCAATGATGGTTCG-3'	SPL2 Native Fw	This work
5'-AGTTGGCGGCGGTCGTGG-3'	SPL2 Native Rev	This work
5'-TATTCGCATCATATATCTGAAATGTTTTAC-3'	VIP1 Native Fw	This work
5'-TTTAATAAGGCAATAATATTAGGTATGTagatatac-3'	VIP1 Native Rev	This work
5'-GGCAGAGGAGAGTTATCACTCC-3'	VTC1 Native Fw	This work
5'-CGAGCCTCTTTAACCTAATGC-3'	VTC1 Native Rev	This work
5'-ACGAGGTTCCCAGTTTCCC-3'	VTC2 Native Fw	This work
5'-ATAGAGAATAGTGGAATAAGGGTGG-3'	VTC2 Native Rev	This work
5'-GTTACATTACCGGCCACATAGTAG-3'	VTC3 Native Fw	This work
5'-TCTTTTTTTATAAATACCACTAAAACTGG-3'	VTC3 Native Rev	This work
5'-TAATTGTCATACGCATTGTTATACC-3'	VTC4 Native Fw	This work
5'-TTCCAGATTGTTTGGAATTACC-3'	VTC4 Native Rev	This work
5'-CTGAGCTTGAATCTATCGAAATG-3'	YAR070C Native Fw	This work
5'-GAAATTATATCCAACAGAAAGCTCAG-3'	YAR070C Native Rev	This work
5'-GAAAAAAAGATGTTCTTTGGTCG-3'	YJL119C Native Fw	This work
5'-CTATAATTGGTTTGACCATTTTTTC-3'	YJL119C Native Rev	This work
5'-GGAGGAGCTGGGTTATCATAAAG-3'	YNL217W Native Fw	This work
5'-CAGAACCACTGTCAACAACGAC-3'	YNL217W Native Rev	This work
GFP-tag lib (Native locus) validation		
5'-TGCCCACACATCTAATCAAACG-3'	CBF1pr Seq Fw	This work
5'-GGAACTAGTAACAAAATACGACACG-3'	CTF19pr Seq Fw	This work
5'-GCGGCTTAATTCTCAAACGTC-3'	DDP1pr Seq Fw	This work
5'-GGTCTAGAAAAGGCTGCTCC-3'	ENA1pr Seq Fw	This work
5'-GGGCTAGGTCTAGAAAAGGC-3'	ENA2pr Seq Fw	This work
5'-GGAGAGCTTGGGATTTACACG-3'	GDE1pr Seq Fw	This work
5'-CGGTACCATTATCATCTCTCAACC-3'	HOR2pr Seq Fw	This work
5'-GGATGTTCTGCCCTATAATGTTACG-3'	PHM6pr Seq Fw	This work
5'-CAGAACGCCTAATCGAATCGTATTATC-3'	PHM8pr Seq Fw	This work
5'-GGTATTGCCAAGAGATTAAACAAGG-3'	PHO11pr Seq Fw	This work
5'-GGTATTGCCAAGAGATTAAACAAGG-3'	PHO12pr Seq Fw	This work
5'-TGAATACGACACAACCTACTTGG-3'	PHO5pr Seq Fw	This work
5'-AGAATTGATCATGCTGGTCACC-3'	PHO8pr Seq Fw	This work
5'-GACTTCGGTAAAATTGATGGAAGC-3'	PHO81pr Seq Fw	This work
5'-TGTTCTAACCATTGGGCTGC-3'	PHO86pr Seq Fw	This work
5'-GCCCTTATTGTTCTGGTTAGTAATCTC-3'	PHO89pr Seq Fw	This work
5'-GCAATTACAACTACAACGCCG-3'	SPL2pr Seq Fw	This work

5'-CCCAAATCTTCTTCTCAATTCGATG-3'	VIP1pr Seq Fw	This work
5'-TGGCAAGAGCGATTAAGTATAGC-3'	VTC1pr Seq Fw	This work
5'-GGTATCTCAAGAGGAAAATGAGCG-3'	VTC2pr Seq Fw	This work
5'-AGAATACTCGCCACGTTATTGC-3'	VTC3pr Seq Fw	This work
5'-CAGGAAACCTCCACTACCAAC-3'	VTC4pr Seq Fw	This work
5'-TGCTGGGTGACTTTATTCACAAG-3'	YNL217Wpr Seq Fw	This work
5'-CACTCTTGAAAAAGTCATGCCG-3'	GFP Seq Rev	This work

Table 2.5: sgRNA sequences.

Target loci	Landing pad sequence
Pho4p sgRNA1	ACTGGCGTCTTTAATCCCCG
Pho4p sgRNA2	CAACTGCTACAATCAAGCCG
LYS2 sgRNA1	GATAAATTCACAATGCTGAG
LYS2 sgRNA2	CTAATAACAATCAACCCACG
VTC3 sgRNA1	CATCAGAAAGAACCCTCAGG
VTC3 sgRNA2	AGAAAGGGGTACTCAAAAGG
CTF19 sgRNA1	ACGCGCACTGAAGCTACAGG
CTF19 sgRNA2	ATGGGCAGAAAACAGTTACG
PHO84 sgRNA1	ACTACCGTGCCAGTAAACGT
PHO84 sgRNA2	GCTTGGGGTCAAATCTCCGG
VIP1 sgRNA1	TGATCATCTTGAATACCGGG
VIP1 sgRNA2	ATGTAAGACCCCTCCGTACG
CBF1 sgRNA1	CGAGGAGCAGAACTACAGCG
CBF1 sgRNA2	AGACGTCATTCCCTCCAGGG
YJL119C sgRNA1	ATACAGTTGATACAGAACGT
YJL119C sgRNA2	CACCACACAGGTTTTACGTG
VTC4 sgRNA1	TAGGTCAAGAACCTCCAGAG
VTC4 sgRNA2	CGAAGATAACGACTTCGATG
SPL2 sgRNA1	TGTGACTGCGATGTTCTACG
SPL2 sgRNA2	GGGTATTTACAGAGAGCCAG
PHO5 sgRNA1	GTCTTAGCCAGACTGACAGT
PHO5 sgRNA2	GTTACAAGCACTCAAAGTGT
PHO81 sgRNA1	GTGCATTCAAAACATCAACG
PHO81 sgRNA2	ACTATTGATACGTTATGGGG
VTC2 sgRNA1	TCCTAAGCAAGCATATGAGG
VTC2 sgRNA2	ACATTGGCGTAGACCAACAC
PHO8 sgRNA1	CAAACCAATACAGACCACAG
PHO8 sgRNA2	AAGTCTACTACTCTCCCCAG
PHM8 sgRNA1	CCACTACGACAGACCCATCG
PHM8 sgRNA2	TCCGGATCCGGTAATTTGGG
GIT1 sgRNA1	ATAGAAAACACCTCTAACAG
GIT1 sgRNA2	TAATAAAATATGATGCCGAG
PHO89 sgRNA1	TAGTAAGCATTAAAACAGCG
PHO89 sgRNA2	CATTGGTGGGATATCATCAG
PHO11 sgRNA	CCAAAAGTATCATGACAACA
ENA2 sgRNA	CACAGCGAAAGAGATCGCCA
YAR070C sgRNA	ACTTTGCACAGGAATCGTGT
PHO12 sgRNA	TGTGAAATAAATGACTTCTA

ENA1 sgRNA	CGTCCACAGATTGAAAACAG
PHM6 sgRNA	TACATCATCGATGCACCTCG
HOR2 sgRNA	ACCGTCGACGTCGAACAAAG
VTC1 sgRNA	AATTCAGTAAACCTACACCA
PHO86 sgRNA	TCTGAAAATCGACTACTAGG
DDP1 sgRNA	TGAAGACATGAGACCCCCTA
YNL217W sgRNA	AAACTGGTGGAGACAACCGA
GDE1 sgRNA	TGATTCGGCATCTCCGAACG

Table 2.6: Media and plate recipes.

Ingredient	Amount
Synthetic complete, phosphate-rich media (per liter)	
Glucose	20 g
YNB with ammonium sulfate, without phosphates	5.6 g
Complete Supplement Mixture	790 mg
Adenine hemisulfate salt	163.5 mg
L-Tryptophan	50 mg
Monobasic potassium phosphate	1 g
Sodium chloride	0.1 g
Synthetic complete, phosphate-free media (per liter)	
Glucose	20 g
YNB with ammonium sulfate, without phosphates	5.6 g
Complete Supplement Mixture	790 mg
Adenine hemisulfate salt	163.5 mg
L-Tryptophan	50 mg
Potassium chloride	547.8 mg
Sodium chloride	0.1 g
Synthetic complete, uracil dropout plate (per liter)	
Glucose	40 g
Yeast synthetic dropout supplement, without uracil	1.92 g
YNB without Amino Acids	6.7 g
Agar	40 g
Note: 20 mL per plate	
Synthetic complete, uracil dropout plate (per liter)	
Glucose	40 g
Yeast synthetic dropout supplement, without uracil	1.92 g
YNB without Amino Acids	6.7 g
Agar	40 g
Uracil (760 μg/mL)	66 mL
5-FOA (10%)	10 mL
Note: 20 mL per plate	

Table 2.7: MICROSTAR mold fabrication: coating and soft bake parameters

#	Layer	Resist	Speed	Time	PB temp.	PB time	SB temp.	SB time
		Spin coating				**Soft bake**		
f1	sieve	SU8-2	2500 rpm	30s	65°C	1min	95°C	1min
f2	chamber	SU8-2	1000 rpm	30s	65°C	1min	95°C	3min
f3	dg	SU8-3025	1875rpm	40s	none	none	95°C	13.3min
f4	flow	AZ 10XT-60	1000rpm	40s	none	none	105°C	6.5min
c1	control	SU8-3025	4144rpm	40s	none	none	95°C	10min

Table 2.8: MICROSTAR mold fabrication: exposure and post exposure bake parameters

#	Layer	Power (mW/cm2)	Time	Dose (mJ/cm2)	PB temp.	PB time	PEB temp.	PEB time
		Exposure				**Post exposure bake**		
f1	sieve	20	3.2s	64	65°C	1min	95°C	1min
f2	chamber	20	6.4s	128	65°C	1min	95°C	1min
f3	dg	20	10.8s	216	65°C	1min	95°C	4.3min
f4	flow	20	15x10s	3000	none	none	none	none
c1	control	20	7.5s	150	65°C	1	95°C	3min

Table 2.9: MICROSTAR mold fabrication: development and hard bake parameters.

#	Layer	Developer	Time	Rinser	Rinsing time	HB temp.	HB time	Final height
		Exposure				**Hard bake**		
f1	sieve	none	none	none	none	none	none	1.2 μm
f2	chamber	PGMEA	1min	2-propanol	1min	135	2 hrs	3.8 μm
f3	dg	PGMEA	2x3.5min	2-propanol	1min	135	2 hrs	40 μm
f4	flow	AZ400K 1:3.5	12min	DI water	1min	115	1 hrs	15 μm
c1	control	PGMEA	2x2.5min	2-propanol	1min	135	2 hrs	20 μm

Table 2.10: MCA mold fabrication: coating and soft bake parameters

#	Layer	Resist	Speed	Time	Start temp.	Ramp time	Step temp.	Step time	End temp.	Ramp time
		Spin coating			Soft bake					
f1	sieve	GM1040	3616 rpm	40s	30°C	20min	130°C	5min	30°C	20min
f2	chamber	GM1050	1803 rpm	40s	30°C	25min	130°C	5min	30°C	25min
f3	dg	GM1070	1933rpm	40s	30°C	50min	130°C	5min	30°C	50min
f4	flow	AZ 10XT-60	1000rpm	40s	none	none	105°C	6.5min	none	none
c1	control	GM1060	1793rpm	40s	30°C	30min	130°C	5min	30°C	30min

Table 2.11: MCA mold fabrication: exposure and post exposure bake parameters

#	Layer	Power (mW/cm2)	Time	Dose (mJ/cm2)	Start temp.	Ramp time	Step temp.	Step time	End temp.	Ramp time
		Exposure			Post exposure bake					
f1	sieve	20	3s	60	30°C	20min	90°C	15min	30°C	90min
f2	chamber	20	6.4s	128	30°C	20min	90°C	20min	30°C	90min
f3	dg	20	16.6s	332	30°C	40min	90°C	40min	30°C	90min
f4	flow	20	15x10s	3000	30°C	none	none	none	none	none
c1	control	20	9.4s	188	30°C	30min	90°C	30min	30°C	90min

Table 2.12: MCA mold fabrication: development and hard bake parameters

#	Layer	Developer	Time	Rinser	Rinsing time	HB temp.	HB time	Final height
		Development				Hard bake		
f1	sieve	none	none	none	none	none	none	1.5 μm
f2	chamber	PGMEA	1+0.5min	2-propanol	1min	135	2 hrs	4.2 μm
f3	dg	PGMEA	2+1min	2-propanol	1min	135	2 hrs	40 μm
f4	flow	AZ400K 1:3.5	12min	DI water	1min	115	1 hrs	15 μm
c1	control	PGMEA	2x1min	2-propanol	1min	135	2 hrs	12 μm

Table 2.13: List of strains analyzed in each MICROSTAR experiment

	exp1	exp2	exp3	exp4	exp5	exp6	exp7	exp8	exp9	exp10
PHO5	■	■	■	■					■	
PHO8	■	■	■	■					■	
PHO11	■	■	■	■					■	
PHO12	■	■	■	■					■	
PHO81	■	■	■	■					■	
PHO84	■	■	■	■					■	■
PHO86	■	■	■	■					■	
PHO89	■	■	■	■					■	
VTC1	■	■	■	■					■	
VTC2	■	■	■	■					■	
VTC3	■	■	■	■	■	■	■	■	■	
VTC4	■	■	■	■					■	
VIP1	■	■	■	■					■	
PHM6	■	■	■	■						
PHM8	■	■	■	■					■	
CTF19					■	■	■	■		
CBF1					■	■	■	■		
DDP1					■	■	■		■	
HOR2					■	■	■	■		
GIT1			■	■	■	■	■	■		
GDE1					■	■	■	■		
ENA2					■	■	■	■		
SPL2					■	■	■	■		
YNL217W					■	■	■	■		
1-324					■		■	■		
BY4741					■	■	■	■		

3 Robustness of the yeast inorganic phosphate gene regulatory network

Contribution: S.C. designed and performed the experiments, analyzed data.

3.1 Introduction

The robustness of gene regulatory networks, which refers to their ability to maintain their functions despite environmental perturbations or genetic variations, is crucial for the stability and adaptability of biological systems (210; 211; 212). Transcription Factors (TFs) regulate gene expression through dynamic interactions with regulatory DNA elements. TFs typically bind to DNA in a sequence-specific manner and can recognize motifs of various lengths. Mutations in the TF binding domain can have deleterious effects by significantly altering binding affinity and specificity. However, certain TFs exhibit robustness to mutations, where amino acid substitutions in the DNA binding domain may have minimal impact on binding specificity. For example, the human helix-loop-helix transcription factor Max binds to DNA through five specific residues, and changes to three of these residues do not affect its binding specificity (46). TFs bind to their target sites by diffusing through the cellular volume, whether directly from the surrounding solution or by migrating from previously bound nonspecific sites. The likelihood of DNA binding typically increases linearly with the concentration of the TF protein, regardless of how it arrives at the target site (213). While some TFs have low cellular concentrations with only a few hundred copies per cell, others exhibit significant variation, with levels that can change by more than 1000-fold in response to growth phase or external stimuli (214; 215). Modulating the cellular concentration of a TF protein can contribute to transcriptional activity by altering the DNA-residence time of the protein, which cells can achieve through a concentration-dependent dissociation mechanism or the promotion of multiprotein-DNA complexes (216). Additionally, TF levels can modulate the exposure of a gene to transcription-related proteins by altering chromosomal structure in a concentration-dependent manner, thereby serving as a regulatory component. The master regulator of the Pho regulon, Pho4, whom blongs to the bHLH family, is considered to be expressed at a constitutive level (193). Despite extensive research on the Pho regulon, it remains unclear how robust the regulon is to various mutations in the binding domain of Pho4, and the extent to which the expression levels of Pho4 may impact the sensitivity of the Pho regulon has yet to be determined. In this chapter, we studied the robustness of the yeast inorganic phosphate gene regulatory network in response to the concentration changes and genetic mutations of the master regulator Pho4 by 1) by tuning the expression level of Pho4 with a Z3EV-mediated inducible system (217); 2) mutating the basic binding domain of Pho4. All the strains were sequence verified by Sanger sequencing.

3.2 Results

3.2.1 Tuning the Pho4 level using a Z3EV inducible system

Construction of the input-out strain library

To determine the robustness of the native phosphate GRN in response to different levels of Pho4, we constructed a Z3EV-mediated inducible phosphate gene regulatory network, where the inputs are different inducible Pho4 cassettes and the outputs are Pho4-regulated promoter cassettes (Figure 3.1). The input cassette comprises two single unit transcription cassettes with two HO locus homology arms. One consists of the REV1 promoter, Z3EV coding sequences and the ScTDH1 terminator and the other consists of Gal1-Z3EV promoter, Pho4 variants including native Pho4, nucleus-localized Pho4, native Pho4 N-terminus tagged with yeast codon optimized mScarlet-i (yomScarlet-i), nucleus-localized Pho4 N-terminus tagged with yomScarlet-i, and the ScTDH1 terminator. The Z3EV protein is an artificial transcription factor, with a fusion of the ZIF268 DNA binding domain, the ligand binding domain of the human estrogen receptor, and the viral protein 16 (VP16). The input level could be tuned by changing the extracellular concentration of the inducer, i.e., β-estradiol. After adding β-estradiol, Z3EV could translocate into nucleus and bind to the Gal1-Z3EV promoter and then the downstream genes could be expressed. The advantages of this induction system (217) are that β-estradiol is not consumed by cells and it has a fast and linear dose response from concentration of 0 to 1 μM. To make the Pho4 constitutively nuclear localized, point mutations (serine to alanine mutation) were made to five phosphorylation sites (SP1, SP2, SP3, SP4, and SP6) of the native Pho4 as described by Komeili and O'Shea (108). The output cassettes consist of insulated native promoter, yoEGFP, and a fixed ADH1 terminator, which were integrated into the Lys2 locus. We selected 23 Pho4 regulated promoters, including CBF1, CTF19, DPP1, ENA2, GDE1, GIT1, HOR2, PHM6, PHO11, PHO12, PHO5, PHO8, PHO81, PHO84, PHO86, PHO89, SPL2, VIP1, VTC1, VTC2, VTC3, VTC4, YNL217W promoters. To construct the input-out library, I started with a commercially available Pho4 knockout (KO) yeast strain where the Pho4 was replaced with a KanMX selection marker and integrated the input and output cassettes in two rounds (Figure 3.1). In the first-round integration, the inducible Pho4 cassette was integrated into HO locus of the parent strain by CRISPR-Cas9 assisted homologous recombination (Figure 2.2). The inducible integration cassette was flanked by two 500 bps HO locus homology arms. In the second-round integration, the output cassettes were integrated into the Lys2 locus.
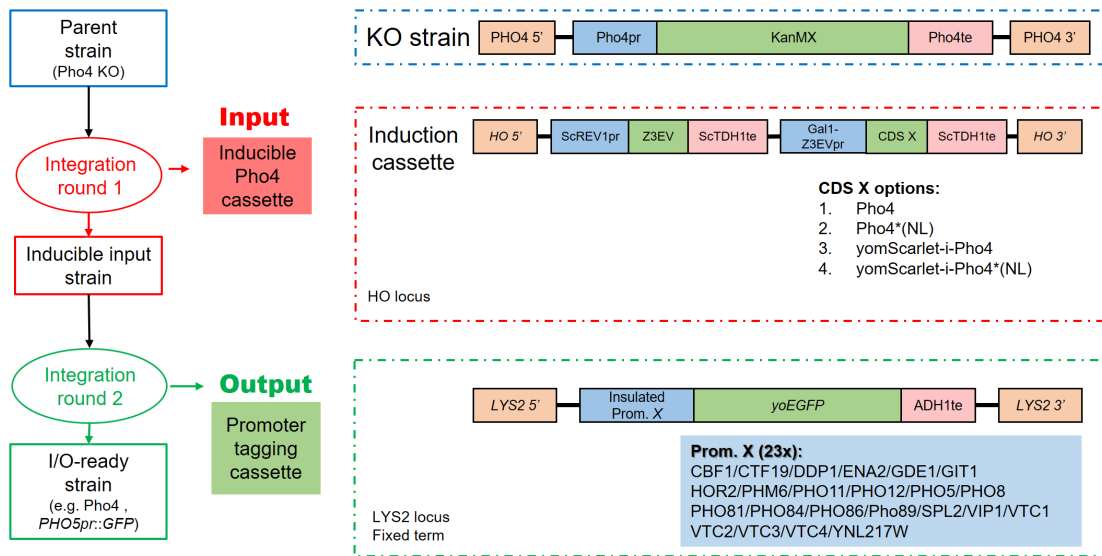
Figure 3.1: **Strain library construction of the Z3EV-mediated inducible GRN.** The parent strain is a yeast Pho4 knockout strain, where Pho4 is replaced with a KanMX selection marker. In the first-round integration, the induction Pho4 cassettes were integrated into HO locus. We included four Pho4 variants in the this study, including Pho4, Pho4*(NLS), yomScarleti-Pho4, yomScarleti-Pho4*(NLS). In the second-round integration, the yoEGFP-tagged promoter cassettes were integrated into the Lys2 locus. We included 23 promoters and generated a total number of 92 inducible input-output strains.

Parental strain validation via colony PCR

I verified the Pho4 KO strain by two colony PCRs (Figure 3.2). The first was external-external colony PCR with primers binding outside the Pho4 or KanMX region, which could result a product of 1880 bp with Pho4 KO strain and a product of 1386 bp with the wild type BY4741 strain. The other was internal-internal colony PCR with primers binding inside the Pho4, which could result no product with Pho4 KO strain and a product of 881 bp with the wild type BY4741 strain.
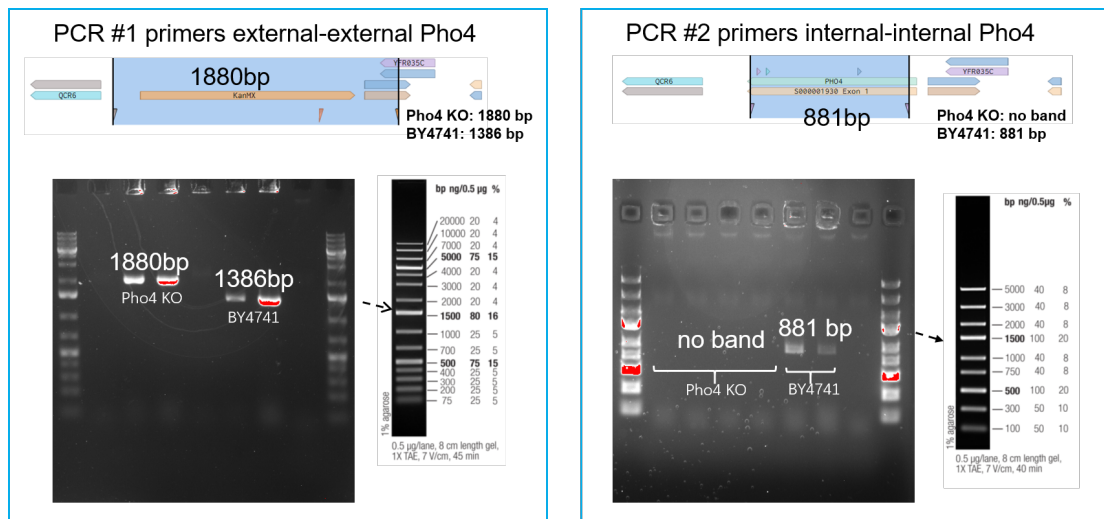
Figure 3.2: **Validation of Pho4 KO strain with colony PCRs.** The first colony PCR was with external-external primers, which could result a product of 1880 bp with Pho4 KO strain and a product of 1386 bp with wild type BY4741 strain. The second colony PCR was with internal-internal primers, which could result no product with Pho4 KO strain and a product of 881 bp with wild type BY4741 strain.

Assembly of the inducible multi-gene cassettes

I constructed four inducible cassettes by using the YTK-toolkit (176) cassette assembly strategy (Figure 3.3). The cassette plasmids were assembled with six part plasmids including assembly connectors, promoter, coding sequence, terminator, bacterial marker and origin or yeast marker and origin. The part plasmids could be assembled via a BsmBI assembly with a part entry vector given in the YTK-toolkit and PCR product or synthesized DNA. Each type of part plasmids has a predefined 4 bp flank overhangs after BsaI digestion which could assemble into a cassette plasmid via a BsaI assembly. The cassette plasmids could assemble into a multigene plasmid via a BsmBI assembly. The multigene plasmid contains two transcription units including the Z3EV expression unit and CDS X expression unit and could then be digested by NotI, resulting in a linear integration cassette flanked by two HO homology arms. I assembled a Z3EV cassette, four CDS X cassettes and a backbone cassette with HO locus homology arms via BsaI assembly. I assembled four multigene cassettes via BsmBI assembly, which consist of HO locus homology arms, Rev1 promoter, Z3EV, ScTDH1 terminator, Gal1-Z3EV promoter, CDS X (Pho4, nucleus localized Pho4, yomScarlet-i N-terminus tagged Pho4, and yomScarlet-i N-terminus tagged nucleus localized Pho4), ScTDH1 terminator, KanR marker and ColE1 origin. Then I digested them with NotI restriction enzyme for further integration to the Pho4 KO strain.

CDS X cassette [2]	1	2	3	4	5	6	7	8a	8b
Pho4	ConL1	Z3EV pr	Pho4	ScTDH1 Te	ConRE	AmpR-ColE1			
Pho4*	ConL1	Z3EV pr	Pho4*	ScTDH1 Te	ConRE	AmpR-ColE1			
yomScarlet-i-Pho4	ConL1	Z3EV pr	yomScarlet-i-Pho4	ScTDH1 Te	ConRE	AmpR-ColE1			
yomScarlet-i-Pho4*	ConL1	Z3EV pr	yomScarlet-i-Pho4*	ScTDH1 Te	ConRE	AmpR-ColE1			
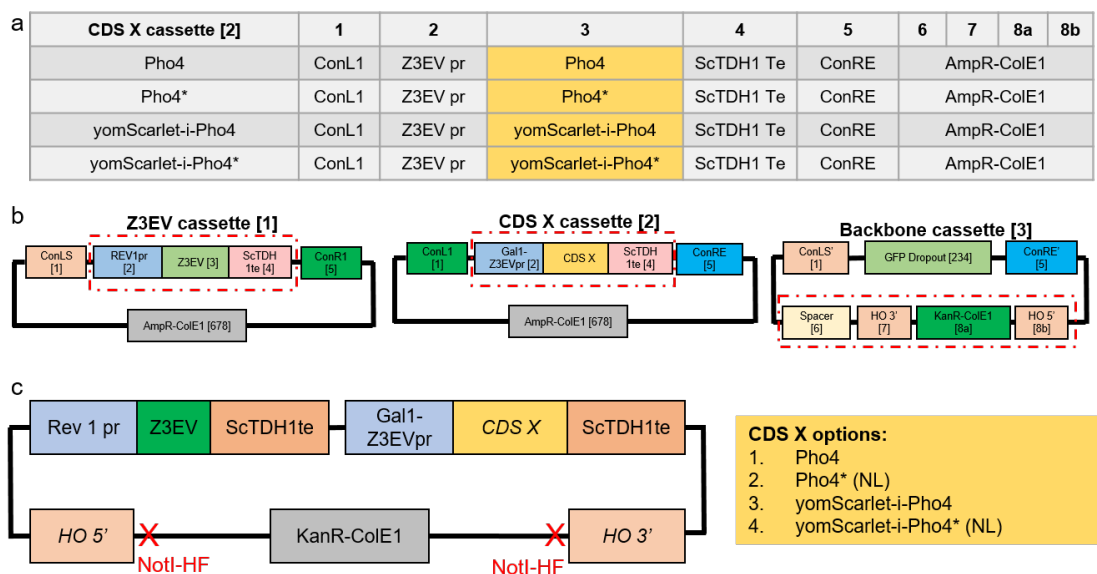
Figure 3.3: **YTK-toolkit cassette assembly strategy (176).** a. Part plasmids assembly. Each type of part plasmids has a predefined 4 bp flank overhangs which could assemble into a cassette plasmid via a BsaI assembly. b. Cassette plasmids assembly. Cassette plasmids could assemble into a multigene plasmid via a BsmBI assembly. c. Multigene plasmids. It contains two transcription units including the Z3EV expression unit and CDS X expression unit and could be digested by NotI restriction enzyme, resulting in a linear integration cassette flanked by two HO homology arms.

CRISPR-Cas9 assisted cloning of the inducible strain library

I integrated four inducible integration cassettes into HO locus of the Pho4 KO strain and the promoter output cassette into Lys2 locus of the inducible strains with the CRISPR- Cas9 assisted cloning strategy (Figure 2.2). Three parts including the integration cassette, Cas9 backbone, and sgRNAs were transformed into yeast Pho4 KO strain by using a high efficiency transformation protocol (203). After transformation, the Cas9 backbone and sgRNA could form a plasmid due to the yeast homologous recombination mechanism and then produce the Cas9-sgRNA protein, which could target the HO locus and make a double-stranded break in the chromosome. The integration cassette will be integrated into the HO locus to repair the double-stranded break due to the yeast homologous recombination mechanism. After the transformation, the yeast cells were plated on a synthetic complete minus uracil (SC-URA) plate for selection. The cells with repaired Cas9-Ura3-sgRNA plasmid can produce the uracil and survive on the SC-URA plate. The colonies on the SC-URA plate will be plated on a SC plus 5-FOA plate for curing the integrated colonies. Ura3 could turn 5-FOA to 5-FU, which is toxic to yeast cells. This negative selection pressure will help yeast cells get rid of the repaired plasmids to avoid the off-target effects. In the end, the colonies on the SC 5-FOA plate were screened by colony PCR and verified by Sanger sequencing. In the end, a total of 92 inducible strains were constructed and sequence verified (Figure 3.4).

Promoters	Inducible Pho4 Variants			
	mScarleti-Pho4	mScarleti-Pho4*	Pho4	Pho4*
VTC3				
PHO89				
SPL2				
PHO8				
VTC4				
CBF1				
PHO5				
VTC2				
VIP1				
PHO84				
GIT1				
PHO81				
CTF19				
YNL217W				
HOR2				
PHO12				
PHO11				
ENA2				
PHM6				
VTC1				
PHO86				
DDP1				
GDE1				
Note: Green represents sequencing verified (92/92).				

Figure 3.4: **Inducible strain libary.** A total number of 92 inducible strains were constructed and sequence verified.

Functional validation of the inducible strains

I carried out a series of preliminary tests using the microscope for initial functional validation of these strains, and get the input-output level of the inducible Pho4 gene regulatory network with different levels of β-estradiol titration. The inducible mScarleti-Pho4 and nuclear localized (NLS) mScarleti-Pho4* strains were induced with different levels of β-estradiol (0, 25, 50, 100 nM) for 6 h in the synthetic complete medium. I observed the mScarlet-i expression under the microscope. Both of them could be induced by β-estradiol and the mScarleti could get expressed. The mScarleti-Pho4 was distributed in the cytoplasm and the mScarleti-Pho4* (NLS) was nucleus-localized as expected (Figure 3.5). From the images captured by microscope, we could see a clear increase of the mScarleti expression level when we increase the β-estradiol concentrations.

Inducible mScarleti-Pho4 (HO)

Inducible mScarleti-Pho4*(NLS) (HO)
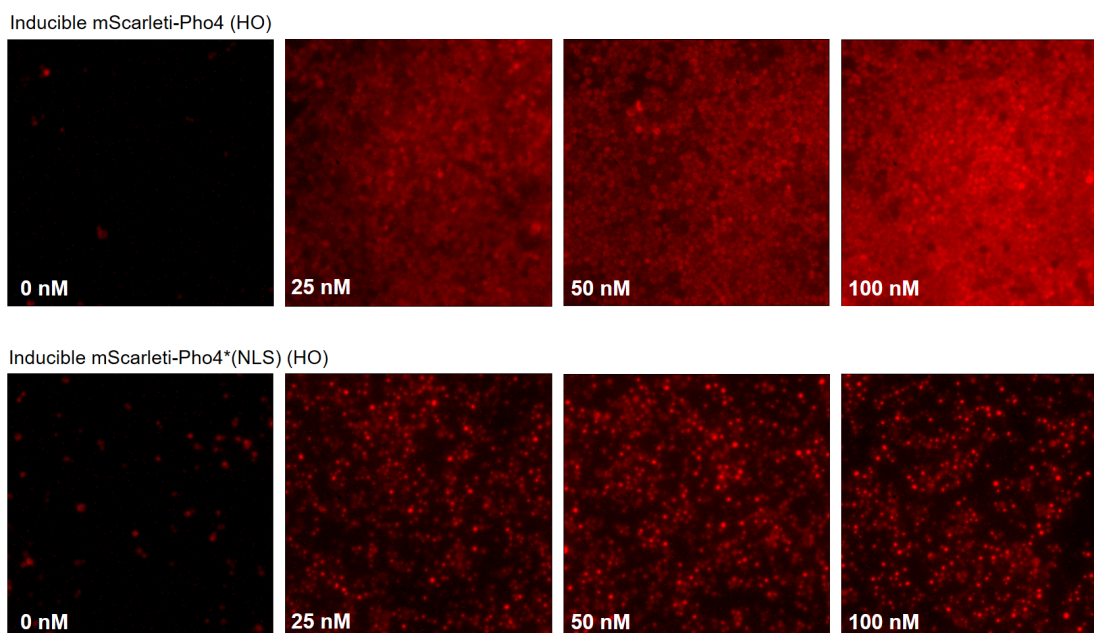
Figure 3.5: **Functional validation of the inducible mScarleti-Pho4 and nuclear localized mScarleti-Pho4.** After 6 h induction with β-estradiol, the mScarleti-Pho4 was distributed in the cytoplasm and the mScarleti-Pho4* (NLS) was nucleus-localized as expected. The Pho4 expression level increased as we increased the β-estradiol concentrations.

Matching the expression level of native Pho4

The inducible strains were further tested in the plate reader with different β-estradiol concentrations, including 0, 1, 2, 3, 4, 5, 10, 25, 50, 100, 500, 1000 nM. The native mScarleti-Pho4 was set as a positive control. With the plate reader data, we determined that to match the native Pho4 level, the β-estradiol concentration would be between 1 to 2 nM (Figure 3.6). The basal expression level of mScarleti-Pho4 and mScarleti-Pho4* (NLS) were lower than the

native mScarleti-Pho4 level, meaning that with the inducible system, we could titrate the Pho4 level below the range of native levels. We further tested the inducible mScarleti-Pho4 on FACS (Becton Dickinson, LSRII SORP), setting the native mScarleti-Pho4 as a positive control and the native BY4741 as a negative control. We incubated these strains for 12 h in the synthetic complete medium with different levels of β-estradiol titrations (0-5 nM). After 12h of pre-culture, we diluted them into OD 0.1 and induced with the same medium condition for another 8 h until reaching steady state. Then we concentrated the culture to OD 0.5 in the same medium and ceased expression on ice for the FACS measurement. It is shown that the inducible mScarleti-Pho4 has no leaky expression without β-estradiol induction. We could tune the inducible Pho4 quite precisely and could match the native expression level with a β-estradiol induction of around 1.9 nM.
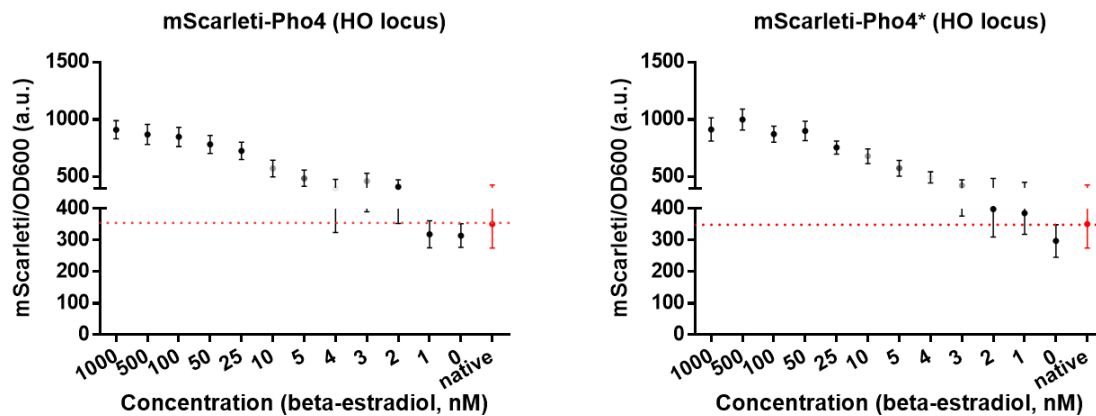
Figure 3.6: **Plate reader test of the inducible mScarleti-Pho4 and mScarleti-Pho4* (NLS) with different β-estradiol titrations.** The inducible mScarleti-Pho4 level could match the native level by fine tuning the β-estradiol concentrations. One could tune the mScarleti-Pho4 level to the levels below or above the native level, with lower or higher β-estradiol concentrations.
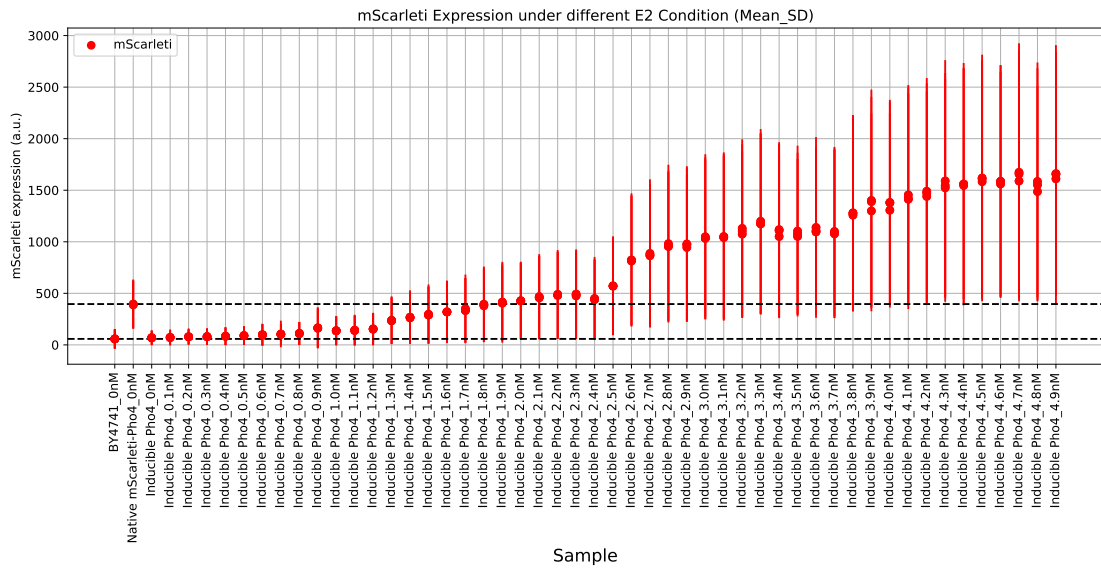
Figure 3.7: **FACS measurement of the inducible mScarleti-Pho4 with different β-estradiol titrations.** The inducible Pho4 could match the native Pho4 level with a β-estradiol induction of around 1.9 nM. The inducible Pho4 could be fine tuned to different levels, either below or above the native level, with lower or higher β-estradiol concentrations.

Functional validation of the dual color strain

The dual color strains were initially tested with a high level of β-estradiol induction (500 nM). The dual color strains were induced with β-estradiol with a concentration of 500 nM for 6 h in the synthetic complete medium. The mScarlet-i and GFP expression were observed under the microscope. Both mScarleti-Pho4 and mScarleti-PHo4*(NLS) could be induced by β-estradiol. The mScarleti-Pho4 was distributed in the cytoplasm and the mScarleti-Pho4* (NLS) was nucleus-localized as expected. The mScarleti-Pho4 showed a certain level of leakage to the nucleus due to the over expression and activated the Pho5pr. The mScarleti-Pho4* (NLS) led to significantly higher activation of Pho5pr compared to the mScarleti-Pho4.

mScarleti-Pho4 + Pho5pr (500 nM induction)
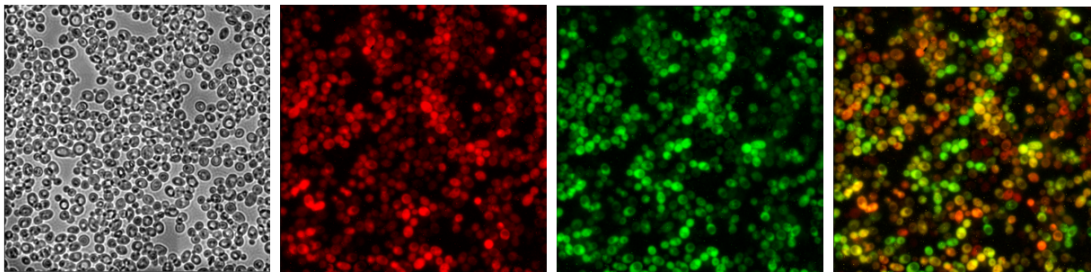
mScarleti-Pho4*(NLS) + Pho5pr (500 nM induction)
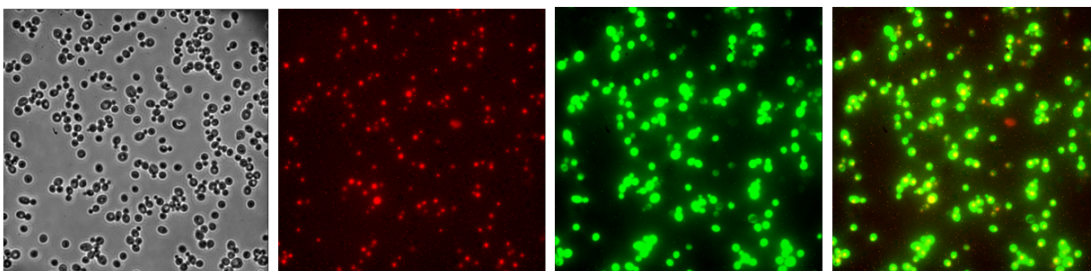
Figure 3.8: **Functional validation of the dual color inducible strains.** The dual color strains were induced with β-estradiol with a concentration of 500 nM for 6 h in the synthetic complete medium. The mScarlet-i and GFP expression were observed under the microscope. Both of them could be induced by β-estradiol. The mScarleti-Pho4 was distributed in the cytoplasm and the mScarleti-Pho4* (NLS) was nucleus-localized as expected. The mScarleti-Pho4 showed a certain level of leakage to the nucleus due to the over expression and activated the Pho5pr. The mScarleti-Pho4* (NLS) led to significantly higher activation of Pho5pr compared to the mScarleti-Pho4.

