DETECTING AND MODIFYING CHROMATIN ORGANIZATION IN THE BUDDING YEAST GENOME WITH SYNTHETIC PROTEINS

by

ORION GARRETT BAXTER BANKS

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2022

DISSERTATION APPROVAL PAGE

Student: Orion Garrett Baxter Banks

Title: Detecting and Modifying Chromatin Organization in the Budding Yeast Genome with Synthetic Proteins

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Chemistry and Biochemistry by:

Victoria DeRose	Chairperson
Michael Harms	Advisor
Laura McKnight	Core Member
Diane Hawley	Core Member
Mathew Barber	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded December 2022

DISSERTATION ABSTRACT

Orion Garrett Baxter Banks

Doctor of Philosophy

Department of Chemistry and Biochemistry

December 2022

Title: Detecting and Modifying Chromatin Organization in the Budding Yeast Genome with Synthetic Proteins

DNA in the nucleus of eukaryotic cells is condensed into chromatin by forming specialized structures called nucleosomes. DNA is wrapped around a set of histone proteins to from nucleosomes, which are eventually positioned by robust, reproducible mechanisms. Mutations in chromatin organizing proteins are common in human diseases, indicating that studying the systems in place to control chromatin organization could be insightful for developing novel therapeutics. *S. cerevisiae* is an ideal model organism for this work, as it is tolerant to dysregulated nucleosome positioning, genetically tractable, and contains homologs of many human proteins. Here, I describe several studies that have furthered our understanding of chromatin organization in the context of nucleosome positioning. First, I detail a set of synthetic proteins that enable nucleosomes to be repositioned at characteristic sites in the yeast genome. Following this is a detailed account of our investigation into the biochemical mechanisms underlying targeted chromatin remodeling at a subset of genome loci by the chromatin remodeler Isw2. Next, a protocol for efficiently detecting nucleosome positions is described, including its application to yeast deficient in chromatin remodeling activity. In chapter five, I present a system for detecting protein-DNA interactions and nucleosome positions using targeted

nuclease digestion with endogenously expressed proteins. I also outline some of the potential applications of this technique, including use with chromatin remodeling proteins, which often prove difficult to map with standard methods.

This dissertation contains previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Orion Garrett Baxter Banks

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Western Washington University, Bellingham, WA

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2022, University of Oregon
Master of Science, Chemistry, 2017, Western Washington University
Bachelor of Science, Biochemistry, 2015, Western Washington University

AREAS OF SPECIAL INTEREST:

Biochemistry
Molecular Biology
Computational Biology
Genomics
Protein Engineering

PROFESSIONAL EXPERIENCE:

Research assistant, Western Washington University (Smirnov lab), 2014-2015

Graduate research and teaching assistant, Western Washington University (Antos lab), 2015-2017

Graduate teaching assistant, University of Oregon, 2017-2018

Graduate research assistant, University of Oregon, 2018-2022

GRANTS, AWARDS, AND HONORS:

Knight Campus Undergraduate Scholars Petrone Scholar, University of Oregon, 2020

Adamson Family First Year Graduate Student award, University of Oregon Institute of Molecular Biology, 2018

Molecular Biology and Biophysics training grant (NIH T32 GM007759), University of Oregon, 2018 - 2020

Graduate Teaching Excellence in Biochemistry award, University of Oregon Department of Chemistry and Biochemistry, 2018

PUBLICATIONS:

McKnight LE, Crandall JG, Bailey TB, **Banks OGB**, Orlandi KN, Truong VN, Donovan DA, Waddell GL, Wiles ET, Hansen SD, Selker EU, McKnight JN. Rapid and inexpensive preparation of genome-wide nucleosome footprints from model and non-model organisms. *STAR Protoc*. (2021) doi: 10.1016/j.xpro.2021.100486

Donovan DA, Crandall JG, Truong VN, Vaaler AL, Bailey TB, Dinwiddie D, **Banks OG**, McKnight LE, McKnight JN. Basis of specificity for a conserved and promiscuous chromatin remodeling protein. *eLife* (2021) doi: 10.7554/eLife.64061.

Donovan DA, Crandall JG, **Banks OGB**, Jensvold ZD, Truong V, Dinwiddie D, McKnight LE, McKnight JN. Engineered Chromatin Remodeling Proteins for Precise Nucleosome Positioning. *Cell Rep*. (2019) doi: 10.1016/j.celrep.2019.10.046.

Nikghalb KD, Horvath NM, Prelesnik JL, **Banks OGB**, Filipov PA, Row RD, Roark TJ, Antos JM. Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *Chembiochem* (2018) doi: 10.1002/cbic.201700517.

Banks, Orion. Towards the Substrate-bound Structure of Streptococcus pneumoniae Sortase A. *WWU Graduate School Collection* (2017) https://cedar.wwu.edu/wwuet/609

ACKNOWLEDGMENTS

I am incredibly grateful to the members of my committee for providing support during my time at University of Oregon. I would like to thank Laura McKnight for her guidance and wisdom, which I found uplifting in difficult times. I am grateful to Mike Harms for his willingness to co-advise me – I learned a tremendous amount from our discussions. I would also like to thank the Genomics Core (GC3F) at University of Oregon for high throughput sequencing services. Last, I would like to thank Ashley Banks for her commitment to being my partner in life and her belief in me. This work was supported by a National Institutes of Health training grant T32 GM007759 (to O.G.B.B), and by NIGMS R01 GM129242 (to J.N.M. and L.E.M).

For Jeff McKnight, whose compassion, humility, and brilliance will never be forgotten. I am fortunate to have known him.

TABLE OF CONTENTS

Chapter Page

Chapter Page

APPENDICES

LIST OF FIGURES

Figure Page

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Chromatin is the intricate web of DNA and proteins that enable eukaryotes to condense and store DNA in the nucleus. Condensing DNA into fibrous strands occurs during mitosis, however, chromatin more often exists in a less condensed state during the remainder of a cell's life cycle (Lieberman-Aiden et al., 2009). Chromatin self-associates to build several levels of structure and can often described as open (euchromatin) or closed (heterochromatin). Euchromatin is most often found in actively transcribed regions of the genome, while heterochromatin is typically found in non-transcribed regions (Cremer at al., 2015). Researchers have observed large changes in chromatin compaction during mammalian (Hota and Bruneau, 2016) and plant (van Zanten et al., 2012) development, and in response to environmental stressors in yeast (Gasch et al., 2000). Aberrant chromatin structure is common in human diseases, and mutations affecting the organization of chromatin are often observed in cancer (Zane et al., 2014) and neurodegenerative disorders (Berson et al., 2018).

Histones are the group of well-conserved DNA binding proteins that associate with DNA to form chromatin. Histones are highly positively charged, which enables them to non-specifically interact with negatively-charged phosphate groups in the backbone of DNA. Historical experiments using x-ray diffraction data described a repeating structure of histone-DNA complexes involving 200bp of DNA (Kornberg, 1974). The crystal structure of the nucleosome core particle shows that nucleosomes are composed of 147bp of DNA wrapped around a histone octamer core (Luger et al., 1996). Histone cores are typically composed from two monomers each of H2A, H2B, H3, and

H4, but can also contain histone variants, which tune the interaction of the histone core with DNA or mark nucleosomes at specific loci (Martire and Banaszynski, 2020). The N-terminal amino acids of histones (known as histone tails) are highly positively charged and known to interact with DNA and stabilize the nucleosome (Szerlong and Hansen, 2011). Histone tails are frequently modified with various post-translation modifications that function as organizational markers, particularly with respect to transcriptional activity, but also DNA repair and chromatin structure inheritance (Cosgrove et al., 2004; Kouzarides, 2007).

Nucleosome positions in chromatin are often observed as "phased" arrays, where nucleosomes are evenly spaced with respect to one another (Singh et al., 2021). Transcription start sites (TSSs) are commonly associated with organized chromatin (Lai et al., 2018), where nucleosome position maps show a pattern of peaks where signal decays as it moves into the gene body. Origins of replication, marked by autonomous replication sequences in yeast, are also found at the center of regions that display evenly spaced nucleosomes (Fennessy and Owen-Hughes, 2016). Additionally, transcription factor (TF) binding sites are commonly associated with nucleosome phasing and TFs compete with nucleosomes to re-bind DNA following replication (Kasinathan et al., 2014; Adams and Workman, 1995; Venkatesh and Workman, 2015).

Previous work has shown that specific nucleosome positions derived from whole-genome analysis cannot be exactly recapitulated in vitro (Kaplan et al., 2008; Krietenstein et al., 2016). Interestingly, when nucleosomes interact with genomic DNA in salt-gradient dialysis (Thåström et al., 2004), some positions are similar to those seen in vivo (Thåström et al., 2004; Kaplan et al., 2008; Krietenstein et al., 2016), suggesting that

genomic DNA sequences are inherently capable of self-patterning nucleosomes to some extent. However, treating these chromatin templates with whole cell extracts is not sufficient to produce all the observed elements of chromatin organization in live cells (Krietenstein et al., 2016). This result is consistent with many observations that replication and transcription directly influence chromatin organization in vivo (Radman-Livaja et al., 2011; Ramachandran and Henikoff, 2016; Vasseur et al., 2016; Stewart-Morgan et al., 2019).

Importantly, neither nucleosome positions nor target sequences alone can fully predict where DBPs bind in the genome (Rhee and Pugh, 2011; Guertin and Lis, 2013). A subset of TFs in several model systems have been shown to target binding sites regardless of restrictions implied by local nucleosome patterning (Balsalobre and Drouin, 2022). These pioneering factors (PFs) can interact with their target sites independently of a nucleosome positioned on or near the sequence (Meers et al., 2019). The occupancy of a PF at a site or the ability of the PF to invade an obstructed target sequence aligns with the similarity of the motif to its prototypical consensus sequence (Meers et al., 2019; Echigoya et al., 2020). PFs and their yeast analogs, general regulatory factors (GRFs) achieve chromatin organization by interfering with nucleosome deposition, but also by recruiting chromatin remodeling proteins (Voss and Hager, 2013).

This dissertation contains previously published co-authored material.

Chapter II:

Donovan, D.A., Crandall, J.G., **Banks, O.G.B.**, Jensvold, Z.D., Truong, V., Dinwiddie, D., McKnight, L.E., McKnight, J.N. (2019). Engineered chromatin remodeling proteins for precise nucleosome positioning. *Cell Reports* 29:2520–2535.

Chapter III:

Donovan, D.A., Crandall, J.G., Truong, V.N., Vaaler, A.L., Bailey, T.B., Dinwiddie, D., **Banks, O.G.B.**, McKnight, L.E., McKnight, J.N. (2021). Basis of Specificity for a Conserved and Promiscuous Chromatin Remodeling Protein. *eLife* 10:e64061

Chapter IV:

McKnight LE, Crandall JG, Bailey TB, **Banks OGB**, Orlandi KN, Truong VN, Donovan DA, Waddell GL, Wiles ET, Hansen SD, Selker EU, McKnight JN. (2021). Rapid and inexpensive preparation of genome-wide nucleosome footprints from model and non-model organisms. *STAR Protoc*. 2(2):100486.

Bridge from Chapter I to II

Considering the relationship between nucleosome positions, DBP binding, and transcription, there is much interest in developing systems to analyze the cellular response to mispositioned or ectopic nucleosomes. Previous work by Jeff McKnight and his colleagues (McKnight et al., 2011) showed that engineered chromatin remodeling proteins (E-ChRPs) were well-suited to accomplish this task. The catalytic domain of the general chromatin remodeler Chd1 was fused to the DNA-binding domain of AraC, which enabled nucleosomes to be repositioned at target sites. While Chd1 typically helps create evenly-spaced arrays of nucleosomes, it was observed that nucleosomes were pulled into atypical positions by the chimeric remodeler. Eventually, this protein fusion system was adapted for use in yeast, where targeted ectopic nucleosome positions were observed using MNase-Seq (McKnight et al., 2016).

As outlined in Chapter II, we expanded on this work by designing novel E-ChRPs using an expanded set of DBDs and proteins fused to the ATPase subunit of Chd1. I was directly involved with analyzing the galactose-inducible forms of Chd1-Ume6 and Chd1-SpyCatcher. To accomplish this, I performed MNase digestions on yeast samples where

expression of Chd1-SpyCatcher was induced in the presence of SpyTagged TFs. I also performed ChIP-Seq on samples from the Chd1-SpyCatcher system, which allowed us to evaluate the behavior of SpyTagged TFs when ectopic nucleosomes were positioned near or within their consensus sequences. After collecting MNase-Seq and ChIP-Seq samples, I generated several of the DNA fragment libraries used in Next-Generation Sequencing analysis. The sections provided establish the use of fusion proteins as tools for repositioning nucleosomes and provide a basis for the protein fusions used in chapter V and VI, which are based on the SpyCatcher-Spytag system.

References Cited

Adams CC, Workman JL. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol*. 1995;15(3):1405-1421. doi:10.1128/MCB.15.3.1405

Balsalobre A, Drouin J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat Rev Mol Cell Biol*. 2022 Jul;23(7):449-464. doi: 10.1038/s41580-022-00464-z. Epub 2022 Mar 9. PMID: 35264768.

Berson A, Nativio R, Berger SL, Bonini NM. Epigenetic Regulation in Neurodegenerative Diseases. *Trends Neurosci*. 2018;41(9):587-598. doi:10.1016/j.tins.2018.05.005

Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol*. 2004 Nov;11(11):1037-43. doi: 10.1038/nsmb851. PMID: 15523479.

Cremer T, Cremer M, Hübner B, Strickfaden H, Smeets D, Popken J, Sterr M, Markaki Y, Rippe K, Cremer C. The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett*. 2015 Oct 7;589(20 Pt A):2931-43. doi: 10.1016/j.febslet.2015.05.037.

Echigoya K, Koyama M, Negishi L, Takizawa Y, Mizukami Y, Shimabayashi H, Kuroda A, Kurumizaka H. Nucleosome binding by the pioneer transcription factor OCT4. *Sci Rep*. 2020 Jul 16;10(1):11832. doi: 10.1038/s41598-020-68850-1. PMID: 32678275; PMCID: PMC7367260.

Fennessy RT, Owen-Hughes T. Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. *Nucleic Acids Res*. 2016 Sep 6;44(15):7189-203. doi: 10.1093/nar/gkw331. Epub 2016 Apr 22. PMID: 27106059; PMCID: PMC5009725.

Gasch AP, Spellman PT, Kao CM, et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell*. 2000;11(12):4241-4257. doi:10.1091/mbc.11.12.4241

Guertin MJ, Lis JT. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev*. 2013;23(2):116-123. doi:10.1016/j.gde.2012.11.008

Hota SK, Bruneau BG. ATP-dependent chromatin remodeling during mammalian development. *Development*. 2016;143(16):2882-2897. doi:10.1242/dev.128892

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*. 2009 Mar 19;458(7236):362-6. doi: 10.1038/nature07667. Epub 2008 Dec 17. PMID: 19092803; PMCID: PMC2658732.

Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods*. 2014 Feb;11(2):203-9. doi: 10.1038/nmeth.2766. Epub 2013 Dec 15. PMID: 24336359; PMCID: PMC3929178.

Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974 May 24;184(4139):868-71. doi: 10.1126/science.184.4139.868. PMID: 4825889.

Kouzarides T. Chromatin modifications and their function. *Cell*. 2007 Feb 23;128(4):693-705. doi: 10.1016/j.cell.2007.02.005. PMID: 17320507.

Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF, Korber P. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*. 2016 Oct 20;167(3):709-721.e12. doi: 10.1016/j.cell.2016.09.045. PMID: 27768892; PMCID: PMC5240917.

Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, Hu G, Ni B, Zhao K. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature*. 2018 Oct;562(7726):281-285. doi: 10.1038/s41586-018-0567-3. Epub 2018 Sep 26. Erratum in: *Nature*. 2018 Nov 6;: PMID: 30258225; PMCID: PMC8353605.

Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi:10.1126/science.1181369

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 1997 Sep 18;389(6648):251-60. doi: 10.1038/38444. PMID: 9305837.

Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol*. 2020 Sep;21(9):522-541. doi: 10.1038/s41580-020-0262-8. Epub 2020 Jul 14. PMID: 32665685; PMCID: PMC8245300.

McKnight JN, Jenkins KR, Nodelman IM, Escobar T, Bowman GD. Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Mol Cell Biol*. 2011 Dec;31(23):4746-59. doi: 10.1128/MCB.05735-11. Epub 2011 Oct 3. PMID: 21969605; PMCID: PMC3232923.

McKnight JN, Tsukiyama T, Bowman GD. Sequence-targeted nucleosome sliding in vivo by a hybrid Chd1 chromatin remodeler. *Genome Res*. 2016 May;26(5):693-704. doi: 10.1101/gr.199919.115. Epub 2016 Mar 18. PMID: 26993344; PMCID: PMC4864466.

Meers MP, Janssens DH, Henikoff S. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Mol Cell*. 2019;75(3):562-575.e5. doi:10.1016/j.molcel.2019.05.025

Radman-Livaja M, Verzijlbergen KF, Weiner A, van Welsem T, Friedman N, Rando OJ, van Leeuwen F. Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biol*. 2011 Jun;9(6):e1001075. doi: 10.1371/journal.pbio.1001075. Epub 2011 Jun 7. PMID: 21666805; PMCID: PMC3110181.

Ramachandran S, Henikoff S. Transcriptional Regulators Compete with Nucleosomes Post-replication. *Cell*. 2016 Apr 21;165(3):580-92. doi: 10.1016/j.cell.2016.02.062. Epub 2016 Apr 7. PMID: 27062929; PMCID: PMC4855302.

Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011 Dec 9;147(6):1408-19. doi: 10.1016/j.cell.2011.11.013. PMID: 22153082; PMCID: PMC3243364.

Singh AK, Schauer T, Pfaller L, Straub T, Mueller-Planitz F. The biogenesis and function of nucleosome arrays. *Nat Commun*. 2021 Dec 1;12(1):7011. doi: 10.1038/s41467-021-27285-6. PMID: 34853297; PMCID: PMC8636622.

Stewart-Morgan KR, Reverón-Gómez N, Groth A. Transcription Restart Establishes Chromatin Accessibility after DNA Replication. *Mol Cell*. 2019 Jul 25;75(2):284-297.e6. doi: 10.1016/j.molcel.2019.04.033. Epub 2019 May 21. Erratum in: *Mol Cell*. 2019 Jul 25;75(2):408-414. PMID: 31126739.

Szerlong HJ, Hansen JC. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem Cell Biol*. 2011 Feb;89(1):24-34. doi: 10.1139/O10-139. PMID: 21326360; PMCID: PMC3125042.

Thåström A, Bingham LM, Widom J. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *J Mol Biol*. 2004 May 7;338(4):695-709. doi: 10.1016/j.jmb.2004.03.032. PMID: 15099738.

van Zanten M, Carles A, Li Y, Soppe WJ. Control and consequences of chromatin compaction during seed maturation in Arabidopsis thaliana. *Plant Signal Behav*. 2012;7(3):338-341. doi:10.4161/psb.19281

Vasseur P, Tonazzini S, Ziane R, Camasses A, Rando OJ, Radman-Livaja M. Dynamics of Nucleosome Positioning Maturation following Genomic Replication. *Cell Rep*. 2016 Sep 6;16(10):2651-2665. doi: 10.1016/j.celrep.2016.07.083. Epub 2016 Aug 25. PMID: 27568571; PMCID: PMC5014762.

Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol*. 2015 Mar;16(3):178-89. doi: 10.1038/nrm3941. Epub 2015 Feb 4. PMID: 25650798.

Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet*. 2014 Feb;15(2):69-81. doi: 10.1038/nrg3623. Epub 2013 Dec 17. PMID: 24342920; PMCID: PMC6322398.

Zane L, Sharma V, Misteli T. Common features of chromatin in aging and cancer: cause or coincidence?. *Trends Cell Biol*. 2014;24(11):686-694. doi:10.1016/j.tcb.2014.07.001

CHAPTER II

E-CHRPS: ENGINEERED CHROMATIN REMODELING PROTEINS FOR PRECISE NUCLEOSOME POSITIONING

*This chapter contains previously published co-authored material.

Donovan, D.A., Crandall, J.G., **Banks, O.G.B.**, Jensvold, Z.D., Truong, V., Dinwiddie, D., McKnight, L.E., McKnight, J.N. (2019). Engineered chromatin remodeling proteins for precise nucleosome positioning. Cell Reports 29:2520–2535.

Conceptualization, D.A.D., L.E.M., and J.N.M.; Methodology, D.A.D., J.G.C., O.G.B.B., Z.D.J., V.T., L.E.M., and J.N.M.; Investigation, D.A.D., J.G.C., O.G.B.B., Z.D.J., V.T., L.E.M., and J.N.M.; Writing – Original Draft, D.A.D., J.G.C., and J.N.M.; Writing – Review & Editing, D.A.D., J.G.C., Z.D.J., L.E.M., and J.N.M.; Visualization, D.D. and J.N.M.; Supervision, L.E.M. and J.N.M.; Project Administration, J.N.M.; Funding Acquisition, J.N.M.

Introduction

The nucleosome is the fundamental repeating unit of chromatin, composed of DNA wrapped around an octamer of histone proteins. While nucleosomes are dynamic structures that are constantly assembled, disassembled, and repositioned in the genome, their positions at gene regulatory elements like transcription start sites (TSSs) show characteristic organization (Lai and Pugh, 2017). Thus, nucleosome positions are thought to have regulatory implications for DNA-dependent processes like transcription, replication and DNA repair (Hauer and Gasser, 2017; MacAlpine and Almouzni, 2013; Venkatesh and Workman, 2015). Because positions of nucleosomes in the genome play a major role in determining DNA sequence accessibility, the ability to precisely manipulate

nucleosome positions would have profound implications for investigating and controlling DNA-dependent processes *in vivo*.

ATP-dependent chromatin remodeling factors couple the hydrolysis of ATP to the movement of nucleosomes along a fragment of DNA (Cairns et al., 1996; Fazzio and Tsukiyama, 2003; Langst et al., 1999; Smith and Peterson, 2005; Stockdale et al., 2006; Tsukiyama et al., 1994). By altering the positions of nucleosomes, this family of enzymes controls the accessibility of underlying DNA *in vivo,* thereby regulating DNA-dependent processes. The CHD and ISWI families of chromatin remodelers contain a conserved catalytic ATPase that drives chromatin remodeling by binding and hydrolyzing ATP (Zhou et al., 2016), and a C-terminal region that interacts with extranucleosomal DNA to modify the direction and outcome of nucleosome repositioning (Gangaraju and Bartholomew, 2007; Hota et al., 2013; McKnight et al., 2011; Ryan et al., 2011).

Previous work established that chromatin remodeling by *S. cerevisiae* Chd1 can be targeted to specific nucleosomes by replacing the native, nonspecific Chd1 DNA binding domain (DBD) with sequence-specific DBDs (McKnight et al., 2011; Nodelman and Bowman, 2013). We previously showed that hybrid Chd1 fusions with exogenous, sequence-specific DBDs predictably move nucleosomes onto their recruitment sequences *in vitro* (McKnight et al., 2011). We recently demonstrated that fusion of Chd1 to the Zn2Cys6 DBD from Ume6, a meiotic repressor from yeast, allows directed nucleosome positioning at target genes across the *S. cerevisiae* genome (McKnight et al., 2016).

Here we have simplified the customizable design and validated the function of sequence-targeted chromatin remodeling proteins using diverse targeting strategies. These engineered chromatin remodeling proteins (E-ChRPs) work with a wide variety of

targeting domains and can occlude target DNA sequences by precisely repositioning nucleosomes onto recruitment motifs. We show that E-ChRPs possessing transcription factor (TF) DNA binding domains can block binding of endogenous transcription factors by incorporating TF binding sites into nucleosomes genome-wide. E-ChRPs can also be directly recruited to DNA-associated TFs through SpyTag/SpyCatcher pairs (Zakeri et al., 2012) allowing for identification and occlusion of TF-bound genomic loci. Finally, we show that positioning of nucleosomes can be achieved by a dCas9-targeted E-ChRP using optimized, noncanonical gRNAs.

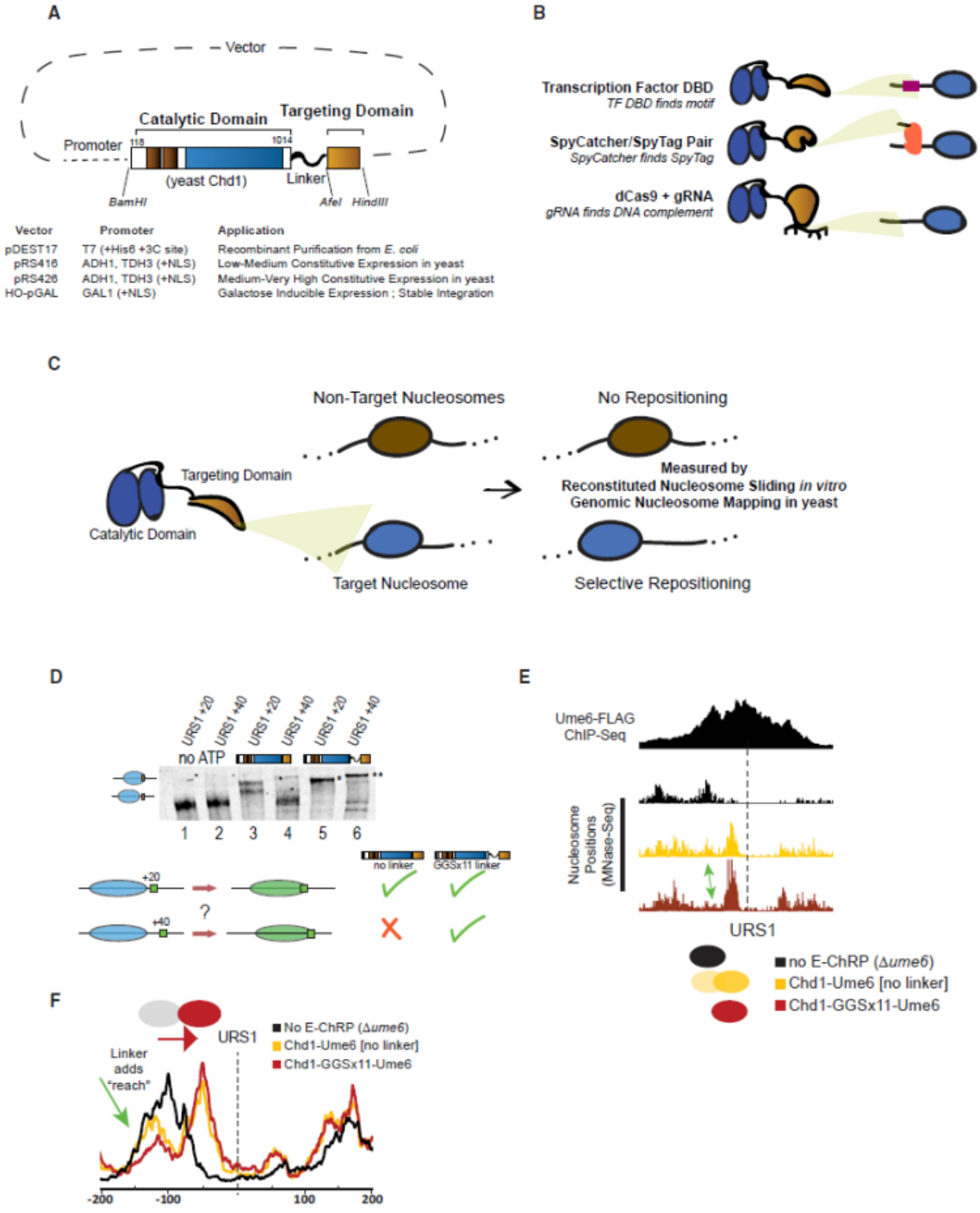
Results

Design

The core E-ChRP design was inspired by previous work (McKnight et al., 2011; McKnight et al., 2016) where individual sequence-specific DNA binding domains (DBDs) replaced the C-terminal nonspecific DNA binding domain of a functional *S. cerevisiae* Chd1 chromatin remodeler fragment (Fig 1A). Yeast Chd1 is an ideal enzyme for engineered chromatin remodeling because it is monomeric, displays robust nucleosome positioning activity on nucleosome substrates derived from multiple organisms, and is less influenced by histone modifications than other chromatin remodelers (Ferreira et al., 2007; Hauk et al., 2010). After the Chd1 catalytic module we incorporated unique restriction sites flanking the targeting domain in vectors allowing for recombinant expression in *E. coli*, constitutive expression from ADH1 or GPD promoters in *S. cerevisiae* (Mumberg et al., 1995), or galactose-inducible expression from the HO locus in *S. cerevisiae* (Voth et al., 2001). This scaffold allows

easy swapping of the C-terminal targeting domain resulting in a simple method to design chromatin remodelers that can be localized to desired nucleosomes. To demonstrate the versatility of the approach, we incorporated and assessed engineered chromatin remodeling through multiple transcription factor DNA binding domains, through SpyCatcher/SpyTag pairs, and through dCas9 targeting (Fig 1B). We first assessed the ability of different E-ChRPs to reposition target-containing mononucleosomes in a purified biochemical assay (Eberharter et al., 2004). To validate *in vivo* function, we introduced E-ChRPs into *S. cerevisiae* and measured global nucleosome positions by MNase-seq. Functional E-ChRPs can position targeted nucleosomes onto recruitment motifs as measured by mononucleosome sliding toward recruitment sequences *in vitro* or target motif occlusion by nucleosomes *in vivo* (Fig 1C).

Figure 1 (next page): Strategies for Optimizing Targeted Nucleosome Positioning by E-ChRPs (A) Architecture of the E-ChRP core where the yeast Chd1 catalytic domain is linked to a targeting domain with a flexible linker. **(B)** Summary of targeting methods used in this work, including sequence-specific DBD targeting to a recognition motif (top), SpyCatcher domain covalently attaching to a SpyTag-containing chromatin-bound protein (middle), and dCas9-bound gRNA interacting with a complementary sequence (bottom). **(C)** Predicted outcome from targeted E-ChRPs, indicating that select nucleosomes are positioned by the E-ChRP onto the recruitment site. **(D)** Nucleosome repositioning *in vitro* by Chd1-Ume6 with and without 11 repeats of glycine-glycine-serine between the Chd1 catalytic domain and Ume6 DBD. Nucleosomes with the Ume6 recognition motif (URS1) located 20 or 40 bp from the nucleosome edge were incubated with Chd1-Ume6(DBD) or Chd1-GGSx11-Ume6(DBD), and nucleosomes were resolved using native PAGE (top). Nucleosome positions before and after remodeling were resolved using 6% native PAGE. Summary of repositioning of each substrate by Chd1 fusions (bottom). **(E)** Genome Browser image showing nucleosome dyad positions at a representative Ume6 binding site (URS1) for a parental strain lacking endogenous Ume6 (black), after introduction of Chd1-Ume6 without (yellow) or with (red) a flexible linker. Dashed line indicates the location of the Ume6 binding motif. Arrow indicates nucleosome that is more efficiently positioned with a Chd1-GGSx11-Ume6(DBD) fusion. **(F)** Average nucleosome positioning (dyad signal) at genomic Ume6 binding sites showing Chd1-GGSx11-Ume6(DBD) more readily positions nucleosomes distal to the recognition element than Chd1-Ume6(DBD) lacking a flexible linker.