Input-output responses under β-estradiol titrations and full P_i starvation

We characterized the input-output responses of the inducible Pho4-promoter strains under β-estradiol titrations and full P_i starvation. We included 6 inducible Pho4 strains (inducible Pho4 with LF Pho5pr, Pho89pr, Pho84pr, VTC1pr, Pho8pr, Pho86pr), 6 native Pho4 strains (native Pho4 with LF Pho5pr, Pho89pr, Pho84pr, VTC1pr, Pho8pr, Pho86pr). We also included the native mScarleti-Pho4 strain as a positive control and a BY4741 strain as a negative control. First, we studied the mScarleti expression in response to different β-estradiol tritrations in the P_i rich medium. The inducible Pho4-promoter strains were incubated in the P_i rich medium with different β-estradiol concentrations, including 0, 1, 1.5, 1.8, 2, 2.2, 2.5, 3 nM. They were induced for 12h until reaching the steady state before the FACS measurement. The result demonstrated that the inducible mScarleti-Pho4 was able to achieve native mScarleti-Pho4 levels within a range of 1 nM to 3 nM (Figure 3.9). In the absence of β-estradiol induction, no leaky expression of mScarleti-Pho4 was observed for any of the inducible strains. As the β-estradiol concentration increased, the expression level of mScarleti-Pho4 also increased. The output responses at steady state were also measured under different β-estradiol titrations in P_i rich and free medium (Figure 3.10). In the P_i rich condition, all the output promoters of the inducible strains were not activated, showing a similar level as the native LF promoter strains or BY4741. In the P_i free condition, the output promoters were activated at different levels, which were linked with the mScarleti-Pho4 concentrations. For the Pho5 promoter, as the mScarleti-Pho4 level increased, the promoter activation also increased when the β-estradiol level was below 1.8 nM. However, when the β-estradiol level exceeded 1.8 nM, the promoter activation remained stable despite the continued increase in mScarleti-Pho4 levels, demonstrating a degree of robustness in response to changes in Pho4 concentration. Regarding the Pho89 promoter, activation began at 1.0 nM β-estradiol and remained stable above this concentration, albeit at a lower level than native promoter activation. For the Pho84 promoter, leaky activation was observed even when mScarleti-Pho4 was not expressed at 0 nM β-estradiol, and this activation was higher than the native level. A significant activation occurred when the β-estradiol level exceeded 1.0 nM. For the VTC1, Pho8, and Pho86 promoters, similar activation patterns were observed. Promoter activation increased with rising mScarleti-Pho4 levels when the β-estradiol level was below 1.8 nM. However, the promoter activation remained stable when the β-estradiol level exceeded 1.8 nM, despite the continued increase in mScarleti-Pho4 levels, again demonstrating robustness in response to changes in Pho4 concentration.
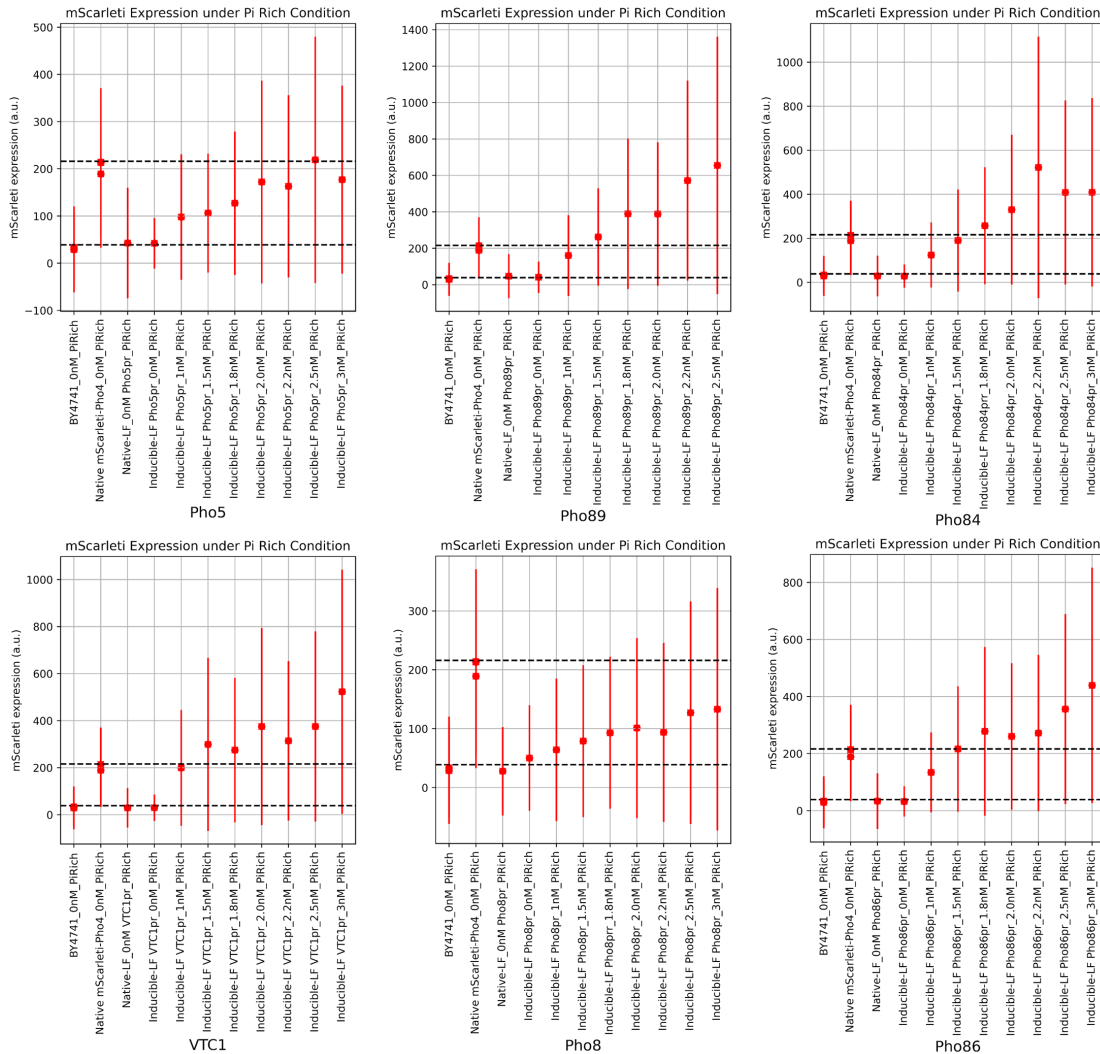
Figure 3.9: **Input responses after 12 h induction under different β-estradiol titrations.** The inducible Pho4-promoter strains were incubated in the P_i rich medium with different β-estradiol concentrations, including 0, 1, 1.5, 1.8, 2, 2.2, 2.5, 3 nM. They were induced for 12h until reaching the steady state before the FACS measurement. The native mScarleti-Pho4 was set as a positive control and the BY4741 was set as a negative control.
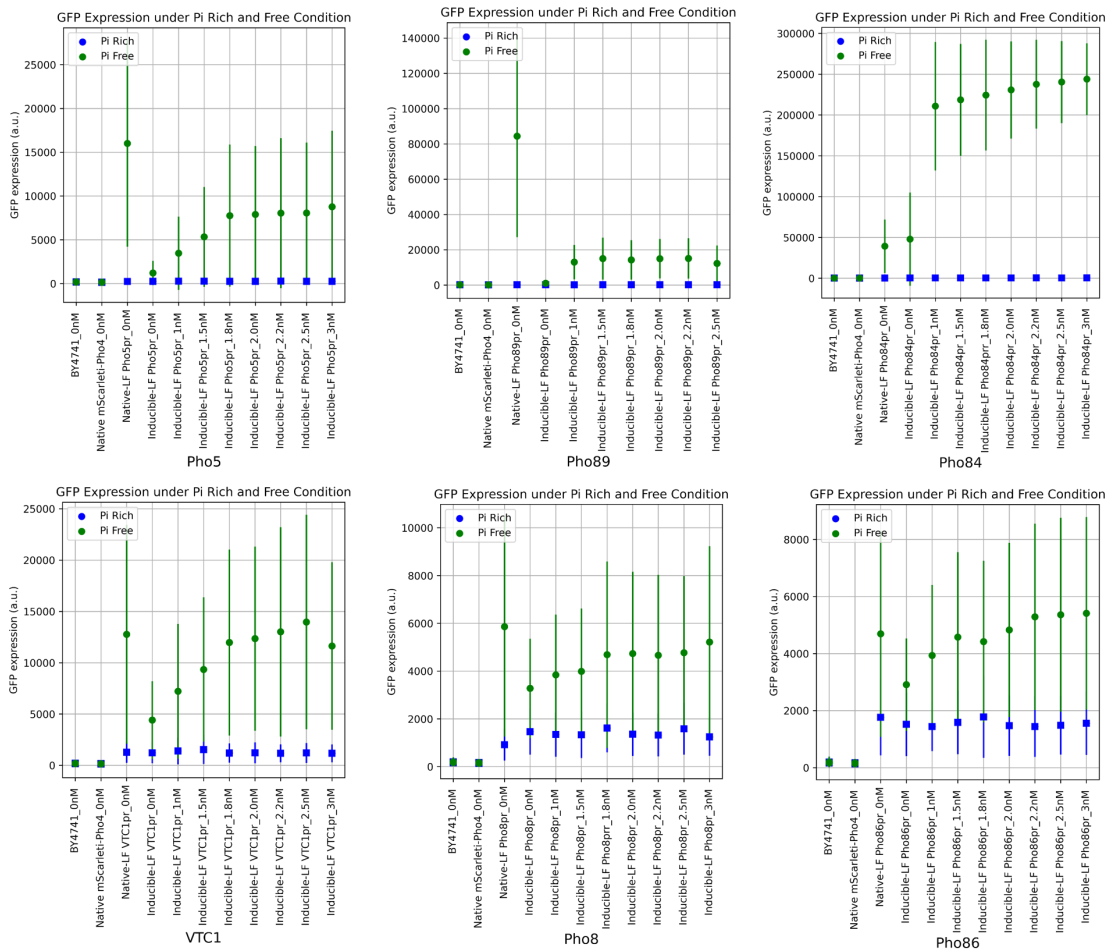
Figure 3.10: **Output responses after 12 h induction under different β-estradiol titrations in P_i rich and free condition.** The inducible Pho4-promoter strains were incubated in the P_i rich and free medium with different β-estradiol concentrations, including 0, 1, 1.5, 1.8, 2, 2.2, 2.5, 3 nM. To induce the promoter activation, the inducible strains were incubated in the P_i rich medium for 12 h and then incubated in the P_i free medium for another 12h to ready steady state before the FACS measurement. The promoters were inactivated in the P_i rich condition and activated in the P_i free condition.

3.2.2 Tuning Pho4-promoter binding affinity and specificity

Construction of the input-out library

To tune the Pho4-promoter binding affinity and specificity, we constructed a library consisting of 276 strains, where the input are Pho4 variants expected with different binding affinity and specificity to the promoters, and the outputs are Pho4-regulated promoter cassettes (Figure 3.11). The input cassette comprises a single unit transcription cassette with HO locus homology arms, the native Pho4 promoter, Pho4 variants N-terminus tagged with yomScarleti, the native Pho4 terminator. We designed 12 Pho4 variants including the Pho4-R14Y, Pho4-R10D, Pho4-R14YE10D, Pho4-E10N, Pho4-H6R, Pho4-H6Y, Pho4-R3K, Pho4-R3Q, Pho4-K2T, Pho4-K2E, Pho4-DBD*, and Pho4*(NLS) (Table 3.1). These Pho4 variants are made by mutating certain residues in the basic binding domain or the translocation domain of Pho4. The output cassettes consist of insulated native promoter, yoEGFP, and a fixed ADH1 terminator, and the Lys2 locus homology arms. We included 23 Pho4 regulated promoters, including CBF1, CTF19, DPP1, ENA2, GDE1, GIT1, HOR2, PHM6, PHO11, PHO12, PHO5, PHO8, PHO81, PHO84, PHO86, PHO89, SPL2, VIP1, VTC1, VTC2, VTC3, VTC4, YNL217W. To construct the input-out library, we started with a commercially available Pho4 knockout (KO) yeast strain where the Pho4 was replaced with a KanMX selection marker and integrated the input and output cassettes in two rounds (Figure 3.11). In the first-round integration, the Pho4 variant cassettes were integrated into HO locus of the parent strain by CRISPR-Cas9 assisted homologous recombination (Figure 2.2). In the second-round integration, the output cassettes were integrated into the Lys2 locus with the same method.

Table 3.1: Pho4 variants with different binding affinity and specificity.

Pho4 variants	Expected BS	Mutation notes
Pho4 R14Y	CAT	R14Y in the basic binding domain
Pho4 E10D	CGC	E10D in the basic binding domain
Pho4 R14Y/E10D	CAT	R14Y, E10D in the basic binding domain
Pho4 E10N	CGC	E10N in the basic binding domain
Pho4 H6R	CAC	H6R in the basic binding domain
Pho4 H6Y	CAC	H6Y in the basic binding domain
Pho4 R3K	CAC	R3K in the basic binding domain
Pho4 R3Q	CAC	R3Q in the basic binding domain
Pho4 K2T	CAC	K2T in the basic binding domain
Pho4 K2E	CAC	K2E in the basic binding domain
Pho4 DBD*	No	K2A, R3A, H6A, E10A, R14A in the basic binding domain
Pho4* NLS	CAC	S->A in the SP1, SP2, SP3, SP4, SP6
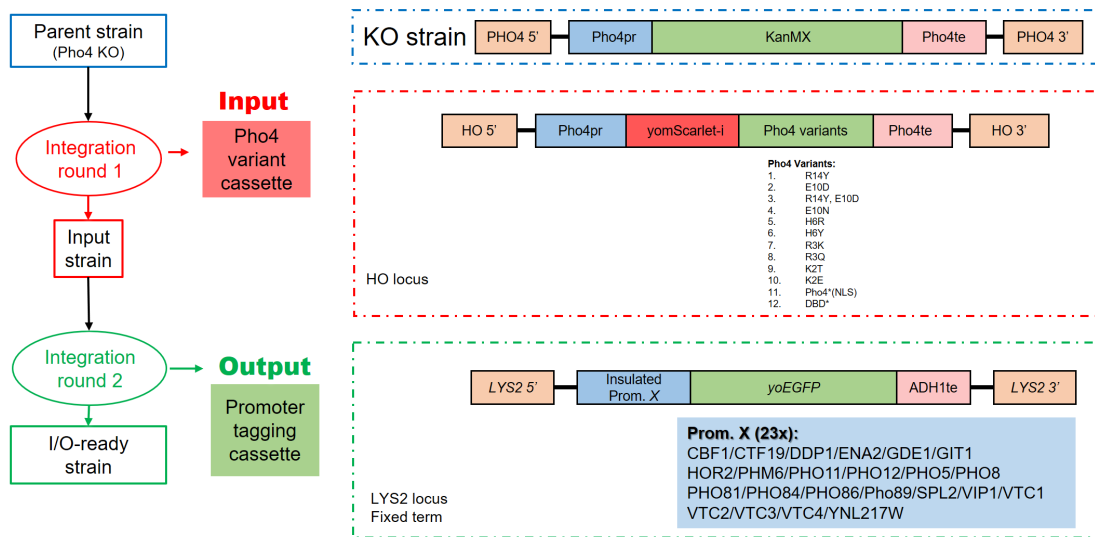
Figure 3.11: **Strain library construction of Pho4 variants and output promoters.** The parent strain is a yeast Pho4 knockout strain, where Pho4 is replaced with a KanMX selection marker. In the first-round integration, the Pho4 variant cassettes were integrated into HO locus. In the second-round integration, the yoEGFP-tagged promoter cassettes were integrated into the Lys2 locus.

3.3   Discussion

In this chapter, we studied the robustness the inorganic phosphate gene regulatory network by tuning the expression level of Pho4 using a Z3EV-mediated inducible system. In the native system, Pho4 is known to be constitutively expressed with a constant concentration. We hypothesize that the native system is built with a certain level of redundancy where the Pho system is robust to change of the Pho4 concentrations, including the Pho4 nuclear localization and Pho regulon genes. With the Z3EV mediated inducible system, we are able to tune the Pho4 expression level, from the basal level similar to the BY4741 strain to the native Pho4 level, and far above the native Pho4 level. This could be achieved by simply adding an inducer, i.e. β-estradiol, which is not consumed by the yeast cells. The native Pho4 level could be matched when induced with around 2 nM β-estradiol concentration. We found that the inducible Pho4 could still respond to the P_i starvation and translocate to the nucleus as the native Pho4. When Pho4 was induced at a relatively low expression level, comparable to or lower than the native Pho4 level, no activation of the regulated promoters was observed. This suggests that Pho4 could be fully phosphorylated and remain in the cytoplasm under these conditions. When switching to the Pi free medium, Pho4 could translocate to the nucleus and activate Pho regulon promoters under full starvation. This suggests that below the native level, Pho4 is robust in term of nucleus localization activity and promoter activation, showing similar responses to Pi changes. When Pho4 was induced at a relatively high expression level, it activated the regulated promoters even without starvation. This means that higher level Pho4 could leak into the nucleus. This phenomenon could be explained by an excess of Pho4 saturating the Pho80/Pho85 CDK complexes, which are responsible for phosphorylating Pho4 in the nucleus (108). As a result, only a subset of Pho4 molecules are phosphorylated and translocated to the cytoplasm and the rest of Pho4 stays in the nucleus. Therefore, it can be concluded that the robustness of Pho4 translocation is limited to a certain range of Pho4 expression levels. We also examined promoter activation under full Pi starvation conditions with a low Pho4 concentration. Different promoters exhibited similar behavior, becoming activated upon Pi starvation and showing increased activation as the Pho4 concentration increased. With a low Pho4 concentration, we could activate the promoter with a strength lower than the native level. Activation reached a plateau when the Pho4 level reached a threshold. This can be explained by the fact that when the Pho4 level is very low, only a portion of binding motifs are occupied, or Pho4 has a shorter residence time on the promoter, leading to lower activation. When the Pho4 level is higher, Pho4 can saturate the binding sites on the promoters or outcompete nucleosomes for binding, resulting in saturated activation and demonstrating robustness to changes in the Pho4 concentration. For example, the Pho5 promoter has a exposed Pho4 site and a nucleosomal Pho4 site (91). At lower Pho4 levels, the Pho4 could only activate the exposed site, leading to a lower activation level. At higher Pho4 levels, the nucleosomal Pho4 site would be activated, leading to a higher or saturated activation level. We also constructed a constitutively nuclear-localized Pho4 by mutating five phosphorylation sites of the Pho4 (108). The nuclear-localized Pho4 could stay in the nucleus and activate Pho regulon promoters even without Pi starvation. The constitutively

nuclear-localized Pho4 could potentially aid in decoupling the effects of Pi starvation and β-estradiol titration by directly adjusting the Pho4 level in the nucleus, thereby activating regulated promoters at different levels without the need for Pi starvation. This hypothesis requires further testing in future studies. Additionally, to study the plasticity of the Pho4 binding domain, a library of 12 Pho4 variants were constructed by mutating the basic domain of the Pho4 based on a previous *in vitro* study(46). They are expected to bind Pho regulated promoters with different binding affinities or specificities. The robustness of Pho networks in response to mutations in the Pho4 binding domain should be investigated beyond the scope of this thesis work.

3.4 Methods

3.4.1 Strain construction

All the plasmids used or generated in this chapter are listed in Table 3.2. All yeast strains utilized in this chapter are listed in Table 3.3. The methods used for plasmid and yeast cloning strategy and protocols in this Chapter are identical to those detailed in Chapter 2.

3.4.2 Plate reader measurement

Yeast cells were grown overnight at 30°C in the synthetic complete medium before adding to the 96 well plate. The overnight culture was diluted to OD600 of 0.1 in the synthetic complete supplemented with different levels of β-estradiol and induced for 12 hours in a standard plate reader (Bio-Tek, Microplate Reader Synergy HT). The OD600 and the fluorescence were measured over time. All the samples were in triplicates. For yoEGFP measurement: excitation and emission wavelengths were 488 nm and 507 nm, respectively. For yomScarlet-i measurement: excitation and emission wavelengths were 569 and 594 nm, respectively. The data was plotted as mean ± s.d. (n=3).

3.4.3 Flow cytometry assay

Overnight cultures were diluted and grown to an OD600 of 0.1. The yeast culture was then diluted to 0.05 into synthetic complete medium or in the P_i-free medium supplemented with different levels of β-estradiol. Cultures were grown for 12-14 hr until reaching steady state. The culture was concentrated to an OD600 of 0.5 and cease expression on ice before FACS measurement (BD LSR Fortessa). All the analysis was conducted with FlowJo and Python.

3.5   Supplemental Information

Supplemental Tables

Table 3.2: Plasmids used or generated in Chapter 3.

Plasmid ID	Strain Name	Source	Identifier
pSC024	[HO] Rev1-Z3EV-Z3EVpr-mScarleti-Pho4	This work	N/A
pSC025	[HO] Rev1-Z3EV-Z3EVpr-mScarleti-Pho4*	This work	N/A
pSC071	[HO] Rev1pr-Z3EV-THD1te-Z3EVpr-PHO4-TDH1te	This work	N/A
pSC072	[HO] Rev1pr-Z3EV-THD1te-Z3EVpr-PHO4*-TDH1te	This work	N/A
pSC415	[HO] Pho4pr-yomScarleti-Pho4 DBD*-Pho4te	This work	N/A
pSC416	[HO] Pho4pr-yomScarleti-Pho4 R14Y-Pho4te	This work	N/A
pSC417	[HO] Pho4pr-yomScarleti-Pho4 E10D-Pho4te	This work	N/A
pSC418	[HO] Pho4pr-yomScarleti-Pho4 R14Y E10D-Pho4te	This work	N/A
pSC419	[HO] Pho4pr-yomScarleti-Pho4 E10N-Pho4te	This work	N/A
pSC420	[HO] Pho4pr-yomScarleti-Pho4 K2T-Pho4te	This work	N/A
pSC421	[HO] Pho4pr-yomScarleti-Pho4 K2E-Pho4te	This work	N/A
pSC422	[HO] Pho4pr-yomScarleti-Pho4 R3K-Pho4te	This work	N/A
pSC423	[HO] Pho4pr-yomScarleti-Pho4 R3Q-Pho4te	This work	N/A
pSC424	[HO] Pho4pr-yomScarleti-Pho4 H6R-Pho4te	This work	N/A
pSC425	[HO] Pho4pr-yomScarleti-Pho4 H6Y-Pho4te	This work	N/A
pSC426	[HO] Pho4pr-yomScarleti-Pho4 NLS*-Pho4te	This work	N/A
pYTK001-096	Yeast MoClo Toolkit	Dueber lab	Addgene# 1000000061
pTY001	LF_PHO5_yoEGFP_tADH1	This work	N/A
pTY002	LF_PHO89_yoEGFP_tADH1	This work	N/A
pTY003	LF_PHM6_yoEGFP_tADH1	This work	N/A
pTY004	LF_PHO84_yoEGFP_tADH1	This work	N/A
pTY005	LF_PHO11_yoEGFP_tADH1	This work	N/A
pTY006	LF_PHO12_yoEGFP_tADH1	This work	N/A
pTY007	LF_VTC1_yoEGFP_tADH1	This work	N/A
pTY008	LF_VTC3_yoEGFP_tADH1	This work	N/A
pTY009	LF_VTC4_yoEGFP_tADH1	This work	N/A
pTY010	LF_VTC2_yoEGFP_tADH1	This work	N/A
pTY011	LF_PHO8_yoEGFP_tADH1	This work	N/A
pTY012	LF_PHO86_yoEGFP_tADH1	This work	N/A

pTY013	LF_VIP1_yoEGFP_tADH1	This work	N/A
pTY014	LF_PHO81_yoEGFP_tADH1	This work	N/A
pTY015	LF_HOR2_yoEGFP_tADH1	This work	N/A
pTY017	LF_CTF19_yoEGFP_tADH1	This work	N/A
pTY018	LF_SPL2_yoEGFP_tADH1	This work	N/A
pTY019	LF_CBF1_yoEGFP_tADH1	This work	N/A
pTY020	LF_DDP1_yoEGFP_tADH1	This work	N/A
pTY022	LF_ENA2_yoEGFP_tADH1	This work	N/A
pTY023	LF_GDE1_yoEGFP_tADH1	This work	N/A
pTY024	LF_GIT1_yoEGFP_tADH1	This work	N/A
pTY027	LF_YNL217W_yoEGFP_tADH1	This work	N/A
pTY055	LYS2 sgRNA1	This work	N/A
pTY056	LYS2 sgRNA2	This work	N/A
pTY057	pWS158 [o]	Ellis lab	Addgene# 90517
pTY058	pWS082 - sgRNA Entry Vector	Ellis lab	Addgene# 90516
pTY103	mScarlet-i Pho4p N-tagged	This work	N/A

Table 3.3: Yeast strains used or generated in Chapter 3.

Strain ID	Strain Name	Source
sSC001	BY4741	Horizon (Cat# YSC1048)
sSC002	Pho4 KO	ATCC
sTY002	mScarleti-Pho4	This work
sSC047	Rev1-Z3EV-Z3EVpr-mScarleti-Pho4	This work
sSC048	Rev1-Z3EV-Z3EVpr-mScarleti-Pho4*	This work
sSC052	Rev1pr-Z3EV-THD1te-Z3EVpr-PHO4-TDH1te	This work
sSC053	Rev1pr-Z3EV-THD1te-Z3EVpr-PHO4*-TDH1te	This work
sSC058	inducible-mScarleti-Pho4:VTC3	This work
sSC059	inducible-mScarleti-Pho4:PHO89	This work
sSC060	inducible-mScarleti-Pho4:PHO5	This work
sSC061	inducible-mScarleti-Pho4:PHO84	This work
sSC062	inducible-mScarleti-Pho4:PHM8	This work
sSC063	inducible-mScarleti-Pho4:CTF19	This work
sSC064	inducible-mScarleti-Pho4*:VTC3	This work
sSC065	inducible-mScarleti-Pho4*:PHO89	This work
sSC066	inducible-mScarleti-Pho4*:PHO5	This work
sSC067	inducible-mScarleti-Pho4*:PHO84	This work
sSC068	inducible-mScarleti-Pho4*:PHM8	This work
sSC069	inducible-mScarleti-Pho4*:CTF19	This work
sSC088	inducible-Pho4:VTC3	This work
sSC089	inducible-Pho4:PHO89	This work
sSC090	inducible-Pho4:PHO5	This work
sSC091	inducible-Pho4:PHO84	This work
sSC092	inducible-Pho4:PHM8	This work
sSC093	inducible-Pho4:CTF19	This work
sSC094	inducible-Pho4*:VTC3	This work
sSC095	inducible-Pho4*:PHO89	This work
sSC096	inducible-Pho4*:PHO5	This work
sSC097	inducible-Pho4*:PHO84	This work
sSC098	inducible-Pho4*:PHM8	This work
sSC099	inducible-Pho4*:CTF19	This work
sSC100	inducible-Pho4:PHO8	This work
sSC101	inducible-Pho4:CBF1	This work
sSC102	inducible-Pho4*:VTC4	This work
sSC103	inducible-Pho4*:VTC2	This work
sSC104	inducible-Pho4*:VIP1	This work

sSC105	inducible-Pho4*:YNL217W	This work
sSC106	inducible-Pho4*:SPL2	This work
sSC107	inducible-Pho4*:PHO8	This work
sSC108	inducible-Pho4*:GDE1	This work
sSC109	inducible-Pho4:VTC4	This work
sSC110	inducible-Pho4:VTC2	This work
sSC111	inducible-Pho4:VIP1	This work
sSC112	inducible-Pho4:PHO11	This work
sSC113	inducible-Pho4:PHO81	This work
sSC114	inducible-Pho4:HOR2	This work
sSC115	inducible-Pho4:SPL2	This work
sSC116	inducible-Pho4:GIT1	This work
sSC117	inducible-Pho4:YNL217W	This work
sSC118	inducible-Pho4:PHO12	This work
sSC119	inducible-Pho4:ENA2	This work
sSC120	inducible-Pho4:PHM6	This work
sSC121	inducible-Pho4:VTC1	This work
sSC122	inducible-Pho4:PHO86	This work
sSC123	inducible-Pho4:DDP1	This work
sSC124	inducible-Pho4:GDE1	This work
sSC125	inducible-Pho4*:CBF1	This work
sSC126	inducible-Pho4*:GIT1	This work
sSC127	inducible-Pho4*:PHO81	This work
sSC128	inducible-Pho4*:HOR2	This work
sSC129	inducible-Pho4*:PHO12	This work
sSC130	inducible-Pho4*:PHO11	This work
sSC131	inducible-Pho4*:ENA2	This work
sSC132	inducible-Pho4*:PHM6	This work
sSC133	inducible-Pho4*:VTC1	This work
sSC134	inducible-Pho4*:DDP1	This work
sSC135	inducible-mScarleti-Pho4:SPL2	This work
sSC136	inducible-mScarleti-Pho4:PHO8	This work
sSC137	inducible-mScarleti-Pho4:VTC4	This work
sSC138	inducible-mScarleti-Pho4:CBF1	This work
sSC139	inducible-mScarleti-Pho4:VTC2	This work
sSC140	inducible-mScarleti-Pho4:VIP1	This work
sSC141	inducible-mScarleti-Pho4:GIT1	This work
sSC142	inducible-mScarleti-Pho4:PHO81	This work
sSC143	inducible-mScarleti-Pho4:YNL217W	This work
sSC144	inducible-mScarleti-Pho4:HOR2	This work

sSC145	inducible-mScarleti-Pho4:PHO12	This work
sSC146	inducible-mScarleti-Pho4:PHO11	This work
sSC147	inducible-mScarleti-Pho4:ENA2	This work
sSC148	inducible-mScarleti-Pho4:PHM6	This work
sSC149	inducible-mScarleti-Pho4:VTC1	This work
sSC150	inducible-mScarleti-Pho4:PHO86	This work
sSC151	inducible-mScarleti-Pho4:DDP1	This work
sSC152	inducible-mScarleti-Pho4:GDE1	This work
sSC153	inducible-mScarleti-Pho4*:SPL2	This work
sSC154	inducible-mScarleti-Pho4*:PHO8	This work
sSC155	inducible-mScarleti-Pho4*:VTC4	This work
sSC156	inducible-mScarleti-Pho4*:CBF1	This work
sSC157	inducible-mScarleti-Pho4*:VTC2	This work
sSC158	inducible-mScarleti-Pho4*:VIP1	This work
sSC159	inducible-mScarleti-Pho4*:GIT1	This work
sSC160	inducible-mScarleti-Pho4*:PHO81	This work
sSC161	inducible-mScarleti-Pho4*:YNL217W	This work
sSC162	inducible-mScarleti-Pho4*:HOR2	This work
sSC163	inducible-mScarleti-Pho4*:PHO12	This work
sSC164	inducible-mScarleti-Pho4*:PHO11	This work
sSC165	inducible-mScarleti-Pho4*:ENA2	This work
sSC166	inducible-mScarleti-Pho4*:PHM6	This work
sSC167	inducible-mScarleti-Pho4*:VTC1	This work
sSC168	inducible-mScarleti-Pho4*:PHO86	This work
sSC169	inducible-mScarleti-Pho4*:DDP1	This work
sSC170	inducible-mScarleti-Pho4*:GDE1	This work

4 Bottom up engineering of an orthogonal gene regulatory network

Contribution: S.C. designed and performed the experiments, analyzed data.

4.1   Introduction

Biological organisms rely on complex gene regulatory networks (GRNs) to respond to changing environmental conditions. The cellular behaviors are usually an integration of multiple gene outputs and their regulation often requires the precise regulation of multiple genes. The complex interplay between the regulated promoters and their corresponding transcription factors allows for the precise regulation of gene expression through creation of feedforward loops, autoregulation, and other motifs (218). Regulated promotors encode information of both the strength and temporal expression of their respective genes (219) and understanding the innate properties of promoters would allow for their precise engineering. Synthetic biology allows us to engineer both the promoters and respective transcription factors of different genes of interest in order to create orthogonal GRNs (220; 190). Many GRNs have been characterized in the past (218; 35) and some pathways have been refactored (221). However, the attempts to reproduce functionally similar circuits has yielded variations in stochastic behavior due to the network architecture (222). To replicate the function of native GRNs, it is necessary to test multiple architectures to reduce the impact of variations in architecture dependent noise, such as promoter orientation (head to head, tail to tail, head to tail, etc.) and genomic location. In addition, replicating larger gene networks remains a challenge because of the sheer number of confounding interactions amongst individual genes and native cellular processes. The function of many genes towards a specific GRN is also often unclear. Thus, it is required to select certain genes and build a clean, modular GRN to get rid of the potential influence of complex native interactions (223) and to better understand how different genes interact and inter-regulate each other. In this work, we bottom up assembled and refactored the yeast phosphate GRN on a yeast artificial chromosome. We designed and characterized a set of synthetic promoters that are regulated by an orthogonal synthetic transcription factor. The synthetic transcription factor (190) consists of separate domains carrying out different molecular functions: i) a synthetic Zn-finger domain enables binding to DNA at specific sequences embedded within the engineered promoters; ii) a protein-protein interaction domain enables cooperative interactions; and iii) the phosphorylation sites that encode regulation of nuclear import and export which comes from Pho4. Their artificial nature makes them ideal for characterization because they are able to get rid of their regulation from the effect of native interactions (223) in spite of variations in genetic or organismal context (224). The successful reverse engineering of the synthetic gene regulatory network would address following questions: i) how to encode the complex functions of native promoters into engineered promoters (abundance levels, timing, rates, decoupling, and how these can be varied in response to different TF concentrations); ii) how to engineer and implement orthogonal transcription factors; iii) how to assemble multiple gene cassettes (including promoter – gene – terminator) on an artificial chromosome to learn what effect chromosome location and relative location has on gene expression.

4.2 Results

4.2.1 Construction of a yeast artificial chromosome scaffold vector

Yeast artificial chromosomes (YACs) (225) are plasmid shuttle vectors capable of replicating and being selected in common bacterial hosts such as E. coli, as well as budding yeast S. cerevisiae. They are relatively small size (~12 kilobases, kb) and of circular form in E. coli but can be linear and of very large size (~several hundreds of kb) when introduced in yeast. Linear YACs generated from circular ones contain all the cis-acting structure elements essential for behaving like natural chromosomes: an autonomously replicating sequences (ARS) for replication; a centromere (CEN) for segregation during cell division; and two telomeres (TELs) for maintenance. Their capability to accept large DNA inserts enables them to reach a minimal size of ~150kb required for chromosome-like stability and for fidelity of transmission in yeast cells. YACs have several advantages (226) over other large capacity vectors: i) accommodation of DNA fragments in size of thousands of kilobases; ii) stable maintenance of cloned eukaryotic DNA due to compatibility with the yeast replication machinery; iii) amenable to large-scale plasmid amplification in E. coli; iv) creation of specific gene changes within exogenous DNA sequences by using the faithful and efficient yeast mechanism of homologous recombination; v) superior to traditional plasmid transformation, alleviating the effects that copy number variation can have on protein expression and allowing for homogeneous protein expression throughout the cell population, even when constructs are integrated in more than a single locus (227). We started from an existing circular YAC named pYAC3 constructed by Burke *et al.* (225) to create the initial scaffold vector and make modifications by homologous recombination with modular DNA fragments pre-designed for integration (Figure 4.1). These DNA fragments have similar structures with transcription units, including promoter, open reading frame (ORF), and terminator, flanked by common homology arms for integration. A Yeast Toolkit (176) has been developed and optimized for the purpose of modular, multipart cloning. All new promoters, homology sequences, ORFs, and terminators were cloned using the specification of this scheme. This approach ensures that all parts are compatible and can be used for other parts within the present kit.
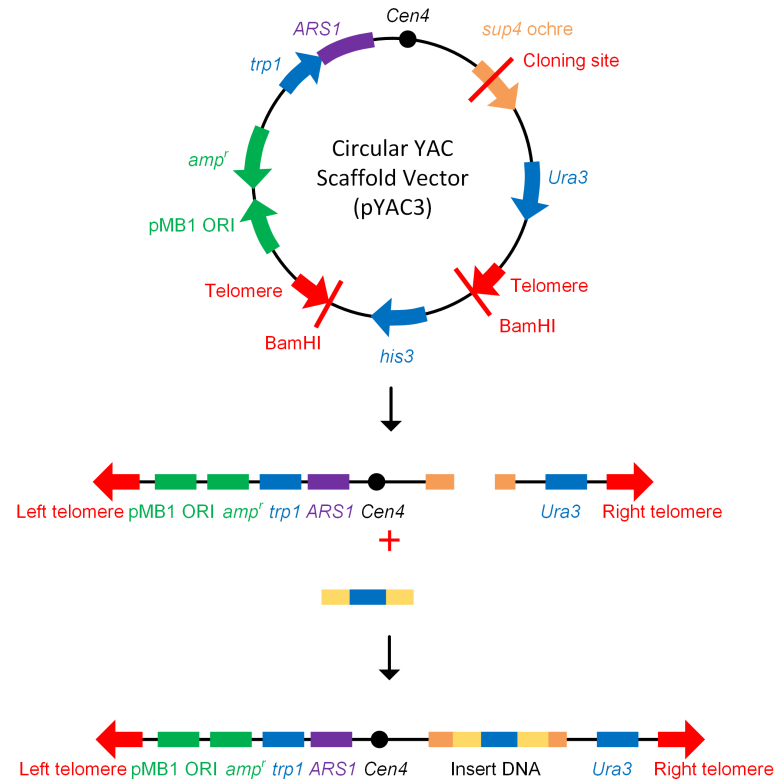
Figure 4.1: **Construction of YAC scaffold vector from a circular YAC (pYAC3).** The circular YAC could be digested with restriction enzymes ligated with DNA fragments with compatible ends to form a linearized YAC scaffold vector, which will be used for exogenous DNA integration via homologous combination. *Figure adapted from Burke et al. Science 236, 806–812 (1987).*

4.2.2 Design of the synthetic transcription factors and promoters

Engineering an synthetic PhoZF orthogonal to the native Pho4

Transcription factors are required to activate the activity of regulated promotors. We first created a synthetic transcription factor (synthetic PhoZF) that is orthogonal to Pho4 to regulate our synthetic regulated promoters (Figure 4.2). It is necessary to retain certain native Pho4 functions, such as nuclear shuttling in response to depletion of inorganic phosphate and interaction with Pho80, in order to match the function of the native Pho4. We hypothesize to replace only the DNA binding domain of the native Pho4 protein with an artificial DNA binding domain. Specifically, the basic Helix-Loop-Helix (bHLH) binding domain of the Pho4 protein consists of the basic region that could bind specific DNA sequence and the HLH region that is the protein-protein interaction domain. The basic binding domain of the Pho4 protein is readily accessible for mutagenesis or replacement with a zinc finger DNA binding domain (190), which has already been extensively characterized *in vitro* in our laboratory (228) and therefore is ready for implementation *in vivo* as a synthetic transcription factor. The synthetic transcription factor consists of separate domains for different molecular functions: i) a synthetic Zn-finger domain enables binding to DNA at specific sequences embedded within the engineered promoters; ii) a protein-protein interaction domain enables cooperative interactions; and iii) the phosphorylation sites that encode regulation of nuclear import and export which come from Pho4. Just like the native Pho4, the phosphate responsive synthetic transcription factor will shuttle between the cytoplasm and nucleus based on the availability of the extracellular inorganic phosphate. We also made a constitutively nuclear localized version, i.e. synthetic PhoZF*(NLS), by mutating five phosphorylation sites (serine to alanine mutations) in the translocation domain (108), as we described in the chapter 3. The expression level of the synthetic PhoZF is tuned by using a Z3EV-mediated inducible system (217) (Figure 4.6). The mScarlet-i was N-terminus fused to the synthetic transcription factor, so the levels of synthetic transcription factor could be measure throughout the characterization experiments.
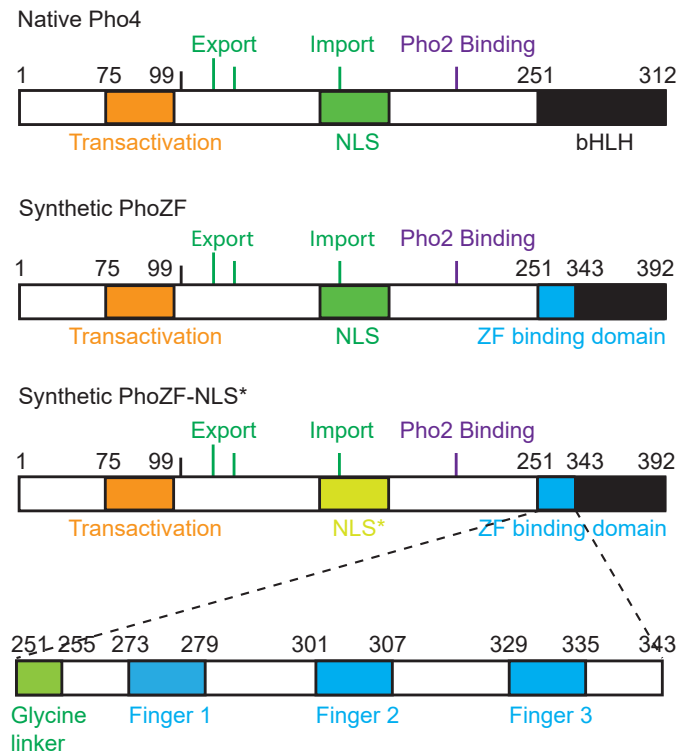
Figure 4.2: **Engineering an synthetic PhoZF orthogonal to the native Pho4.** To design the synthetic transcription factor PhoZF, the basic domain in the bHLH domain of native Pho4 was replaced by a Zinc-Finger binding domain. The native Transaction domain, nuclear localization domain, and Pho2 binding domain were kept as the native Pho4. To obtain a constitutively nuclear localized version, five phosphorylation sites were mutated as previously described (108).

Functional test of the synthetic PhoZF and PhoZF(NLS*)

We characterized the function of the synthetic PhoZF and PhoZF (NLS*) in response to the β-estradiol induction and P_i starvation (Figure 4.3, 4.4). Both the synthetic PhoZF (Figure 4.3a) and PhoZF(NLS*) (Figure 4.3b) strains were induced with β-estradiol with a concentration of 500 nM for 12 h in the synthetic complete medium. The mScarlet-i expression was observed under the microscope. Both of them could be induced by β-estradiol. The mScarleti-PhoZF was distributed in the cytoplasm and the mScarleti-PhoZF* (NLS) was nucleus-localized as expected. In a separate experiment, the synthetic PhoZF strain was induced with β-estradiol with a concentration of 10 nM for 12 h in the synthetic complete medium. The mScarleti-PhoZF was distributed in the cytoplasm. After 12 h of induction, the culture medium was switched to P_i free medium and incubated for 6 h. The mScarleti-PhoZF was translocated to the nucleus (Figure 4.4), showing that the mScarleti-PhoZF could respond to the P_i availability similar to the native Pho4.
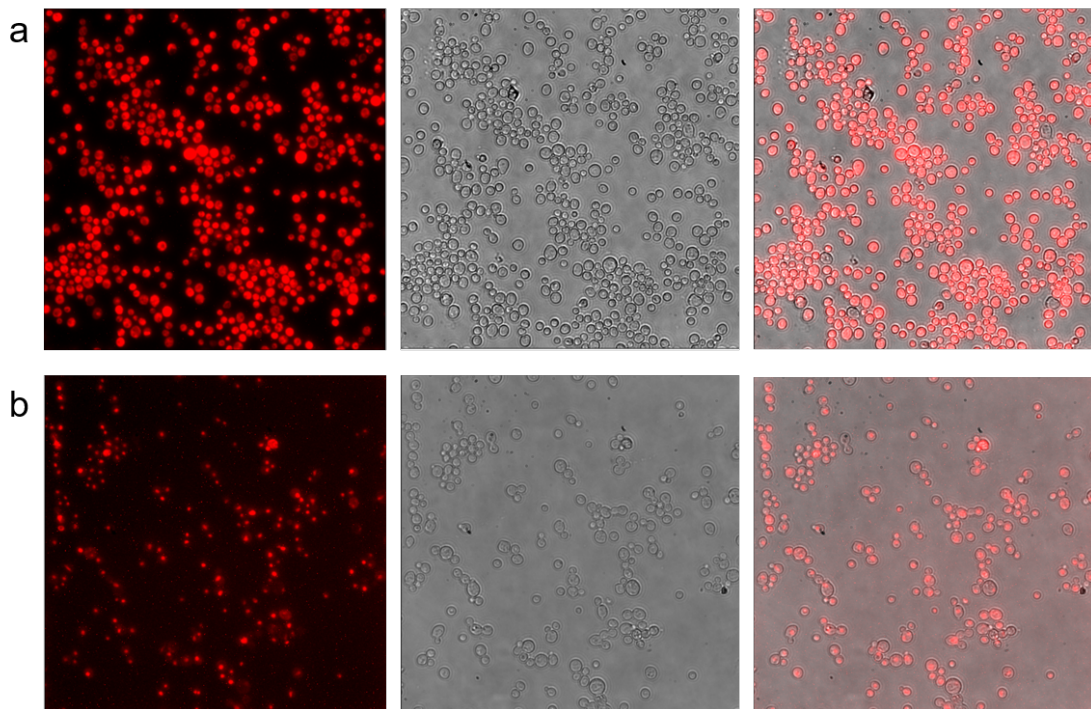
Figure 4.3: **Functional test of synthetic PhoZF and PhoZF* (NLS) under β-estradiol induction.** Both the synthetic PhoZF (a) and PhoZF(NLS*) (b) strains were induced with β-estradiol with a concentration of 500 nM for 12 h in the synthetic complete medium. The mScarlet-i expression was observed under the microscope. Both of them could be induced by β-estradiol. The mScarleti-PhoZF was distributed in the cytoplasm and the mScarleti-PhoZF* (NLS) was nucleus-localized as expected.
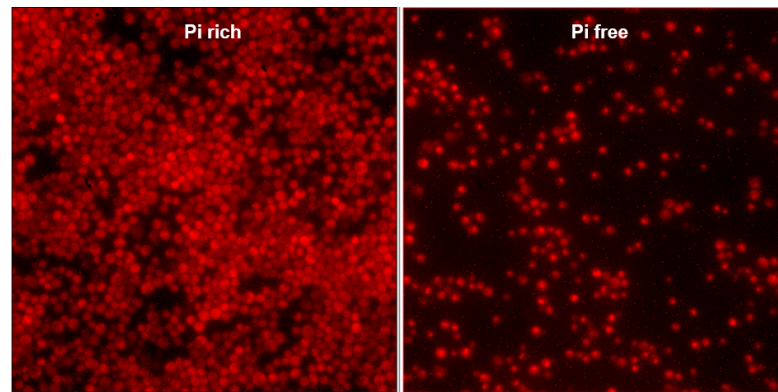
Figure 4.4: **Functional test of synthetic PhoZF under P_i starvation.** The synthetic PhoZF strain was induced with β-estradiol with a concentration of 10 nM for 12 h in the synthetic complete medium. The mScarleti-PhoZF was distributed in the cytoplasm. After 12 h of induction, it was switched to P_i free medium and incubated for 6 h. The mScarleti-PhoZF was translocated to the nucleus, showing that the mScarleti-PhoZF could respond to the P_i availability similar to the native Pho4.
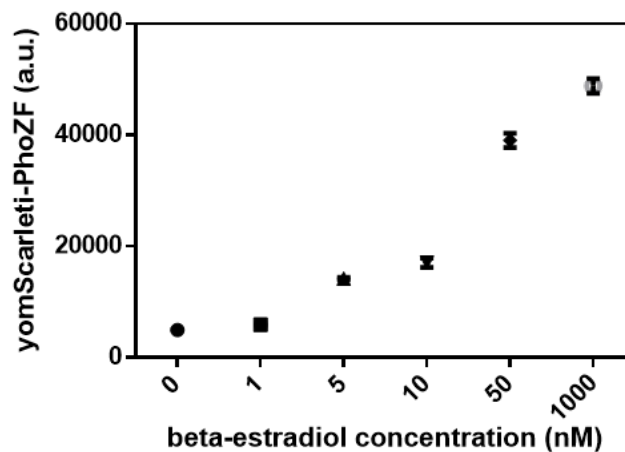
Figure 4.5: **Expression of yomScarleti-PhoZF under different β-estradiol tritration.** To tune the expression level of the synthetic PhoZF, the inducible PhoZF strain was induced with different β-estradiol tritrations, including with a concentration of 1 nM, 5 nM, 10 nM, 50nM, 1000 nM. The PhoZF level increased as the β-estradiol concentration increased.

Tuning the input-output levels

We engineered the synthetic promoters by replacing the binding motifs in the scaffold promoters (CYC1, Pho5, Pho84, Pho81, Pho8, Pho89, Pho12, Pho11, Pho86, VTC1, VTC2, VTC3, VTC4 pr) with the Zinc-Finger binding motifs (Table 4.3). The yoEGFP was used to measure the translational output of the synthetic promoters. The concurrent use of these two fluorescent reporters allow us to study the relationship between the expression level of the synthetic transcription factor and the corresponding output of the synthetic promoters. To tune the output level, we generated synthetic promoters with different ZF binding strengths and different ZF operator numbers. We can precisely engineer the binding sequences to create different ZF-DNA binding strengths (228) and multimerize the binding sequences to create promoters with repeat operators that would correspondingly recruit greater number of synthetic transcription interactions and activation domains (190).
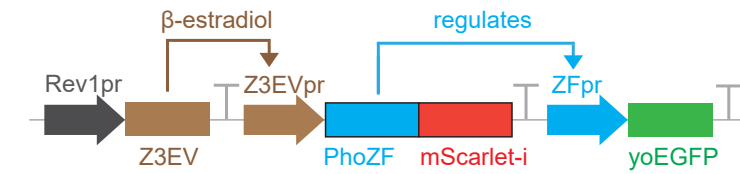
Figure 4.6: **Tuning the input-output level of synthetic PhoZF and ZF promoters.**
The expression level of PhoZF was tuned by the Z3EV mediated inducible system.
By adding β-estradiol, the consititutively expressed Z3EV could translocate to the
necleus and activate the PhoZF expression. The PhoZF could then translocate to the
necleus in response to inorganic Pi starvation, and activate the ZF promoters. The ZF
promoter activation level could be tuned by changing the binding affinities between
the PhoZF and ZF promoters, and binding site numbers.

4.2.3 Construction and characterization of synthetic PhoZF and ZF promoters on YAC

Strain library design and construction

We designed a synthetic mScarleti-PhoZF by substituting the basic binding domain of the native Pho4 with the ZF43-8 domain (190), and also a nuclear localized mScarleti-PhoZF* (NLS) version as previously mentioned. We designed an initial library of ZF promoters with CYC1, Pho5, Pho84, Pho81, Pho8 promoters as the scaffold promoters, respectively. The native binding sites of theses promoters were replaced by the ZF43-8 binding sites. We took the Pho4 KO strain as a parental strain for construction of the synthetic PhoZF and ZF promoters on YAC (Figure 4.7). In the first round, the YAC4 vector was linierized and integrated into the parental strain with a URA3 and a TRP1 selection marker. Before transformation, the YAC4 vector was sequencing verified. In the second round, the inducible mScarleti-PhoZF and mScarleti-PhoZF* (NLS) were integrated into the SUP4 locus of the YAC using Leu2 as the selection marker. In the thrid round, the ZF promoter cassettes were integrated into the URA3 locus using the HygromycinR as the selection marker. We constructed a library of 16 strains for the initial characterization with inducible mScarleti-PhoZF/PhoZF*(NLS) in the SUP4 locus and ZF promoters in the URA3 locus on YAC. In this library, we included minimal CYC1 promoter, CYC1-ZF43-8:1x, CYC1-ZF43-8:2x, CYC1-ZF43-8:4x, Pho5-ZF43-8:2x, Pho84-ZF43-8:4x, Pho81-ZF43-8:1x, Pho8-ZF43-8:1x promoters. In an extented library, we integrated other 8 ZF promoters into the PhoZF strain. These include Pho89-ZF:2x, Pho12-ZF:2x, Pho11-ZF:2x, Pho86-ZF:2x, VTC1-ZF:1x, VTC2-ZF:3x, VTC4-ZF:2x. All the consensus binding sides in the scaffold promoters were replaced by the ZF43-8 binding motifs (190). The synthetic PhoZF was expected to be induced and activate the ZF promoters and generate different output strengths.

Characterization of the input-output function in the plate reader

We characterized the input-output function of the yomScarlet-PhoZF and ZF promoter in the plate reader. The synthetic PhoZF strain was induced with β-estradiol with a concentration of 100 nM for 16 h in the synthetic complete medium. The mScarleti-PhoZF expression level was consistant across different ZF promoter strains (Figure 4.8). With 100 nM β-estradiol induction, the mScarleti-PhoZF expression level was high enough to leak into the nucleus, activating the ZF promoters at varying levels. The activation of CYC1 promoters increased with the increasing of ZF binding sites. The steady state expression was also analysed. The mScarleti-PhoZF expression level at the steady state was consistant across different ZF promoter strains. Compared to the CYC1 minimal promoter without a ZF binding site (CYC1-ZF:0X), all the other ZF promoters exhibited varying fold changes in activation levels. By numerating the binding sites, the CYC1 promoter exhibited clear fold changes. These ZF promoters could be further engineered to fine tune the activation level by changing the binding site numbers or altering the binding motifs.
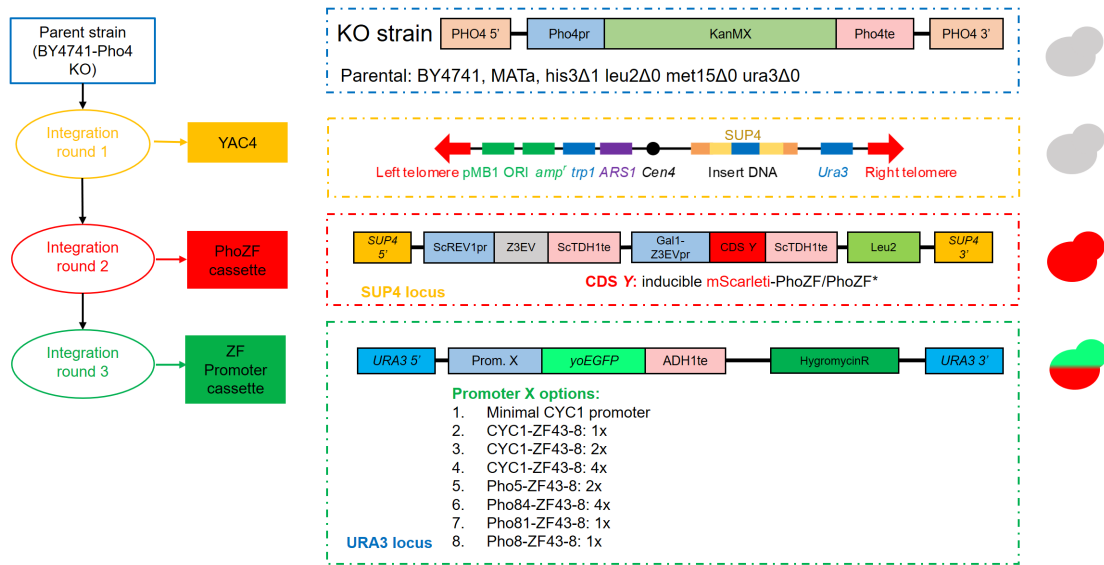
Figure 4.7: **Construction of synthetic PhoZF and ZF promoters on YAC.** The Pho4 KO strain was used as the parental strain for YAC integration. The circular YAC4 vector was linearized before integration. The linearized YAC has two selection markers, including TRP1 and URA3 markers, on the left arm and right arm, respectively. After integration of YAC, the inducible PhoZF cassette was integrated into SUP4 locus on YAC. The PhoZF integration cassette has two SUP4 homology arms and a Leu2 selection marker. We constructed two input strains in this study, including the inducible mScarleti-PhoZF and mScarleti-PhoZF*(NLS). The ZF promoter cassettes were integrated to URA3 locus. The ZF promoter cassettes have two URA3 homology arms and a Hygromycin resistance selection marker. A library of 16 ZF promoters were included in this study.
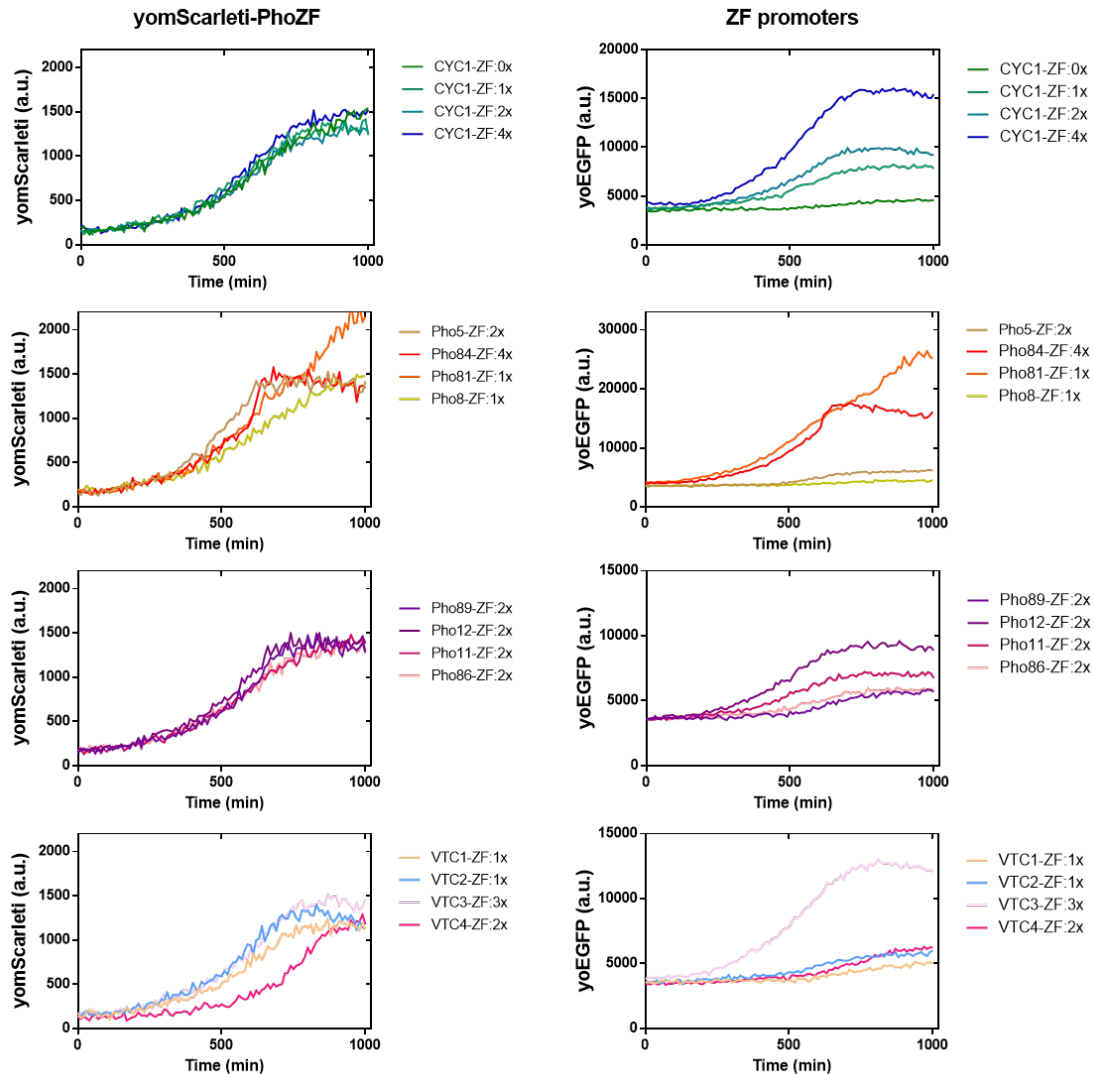
Figure 4.8: **Expression of synthetic PhoZF and ZF promoters over time with 100 nM β-estradiol induction.** The synthetic PhoZF strain was induced with β-estradiol with a concentration of 100 nM for 16 h in the synthetic complete medium. The mScarleti-PhoZF expression level was consistant across different ZF promoter strains. The mScarleti-PhoZF expression level was high enough to leak into the nucleus, activating the ZF promoters at different levels.
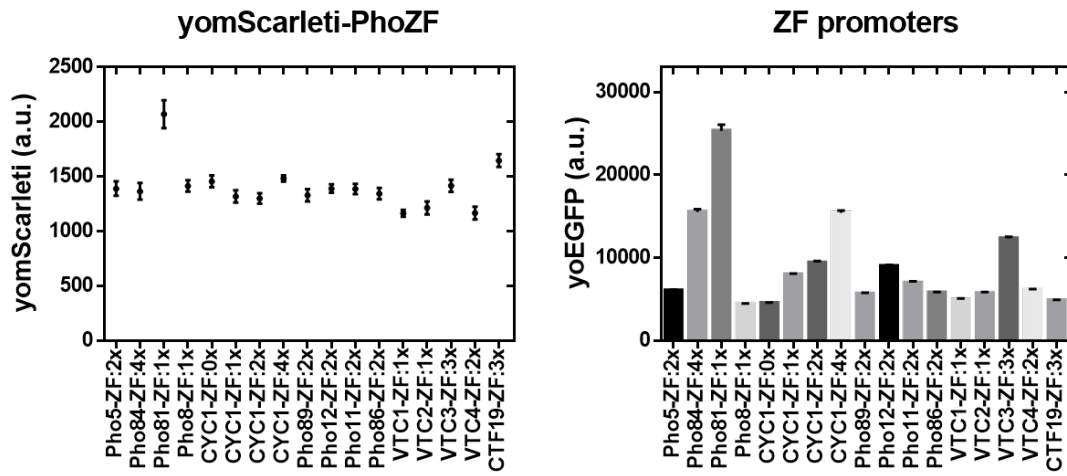
Figure 4.9: **Expression of synthetic PhoZF and ZF promoters at steady state with 100 nM β-estradiol induction.** The mScarleti-PhoZF expression level at the steady state was consistant across different ZF promoter strains. The mScarleti-PhoZF expression level was high enough to leak into the nucleus, activating the ZF promoters at varying levels. Compared to the CYC1 minimal promoter without a ZF binding site (CYC1-ZF:0X), all the other ZF promoters exhibited varying fold changes in activation levels.

4.3   Discussion

In this chapter, we detailed the efforts to construct an orthogonal transcriptional gene regulatory network using a zinc finger synthetic transcription factor and zinc finger synthetic promoters on a yeast artificial chromosome. The synthetic transcription factor PhoZF was designed by replacing the basic domain of the native Pho4 with a Zinc Finger binding domain ZF43-8, characterized by Khalil et al. (190). A library of ZF binding domains and binding motifs were carefully studied *in vivo* and we selected the ZF43-8 zinc finger residue since it showed a high orthogonality which led to a clear fold activation of the ZF43-8 promoter. The transaction domain remained as the native Pho4 and the translocation domain remained also identical to the native Pho4, allowing PhoZF to be responsive to Pi starvation. In addition, PhoZF was mutated at specific amino acid residues in the phosphorylation domain to create a phosphate non-responsive version, PhoZF*(NLS) (229). PhoZF expression was controlled by the Z3EV system described in Chapter 2, with the expression level modulated by adding the inducer β-estradiol. The expression level of PhoZF and PhoZF*(NLS) could be tuned from the basal level to the native Pho4 level, and far above the native levels, by adding different β-estradiol concentrations. To evaluate the orthogonality of PhoZF to native promoters, we integrated all native Pho4-regulated promoters into inducible PhoZF and PhoZF*(NLS) strains and assessed their functionality using a plate reader. Neither PhoZF nor PhoZF*(NLS) activated the native promoters (data not shown), which can be attributed to the replacement of the basic binding domain in PhoZF with the ZF binding domain, as it is not expected to bind native Pho4-regulated promoters. This orthogonality of the PhoZF to native promoters ensures that it will not activate the native Pho system. The synthetic promoters were designed by replacing consensus binding motifs in native Pho4-regulated promoters (Pho5, Pho89, Pho84, Pho81, Pho8, Pho11, Pho12, Pho86, VTC1-4 promoters) and a minimal CYC1 promoter characterized by Bashor et al. (230), with zinc finger binding motifs. We designed and integrated a total of 16 synthetic promoters into inducible PhoZF strains, featuring different numbers of binding sites. Here we included multiple native promoters as the scaffold promoters for further fine tuning in order to match each native promoter output. We employed a yeast artificial chromosome to carry the synthetic PhoZF and ZF promoters, integrating them into the SUP4 and URA3 loci of the YAC. The performance of the synthetic PhoZF and ZF promoters was preliminarily characterized with 100 nM β-estradiol induction in a plate reader. The mScarleti-PhoZF could be induced by β-estradiol to a very high level that is leaky to the nucleus without Pi starvation and activated the ZF promoters at various levels. We observed a clear fold changes for the CYC1 promoters with different numbers of binding sites. For the CYC1 promoters, as the ZF binding site number increased, the activation strength also increased. As for the other promoters, we observed different activation levels. However, we have not compared these ZF promoters with the native promoter performance. In principle, the promoter activation fold change can be further optimized to individually match native promoter levels by modifying the binding affinity between PhoZF and ZF motifs (190). Modification of 1-2 bases in the binding motifs could potentially allow for fine tuning of the promoter output. In order to match the individual native promoters, a library of point mutations can be constructed for

large scale screening and one could select the mutants which recapitulate the native levels. In addition, the influence of orientation and location on activation of dual promoters should also be studied before assembly them onto a single YAC. Insulation of different expression cassettes might be necessary to avoid potential interactions among adjacent gene cassettes. The YAC is a vector capable of carrying over 1 MB of DNA insert. However, it is key to come up with a safer cloning strategy for the large scale multi-cassette DNA insertion. We should avoid using similar scaffold promoter for driving different genes since using similar scaffold promoter multiple times would lead to homologous recombination of similar sequence, thus potentially losing some cassettes. To assemble many gene cassettes onto a single YAC, we should use different scaffold promoters for individual genes. To make promoters more orthogonal, we could design the minimal and fully synthetic promoters based on a bottom up approach (231). In the future work, we plan to carry all synthetic PhoZF and Pho regulon genes controlled by the ZF promoters on a single YAC to match the input-output level of native Pho networks, ultimately constructing a complete synthetic gene regulatory network to replace the native ones.

4.4 Methods

4.4.1 Strain construction

All the plasmids used or generated in this chapter are listed in Table 4.1. All yeast strains utilized in this chapter are listed in Table 4.2. The methods used for plasmid and yeast cloning strategy and protocols in this Chapter are identical to those detailed in Chapter 2.

4.4.2 Plate reader measurement

The methods used for plate reader measurement in this chapter are identical to those detailed in Chapter 3.

4.5   Supplemental Information

Supplemental Tables

Table 4.1: Plasmids used or generated in Chapter 4.

Plasmid ID	Strain Name	Source	Identifier
pYTK001-96	Yeast MoClo Toolkit	Dueber lab	Addgene# 1000000061
pSC010	pSC010	ATCC	N/A
pSC079	YAC4 Homology 3'[7]	This work	N/A
pSC080	YAC4 Homology 5'[8b]	This work	N/A
pSC082	YAC4 Leu2 backbone cassette	This work	N/A
pSC267	YAC4 [URA3 locus] HygromycinR backbone	This work	N/A
pSC343	[YAC4 SUP4] Z3EVpr–mScarleti-sPho4_basic->Linker-ZF43-8–TDH1te	This work	N/A
pSC344	[YAC4 SUP4] Z3EVpr–mScarleti-sPho4*_basic->Linker-ZF43-8–TDH1te	This work	N/A
pSC275	YAC4 [URA3 locus] SPr-CYC1minimal-yoEGFP-tTDHI	This work	N/A
pSC276	YAC4 [URA3 locus] SPr_ZF43-8:2x-linker5bp-yoEGFP-tTDHI	This work	N/A
pSC281	YAC4 [URA3 locus] SPr_ZF43-8:4x-linker5bp-yoEGFP-tTDHI	This work	N/A
pSC282	YAC4 [URA3 locus] Pho89pr-GFP-ADH1te	This work	N/A
pSC283	YAC4 [URA3 locus] Pho8pr-GFP-ADH1te	This work	N/A
pSC284	YAC4 [URA3 locus] CBF1pr-GFP-ADH1te	This work	N/A
pSC285	YAC4 [URA3 locus] Pho5pr-GFP-ADH1te	This work	N/A
pSC286	YAC4 [URA3 locus] Pho84pr-GFP-ADH1te	This work	N/A
pSC287	YAC4 [URA3 locus] Pho81pr-GFP-ADH1te	This work	N/A
pSC288	YAC4 [URA3 locus] Pho12pr-GFP-ADH1te	This work	N/A
pSC289	YAC4 [URA3 locus] Pho11pr-GFP-ADH1te	This work	N/A
pSC290	YAC4 [URA3 locus] Pho81pr-GFP-ADH1te	This work	N/A
pSC291	YAC4 [URA3 locus] VTC3pr-GFP-ADH1te	This work	N/A
pSC292	YAC4 [URA3 locus] SPL2pr-GFP-ADH1te	This work	N/A
pSC293	YAC4 [URA3 locus] VTC4pr-GFP-ADH1te	This work	N/A
pSC294	YAC4 [URA3 locus] VTC2pr-GFP-ADH1te	This work	N/A
pSC295	YAC4 [URA3 locus] VIP1pr-GFP-ADH1te	This work	N/A
pSC296	YAC4 [URA3 locus] GIT1pr-GFP-ADH1te	This work	N/A
pSC297	YAC4 [URA3 locus] CTF19pr-GFP-ADH1te	This work	N/A
pSC298	YAC4 [URA3 locus] YNL217Wpr-GFP-ADH1te	This work	N/A

pSC299	YAC4 [URA3 locus] HOR2pr-GFP-ADH1te	This work	N/A
pSC300	YAC4 [URA3 locus] ENA2pr-GFP-ADH1te	This work	N/A
pSC301	YAC4 [URA3 locus] PHM6pr-GFP-ADH1te	This work	N/A
pSC302	YAC4 [URA3 locus] VTC1pr-GFP-ADH1te	This work	N/A
pSC303	YAC4 [URA3 locus] DDP1pr-GFP-ADH1te	This work	N/A
pSC304	YAC4 [URA3 locus] GDE1pr-GFP-ADH1te	This work	N/A
pSC335	YAC4 [URA3 locus] SynPho5pr_ZF43-8:2x_yoEGFP	This work	N/A
pSC336	YAC4 [URA3 locus] SynPho84pr_ZF43-8:4x_yoEGFP	This work	N/A
pSC337	YAC4 [URA3 locus] SynPho81pr_ZF43-8:1x_yoEGFP	This work	N/A
pSC338	YAC4 [URA3 locus] SynPho8pr_ZF43-8:1x_yoEGFP	This work	N/A

Table 4.2: Yeast strains used or generated in Chapter 4.

Strain ID	Strain Name	Source
sSC001	BY4741	Horizon (Cat# YSC1048)
sSC002	Pho4 KO	ATCC
sSC302	[YAC4 SUP4]–Z3EVpr-mScarleti-sPho4_bHLH->Linker-ZF43-8–TDH1te	This work
sSC303	[YAC4 SUP4]–Z3EVpr-mScarleti-sPho4*_bHLH->Linker-ZF43-8–TDH1te	This work
sSC305	[SynGRN] SPr-CYC1minimal-yoEGFP-tTDHI	This work
sSC306	[SynGRN] SPr_ZF43-8:2x-linker5bp-yoEGFP-tTDHI	This work
sSC308	[SynGRN] SPr_ZF43-8:1x-linker5bp-yoEGFP-tTDHI	This work
sSC311	[SynGRN] SPr_ZF43-8:4x-linker5bp-yoEGFP-tTDHI	This work
sSC405	[SynGRN] SynPho5pr_ZF43-8:2x_yoEGFP	This work
sSC406	[SynGRN] SynPho84pr_ZF43-8:4x_yoEGFP	This work
sSC407	[SynGRN] SynPho81pr_ZF43-8:1x_yoEGFP	This work
sSC408	[SynGRN] SynPho8pr_ZF43-8:1x_yoEGFP	This work
sSC532	[SynGRN]-SynPho89pr_ZF43-8:2x_yoEGFP	This work
sSC533	[SynGRN]-SynPho12pr_ZF43-8:2x_yoEGFP	This work
sSC534	[SynGRN]-SynPho11pr_ZF43-8:2x_yoEGFP	This work
sSC535	[SynGRN]-SynPho86pr_ZF43-8:2x_yoEGFP	This work
sSC536	[SynGRN]-SynVTC1pr_ZF43-8:1x_yoEGFP	This work
sSC537	[SynGRN]-SynCBF1pr_ZF43-8:2x_yoEGFP	This work
sSC538	[SynGRN]-SynVTC2pr_ZF43-8:1x_yoEGFP	This work
sSC539	[SynGRN]-SynVTC3pr_ZF43-8:3x_yoEGFP	This work
sSC540	[SynGRN]-SynVTC4pr_ZF43-8:2x_yoEGFP	This work
sSC541	[SynGRN]-SynCTF19pr_ZF43-8:3x_yoEGFP	This work

Table 4.3: Sequence of the synthetic promoters in Chapter 4.

Synthetic promoters	Sequence	Note
CYC1minimal	cagatccgccaggcgtgtatatatagcgtggatggccaggcaactttagtgctgacacatacaggcatatatatatgtgtgcgacgacacatgatcatatggcatgcatgtgctctgtatgtatataaaactcttgtttcttcttttctctaaatattctttccttatacattaggacctttgcagcataaattactatacttctatagacacacaaacacaaatacacacactaatctagataaaaaggt	Scaffold promoter without ZF BS (230)
CYC1_ZF43-8:1x	aGAGTGAGGAcggattccagatccgccaggcgtgtatatatagcgtggatggccaggcaactttagtgctgacacatacaggcatatatatatgtgtgcgacgacacatgatcatatggcatgcatgtgctctgtatgtatataaaactcttgtttcttctttctctaaatattctttccttatacattaggacctttgcagcataaattactatacttctatagacacacaaacacaaatacacacactaatctagatattaaaggt	CYC1minimal promoter with 1x ZF43-8 BS: GAGT-GAGGA
CYC1_ZF43-8:2x	aGAGTGAGGActcgaGAGTGAGGAcggattccagatccgccaggcgtgtatatatagcgtggatggccaggcaactttagtgctgacacatacaggcatatatatatgtgtgcgacgacacatgatcatatggcatgcatgtgctctgtatgtatataaaactcttgtttcttcttttctctaaatatctttccttatacattaggacctttgcagcataaattactatacttctatagacacacaaacacaaatacacacactaatctagatattaaaggt	CYC1minimal promoter with 2x ZF43-8 BS: GAGT-GAGGA
CYC1_ZF43-8:4x	aGAGTGAGGActcgaGAGTGAGGActcgaGAGTGAGGActcgaGAGTGAGGAcggattccagatccgccaggcgtgtatatatagcgtggatggccaggcaactttagtgctgacacatacaggcatatatatgtgtgcgacgacacatgatcatatggcatgcatgtgctctgtatgtatataaaactcttgtttcttctttctctaaatattctttccttatacattaggacctttgcagcataaattactatacttctatagacacacaaacacaaatacacacactaatctagatattaaaggt	CYC1minimal promoter with 4x ZF43-8 BS: GAGT-GAGGA

| SynPho5pr_ZF43-8:2x | taaatacaatgttccttggttatcccatcgccaataattttattttacca ctgttgaagaagcgaaagaaaaaaaaagggaaaatcaaaacattcc ctgtgcactaatagaagaaaacaagagactccgtccctctttagtgag aaaattgaccagagatggttttttgtccatcttttcgcaaaaaattagttc tattttttacacatcggactgataagttactactgcacattggcattagc taggagggcatccaagtaataattgcgagaaacgtgacccaactttgt tgtaggtccgctccttctaataatcgcttgtatctctacatatgttctattt actgaccgaaagtagctcgctacaataataatgttgacctgatgtcag tccccacgctaatagcggcgtgtcgcacgctctctttacaggacgccg gagaccggcattacaaggatccgaaagttgtattcaacaagaatgcg caaatatgtcaacgtatttggaagtcatcttatgtgcgctgctttaatgt tttctcatgtaagcggacgtcgtctataaacttcaaacgaaggtaaaa ggttcatagcgcttttttctttgtctgcacaaagaaatatatattaaattag GAGTGAGGAttcgcatagaacgcaactgcacaatgccaaaaaa agtaaaagtgattaaaagagttaattgaataggcaatctctaaatgaa tcgatacaaccttggcactcaGAGTGAGGAggactagcacaga ctaaatttatgattctggtccctgttttcgaagagatcgcacatgccaaa ttatcaaattggtcaccttacttggcaaggcatatacccatttgggata agggtaaacatctttgaattgtcgaaatgaaacgtatataagcgctgat gttttgctaagtcgaggttagtatggcttcatctctcatgagaataagaa caacaacaaatagagcaagcaaattcgagattaccaggt | Native BS of the Pho5 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |
| SynPho84pr_ZF43-8:4x | gaaactaatttatcagccgctcgtttatcaaccgttattaccaaattatg aataaaaaaaccatattattatgaaaagacacaaccggaagggag atcacagaccttgaccaagaaaacatgccaagaaatgacagcaatc agtattacgGAGTGAGGAggtgctgttataggcgccctatacgtg cagcatttgctcgtaagggcccttttcaactcatctagcggctatgaaga aaatgttgcccggctgaaaaacacccgttcctctcactgccgcaccgc ccgatgccaatttaatagttcGAGTGAGGAgacgtgttatttcca gGAGTGAGGAgggcggaaattagcgacggcaattgattatggtt cgccgcagtccatcgaaatcagtgagatcggtgcagttatgcaccaa atgtcgtgtgaaaggctttccttatccctcttctcccgttttgcctgcttat tagctagattaaaGAGTGAGGAcgtattactcattaattaaccga cctcatctatgagctaattattattcctttttggcagcatgatgcaacca cattgcacaccggtaatgccaacttagatccacttactattgtggctcg tatacgtatatatataagctcatcctcatctcttgtataaagtaaagttct aagttcacttctaaattttatctttcctcatctcgtagatcaccagggca cacaacaaacaaaactccacgaatacaatccaaggt | Native BS of the Pho84 promoter were substituted with 4x ZF43-8 BS: GAGTGAGGA |

| SynPho81pr_ZF43-8:1x | tttaaaaatgagccatgagcttgttcggagaagagtcgcctaaggtac agtgtgcacagttttgcttttcaacgatgaatgcttgagcgtggcgaag aaaaccttgcgggcttttcgaaaatgtaaggaataaaaatgaagtttc ctgatttataagcaagcaatcactgaacgagaatatcttgtagttaagt gactatcaaacgagcatgagggttacaaagaacttccgtttcaaaaat gaatataatcgtacgtttaccttgtggcagcactagctaacgctacgtg aatgaacgtaccgtgccctattattcttgcttgtgctatctcaagaattg cattttgtaataacaactgcatgggaaaaattatatagatttтctactatt atgtccgcctaagtcagttaaccatctttatcacaaaatatacaattaa ccaactacttaatcaattcggttatattgcttagtatatacgtctttggca cgcgattgaaacgcgctaattgcatcagcctatctttctatgcaagaat gcaagaaaaattgatgtgatgtgccttatcacaattcattacctcctatt tcctctgcagcaacaagtttccttgattataaaggtctttagcgtgaga ggtacaggtgttatggGAGTGAGGAcgaataagggcagaaatta atcaaatttatcaactatttggcgatggctcgagacaggtatagaacc actactaggtgatattgaggcttttgtacaatttatagcaagttttttgaga gtcccttcaagtttgttacataatcttctttgtgcaacgtacaagagcaa agtagaaaaatttggtttttattttttttaagcaacatcagctgcactagtt gagctttтgacaagacatactgctcaaaaaatcttcataacattattttt cggttccacagtgattgagctttttgagagaataacccttтggaggcaa catagatagataaacgtgcaggt | |
| SynPho8pr_ZF43-8:1x | aagaaggaaaaattatattaagcgtgcgggtaaaggcaaggaagaa tcaagtaagacctcaagaatggcactataagtgtggtattataatctgt gtaatcctaatttgagctctacacaataccattcgacggttaacagcta ctgcatcaccgtccagtcatgtcgtacaacggaatagggctcaagtcg gcaaaagggtcatctacgtcgggcGAGTGAGGAcagcgatcac ttgctagcaacaataggcgcagaccacagggtagtcaacagcagcg gcaacaacgacaaaatgcgatcaaaaaggccagccatgacaaggc aagcaggcctcttgctgtgcagaaacagatagagactcatatggaga aacgtgagattgaagtacaagttagcgagctacgggaccgactggag gaggaagaaacgctctcggaagagcagattgacaagaaatgtgaag cgttgagggcaaaactgacgaacgagtggcaagaacagcagcggat gtcctctttgtacacccctcgtaaggcgcgtctaacggaagagcagca tcgacatgaatagcagcattgacgatagcgataagcttcgcgcgtag aggaaaagtaaagggatttttagtatataaagaaagaagtgtatctaaa cgtttatatttttttcgtgctccacattttgccagcaagtggctacataaac atttacatatcagcatacgggacattatttgaacgcgcattagcagcg gt | |

| SynPho89pr _ZF43-8:2x | tatggtttgttgaattcacagtactctcgtcagctttgtcacgtattcaac tgtaacctctttgaatctcaacactacgttgtcttcgcgtgtactaaatgt atcgattatataatagtatttaaaaatcatttaaaattgcaccatgtact gtataatgattttctcaagacgatatttcgagttttgccgatgggtcttat aatctcttttgatagagatacctacttttaacaatatcaaaattactgat ataggttaattttttgttcacgagtaagtacttcagtgacactcgtcagta aagaacattagcggaatcgttataatttcattccttacatccgtttaattt cattaattcgatgcatctttctttggcgaagccagaaaatctagtttcag aaatctcggtttcacccgcaaaaaagtttaaatttcacagatcgcgcc acaccgatcacaaaacggcttcaccacaagggtgtgtggctgtgcga tagaccttttttttctttttctgctttttcgtcatcccGAGTGAGGAgt gccattaatttgttagtgggcccttaaatgtcgaaatattgctaaaaatt ggcccgagtcattgaaaggctttaagaatataccgtacaaaggagttt atgtaatcttaataaattgcatatgacaatgcagGAGTGAGGAgg agacaaatagtaataatactaatctatcaatactagatgtcacagcca ctttggatccttctattatgtaaatcattagattaactcagtcaatagca gatttttttttacaatgtctactgggtggacatctccaaacaattcatgtc actaagcccggttttcgatatgaagaaaattatatataaacctgctgaa gatgatctttacattgaggttattttacatgaattgtcatagaatgagtga catagatcaaaggtgagaatactggagcgtatctaatcgaatcaatat aaacaaagattaagcaaaaggt | Native BS of the Pho89 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |

| SynPho12pr _ZF43-8:2x | tgctgtgaccgtttccaatacggaaagaaacggtactgggagcagga gttgcttttacagatatgaacaatgccaatagagccgcacatgtaatta ctggttcacactcgtggggcccacacgattcctgtgcaaagtttgaca agaggatggagtttcacgtaaatgctgccaaaggtgatgcggttttgtt tttgggcagcctctaccatgttgcaagtgcgaaccatactgtggccac atagattacaaaaaaagtccaggatatcttgcaaacctagcttgttttg taaacgacattgaaaaaagcgtattaaggtgagaaaatcaagattat ctaagccgatgaaaaatgaaaggtatgatttcagccacaaatatatag tagttattttatacatcaagatgagaaaataaagggatttttttcgttctttt atcattttctctttctcacttccgactacttcttatatctactttcatcgtttc attcatcgtgggtgtctaataaagtttttaatgacagagataaccttgata agctttttcttatacgctgtgtcacgtatttattaaattacGAGTGAG GAttcgcataacattctgtagttcatgtgtactaaaaaaaaaggaaat aggaaggaaagagtaaaaagttaatagaaaacagaacacatcccta aacgaagccgcacaatcttggcgttcaGAGTGAGGAggtttaaa aaggcaaattacacagaatttcagaccctgtttaccggagagattcca tattccgcacgtcacattgccaaattggtcatctcaccagatatgttata cccgttttggaatgagcataaacagcgtcgaattgccaagtaaaacgt atataagctcttacatttcgatagattcaagctcagtttcgccttggttgt aaagtaggaagaagaagaagaagaagaggaacaacaacagcaaa gagagcaagaacatcatcagaaataccaggt | Native BS of the Pho12 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |

| SynPho11pr_ZF43-8:2x | ggtgctgtgaccgtttccaatacggaaagaaacggtactgggagcag gagttgcttttacagatatgaacaatgccaatagagccgcacatgtaa ttactggttcacactcgtgggggcccacacgattcctgtgcaaagtttga caagaggatggagtttcacgtaaatgctgccaaaggtgatgcggtttt gtttttgggcagcctctaccatgttgcaagtgcgaaccatactgtggcc acatagattacaaaaaaagtccaggatatcttgcaaacctagcttgttt tgtaaacgacattgaaaaaagcgtattaaggtgaaacaatcaagatt atctatgccgatgaaaaatgaaaggtatgatttctgccacaaatatata gtagttattttatacatcaagatgagaaaataaagggattttttcgttctt ttatcattttctctttctcacttccgactacttcttatatctactttcatcgtt tcattcatcgtgggtgtctaataaagttttaatgacagagataaccttga taagctttttcttatacgctgtgtcacgtatttattaaattac**GAGTGA GGA**ttcgcataacattctgtagttcatgtgtactaaaaaaaaagaaat aggaaggaaagagtaaaaagttaatagaaaacagaacacatcccta aacgaagccgcacaatcttggcgttca**GAGTGAGGA**ggtttaaa aaggcaaattacacagaatttcagaccctgtttaccggagagattcca tattccgcacgtcacattgccaaattggtcatctcaccagatatgttata cccgtttggaatgagcataaacagcgtcgaattgccaagtaaaacgt atataagctcttacatttcgatagattcaagctcagtttcgccttggttgt aaagtaggaagaagaagaagaagaagaggaacaacaacagcaaa gagagcaagaacatcatcagaaataccaggt | Native BS of the Pho11 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |
| SynPho86pr_ZF43-8:2x | aaacagtgatggcttcatgctttcagtttcattattggcacgagtagca gtcgttgagcccatcagagtgcaactttggttaaatgtcgtaaattgtat gattgaatcatccatgcaccaatgtcctccacgggatcgccactttttt cttcatcacgcccattttgttgatccgacgttctgtatcaactgtatac agattcgttgcttctcgcaccacacaggttttacgtgcggccaagacc gttgtcaagtggttcattattgtcgacccactgataaattctattctaatc aattatctcattgataggctctgcactcttgggcatgctgtattgcgagt gaaaaagcgcaagaccgaggagagacaaccctgttctcccattatac aacacaca**GAGTGAGGA**aaagacgcaaacggcctcgtctgag gatcgtagccatcaagaggaagcggaggcgtcgaagaccacatcgc atcgagcgtccactctcaaatatgtacccaataatggaaatccagatg gtggctgttccacttgctcttcccagtccaactgcactggtgcattatca gcagcagcaacagcagctcccgcaacatcatccctggtacgacctct cgctttctgaagaagccctctcgacctgctgttgtagttgacctataact gaaaattcacacccgcattacccagcgcgcccg**GAGTGAGGA**c tctttatggaaactataatgaataacgaccaaagaacatctttttttctc tctcacacaaataaagctttttcatgtttatagatatatagcacatactttt actaagttttctagattgattacgactcttccagttgcacatttttttcatag cgaaagtacaagagtaggaagagtggcacaggt | Native BS of the Pho86 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |

| SynVTC1pr_ ZF43-8:1x | aataatttaacatgaacattttaagctgtccttgtaagaaggccatgat acagattccccccctgccccgaaatggggtctaggcaacattgttctat aaaggaacgcaactttatgtggcaacattaaagcaaatacatatctac ttgtgaagccaaagaattgtaattaaatagcactcattattattttttgtct tctagtttatagttgaatataccatctttttttagattcccaggatgcgcct atatcgttcttcaatgagaaattagcgttttcttcgttgagtccttcaatg ttcaacgtcgttaccgaggcaatatatttatacaaagaagagtgctttt caagttttctaggaccagtttttataagaaggttggaagttgattcggtc gataaattgatttcattgtcatcttttaaactatcataactttcttgaatta acgtaactagttttctttcgggcagaggagagttatcactccttccatcc tttatttcaatctgcatactcattttttttgttgactttcgtaatcacttacttt cacttatcctaaataatttacagtgcgttttgtaaaactgcaagacggt atccgctaatagagagaagaatacgtattttctagtaaagaaatttaca aaaaaggaggtatattggcatcgctattttcggatagaatggcaagag cgattaagtatagcaaactgggaaatgtatagagacaagaaataatg agactctctcttcaagtatctttgagcacgtatcagataactaatggtta tatcgcttgttacggtcggtgatacttcactttgaaatagatattagGA GTGAGGAtctcggagagaagaaaatgaatgctgcattatacaaa atagataacagtatttaagagtccatatttccataaccagcgacttgca tacacgtaacatacgatttactacattatcgaatacgattaaacactac gccagatttccacaatggt | Native BS of the VTC1 promoter was substituted with 1x ZF43-8 BS: GAGT-GAGGA |

| SynCBF1pr_ ZF43-8:2x | gttggggatgaagggggggaatgaaaagcaccaaacctaagcagcaag ataaaaaggaaaatttgaagaaagatgaggtcaagaacggcgacaa ggacaaagtcattgaagaggctaccacgactaatgtggacagcatat tggaaaaaccaacaccaacctccacaaaagtagaggacaacttaag cgaagatgattccacaatgaaggaattaaagtcgatgctaaactcca cccctactacaccaacacataatggcccccactccgctccctgcaaag gctggcactcaattagataagcgtatgagcgacctatcactcaaatcg gaaacaccggcttctacaaagaactttagcgcaccgaatgtttcgtca tcaagcaattcactgagaagcctcggaagtccaagcgtgtcatcctcg aaaaagaagaaggtaattcaccaccacttggacattaccaacagcgt tacaaacatgtcatccgtgagcgcctttatctcaagatagacgttagga cttctttaattccctcttttatgctttagtatcgtcatattctttcataccctt ccttataagcttttatataaaatatcccaccgggaaaagataaagaa cggaagataaatcaaacaatgaaataaagcacacacatataaatcc acatatgtacttatatattttacttcaaacaataaacaaacattttattat ttattttttccccgtacatcgatgtacgtgatcggttttttctcattcctctcgg atccttcgggaccaccatgtctcacccaccaaggtcaaaaaaaaata aaatcggacgagaaaagtattgggcaaacaacaattatcGAGTG AGGActtatGAGTGAGGAtttttatttaatgtagaatgtaaacca ttataccttaaagataaacactttgttaattaatcattttagttggcggtc attttcatttctgctacaagtaccaacatcaagtgcttaaaatataatac ggttttctacactttttattaacgggt | Native BS of the CBF1 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |

| SynVTC2pr_ZF43-8:1x | cgagtccagtgctaagaacgatgacaatgtcaacgagatttttttcacc ttggcgaagttaatccaagaaaaaatcgacagtaacaagcttgttggc gtcggtaacggtaaagagggcaatattagcatcaatagtgggagcgg aaacagttctaaatcaaattgctgttgaagaaaagaagatttttgcttct ttgagaattaatcgtgaactggaattagagctataaatcgcttgattctt ttttcccttttctcaatattcgcaggttattttttcctttatatatatacattg tttgtatctttttttgattaccatacttactattttttatattagtcatattaga tgtcacttttatattatctaaacaattcaaacttctcacactaacttttttca gcatacttctactgttgttctctgctctggaaggtattttcatcaggcgat tggtaaaaaatcaagaaagaatctaaaaaatagtgcaacaggaaaa cgaggttcccagtttccctgggaattcttatccaagtatatattttactttt aagattcagcactttttatttcgttctaatgttttttaccggtagtaaagtttt gcttcttgaatccactattcaactacttgcgcttatgtaaggttcctgtat ggttgcataatcactcaattcgaaaaaaaaaaccGAGTGAGGAct gcttggaaatgaatccgcactgcggtatagcacttccagaaagatact gactgcgaaattcgctttatggagtgtttaagcaattggtatcttaccca ttaagagtcatttctttcttcttgccaacttttttccatcctattatgacctc aaaaaaggcaggctctcttcgtcttatcttttctatcggaagcgaattac tgggaaagaggcataggtataaacagataaaatagacgaaaagagt ttagaaaaaaaagtagaaagaacgactacacctcaacataacgaca cttttttgacggt | Native BS of the VTC2 promoter was substituted with 1x ZF43-8 BS: GAGT-GAGGA |

| SynVTC3pr_ZF43-8:3x | ttgcggtcacttatcgccgaattaggattagaggatgagatcgagatta gattaagaaaatcctgtgtgatatcgttttggaggatatcaccattttta gtcttactgttattttccttcatcaagatattttggtacgtctggatctcttg taaaagagtatttctccttgtcaacagcgcatccttctcctcctgtagct tcagtgcgcgtacctgatcatcatctaatagcgatagttgttgcttctcg tccccatcaatgtttacatctgcttcaccagcatgatgtgtacccacag cgtcttccagacttaaatgcgaattcgatgtgtcgtgcgaattcgtcgta tcagacgtaaaatccatatttgcgttctttgccgacctagtatcaacaa gatcacactagctccaaggtatatagctgtttatataagtctaaattac gcgtttctgattttcggcggactccaactattagatagtcaccattacc ggttacattaccggccacatagtagaacgccagagtatatatgccgtg accttcacggtttggatgttaatgagtaacttgcacagctattggtattt ggacttcttggcacattttccctgtacagctttctttttttcaatgtgcg agggccGAGTGAGGAgcttaatacaacacagtgcagtcgcaag GAGTGAGGAcatgcgattttgaaagcacaaggcagaGAGTG AGGAgaaacatacacacacgcacacacccaactttcgagtgggttt catccaggaagaactttaaatagaggccagagagcggcttacatca gacatcttttctcgaacttacttcatcttctcaatatcgagagggcccat tggcgggtaataatacgcataagagtgatctaatccaagtagtacctc cttagcgctttttttattatcactataggtcatttggctattagagcgaac agcagaatttgtccttggttttcagagtttgaaaggt | Native BS of the VTC3 promoter were substituted with 3x ZF43-8 BS: GAGTGAGGA |

| SynVTC4pr_ZF43-8:2x | caatcaacttcctgccaatatggtccacctttactccatagtggaggagtgtgatgcaagatttgacgaaaaaacaattgaggaaatgctagagatcatctctggctacgcatgagtgtacagtgctttgaattgatttggactggcacttccgaaaaacagaaattactttgttgtgtaacaatataacatgttcaatgagtatgcatgtattattgaacttaagcatatagaaatatttacaagaatatactttattttttagagtttattgcgcacactttgggttagtctcgattttttttatttccattggaggaaggttattatttggaaatttagtgctgcgttataaacgtctgtaaagccagtaccacaagcaacttcggtgagatttaactttgtttattctgtatttttgtcgaggcttaaaagggtatactagttgagggcgttacagtaatcccttcatggcaagaataaagtaatagttatataataattgtcatacgcattgttataccagttgacgagtaaaaaaagacgcaccgtgacttcaagagctgtcgcaaacaagttcgacaatatttaacttataaacatattaaagctttcaaaaatatagtctttcgaagagctttttcatactaagagaagcgttgctaatattccgaattaatcacgcaattgcctctctgctttccacgtaacccatcataggaaattgtattaatatgttacttactggaattgcacaggaaaacaccgacaaaaaattggtttgtctgcgttttgacggagagctactgacttgtaggttcacttttttcacacaccgcactatgcatatatgttgcattttattattcatccgGAGTGAGGAgctgcaGAGTGAGGAaatgaaaaagaagcaatactatctttatatattgaggcaatcagaaagaagttacaggctaacaatcaaatcggccaataaaagagcataacaaggcaggaacagctggt | Native BS of the VTC4 promoter were substituted with 2x ZF43-8 BS: GAGTGAGGA |

| SynCTF19pr_ZF43-8:3x | ccagtaaaatttaacctgtcaaagttatccaatctttgagcctcatctag acattcttctaaggtgtttttgaactgctcagaatttattttttggatttttt cagctgatttagtgttttcctctaaatcatctaattttcgcaagacagcat tatatttggaaatttgaaatgtgtaaacctttctagttctttatccaagg cctccacaaaatcagattcattcctttcggaccaactatccacggaac tacggccatcatgtatgacgctttcctttagtaatttcttcaacctttcat agtcgatatatgaatccttccagggaggatatacgtcgttagccagttt aataccaaatagcattttcaaactctgaaaaccaaggacaaattctgc tgttcgctctaatagccaaatgacctatagtgataataaaaaaagcgc taaggaggtactacttggattagatcactcttatgcgtattattacccgc caatgggccctctcgatattgagaagatgaagtaagttcgagaaaag atgtctgatgtaagccgctctctggcctctatttaaaagttcttcctggat gaaacccactcgaaagttgggtgtgtgcgtgtgtgtatgtttcGAGT GAGGAtctgccttgtgctttcaaaatcgcatgGAGTGAGGAct tgcgactgcactgtgttgtattaagcGAGTGAGGAggccctcgca cattgaaaaaaagaaaagctgtacagggaaaaatgtgccaagaagt ccaaataccaatagctgtgcaagttactcattaacatccaaaccgtga aggtcacggcatatatactctggcgttctactatgtggccggtaatgta accggtaatggtgactatctaatagttggagtccgccgaaaaatcaga aacgcgtaatttagacttatataaacagctatataccttggagctagtg tgatcttgttgatactaggtcggcaaagaacgcaaatggt | Native BS of the CTF19 promoter were substituted with 3x ZF43-8 BS: GAGTGAGGA |

5 Systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and *in vivo*

This chapter has been published in *Nature Communications* 2022.

Authors: Amir Shahein, Maria López-Malo, Ivan Istomin, Evan J. Olson, Shiyu Cheng, Sebastian J. Maerkl

Contribution: S. C. investigated and verified the Z3EV inducible system and minCYC1 promoter utilized in this study. S. C. developed the cloning protocols and performed the cloning of the plasmids and yeast strains utilized in this study.

Reference: Amir Shahein, Maria López-Malo, Ivan Istomin, Evan J. Olson, Shiyu Cheng, Sebastian J. Maerkl. Systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and *in vivo* establishes their functional relevance. Nature Communications, 13, 5273 (2022).

Chapter 5. Systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and *in vivo*

5.1   Abstract

Binding to binding site clusters has yet to be characterized in depth, and the functional relevance of low-affinity clusters remains uncertain. We characterized transcription factor binding to low-affinity clusters *in vitro* and found that transcription factors can bind concurrently to overlapping sites, challenging the notion of binding exclusivity. Furthermore, small clusters with binding sites an order of magnitude lower in affinity give rise to high mean occupancies at physiologically-relevant transcription factor concentrations. To assess whether the observed *in vitro* occupancies translate to transcriptional activation *in vivo*, we tested low-affinity binding site clusters in a synthetic and native gene regulatory network in *S. cerevisiae*. In both systems, clusters of low-affinity binding sites generated transcriptional output comparable to single or even multiple consensus sites. This systematic characterization demonstrates that clusters of low-affinity binding sites achieve substantial occupancies, and that this occupancy can drive expression in eukaryotic promoters.

5.2   Introduction

DNA regulatory sequences, especially in eukaryotic organisms, frequently contain clusters of proximal binding sites (232; 233; 234). Both theory and experiment suggest that this clustering of binding sites is necessary for effective regulatory function (235; 236; 232). Accordingly, binding site clusters, in particular those composed of binding sites for the same transcription factor (homotypic clusters) or of low-affinity sites, have received increasing attention with several functionally important examples identified to date (237; 238; 239). For instance, clusters of low-affinity binding sites were found to be critical for the precise temporal control of gene expression (240), spatial control and patterning in development (237; 241; 242), as well as robustness to mutations (237), features that are hallmarks of native biological systems and highly relevant for engineering synthetic systems. However, our understanding of how such outcomes can be encoded in eukaryotic regulatory sequences and transduced by transcription factors, co-factors, and other biological machinery into function is still in its infancy. In prokaryotes (235; 243; 236; 244) and eukaryotes (245; 246; 247), biophysical models have had success in predicting gene expression based on sequence information alone, but we are missing comparable insights where DNA regulatory sequences contain complex clusters of binding sites. In order to build a better understanding of eukaryotic transcription, we argue that it is important to characterize and develop a quantitative understanding of how transcription factors bind to multiple proximal binding sites varying in affinity, multiplicity, and density (248).

The bulk of existing research on TF-DNA interactions has revolved around single binding sites, often focusing on consensus sequence or high-affinity binding sites. We currently lack a detailed, biochemical analysis of the binding of transcription factors to binding site clusters. For instance, while several reports identified that transcription factors act synergistically when binding to clusters of low-affinity sites (246; 249), whether this synergy is a biochemical property of the TF-DNA interaction or depends on additional *in vivo* complexity remains unknown. Current biochemical approaches for studying transcription factor-DNA interactions either do not provide a direct readout of the occupancy of multiple transcription factors interacting on DNA (250; 251), or lack the throughput necessary to tackle the larger sequence space that emerges when working with multiple sites (252; 249). Furthermore, current methods tend to be biased towards high-affinity interactions, due to the process of isolating bound molecules which can lead to dissociation of lower-affinity, transient binding events.

To fill this knowledge gap, we applied the quantitative, high-throughput MITOMI assay (45) to study binding site clusters. The MITOMI assay performs hundreds to thousands of parallel interaction measurements on a single microfluidic device by surface immobilizing transcription factors exposed to solution phase short DNA targets. MITOMI has been applied to the detailed quantitative characterization of protein - DNA (45; 46; 253; 228; 254; 50; 255), protein - RNA (256), and protein - protein interactions (257), as well as small molecule drug discovery (258). The MITOMI method has also been extended to enable the high-throughput analysis of

molecular association and dissociation kinetics (259; 260).

We slightly modified the original MITOMI assay by inverting the assay geometry using surface immobilized 90 bp-long target DNA strands and solution-phase transcription factors in order to optimize the quantitative characterization of binding site clusters. Reconfiguring MITOMI in this way allows for regulatory sequences ranging from single binding sites to large complex clusters to be encoded on DNA, and for the equilibrium occupancy of multiple transcription factor molecules to be observed. Given the throughput of our assay, we designed a library of DNA sequences that enabled direct assessment of how transcription factor binding is affected by changes to cluster configuration, including site affinities, multiplicities, and densities. We assessed transcription factors from two of the largest transcription factor families: zinc fingers and basic helix loop helix (bHLH).

We found that when DNA targets contain binding sites which overlap and share common basepairs, binding is only partially reduced compared to non-overlapping sites, suggesting that transcription factor molecules can withstand significant steric clash and simultaneously occupy two sites sharing basepairs. This effect, occupancy despite clash, decreased as the cluster density increased to require greater levels of steric clash in states where transcription factors bind neighboring sites, or as transcription factor occupancy on DNA became less energetically favorable due to lower site affinities. Binding to low-density clusters where sites do not share neighboring basepairs appeared largely independent, lacking prominent levels of binding synergy.

We also were able to show that, compared to an individual high-affinity binding site, clusters containing as few as three weak binding sites each an order of magnitude lower in affinity than the consensus sequence will reach greater occupancy levels *in vitro* at transcription factor concentrations that likely occur *in vivo*. Although clusters of weak binding sites achieved high-occupancies *in vitro*, it remained unknown whether occupancy from low-affinity, transient binding events would translate to an *in vivo* setting and give rise to functional regulatory elements. For instance, many past reports *in vivo* have considered dwell times to be an important factor determining functional gene regulation (261; 262; 263; 264; 265). To determine the functional relevance of weak binding site clusters *in vivo* we generated and characterized a synthetic library of minCYC1 promoters containing multiple Zif268 binding sites, and were able to show that the gene expression level driven by clusters of low-affinity sites can match those achieved by high-affinity, consensus binding sites. Finally, to determine whether clusters of weak binding sites could function in a native gene regulatory network we tested weak binding site clusters in the inorganic phosphate regulatory network. We replaced native binding regions containing high-affinity Pho4 and Pho2 binding sites with clusters of low-affinity Pho4 sites in the PHO5 promoter and found that in the context of this native regulatory system controlled by physiological levels of Pho4 transcription factor, gene expression can be recovered with binding site clusters consisting of individual binding sites an order of magnitude lower in affinity than the consensus site.

5.3 Results

5.3.1 iMITOMI Development and Characterization

We developed an inverted MITOMI assay (iMITOMI), by adapting the original MITOMI platform (45) to generate quantitative measurements of transcription factors binding at equilibrium to longer DNA targets containing binding site clusters of up to 6 distinct binding sites (Figure 5.1A-H) (191). We also tested and characterized these binding site clusters *in vivo* in the eukaryotic model organism *S. cerevisiae* (Figure 5.1I-J). In our assay, 90 bp-long double stranded DNA (dsDNA) is immobilized on the chip's surface, whereas protein is maintained free in solution and titrated across different regions of the chip (Figure 5.1D-H, Figure 5.8). This is in contrast to the original MITOMI assay in which transcription factor was immobilized to the surface and 30 bp-long target DNA was added in solution. Inverting the assay chemistry was important to study binding site clusters for several reasons. Using surface-immobilized DNA allows for multiple transcription factor molecules to bind and the resulting increase in fluorescence to be measured, whereas binding of multiple immobilized transcription factors to individual free molecules of DNA would present spatial constraints, and could lead to skewed affinity measurements due to avidity. The iMITOMI chip contains 768 unit cells, each consisting of a DNA spotting chamber, into which DNA targets are introduced by spotting during fabrication, and a MITOMI detection area, wherein binding occurs and detection is conducted (Figure 5.1D-F). The spotting chamber and detection area can be separated from each other by actuating a neck valve, and a sandwich valve can be actuated to isolate the individual reaction chambers. A circular MITOMI button valve lies above each detection area, and can be used to mechanically trap molecules bound to the surface in the detection area when this valve is pressurized.

The DNA in each unit cell is individually programmable with a DNA spotter, and fluid flow to each row of unit cells in the chip can be controlled through actuation of a combination of eight multiplexing valve control lines. We first generate a specific surface chemistry, and immobilize Cy5-tagged dsDNA from each spotting chamber under the button valve in the corresponding detection area before imaging the chip to quantify the surface-bound DNA (Figure 5.1D). We then flow recombinant mScarlet-tagged transcription factor into the chip (Figure 5.9), using a different concentration in each of the eight pairs of rows. After the DNA and protein are allowed to bind and reach equilibrium, we image the chip to quantify the concentration of free protein at equilibrium in each chamber. Then, by actuating the button valve to isolate DNA-bound protein, and after washing away free protein, we image the chip once more to quantify the amount of protein that was in complex with DNA at equilibrium. We can thus directly measure the equilibrium concentrations of free as well as bound protein. We then use these measurements to generate binding curves (Figure 5.1H) by relating the bound protein signal (normalized by the DNA signal) to the free protein signal at equilibrium. For each DNA target, a Markov-Chain Monte Carlo sampler is used to fit a 2-parameter saturation binding curve model.
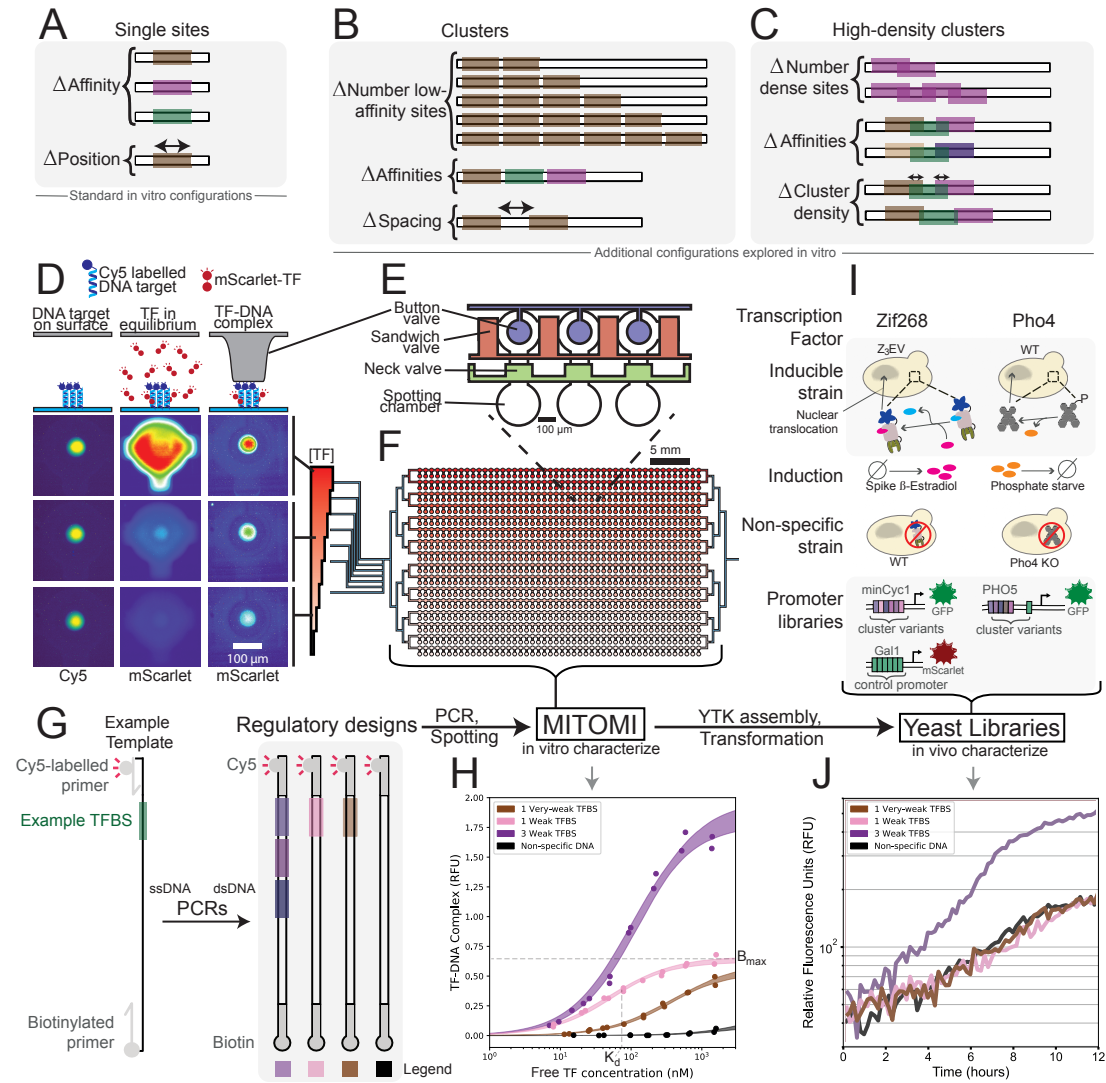
Figure 5.1: **Description of the *in vitro* and *in vivo* measurements.**
Caption on next page.

Figure 5.1: **(A-C)** DNA libraries and parameters explored *in vitro* using iMITOMI, including clusters of multiple binding sites **(B)**, and high-density clusters of overlapping binding sites **(C)**. **(D-F)** Schematic of the high-throughput, *in vitro* iMITOMI method. The iMITOMI chip consists of two-layers, where microvalves in the control layer **(E)** control the flow of fluid in a flow-layer containing 768 programmable reaction chambers **(F)**. **(D)** Image rows represent different TF concentrations, columns represent stages of the assay. DNA is immobilized on the chip's surface, in each assay chamber, under the button valves (left column). Fluorescently-tagged transcription factor molecules are flowed at different concentrations (red color gradient) into different rows of the chip. The free (middle column) and bound (right column) transcription factor signal is quantified at equilibrium. **(G)** To obtain the dsDNA targets, 90 bp-long single-stranded templates are amplified with biotinylated and Cy5-tagged primers. **(H)** Example 2-parameter saturation binding curves for DNA targets containing from 0 to 3 binding sites spanning a wide affinity-range. Shaded regions represent the 5 to 95% confidence intervals identified by exploring model parameter space with a Markov Chain Monte Carlo sampler. **(I)** *In vitro*-characterized clusters are assembled into expression cassettes, and chromosomally integrated into *S. cerevisiae* for characterization **(J)**. Zif268 regulatory sequence designs were characterized in the minCYC1 promoter using an inducible Z_3EV system, where a wildtype (WT) strain enabled quantification of the non-specific response **(I, left)**. Pho4 regulatory sequence designs were characterized in the PHO5 promoter, in a WT strain induced by phosphate starvation, where a Pho4 knockout strain enabled quantification of the non-specific response **(I, right)**.

To validate our method, we conducted several experiments to ascertain that the affinities and specificities measured with iMITOMI correspond to previous MITOMI measurements, which were validated against data obtained with other methods generally considered a gold standard (266; 45; 259). Affinity and specificity measurements taken for single binding site targets were found to be consistent with previous results obtained using standard MITOMI across a wide affinity range, both for Zif268 (267; 228) and Pho4 (45) (Figure 5.8). To test whether using surface-bound DNA might influence transcription factor binding, we placed single binding sites for Zif268 and Pho4 in different positions of the DNA target. No difference in either K_D or B_{max} was observed as the binding site is moved closer towards the chip surface (Figure 5.10), indicating that the distance from the chip surface did not introduce systematic differences in the ability of a transcription factor to bind to the target DNA strand.

5.3.2 *In vitro* characterization of binding site clusters

We applied our method to characterize the binding of Pho4 from the bHLH family and Zif268 from the zinc finger family to DNA targets composed of multiple binding sites, ranging from from 1 to 6 sites for Zif268 (Figure 5.2A-E) and from 1 to 5 sites for Pho4 (Figure 5.2F-J). We characterize compositions of weak sites (W) and very-weak sites (V), in the range of one, or two orders of magnitude lower affinity than the consensus site, respectively.

For each transcription factor, the bound transcription factor signal (normalized by DNA fluorescence) at saturating transcription factor concentrations (which we refer to as the saturation fluorescence, B_{max}) was similar for single strong, medium, and weak binding sites (Figure 5.10). As the number of weak binding sites in a DNA target is increased, we observe a linear increase in the saturation fluorescence (R^2 of 0.99 for Zif268 and 0.95 for Pho4), with a step size corresponding to the saturation fluorescence for single binding site targets (Figure 5.11). This suggests that the binding sites are saturating with transcription factor, and that the step increase in B_{max} (for Zif268, the slope of the line in Figure 5.11E) corresponds to the fluorescence signal resulting from an occupancy of 1 TF bound per DNA molecule (units of RFU per TF molecule bound on average). Therefore normalizing by this value converts our bound transcription factor signal (RFU) into a measure of mean occupancy, $\langle N \rangle$, which is simply the average number of transcription factors bound per molecule of DNA (Figure 5.2C-E, H-J). In the Pho4 DNA library, the fourth binding site we introduced was a very-weak site (Figure 5.10), and thus not fully saturated in the assay. Accordingly, in this case we fit our linear regression to the targets ranging from 1 to 3 binding sites to obtain the step increase in saturation fluorescence for normalization (Figure 5.2J).

Our results show that binding to clusters of sites an order of magnitude weaker than the consensus binding sequence can be saturated, and that all binding sites in even closely-spaced clusters can be concurrently occupied by transcription factors, resulting in a maximum occupancy $\langle N_{max} \rangle$ that consistently increases by 1 with each additional site in a DNA target (Figure 5.2E, J, upper panels). In contrast, the measured affinity (K_D) is determined by the individual binding site affinities present in the target cluster (Figure 5.2E, J, middle panels).
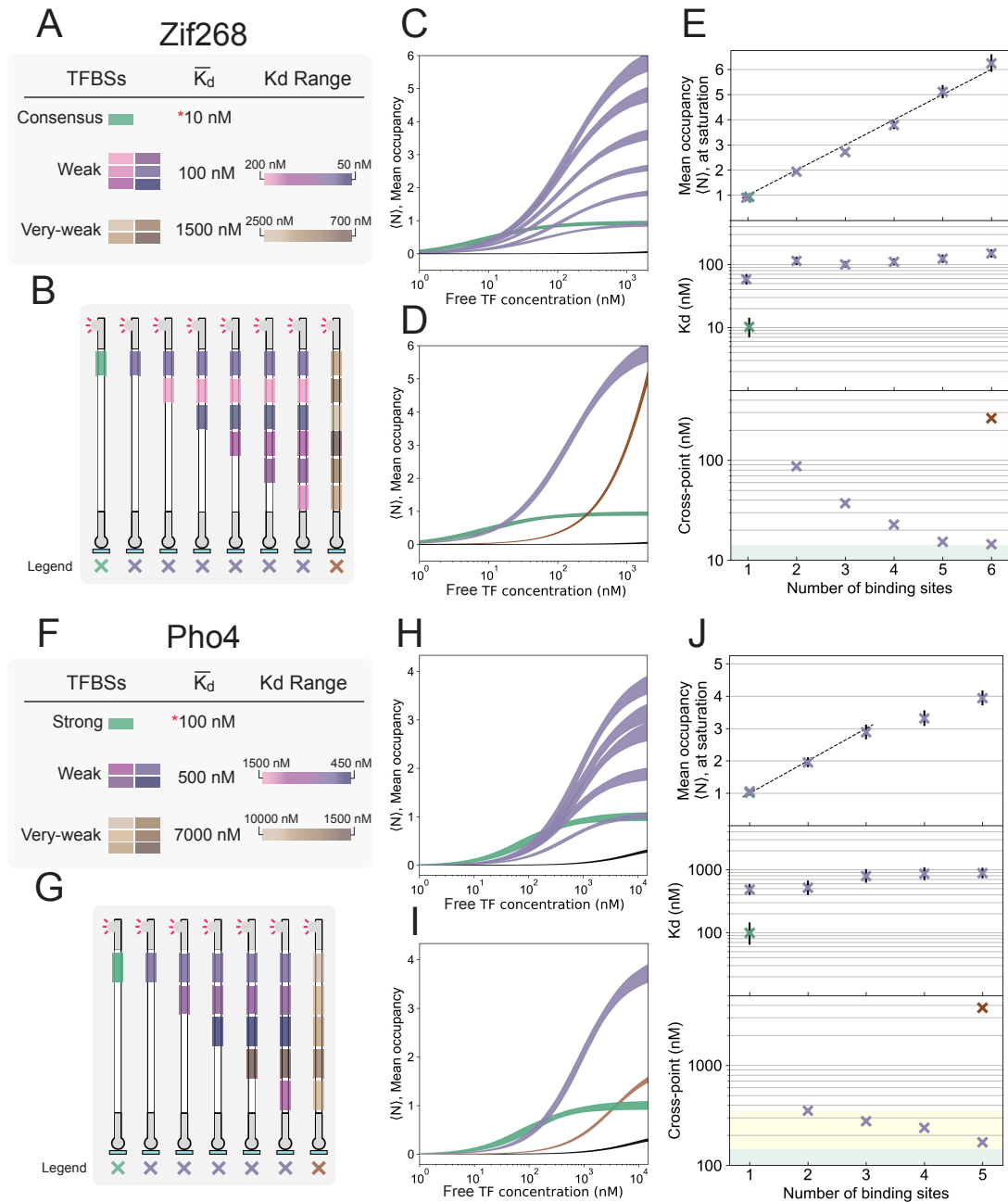
Figure 5.2: ***In vitro* characterization of low-affinity binding site clusters.**
Caption on next page.

Figure 5.2: **(A, F)** Legend for the Zif268 and Pho4 clusters analyzed. Binding sites were grouped into different affinity classes. Consensus or strong sites, weak sites approximately an order of magnitude lower affinity, and very-weak sites approximately two orders of magnitude lower affinity than the consensus site. **(B, G)** Schematic of the Zif268 and Pho4 DNA targets characterized. **(C, H)** Amount of TF in complex with DNA as a function of free TF concentration at equilibrium (saturation binding curves) for DNA targets containing from 1-6 (Zif268, **(C)**) or 1-5 (Pho4, **(H)**) low-affinity binding sites (purple) compared against 1 strong binding site (green). Shaded regions represent 5 to 95% confidence intervals. Further data and analysis available in Figure 5.11. **(D, I)** Saturation binding curves comparing the largest cluster of very weak sites (brown) against the largest cluster of weak sites and a single strong binding site. **(E, J, top)** Mean occupancies at saturating transcription factor concentrations, $\langle N \rangle_{max}$ (R^2 of 0.99 for Zif268, R^2 of 0.95 for Pho4). **(E, J, middle)** Measured affinity values (K_D). **(E, J, bottom)** Cross points: the transcription factor concentration at which the mean occupancy of a DNA target composed of multiple medium or weak binding sites reaches the same mean occupancy as the consensus or strong binding site target. The yellow shaded region represents a range of expected *in vivo* transcription factor concentrations, and the green shaded region represents the characterized K_D for the strong binding site. Error bars represent 5 to 95% confidence intervals.

Having shown that clusters of weak and very-weak binding sites can be fully occupied at high transcription factor concentrations of 1 μM, we next asked whether weak and very-weak clusters also lead to sufficient occupancy at lower transcription factor concentrations. The *in vivo* nuclear concentration of Pho4 in *S. cerevisiae* has previously been estimated to be approximately 355 nM based on single-cell fluorescence experiments (245). Estimates for zinc-fingers in yeast range from 538 to 3334 copies per cell, which translates to a nuclear concentration of roughly 213 nM to 1322 nM (268; 269). As single consensus sequence targets are known to be physiologically functional in a cellular milieu, we determined the concentrations at which low-affinity binding site clusters exhibit similar occupancy levels as the corresponding single consensus sequence (Figure 5.2E, J, lower). In other words, we plotted the concentrations at which the low-affinity cluster saturation binding curves cross the single consensus target saturation binding curve. It can be readily seen that similar occupancies can be reached by all low-affinity clusters, with cross-over occurring at transcription factor concentrations of as low as 14 nM for a cluster of 6 low-affinity Zif268 binding sites, and 170 nM for a cluster of 5 low-affinity Pho4 binding sites. These values are well within the range of physiological concentration estimates and indicate that low-affinity binding site clusters may achieve comparable occupancy levels as a single high-affinity consensus site *in vivo*.

We exploited our platform to explore the impact of binding site proximity on binding affinity (Figure 5.3). First, we set out to determine whether proximal binding sites exhibit synergistic interaction as has been reported *in vivo* (246). We constructed and characterized a library of targets with two low-affinity binding sites, where the position of one site was held constant, while the position of the second site was moved in intervals to adjust the spacing between sites (Δ). For Zif268, binding remained largely unchanged across the library, barring one unexplained outlier (Figure 5.3A, 5bp spacing). In the case of Pho4, we observed a decrease in $\langle N \rangle_{max}$ for gap distances centered around ~ 3 bp, possibly indicating steric competition in this orientation of transcription factor molecules (Figure 5.3B). The K_D values remained largely constant as a function of distance, and corresponded to the individual binding site affinities. We therefore did not observe a large degree of cooperativity between adjacent binding sites (Figure 5.11G), implying that the reported synergism may be a product of higher-order complexity *in vivo*. Together, our results suggest that the equilibrium binding we observe to different binding sites in targets composed of multiple sites is largely independent, possibly with mild steric effects at close distances.

Overlapping binding sites occur frequently in regulatory DNA of prokaryotes and eukaryotes (270; 243; 271; 272; 244; 273), and are thought to influence how binding occurs among transcription factors (246; 271; 274; 275; 276), RNA polymerase, other regulatory elements (277; 271; 278; 279; 280), and nucleosomes (251). In eukaryotes, the relatively low specificity of transcription factors gives rise to regulatory sequences that contain high-density clusters of binding sites tending to overlap with one another (243; 271; 244). As a result of high binding site densities, De Boer et. al. found that shifting a site by one basepair across regions of a yeast promoter impacts expression primarily through the disruption or creation of alternate
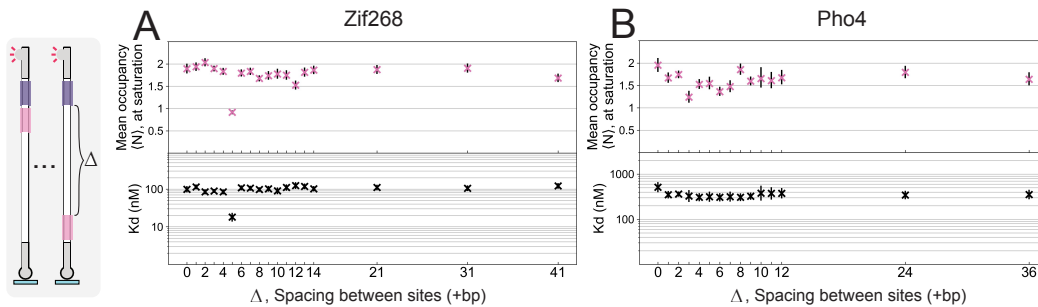
Figure 5.3: **Effect of distance between two adjacent binding sites. (A, B)** Affinities (K_d), and mean occupancies at saturation ($\langle N \rangle_{\max}$), for a library of DNA targets for Zif268 (A) and Pho4 (B), with different distances between two weak binding sites.

transcription factor binding sites ((244)). While competition between transcription factors binding to nearby but distinct binding sites is typically thought of as a graded function of distance, as a rule of thumb overlapping binding sites are often assumed to elicit exclusive transcription factor binding due to steric clash (275; 281), and consequently binding models tend to ignore states where two transcription factors are bound simultaneously to sites sharing common basepairs (246).

To our knowledge, binding of transcription factors to high-density clusters has not been characterized in detail. Thus we designed DNA targets containing from 1 to 4 binding sites of varying affinity for Zif268, where one or three basepairs are shared between neighbouring binding sites (Figure 5.4A). An alignment of the protein-DNA complexes to the DNA sequences together with residue-clash prediction suggests that the protein molecules will exhibit significant steric clash as binding site motifs start to overlap and share common basepairs (Figure 5.4B, C). Yet surprisingly, we discovered that the mean occupancy of transcription factor molecules binding to two binding sites that share basepairs can exceed a value of 1 at the upper range of physiologically relevant transcription factor concentrations (Figure 5.4D). This demonstrates that occupancy despite clash can occur, where two transcription factor molecules bind at once and exhibit steric interference on binding sites sharing common basepairs, and that this might occur in high-density clusters in the genome.