A
Vector
Catalytic Domain
Targeting Domain
Promoter
118
1014
(yeast Chd1)
Linker
BamHI
AfeI
HindIII
Vector
Promoter
Application
pDEST17
T7 (+His6 +3C site)
Recombinant Purification from E. coli
pRS416
ADH1, TDH3 (+NLS)
Low-Medium Constitutive Expression in yeast
pRS426
ADH1, TDH3 (+NLS)
Medium-Very High Constitutive Expression in yeast
HO-pGAL
GAL1 (+NLS)
Galactose Inducible Expression ; Stable Integration
B
Transcription Factor DBD
TF DBD finds motif
SpyCatcher/SpyTag Pair
SpyCatcher finds SpyTag
dCas9 + gRNA
gRNA finds DNA complement
C
Non-Target Nucleosomes
No Repositioning
Targeting Domain
Catalytic Domain
Measured by
Reconstituted Nucleosome Sliding in vitro
Genomic Nucleosome Mapping in yeast
Target Nucleosome
Selective Repositioning
D
URS1 +20
URS1 +40
no ATP
1
2
3
4
5
6
no linker
GGSx11 linker
+20
+40
?
E
Ume6-FLAG
ChIP-Seq
Nucleosome
Positions
(MNase-Seq)
URS1
no E-ChRP (Δume6)
Chd1-Ume6 [no linker]
Chd1-GGSx11-Ume6
F
URS1
No E-ChRP (Δume6)
Chd1-Ume6 [no linker]
Chd1-GGSx11-Ume6
Linker
adds
"reach"
-200
-100
100
200

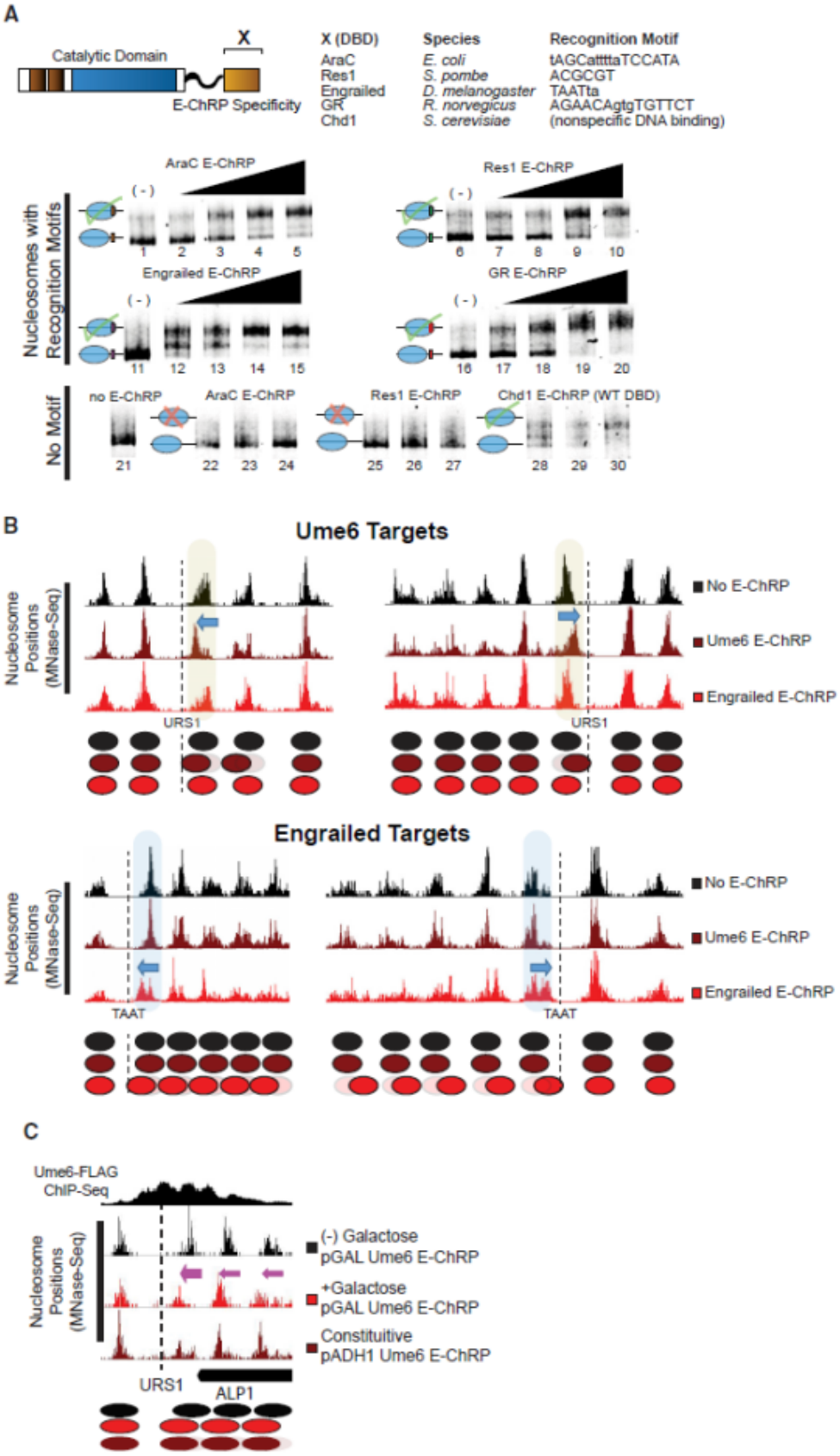
Development and Optimization of a Targeted Remodeler Core

We previously demonstrated that fusion of a foreign DNA binding domain to the Chd1 catalytic core leads to occlusion of a recruitment motif by targeted and directional repositioning of nucleosomes (McKnight et al., 2011; McKnight et al., 2016). Though functional both *in vitro* and *in vivo*, remodeler fusions where the DNA binding domain was directly fused to the Chd1 core resulted in a limited "reach" and nucleosomes residing further than 20 base pairs from the DNA recognition element were not efficiently moved. In addition, the creation of new remodeler fusion proteins was previously cumbersome and lacked versatility. To address these limitations we first created the E-ChRP scaffold (Fig 1A), which consists of the catalytic core of the yeast Chd1 protein followed by a flexible linker including 11 repeats of the glycine-glycine-serine sequence that was previously shown to extend the Chd1 reach (Nodelman and Bowman, 2013). We next created an array of plasmids for recombinant bacterial expression or yeast constitutive or inducible expression allowing for one-step cloning of a desired fusion domain (Fig 1A).

We examined whether the addition of a flexible linker between the Chd1 remodeler core and DNA binding domain increased the reach of these E-ChRPs. We tested the ability of an E-ChRP with a DBD from the *S. cerevisiae* meiotic repressor Ume6 to move mononucleosomes containing a recognition motif, URS1 (Park et al., 1992), 20 or 40 base pairs from the nucleosome edge (Fig 2A). Without a flexible linker, the Chd1-Ume6 E-ChRP was active exclusively when the motif was 20bp away (compare lanes 1 with 3, 2 with 4). In contrast, addition of eleven repeats of glycine-glycine-serine (GGSx11) allowed the remodeler to efficiently mobilize both nucleosome substrates

(compare lanes 1 with 5, 2 with 6). Additionally, the final location of the positioned nucleosomes was dependent on the location of the recognition motif (Fig 2A, compare lanes 5 and 6). Consistent with this increased reach *in vitro*, the Ume6 E-ChRP containing a GGSx11 linker positioned a larger fraction of distal nucleosomes onto target sequences across the *S. cerevisiae* genome (Fig 2B, C). Because the flexible linker led to more robust E-ChRP activity and our design was compatible *in vitro* and *in vivo*, we employed this general scaffold in all subsequent experiments.

Figure 2 (next page): E-ChRPs with Distinct TF DBDs Specifically Position Target Nucleosomes *In Vitro* and *In Vivo* (A) Nucleosome sliding assay demonstrating functionality of increasing concentrations of E-ChRPs containing AraC DBD (lanes 1–5 and 22–24), Res1 DBD (lanes 6–10 and 25–27), engrailed DBD (lanes 11–15), glucocorticoid receptor DBD (lanes 16–20), or Chd1 endogenous DBD (lanes 28–30) *in vitro*. Nucleosomes in lanes 1–20 possess recognition motifs in extranucleosomal DNA for the respective E-ChRP (Ades and Sauer, 1994; Alroy and Freedman, 1992; Anderson et al., 1995; Ayté et al., 1995; Khan et al., 2018; Niland et al., 1996), while lanes 21–30 have no recognition motif. Lower electrophoretic mobility indicates repositioning of nucleosomes away from their end positions.**(B)** Yeast genomic nucleosome dyad positions are shown at representative Ume6 targets (URS1, top) or engrailed targets (TAAT, bottom) in the presence or absence of Ume6 E-ChRP or engrailed E-ChRP. Motif-proximal nucleosomes are highlighted next to indicated motifs, with blue arrows showing direction of nucleosome movement. Cartoon representations of nucleosome positions are provided for each locus. **(C)** Nucleosome dyad signal at a representative locus in yeast demonstrating positioning of nucleosomes toward recruitment motif (dashed line) by galactose inducible Ume6 E-ChRP or a constitutively expressed E-ChRP under the ADH1 promoter.


A
Catalytic Domain
X
E-ChRP Specificity
X (DBD)
AraC
Res1
Engrailed
GR
Chd1
Species
E. coli
S. pombe
D. melanogaster
R. norvegicus
S. cerevisiae
Recognition Motif
tAGCatttttaTCCATA
ACGCGT
TAATta
AGAACAgtgTGTTCT
(nonspecific DNA binding)
Nucleosomes with Recognition Motifs
AraC E-ChRP
Res1 E-ChRP
Engrailed E-ChRP
GR E-ChRP
No Motif
no E-ChRP
Chd1 E-ChRP (WT DBD)
B
Ume6 Targets
Nucleosome Positions (MNase-Seq)
No E-ChRP
Ume6 E-ChRP
Engrailed E-ChRP
URS1
Engrailed Targets
TAAT
C
Ume6-FLAG ChIP-Seq
(-) Galactose pGAL Ume6 E-ChRP
+Galactose pGAL Ume6 E-ChRP
Constituitive pADH1 Ume6 E-ChRP
ALP1

Remodeler Fusions are Highly Versatile *in vitro* and *in vivo*

We next tested mononucleosome targeting of multiple E-ChRPs with various DNA binding domains. We fused the DBD from *E. coli* AraC, *S. pombe* Res1, *D. melanogaster* Engrailed, or *R. norvegicus* Glucocorticoid Receptor to the E-ChRP scaffold. To determine if these E-ChRPs were functional on target nucleosomes *in vitro,* we generated end-positioned mononucleosomes assembled on the 601- positioning sequence (Lowary and Widom, 1998) with 125 base pairs of flanking DNA. The extranucleosomal DNA either contained or lacked a consensus binding motif corresponding each different fusion tested. E-ChRPs containing DNA binding domains were able to mobilize nucleosomes containing their well-defined recruitment motifs (Ades and Sauer, 1994; Alroy and Freedman, 1992; Ayte et al., 1995; Niland et al., 1996) as measured by a native PAGE nucleosome sliding assay (Fig 2A). These E-ChRPs were inactive on nucleosomes lacking their respective motifs (Fig. 2A, lanes 22-27), demonstrating specificity for target substrates *in vitro*. Fusion of the native, sequence-nonspecific DBD from Chd1 to our E-ChRP scaffold showed no apparent discrimination against DNA sequences and was capable of fully mobilizing the nonspecific mononucleosome control (Fig. 2A, lanes 28-30).

To determine whether E-ChRPs can be differentially targeted to specific subsets of nucleosomes *in vivo*, we introduced E-ChRPs into *S. cerevisiae* on a constitutive, ADH1-driven expression plasmid. When the E-ChRP contained a Ume6 DNA binding domain, nucleosomes were repositioned toward Ume6 binding motifs across the genome, but no nucleosome changes were detected at other genomic loci. Conversely, an E-ChRP containing the Engrailed DNA binding domain moved nucleosomes onto Engrailed

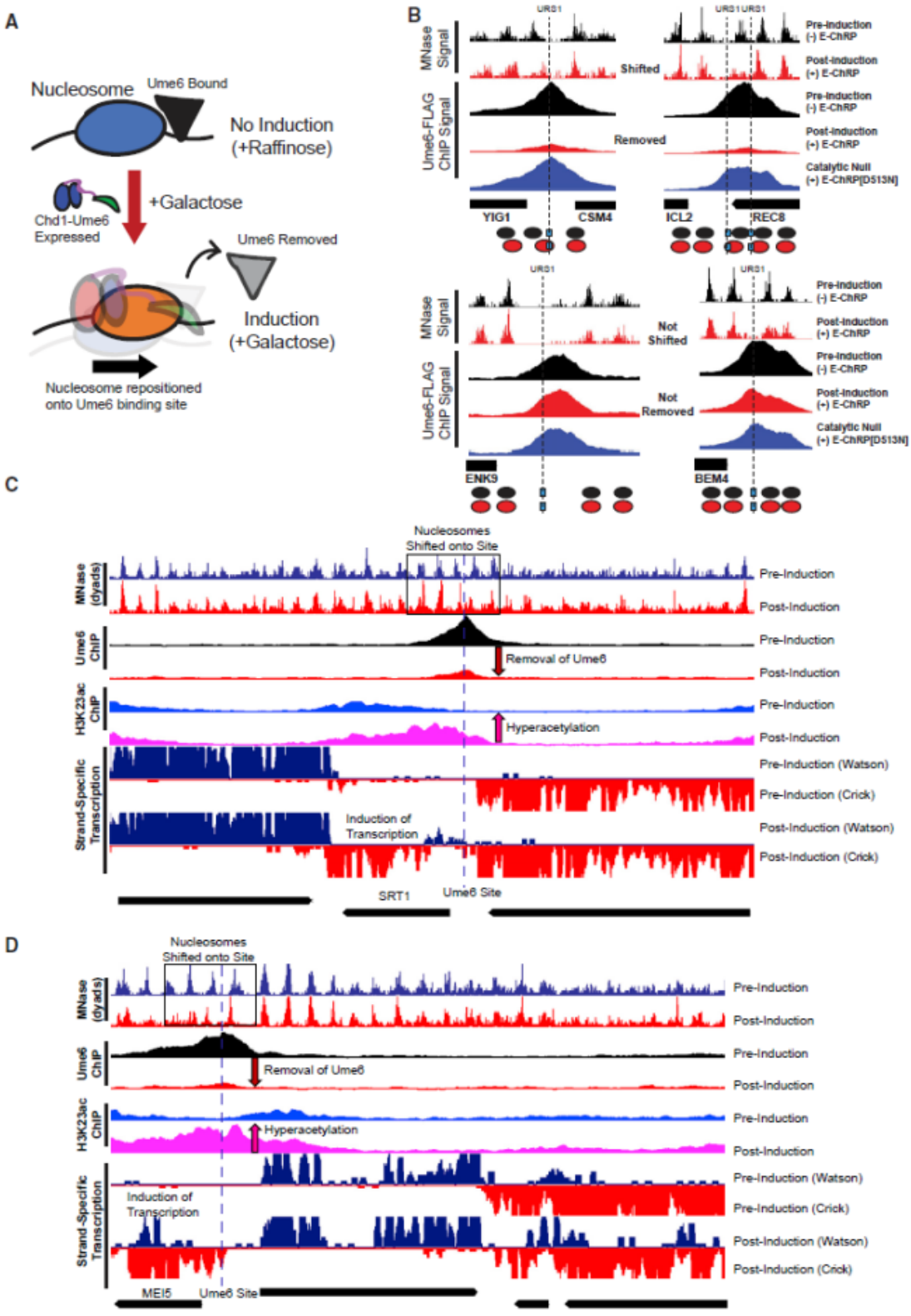
motifs in the yeast genome without altering nucleosome positions at Ume6 binding motifs (Fig 2B). We also introduced Ume6 and Engrailed E-ChRPs into yeast under the high-expression GPD (TDH3) promoter on a 2μm plasmid (Mumberg et al., 1995). Expression of the Ume6 E-ChRP from this construct resulted in positioned nucleosomes at target sites without identified off-target activity similar to an ADH1-driven E-ChRP. However, introduction of this higher expression plasmid containing an Engrailed E-ChRP only produced viable transformants in which the E-ChRP construct was deleted, truncated or mutated. This obligate inactivation of the Engrailed E-ChRP at high expression levels may result from promiscuous action of the Engrailed E-ChRP at tens of thousands of potential target sequences, which would presumably disrupt global nucleosome positioning in a deleterious and pleiotropic manner. Importantly, even when driven from the GPD (TDH3) promoter neither the Ume6 E-ChRP nor the Engrailed E-ChRP was active at target sites when the Chd1 remodeler core contained a catalytically-inactive Walker B (D513N) substitution (Hauk et al., 2010; Walker et al., 1982). Taken together, these results suggest that E-ChRPs can be specifically targeted through multiple distinct DNA binding domains that recognize sequence motifs with high or low complexity both *in vitro* and *in vivo*.

E-ChRPs can Inducibly Remove Transcription Factors

To gain temporal control of E-ChRPs *in vivo*, we introduced the Ume6 E-ChRP under the control of a galactose-inducible promoter integrated at the HO locus in yeast (Voth et al., 2001). Prior to addition of galactose, endogenous Ume6 associates with its consensus sequence across the genome and cooperates with the ISW2 complex to position motif-proximal nucleosomes, leaving 30 bp between the nucleosome edge and

URS1 motif (Goldmark et al., 2000; McKnight et al., 2016). After galactose induction of the Ume6 E-ChRP, a majority of nucleosomes nearest the URS1 site are efficiently repositioned to occlude the URS1 motif within two hours (Fig 3A,B). This galactose-inducible approach allows for more complete remodeling of nucleosomes than the same E-ChRP under the control of a constitutively active ADH1 promoter, potentially commensurate with differing expression levels under these distinct promoters (Fig 3A). We reasoned that because the post-induction nucleosome position results in the Ume6 recruitment motif becoming buried within nucleosomal DNA, remodeling by the Ume6 E-ChRP should interfere with binding of endogenous Ume6 (Fig 3B). To test this possibility, we tagged endogenous Ume6 with a FLAG epitope and measured Ume6-FLAG binding by ChIP-seq before and after induction of the Ume6 E-ChRP. Prior to induction, reproducible Ume6-FLAG binding was observed at URS1 sites across the genome. After induction of the Ume6 E-ChRP, which shifted nucleosomes over URS1 sites, Ume6 binding (as measured by Ume6-FLAG ChIP signal) was strongly reduced or eliminated at many genomic locations.

Figure 3 (next page). E-ChRPs Can Inducibly Remove Endogenous Ume6 from Chromatin (A) Cartoon depiction of Ume6 E-ChRP activity blocking association of endogenous Ume6 at target sites. **(B)** Representative examples of E-ChRP targets where endogenous Ume6 is removed (top) or not removed (bottom) after galactose induction of Ume6 E-ChRP. The catalytic null E-ChRP retains a Ume6 DBD but has a Walker B (D513N) substitution in the Chd1 catalytic core. **(C and D)** Representative Genome Browser images showing E-ChRP-dependent nucleosome dyad movement (within black rectangles) onto Ume6 sites (dashed line) with associated reduction of Ume6-FLAG ChIP, increase in acetylation, and transcriptional induction for the SRT1 locus (C) and MEI5 locus (D).


A
Nucleosome
Ume6 Bound
No Induction
(+Raffinose)
+Galactose
Chd1-Ume6
Expressed
Ume6 Removed
Induction
(+Galactose)
Nucleosome repositioned
onto Ume6 binding site
B
MNase Signal
Ume6-FLAG ChIP Signal
URS1
Shifted
Removed
Pre-Induction
(-) E-ChRP
Post-Induction
(+) E-ChRP
Catalytic Null
(+) E-ChRP[D513N]
YIG1
CSM4
ICL2
REC8
Not Shifted
Not Removed
ENK9
BEM4
C
Nucleosomes
Shifted onto Site
MNase (dyads)
Ume6 ChIP
Removal of Ume6
H3K23ac ChIP
Hyperacetylation
Strand-Specific Transcription
Induction of
Transcription
Pre-Induction
Post-Induction
Pre-Induction (Watson)
Pre-Induction (Crick)
Post-Induction (Watson)
Post-Induction (Crick)
SRT1
Ume6 Site
D
Nucleosomes
Shifted onto Site
MNase (dyads)
Ume6 ChIP
Removal of Ume6
H3K23ac ChIP
Hyperacetylation
Strand-Specific Transcription
Induction of
Transcription
Pre-Induction
Post-Induction
Pre-Induction (Watson)
Pre-Induction (Crick)
Post-Induction (Watson)
Post-Induction (Crick)
MEI5
Ume6 Site

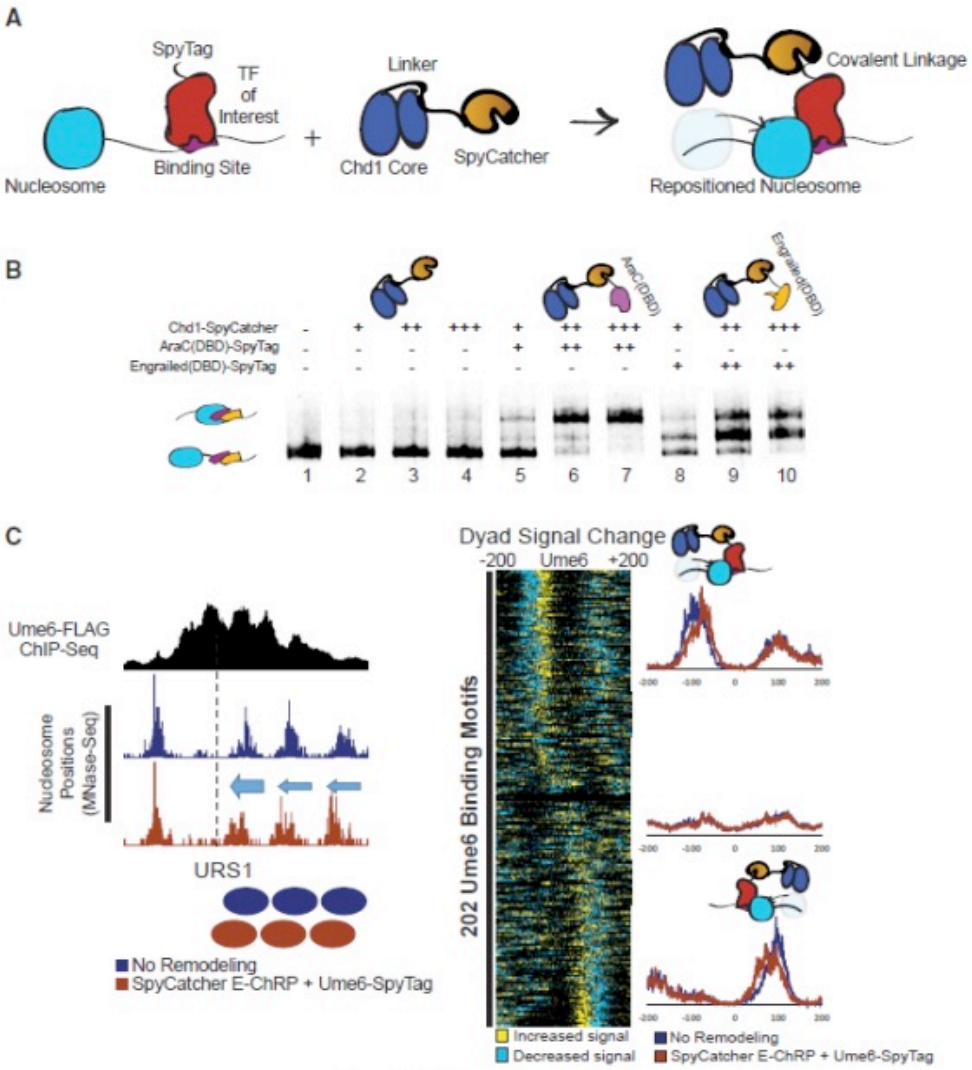
When we sorted Ume6 binding sites based on whether proximal nucleosome positions were shifted after Ume6 E-ChRP induction, we noticed that loss of Ume6-FLAG signal was strikingly reduced where nucleosomes were shifted, but minimally reduced where nucleosomes were not shifted (Fig 3C). To verify that this reduction in Ume6-FLAG signal was not due to direct binding competition between endogenous Ume6-FLAG and the E-ChRP Ume6 DNA binding domain, we measured Ume6-FLAG ChIP signal in the presence of a catalytically inactive Ume6 E-ChRP. This construct, which cannot move nucleosomes but retains the Ume6 DNA binding domain, did not similarly reduce Ume6-FLAG signal. Thus, E-ChRPs can inducibly move nucleosomes over target sequences to restrict access of the underlying DNA to endogenous DNA binding factors.

SpyCatcher E-ChRPs Allow Simple Targeting to Chromatin-Bound Loci

One limitation of the above-described E-ChRPs is their need to compete with endogenous factors for binding sites. To circumvent this problem, we created an E-ChRP where the SpyCatcher protein is fused in place of a DNA binding domain in the Chd1 E-ChRP scaffold. SpyCatcher specifically recognizes a short (~1kDa) SpyTag epitope, forming an isopeptide linkage that allows covalent protein fusions to be created *in vitro* and *in vivo* (Zakeri et al., 2012). This fusion provides two major improvements to the E-ChRP system. First, by simply appending SpyTag to different chromatin-binding factors of interest, nucleosome positioning can be achieved by a single SpyCatcher E-ChRP without the need to design new DBD fusions. Second, by tagging a transcription factor at its endogenous locus, the protein becomes a targetable element for the SpyCatcher E-ChRP only when bound to chromatin (Fig 4A). This bypasses the requirement of a vacant

DNA binding site to target a DBD-containing E-ChRP, allowing access to sequences in the genome that could otherwise be blocked by a stably-bound transcription factor. In sum, this strategy produces a single SpyCatcher E-ChRP that can be targeted to any chromatin-bound protein of interest in the genome by simple attachment of a short SpyTag. To validate the function of the SpyCatcher E-ChRP design, we purified recombinantly-expressed Chd1-SpyCatcher and two SpyTag-containing DNA binding domains. Mononucleosomes harboring recognition sequences for each DNA binding domain in the extranucleosomal DNA were incubated with the SpyCatcher E-ChRP with and without addition of SpyTag-Engrailed(DBD) or SpyTag-AraC(DBD).

Figure 4 (next page). Development and Validation of E-ChRPs Containing SpyCatcher/SpyTag Pairs (A) Cartoon representation for introducing a Chd1-SpyCatcher E-ChRP into cells containing SpyTagged, chromatin-bound proteins. The SpyCatcher domain forms a covalent isopeptide bond with SpyTag, allowing localization of E-ChRP activity to endogenously bound chromatin proteins. **(B)** Nucleosome sliding assay demonstrating that a single SpyCatcher E-ChRP cannot position nucleosomes without a SpyTag-containing DBD (lanes 2–4) but can use a SpyTagged AraC DBD (lanes 5–7) or engrailed DBD (lanes 8–10) to reposition nucleosomes containing respective DBD recognition motifs. The AraC and engrailed recognition elements are located 7 and 11 nt from the nucleosome edge, respectively. **(C)** Representative motif in yeast where ADH1-driven SpyCatcher E-ChRP can reposition nucleosomes (dyads) at a Ume6 binding site in the presence of Ume6-SpyTag (left) and genomic analysis of nucleosome dyad positioning by SpyCatcher E-ChRP at 202 intergenic instances of the Ume6 recognition sequence in cells containing SpyTagged Ume6 (right). **(D)** Genomic analysis of nucleosome dyad positions in Reb1-SpyTagged cells (left) or Ume6-SpyTagged cells (right) before and after 2 h induction of galactose inducible SpyCatcher E-ChRP. Heatmaps show change in nucleosome dyad signal after induction of SpyCatcher E-ChRP, and individual traces show average positions of nucleosomes in each cluster before and after SpyCatcher E-ChRP induction for each SpyTag-DBD strain.


A
SpyTag
TF of Interest
Binding Site
Nucleosome
Linker
Chd1 Core
SpyCatcher
Covalent Linkage
Repositioned Nucleosome
B
AraC(DBD)
Engrailed(DBD)
Chd1-SpyCatcher
AraC(DBD)-SpyTag
Engrailed(DBD)-SpyTag
1 2 3 4 5 6 7 8 9 10
C
Ume6-FLAG ChIP-Seq
Nucleosome Positions (MNase-Seq)
URS1
No Remodeling
SpyCatcher E-ChRP + Ume6-SpyTag
Dyad Signal Change
-200 Ume6 +200
202 Ume6 Binding Motifs
Increased signal
Decreased signal
No Remodeling
SpyCatcher E-ChRP + Ume6-SpyTag
D
SpyCatcher E-ChRP
Reb1-SpyTag
-250 Motif +250
940 Intergenic Reb1 Motifs
Immobile Cluster
n=405
Mobile Cluster
n=537
loss gain
change in nucleosome dyad signal after E-ChRP induction
Pre-induction Positions
Post-induction Positions
Ume6-SpyTag
202 Intergenic Ume6 Motifs
Mobile Cluster 1
n=38
Immobile Cluster
n=130
Mobile Cluster 2
n=33

The SpyCatcher E-ChRP has no activity on nucleosome substrates in the absence of a SpyTag-DBD pair (Fig 4B, lanes 2-4), since SpyCatcher has no intrinsic DNA binding affinity. However, addition of either SpyTag-AraC(DBD) (Fig 4B, lanes 5-7) or SpyTag-Engrailed(DBD) (Fig 4B, lanes 8-10) resulted in robust repositioning of mononucleosomes *in vitro*, demonstrating the versatility of this system.