Since a simple one-site saturation binding model fails to describe this phenomenon, we extended our description of binding to a statistical mechanical model. First we attempted to model binding to high-density clusters using an exclusive binding model (full steric occlusion) where two transcription factors cannot bind concurrently to overlapping binding sites, resulting in just three states of the system for a 2-site cluster (Figure 5.4D, Figure 5.12). We parameterized the model using binding energies (ε_{bs}) derived from an independent characterization of the individual binding sites. As expected, for a two-site cluster the exclusive model fails to account for mean occupancies above 1. We incorporated an additional permissive state into the model, where two transcription factors can bind simultaneously, and

fit a floating interaction energy between them (ε_{clash}) (Figure 5.4D,E, Figure 5.12, Table 5.4). This 2-parameter (1 floating) model described the experimental data better than the exclusive model. Both the Akaike information criterion (AIC), and Bayesian information criterion (BIC) decreased significantly for the permissive model, which are measures of goodness-of-fit, and useful for conducting model selection when models differ in their number of parameters (penalizing the permissive model for its additional parameter, Figure 5.12). As the number of basepairs shared between binding sites increases, the mean occupancy decreases (as a function of concentration), and the probability of the 2 transcription factor-bound state is reduced, although it is still higher than initially expected at a large amount of overlap (3 basepairs, or 33% of the motif, Figure 5.4F). Furthermore, as we decrease site affinities, we observe a large decrease in occupancy in overlapping sites, suggesting that the transcription factor's affinity for its binding site (ε_{bs}) is important for compensating for the steric interference. We further investigated binding to high-density clusters where individual sites differ in their affinities (Figure 5.4G, H). In this case, binding is dominated by the stronger binding sites in a given cluster. For the clusters of three mixed-affinity binding sites, the maximum occupancies and affinities were chiefly determined by the number of strong, non-overlapping binding sites present in the cluster.
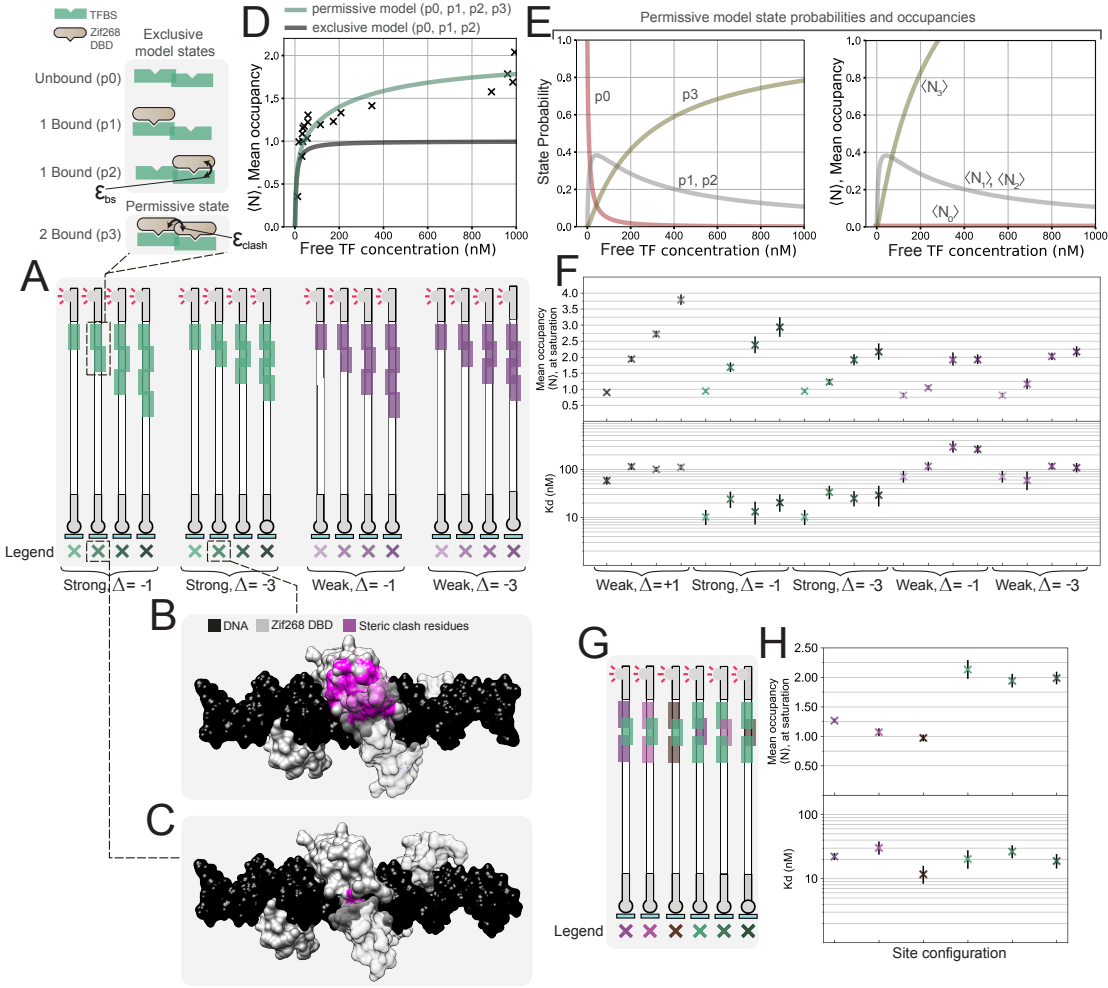
Figure 5.4: ***In vitro* characterization of high-density binding site clusters.**
Caption on next page.

Figure 5.4: **(A)** Library of DNA targets analyzed containing Zif268 high density binding site clusters, where basepairs are shared between neighboring binding sites. Clusters contain up to four strong or weak binding sites, and neighboring sites share 1 or 3 common basepairs. **(B, C)** Aligned crystal structures, highlighting Zif268 residues expected to exhibit steric clash in magenta. 3 basepairs (33% of the Zif268 motif) shared in **(B)**, and 1 basepair (11% of the motif) shared in **(C)**). **(D-E)** Statistical mechanical analysis of binding to consensus binding sites sharing a common basepair. **(D, left)** Legend of the states corresponding to two different binding models, an exclusive binding model (top three states), and a permissive binding model (all four states). **(D, right)** Binding data and model predictions, illustrating that two transcription factors can bind at once to binding sites sharing basepairs. **(E)** Predicted probability **(left)** and occupancy **(right)** for each permissive model state as a function of transcription factor concentration. **(F)** Experimental data for the full library of high-density clusters (data fit with a 2-parameter SBC similar to above). Occupancies at saturation, $\langle N \rangle_{\max}$, and affinities, K_D. **(G)** DNA targets containing mixed-affinity high-density clusters where 3 basepairs are shared between neighboring binding sites. **(H)** Corresponding experimental data for mixed-affinity clusters.

5.3.3   *In vivo* validation of low-affinity binding site cluster function in a synthetic gene regulatory system

Experimental methods and computational analyses designed to study gene regulation have traditionally been biased towards single sites of high-affinity or small clusters of high-affinity binding sites. Homotypic clusters of low-affinity sites are prevalent in eukaryotic regulatory sequences (232), yet the level of binding to these clusters in the *in vivo* context and the resulting impact on gene expression is uncertain. Having shown *in vitro* that at physiologically-relevant transcription factor concentrations an equal or greater mean transcription factor occupancy can be achieved by low-affinity clusters compared to an individual consensus site, we set out to investigate whether low-affinity site clusters are functional in the context of eukaryotic promoters *in vivo*. Although our biochemical analysis showed that low-affinity binding site clusters are able to achieve high mean occupancies, several questions remained as to whether low-affinity binding site clusters would be functional *in vivo*. First, it is not a priori known whether high mean occupancies of transcription factors alone are sufficient to recruit the necessary regulatory machinery to a promoter and give rise to transcriptional activation. Consensus binding sites are able to achieve high mean occupancies of a single transcription factor, and this occupancy is further characterized by long-dwell times of the transcription factor on its binding site (259). Low-affinity clusters on the other hand achieve a similar mean occupancy in non-saturating conditions through a time-averaged occupation of several transcription factors bound to the cluster, and each transcription factor - DNA interaction exhibits a considerably shorter dwell time. It is a matter of debate whether high occupancies or long dwell times are necessary for *in vivo* function of regulatory regions. Second, we estimated that *in vivo* transcription factor concentrations may be high enough to give rise to sufficiently high occupancies of low-affinity clusters, but this assumption remained to be tested *in vivo*.

We first addressed the question of whether low-affinity binding site clusters can be functional *in vivo* by generating synthetic minimal promoters, driven by an engineered, exogenous Zif268 transcription factor. We inserted the 90 bp regulatory sequences that were characterized in our *in vitro* work approximately 200 bp upstream of the transcription start site in the minimal CYC1 promoter (Figure 5.5A). These sequences characterized *in vitro* were originally designed into a DNA backbone selected to prevent binding to known yeast transcription factors. We varied similar regulatory sequence parameters and maintained similar affinity-class definitions as for our low-affinity *in vitro* binding characterization. We used the yeast toolkit (YTK) 2.0 workflow (176) to assemble integration cassettes with our engineered minCYC1 promoters upstream of GFP. We integrated these synthetic cassettes into the Lys2 locus of a modified Z_3EV-containing parental yeast strain (Figure 5.5B) (282). Z_3EV is a synthetic transcriptional activator with three domains: the Zif268 DNA-binding domain, an estrogen receptor element that drives nuclear translocation in the presence of β-estradiol, and a VP16 domain that recruits transcriptional machinery to activate transcription. We also included a control promoter into the parental strain which drives mScarlet expression under the regulation of 6 consensus Zif268 binding sites in a synthetic GAL1 promoter backbone. The same parental strain was used for all strains
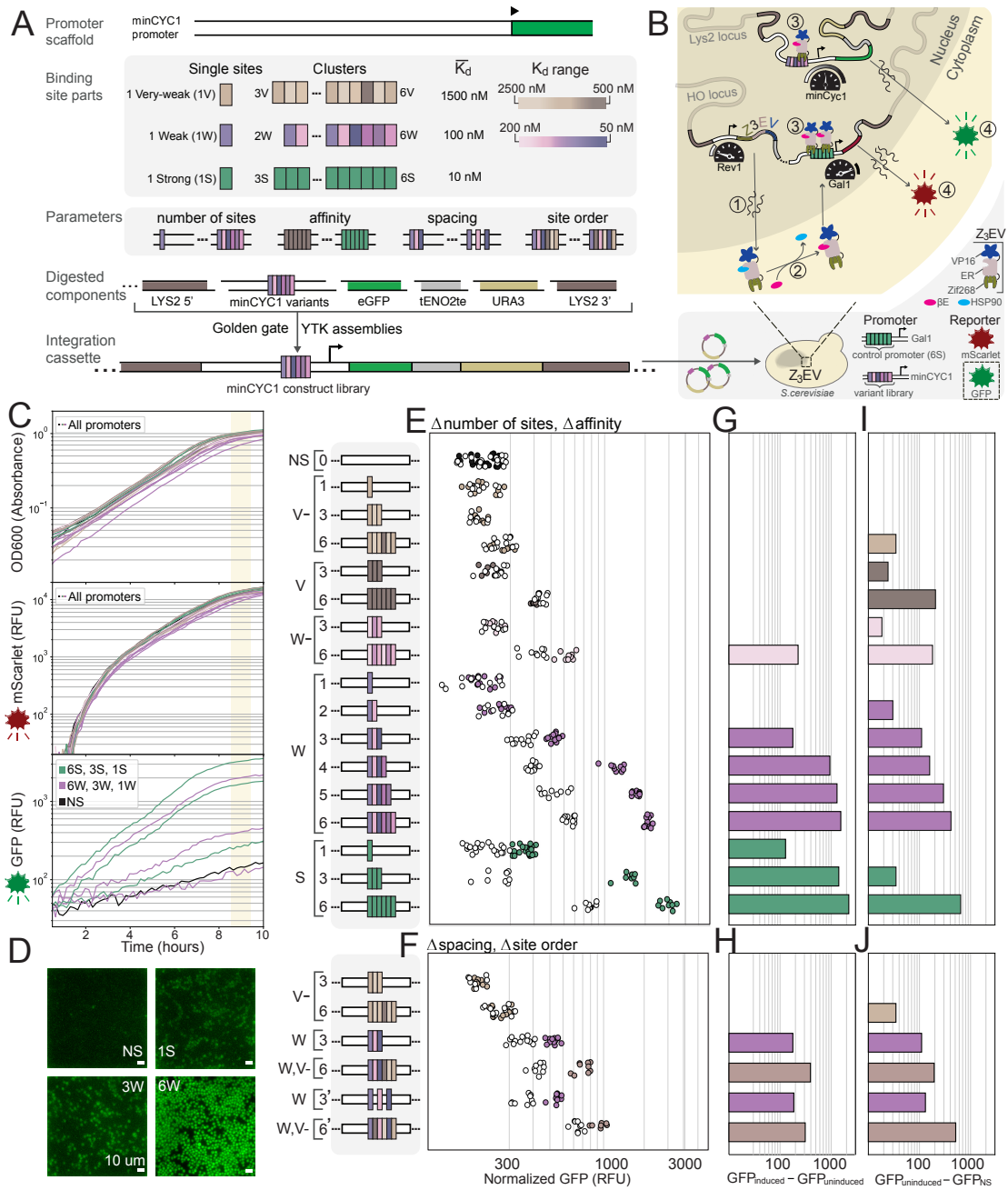
Figure 5.5: *In vivo* **characterization in a synthetic, minimal gene regulatory system.**
Caption on next page.

Figure 5.5: **(A)** Promoter legend. *In vitro* characterized Zif268 binding regions were inserted into a minCYC1 minimal promoter scaffold. Promoters were assembled into multi-cassette plasmids of standard parts driving GFP production, using the yeast-toolkit workflow (176), and integrated into the LYS2 locus of a Z_3EV parental strain (282). **(B)** Z_3EV is a synthetic transcription factor containing a Zif268 DNA-binding domain, an estrogen-responsive (ER) domain responsible for nuclear translocation upon β-estradiol induction and binding, and a VP16 activation domain. The parental strain was engineered to contain a synthetic GAL1 control promoter with 6 consensus binding sites driving mScarlet production, to quality-control Z_3EV activity. **(C)** All strains displayed consistent growth (OD600, upper), and mScarlet production (middle). Strains varied in their GFP production depending on the minCYC1 promoter variant (bottom). Strains were characterized with single-cell fluorescence microscopy **(D)** and on a multi-mode platereader **(E-J)**. **(E-F)** Strains were characterized either without (white, uninduced), or in the presence of saturating (200 nM) β-estradiol concentrations (colored, induced), over three independent experimental measurements. **(E)** The number of binding sites was varied from 1-6, and six distinct affinity classes were characterized (consensus binding sites (S), weak sites (W, 10X lower-affinity, W-, 20X lower-affinity), very-weak sites (V, 50X lower-affinity, V-, 100X lower-affinity), as well as a non-specific DNA target (NS). Promoters are organized from weakest (top) to strongest binding sites (bottom). **(F)** We varied the spacing between sites in clusters, characterized clusters with binding sites from mixed affinity classes, and changed the ordering of binding sites in mixed-affinity clusters. **(G-H)** A Z_3EV-specific response was quantified as the difference in fluorescence intensity between induced and uninduced conditions. **(I-J)** A non-specific response was quantified as the difference in fluorescence intensity in uninduced conditions between a given promoter strain and the NS promoter-containing strain. The non-specific response also showed strong correlation with gene expression measurements of the same promoter library integrated in a BY4741 yeast strain lacking Z_3EV (Figure 5.13).

generated, allowing us to validate the correct functioning of the inducible gene expression system to ensure consistency across strains.

Yeast strains were either induced with a high concentration of β-estradiol (200 nM), or not uninduced (0 nM), and measured in the mScarlet and GFP channels on a multi-mode platereader (Figure 5.5C) and cross-validated by direct fluorescence microscopy imaging (Figure 5.5D, Figure 5.13). All strains exhibited highly consistent growth rates and mScarlet expression from the control-promoter (Figure 5.5C, Figure 5.13E-H). In contrast, the cluster variants in our engineered promoter library led to a wide range of gene expression levels (GFP) that depended on cluster configuration and binding site affinities. We normalized GFP intensities by cell densities, and averaged this normalized gene expression signal for all readings between 8.5 hours and 9.5 hours after induction (Figure 5.5E, F). Each datapoint in Figure 5.5E corresponds to an independent replicate, where generally around three replicates were characterized in each experiment, and the experiment was repeated from three to six times for each strain. All promoters containing 1, 3, or 6 consensus binding sites gave rise to functional activation and expression of GFP. The minCYC1 promoter containing a single consensus binding site achieved only modest expression levels, whereas 3 consensus binding sites led to very high expression of GFP (Figure 5.5E-H). Increasing the number of consensus binding sites from 3 to 6 led to slightly higher gene expression levels, potentially indicating that expression levels are near the upper limit of what the minCYC1 promoter can support. The output of a promoter containing no binding sites (NS) on the other hand remained low, both in the induced and uninduced conditions. We tested binding site clusters containing between 1 and 6 very weak binding sites with average K_D values 2 orders above the K_D of the consensus sequence. None of these promoters led to a measureable increase in gene expression when comparing the induced to the non-induced state, with the possible exception of a promoter containing 6 binding sites with slightly higher binding affinities (W-6). We started observing a difference in gene expression between the induced and non-induced conditions, when testing clusters of weak binding sites with average K_D values one order of magnitude above the K_D of the consensus sequence. Clusters of as few as 3 weak binding sites (W3) surpassed the single consensus binding site in its ability to activate gene expression, while clusters of 4, 5, or 6 weak binding sites achieved a similar level of gene expression compared to clusters composed of three consensus binding sites. Therefore, we were able to show in this minimal synthetic gene regulatory system, that small clusters of low-affinity binding sites can drive gene expression equivalently to clusters of consensus binding sites.

Interestingly, we found that not only did promoters vary in their induced levels of GFP-expression, but they also varied significantly in uninduced gene expression in a way that depended consistently on the promoter class. Increasing numbers of binding sites within a cluster gave rise to increasing levels of leakiness in the uninduced condition (Figure 5.5I-J). The highest level of leakiness was observed for a cluster of 6 consensus sequences (S6), and the level of leakiness scaled with the number of weak binding sites in a cluster (W1-6). Even non-functional clusters of very weak binding sites gave rise to increased leakiness (V6
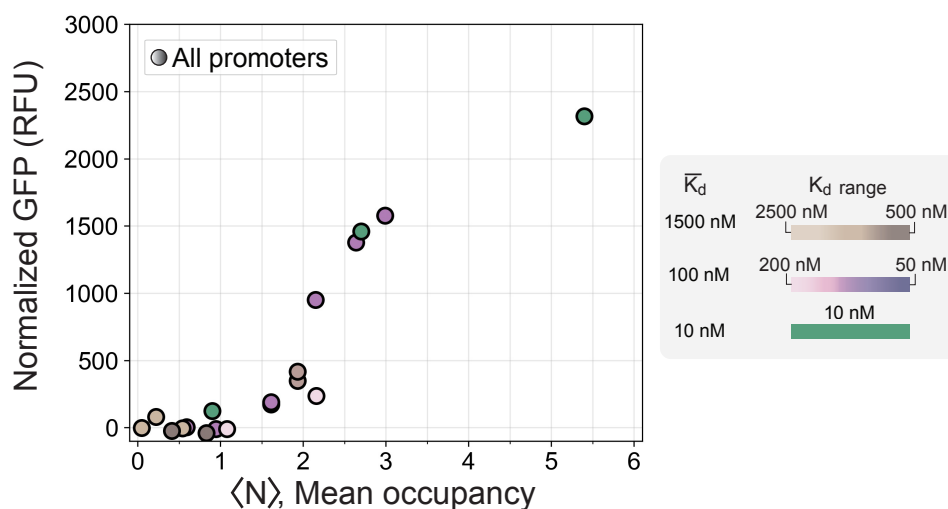
Figure 5.6: **Thresholded gene activation.** Thresholded relationship of gene expression (GFP normalized by OD600) as a function of predicted mean occupancy (based on *in vitro* characterization data), for the full Z_3EV strain library. Color scheme is the same as in Figure 5.5.

and W-6). The one exception to this general observation was the promoter containing 3 consensus sequences (S3) which achieved high levels of induction without any significant increase in leakiness. We reasoned that this response could either be specific and due to Z_3EV transcription-factor leak under non-inducing conditions, or non-specific, resulting from the binding of alternative endogenous transcription factors. Therefore, we characterized the entire promoter library in a wildtype BY4741 yeast strain lacking Z_3EV, and discovered that the level of expression correlated strongly with uninduced gene expression in the Z_3EV parental strain (R^2 of 0.9, Figure 5.13). This confirmed that the leakiness in the uninduced state was due to non-specific binding of one or more endogenous transcriptional activators.

Our library of well-characterized regulatory sequences where mean occupancy and affinity vary independently is useful for trying to understand which binding properties predominantly drive transcription from clusters. In eukaryotes, there are limited and conflicting examples where gene expression has been explained using a reduced set of interpretable binding properties (236; 264; 244; 246; 245; 247). We used our *in vitro*-characterization data to predict the mean occupancies of transcription factor molecules binding to clusters in the Z_3EV promoter library *in vivo*. We compared these predictions to the TF-specific promoter responses (Figure 5.6).

Our model suggests that the minCYC1 promoter regulates gene expression through a simple thresholded relationship with the average number of bound transcription factor molecules, where up until an average of around 1 transcription factor bound to the promoter, gene expression is largely silent, and after this threshold, expression depends primarily on the

average number of transcription factor molecules bound rather than on their affinity for their binding sites. For instance, a cluster composed of 5 weak binding sites (W5) or 3 strong binding sites (S3) differ in affinity by an order of magnitude, but show similar levels of gene expression which can be explained by their predicted mean occupancies. These results do not vary qualitatively within a reasonable range of expected parameter values (Figure 5.13). Thresholding can also be observed qualitatively, since one weak site (W1), or two weak sites (W2) show no impact on expression, but each additional weak site from then on increases expression significantly (Figure 5.5E). Furthermore, adding 3 very-weak sites to a very-weak cluster (moving from V-3 to V-6) does not increase transcription factor-specific expression, while appending these same 3 very-weak sites just after a cluster that has a higher starting occupancy (moving from W3 to W3V-3) causes a transcription factor-specific increase in promoter output (Figure 5.5F). We observed no difference in expression level when we changed spacing between binding sites (W3 compared to W'3), and little to no difference when we changed the ordering of sites in promoters containing binding sites from different affinity classes (W3V-3 compared to W3V-'3V).

5.3.4   *In vivo* validation of low-affinity binding site cluster function in a native gene regulatory system

Having established that clusters of low-affinity binding sites are functional in a synthetic gene regulatory network, we next addressed the question of whether low-affinity binding sites clusters would also be functional in a native promoter regulated by a transcription factor expressed at physiological concentrations. We turned to the well-studied inorganic phosphate regulatory network in *S. cerevisiae* to address this question. The PHO5 promoter is one of the best understood promoters of the inorganic phosphate regulatory network in yeast. It is activated in response to phosphate starvation, which causes the master regulator Pho4 to localize to the nucleus leading to binding to a Pho4 binding site and an overlapping Pho2 site located in a nucleosome-free region (NFR) (Figure 5.7A). This binding results in the displacement of a nearby nucleosome (in the -2 position), which in turn makes a higher-affinity Pho4 site and two nearby Pho2 binding sites available for binding which are nominally in a nucleosome-occluded region (NOR). This mechanism is believed to decouple the promoter's activation threshold from its dynamic range, where as a first approximation the Pho4 and Pho2 binding sites in the NFR are considered to confer the promoter with its threshold and all binding sites in the NFR and NOR determine gene expression levels after that threshold is met (283; 245).

Significant effort has been invested to uncover the functional relevance of the individual binding sites in the PHO5 promoter. A Pho2 binding site in the NFR has been found to recruit chromatin remodellers and to promote cooperative binding of Pho4, while the Pho2 sites in the NOR have been identified as necessary for efficient activation by Pho4 (284; 285). Ablation of Pho2 binding sites has been found to decrease gene expression up to 10-fold (245; 284), and a Pho2 knockout strain exhibits close to no PHO5 promoter expression compared to a wildtype strain (286). However, Pho4 overexpression has also been shown to compensate partially for the loss of expression due to Pho2 binding site ablation (284), or even for the loss of expression in Pho2 knockout strains (286; 287; 288), presumably through an increase in occupancy of the native Pho4 binding sites. Based on this picture, and since weak clusters of Pho4 binding sites attain similar to greater occupancies than the individual consensus or native sites at relevant transcription factor concentrations *in vitro* (Figure 5.2), we hypothesized that weak binding site clusters might achieve sufficiently high Pho4 occupancies in the PHO5 promoter and reproduce part of the functionality conferred by the native regulatory sequence, both in the NFR and NOR.

We replaced the native Pho4 binding sites together with their neighboring Pho2 sites, in the NFR or NOR, with clusters of 5 low-affinity Pho4 binding sites (weak or very-weak) (Figure 5.7A-C) that we had characterized *in vitro* (Figure 5.2). We conducted this replacement either in a native promoter scaffold (WT), or a scaffold entirely depleted of known Pho2 binding sites (delPho2), in order to decouple the Pho2-specific effect from the effect of Pho4 occupancy. PHO5 promoters were assembled into a cassette driving GFP reporter expression (Figure

5.7C), and integrated into the LYS2 locus of either a wildtype yeast strain (BY4741) or a Pho4 knockout strain lacking Pho4 (Figure 5.7D). All strains were sequence-verified, and exhibited consistent growth irrespective of their PHO5 promoter variant (Figure 5.7E). The positioning of nucleosomes in the PHO5 promoter has been characterized extensively and found to be robust to promoter context, such that it remains consistent when the promoter is located in an alternate locus (e.g. the LYS2 locus) and for a large degree of modification to the promoter sequence (288; 245).

Our results on deleting native Pho4 binding sites in the NFR and NOR regions are in concordance with previous characterizations of the PHO5 promoter (283; 245) (Figure 5.7E-J, Figure 5.147). Indeed, we observed greater expression for the WT promoter with a native NFR but an ablated NOR Pho4 site (WT N-A), compared to the promoter where both NFR and NOR Pho4 sites were ablated (WT A-A). This level of expression is then scaled with the addition of the native NOR Pho4 site (WT N-N). On the other hand, the promoter with an ablated Pho4 site in the NFR and a native Pho4 site in the NOR (WT A-N), resulted in similar expression to the ablated-ablated promoter (A-A), presumably because the nucleosome was not displaced effectively upon induction.

Compared to the wildtype A-N promoter, replacing the Pho4 binding site in the NFR (which also deleted the Pho2 binding site) with a cluster of low-affinity Pho4 binding sites (WT C-N) recovered Pho4-specific expression. Furthermore, compared to the N-A promoter, replacing the NOR (which also deleted the Pho2 binding sites) with a cluster of low-affinity sites (WT N-C) also improved expression compared to the N-A promoter. Promoters with no native Pho4 binding sites and a low-affinity binding site cluster in the NFR (C-A) or NOR (A-C) lost function similar to promoters with the corresponding native Pho4 binding sites present (N-A and A-N, respectively). The wildtype C-N promoter showed significantly higher expression than wildtype C-A promoter (and A-N showed ablated levels of expression), suggesting that the low-affinity cluster in the NFR may contribute to nucleosome displacement to uncover the NOR. Our results show that low-affinity clusters can generate similar expression levels of the PHO5 promoter compared to what is conferred by the native cis-regulatory sequence of the NFR and the NOR. Partial recovery of function was expected based on the aforementioned work demonstrating that overexpressing Pho4 (presumably increasing its concentration and occupancy) will only partially compensate for Pho2 site ablation (284; 286).

Indeed, when all Pho2 binding sites are deleted from the promoter, gene expression from the native promoter (delPho2 N-N) drops by around 6-fold, consistent with previous reports (284; 245). Remarkably, in this Pho2-depleted scaffold, replacing the native Pho4 sites in either the NFR or NOR with clusters of low-affinity binding sites (delPho2 C-N, or delPho2 N-C, respectively) actually resulted in a similar or greater level of gene expression than what was achieved by the native binding sites lacking the neighboring Pho2 sites. This implies that the Pho4-specific impact on gene expression of low-affinity Pho4 clusters is equal to or greater

than for the native Pho4 binding sites and its potential co-operative interaction with Pho2. This then suggests that low-affinity clusters attain significant occupancy, which we previously established *in vitro* and in a minimal promoter *in vivo*, and can be extended to native and more complex promoter settings.

Expression in the delPho2 C-N strain does not decrease compared to the WT C-N strain (containing all Pho2 binding sites except for the one replaced by the cluster in the NFR), whereas compared to the WT N-C the delPho2 N-C strain shows lower expression, suggesting that the Pho2 binding site in the NFR is of greater importance to gene expression as has been suggested in the literature (284). Indeed, our results suggest that ablating the nucleosome-free Pho2 site hinders the efficacy of the NFR, likely in part by reducing cooperative binding to Pho4 (284). However, our results suggest that the function of this regulatory region can be compensated for by strengthening the region's Pho4 mean occupancy through a low-affinity Pho4 binding cluster, leading to a higher level of gene expression.

Interestingly, similar to the Z_3EV promoter system, non-specific leaky expression correlated with the number of low-affinity binding sites (Figure 7K). Dual-cluster replacement resulted in significant non-specific expression, which did not depend on Pho4. This is evidenced by the fact that expression did not depend on induction in the wildtype background. Furthermore, these strains were the only ones to exhibit a high level of expression at timepoint zero of induction, and their expression did not increase over time. This was the case for dual-cluster replacement both in the native promoter scaffold and in the scaffold depleted of Pho2 binding sites. In further agreement, these promoters showed significant expression in the δPho4 background (Figure 5.7L). Replacement of the NOR alone by a weak-cluster was enough to drive a moderate level of non-specific expression, while this was not the case for cluster-replacement in the NFR. Aside from the above discussed strains, all other strains showed little to no non-specific expression. They exhibited consistent, ablated levels of gene expression in the absence of induction, and in the δPho4 background.
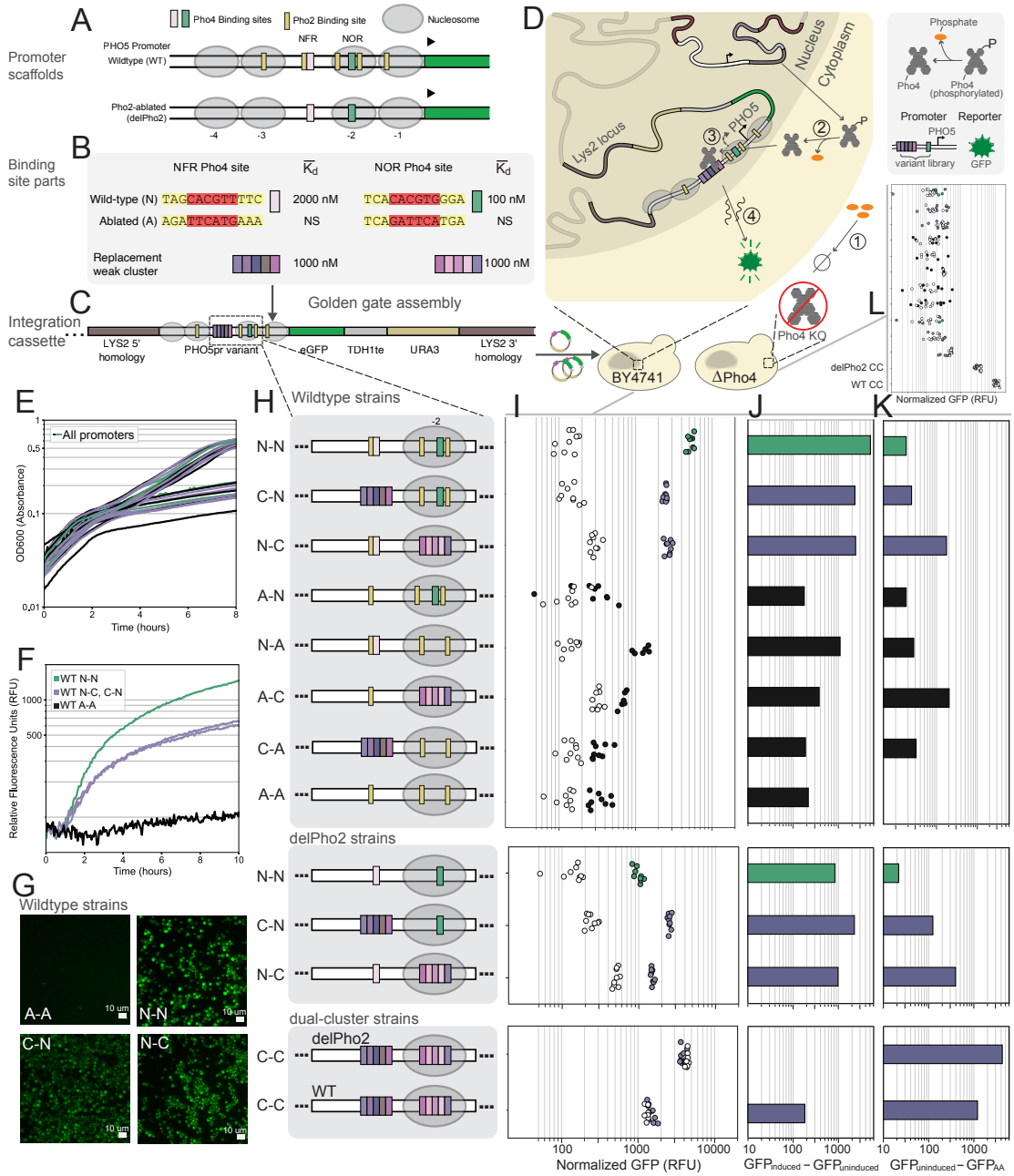
Figure 5.7: *In vivo* **characterization in a native gene regulatory system.**
Caption on next page.

Figure 5.7: **(A)** PHO5 promoter scaffolds in which native Pho4 binding sites are replaced with low-affinity Pho4 clusters. **(B)** Legend of Pho4 binding site parts. Native binding regions containing Pho4 and Pho2 sites either have their Pho4 binding sites ablated, or are replaced with clusters of low-affinity Pho4 binding sites. **(C)** Promoters were assembled into a multi-cassette plasmid and integrated into the LYS2 locus of either BY4741 or Pho4 knockout (δPho4) yeast strains. **(D)** Schematic of the response (induction) to phosphate starvation. **(E)** OD600 time series measurements for the full BY4741 library of promoters, both for uninduced and induced conditions. **(F)** Induction time series measurements for select cluster-containing strains. **(G)** Strains were also characterized by single-cell fluorescence microscopy. **(H-J)** A Pho4-specific response was quantified as the difference in fluorescence intensity between induced and uninduced conditions. **(K)** A non-specific response was quantified as the difference in fluorescence intensity in uninduced conditions between a given promoter strain and the ablated-ablated (AA) promoter strain. **(L)** The non-specific response can also be seen in a Pho4 knockout strain, with a similar relationship across promoter variants (Figure 5.14).

5.4 Discussion

Our current understanding of how transcription factors bind to genomic regulatory regions to drive transcription has developed largely through experiments that characterize binding to individual binding sites, often with a focus on high-affinity consensus sites. The functional relevance of low-affinity binding sites, particularly clusters of low-affinity binding sites is less-well understood (238). Methodological challenges associated with measuring binding of multiple transcription factors to low-affinity clusters has prevented precise measurements of how binding to these clusters compares with binding to individual consensus binding sites, creating uncertainty about the significance of low-affinity clusters. Indeed, through prior state-of-the-art biochemical methods used to measure collective binding in high-throughput, occupancy appeared driven primarily by strong binding sites (250). Reports in the literature identified *in vivo* examples where low-affinity clusters are functionally important (237; 238; 239; 240; 241; 242), or exhibit properties like binding synergy (246; 249). But, the lack of a systematic analysis *in vitro* or *in vivo* limited us to inductive reasoning when trying to apply these findings to new contexts. By combining *in vitro* biochemical analysis with *in vivo* functional studies of gene regulation in the eukaryotic model system *S. cerevisiae* we were able to provide insights into how transcription factors interact with low-affinity clusters and were able to show that these clusters are functional *in vivo.*

We applied a high-throughput *in vitro* method to characterizing collective transcription factor binding to DNA sequences consisting of arrays of binding sites ranging over orders of magnitude in affinity. This permitted us to develop a quantitative biophysical understanding of how transcription factors interact with binding site clusters, particularly to clusters consisting of low-affinity binding sites. Clusters of low-affinity binding sites achieved similar levels of occupancy as single high-affinity binding sites and we predicted that this level of occupancy can be achieved at physiologically relevant transcription factor concentrations. We verified this effect with two distinct transcription factors representing two of the largest transcription factor structural classes: zinc finger and bHLH. Small clusters of 2-6 binding sites with individual binding site affinities one order of magnitude lower than the consensus sequence were able to achieve this effect. We furthermore challenged the notion that overlapping binding sites are mutually exclusive due to steric effects, and thus would allow only one transcription factor to bind to overlapping sites at any given time. In fact, clusters of strong binding sites with a one bp overlap of a 9 bp consensus sequence were not mutually exclusive and were simultaneously bound by several transcription factors. Even a 3 bp overlap of a 9bp consensus motif permitted some co-occurring binding, although here the steric factor strongly dominated. Similar trends were observed for weak binding site clusters with 1 and 3 bp overlapping sequences, indicating that strong binding sites permit and support partial binding of a transcription factor, allowing full occupancy of overlapping binding sites.

The high levels of occupancy achieved by low-affinity binding site clusters as determined by our *in vitro* characterization suggested that low-affinity binding site clusters may indeed be

functional *in vivo*. We therefore first tested low-affinity binding site clusters in an engineered synthetic system *in vivo*. We placed low-affinity Zif268 binding site clusters in the minimal CYC1 promoter and used a β-estradiol inducible Zif268 transcription factor (Z_3EV) to test whether these promoters could give rise to expression. We found that binding site clusters of 3 to 6 weak binding sites with individual affinities one order of magnitude below the consensus sequence were able to activate transcription to similar levels as a promoter containing 3 consensus sequence binding sites. Therefore, in this synthetic system, small clusters of low-affinity binding sites are functional. We also observed a concurrent increase in promoter leakiness which we characterized and attributed to non-specific interactions of an endogenous transcription factor binding to these binding site clusters. Although clusters of low-affinity binding sites were functional in the synthetic Z_3EV-minCYC1 system, it was possible that this was due to high, super-physiological expression levels of the Z_3EV transcription factor. To control for this uncertainty we tested whether low-affinity binding site clusters could substitute single high-affinity binding sites in a native gene regulatory network. We substituted native binding sites in the nucleosome-free and nucleosome-occluded regions of the well-characterized PHO5 promoter, which is regulated by endogenous levels of the master regulator Pho4. In this context low-affinity binding site clusters were also able to functionally replace single high-affinity binding sites, recovering a large proportion of the native PHO5 gene expression level. This was particularly impressive considering that substitution of the native Pho4 binding sites in these two locations also ablated the neighboring Pho2 binding sites, which are known to further strengthen the effect of Pho4 binding to these single target sites and contribute to transcriptional activation of this promoter.

These findings have several important implications on our current understanding of transcriptional regulation, computational methods for predicting gene regulatory function, gene regulatory network evolution, and engineering of synthetic gene regulatory networks. We have shown that low-affinity binding site clusters are effective at activating transcription *in vivo* suggesting that for computational approaches to gene regulatory function prediction and transcription factor binding prediction based on ChIP-seq data, that low-affinity binding sites warrant closer consideration, and that clusters of low-affinity binding sites can be highly functional.

On a more fundamental level, it was not known whether transcription factor dwell time or occupancy is the critical parameter that determines whether a binding region can give rise to functional gene expression. A single high-affinity binding site and clusters of low-affinity binding sites can both give rise to high occupancies at physiologically relevant transcription factor concentrations. But how this occupancy is achieved is fundamentally different between these two cases, with a single high-affinity binding site achieving high occupancy due to long transcription factor dwell times (low off-rates), whereas low-affinity binding site clusters achieve high occupancy as an ensemble average of many interactions of short dwell times (high off-rates). Our findings show that both systems can give rise to transcriptional activation and suggests that long transcription factor dwell times are not required for transcriptional

activation. This may not apply to transcriptional repression where it is possible that long dwell times are critical for efficient repression. It will be interesting to assess whether low-affinity binding site clusters can function as effective repressive regulatory regions or not. Finally, whether there is an observable difference in regards to gene expression noise between low-affinity binding site clusters and single high affinity binding sites would also warrant further exploration (289).

From an evolutionary perspective, low-affinity binding site clusters are of considerable interest as they could serve as a functional intermediate between a non-regulated promoter and a highly evolved promoter containing one or more consensus sequence binding sites. It is interesting that clusters of low-affinity binding sites are functional, but that they also seem to come with an additional cost which is non-specific regulation or cross-talk. Small clusters of low-affinity binding sites thus might be easier to form from a non-specific sequence background, but are then under continued pressure to further evolve to more specific high-affinity binding sites. It is striking that as few as 3 low-affinity binding sites are sufficient to generate a highly functional regulatory region. Likewise, when approaching this notion from a transcription factor centric perspective, there exist many more possible low-affinity binding sites for any given transcription factor than specific sites, which also enables regulatory regions to evolve more readily but also is the reason for non-specific interactions and cross-talk. Transcriptional regulation has traditionally been considered as being encoded in strong, well-defined binding sites. Results from de Boer et. al. supported a more finer-grained view of transcription by demonstrating that gene expression in promoters of *S. cerevisiae* is largely driven through the collective action of many weak regulatory contributions (244; 290). Holding total occupancy constant, many weaker sites may allow for a greater number of binding sites and different transcription factors (290), increasing the potential of network connectedness, or the average degree of nodes found in a eukaryotic gene regulatory network.