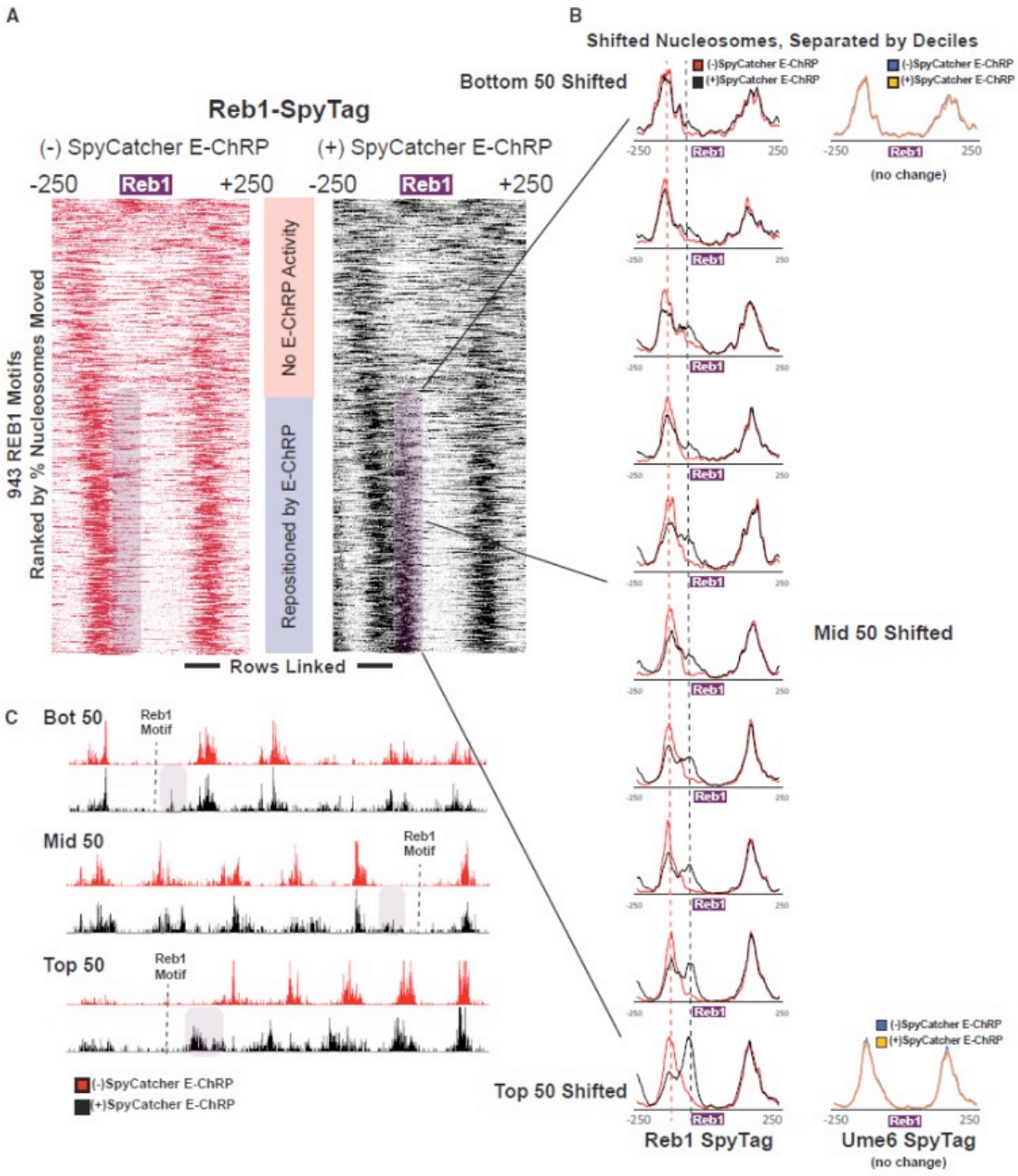
We next introduced the SpyCatcher E-ChRP under a constitutive ADH1 promoter into *S. cerevisiae* cells where a C-terminal SpyTag was added to full-length Ume6 at the endogenous locus. As expected, we observed repositioned nucleosomes at URS1 sites across the genome indicating chromatin remodeling at Ume6-bound loci (Fig 4C).

To achieve temporal control of this modular system *in vivo*, we appended SpyTag to the C-terminus of either Ume6 or Reb1, a yeast general regulatory factor, in a strain harboring a galactose-inducible SpyCatcher E-ChRP at the HO locus. After induction of SpyCatcher E-ChRP expression, nucleosomes were shifted toward Ume6 binding sites in cells containing Ume6-SpyTag or toward Reb1 binding sites in cells containing Reb1-SpyTag (Fig 4D). Interestingly, the fraction of shifted nucleosomes was generally low at Ume6 binding sites in Ume6-SpyTag cells but comparatively higher at Reb1 binding sites in Reb1-SpyTag strains (Fig 4D). This difference could be explained by higher occupancy or stability of Reb1 than Ume6 binding at target sites, which would allow a greater fraction of Reb1-tethered SpyCatcher to mobilize motif-proximal nucleosomes. Consistent with this possibility, the cellular abundance of Ume6 is significantly lower than that of Reb1 (Kulak et al., 2014). For Reb1-SpyTag strains, the positioning of a single motif-proximal nucleosome by the SpyCatcher E-ChRP initiated the shift of an entire array of nucleosomes toward the target motif, consistent with previous

observations that the positioning of a “barrier nucleosome” influences and constrains positions of an entire array of nucleosomes (Mavrich et al., 2008; McKnight et al., 2016).

Interestingly, the positioning of nucleosomes appeared to occur on only the 5' side of the Reb1 recognition sequence, suggesting the orientation of Reb1 binding impacts the ability of Chd1 to reach nucleosomes near binding sites (Fig 4D). This restriction could be explained by a constrained C-terminus of Reb1 when bound to chromatin, which is consistent with similarly constrained Reb1-MNase cleavage patterns seen in previous ChEC-seq experiments (Zentner et al., 2015). Unexpectedly, the fraction of nucleosomes shifted at individual Reb1 binding sites varied greatly in our data set, with some sites exhibiting repositioning of nearly 100% of motif-proximal nucleosomes in the population and others having a much smaller fraction moved (Fig 5A-C). These differences are not explained by initial nucleosome occupancy or location differences (Fig 5B), but are possibly related to relative Reb1 occupancy at different genomic locations.

Figure 5 (next page). E-ChRP Targeting to Chromatin-Bound Reb1 Provides Differential Occupancy Information at Reb1 Motifs (A) Nucleosome dyad signal at 943 intergenic Reb1 binding motifs in Reb1-SpyTag strains before (left) and after (right) 2 h induction of SpyCatcher E-ChRP. Rows are ordered by change in nucleosome positioning after galactose induction. Purple shading highlights the region to which nucleosomes are moved by SpyCatcher E-ChRP in the Reb1-SpyTag strain. **(B)** The purple mobile fraction from (A) was split into deciles (~50 motifs per decile) showing average positioning by SpyCatcher E-ChRP for each decile. Dashed lines indicate the pre-induction, unremodeled position (red) or post-induction, remodeled position (black). Ume6-SpyTag control traces are provided for the top and bottom deciles demonstrating that SpyCatcher E-ChRP cannot function at Reb1 sites in the presence of Ume6-SpyTag instead of Reb1-SpyTag. **(C)** Genome Browser images for representative loci showing nucleosome dyad positioning by SpyCatcher E-ChRP in a Reb1-SpyTag strain for the top, middle, and bottom deciles. Purple shading indicates the motif-proximal, repositioned nucleosomes. Dashed lines indicate the location of Reb1 motif.


A
Reb1-SpyTag
(-) SpyCatcher E-ChRP
(+) SpyCatcher E-ChRP
-250
Reb1
+250
943 REB1 Motifs
Ranked by % Nucleosomes Moved
No E-ChRP Activity
Repositioned by E-ChRP
Rows Linked
B
Shifted Nucleosomes, Separated by Deciles
Bottom 50 Shifted
(-)SpyCatcher E-ChRP
(+)SpyCatcher E-ChRP
(no change)
Mid 50 Shifted
Top 50 Shifted
Reb1 SpyTag
Ume6 SpyTag
C
Bot 50
Mid 50
Top 50
Reb1 Motif
(-)SpyCatcher E-ChRP
(+)SpyCatcher E-ChRP

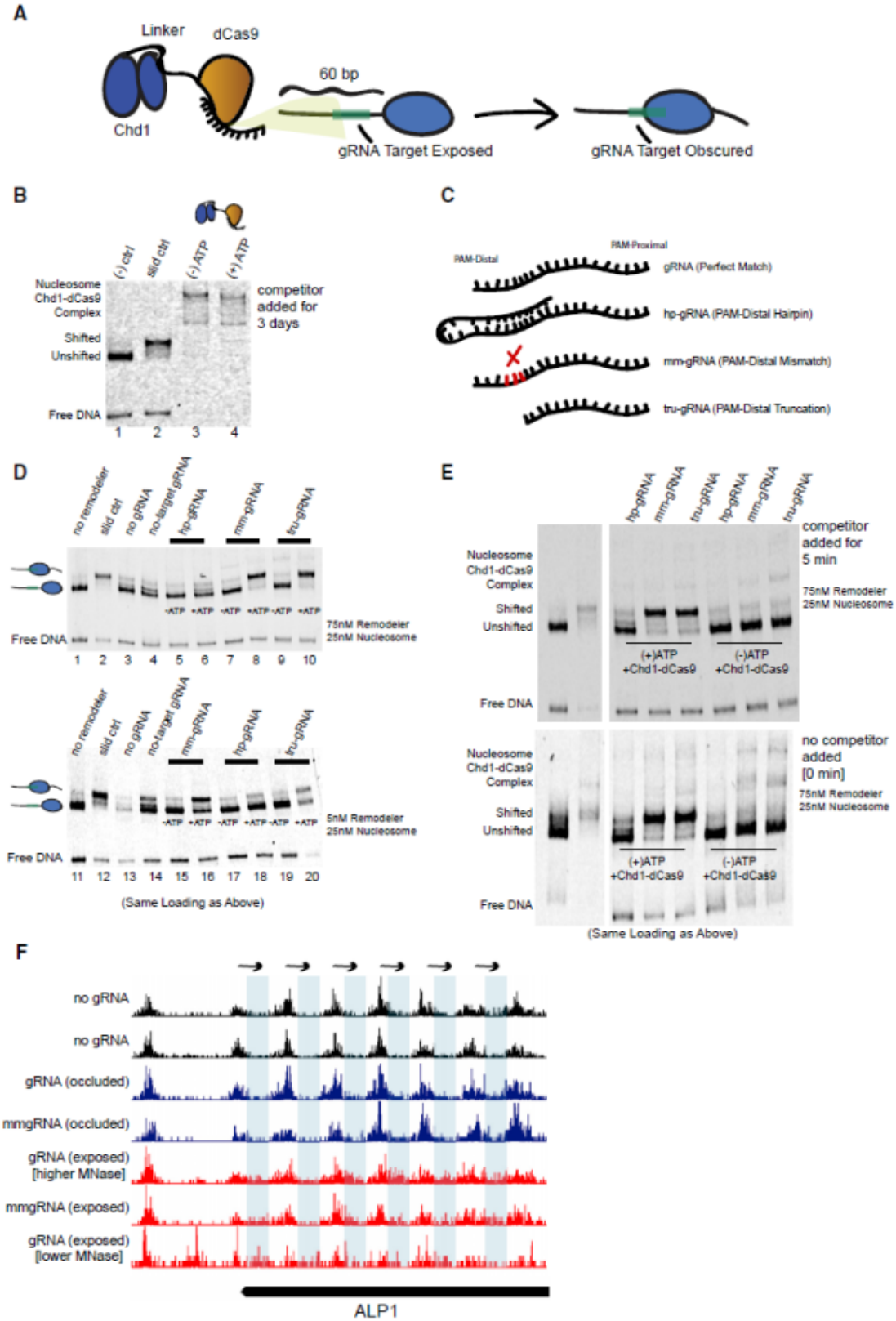
To better validate the ability of SpyCatcher E-ChRP to identify fractional Reb1 occupancy at Reb1 binding sites, we compared our data set to crosslinking ChIP, CUT&RUN (Skene and Henikoff, 2017), ORGANIC (Kasinathan et al., 2014) and ChEC-Seq (Zentner et al., 2015) data sets. There was striking correlation between our data, ORGANIC and ChEC-Seq, with some motifs exclusively showing Reb1 occupancy when measured by these three methods. Relative Reb1 occupancies mapped by CUT&RUN and standard ChIP were less correlated with our data suggesting that formaldehyde-free binding profiles similarly capture relative TF occupancy. Minimally, the observation that all nucleosomes are shifted at some Reb1 binding sites in a population of cells argues that some Reb1 sites are nearly 100% occupied, as E-ChRP-derived nucleosome movement cannot be observed without Reb1 binding (Fig 5C). While these Reb1 occupancy estimates are conflated with presence, accessibility and relative occupancy of motif-proximal nucleosomes, our data suggest that SpyCatcher E-ChRPs can serve as a relative measure of protein localization in cells that is orthogonal to ChIP, allowing for a lower-limit estimate of SpyTagged protein occupancy at individual binding motifs in the genome.

dCas9-targeted Nucleosome Positioning with Nonstandard gRNAs

While the E-ChRPs described above show robust nucleosome positioning activity when targeted through various DNA binding domains or through SpyCatcher/SpyTag pairs, their ability to alter nucleosome positions depends on the interaction between pre-existing DNA binding domains with defined DNA motifs. To overcome this limitation and allow targeted positioning of single nucleosomes by design, we created a dCas9 E-ChRP. This construct allows versatile targeting to specific nucleosomes by designing

proximal gRNAs. We recombinantly expressed the dCas9 E-ChRP in *E. coli* and purified the ~300kDa fusion protein. To test its ability to move gRNA-targeted nucleosomes we reconstituted end- positioned mononucleosomes and designed gRNAs with or without complementarity to the extranucleosomal DNA. Successful gRNA-stimulated chromatin remodeling would result in movement of the nucleosome toward the target sequence, producing a slower-migrating centrally-positioned nucleosome (Fig 6A). While nucleosomes were efficiently moved toward the center of DNA fragments with control Chd1 protein, introduction of Chd1-dCas9 and complementary gRNA resulted in supershifted complexes with unresolved nucleosome positions.

Figure 6 (next page). Remodeling Can Be Targeted Using a dCas9 E-ChRP with Canonical and Noncanonical gRNA Substrates (A) Cartoon depiction of dCas9-targeted chromatin remodeling and predicted repositioning of target nucleosome. **(B)** Nucleosome sliding assay showing irreversible association of the dCas9 E-ChRP with target nucleosomes and free DNA. The ‘‘slid ctrl’’ includes ATP and a control Chd1 protein capable of positioning nucleosomes toward the center of the DNA fragment. Excess unlabeled competitor DNA was added 3 days prior to loading. Nucleosome concentration was 25 nM, and E-ChRP concentration was 75 nM. **(C)** Cartoon depiction of canonical and nonstandard gRNA protospacers. **(D)** Nucleosome sliding assay demonstrating robust positioning of nucleosomes by a dCas9 E-ChRP targeted with nonstandard gRNAs in single-turnover (top) or multi-turnover (bottom) conditions. The slid ctrl (lanes 2 and 12) includes ATP and a control Chd1 protein. The ‘‘no gRNA’’ lanes (3 and 13) contain a dCas9 E-ChRP, ATP, and no gRNA. No-target gRNA samples (lanes 4 and 14) contain a dCas9 E-ChRP and a gRNA without any sequence complementarity to the substrate nucleosome. Note that the order of lanes between the upper and lower panels is similar except for loading of hp-gRNA (lanes 5, 6, 17, and 18) and mmgRNA (lanes 7, 8, 15, 16). **(E)** Nucleosome sliding assay demonstrating lack of stable association of dCas9 E-ChRPs with target nucleosomes or DNA in the presence of nonstandard gRNAs. For the upper gel, competitor DNA was added prior to loading. The bottom gel contains the identical reactions in the same order as the upper gel, but no competitor DNA was added before loading. **(F)** Genome Browser image showing both canonical and mismatched gRNAs allow repositioning of nucleosome dyads (note appearance of dyad signal in blue rectangle regions) by a dCas9 E-ChRP at the ALP1 locus when the gRNA sequence is exposed but not when gRNA sequence is occluded in the absence of dCas9 E-ChRP.


A
Linker
dCas9
Chd1
60 bp
gRNA Target Exposed
gRNA Target Obscured
B
(-) ctrl
sliid ctrl
(-) ATP
(+) ATP
Nucleosome
Chd1-dCas9
Complex
Shifted
Unshifted
Free DNA
competitor
added for
3 days
1 2 3 4
C
PAM-Proximal
PAM-Distal
gRNA (Perfect Match)
hp-gRNA (PAM-Distal Hairpin)
mm-gRNA (PAM-Distal Mismatch)
tru-gRNA (PAM-Distal Truncation)
D
no remodeler
sliid ctrl
no gRNA
no-target gRNA
hp-gRNA
mm-gRNA
tru-gRNA
-ATP +ATP -ATP +ATP -ATP +ATP
75nM Remodeler
25nM Nucleosome
Free DNA
1 2 3 4 5 6 7 8 9 10
5nM Remodeler
25nM Nucleosome
11 12 13 14 15 16 17 18 19 20
(Same Loading as Above)
E
hp-gRNA
mm-gRNA
tru-gRNA
competitor
added for
5 min
75nM Remodeler
25nM Nucleosome
(+)ATP
+Chd1-dCas9
(-)ATP
+Chd1-dCas9
no competitor
added
[0 min]
(Same Loading as Above)
F
no gRNA
no gRNA
gRNA (occluded)
mmgRNA (occluded)
gRNA (exposed)
[higher MNase]
mmgRNA (exposed)
gRNA (exposed)
[lower MNase]
ALP1

Even in the presence of 1000-fold competitor DNA for 3 days, the Chd1-dCas9 fusion protein would not release from gRNA-targeted nucleosomes (Fig 6B). This inability of dCas9 to release from target sequences is consistent with the ability of dCas9 to specifically bind and interfere with transcription in cells due to stable R-loop formation (Jinek et al., 2012; Laughery et al., 2019; Qi et al., 2013). Importantly, the inability of the dCas9 E-ChRP to release from substrate prevents its utility for precise, gRNA-targeted nucleosome positioning.