Synthetic promoters to date have predominantly made use of consensus binding sites, high-affinity binding sites, clusters thereof (190), and most recently clusters of lower affinity binding sites in combination with transcription factor cooperativity (291). As engineering synthetic promoters and entire gene regulatory networks matures it is likely that increasingly precise gene expression levels will become necessary. This to some extent may be achievable with relatively few single binding sites of various affinity, but it may be worthwhile considering the use of low-affinity binding site clusters. Low-affinity binding site clusters may be more resilient to mutations for example. First, a single deleterious mutation in a cluster of several low-affinity binding sites will have a much smaller impact on the resulting gene expression level as a mutation in a single high affinity binding site. Second, a single base change in a single high-affinity binding site will result in a large decrease in binding affinity, whereas a single base change in a low-affinity site will lead to a much smaller relative decrease in affinity, or may even be silent in regards to affinity.

In summary we conducted a systematic characterization of low-affinity binding site clusters

by conducting a quantitative biochemical analysis *in vitro* as well as functional studies *in vivo*. This work provided insights into how transcription factors bind to such clusters and established that a small number of low-affinity binding sites in a local cluster can be highly functional *in vivo*. These insights improve our current understanding of gene regulatory networks, and our ability to engineer sophisticated gene regulatory networks *de novo*.

5.5   Supplemental Information

Supplemental Figures
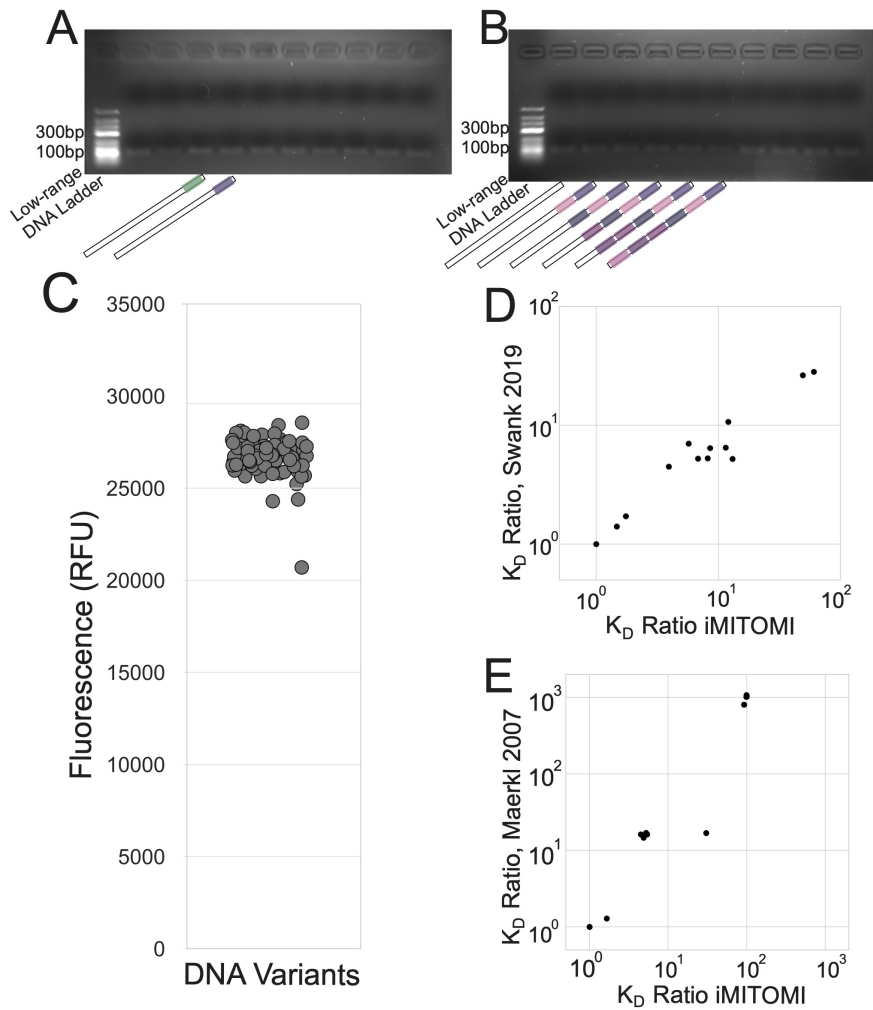
Figure 5.8: **(A, B)** Agarose gels for iMITOMI 90 bp PCR products (see Section 5.6.1) corresponding to the multi-weak Zif268 DNA target library characterized in (Figure 5.2C). **(C)** Fluorescence measurements of concentration-equalized iMITOMI DNA targets in a spotting plate before spotting. **(D, E)** Similar affinities and specificities were obtained when single site DNA targets were characterized on iMITOMI for binding to Zif268, and Pho4, as compared to (267) and (45), respectively. The K_D Ratio represents the K_D divided by the K_D of the strongest binding site analyzed in this study.
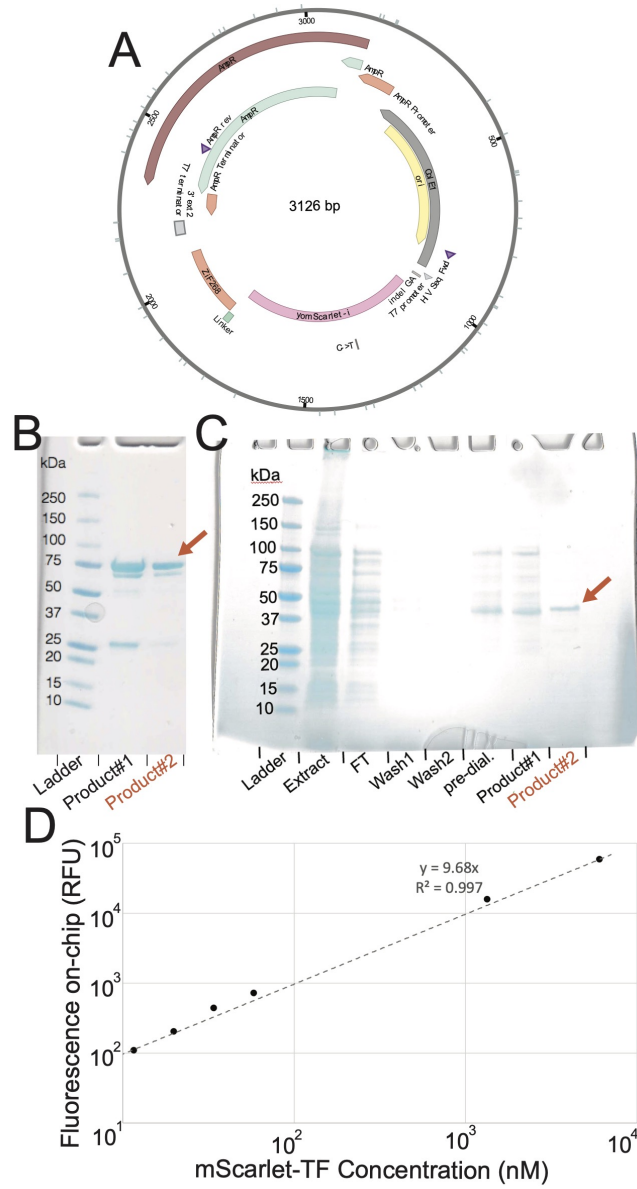
Figure 5.9: Characterization and calibration for purified proteins.
Caption on the next page.

Figure 5.9: **(A)** Plasmid for 5'-mScarlet-Zif268-6xHis-3' construct under a T7 promoter. A similar construct was used for the Pho4 transcription factor. **(B, C)** Proteins were his-tag purified and ran on a denaturing PAGE gel for Pho4 **(B)** and Zif268 **(C)**. **(B)** Compared to Product #1, Product #2 was additionally buffer exchanged in an Amicon spin-column. **(C)** Product #2 is from an entirely different purification batch. In both cases the higher purity Product #2 was used for all iMITOMI experiments in this study. **(D)** Calibration curve relating fluorescence measurements on-chip to mScarlet-transcription factor concentration quantitated by a Bradford assay.
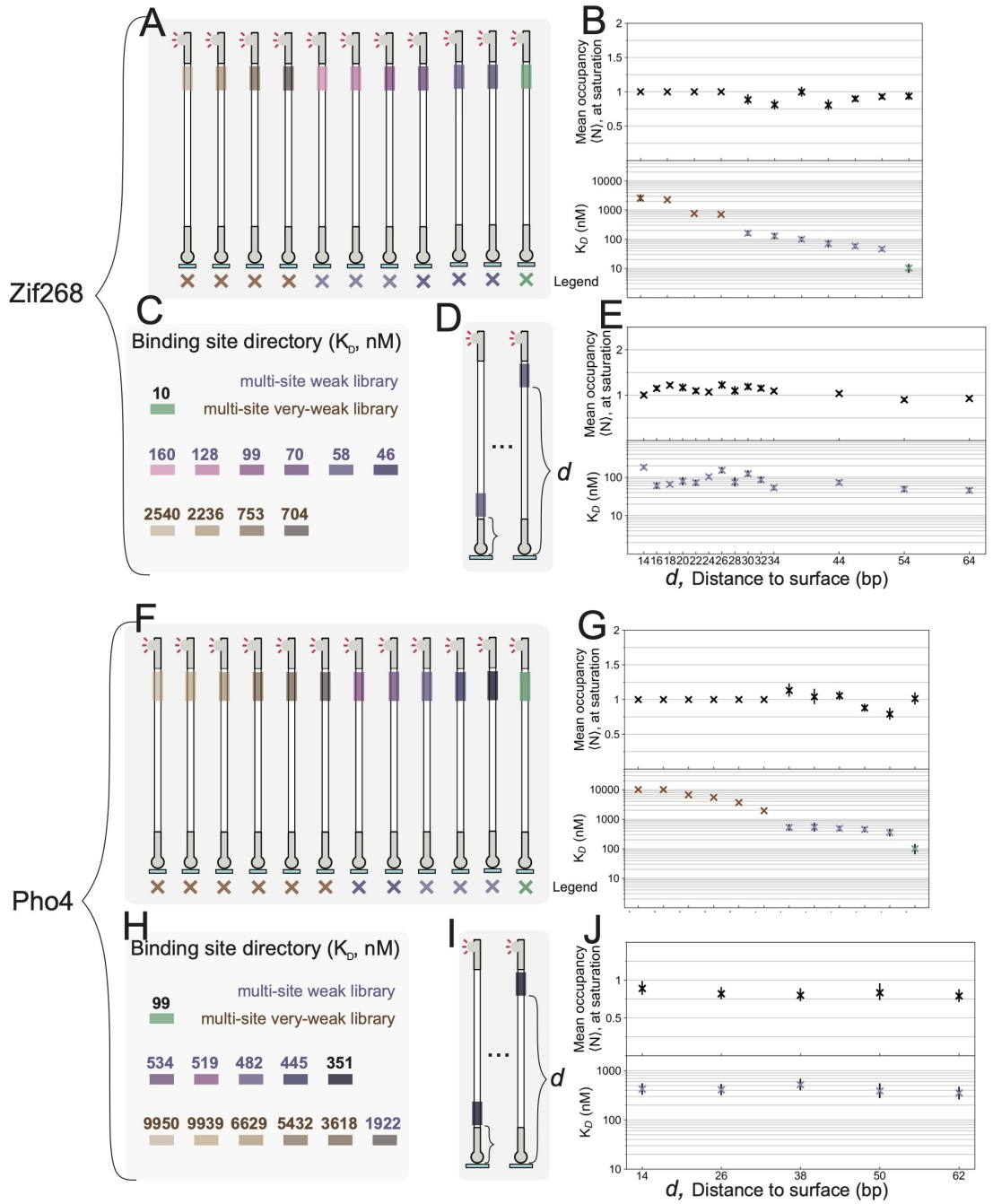
Figure 5.10: Binding site directories and single site positions scans.
Caption on the next page.

Figure 5.10: **(A)** Single site DNA targets for binding sites used to build Zif268 clusters throughout the study. **(B)** Mean occupancies at saturation (upper) and K_D values (lower) for targets in **(A)**, in the same order from left to right. For single weak binding sites, the $\langle N \rangle_{\max}$ parameter was fixed. **(C)** Color-coded Zif268 binding site directory with affinity values in nanomolar. **(D)** DNA targets corresponding to the position scan in **(E)**, where a Zif268 weak binding site's distance (d) from the chip surface was varied across targets in the library. **(E)** Characterization data for the Zif268 single site position scan library. **(F)** Single site DNA targets for binding sites used to build Pho4 clusters throughout the study. **(G)** Mean occupancies at saturation and K_D values for targets in **(F)**. **(H)** Color-coded Pho4 binding site directory with affinity values in nanomolar. **(I)** DNA targets corresponding to a Pho4 binding site position scan. **(J)** Characterization data for the Pho4 position scan library. Similarly to **(E)**, the distance to the chip's surface did not significantly impact binding.
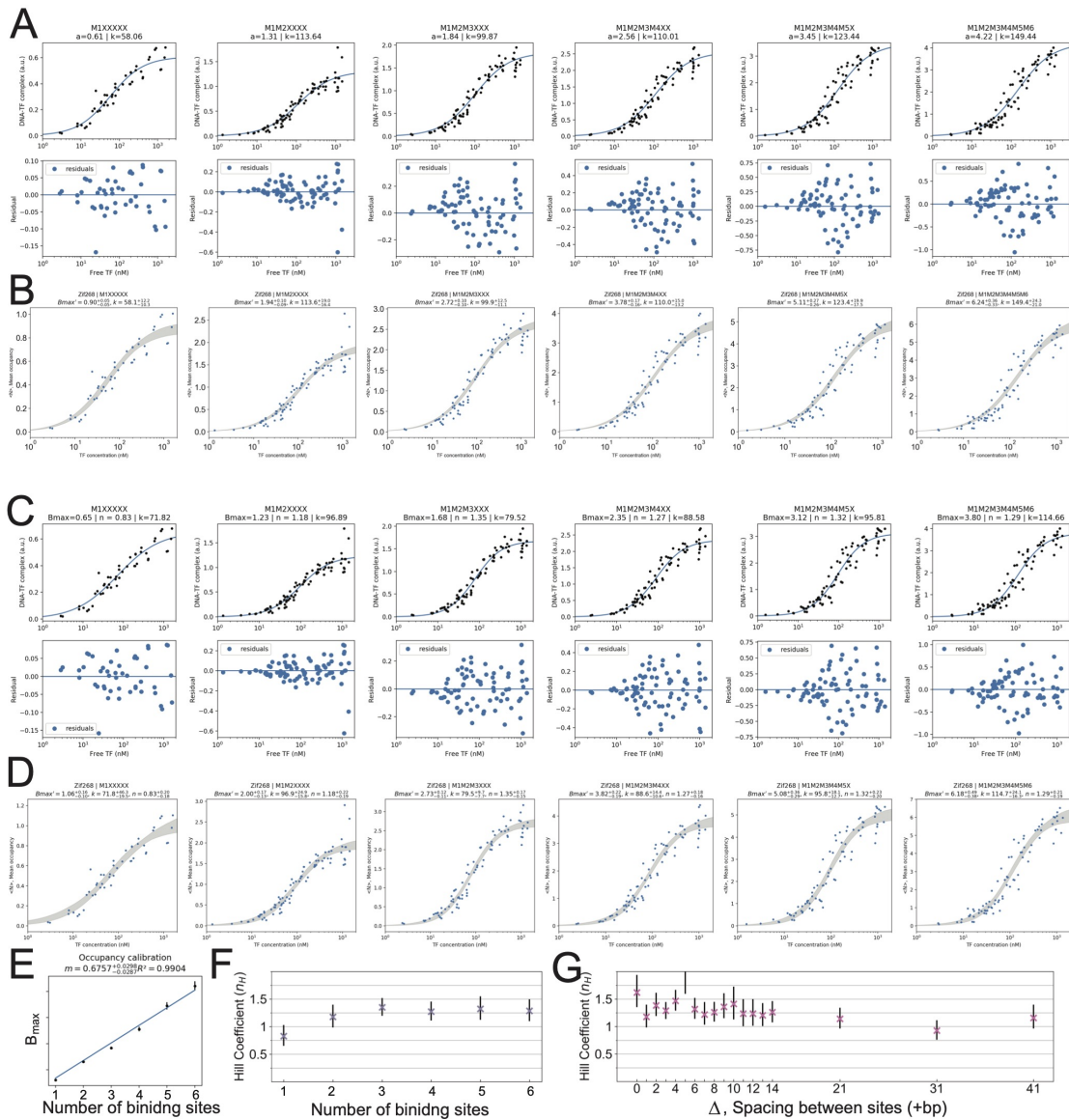
Figure 5.11: Data and model fits (SBC, Hill) for the Zif268 multi-weak library in Figure 5.2.

Caption on the next page.

Figure 5.11: **(A)** Data was fit with a 2-parameter saturation binding curve (upper) (Equation 5.7). Residuals for the fit (lower). **(B)** Parameter space was explored using Markov chain Monte Carlo (MCMC) to obtain a 5 to 95% confidence interval range. Data is normalized by the step increase in B_{max} (Equation 5.9). **(C, D)** Similar plots to **(A, B)** instead using a Hill model (5.8). **(E)** Step increase in B_{max} with additional binding sites for the saturation binding model fits in **(A)**. **(F, G)** Hill coefficients (n_H) for the multi-site weak library in **(C, D)** and for the two-site gap-scan library targets (Figure 5.3) fit with a Hill model.
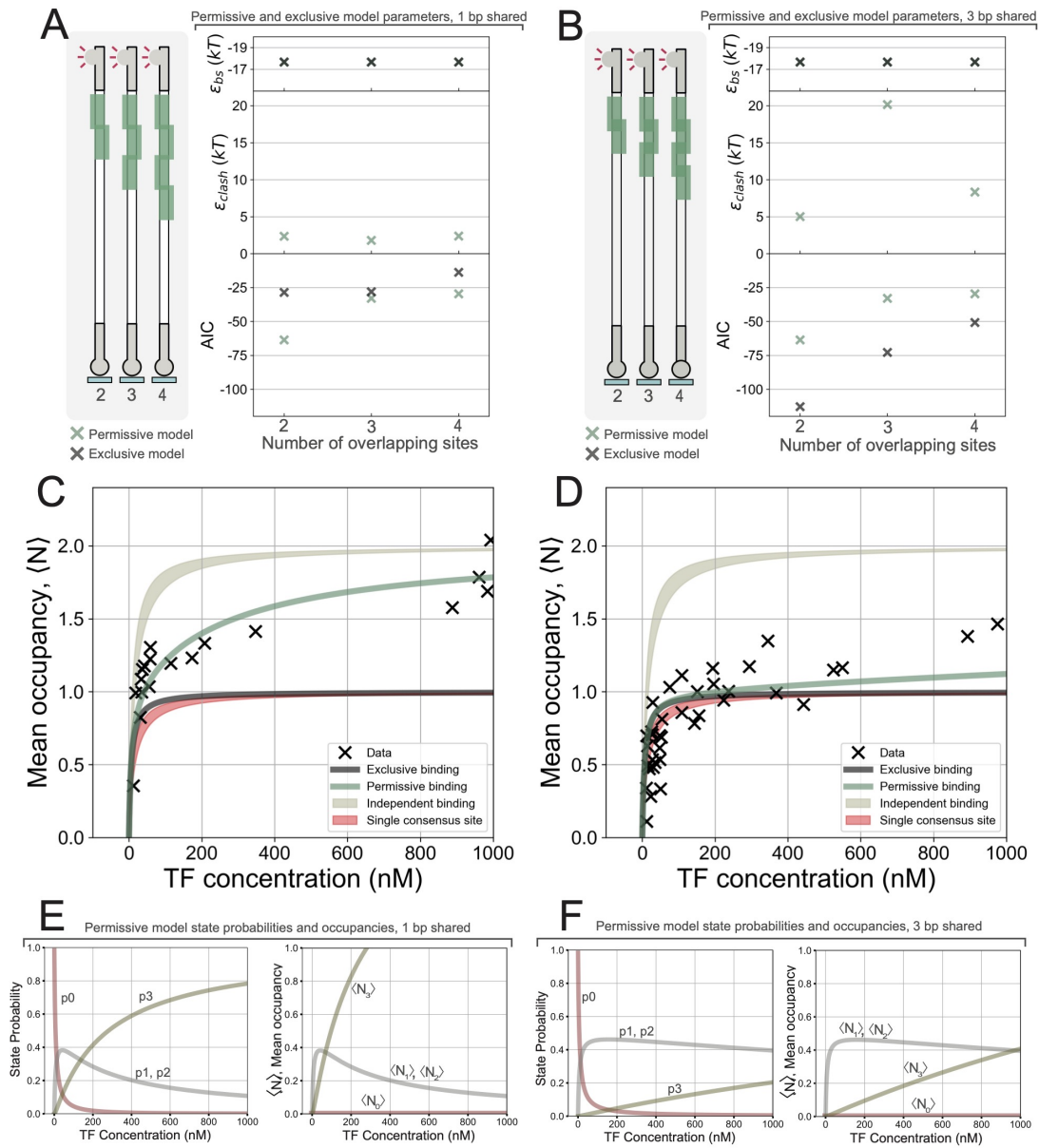
Figure 5.12: Modeling of binding to high-density clusters.
Caption on the next page.

Figure 5.12: **(A)** DNA targets for high-density clusters with 1 shared basepair betweeen neighboring consensus binding sites (left). Corresponding parameters and Akaike information criteria (AIC) for exclusive and permissive models of binding (right). Energies are reported in units of kT (product of Boltzmann constant and temperature). **(B)** DNA targets (left) and model parameters (right) for high-density clusters with 3 shared basepairs between neighboring consensus binding sites. **(C)** Characterization data for the DNA target in **(A)** with two consensus binding sites sharing 1 basepair. Exclusive and permissive model predictions in grey and green respectively. 95% confidence intervals for the characterized single consensus binding site (red), and a model of independent binding to two consensus binding sites (beige). **(D)** Corresponding data and models for the DNA target in **(B)** with two consensus binding sites sharing 3 basepairs. **(E, F)** State probabilities and occupanices for the permissive model, for the 2 consensus site DNA targets with 1 shared basepair **(E)** and with 3 shared basepairs **(F)**.
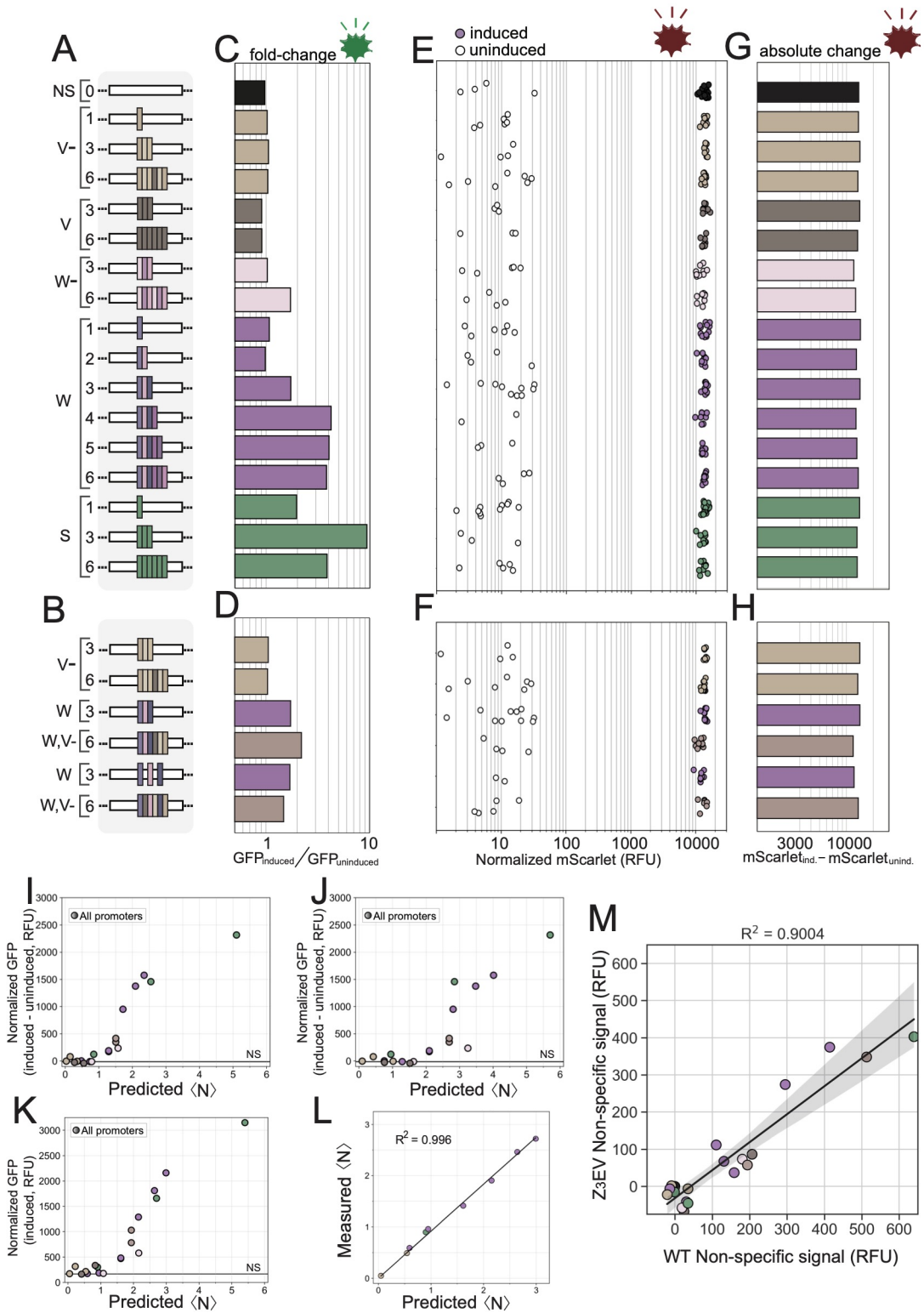
Figure 5.13: Alternate analyses and control data for the Zif268 *in vivo* results in Figures 5.5 and 5.6.

Caption on the next page.

Figure 5.13: **(A, B)** Legends for the characterized minCYC1 promoter libraries. **(C, D)** Alternate data analysis to Figure 5.5G and 5.5H, quantifying the Zif268-specific impact on gene expression as GFP mean fold-change between induced and uninduced strains. **(E, F)** Induced and uninduced mScarlet expression levels driven by the control promoter. **(G, H)** Absolute difference between induced and uninduced mScarlet expression levels. **(I, J)** Alternate data analysis to Figure 5.6, varying the transcription factor concentration from 73 nM (85% consensus saturation) **(I)** to 245 nM (95% consensus saturation) **(J)**. **(K)** Changing the measure of specific gene expression from the difference between mean induced and mean uninduced GFP signal, to mean induced GFP signal. **(L)** Relationship between predicted mean occupancy and measured mean occupancy, for the *in vitro*-characterized DNA targets that were used in the Zif268 *in vivo* study. Predictions were based on an independent statistical thermodynamic binding model similar to Equation 5.24, but with a variable number of binding sites, parameterized by data collected through characterization of single-site DNA targets, as explained in equation 5.6.6. R^2 of 0.996. **(M)** Relationship between the non-specific signal for the Z_3EV strains containing the Z_3EV transcription factor, and the WT (BY4741) strain containing no transcription factors with a Zif268 DNA binding domain. In each case the non-specific signal was quantified as the absolute mean difference between a given promoter's GFP expression and the GFP expression of the nonspecific (NS) promoter containing no binding sites (similar to Figure 5.5I and 5.5J), both in the respective strain's background. R^2 of 0.9, suggesting that the expression in uninduced strains is due to an endogenous transcription factor.
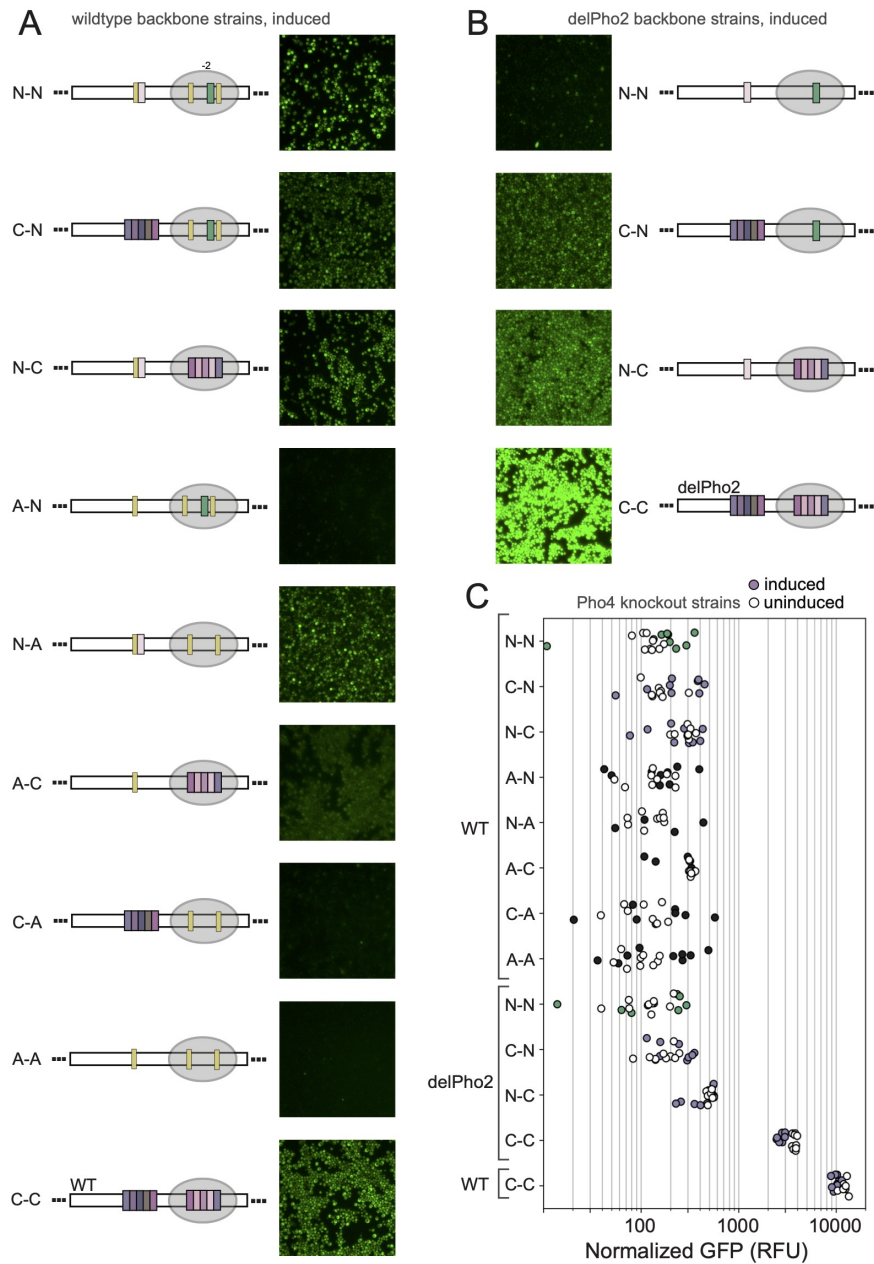
Figure 5.14: Fluorescence microscope images for the BY4741 strain library, and induced and uninduced platereader measurements for the Pho4 knockout strains. Caption on the next page.

Figure 5.14: **(A)** Fluorescence microscope images of the BY4741 strains with wildtype PHO5 promoter backbones induced by phosphate starvation. **(B)** Fluorescence microscope images of the BY4741 strains with delPho2 (all Pho2 sites ablated) PHO5 promoter backbones induced by phosphate starvation. **(C)** Platereader measurements following induction or in the absence of induction, for Pho4 knockout strains, enlarged from Figure 5.7L. Platereader data was processed similarly to the other platereader measurements in Figure 5.7.

5.6 Materials and methods

5.6.1 Purified protein and iMITOMI DNA target production

DNA target PCRs

Initial materials for producing iMITOMI DNA target libraries included tagged primers (5' Cy5-tagged: CTG/iCy5/TCGGCCGCTAACA, and 3' biotin-tagged: /5Biosg/GTCATACCGCCGGA) ordered from IDT. Furthermore, DNA targets of interest were ordered as single-stranded 90-basepair long primers from IDT, which had 5' and 3' ends complementary to the tagged primers. For each library member, two 50uL PCR reactions were prepared according to the following recipe:

Table 5.1: iMITOMI targets' PCR recipe

Reagent	Initial conc.	Volume to add (per 50 μL rxn)
Biotin-tagged primer	100 μM	0.25 μL
Cy5-tagged primer	100 μM	0.25 μL
90bp ssDNA target primer	500 nM	0.25 μL
DreamTaq Green PCR Master Mix	2X	25 μL
Invitrogen UltraPure Distilled Water		24.25 μL

Table 5.2: iMITOMI targets' PCR cycling conditions

Step	Temperature	Duration
1.	95°C	1 minute
2.	95 °C	30 seconds
3.	50°C	30 seconds
4.	72°C	1 minute
Steps 2 to 4 are cycled a total of 30 times		
5.	72°C	10 minutes
6.	4°C	∞

Two 50 μL PCR products were mixed with 500 μL of DNA binding buffer, combined into one purification column (Zymo DNA Clean & Concentrator 50), and PCR purified according to kit instructions. Samples were eluted into 80 μL of Invitrogen UltraPure Distilled Water. DNA targets were run on an agarose gel for characterization, and only products with a single band of expected size were advanced to spotting.

Spotting plate preparation

75 μL of each DNA target was introduced into independent wells of a 384 well plate. Plates were measured for fluorescence with an excitation wavelength of 570±9 nM and an emission wavelength of 593±9 nM, using a Biotek Synergy Mx Multi-mode reader. Based on these measurements, new fluorescence-equalized wells were prepared by diluting an appropriate amount of DNA with water to reach 72 μL (no water was added for the well with the lowest fluorescence intensity, the others were equalized to this fluorescence intensity ± 5%, Supp. Figure 5.8C. 24 μL of 2% Bovine Serum Albumin (in H_2O) was added to each well. The plate was sealed with aluminum and kept at -20°C until time of use.

His-tag protein purification

For each transcription factor, bacteria containing the corresponding plasmid were overnight cultured with ampicilin (Amp), and innoculated into two 1 L baffled flasks each containing 500 mL LB and Amp (100 ug/ml). Cultures were grown at 37°C at 260 rpm for 2 hours, induced with IPTG (1 mM) and grown for an additional 3 hours. Cells were centrifuged at 3220 g at 4°C for 10 min in 50 mL falcon flasks. Cell pellets were resuspended in buffer A (sequential transfers, using total volume of 7.5 mL), and sonicated (Vibra cell 75186 sonicator, probe tip diameter: 6 mm) with 20 s ON 20 s OFF for four cycles using 70% amplitude. Lysate was distributed to 2 mL tubes and centrifuged for 20 minutes at 4°C at 20,000 g. Supernatant was removed and dispensed into a regenerated Ni Sulfate column. The column was washed with 25 mL buffer A, then protein was eluted in 5 mL buffer B. The sample was dialysed overnight in 1 L of HT buffer at 4°C with magnetic stirring, followed by 3 hours of dialysis with HT stock buffer under similar conditions. The sample was run on an SDS-PAGE gel in 5.9 for characterization.

Purification buffer recipes

Table 5.3: Purification buffer recipes

Component	Final conc.
Buffer A	
NH4Cl	1 M
HEPES	50 mM
MgCl	10 mM
B-ME (fresh)	7 mM
Buffer B	
Imidazole	0.5 M
HEPES	50 mM
MgCl	10 mM
KCl	100 mM

B-ME (fresh)	7 mM
HT buffer	
HEPES	50 mM
MgCl	10 mM
KCl	100 mM
B-ME (fresh)	7 mM
HT stock buffer	
HEPES	50 mM
MgCl	10 mM
KCl	100 mM
Glycerol	60%
B-ME (fresh)	7 mM

5.6.2 Control and flow wafer fabrication

Mold-fabrication using silicon wafers was performed through a photolithography process in the CMi cleanrooms at EPFL, as described previously (292; 45). In brief, for the control layer, a silicon wafer was primed in a Tepla 300, and SU-8 photoresist (GM 1070) was spin coated onto the wafer using a Sawatec spin coater to reach a height of 30 μm. After a soft bake the wafer was exposed (365 nm illumination, 20 mW/cm^2 light intensity) using a chrome mask for 10 s on a MABA6 mask aligner. The wafer was post exposure baked and then developed using propylene glycol methyl ether acetate (PGMEA), followed by a hard bake. For the flow layer the silicon wafer was treated with hexamethyldisilizane (HMDS) vapor using a YesIII oven. AZ 9260 photoresist was spin coated onto the wafer using an EVG150 modular cluster tool to reach a height of around 14 μm. After the wafer was baked it was left for a one hour relaxation period, followed by UV exposure using an MABA6 mask aligner. The total dose was 660 mJ/cm split into two exposures of 18s with a 10s wait period in between (20 mW/cm^2 light intensity). The wafer was developed again using the EVG150 (AZ 400K developer), and then baked at 160 °C for two hours. Designs of microfluidic devices are available on our group's website: http://lbnc.epfl.ch.

5.6.3 PDMS chip fabrication

Wafers were first treated with chlorotrimethylsilane (TMCS) vapor to facilitate removal of cured polydimethylsiloxane (PDMS) from the wafers. PDMS curing agent and elastomer (amounts of 1 g and 20 g for the flow layer, 10 g and 50 g for the control layer, respectively) were mixed (2000 rpm for 1 min) and defoamed (2200 rpm for 2 min) using a Thinky ARE-250

centrifugal mixer. PDMS from the control layer was poured on the control wafer and left to degas in a vacuum chamber, followed by baking for 20 minutes. For the flow layer, 4 mL of PDMS was spin-coated onto the flow-wafer using a SCS G3P-8 spin coater (1800 rpm for 35s), followed by baking for 20 minutes. Control inlets were punched using a Schmidt Press manual hole puncher and 21-gauge (OD 0.04") pins (Technical Innovations, Inc.), and the control PDMS blocks were aligned to the flow layer under a stereoscope using coaxial illumination. Aligned chip layers were bonded at 80°C for 90 minutes. Next, PDMS chips were cut off the flow layer, and flow inlets and outlets were punched using a 900 mm pin.

5.6.4   Epoxy slide coating and microarraying

VWR slides with cut edges and plain ends (631-1550) were cleaned in a solution of 600 mL miliQ water, 120 mL 25% Ammonia solution, and 150 mL H_2O_2 (on a hotplate at 80°C). Slides were then washed with miliQ water and dried. Next, slides were left to incubate for 20 minutes in a bath of 891 mL toluene and 5 mL of (3-Glycidyloxypropyl)trimethoxysilane (GPS, Sigmaaldrich Cat. 440167). Slides were rinsed with toluene and dried, followed by baking at 120°C for 120 minutes. Epoxy-coated glass slides were spotted with the purified 90 bp DNA targets using a Genetix Qarray[2] robot with an MP2.5 pin (Arrayit). Spotting chambers were aligned to DNA spots on the glass slide under a microscope, and the device was baked for 8 hours at 80°C.

5.6.5   iMITOMI experiments

Control lines were filled with tap water and pressurized at 172 kPa. The the neck valve was pressurized to prevent premature target resolubilization. Flow lines were pressurized at 48 kPa. Surface chemistry was conducted in the remainder of the chip (45), by sequentially patterning the surface using 15 minute flow steps with biotin-BSA, then neutravidin, followed by actuation of the button valve and a second biotin-BSA flow. PBS washes were conducted between each flow step for 5 minutes. This resulted in available neutravidin binding sites under the button valve in the detection chamber, whereas the remainder of neutravidin binding sites in the chip's flow channels were passivated by biotin-BSA. DNA targets in spotting chambers were resolubilized by pressurizing the PBS flow line, closing the outlet and opening the neck valve to force air out of the PDMS and PBS into the spotting chambers. After a PBS wash with the neck valve closed to prevent any contamination, the sandwich valves were pressurized to isolate chambers from one another, and the neck and button valves were opened to allow the biotinylated Cy5-tagged DNA targets to diffuse to and bind under the button valves in the detection chambers for 90 minutes. The button valve was then closed, and the chip was washed with PBS. The chip surface was blocked using Promega wheat-germ extract to reduce non-specific binding of concentrated fluorescent protein solutions. The chip was multiplexed into 8 sections using the multiplexing valves. Different concentrations of purified mScarlet-tagged transcription factor were flowed into different sections of the chip, followed by incubation for 60 minutes with the button valves open to allow binding to reach equilibrium.

The chip was scanned to quantify the free protein concentration at equilibrium (compared against a background scan after the subsequent wash step). The button valve was closed to isolate the bound protein at equilibrium, and the chip was washed, and scanned again to quantify the bound protein and amount of DNA present (compared against a background scan before fluorescent protein or fluorescent DNA was introduced into the detection chamber, respectively).

5.6.6 iMITOMI data analysis and modeling

Image analysis and data processing

Images from a given scan were stitched together using ImageJ and ROIs were processed using GenePix, which facilitated automatic and manual feature alignment, as well as conversion of ROI pixel intensities to mean signal intensities.

Quantification of free and bound protein signals

A normalized bound signal intensity was quantified as:

$$S_{bound} = \frac{TF_{bound} - TF_{background}}{DNA_{immobilized} - DNA_{background}} \tag{5.1}$$

S_{bound} corresponds to the bound transcription factor at equilibrium (background corrected) normalized by the amount of DNA at equilibrium (background corrected). The values used in this calculation were mean signal intensities from an ROI under the button valve, at the appropriate assay stage discussed in section 5.6.5. A signal S_{free} corresponding to free protein at equilibrium was quantified using an ROI next to the button valve as:

$$S_{free} = TF_{free} - TF_{background} \tag{5.2}$$

This signal was calibrated against the calibration curves from Figure 5.9 to obtain absolute TF concentrations, [TF].

Saturation binding curve parameter estimation

According to mass action kinetics, at equilibrium a 1:1 transcription factor (TF) to binding site (BS) binding reaction follows the relationship
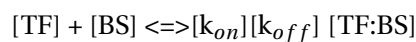
$$[TF] + [BS] <=>[k_{on}][k_{off}] \ [TF{:}BS]$$

$$k_{on} \cdot [TF][BS] = k_{off} \cdot [\text{TF:BS}]$$

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[TF][BS]}{[\text{TF:BS}]}$$

$$[\text{TF:BS}] = \frac{[TF][BS]}{[K_D]} \tag{5.3}$$

Furthermore, the proportion of binding sites bound by transcription factor molecules, P_{bound}, can be defined as follows

$$P_{bound} = \frac{[\text{TF:BS}]}{[\text{TF:BS}] + [BS]} \tag{5.4}$$

Through substituting Equation 5.3 into Equation 5.4 and rearranging, it follows that

$$P_{bound} = \frac{[TF]}{K_D + [TF]} \tag{5.5}$$

This derives a relationship to model the proportion of binding sites bound by transcription factor, based on the concentration of free transcription factor and the binding reaction's dissociation constant (K_D, representing the affinity).

The same functional form as in Equation 5.5 can be used similarly to describe a binding reaction to DNA targets containing multiple binding sites, if all binding site affinities are assumed equivalent, and binding is assumed to be independent. To see this, it is useful to consider that [BS] was used in the above analysis instead of [DNA], and the analysis is not sensitive to if some constant number of [BS] are distributed on fewer molecules of DNA. The B_{max} parameter represents the fluorescence intensity at transcription factor concentrations for a given DNA target, and using it to scale P_{bound} allows for a rescaling from proportion of binding sites bound to S_{bound}, that is

$$S_{bound} = P_{bound} \cdot B_{max} \tag{5.6}$$

Therefore, experimental data characterizing saturation binding to a given DNA target can be fit according to the two parameter (K_D, B_{max}) model in equation 5.7

$$S_{bound} = \frac{B_{max} \cdot [TF]}{K_D + [TF]} \tag{5.7}$$

where [TF] is in absolute concentration units (nM) to obtain a K_D estimate in nM. As discussed, B_{max} represents the fluorescence level at saturating transcription factor concentrations.

In order to model cooperative binding, a common phenomenological modification to Equation 5.7 is the incorporation of the Hill coefficient (n_H), even for 1:1 binding reactions, which produces the Hill model:

$$S_{bound} = \frac{B_{max} \cdot [TF]^{n_H}}{K_D + [TF]^{n_H}} \qquad (5.8)$$

For a given DNA target, data were combined across two to six experimental days and an initial fit to Equation 5.7 was obtained through non-linear least squares minimization with the Lmfit package in Python (293). To explore the parameter space around this fit initial parameter set and obtain confidence intervals, we used Markov Chain Monte Carlo to conduct sampling of the posterior probability distribution given the experimental data and assuming a uniform prior (Figure 5.11B and C).

To obtain mean occupancies $\langle N \rangle$, S_{bound} (units RFUs) was normalized (calibrated) according to the B_{max} increase from adding additional binding sites (or equivalently, the RFUs per 1 TF molecule bound to DNA on average), as described in the results section of the main text. This converts equation 5.7 to the more interpretable form in equation 5.9

$$\langle N \rangle = \frac{n_{sites} \cdot [TF]}{K_D + [TF]} \qquad (5.9)$$

where the mean occupancy $\langle N \rangle$ represents the average number of transcription factor molecules bound per DNA molecule, and n_{sites} represents the number of defined binding sites on the DNA target. Upon normalization, B_{max} on the right hand side is converted to n_{sites} since the B_{max} of a given target can be thought of as the B_{max} from each binding site, multiplied by the number of binding sites.

Statistical mechanical modeling

Statistical mechanical models of binding were built starting from the useful framework developed by Philips et. al. and presented in the Physical Biology of the Cell textbook (294). We recommend referring to the textbook for a detailed discussion of the modeling framework. Essentially, microstates of the binding system were grouped into states, where each state represents all of the possible microstates of the system where a particular set of DNA binding sites are bound by transcription factor molecules (e.g. the unbound state represents all microstates of the system where the DNA is not bound by any transcription factor molecule). Here a microstate signifies a particular distinguishable arrangement of indistinguishable transcription factor molecules among positions of a spatially discretized environment (e.g. the unbound state is composed of many microstates, which differ from one another based on which positions in the lattice-modeled solution are occupied). Each state can be attributed a total energy, by accounting for the energies from all particles in one of its microstates. This

total energy does not differ between the different microstates of a given state. Furthermore, each state is attributed a multiplicity, by counting the number of different microstates that can give rise to the state.

In order to calculate the probability of any state, first recognize that

$$p_i \propto e^{-E_i/(kT)}$$

$$\beta = \frac{1}{kT}$$

$$p_i \propto e^{-\beta E_i} \tag{5.10}$$

where p_i is the probability of microstate i, E_i is the total energy from all particles in microstate i, and the constant kT is a product of the Boltzmann constant k and the absolute temperature T and usually written as β. The term $e^{-\beta E_i}$ is the Boltzmann factor. Equation 5.11 relates the ratio of probabilities for two different microstates to the ratio of their Boltzmann factor's, which depends on the difference between their total energies. This captures the typical intuition that lower-energy configurations are more favorable.

$$\frac{p_i}{p_j} = e^{\beta(E_j - E_i)} \tag{5.11}$$

To determine the probability for a given microstate (Equation 5.12), its Boltzmann factor must be normalized by the sum of the Boltzmann factors from every possible microstate (the partition function, Z, Equation 5.13), such that the microstate probabilities sum to 1.

$$p_i = \frac{e^{-\beta E_i}}{Z} \tag{5.12}$$

$$Z = \Sigma_i \; e^{-\beta E_i} \tag{5.13}$$

In order to calculate the probability of any given state (group of microstates), it follows that the state's multiplicity can be multiplied by the state's Boltzmann factor. This produces what is effectively the sum over Boltzmann factors from all microstates belonging to the state, a term referred to as the state's weighted multiplicity (294), which accordingly when normalized by the partition function produces the state's probability.

For illustration, it is useful to first consider how to count the number of microstates (how to

calculate the multiplicity) for the unbound state

$$\frac{\Omega!}{L!(\Omega - L)!} \tag{5.14}$$

Where Ω represents the number of discrete positions where a transcription factor could be located in solution, and L represents the number of transcription factor molecules. Transcription factor molecules can be placed in solution in $\frac{\Omega!}{(\Omega - L)!}$ different configurations, but they can be rearranged indistinguishably $L!$ ways.

Furthermore, the following related approximation will enable us to simplify our analysis

$$\frac{x!}{y!(x - y)!} \approx \frac{x^y}{y!}, \quad \text{if } x \gg y \tag{5.15}$$

More generally, state multiplicites evaluate to

$$\frac{\Omega!}{(L - n)! \ (\Omega - (L - n))!} \tag{5.16}$$

Where n represents the number of TF molecules bound to the DNA molecule

Which according to Approximation 5.15 simplifies to

$$\frac{\Omega^{L-n}}{(L - n)!} \tag{5.17}$$

Since we assume that in our binding system $\Omega \gg L$

Table 5.4 and Equations 5.18 to 5.25 demonstrate how to model a simple system of transcription factor molecules binding to DNA having a single binding site, in order to illustrate how the model in Equation 5.5 that we derived from mass action kinetics can be equivalently derived through statistical mechanics. This connection ultimately allowed us to coherently integrate both mass action kinetic and statistical mechanical modeling in our study.

Table 5.4: 1:1 binding through the lens of statistical mechanics

State	Energy, E_i	Multiplicity	Relative weighted multiplicity
—	$L \cdot \varepsilon_{sol}$	$\frac{\Omega!}{L!(\Omega - L)!} \approx \frac{\Omega^L}{L!}$	1
$\underline{O}$	$\varepsilon_{bs} + (L - 1) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-1)!(\Omega - (L-1))!} \approx \frac{\Omega^{L-1}}{(L-1)!}$	$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs}}$

The total energy, E_i, of the unbound state (_) is signified by the sum of an ε_{sol} energy term from each of the L particles in solution. For the bound state ($\underline{O}$), one particle less will be at an

energy of ε_{sol}, and instead will have the energy ε_{bs}, representing its specific interaction to a single binding site on DNA. Multiplicities are weighted according to their Boltzmann factors as follows, for the unbound state the weighted multiplicity is

$$\frac{\Omega^L}{(L)!} \cdot e^{-\beta L \varepsilon_{sol}} \tag{5.18}$$

And for the bound state, the weighted multiplicity is

$$\frac{\Omega^{L-1}}{(L-1)!} \cdot e^{-\beta((L-1)\varepsilon_{sol}+\varepsilon_{bs})} \tag{5.19}$$

Furthermore, following the convention in (294), we normalize the weighted multiplicities and partition function by the weighted multiplicity of the unbound state, to represent these terms as relative weighted multiplicities (Table 5.4, Equation 5.20). For the bound state the relative weighted multiplicity simplifies to

$$\frac{\frac{\Omega^{L-1}}{(L-1)!} \cdot e^{-\beta((L-1)\varepsilon_{sol}+\varepsilon_{bs})}}{\frac{\Omega^L}{(L)!} \cdot e^{-\beta L \varepsilon_{sol}}}$$

$$= \frac{L}{\Omega} \cdot e^{-\beta(\varepsilon_{bs}-\varepsilon_{sol})} \tag{5.20}$$

Furthermore, as illustrated in (294),

$$[TF] = \frac{L}{\Omega \cdot V_{box}}$$

$$[TF]_o = \frac{1}{V_{box}}$$

$$\frac{[TF]}{[TF]_o} = \frac{L}{\Omega} \tag{5.21}$$

where [TF] is the concentration of transcription factor, and the constant $[TF]_o$ is a reference concentration that corresponds to if each position in solution was occupied by a transcription factor molecule (294). Furthermore, V_{box} is the volume of one discrete lattice position in solution. This allows the bound state's relative multiplicity to be expressed as

$$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs}}, \quad \text{where } \Delta \varepsilon_{bs} = \varepsilon_{bs} - \varepsilon_{sol} \tag{5.22}$$

Then, the probability of any state can be determined by normalizing by the partition function (expressed relative to the unbound state) Z, which is the sum of all relative weighted multiplicities

$$Z = 1 + \frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs}} \tag{5.23}$$

Therefore, the probability of the bound state is

$$p_{bound} = \frac{\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs}}}{1 + \frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs}}} \tag{5.24}$$

We used the single-site binding model in Equation 5.24, which is equivalent to the expression for mean occupancy for single-site targets, to obtain energy parameters for particular binding sites, by fitting the corresponding single-site DNA targets' binding data (normalized by the increase in B$_{max}$ with additional sites) to the model (using non-linear least-squares minimization, followed by MCMC, as before). These binding site energies were used to parameterize higher-order statistical mechanical models of binding to DNA targets containing multiple binding sites (e.g. for modeling exclusive and permissive binding to high-density clusters).

Equation 5.24 is the statistical mechanical equivalent to Equation 5.5, considering that

$$K_D = [TF]_o \cdot e^{\beta \Delta \varepsilon_{bs}} \tag{5.25}$$

As discussed, this connection permitted us to move between statistical mechanical and mass action modeling frameworks in our study.

Exclusive binding model to describe binding to high-density clusters

In exclusive binding models, transcription factors were not permitted to be bound simultaneously to overlapping binding sites (sites sharing common basepairs), as illustrated in the main manuscript Figure 5.4D and Table 5.5. For example, for a cluster with two binding sites, only three states (unbound, left site bound, or right site bound) were considered in the exclusive model. The states, energies, and multiplicites for an exclusive model of binding to a DNA target containing two binding sites are shown in Table 5.5.

In order to calculate the mean occupancy of transcription factors on DNA, we can calculate a probability-weighted average of the number of transcription factors bound in each state.

Table 5.5: Exclusive binding model for a high-density cluster with two overlapping sites

State	Energy, E_i	Multiplicity	Relative weighted multiplicity
—	$L \cdot \varepsilon_{sol}$	$\frac{\Omega!}{L!(\Omega - L)!} \approx \frac{\Omega^L}{L!}$	1
O_	$\varepsilon_{bs1} + (L-1) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-1)!(\Omega - (L-1))!} \approx \frac{\Omega^{L-1}}{(L-1)!}$	$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs1}}$
_O	$\varepsilon_{bs2} + (L-1) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-1)!(\Omega - (L-1))!} \approx \frac{\Omega^{L-1}}{(L-1)!}$	$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs2}}$

Therefore, for the example of exclusive binding to two overlapping sites (states presented in Table 5.5)

$$\langle N \rangle = 0 \times \frac{1}{Z} + 1 \times \frac{\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs1}}}{Z} + 1 \times \frac{\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs2}}}{Z}, \quad \text{where} \tag{5.26}$$

$$Z = 1 + \frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs1}} + \frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs2}} \tag{5.27}$$

Permissive binding model to describe binding to high-density clusters

In permissive binding models, transcription factors were permitted to be bound simultaneously to overlapping binding sites, as illustrated in the main manuscript Figure 5.4D and Table 5.6. For example, for a cluster with two overlapping binding sites, four states were permitted (unbound, left site bound, right site bound, or both sites bound).

Table 5.6: Permissive binding model for a high-density cluster with two overlapping sites

State	Energy, E_i	Multiplicity	Relative weighted multiplicity
—	$L \cdot \varepsilon_{sol}$	$\frac{\Omega!}{L!(\Omega - L)!} \approx \frac{\Omega^L}{L!}$	1
O_	$\varepsilon_{bs1} + (L-1) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-1)!(\Omega - (L-1))!} \approx \frac{\Omega^{L-1}}{(L-1)!}$	$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs1}}$
_O	$\varepsilon_{bs2} + (L-1) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-1)!(\Omega - (L-1))!} \approx \frac{\Omega^{L-1}}{(L-1)!}$	$\frac{[TF]}{[TF]_o} \cdot e^{-\beta \Delta \varepsilon_{bs2}}$
OO	$\varepsilon_{bs1} + \varepsilon_{bs2} + \varepsilon_{clash} + (L-2) \cdot \varepsilon_{sol}$	$\frac{\Omega!}{(L-2)!(\Omega - (L-2))!} \approx \frac{\Omega^{L-2}}{(L-2)!}$	$(\frac{[TF]}{[TF]_o})^2 \cdot e^{-\beta(\Delta \varepsilon_{bs1} + \Delta \varepsilon_{bs2} + \varepsilon_{clash})}$

For binding to larger clusters where two or more transcription factors can be bound at once (i.e. the state with both sites bound in the permissive model in Table 5.6), this must be taken into account in these states' multiplicities. For illustration, consider the multiplicity of a state

with two transcription factors bound at once. According to Equation 5.16 and Equation 5.17.

$$\frac{\Omega!}{(L-2)! \ (\Omega - (L-2))!} = \frac{\Omega^{L-2}}{(L-2)!} \tag{5.28}$$

Which when normalized by the multiplicity of the unbound state results in the relative multiplicity

$$\frac{\frac{\Omega^{L-2}}{(L-2)!}}{\frac{\Omega^L}{L!}} = \frac{L(L-1)}{\Omega^2} \tag{5.29}$$

The term on the RHS of Equation 5.29 can be re-written as

$$\frac{\frac{L!}{(L-2)!}}{\Omega^2} \tag{5.30}$$

Which according to Approximation 5.15

$$= \frac{L^2}{\Omega^2} \quad if \ L \gg 2 > 0 \tag{5.31}$$

More generally, a state with n transcription factors bound at once will have the relative weighted multiplicity

$$\left(\frac{L}{\Omega}\right)^n, \quad \text{if} \ L \gg n > 0$$

$$= \left(\frac{[TF]}{[TF]_o}\right)^n \tag{5.32}$$

Permissive model states possess an additional energy parameter ε_{clash} (Table5.6, bottom) for each set of shared basepairs (at the junction of two overlapping sites) where two transcription factors are bound simultaneously. For instance, for a cluster with three binding sites where the first site overlaps the second, and the second overlaps the third, the state with three transcription factors bound simultaneously would possess $2 \cdot \varepsilon_{clash}$ in its energy term. ε_{clash} can be considered as composed of both an energy associated with steric interference between transcription factor molecules, and some loss of binding to the shared basepairs, lumped into a single parameter.

Model selection

To compare models of binding to high-density clusters, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were computed using the Lmfit package (293). The

Chapter 5. Systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and *in vivo*

AIC and BIC can be computed as:

$$AIC = N \cdot ln(\frac{\chi^2}{N}) + 2N_{vars} \tag{5.33}$$

$$BIC = N \cdot ln(\frac{\chi^2}{N}) + ln(N) \cdot N_{vars} \tag{5.34}$$

$$\tag{5.35}$$

Where χ^2 is the chi-square, N is the number of data points, and N_{vars} is the number of floating parameters. For clusters where 1 basepair was shared between neighboring binding sites, the lower AIC and BIC suggest that the permissive model better represents binding than the exclusive model (Figure 5.12). On the other hand, as the number of shared basepairs increased to 33% of the motif (3bp), binding became largely exclusive, as reflected by the lower AIC and BIC for the exclusive model.

5.6.7 *In vivo* data analysis and modeling

Modeling independent binding to clusters using parameters derived from single site DNA targets

Our results identified that binding sites with a positive gap distance in lower-density clusters exhibit largely independent binding. Therefore to model binding to these clusters, accounting for the difference in affinities between different sites, we used an independent binding model parameterized by energies derived from characterization of the individual binding sites on single site DNA targets. We used a statistical mechanical model that accounts for the different states of the system, although a similar result could be obtained by summing the occupancy contribution from each individual binding site in isolation. To predict the mean occupancy of Zif268 clusters *in vivo*, we simulated a concentration range from 85% to 95% saturation of the consensus binding site (73 nM to 245 nM), confirming our results under a wide range of concentrations expected for the Z_3EV transcription factor. To check the accuracy of our model and premise of binding independence, for clusters that we characterized fully *in vitro* we compared our model predictions to measured occupancies (Figure 5.6M), which showed near perfect agreement and an R^2 value of 0.996. To quantify induced and uninduced levels of gene expression from platereader timeseries measurements, the signal in the appropriate fluorescence or OD600 channel from blank wells containing media alone was subtracted. Then the fluorescence timeseries measurements were normalized by their corresponding OD600 values. To represent induced and uninduced levels for each replicate, timeseries measurements were averaged across a time window before the strains reached stationary phase, and all data was plotted in strip plots using Seaborn and Matplotlib in Python. Target transcription factor-specific and non-specific signals were then quantified as explained in the

results section.

5.6.8 Yeast strain generation and characterization

Media and growth conditions

Cells were incubated for yeast transformation at 30°C shaking at 250 rpm in yeast extract peptone dextrose (YPD) medium (Sigma-Aldrich) (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose). Plates contains 20 g/L of agar (10752-36, Alfa). Promoter library yeast strains were cultured in synthetic complete (SC) medium lacking uracil (1.92 g of yeast synthetic dropout medium supplement without uracil, Y1501-20G, Sigma-Aldrich), 6.7 g Yeast Nitrogen Base (YNB) without amino acids (Y0626-1KG, Sigma-Aldrich), were dissolved in 960 mL of deionized water and autoclaved. After sterilization we added 40 mL 50% filter sterilized glucose solution to a final volume of 1 L. For plate reader experiments of PHO5 promoter library strains we use a different recipe of SC, SC phosphate-free (PF) and SC phosphate-rich (PR) medium, a modified recipe from that described in (283). In this case we add per 1 L, 20 g of glucose, 5.6 g YNB with ammonium sulfate, without phosphates, without sodium chloride (MP 4027-812), 0.79 g Complete Supplement Mixture (CSM) (MPB-114500022) and 0.1 g sodium chloride. The amount of adenine and tryptophan was supplemented over the CSM to final concentrations of 0.13 g/L adenine and 0.1 g/L tryptophan to suppress autofluorescence. SC PF medium contains 0.55 g/L of potassium chloride instead of 1 g/L of monobasic potassium phosphate that contains SC PR medium.

SC PF medium recipe (per Litre)

Ingredient	Amount
Nitogen source	
Ammonium sulfate	5 mg
Carbon Source	
Dextrose	20 g
Amino acids	
Adenine	130 mg
L-Arginine	50 mg
L-Aspartic Acid	80 mg
L-Histidine HCl	20 mg
L-Isoleucine	50 mg
L-Leucine	100 mg
L-Lysine HCl	50mg
L-Methionine	20 mg
L-Phenylalanine	50 mg
L-Threonine	100 mg
L-Tryptophan	100mg
Uracil	20 mg

L-Tyrosine	50 mg
L-Valine	140 mg
Vitamins	
Biotin	2 µg
Calcium Pantothenate	0.4 mg
Folic Acid	2 µg
Inositol	2 mg
Niacin	0.4 mg
P-Aminobenzoic Acid	0.2 mg
Pyridoxine Hydrochloride	0.4 mg
Riboflavin	0.2 mg
Thiamine Hydrochloride	0.4 mg
Trace elements	
Boric Acid	0.5 mg
Copper Sulfate	0.04 mg
Potassium Iodide	0.1 mg
Ferric Chloride	0.2 mg
Manganese Sulfate	0.4 mg
Sodium Molybdate	0.2 mg
Zinc Sulfate	0.4 mg
Salts	
Magnesium Sulfate	0.5 mg
Calcium Chloride	0.1 mg
Potassium chloride	550 mg
Sodium chloride	100 mg

Promoter library and yeast strain construction

Promoter variants were ordered as gene fragments from TWIST Bioscience. The PHO5 promoter library integration cassette consisted of two LYS2 homology arms (each 500 bp), PHO5 promoters variants, yeast codon optimized enhanced green fluorescent protein (yoEGFP), TDH1 terminator sequence (177 bp), and URA3 auxotrophic marker. The minimal CYC1 promoter library integration cassette consisted of two LYS2 homology arms (each 500 bp), CYC1 minimal promoter variants, yoEGFP ENO2 terminator and URA3 auxotrophic marker. All integration cassettes were constructed hierarchically by using modular part plasmids following the YeastToolkit (YTK) workflow (176).

Briefly, multiple YTK part plasmids were assembled into cassette plasmids via BsaI goldengate

assembly and then multiple cassette plasmids were assembled to the multi-cassette plasmids via BsmBI goldengate assembly. Goldengate reactions are prepared as following: add the volume necessary to have 100 ng of each plasmid, except for the backbone that was added 25 ng, 2 μL of 10x T4 ligase buffer (NEB), 1 μL of T4 ligase (NEB), 1 μL of BsaI or BsmBI (NEB), and adding water to final volume of 20 μL. Thermocycler setup: (BsmBI assembly: 45°C for 2 min/BsaI assembly: 42 °C for 2 min, 16°C for 5 min) x 25 cycles, followed by a final digestion step at 60°C for 10 min and a heat inactivation at 80°C for 20 min. To assemble PHO5 promoter library cassette plasmids (from pSC189 to pSC201), assembly connector ConLS part plasmid(pYTK002), PHO5 promoters variant part plasmids (from pSC170 to pSC182), yoEGFP part plasmid (pMC068), TDH1 terminator part plasmid (pYTK056), assembly connector ConRE part plasmid(pYTK072) were assembled into the backbone plasmid (pYTK095) via a BsaI goldengate assembly. PHO5 promoter library integration cassette plasmids (from pML001 to pML013) were constructed by BsmBI assembly of cassettes plasmids (from pSC189 to pSC201) into the backbone plasmid that contains two LYS2 homology arms and URA3 auxotrophic marker (pSC109).

To assemble the minimal CYC1 promoter library cassette plasmids (from pAS009 to pAS016 and from pML022 to pML033), assembly connector ConLS part plasmid(pYTK002), CYC1 minimal promoters variant part plasmids (from pAS001 to pAS008 and from pAS058 to pAS069), yoEGFP part plasmid (pMC068), ENO2 terminator part plasmid (pYTK055), assembly connector ConRE part plasmid (pYTK072) were assembled into the backbone plasmid (pYTK095) via a BsaI goldengate assembly. CYC1 minimal promoter library integration cassette plasmids (from pAS017 to pAS024 and from pML034 to pML045) were constructed by BsmBI assembly of cassettes plasmids (from pAS009 to pAS016 and from pML022 to pML033), into the backbone plasmid that contains two LYS2 homology arms and URA3 auxotrophic marker (pSC109).

For all plasmid cloning we used NEB 10-beta Competent *E. coli* (High Efficiency) (New England Biolabs, Cat C3019H). We followed the transformation protocol from the manufacturer. We conducted bacterial selection and growth in Lysogeny Broth (LB) plates or LB medium at 37°C with supplement of appropriate antibiotics (chloramphenicol 34 μg/mL, ampicillin 100 μg/mL or kanamycin 50 μg/mL). All plasmids were sequence-verified by sanger sequencing (Microsynth Ecoli NightSeq).

The PHO5 promoter library yeast strains were made by transforming *S. cerevisiae* strains BY4741 and ΔPho4 (Yeast Knockout Library, Horizon Discovery Ltd.), with the integration cassettes and we obtained 26 variant strains (from sML034 to sML059). The minCYC1 promoter library integration cassettes were used to transform the parental strains BY4741 and sSC051 (S.C. et al., unpublished data) to obtain 40 variant strains (from sAS001 to sAS008 and from sML073 to sML104). In brief, sSC051 has integrated into HO locus the yeast codon optimized mScarlet-i (yomScarlet-i) under the control of a modified GAL1 promoter with six Zif268 biding sites as described (295). Upstream, in the same HO locus, this strain has integrated

the artificial transcription factor Z3EV (a fusion of the mouse transcriptional factor Zif268 DNA binding domain, the ligand binding domain of the human estrogen receptor and viral protein 16) that is selectively relocated from the cytoplasm to the nucleus in the presence of β-estradiol. All promoter library yeast strains were made by homologous recombination of the integration cassettes, into LYS2 locus, using the lithium acetate/polyethylene glycol (PEG) method (203). For yeast integration all integration cassette plasmids were digested with NotI and after that the DNA was purified (ZYMO DNA clean and concentrator-25 kit). The transformants were selected in SC medium lacking uracil plates at 30°C. We screened transformants for correct integration by colony PCR followed by sanger sequencing of the colony PCR product (Microsynth).

Yeast characterization

Promoter library strains were grown overnight in 3 ml in SC medium lacking uracil at 30°C shaking at 250 rpm. Cultures were diluted to an optical density (O.D) of 0.175 in fresh SC medium lacking uracil. Cells grown until log phase, OD ~ 0.8, then PHO5 promoter library strains were washed twice in SC phosphate-free (PF) medium and diluted to starting OD of 0.1 – 0.2 in SC PF medium and SC phosphate-rich (PR) medium. Minimal CYC1 promoter strains were diluted to starting OD of 0.1 – 0.2 in SC medium lacking uracil and SC medium lacking uracil with β-estradiol at 200 nM. Cells were cultured in 96-well plates with clear flat bottom (Nunc) and covered with a gas-permeable membrane (Breathe-Easy membrane, Sigma-Aldrich). OD, GFP and mScarlet was measured every 10 min for 20 – 24 hours on a plate reader (BioTek SynergyMx). 96-well plates were incubated at 30°C with continuous agitation while the plate reading is not measuring. Background of the media was subtracted , then data was normalized dividing by the OD at each time point.

The same growth and inducing conditions were used to induce promoter library for characterization by fluorescence microscopy. The only difference being that 96-well plates with clear conical bottoms (Nunc) were used for growing the yeast for microscope analysis. Strains were imaged at log phase, at an OD of around 0.8, using a Nikon Eclipse Ti fluorescent microscope. Sample visualization was conducted under a 60x objective through the NIS-Elements software (Nikon Instruments). Cells were imaged in bright-field and fluorescence modes. Images were processed and data analyzed using Image J.

6 Conclusion and outlook

In this thesis, we performed the quantitative characterization of the native inorganic phosphate gene regulatory network and investigated the robustness of the native Pho network in response to the changes in the concentration of master regulator Pho4. We assembled a synthetic gene regulatory network using the synthetic transcription factor and synthetic promoters on the yeast artificial chromosome. We also performed a systematic analysis of low-affinity transcription factor binding site clusters *in vitro* and in a synthetic and native gene regulatory network *in vivo* in *S. cerevisiae*.

In chapter 2, we conducted a system-level analysis of the Pho regulon at a single cell level with three state-of-the-art high throughput microfluidic devices. We characterized the master regulator Pho4 and 23 regulated promoters in the Pho regulon. We found that the Pho regulon can attain three well-defined states: a fully activated state, a plateau state, and an off state, with a bistable transition state between the off and the plateau state. The plateau state is perfectly adapted and robust to external P_i concentration changes over a range of 20 - 100 μM, or 20 - 200 μM if the bistable region is included. We characterized nuclear localization of Pho4 over a wide range of external P_i concentrations and discovered that, rather than attaining two nuclear localization states, Pho4 can attain three distinct nuclear localization states. In the plateau region Pho4 is partially localized to the nucleus, and this localization level is robust to external P_i concentration. Having three defined nuclear Pho4 concentrations of zero, intermediate, and high, also simplifies Pho regulon promoter evolution. Decoupling is used to prevent promoter activation in the plateau region, which is easily accomplished because nuclear Pho4 concentrations are sufficiently low to prevent outcompeting nucleosomes. Only when nuclear Pho4 concentrations are high can Pho4 compete with nucleosomes leading to full promoter activation. A carrier transport model recently explained the existence of different affinity transporters (200). We realized that the carrier model also predicts that internal substrate concentrations are constant and robust to external substrate concentrations. Therefore a simple system of two transporters with different affinities could establish two robust [iP_i], which then simply need to establish two Pho4 phosphorylation states through the Pho80p/85p kinase system. These two phosphorylation states in turn determine whether Pho4 is sequestered to the cytoplasm, or nuclear localized to intermediate levels. When iP_i levels fall further, Pho4 is completely dephosphorylated and fully nuclear localized. This model is compatible with all known Pho regulon phenotypes, including the observation of commitment or hysteresis that occurs when cells are returned to P_i rich conditions.

In Chapter 3, we investigated the robustness of the inorganic phosphate gene regulatory network by tuning the expression level of Pho4 using a Z3EV-mediated inducible system. By simply adding the β-estradiol as the inducer, we are able to tune the Pho4 expression level from the basal level similar to the BY4741 strain, to the native Pho4 level, and far above the native Pho4 level. The inducible Pho4 could respond to the P_i starvation and translocated to the nucleus. When Pho4 was induced at a relatively high expression level, it could leak into the nucleus and activate the regulated promoters. This phenomenon could be explained by an excess of Pho4 saturating the Pho80/Pho85 CDK complexes, which are responsible for

phosphorylating Pho4 in the nucleus (229). As a result, only a subset of Pho4 molecules are phosphorylated and translocated to the cytoplasm. When Pho4 was induced at a relatively low expression level, comparable to or lower than the native Pho4 level, no activation of the regulated promoters was observed. This suggests that Pho4 could be fully phosphorylated and remain in the cytoplasm under these conditions. Therefore, it can be concluded that the robustness of Pho4 translocation is limited to a certain range of Pho4 expression levels. We examined promoter activation under full iP_i starvation conditions with a low Pho4 concentration. Different promoters exhibited similar behavior, becoming activated upon iP_i starvation and showing increased activation as the Pho4 concentration increased. With a low Pho4 concentration, we could activate the promoter with a strength lower than the native level. Activation level reached a plateau when the Pho4 level reached a threshold. This can be explained by the fact that when the Pho4 level is very low, only a portion of binding motifs are occupied, or Pho4 has a shorter residence time on the promoter, leading to lower activation. When the Pho4 level is higher, Pho4 can saturate the binding sites on the promoters or outcompete nucleosomes for binding, resulting in saturated activation and demonstrating robustness to changes in Pho4 concentration. The constitutively nuclear-localized Pho4 could potentially aid in decoupling the effects of P_i starvation and β-estradiol titration by directly adjusting the Pho4 level in the nucleus, thereby activating regulated promoters without the need for starvation. This hypothesis requires further testing in future studies. Additionally, the robustness of Pho networks in response to mutations in the Pho4 binding domain should be investigated beyond the scope of this thesis work.

In Chapter 4, we detailed the efforts to construct an orthogonal transcriptional gene regulatory network using a zinc finger synthetic transcription factor and zinc finger synthetic promoters on a yeast artificial chromosome. The synthetic transcription factor PhoZF was designed by replacing the basic domain of the native Pho4 with a Zinc Finger binding domain ZF43-8, characterized by Khalil et al. (190). The translocation domain remained identical to native Pho4, allowing PhoZF to be responsive to iP_i starvation. PhoZF was mutated at specific amino acid residues in the phosphorylation domain to create a phosphate non-responsive version, PhoZF*(NLS) (229). PhoZF expression was controlled by the Z3EV system described in Chapter 2, with the expression level modulated by adding the inducer β-estradiol. To evaluate the orthogonality of PhoZF to native promoters, we integrated all native Pho4-regulated promoters into inducible PhoZF and PhoZF*(NLS) strains and assessed their functionality using a plate reader. Neither PhoZF nor PhoZF*(NLS) activated the native promoters (data not shown), which can be attributed to the replacement of the basic binding domain in PhoZF with the ZF binding domain, as it is not expected to bind native Pho4-regulated promoters. The synthetic promoters were designed by replacing consensus binding motifs in native Pho4-regulated promoters (Pho5, Pho89, Pho84, Pho81, Pho8, Pho11, Pho12, Pho86, VTC1-4 promoters) and a minimal CYC1 promoter characterized by Bashor et al. (230), with zinc finger binding motifs. We designed and integrated a total of 16 synthetic promoters into inducible PhoZF strains, featuring different numbers of binding sites. The promoter strength can be further adjusted by modifying the binding affinity between PhoZF and ZF motifs (190). We employed a yeast

artificial chromosome to carry the synthetic PhoZF and ZF promoters, integrating them into the SUP4 and URA3 loci of the YAC. The performance of the synthetic PhoZF and ZF promoters was characterized with 100 nM β-estradiol induction in a plate reader. The mScarleti-PhoZF could be induced by β-estradiol, activating the ZF promoters at various levels. The promoter activation fold change can be further optimized to individually match native promoter levels. The YAC is a vector capable of carrying over 1 MB of DNA insert. In the future work, we plan to carry all synthetic PhoZF and Pho regulon genes controlled by the ZF promoters on a single YAC to match the input-output level of native Pho networks, ultimately constructing a complete synthetic gene regulatory network to replace the native ones.

In Chapter 5, we applied a high-throughput *in vitro* method to characterizing collective transcription factor binding to DNA sequences consisting of arrays of binding sites ranging over orders of magnitude in affinity. This permitted us to develop a quantitative biophysical understanding of how transcription factors interact with binding site clusters, particularly to clusters consisting of low-affinity binding sites. Clusters of low-affinity binding sites achieved similar levels of occupancy as single high-affinity binding sites and we predicted that this level of occupancy can be achieved at physiologically relevant transcription factor concentrations. We verified this effect with two distinct transcription factors representing two of the largest transcription factor structural classes: zinc finger and bHLH. Small clusters of 2-6 binding sites with individual binding site affinities one order of magnitude lower than the consensus sequence were able to achieve this effect. We furthermore challenged the notion that overlapping binding sites are mutually exclusive due to steric effects, and thus would allow only one transcription factor to bind to overlapping sites at any given time. In fact, clusters of strong binding sites with a one bp overlap of a 9 bp consensus sequence were not mutually exclusive and were simultaneously bound by several transcription factors. Even a 3 bp overlap of a 9bp consensus motif permitted some co-occurring binding, although here the steric factor strongly dominated. Similar trends were observed for weak binding site clusters with 1 and 3 bp overlapping sequences, indicating that strong binding sites permit and support partial binding of a transcription factor, allowing full occupancy of overlapping binding sites. The high levels of occupancy achieved by low-affinity binding site clusters as determined by our *in vitro* characterization suggested that low-affinity binding site clusters may indeed be functional *in vivo*. We therefore first tested low-affinity binding site clusters in an engineered synthetic system *in vivo*. We placed low-affinity Zif268 binding site clusters in the minimal CYC1 promoter and used a β-estradiol inducible Zif268 transcription factor (Z_3EV) to test whether these promoters could give rise to expression. We found that binding site clusters of 3 to 6 weak binding sites with individual affinities one order of magnitude below the consensus sequence were able to activate transcription to similar levels as a promoter containing 3 consensus sequence binding sites. Therefore, in this synthetic system, small clusters of low-affinity binding sites are functional. In summary we conducted a systematic characterization of low-affinity binding site clusters by conducting a quantitative biochemical analysis *in vitro* as well as functional studies *in vivo*. This work provided insights into how transcription factors bind to such clusters and established that a small number of low-affinity

binding sites in a local cluster can be highly functional *in vivo.* These insights improve our current understanding of gene regulatory networks, and our ability to engineer sophisticated gene regulatory networks *de novo.*

Bibliography

[1] Davidson, E. & Levine, M. Gene regulatory networks. *PNAS* **5**, 4935 (2004).

[2] Davidson, E., Rast, J. P. & Oliveri, P. A Genomic Regulatory Network for Development. *Science* **295**, 1669–1678 (2002).

[3] Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology* **2**, 1–7 (2014).

[4] Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).

[5] Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Research* **25**, 1491–1498 (2015).

[6] Kim, J. & Eberwine, J. RNA: State memory and mediator of cellular phenotype. *Trends in Cell Biology* **20**, 311–318 (2010).

[7] Dueck, H. *et al.* Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biology* **16**, 1–17 (2015).

[8] Grewal, S. I. & Moazed, D. Heterochromatin and epigenetic control of gene expression. *Science* **301**, 798–802 (2003).

[9] Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K. & Vingron, M. Histone modification levels are predictive for gene expression. *PNAS* **107**, 2926–2931 (2010).

[10] Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**, 245–254 (2003).

[11] Gaydos, L. J., Wang, W. & Strome, S. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345**, 1515–1518 (2014).

[12] Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews Molecular Cell Biology* **18**, 31–42 (2016).

[13] Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nature Reviews Molecular Cell Biology* (2022).

Bibliography

[14] De Jong, H. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology* **9**, 67–103 (2002).

[15] Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* **9**, 770–780 (2008).

[16] Goffeau, A. *et al.* Life with 6000 Genes. *Science* **274**, 546–567 (1996).

[17] Botstein, D. & Fink, G. R. Yeast: An experimental organism for modern biology. *Science* **189**, 1439–1442 (1988).

[18] Nielsen, J. Yeast Systems Biology: Model Organism and Cell Factory. *Biotechnology Journal* **14**, 1800421 (2019).

[19] Boone, C. Yeast systems biology: Our best shot at modeling a cell. *Genetics* **198**, 435–437 (2014).

[20] Hohmann, S. Nobel yeast research. *FEMS Yeast Research* **16**, 1–3 (2016).

[21] Yu, R. & Nielsen, J. Yeast systems biology in understanding principles of physiology underlying complex human diseases. *Current Opinion in Biotechnology* **63**, 63–69 (2020).

[22] DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).

[23] Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *PNAS* **95**, 14863–14868 (1998).

[24] Hughes, T. R. *et al.* Functional Discovery via a Compendium of Expression Profiles. *Cell* **102**, 109–126 (2000).

[25] Huh, W.-k. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

[26] Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).

[27] Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).

[28] Giaever, G., Chu, L., Angela M.and Ni & Connelly, C. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002).

[29] Giaever, G., Chu, L., Angela M.and Ni & Connelly, C. The yeast deletion collection: A decade of functional genomics. *Genetics* **197**, 451–465 (2014).

[30] Tsien, R. Y. The green fluorescent protein. *Annual Review of Biochemistry* **67**, 509–544 (1998).

[31] Cherry, J. M. *et al.* SGD: Saccharomyces Genome Database. *Nucleic Acids Research* **26**, 73–79 (1998).

[32] Cherry, J. M. *et al.* Saccharomyces Genome Database: The genomics resource of budding yeast. *Nucleic Acids Research* **40**, 700–705 (2012).

[33] Consortium, T. G. O. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).

[34] Hughes, T. R. & de Boer, C. G. Mapping yeast transcriptional networks. *Genetics* **195**, 9–36 (2013).

[35] De Boer, C. G. & Hughes, T. R. YeTFaSCo: A database of evaluated yeast transcription factor sequence specificities. *Annual Review of Biochemistry* **40**, 169–179 (2012).

[36] Weirauch, M. T. & Hughes, T. R. A catalogue of eukary- otic transcription factor types, their evolutionary origin, and species distribution. *Subcell. Biochem.* **52**, 25–73 (2011).

[37] Chua, G., Robinson, M. D., Morris, Q. & Hughes, T. R. Transcriptional networks: Reverse-engineering gene regulation on a global scale. *Current Opinion in Microbiology* **7**, 638–646 (2004).

[38] MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**, 1–14 (2006).

[39] Geertz, M. & Maerkl, S. J. Experimental strategies for studying transcription factor-DNA binding specificities. *Briefings in Functional Genomics* **9**, 362–373 (2010).

[40] Jolma, A. & Taipale, J. Methods for analysis of transcrip- tion factor DNA-binding specificity in vitro. *Subcell. Biochem.* **52**, 155–173 (2011).

[41] Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

[42] Djordjevic, M. SELEX experiments: New prospects, applications and data analysis in inferring regulatory pathways. *Biomolecular Engineering* **24**, 179–189 (2007).

[43] Bulyk, M. L., Huang, X., Choo, Y. & Church, G. M. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7158–63 (2001).

[44] Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnology* **24**, 1429–1435 (2006).

Bibliography

[45] Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science (New York, NY)* **315**, 233–237 (2007).

[46] SJ, M. & SR, Q. Experimental determination of the evolvability of a transcription factor. *PNAS* **106**, 18650–5 (2009).

[47] Badis, G. *et al.* Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* **324**, 1720–1723 (2009).

[48] Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Research* **19**, 556–566 (2009).

[49] Gordân, R. *et al.* Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biology* **12**, R125 (2011).

[50] Fordyce, P. M. *et al.* De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature biotechnology* **28**, 970–5 (2010).

[51] Horak, C. E. & Snyder, M. ChIP-chip: A genomic approach for identifying transcription factor binding sites. *Methods in Enzymology* **350**, 469–483 (2002).

[52] Ren, B. *et al.* Genome-Wide Location and Function of DNA Binding Proteins. *Science* **290**, 2306–2309 (2000).

[53] Hanlon, S. E. & Lieb, J. D. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Current Opinion in Genetics and Development* **14**, 697–705 (2004).

[54] Johnson, D. S., Mortazavi, A. & Myers, R. M. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **316**, 1497–1502 (2007).

[55] Park, P. J. ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**, 669–680 (2009).

[56] Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in biochemical sciences* **25**, 99–104 (2000).

[57] Nelson, J. D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature protocols* **1**, 179–185 (2006).

[58] de Wit, E. & De Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes & development* **26**, 11–24 (2012).

[59] Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic acids research* **46**, D794—-D801 (2018).

[60] Celniker, S. E. *et al.* Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).

[61] van Steensel, B. & Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature biotechnology* **18**, 424–428 (2000).