To promote release of the dCas9 E-ChRP from nucleosome substrates, we used gRNAs with noncanonical structures (Fig 6C) including a truncated gRNA (Fu et al., 2014) containing only 14nt of complementarity to target sequences, a gRNA with a PAM-distal hairpin (Josephs et al., 2015) that has predicted self-annealing capacity and reduced affinity for target sequences, and a 20nt gRNA with a PAM-distal 3nt- mismatch (mm-gRNA) that would result in an R-loop with a frayed end. Both the tru-gRNA and the mm-gRNA allowed for efficient targeted repositioning of nucleosomes toward the gRNA binding site either through direct Chd1-dCas9 fusion or introduction of Chd1-SpyCatcher and dCas9-SpyTag pairs (Fig 6D, lanes 1-10). These noncanonical gRNAs promoted multi-turnover catalysis by the dCas9 E-ChRP demonstrating that the weakened dCas9/gRNA complexes were stable enough to promote specific enzymatic activity but weak enough to readily and repeatedly disengage from its substrate (Fig 6D, lanes 11-20). Strikingly, dCas9-Chd1 targeted through weakened gRNAs did not require any competitor DNA to disengage from nucleosome substrates (Fig 6E). We believe this ability to readily dissociate from DNA targets while providing enough dwell time and specificity for targeted nucleosome positioning provides a facile method to alter

nucleosome positions by design. Furthermore, readily-dissociating mm-gRNAs can likely be employed for dCas9-targeted epigenome editing in cells, since current epigenome editing with canonical gRNAs likely leads to a combination of local epigenetic modification and stably-bound, mutagenic R-loop formation from strong dCas9 binding (Laughery et al., 2019).

Discussion

In conclusion, we have created and validated the use of E-ChRPs as an easy and versatile method for altering the positions of specific nucleosomes both *in vitro* and *in vivo*. We have demonstrated that E-ChRPs have widespread compatibility with various DNA binding domains and have created a single SpyCatcher E-ChRP that can be inducibly attached to chromatin-associated factors to move adjacent nucleosomes. We have shown that induced positioning of nucleosomes by E-ChRPs can establish new nucleosomal arrays, occlude transcription factor binding motifs across the genome, and report on relative transcription factor occupancy at target motifs. Finally, we have optimized a dCas9-targeted E-ChRP by creating weakened, noncanonical guide RNAs with PAM-distal mismatches to robustly position and release from targeted nucleosomes *in vitro*. We envision future research can employ E-ChRPs to probe questions directly relating the position of nucleosomes to downstream biological processes and can lead to insight into how cells can tolerate or correct ectopic nucleosome positioning events. We further expect weakened gRNAs will be an effective strategy to target specific epigenetic changes or nucleosome positioning changes in cells while limiting indirect consequences of stably- or irreversibly-bound dCas9 (Laughery et al., 2019). Finally, the ability to position nucleosomes onto target sequences may potentially lead to the development of

E-ChRPs that block oncogenic or other disease-related transcription factors from accessing binding sites genome-wide.

Limitations

While the E-ChRPs described in this work are highly versatile allowing for multiple targeting schemes, there are some limitations in the ability of E-ChRPs to position target nucleosomes. First, to create a functional Chd1-TF(DBD) fusion, the boundary of the DNA binding domain for the specific transcription factor must be known, and it must fold in the context of the fusion protein. While all fusions we have tested have been functional to date, we focused on well-behaved and well-studied DNA binding domains. Second, there is still a limitation on how far E-ChRPs can "reach". Based on our *in vivo* mapping results, if a nucleosome edge is initially beyond ~75 base pairs from the E-ChRP recruitment site, nucleosome repositioning activity is less favorable. Moreover, E-ChRPs do not appear to have any *de novo* nucleosome deposition activity, so exaggerated nucleosome-free regions of the genome would not permit nucleosome positioning. While we are very interested in the ability of SpyCatcher E-ChRPs to position different fractions of nucleosomes at different sites in the genome and we speculate this is due to relative SpyTagged TF occupancy, our method is blind to TF binding sites where there are no motif-proximal nucleosomes. Further, it would be challenging to use E-ChRPs in isolation to define TF binding landscape due to the relatively noisy signal generated from fractional repositioned nucleosomes. This is especially true for sites where small fractions of nucleosomes are repositioned. Finally, while the determination that dCas9 E-ChRPs work readily with mismatched gRNAs may allow better assessment of targeted chromatin modification effects (since dCas9 does not

remain stably associated with the target sequence), we note that mismatched gRNAs naturally possess a lower capacity for target specificity. Nevertheless, the robust activity we see with dCas9 E-ChRPs and mismatched gRNAs highlights, and may be generally useful for assessing, metastable off-target locations of gRNAs.

Bridge from Chapter II to III

In this Chapter, we described the use of E-ChRPs to precisely reposition nucleosomes at target sites throughout the yeast genome. This activity is similar to that described for targeted nucleosome remodeling by Isw2. Isw2 recruitment to TF-bound genome loci was discovered as a mechanism for +1 nucleosome positions to be defined through ChRP activity. While this is believed to occur through binding at sites within DNA exposed by NFRs, specifically those formed by Reb1, this does not align with recruitment of Isw2 by Ume6, which does not have GRF activity and is not directly associated with NFR formation. As such, our lab set out to determine the biochemical mechanism of Isw2 interaction with Ume6.

For this study, I was involved in addressing the comments we received after submitting the manuscript for review. I processed and visualized ChIP-Seq and MNase-Seq data derived from yeast strains with mutated versions of the Isw2 adaptor protein Itc1. The sections provided establish the functionality of Isw2 as a targeted chromatin remodeling protein, which is relevant to the results presented in Chapter VI.

References Cited

Ades, S.E., and Sauer, R.T. (1994). Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50. Biochemistry *33*, 9187-9194.

Alroy, I., and Freedman, L.P. (1992). DNA binding analysis of glucocorticoid receptor specificity mutants. Nucleic Acids Res *20*, 1045-1052.

Anderson, S.F., Steber, C.M., Esposito, R.E., and Coleman, J.E. (1995). UME6, a negative regulator of meiosis in Saccharomyces cerevisiae, contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1 DNA sequence in a zinc-dependent manner. Protein Sci *4*, 1832-1843.

Ayte, J., Leis, J.F., Herrera, A., Tang, E., Yang, H., and DeCaprio, J.A. (1995). The Schizosaccharomyces pombe MBF complex requires heterodimerization for entry into S phase. Mol Cell Biol *15*, 2589-2599.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249- 1260.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. Nucleic Acids Res *43*, D662-669.

Eberharter, A., Langst, G., and Becker, P.B. (2004). A nucleosome sliding assay for chromatin remodeling factors. Methods Enzymol *377*, 344-353.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth Biol *2*, 604-613.

Fazzio, T.G., and Tsukiyama, T. (2003). Chromatin remodeling in vivo: evidence for a nucleosome sliding mechanism. Mol Cell *12*, 1333-1340.

Ferreira, H., Flaus, A., and Owen-Hughes, T. (2007). Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. J Mol Biol *374*, 563-579.

Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. Bioinformatics *32*, 2089-2095.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol *32*, 279-284.

Gangaraju, V.K., and Bartholomew, B. (2007). Dependency of ISW1a chromatin remodeling on extranucleosomal DNA. Mol Cell Biol *27*, 3217-3225.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M., and Tsukiyama, T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell *103*, 423-433.

Hauer, M.H., and Gasser, S.M. (2017). Chromatin and nucleosome dynamics in DNA damage and repair. Genes Dev *31*, 2204-2221.

Hauk, G., McKnight, J.N., Nodelman, I.M., and Bowman, G.D. (2010). The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. Mol Cell *39*, 711-723.

Hota, S.K., Bhardwaj, S.K., Deindl, S., Lin, Y.C., Zhuang, X., and Bartholomew, B. (2013). Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. Nat Struct Mol Biol *20*, 222-229.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Josephs, E.A., Kocak, D.D., Fitzgibbon, C.J., McMenemy, J., Gersbach, C.A., and Marszalek, P.E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. Nucleic Acids Res *43*, 8924-8941.

Kasinathan, S., Orsi, G.A., Zentner, G.E., Ahmad, K., and Henikoff, S. (2014). High-resolution mapping of transcription factor binding sites on native chromatin. Nat Methods *11*, 203-209.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S.R., Tan, G.*, et al.* (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res *46*, D1284.

Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic- sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods *11*, 319-324.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat Rev Mol Cell Biol *18*, 548-562.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357- 359.

Langst, G., Bonte, E.J., Corona, D.F., and Becker, P.B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. Cell *97*, 843-852.

Laughery, M.F., Mayes, H.C., Pedroza, I.K., and Wyrick, J.J. (2019). R-loop formation by dCas9 is mutagenic in Saccharomyces cerevisiae. Nucleic Acids Res *47*, 2389-2401.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol *119*, 1-16.

MacAlpine, D.M., and Almouzni, G. (2013). Chromatin and DNA replication. Cold Spring Harb Perspect Biol *5*, a010207.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res *18*, 1073-1083.

McKnight, J.N., Boerma, J.W., Breeden, L.L., and Tsukiyama, T. (2015). Global Promoter Targeting of a Conserved Lysine Deacetylase for Transcriptional Shutoff during Quiescence Entry. Mol Cell *59*, 732-743.

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T., and Bowman, G.D. (2011). Extranucleosomal DNA binding directs nucleosome sliding by Chd1. Mol Cell Biol *31*, 4746-4759.

McKnight, J.N., and Tsukiyama, T. (2015). The conserved HDAC Rpd3 drives transcriptional quiescence in S. cerevisiae. Genom Data *6*, 245-248.

McKnight, J.N., Tsukiyama, T., and Bowman, G.D. (2016). Sequence-targeted nucleosome sliding in vivo by a hybrid Chd1 chromatin remodeler. Genome Res *26*, 693-704.

Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119-122.

Niland, P., Huhne, R., and Muller-Hill, B. (1996). How AraC interacts specifically with its target DNAs. J Mol Biol *264*, 667-674.

Nodelman, I.M., and Bowman, G.D. (2013). Nucleosome sliding by Chd1 does not require rigid coupling between DNA-binding and ATPase domains. EMBO Rep *14*, 1098-1103.

Park, H.D., Luche, R.M., and Cooper, T.G. (1992). The yeast UME6 gene product is required for transcriptional repression mediated by the CAR1 URS1 repressor binding site. Nucleic Acids Res *20*, 1909-1915.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Rodriguez, J., McKnight, J.N., and Tsukiyama, T. (2014). Genome-Wide Analysis of Nucleosome Positions, Occupancy, and Accessibility in Yeast: Nucleosome Mapping, High-Resolution Histone ChIP, and NCAM. Curr Protoc Mol Biol *108*, 21 28 21-16.

Ryan, D.P., Sundaramoorthy, R., Martin, D., Singh, V., and Owen-Hughes, T. (2011). The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. EMBO J *30*, 2596-2609.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Smith, C.L., and Peterson, C.L. (2005). A conserved Swi2/Snf2 ATPase motif couples ATP hydrolysis to chromatin remodeling. Mol Cell Biol *25*, 5880-5892.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J Biol Chem *281*, 16279-16288.

Tsukiyama, T., Becker, P.B., and Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature *367*, 525-532.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol *16*, 178-189.

Voth, W.P., Richards, J.D., Shaw, J.M., and Stillman, D.J. (2001). Yeast vectors for integration at the HO locus. Nucleic Acids Res *29*, E59-59.

Walker, A., Taylor, J., Rowe, D., and Summers, D. (2008). A method for generating sticky-end PCR products which facilitates unidirectional cloning and the one-step assembly of complex DNA constructs. Plasmid *59*, 155-162.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J *1*, 945-951.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A *109*, E690-697.

Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat Commun *6*, 8733.

Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. Annu Rev Biophys *45*, 153-181.

CHAPTER III

BASIS OF SPECIFICITY FOR A CONSERVED AND PROMISCUOUS CHROMATIN REMODELING PROTEIN

*This chapter contains previously published co-authored material.

Donovan, D.A., Crandall, J.G., Truong, V.N., Vaaler, A.L., Bailey, T.B., Dinwiddie, D., **Banks, O.G.B.**, McKnight, L.E., McKnight, J.N. (2021). Basis of Specificity for a Conserved and Promiscuous Chromatin Remodeling Protein. eLife 10:e64061

Conceptualization, D.A.D. and J.N.M. ; Methodology, D.A.D., J.G.C., V.N.T., A.L.V., T.B.B., L.E.M., and J.N.M.; Investigation, D.A.D., J.G.C., V.N.T., A.L.V., T.B.B., L.E.M., and J.N.M.; Writing – Original Draft, D.A.D. and J.N.M.; Writing – Reviewing & Editing, D.A.D., J.G.C., A.L.V., O.G.B.B, L.E.M., and J.N.M.; Visualization, D.D., O.G.B.B., and J.N.M.; Supervision, L.E.M. and J.N.M.; Project Administration, J.N.M.; Funding Acquisition, J.N.M.

Introduction

Chromatin consists of the nucleic acids and proteins that make up the functional genome of all eukaryotic organisms. The most basic regulatory and structural unit of chromatin is the nucleosome. Each nucleosome is defined as an octamer of histone proteins, which is wrapped by approximately 147 base pairs of genomic DNA (Luger et al., 1997; Kornberg, 1974). The specific positioning of nucleosomes on the underlying DNA can have significant effects on downstream processes, such as promoter accessibility and molecular recruitment, which ultimately serve to alter gene expression (Lai and Pugh, 2017). Despite decades of research, the mechanisms leading to precise nucleosome locations in cells are still being defined.

Nucleosome positioning is dynamically established by a group of enzymes known as ATP-dependent chromatin remodeling proteins (ChRPs) (Zhou et al., 2016). Extensive biochemical and structural characterization has been performed on this group of proteins from various families (Clapier et al., 2017). The chromodomain-helicase-DNA binding (CHD) and imitation switch (ISWI) families of ChRPs have been characterized as nonspecific nucleosome sliding and spacing factors in vitro (Stockdale et al., 2006; Hauk et al., 2010; McKnight et al., 2011; Kagalwala et al., 2004; Lusser et al., 2005; Tsukiyama et al., 1999; Pointner et al., 2012). In yeast, flies, and mammals, ChRPs generate evenly spaced nucleosome arrays at transcription start sites and organize genomic chromatin at other defined boundaries (Pointner et al., 2012; Lee et al., 2007; Mavrich et al., 2008a; Valouev et al., 2011; Krietenstein et al., 2016; Wiechens et al., 2016; Baldi et al., 2018; Gkikopoulos et al., 2011; Zhang et al., 2011). However, relatively little is known about the in vivo biological regulation of these spacing factors, and it is not understood how they can accurately and reproducibly position nucleosomes throughout the genome in different cellular contexts.

A widely accepted model is that ChRPs pack nucleosome arrays against a noninteracting barrier, such as an unrelated DNA binding protein or another nucleosome (Krietenstein et al., 2016; Zhang et al., 2011; Mavrich et al., 2008b). In this way, general regulatory factors (GRFs) could establish chromatin landscapes with differing nucleosome arrays in response to changes in the cellular environment. In support of this model, nucleosome arrays near GRFs and other DNA binding elements appear to be phased relative to the binding motifs of the sequence-specific DNA binding factors in cells and in biochemically reconstituted cell-free systems (Krietenstein et al., 2016; Baldi

et al., 2018; Yan et al., 2018). This model suggests that boundaries of nucleosome arrays are determined by the binding of barrier factors. Implicit in this barrier model are the assumptions that ChRPs act as nonspecific nucleosome spacing machines throughout the genome and that specific ChRP and GRF interactions are not required to establish nucleosome positions. While this model provides a good explanation for how phased nucleosome arrays can be established throughout the genome by a combination of DNA binding factors and nonspecific chromatin remodeling factors, the fundamental assumptions of the barrier model have not been thoroughly tested.

It has been shown through genetic and recent biochemical experiments that members of the ISWI family of ChRPs functionally interact with transcription factors in vivo (Krietenstein et al., 2016; Gelbart et al., 2005; Goldmark et al., 2000; Fazzio et al., 2001; Yadon et al., 2013). One of the most well-defined interacting partners of ISWI proteins is the meiotic repressor unscheduled meiotic gene expression (Ume6), which is found in yeasts. It has been previously demonstrated that Ume6 and Isw2, an ISWI-containing ChRP complex in Saccharomyces cerevisiae (homologous to the ATP-dependent chromatin assembly factor [ACF] complex in humans and flies), share genetic targets of repression and likely interact physically (Goldmark et al., 2000). While interactions with sequence-specific DNA binding proteins can potentially determine precise nucleosome targeting and final nucleosome positions (Donovan et al., 2019; Bowman and McKnight, 2017; McKnight et al., 2016), the mechanisms through which physical interactions between Isw2 and any genomic recruitment factor like Ume6 influence nucleosome positioning activity in cells have not been defined. For example, it is not known how these physical interactions occur or what role they play in the

biochemical outcomes of chromatin remodeling reactions and the resulting downstream biological outputs.

In this work, we have successfully identified the mechanism of interaction between Isw2 and Ume6 in S. cerevisiae. By taking a protein dissection approach combined with genome-wide nucleosome profiling, we have identified a previously uncharacterized helical domain in Ume6 that allows for Isw2 binding, specific genomic recruitment, and precise nucleosome positioning outcomes. We further demonstrate that conserved attributes of this helical domain are observed in the cell cycle regulator Swi6, which we have identified as a new Isw2-recruitment adapter protein that allows for specific nucleosome positioning at Mbp1/Swi6 (MBF) and Swi4/Swi6 (SBF) targets. We have also determined that the transcription factor-interacting interface of Isw2/ACF-like remodeling complexes contains a few key and highly conserved residues within the WAC (WTSF/Acf1/cbp146) domain. Finally, we show that these residues, which are essential for directional, sequence-specific remodeling, were lost in the evolution of the Drosophila lineage, where extensive biochemical, genetic, and genomic characterization has been performed on the ITC1 ortholog ACF.

Results

Isw2 activity in cells is inconsistent with known biochemistry and the barrier model for nucleosome packing

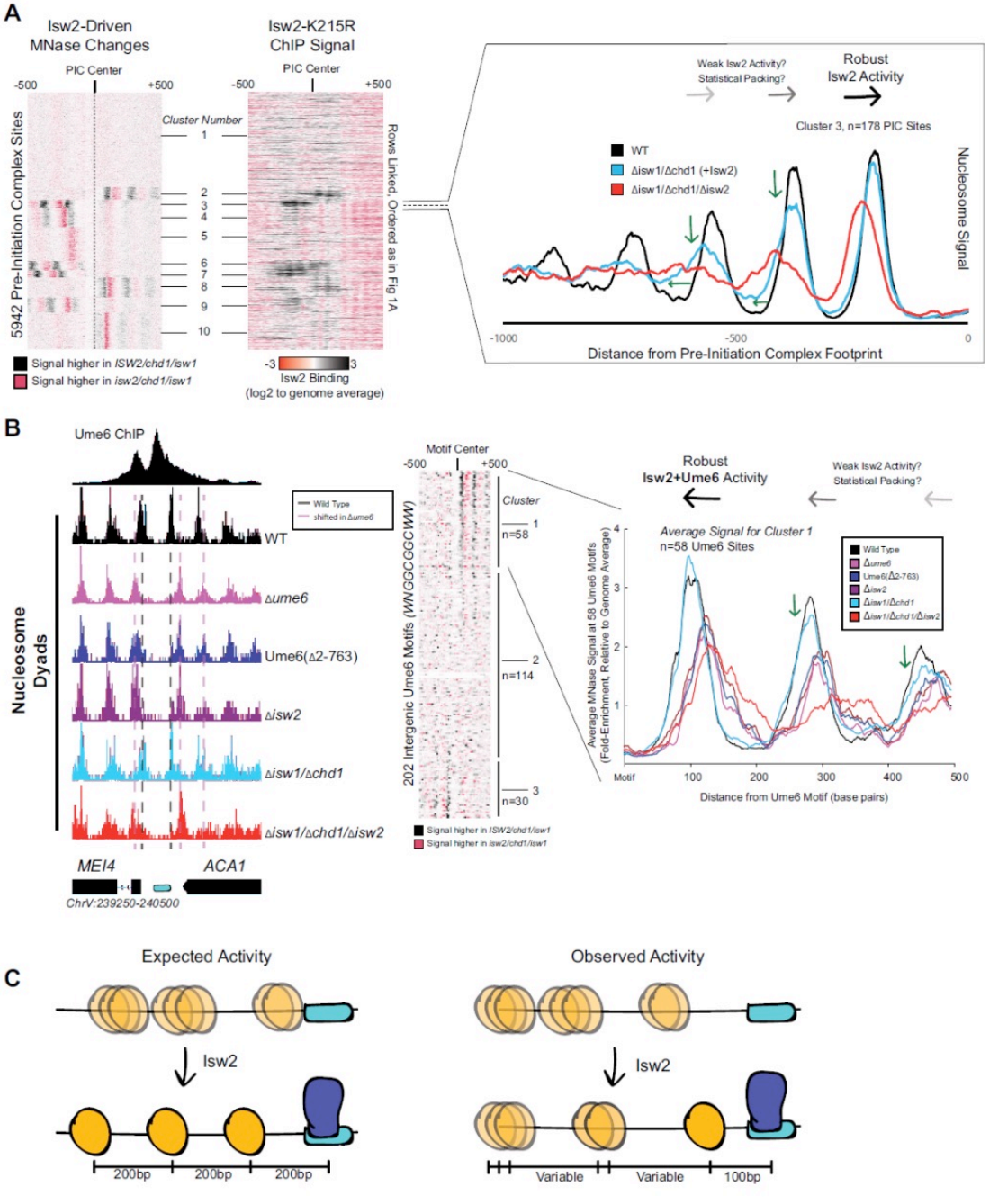
We wished to understand how the conserved Isw2 protein complex in yeast behaves genome-wide and at specific promoter nucleosomes at target sites. Yeast Isw2 has been characterized extensively in biochemical assays, which all suggest that it has

nonspecific DNA binding, ATP hydrolysis, nucleosome sliding, mononucleosome centering, and nucleosome spacing activities (Stockdale et al., 2006; Kagalwala et al., 2004; Lusser et al., 2005; Tsukiyama et al., 1999; Dang and Bartholomew, 2007; Dang et al., 2006; Hota et al., 2013; Kassabov et al., 2002; Zofall et al., 2004; Zofall et al., 2006). These nonspecific nucleosome mobilizing activities suggest that the Isw2 protein should be able to organize nucleosome arrays against a barrier across the genome in yeast cells since (1) it is estimated that there are enough Isw2 molecules for every 10–20 nucleosomes in the genome (Gelbart et al., 2005), (2) Drosophila melanogaster ACF, an Isw2 ortholog, can organize nucleosomes into evenly spaced arrays (Baldi et al., 2018), and (3) other nonspecific and related nucleosome spacing factors can globally space nucleosomes across the genome in yeast and other organisms (Pointner et al., 2012; Wiechens et al., 2016; Gkikopoulos et al., 2011; Zhang et al., 2011). To first determine how Isw2 positions nucleosomes in S. cerevisiae, we examined nucleosome positioning activity in an isw1/chd1 deletion background to remove known and potentially overlapping global spacing factors and highlight 'isolated positioning activity' by Isw2. When examining the positioning of nucleosomes with and without Isw2 at all yeast pre-initiation complex sites (PICs), it is evident that Isw2 activity is specialized at only a subset of target sites (Figure 7A). As seen previously (Gkikopoulos et al., 2011; Ocampo et al., 2016), no global nucleosome spacing or organizing activity is detected by Isw2 alone. Close inspection of Isw2-targeted PICs suggests that Isw2 can only organize a single PIC-proximal nucleosome, while subsequent nucleosomes become more poorly phased as the distance from the initially positioned nucleosome increases (Figure 7A). Importantly, the PICs that display specific Isw2-directed activity are bound by Isw2,

while those lacking any detectable nucleosome organization by Isw2 are unbound (Figure 7A, middle panel).

It has been shown that Isw2 associates with sequence-specific DNA binding factors, such as the transcriptional repressor Ume6 (Goldmark et al., 2000; Fazzio et al., 2001). Isw2 activity at Ume6-bound loci has been previously characterized as precise, with Isw2 reproducibly moving nucleosomes until the predicted edge of the nucleosome core particle is 30 base pairs from the center of the Ume6 binding motif (McKnight et al., 2016).

Figure 7 (next page): Isw2 is a Specialist Remodeler that Positions Single Nucleosomes at Target Sites. (**A**) (Left) Clustered heatmap showing differences in nucleosome dyad signal between *isw2*/*isw1*/*chd1* and *ISW2*/*isw1*/*chd1* strains at 5942 pre-initiation complex sites (PICs). Black indicates positions where Isw2 preferentially positions nucleosomes compared to the strain lacking Isw2. (Middle) Heatmap of ISW2(K215R) ChIP signal, with rows linked to the PIC data on the left, shows that Isw2-dependent nucleosome changes overlap with regions where Isw2 is present. (Right) Average nucleosome dyad signal for wild type (WT) (black), *isw1*/*chd1* (cyan), and *isw2*/*isw1*/*chd1* (red) strains for the 178 PIC sites in cluster 3. Black arrows denote Isw2-driven nucleosome shifts. Green arrows indicate rapid decay of positioning at PIC-distal nucleosomes in the *ISW2*/*isw1*/*chd1* mutant. (**B**) (Left) Genome Browser image showing nucleosome dyad signal at a *u*nscheduled *m*eiotic gene *e*xpression (Ume6) motif (cyan rectangle) for indicated strains. Vertical gray dashed line denotes the motif-proximal WT nucleosome positions while vertical pink dashed line indicates the nucleosome positions in the absence of Ume6 or Isw2. (Center) Clustered heatmap showing the difference in nucleosome dyad signal between *isw2*/*isw1*/*chd1* and *ISW2*/*isw1*/*chd1* strains at 202 intergenic Ume6 motifs. Black indicates positions where Isw2 preferentially positions nucleosomes compared to strains lacking Isw2. (Right) Average nucleosome dyad signal for indicated strains at Ume6 motifs in cluster 1. Black arrows indicate direction of nucleosome positioning by Isw2. Green arrows signify decreased positioning of motif-distal nucleosomes in the *ISW2*/*isw1*/*chd1* strain (cyan) compared to WT (black). (**C**) (Left) Cartoon depicting the expected activity of Isw2 at barrier elements according to current biochemical data and nucleosome positioning models. Isw2 is thought to move nucleosomes away from bound factors and space nucleosomes with an approximately 200 base pair repeat length. (Right) Cartoon of the observed activity of Isw2 at target sites where only a motif proximal single nucleosome is precisely positioned but distal nucleosomes are not well-spaced by Isw2.


A
Isw2-Driven MNase Changes
Isw2-K215R ChIP Signal
PIC Center
-500
+500
Cluster Number
5942 Pre-Initiation Complex Sites
Rows Linked, Ordered as in Fig 1A
Signal higher in ISW2/chd1/isw1
Signal higher in isw2/chd1/isw1
Isw2 Binding
(log2 to genome average)
Weak Isw2 Activity?
Statistical Packing?
Robust
Isw2 Activity
Cluster 3, n=178 PIC Sites
WT
Δisw1/Δchd1 (+Isw2)
Δisw1/Δchd1/Δisw2
Nucleosome Signal
Distance from Pre-Initiation Complex Footprint
B
Ume6 ChIP
Nucleosome Dyads
Wild Type
shifted in Δume6
WT
Δume6
Ume6(Δ2-763)
Δisw2
Δisw1/Δchd1
Δisw1/Δchd1/Δisw2
MEI4
ACA1
ChrV:239250-240500
Motif Center
202 Intergenic Ume6 Motifs (WNGGCGGCWW)
Cluster
n=58
n=114
n=30
Robust
Isw2+Ume6 Activity
Weak Isw2 Activity?
Statistical Packing?
Average Signal for Cluster 1
n=58 Ume6 Sites
Average MNase Signal at 58 Ume6 Motifs
(Fold-Enrichment, Relative to Genome Average)
Motif
Distance from Ume6 Motif (base pairs)
C
Expected Activity
Observed Activity
Isw2
200bp
Variable
100bp

Because of the connection to Ume6, we examined nucleosome positions in an isw1/chd1 background in the presence and absence of Isw2 to determine whether Isw2 is similarly restricted at known target sites. Again, we determined that Isw2 is efficient at positioning the Ume6-proximal nucleosome but positioning of nucleosomes decays rapidly as the distance from the proximal nucleosome increases, suggesting that Isw2 may only position single nucleosomes at target sites (Figure 7B, clusters 1 and 3). Nucleosomes also appear to always be positioned toward Ume6 motifs as nucleosome positions in the absence of Isw2 are always more distal to the Ume6 motif than when Isw2 is present. Finally, these nucleosomes are positioned with the dyad only separated from the Ume6 motif by 100 nucleotides rather than the ~200 nucleotides that would be expected between dyads in a nucleosome array based on Isw2 preferentially leaving 60 base pairs of linker DNA between nucleosomes in vitro (Kagalwala et al., 2004; Tsukiyama et al., 1999). Of note, a subset of Ume6-bound sites do not display Isw2- dependent nucleosome remodeling (Figure 7B, cluster 2). We have observed slightly reduced chromatin immunoprecipitation (ChIP) signal for Ume6 at these sites. We speculate that for cluster 2 sites where Ume6 is bound, the Isw2 complex might in fact be recruited but that the nearest nucleosome is too distant from the recruitment site, making it out of reach of the remodeler and thus resulting in no change in nucleosome position at these sites.