[62] Aughey, G. N. & Southall, T. D. Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdisciplinary Reviews: Developmental Biology* **5**, 25–37 (2016).

[63] Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols* **2010**, pdb—-prot5384 (2010).

[64] Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**, 877–885 (2007).

[65] Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current protocols in molecular biology* **109**, 21–29 (2015).

[66] Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

[67] Fields, S. & Song, O.-k. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).

[68] Chien, C.-T., Bartel, P. L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Sciences* **88**, 9578–9582 (1991).

[69] Sieweke, M. Detection of transcription factor partners with a yeast one hybrid screen. In *Transcription Factor Protocols*, 59–77 (Springer, 2000).

[70] Deplancke, B., Dupuy, D., Vidal, M. & Walhout, A. J. M. A gateway-compatible yeast one-hybrid system. *Genome research* **14**, 2093–2101 (2004).

[71] Cottier, S. *et al.* The yeast three-hybrid system as an experimental platform to identify proteins interacting with small signaling molecules in plant cells: potential and limitations. *Frontiers in plant science* **2**, 101 (2011).

[72] Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods* **15**, 47–51 (2018).

[73] Herzenberg, L. A. *et al.* The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clinical chemistry* **48**, 1819–1827 (2002).

Bibliography

[74] Grégori, G. *et al.* Hyperspectral cytometry at the single-cell level using a 32-channel photodetector. *Cytometry Part A* **81**, 35–44 (2012).

[75] de Boer, C. G., Vaishnav, E. D., Sadeh, R., Abeyta, E. L. & Friedman, A., Nir Regev. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology* **38**, 56–65 (2020).

[76] Whitesides, G. M. The origins and the future of microfluidics. *Proceedings of the National Academy of Sciences* **442**, 368–373 (2006).

[77] Xia, Y. & Whitesides, G. M. Soft lithography. *Angew. Chem. Int. Ed. Engl.* **37**, 550–575 (1998).

[78] McDonald, J. C. *et al.* Fabrication of microfluidic systems in poly (dimethylsiloxane). *Electrophoresis* **21**, 27–40 (2000).

[79] Sackmann, E. K., Fulton, A. L. & Beebe, D. J. The present and future role of microfluidics in biomedical research. *Nature* **507**, 181–189 (2014).

[80] Shen, Y., Yalikun, Y. & Tanaka, Y. Recent advances in microfluidic cell sorting systems. *Sensors and Actuators B: Chemical* **282**, 268–281 (2019).

[81] Denervaud, N. *et al.* A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proceedings of the National Academy of Sciences* **110**, 15842–15847 (2013).

[82] Ahrberg, C. D., Manz, A. & Chung, B. G. Polymerase chain reaction in microfluidic devices. *Lab on a Chip* **16**, 3866–3884 (2016).

[83] Guo, M., Rotem, A., Heyman, J. & Weitz, D. Droplet microfluidics for high-throughput biological assays. *Lab Chip* **12**, 2146–2155 (2012).

[84] Joensson, H. N. & Andersson Svahn, H. Droplet microfluidics–a tool for single-cell analysis. *Angewandte Chemie International Edition* **51**, 12176–12192 (2012).

[85] Low, L. A., Mummery, C., Berridge, B. R., Austin, C. P. & Tagle, D. A. Organs on chips into the next decade. *Nature Reviews Drug Discovery* **20**, 345–361 (2021).

[86] Leung, C. M. *et al.* A guide to the organ on a chip. *Nature Reviews Methods Primers* **2**, 2662–8449 (2022).

[87] Breslauer, D. N., Lee, P. J. & Lee, L. P. Microfluidics-based systems biology. *Molecular BioSystems* **2**, 97–112 (2006).

[88] Bennett, M. R. & Hasty, J. Microfluidic devices for measuring gene network dynamics in single cells. *Nature Reviews Genetics* **10**, 628–638 (2009).

[89] Kim, J. *et al.* A programmable microfluidic cell array for combinatorial drug screening. *Lab Chip* **12**, 1813–1822 (2012).

[90] Cookson, S., Ostroff, N., Pang, W.-L., Volfson, D. & Hasty, J. Monitoring dynamics of single-cell gene expression over multiple cell cycles. *Molecular Systems Biology* **1**, 0024 (2005).

[91] Rajkumar, A. S., Denervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nature genetics* **45**, 1207–1215 (2013).

[92] Zhu, Z., Frey, O., Ottoz, D. S., Rudolf, F. & Hierlemann, A. Microfluidic single-cell cultivation chip with controllable immobilization and selective release of yeast cells. *Lab Chip* **12**, 906–915 (2012).

[93] Gao, Z. *et al.* Comparative analysis of yeast replicative lifespan in different trapping structures using an integrated microfluidic system. *Advanced Materials Technologies* **5**, 2000655 (2020).

[94] Lee, S.-S., Vizcarra, I. A., Huberts, D. H., Lee, L. P. & Heinemann, M. Whole lifespan microscopic observation of budding yeast aging through a microfluidic dissection platform. *Proceedings of the National Academy of Sciences* **109**, 4916–4920 (2012).

[95] Zhang, Y. *et al.* Single cell analysis of yeast replicative aging using a new generation of microfluidic device. *PLOS ONE* **7**, 1–10 (2012).

[96] Li, Y. *et al.* Multigenerational silencing dynamics control cell aging. *Proceedings of the National Academy of Sciences* **114**, 11253–11258 (2017).

[97] Li, Y. *et al.* A programmable fate decision landscape underlies single-cell aging in yeast. *Science* **369**, 325–329 (2020).

[98] Taylor, R. J. *et al.* Dynamic analysis of mapk signaling using a high-throughput microfluidic single-cell imaging platform. *Proceedings of the National Academy of Sciences* **106**, 3758–3763 (2009).

[99] Bennett, M. R. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–1122 (2008).

[100] Dertinger, S., Chiu, D., Jeon, N. & Whitesides, G. Generation of gradients having complex shapes using microfluidic networks. *Analytical chemistry* **73**, 1240–1246 (2001).

[101] Walker, G. M., Monteiro-Riviere, N., Rouse, J. & O'Neill, A. T. A linear dilution microfluidic device for cytotoxicity assays. *Lab on a Chip* **7**, 226–232 (2007).

[102] Ming Yip, H., Cheng, S., Olson, E. J., Crone, M. & Maerkl, S. J. Perfect adaptation achieved by transport limitations governs the inorganic phosphate response in s. cerevisiae. *PNAS* **120**, 1–9 (2023).

[103] Austin, S. & Mayer, A. Phosphate homeostasis–a vital metabolic equilibrium maintained through the inphors signaling pathway. *Frontiers in Microbiology* **11**, 1367 (2020).

Bibliography

[104] Oshima, Y. The phosphatase system in *Saccharomyces cerevisiae. Genes & genetic systems* **72**, 323–334 (1997).

[105] Yoshida, K., Ogawa, N. & Oshima, Y. Function of the pho regulatory genes for repressible acid phosphatase synthesis in saccharomyces cerevisiae. *Molecular Genetics and Genomics* **217**, 40–46 (1989).

[106] Yoshida, K., Kuromitsu, Z., Ogawa, N. & Oshima, Y. Mode of expression ofthe positive regulatory genes pho2 and pho4 ofthe phosphatase regulon in saccharomyces cerevisiae. *Molecular Genetics and Genomics* **217**, 31–39 (1989).

[107] O'Neill, E. M., Kaffman, A., Jolly, E. R. & O'Shea, E. K. Regulation of pho4 nuclear localization by the pho80-pho85 cyclin-cdk complex. *Science* **271**, 209–212 (1996).

[108] Komeili, A. & O'Shea, E. K. Roles of phosphorylation sites in regulating activity of the transcription factor pho4. *Science* **284**, 977–980 (1999).

[109] Jeffery, D. A., Springer, M., King, D. S. & O'Shea, E. K. Multi-site phosphorylation of pho4 by the cyclin-cdk pho80-pho85 is semi-processive with site preference. *Journal of molecular biology* **306**, 997–1010 (2001).

[110] Schneider, K. R., Smith, R. L. & O'Shea, E. K. Phosphate-regulated inactivation of the kinase pho80–pho85 by the cdk inhibitor pho81. *Science* **266**, 122–126 (1994).

[111] Springer, M., Wykoff, D. D., Miller, N. & O'Shea, E. K. Partially phosphorylated pho4 activates transcription of a subset of phosphate-responsive genes. *PLoS biology* **1**, e28 (2003).

[112] Secco, D., Wang, C., Shou, H. & Whelan, J. Phosphate homeostasis in the yeast saccharomyces cerevisiae, the key role of the spx domain-containing proteins. *FEBS Letters* **586**, 289–295 (2012).

[113] Jensen, L. T., Ajua-Alemanji, M. & Culotta, V. C. The saccharomyces cerevisiae high affinity phosphate transporter encoded by pho84 also functions in manganese homeostasis. *Journal of Biological Chemistry* **278**, 42036–42040 (2003).

[114] Hürlimann, H. C., Pinson, B., Stadler-Waibel, M., Zeeman, S. C. & Freimoser, F. M. The spx domain of the yeast low-affinity phosphate transporter pho90 regulates transport activity. *EMBO reports* **10**, 1003–1008 (2009).

[115] Hürlimann, H. C., Stadler-Waibel, M., Werner, T. & Freimoser, F. M. Pho91 is a vacuolar phosphate transporter that regulates phosphate and polyphosphate metabolism in saccharomyces cerevisiae. *Molecular Biology of the Cell* **18**, 4438–4445 (2007).

[116] Persson, B. L. *et al.* Phosphate permeases of *Saccharomyces cerevisiae. Biochimica et Biophysica Acta* **1365**, 23–30 (1998).

[117] Levy, S., Kafri, M., Carmi, M. & Barkai, N. The competitive advantage of a dual-transporter system. *Science* **334**, 1408–1412 (2011).

[118] Wykoff, D. D., Rizvi, A. H., Raser, J. M., Margolin, B. & O'Shea, E. K. Positive feedback regulates switching of phosphate transporters in s. cerevisiae. *Molecular cell* **27**, 1005–1013 (2007).

[119] Sethuraman, A., Rao, N. N. & Kornberg, A. The endopolyphosphatase gene: essential in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **98**, 8542–8547 (2001).

[120] Gerasimaite, R. & Mayer, A. Enzymes of yeast polyphosphate metabolism: structure, enzymology and biological roles. *Biochemical Society Transactions* **44**, 234–239 (2016).

[121] Gerasimaite, R. & Mayer, A. Ppn2, a novel zn(2+)-dependent polyphosphatase in the acidocalcisome-like yeast vacuole. *Journal of Cell Science* **130**, 1625–1636 (2017).

[122] Gerasimaite, R., Sharma, S., Desfougères, Y., Schmidt, A. & Mayer, A. Coupled synthesis and translocation restrains polyphosphate to acidocalcisome-like vacuoles and prevents its toxicity. *Journal of cell science* **127**, 5093–5104 (2014).

[123] Vardi, N., Levy, S., Gildor, T. & Pilpel, Y. Sequential feedback induction stabilizes the phosphate starvation response in budding yeast. *Cell reports* **9**, 1122–1134 (2014).

[124] Ogawa, N., DeRisi, J. & Brown, P. O. New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. *Molecular biology of the cell* **11**, 4309–4321 (2000).

[125] Reddy, V. S., Singh, A. K. & Rajasekharan, R. The saccharomyces cerevisiae phm8 gene encodes a soluble magnesium-dependent lysophosphatidic acid phosphatase*. *Journal of Biological Chemistry* **283**, 8846–8854 (2008).

[126] Xu, Y.-F. *et al.* Nucleotide degradation and ribose salvage in yeast. *Molecular Systems Biology* **9**, 665 (2013).

[127] Patton-Vogt, J. Transport and metabolism of glycerophosphodiesters produced through phospholipid deacylation. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1771**, 337–342 (2007).

[128] Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nature Reviews Microbiology* **12**, 381–390 (2014).

[129] Meng, F. & Ellis, T. The second decade of synthetic biology: 2010–2020. *Nature Communications* **11**, 1–4 (2020).

[130] Voigt, C. A. Synthetic biology 2020 – 2030 : six commercially-available products that are changing our world. *Nature Communications* **11**, 10–15 (2020).

Bibliography

[131] Benner, S. A. & Sismour, A. M. Synthetic biology. *Nature Reviews Genetics* **6**, 533–543 (2005).

[132] Jackson, D. A., Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *PNAS* **69**, 2904–2909 (1972).

[133] Saik, R. K. *et al.* Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science* **239**, 487–491 (1987).

[134] Goeddel, D. V., Kleid, D. G. & Bolivar, F. Expression in Escherichia coli of chemically synthesized genes for human insulin. *PNAS* **76**, 106–110 (1979).

[135] Sanger, F. *et al.* Nucleotide sequence of bacteriophage X174DNA. *Nature* **265**, 687–695 (1977).

[136] Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997).

[137] Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2**, 1–14 (2006).

[138] Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).

[139] Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **22**, 867–870 (2004).

[140] Deans, T. L., Cantor, C. R. & Collins, J. J. A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell* **130**, 363–372 (2007).

[141] Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

[142] Nem, M. N.-e., Fung, E., Wong, W. W., Suen, J. K. & Bulter, T. A synthetic gene – metabolic oscillator. *Nature* **435**, 118–122 (2005).

[143] Stricker, J. *et al.* A fast , robust and tunable synthetic gene oscillator. *Nature* **456**, 516–520 (2008).

[144] Tigges, M., Marquez-lago, T. T. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).

[145] Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnology* **25**, 795–801 (2007).

[146] Sohka, T. *et al.* An externally tunable bacterial band-pass filter. *PNAS* **106**, 10135–10140 (2009).

[147] Maung, N. W. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *PNAS* **104**, 14283–14288 (2007).

[148] Ajo-franklin, C. M. *et al.* Rational design of memory in eukaryotic cells. *Genes Development* **21**, 2271–2276 (2007).

[149] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & Van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31**, 69–73 (2002).

[150] Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186 (2002).

[151] Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *PNAS* **100**, 7714–7719 (2003).

[152] Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).

[153] Zhang, W., Mitchell, L. A., Bader, J. S. & Boeke, J. D. Synthetic Genomes. *Annual Review of Biochemistry* **89**, 77–101 (2020).

[154] Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).

[155] Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253–1 (2016).

[156] Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **6213**, 787–793 (2014).

[157] Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–59 (2014).

[158] Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).

[159] Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).

[160] Mercy, G. *et al.* 3D organization of synthetic and scrambled chromosomes. *Science* **355**, eaaf4597 (2017).

[161] Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018).

Bibliography

[162] Pretorius, I. S. & Boeke, J. D. Yeast 2.0—connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Research* **18**, 1–15 (2018).

[163] Editorial. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature communications* **9**, 1939 (2018).

[164] Blount, B. A. *et al.* Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nature communications* **9**, 1932 (2018).

[165] Wang, J. *et al.* Ring synthetic chromosome V SCRaMbLE. *Nature communications* **9**, 3783 (2018).

[166] Shen, M. J. *et al.* Heterozygous diploid and interspecies SCRaMbLEing. *Nature communications* **9**, 1–8 (2018).

[167] Liu, W. *et al.* Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. *Nature communications* **9**, 1–12 (2018).

[168] Wu, Y. *et al.* In vitro DNA SCRaMbLE. *Nature communications* **9**, 1–9 (2018).

[169] Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K. & Mueller-Roeber, B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nature communications* **9**, 1–10 (2018).

[170] Jia, B. *et al.* Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nature communications* **9**, 1–10 (2018).

[171] Luo, Z. *et al.* Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. *Nature communications* **9**, 1–10 (2018).

[172] Shao, Y. *et al.* Creating a functional single-chromosome yeast. *Nature* **560**, 331–335 (2018).

[173] Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).

[174] Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26**, 787–793 (2008).

[175] Smolke, C. D. Building outside of the box: IGEM and the BioBricks Foundation. *Nature Biotechnology* **27**, 1099–1102 (2009).

[176] Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synthetic Biology* **4**, 975–986 (2015).

[177] Rantasalo, A., Kuivanen, J., Penttilä, M., Jäntti, J. & Mojzita, D. Synthetic Toolkit for Complex Genetic Circuit Engineering in Saccharomyces cerevisiae. *ACS Synthetic Biology* **7**, 1573–1587 (2018).

[178] Fonseca, J. P. *et al.* Synthetic Toolkit for Complex Genetic Circuit Engineering in Saccharomyces cerevisiae. *ACS Synthetic Biology* **8**, 2593–2606 (2019).

[179] Costello, A. & Badran, A. H. Synthetic Biological Circuits within an Orthogonal Central Dogma. *Trends in Biotechnology* **39**, 59–71 (2021).

[180] Liu, C. C., Jewett, M. C., Chin, J. W. & Voigt, C. A. Toward an orthogonal central dogma. *Nature Chemical Biology* **14**, 103–106 (2018).

[181] Carroll, D. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Genetics* **188**, 773–782 (2011).

[182] Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *PNAS* **93**, 1156–1160 (1996).

[183] Pavletich, N. P. & Pabo, C. . Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A. *Science* **252**, 809–817 (1991).

[184] Pabo, C. O., Peisach, E. & Grant, R. A. Design and selection of Novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* **70**, 313–40 (2001).

[185] Park, K.-s. *et al.* Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nature Biotechnology* **21**, 1208–1215 (2003).

[186] Beerli, R. R., Dreier, B. & Barbas, C. F. Positive and negative regulation of endogenous genes by designed transcription factors. *PNAS* **97**, 1495–1500 (2000).

[187] Maeder, M. L. *et al.* Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Molecular Cell* **31**, 294–301 (2008).

[188] Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F. & Joung, J. K. Oligomerized pool engineering (OPEN): An 'open-source' protocol for making customized zinc-finger arrays. *Nature Protocols* **4**, 1471–1501 (2009).

[189] Sander, J. D. *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods* **8**, 67–69 (2011).

[190] Khalil, A. S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).

[191] Istomin, I. *Microfluidic iMITOMI platform to study the architecture of low-affinity transcription factor binding site clusters.* Ph.D. thesis, EPFL (2020).

[192] Austin, S. & Mayer, A. Phosphate Homeostasis - A Vital Metabolic Equilibrium Maintained Through the INPHORS Signaling Pathway. *Frontiers in Microbiology* **11**, 1367 (2020).

Bibliography

[193] Oshima, Y., Ogawa, N. & Harashima, S. Regulation of phosphatase synthesis in saccharomyces cerevisiae–a review. *Gene* **179**, 171–177 (1996).

[194] Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).

[195] Ferrell, J. E. Perfect and near-perfect adaptation in cell signaling. *Cell systems* **2**, 62–67 (2016).

[196] Muzzey, D., Gomez-Uribe, C. A., Mettetal, J. T. & van Oudenaarden, A. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* **138**, 160–171 (2009).

[197] Ricci-Tam, C., Valleriani, A., Höfer, T., Ravarani, C. N. & Fierz, B. Decoupling transcription factor expression and activity enables dimmer switch gene regulation. *Science* **372**, 292–295 (2021).

[198] Wang, P. *et al.* Robust growth of escherichia coli. *Current biology* **20**, 1099–1103 (2010).

[199] Lam, F., Steger, D. & O'Shea, E. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).

[200] Bosdriesz, E., Molenaar, D., Teusink, B., Bruggeman, F. J. & Schaap, P. J. Low affinity uniporter carrier proteins can increase net substrate uptake rate by reducing efflux. *Scientific reports* **8**, 5576 (2018).

[201] Bosdriesz, E. *et al.* Low affinity uniporter carrier proteins can increase net substrate uptake rate by reducing efflux. *Scientific Reports* **8**, 5576 (2018).

[202] Vardi, N., Levy, S., Assaf, M., Carmi, M. & Barkai, N. Budding yeast escape commitment to the phosphate starvation program using gene expression noise. *Current biology* **23**, 2051–2057 (2013).

[203] Gietz, R. D. & Schiest, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 31–34 (2007).

[204] Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

[205] Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**, 580–584 (2002).

[206] Dertinger, S. K. W., Chiu, D. T., Jeon, N. L. & Whitesides, G. M. Generation of gradients having complex shapes using microfluidic networks. *Analytical Chemistry* **73**, 1240–1246 (2001).

[207] McQuin, C. *et al.* Cellprofiler 3.0: Next-generation image processing for biology. *PLOS Biology* **16**, 1–17 (2018).

[208] Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nature Methods* (2019).

[209] Bosdriesz, E. *et al.* Low affinity uniporter carrier proteins can increase net substrate uptake rate by reducing efflux. *Scientific Reports* **8**, 5576 (2018).

[210] Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Frontiers in genetics* **6**, 322 (2015).

[211] Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genetics* **24**, 355–361 (2000).

[212] MacNeil, L. T. & Walhout, A. J. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* **21**, 645–657 (2011).

[213] Halford, S. E. & Marko, J. F. How do site-specific dna-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).

[214] AliAzam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* **181**, 6361–6370 (1999).

[215] Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–195 (2010).

[216] Koşar, Z. & Erbaş, A. Can the Concentration of a Transcription Factor Affect Gene Expression? *Frontiers in Soft Matter* **2**, 1–6 (2022).

[217] McIsaac, R. S., Gibney, P. A., Chandran, S. S., Benjamin, K. R. & Botstein, D. Synthetic biology tools for programming gene expression without nutritional perturbations in Saccharomyces cerevisiae. *Nucleic Acids Research* **42**, 1–8 (2014).

[218] Lee, T. I. *et al.* Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799–804 (2002).

[219] Hansen, A. S. & O'Shea, E. K. Cis Determinants of Promoter Threshold and Activation Timescale. *Cell Reports* **12**, 1226–1233 (2015). URL http://dx.doi.org/10.1016/j.celrep.2015.07.035.

[220] Elowitz, M. & Lim, W. A. Build life to understand it. *Nature* **468**, 889–890 (2010).

[221] Temme, K., Zhao, D. & Voigt, C. A. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *PNAS* **109**, 7085–7090 (2012).

[222] Çağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. & Süel, G. M. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139**, 512–522 (2009).

Bibliography

[223] Lu, T. K., Khalil, A. S. & Collins, J. J. Next-generation synthetic gene networks. *Nature Biotechnology* **27**, 1139–1150 (2009).

[224] Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology-identifying causes of failure of synthetic biological systems. *Biotechnology journal* **7**, 856–866 (2012).

[225] Burke, D. T., Carle, G. F. & Olson, M. V. Cloning of large segments of exogenous dna into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812 (1987).

[226] Bruschi, C. V., Gjuracic, K. & Tosato, V. Yeast artificial chromosomes. *eLS* 1–10 (2006).

[227] Ryan, O. W. *et al.* Selection of chromosomal dna libraries using a multiplex crispr system. *eLife* **3**, e03703 (2014).

[228] Blackburn, M. C., Petrova, E., Correia, B. E. & Maerkl, S. J. Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. *Nucleic acids research* **44**, e68–e68 (2015).

[229] Komeili, A. & O'Shea, E. K. Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science* **284**, 977–980 (1999).

[230] Bashor, C. J. *et al.* Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* **364**, 593–597 (2019).

[231] Levo, M. & Segal, E. In pursuit of design principles of regulatory sequences. *Nat Rev Genet* **15**, 453–468 (2014).

[232] Gotea, V. *et al.* Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Research* **20**, 565–577 (2010).

[233] Ezer, D., Zabet, N. R. & Adryan, B. Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. *Computational and structural biotechnology journal* **10**, 63–69 (2014).

[234] Lifanov, A. P., Makeev, V. J., Nazina, A. G. & Papatsenko, D. A. Homotypic regulatory clusters in drosophila. *Genome Research* **13**, 579–588 (2003).

[235] Wunderlich, Z. & Mirny, L. A. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics* **25**, 434–440 (2009).

[236] Estrada, J., Wong, F., DePace, A. & Gunawardena, J. Information integration and energy expenditure in gene regulation. *Cell* **166**, 234–244 (2016).

[237] Crocker, J. *et al.* Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).

[238] Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annual review of cell and developmental biology* **35**, 357–379 (2019).

[239] Crocker, J., Noon, E. P.-B. & Stern, D. L. The soft touch: low-affinity transcription factor binding sites in development and evolution. In *Current Topics in Developmental Biology*, vol. 117, 455–469 (Elsevier, 2016).

[240] Gaudet, J. & Mango, S. E. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295**, 821–825 (2002).

[241] Farley, E. K. *et al.* Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).

[242] Jiang, J. & Levine, M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741–752 (1993).

[243] Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

[244] de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology* **38**, 56–65 (2020).

[245] Rajkumar, A. S., Denervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nature Genetics* **45**, 1207–1215 (2013).

[246] Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008).

[247] Mogno, I., Kwasnieski, J. C. & Cohen, B. A. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome research* (2013).

[248] Weingarten-Gabbay, S. & Segal, E. The grammar of transcriptional regulation. *Human Genetics* **133**, 701–711 (2014).

[249] Castellanos, M., Mothi, N. & Muñoz, V. Eukaryotic transcription factors can track and control their target genes using dna antennas. *Nature Communications* **11**, 1–13 (2020).

[250] Levo, M. *et al.* Unraveling determinants of transcription factor binding outside the core binding site. *Genome research* **25**, 1018–1029 (2015).

[251] Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).

[252] Kim, S. *et al.* Probing allostery through DNA. *Science* **339**, 816–819 (2013).

Bibliography

[253] Rockel, S., Geertz, M., Hens, K., Deplancke, B. & Maerkl, S. J. iSLIM: a comprehensive approach to mapping and characterizing gene regulatory networks. *Nucleic acids research* **41**, e52 (2013).

[254] Le, D. D. *et al.* Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proceedings of the National Academy of Sciences* **115**, E3702–E3711 (2018).

[255] Aditham, A. K., Markin, C. J., Mokhtari, D. A., DelRosso, N. & Fordyce, P. M. High-throughput affinity measurements of transcription factor and DNA mutations reveal affinity and specificity determinants. *Cell Systems* **12**, 112–127 (2021).

[256] Martin, L. L. *et al.* Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nature Methods* **9**, 1192–1194 (2012).

[257] Gerber, D., Maerkl, S. J. & Quake, S. R. An in vitro microfluidic approach to generating protein-interaction networks. *Nature Methods* **6**, 71–74 (2009).

[258] Einav, S. *et al.* Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. *Nature Biotechnology* **26**, 1019–1027 (2008).

[259] Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences* **109**, 16540–16545 (2012).

[260] Bates, S. R. & Quake, S. R. Highly parallel measurements of interaction kinetic constants with a microfabricated optomechanical device. *Applied Physics Letters* **95**, 73705–73705 (2009).

[261] Popp, A. P., Hettich, J. & Gebhardt, J. C. M. Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Research* **49**, 6249–6266 (2021).

[262] Clauß, K. *et al.* DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Research* **45**, 11121–11130 (2017). URL http://academic.oup.com/nar/article/45/19/11121/4085842.

[263] Callegari, A. *et al.* Single-molecule dynamics and genome-wide transcriptomics reveal that nf-kb (p65)-DNA binding times can be decoupled from transcriptional activation. *PLoS Genetics* **15**, e1007891 (2019).

[264] Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2012).

258

[265] Loffreda, A. *et al.* Live-cell p53 single-molecule binding is modulated by c-terminal acetylation and correlates with transcriptional activity. *Nature Communications* **8**, 1–12 (2017).

[266] Rastogi, C. *et al.* Accurate and sensitive quantification of protein-DNA binding affinity. *Proceedings of the National Academy of Sciences* **115**, E3692–E3701 (2018).

[267] Swank, Z., Laohakunakorn, N. & Maerkl, S. J. Cell-free gene-regulatory network engineering with synthetic transcription factors. *Proceedings of the National Academy of Sciences* **116**, 5892–5901 (2019).

[268] Biggin, M. D. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell* **21**, 611–626 (2011).

[269] Alberts, B. *et al.* Molecular biology of the cell. In *Molecular Biology of the Cell* (2002).

[270] Hermsen, R., Tans, S. & Ten Wolde, P. R. Transcriptional regulation by competing transcription factor modules. *PLoS computational biology* **2** (2006).

[271] Panne, D. The enhanceosome. *Current Opinion in Structural Biology* **18**, 236–242 (2008).

[272] Ezer, D., Zabet, N. R. & Adryan, B. Physical constraints determine the logic of bacterial promoter architectures. *Nucleic acids research* **42**, 4196–4207 (2014).

[273] Hoch, M., Gerwin, N., Taubert, H. & Jackle, H. Competition for overlapping sites in the regulatory region of the drosophila gene kruppel. *Science* **256**, 94–97 (1992).

[274] Guturu, H., Doxey, A. C., Wenger, A. M. & Bejerano, G. Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20130029 (2013).

[275] Fry, C. J. & Farnham, P. J. Context-dependent transcriptional regulation. *Journal of Biological Chemistry* **274**, 29583–29586 (1999).

[276] Stanojevic, D., Small, S. & Levine, M. Regulation of a segmentation stripe by overlapping activators and repressors in the drosophila embryo. *Science* **254**, 1385–1387 (1991).

[277] Ptashne, M. Principles of a switch. *Nature Chemical Biology* **7**, 484–487 (2011).

[278] Dhiman, A. & Schleif, R. Recognition of overlapping nucleotides by arac and the sigma subunit of RNA polymerase. *Journal of Bacteriology* **182**, 5076–5081 (2000).

[279] Humphrey, G. W., Englander, E. W. & Howard, B. H. Specific binding sites for a pol iii transcriptional repressor and pol ii transcription factor yy1 within the internucleosomal spacer region in primate alu repetitive elements. *Gene Expression The Journal of Liver Research* **6**, 151–168 (1996).

Bibliography

[280] Lee, N. L., Gielow, W. O. & Wallace, R. G. Mechanism of arac autoregulation and the domains of two overlapping promoters, pc and pbad, in the l-arabinose regulatory region of escherichia coli. *Proceedings of the National Academy of Sciences* **78**, 752–756 (1981).

[281] Lee, J. Y. *et al.* Multiple sgrnas with overlapping sequences enhance crispr/cas9-mediated knock-in efficiency. *Experimental & Molecular Medicine* **50**, 1–9 (2018).

[282] McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research* **41**, e57–e57 (2013).

[283] Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).

[284] Barbaric, S., Munsterootter, M., Goding, C. & Horz, W. Cooperative pho2-pho4 interactions at the pho5 promoter are critical for binding of pho4 to uasp1 and for efficient transactivation by pho4 at uasp2. *Molecular and Cellular Biology* **18**, 2629–2639 (1998).

[285] Barbaric, S., Münsterkötter, M., Svaren, J. & Hörz, W. The homeodomain protein pho2 and the basic-helix-loop-helix protein pho4 bind DNA cooperatively at the yeast pho5 promoter. *Nucleic Acids Research* **24**, 4479–4486 (1996).

[286] Fascher, K.-D., Schmitz, J. & Hörz, W. Role of trans-activating proteins in the generation of active chromatin at the pho5 promoter in s. cerevisiae. *The EMBO Journal* **9**, 2523–2528 (1990).

[287] Nourani, A., Utley, R. T., Allard, S. & Côté, J. Recruitment of the nua4 complex poises the pho5 promoter for chromatin remodeling and activation. *The EMBO Journal* **23**, 2597–2607 (2004).

[288] Korber, P. & Barbaric, S. The yeast pho5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Research* **42**, 10888–10902 (2014).

[289] Azpeitia, E. & Wagner, A. Short residence times of DNA-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system. *Frontiers in Molecular Biosciences* **7**, 67 (2020).

[290] Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome research* **16**, 962–972 (2006).

[291] Bashor, C. J. *et al.* Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* **364**, 593–597 (2019).

[292] Laohakunakorn, N., Lavickova, B., Swank, Z., Laurent, J. & Maerkl, S. Steady-state cell-free gene expression with microfluidic chemostats. *protocols.io* (2019). Https://dx.doi.org/10.17504/protocols.io.46wgzfe.

[293] Newville, M., Stensitzki, T., Allen, D. & Ingargiola, A. Lmfit: Non-linear least-square minimization and curve-fitting for python, doi: 10.5281/zenodo. 11813 (2014).

[294] Phillips, R., Kondev, J., Theriot, J., Garcia, H. G. & Orme, N. *Physical biology of the cell* (Garland Science, 2012).

[295] McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic acids research* **41**, e57–e57 (2012).

SHIYU CHENG

Ecole Polytechnique Federale de Lausanne, Station 17, BM 2111, CH 1015, Lausanne, Switzerland
(+41) 077 918 6057 / shiyu.cheng@epfl.ch / chengshiyuws@gmail.com

EDUCATION

Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland *10/2018 - 09/2023*
Doctor of Philosophy in Biotechnology and Bioengineering
School of Engineering, Advisor: Prof. Sebastian Maerkl

University of Chinese Academy of Sciences (UCAS), China *09/2015 - 03/2017*
Master of Science in Polymer Chemistry and Physics (dropout, on 20/3/2017)
School of Chemistry and Chemical Engineering, Advisor: Prof. Xingyu Jiang

Tianjin University (TJU), China *09/2011 - 07/2015*
Bachelor of Engineering in Materials/Polymer Science and Engineering (obtained, on 13/07/2015)
School of Materials Science and Engineering, Advisor: Prof. Yongtao Shen & Prof. Xubo Yuan

RESEARCH EXPERIENCE

Doctoral Assistant, School of Engineering (EPFL), Advisor: Prof. Sebastian Maerkl *10/2018 - 09/2023*
Topic: Quantitative mapping the yeast phosphate gene regulatory network and reverse engineering an orthogonal GRN
Research Assistant, National Center for Nanoscience & Technology, Advisor: Prof. Xingyu Jiang *07/2014 - 09/2018*
Topic: Polymeric vascular scaffold & metal-polymer conductor enabled implantable bioelectronics
Research Assistant, School of Materials Science & Engineering (TJU), Advisor: Prof. Yongtao Shen *09/2012 - 06/2014*
Topic: Controllable preparation of carbon nanoscroll via solvent sonication method

RESEARCH INTERESTS

Gene Regulatory Network Engineering, Genetic Engineering, Biomedical Engineering

PEER-REVIEWED PUBLICATIONS

(Note: #Equal contribution)

14. Hon Ming Yip#, Shiyu Cheng#, Evan J. Olson#, Michael Crone, and Sebastian J. Maerkl*. Perfect adaptation achieved by transport limitations governs the inorganic phosphate response in S. cerevisiae. **PNAS**, 120 (2), e2212151120 (2023).

13. Amir Shahein, Maria Lopez, Ivan Istomin, Evan J. Olson, Shiyu Cheng, Sebastian Maerkl*. Systematic analysis of low-afinity transcription factor binding site clusters establishes their functional relevance in vitro and in vivo, **Nature Communications**, 12, 5273 (2022).

12. Chunliang Zhang, Qian Xie, Ruitao Cha, Li Ding, Liujun Jia, Lei Mou, Shiyu Cheng, Nuoxin Wang, Zulan Li, Yang Sun, Chuanjue Cui, Yu Zhang, Yan Zhang, Fengshan Zhou, Xingyu Jiang*. Anticoagulant hydrogel tubes with poly(e-caprolactone) sheaths for small-diameter vascular grafts, **Advanced Healthcare Materials**, 2100839 (2021).

11. Simoni Giaveri, Adeline M. Schmitt, Laura Roset Julia, Anna Murello, Vincenzo Scamarcio, Laure Menin, Shiyu Cheng, Daniel Ortiz, Luc Patiny, Sreenath Bolisetty, Raffaele Mezzenga, Sebastian J. Maerkl*, Francesco Stellacci*. Nature-inspired Circular-economy Recycling (NaCRe) for Proteins: Proof of Concept, **Advanced Materials**, 2104581 (2021).

10. Chen Hang#, Li Ding#, Shiyu Cheng#, Jie Qi, Xiaoyan Liu, Qian Liu, Yan Zhang, Xingyu Jiang. A soft, absorbable temporary epicardial pacing wire, **Advanced Materials**, 2101447 (2021).

9. Li Ding#, Chen Hang#, Shiyu Cheng#, Liujun Jia, Lei Mou, Lixue Tang, Chunliang Zhang, Yangzhouyun Xie, Wenfu Zheng, Yan Zhang, Xingyu Jiang. A soft, conductive external stent inhibits intimal hyperplasia in vein grafts by electroporation and mechanical restriction, **ACS Nano**, 14, 16770-16780 (2020).

8. Shiyu Cheng#, Chen Hang#, Li Ding#, Liujun Jia, Lixue Tang, Lei Mou, Jie Qi, Ruihua Dong, Wenfu Zheng, Yan Zhang, Xingyu Jiang. Electronic blood vessel, **Matter (Cell Press)**, 3, 1664-1684 (2020). (Featured in Eurekalert! (AAAS), IEEE Spectrum, Science Daily, Medicalxpress, New Scientist, New atlas, etc.)

7. Ruihua Dong, Xiaoyan Liu, Shiyu Cheng, Lixue Tang, Mian Chen, Leni Zhong, Zhen Chen, Shaoqin Liu, and Xingyu Jiang. Highly stretchable, biocompatible metal-polymer conductor electrode array for electrophysiology, **Advanced Healthcare Materials**, 2000641 (2020).

6. Lixue Tang, Shiyu Cheng, Luyao Zhang, Hanbing Mi, Lei Mou, Shuaijian Yang, Zhiwei Huang, Xinghua Shi, Xingyu Jiang. Printable metal-polymer conductors for highly stretchable bio-devices, **iScience**, 4, 302311 (2018). (Front cover, Lead author in device characterization & bio-applications)

5. Nuoxin Wang, Yunhu Peng, Wenfu Zheng, Lixue Tang, Shiyu Cheng, Junchuan Yang, Shaoqin Liu, Wei Zhang, Xingyu Jiang. A strategy for rapid construction of blood vessel-like structures with complex cell alignments, **Macromolecular**

Bioscience, 18, 1700408 (2018). (Front cover)

4. Shiyu Cheng#, Yu Jin#, Nuoxin Wang, Feng Cao, Wei Zhang, Wei Bai, Wenfu Zheng, Xingyu Jiang. Self-adjusting, polymeric multi-layered roll that can keep the shapes of the blood vessel scaffolds during biodegradation, **Advanced Materials**, 29 (28), 201700171 (2017). (Frontispiece, Featured in Advanced Science News, Nature Reviews Materials)

3. Nuoxin Wang, Wenfu Zheng, Shiyu Cheng, Wei Zhang, Shaoqiu Liu, Xingyu Jiang. In vitro evacuation of essential mechanical properties and cell behaviors of a novel PLGA-based tubular scaffold for small-diameter vascular tissue engineering, **Polymers**, 9 (8), 318 (2017).

2. Nuoxin Wang, Lixue Tang, Wenfu Zheng, Shiyu Cheng, Yifeng Lei, Lingmin Zhang, Binfeng Hu, X. Liu, Wei Zhang, Xingyu Jiang. A strategy for rapid and facile fabrication of controlled, layered blood vessel-like structures, **RSC Advances**, 6, 55054 (2016).

1. Yongtao Shen, Ke Deng, Songlin Yang, Bo Qin, Shiyu Cheng, Ningbo Zhu, Jiejin Ding, Dahui Zhao, Ji Liu, Qingdao Zeng and Chen Wang. Triangular-shaped molecular random tiling and molecular rotation in two-dimensional glassy networks, **Nanoscale**, 6, 7221 (2014).

REVIEWS & CHAPTERS

1. Shiyu Cheng, Jinqi Deng, Wenfu Zheng, Xingyu Jiang, Microfluidics for Biomedical Applications. In R. Narayan (Ed.), Encyclopedia of Biomedical Engineering, vol. 1, pp. 368-383. Elsevier (2019). [link]

TEACHING

2019-2020 Instructor of EPFL iGEM Project: Espress'EAU, On-site water quality test, Gold Metal;

2018-2019 Instructor of EPFL iGEM Project: ViTEST, Field-deployable cell-free detection of grapevine phytoplasma diseases, Grand Prize Winner (Overgrad, #1 out of 375 teams), Best Environment Project & Best Integrated Human Practices, 5 Nominations, Gold Medal, Featured in EPFL News.

CONFERENCES ORGANIZED

2019 EPFL-ETHZ Summer School: Quantitative Biology: Bridging the Gap between Quantitative and Experimental Approaches, Fiesch, Switzerland. (Co-organizer: Zoe Swank, Ming Yip, Jan Mikelson, Joaquin Gutierrez Mena)

CONFERENCES ATTENDED

2019 Nov 27, Lausanne (Switzerland), EPFL Bioengineering Day 2019, Poster presentation.
2017 Sept 27-30, Yokohama (Japan), A3 Foresight Symposium 2017 on Nanoscale Imaging of Cancers, Poster presentation.
2017 Aug 29-31, Beijing, The 7th International Conference on Nanosci & Tech (ChinaNANO 2017), Poster presentation.
2016 Jul 31-Aug 3, Beijing, A3 Foresight Symposium 2016 on Nanoscale Imaging of Cancers, Volunteer.
2016 Jul 8-11, Beijing, Precision Nanomedicine Symposium 2016, Volunteer.
2015 Sept 3-5, Beijing, The 6th International Conference on Nanoscience and Technology (ChinaNANO 2015), Volunteer.
2014 Aug 4-7, Beijing, The 29th Chinese Chemical Society (CCS) Congress, Attendee.

AWARDS/HONORS

2018-2019   iGEM2019, Grand Prize Winner (Overgrad);
2016-2017   Award for First Class Scholarship for Graduate Students (top 3%), NCNST, CAS;
2012-2014   School Excellent Award, Students Innovation and Entrepreneurship Training Program, TJU;
2012-2014   Award for ZHENGDONG Innovation Scholarship, TJU;
2012-2013   Award for Advanced Individual in Technological Innovations, TJU;

ACADEMIC REFEREES

Sebastian Josef Maerkl, Ph.D., Associate Professor (Advisor, PhD Research)
Institute of Bioengineering, School of Engineering, Ecole Polytechnique Federale de Lausanne (EPFL)
Address: Station 17, BM 5130, Lausanne, CH-1015, Switzerland
Email: sebastian.maerkl@epfl.ch; sebastian.maerkl@gmail.com
URL: http://lbnc.epfl.ch

Xingyu Jiang, Ph.D., Chair Professor & Department Head (Advisor, Masters Research)
Department of Biomedical Engineering, Southern University of Science and Technology (SUSTech), Shenzhen
Address: Faculty Research Building 2, 1088 Xueyuan Rd, Nanshan District, Shenzhen, Guangdong 518055, China
Email: xingyujiang@nanoctr.cn; jiang@sustech.edu.cn
URL: http://www.jiangxingyu.com