The observations that (1) Isw2 is solely required to move single nucleosomes at target sites, (2) Isw2 does not have global nucleosome spacing/organizing activity, and (3) Isw2 moves nucleosomes within 100 nucleotides of bound Ume6 suggest that Isw2 behavior in cells is distinct from our understanding of Isw2 activity from decades of biochemical characterization. Similarly, these specific movements toward Ume6 (a

barrier) are inconsistent with previous biophysical studies, where ISWI proteins were shown to move nucleosomes away from inert DNA-bound factors (Li et al., 2015). Because of these inconsistencies, we wished to know if Isw2 followed the 'barrier model' for positioning nucleosomes at Ume6-bound targets. To initially test this, we created a variant Ume6 construct where all residues were deleted except for the DNA binding domain. This Ume6(Δ2–763) construct binds to the same targets as full-length Ume6. However, Isw2 does not appear to have any activity on global Ume6-proximal nucleosomes in the presence of the Ume6 DNA binding domain alone as nucleosomes in this strain occupy identical positions when Ume6 or Isw2 are completely absent (Figure 7B). In the presence of full-length Ume6, the Isw2 complex appears to be necessary and sufficient for moving motif-proximal nucleosomes as nucleosome positions in the ISW2/isw1/chd1 strain could achieve identical motif-proximal nucleosome positions as the wild-type strain. Additionally, the CHD1/isw1/isw2 and ISW1/chd1/isw2 strains were unable to move any Ume6-proximal nucleosomes, which strongly argues that Ume6 is not acting as a passive barrier against which nucleosome spacing factors can pack nucleosomes. Instead, these data are more consistent with the recent characterization of Isw2 as a 'puller' (Kubik et al., 2019), with Ume6 being a DNA- bound factor that may immobilize Isw2 to create leverage for 'pulling'. Consistent with this immobilized pulling model and consistent with the directional movement of single nucleosomes toward Ume6-bound sites, artificially tethered chromatin remodeling proteins were previously shown to always move nucleosomes toward target sites (Donovan et al., 2019). We suspected that Ume6 and Isw2 likely interact in a specific fashion to faithfully

select and precisely move single-target nucleosomes toward a recruitment motif (Figure 7C).

A small helical epitope is necessary and sufficient for Isw2-directed nucleosome positioning at Ume6 targets

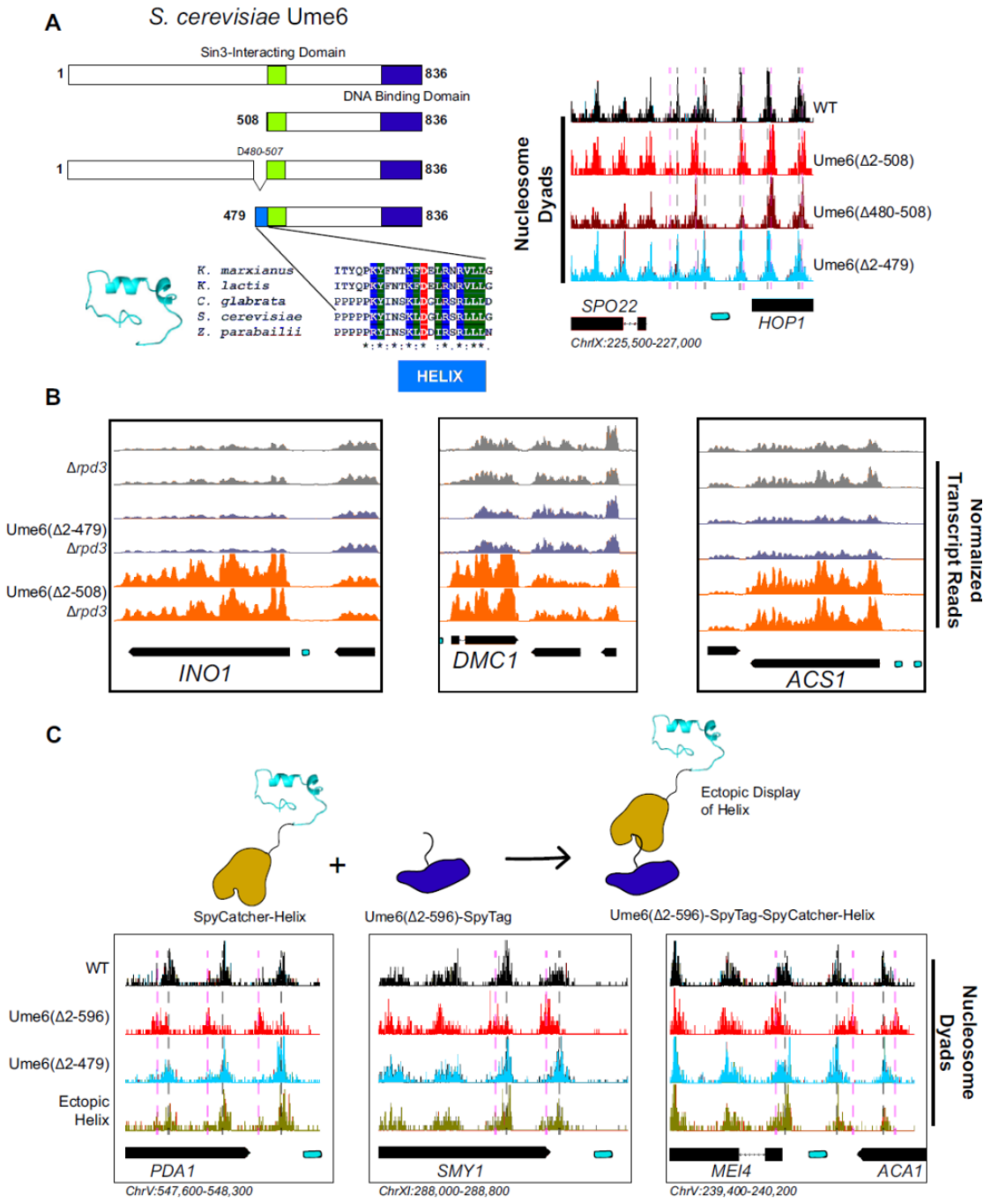
To determine which region(s) on Ume6 are required for specific nucleosome positioning by Isw2, we initially created a panel of N-terminal Ume6 truncations to determine when nucleosome positioning by Isw2 is lost. This initial truncation panel was necessary due to the poor overall conservation of the Ume6 protein even within related yeasts, as well as the disordered structure predicted by Phyre2 (Kelley et al., 2015). Our truncation panel indicated that Isw2 activity was retained if the N- terminus was deleted to residue 322 but lost when deleted to residue 508. Closer inspection of the residues between 322 and 508 revealed a conserved region with a proline-rich segment followed by a predicted alpha helix, altogether spanning Ume6 residues 479–508 (Figure 8A). Deletion of residues 2–479 preserved Isw2-positioned nucleosomes at Ume6 sites, while an internal deletion of 480–507 in the context of an otherwise full-length Ume6 abrogated nucleosome positioning by Isw2.

Importantly, Ume6 Δ2–479 and Ume6 Δ2–508 showed identical binding as measured by ChIP, indicating that the loss of nucleosome positioning is not due to the loss of Ume6 binding. Since this region is proximal to the characterized Sin3-binding domain in Ume6 (Washburn and Esposito, 2001), we wished to validate that the newly determined Isw2- recruitment helix is independent from the Sin3-binding domain. Ume6 recruits both Isw2 and Sin3-Rpd3 for full repression of target genes (Goldmark et al., 2000; Fazzio et al., 2001). If either Isw2 or Sin3-Rpd3 is present, there is partial

repression at Ume6- regulated genes. However, if Sin3-Rpd3 and Isw2 are both lost, Ume6 targets are fully de-repressed.

We examined transcriptional output at Ume6 genes in Ume6(Δ2–479) +/– Rpd3 and Ume6(Δ2–508) +/– Rpd3. Transcription was modestly increased at Ume6 targets in Ume6(Δ2–508)/RPD3+ compared to Ume6(Δ2–479)/RPD3+, which would be expected if only Isw2 is lost when residues 479–508 are deleted.

Figure 8 (next page): A Small Predicted Helix is the Isw2 Recruitment Epitope in Ume6. (A) (Top left) Schematic diagram of Ume6 truncation and deletion constructs used to identify the Isw2-recruitment epitope, with the known Sin3-interacting domain depicted as a green square, the DNA binding domain as a dark blue rectangle, and the putative Isw2-recruitment helix as a light blue rectangle. (Bottom left) Modeled helical peptide (by Phyre2) and sequence conservation of the identified Isw2-recruitment motif in Ume6 constructs from other yeasts. Asterisks denote invariant residues. (Right) Nucleosome dyad signal for Ume6 truncation and deletion strains indicates deletion of the region from residues 480 to 507 completely abrogates nucleosome positioning by Isw2 at Ume6 target sites. Vertical dashed gray lines denote wild-type (WT) positions of nucleosomes while vertical dashed pink lines indicate *isw2* or *ume6*-deficient positions of nucleosomes. (**B**) Genome Browser image showing transcript abundance at three Ume6 target sites for yeast strains lacking Rpd3 with WT Ume6 (gray), Ume6(Δ2–479) (blue), and Ume6(Δ2–508) (orange). Grossly increased transcription is seen when residues 480–507 are deleted, consistent with expected transcriptional increase associated with loss of Isw2 and Rpd3. Upstream repression sequence (URS) sites are indicated as cyan rectangles. No significant increase in transcription is detected when Ume6 residues 2–479 are deleted. Biological replicates are shown to highlight reproducibility. (**C**) (Top) Cartoon schematic for ectopic display of the Isw2-recruiting helix (residues 480–507) to the C-terminus of a truncated Ume6 construct lacking Isw2-directed nucleosome positioning. A short SpyTag is appended to the C-terminus of the Ume6 construct and residues 480–507 are fused to the SpyCatcher domain and introduced on a yeast expression vector. (Bottom) Nucleosome dyad signal demonstrating recovery of Isw2-directed nucleosome positions at a subset of Ume6 target genes by the ectopically displayed helical element. Vertical dashed gray lines denote WT positions of nucleosomes while vertical dashed pink lines indicate *isw2* or *ume6*-deficient positions of nucleosomes. URS sites are indicated as cyan rectangles.


A
S. cerevisiae Ume6
Sin3-Interacting Domain
DNA Binding Domain
D480-507
K. marxianus
K. lactis
C. glabrata
S. cerevisiae
Z. parabailii
HELIX
Nucleosome Dyads
WT
Ume6(Δ2-508)
Ume6(Δ480-508)
Ume6(Δ2-479)
SPO22
HOP1
ChrIX:225,500-227,000
B
Δrpd3
Ume6(Δ2-479) Δrpd3
Ume6(Δ2-508) Δrpd3
INO1
DMC1
ACS1
Normalized Transcript Reads
C
SpyCatcher-Helix
Ume6(Δ2-596)-SpyTag
Ume6(Δ2-596)-SpyTag-SpyCatcher-Helix
Ectopic Display of Helix
WT
Ume6(Δ2-596)
Ume6(Δ2-479)
Ectopic Helix
PDA1
SMY1
MEI4
ACA1
ChrV:547,600-548,300
ChrXI:288,000-288,800
ChrV:239,400-240,200
Nucleosome Dyads

More convincingly, only a modest increase in transcription was seen at Ume6 targets in the Ume6(Δ2–479)/Δrpd3 strain, suggesting that Isw2 is still present, while the Ume6(Δ2– 508)/Δrpd3 strain displayed extreme induction of Ume6-regulated genes, suggesting that both Isw2 and Rpd3 activity are absent (Figure 8B).

Finally, we wanted to know if the predicted helix consisting of Ume6 residues 479–508 was sufficient to bring Isw2 nucleosome positioning activity to Ume6 target sites. To test this, we employed the SpyCatcher/SpyTag system (Zakeri et al., 2012), which creates a spontaneous covalent bond between a short SpyTag peptide and a SpyCatcher domain. We fused the SpyTag peptide to the C-terminus of Ume6(Δ2–596), a construct that is incapable of positioning motif-proximal nucleosomes. We then appended Ume6 residues 479–508 to the C-terminus of the SpyCatcher domain and introduced this fusion on a yeast expression plasmid driven by the ADH1 promoter. In yeast cells, this would create a fusion protein where the helical element is ectopically displayed on the C-terminus of a DNA binding competent but nucleosome-positioning- deficient construct, connected via a SpyTag-SpyCatcher linker. This fusion protein was capable of fully recapitulating Isw2-positioned nucleosomes at a subset of Ume6 sites (Figure 8C). Perhaps not surprisingly, considering the non-native positioning of the recruitment helix in this fusion construct, not all Ume6 sites were able to gain proper nucleosome positioning with this chimeric system. We conclude that the region spanning residues 479–508 in Ume6 is a yeast-conserved Isw2-recruitment domain and is required and sufficient for recruiting Isw2 nucleosome positioning activity to Ume6 targets.

Discussion

An interacting barrier model for nucleosome array establishment

Collectively, our results give rise to an 'interacting barrier model' as an alternative means of genomic nucleosome positioning by introducing a targeted interaction between an epitope contained within condition-specific transcription factors and ISWI-type ChRPs. We show that a recruitment factor, the sequence-specific repressor Ume6, harbors a helical domain that interacts with the N-terminus of the Isw2 accessory protein Itc1. Further, we reveal this geometrically restricts the binding of the Isw2 catalytic subunit to a motif-proximal nucleosome. The complex then remodels the nucleosome, repositioning it to a specific distance from the Ume6 recognition motif. At this point and for reasons to be elucidated, this complex is strained or inactivated, and it fails to remodel any further, leaving the nucleosome in a precise location with respect to the bound recruitment factor. The activity of Isw2 and the interacting barrier sets the absolute phase of a nucleosome array that is propagated by true nonspecific spacing activities of Chd1 and Isw1 in yeast, as previously described (Gkikopoulos et al., 2011; Zhang et al., 2011; Ocampo et al., 2016). This 'interacting barrier model' of chromatin organization is more comparable to the factor-targeted activities of SWI/SNF than the nonspecific array spacing of CHD family remodelers and is potentially conserved through humans based on conservation of key interacting residues in Itc1 and the observation that Isw2 orthologs can precisely position nucleosomes adjacent to specific factors in the human genome (Wiechens et al., 2016). Together, we show that coupling between an epitope on an interacting barrier and a conserved chromatin remodeling

protein leads to robust, directional, and specific nucleosome organization at genomic regulatory elements.

Small epitopes in transcription factors organize large chromatin domains

Our data suggest that some small peptide domains embedded within transcription factors can nucleate nucleosome arrays of over 1 kb in length in vivo through an interaction with evolutionarily conserved ChRPs. Unlike the arrays established by nonspecific ChRPs, these nucleosome arrays are organized in a sequence-specific and directional manner. Establishing large swaths of chromatin structure by appending a small epitope on a genome-associated protein creates opportunity for diversity with few evolutionary constraints. Only changes in relatively small DNA binding motifs and the small peptide sequences with which they interact can have a large impact on chromatin structure.

Supporting this notion, we were able to identify a strikingly similar motif to that found in Ume6 in the unrelated cell cycle regulator Swi6, which we identified as a new Isw2- recruitment adapter for Swi4 and Mbp1. For these reasons, we find it likely that more ChRP-interacting motifs will be discovered in multiple transcription factors from a variety of organisms, and these motifs may play a significant role in sequence-specific nucleosome positioning for precisely phased and tunable nucleosome arrays in eukaryotic genomes. Importantly, the identification of such epitopes in human cells could lead to the development of targeted drugs to specifically disrupt defined remodeler–transcription regulator interactions.

Bridge from Chapter III to IV

In Chapter III, we investigated the biochemical basis for targeted chromatin remodeling by Isw2. In the course of the work described in Chapters II and III, we began developing an optimized protocol to collect mononucleosomes, described in Chapter IV. Considering our lab's interest in precisely mapping nucleosome positions, we commonly performed MNase-Seq to collect nucleosome-size fragments from yeast cultures. However, the process of MNase digestion was not well-optimized, leading us to design a shorter process, termed rapid MNase, that simplified MNase-Seq. While the time to perform the procedure was decreased, we also found that it was widely applicable, allowing us to collect nucleosomes from log-phase and quiescent yeast, *c. elegans*, and wild mushrooms, among other organisms.

For this work, I was responsible for generating sets of MNase-digested samples for NGS using both the original MNase-Seq protocol and our rapid MNase protocol. I performed these experiments for wild-type yeast and for a strain harboring several ChRP deletions, which showed disrupted nucleosome patterning. I was also involved in reviewing and editing the final manuscript. The sections provided detail our development of an efficient method for analyzing nucleosome positions in various organisms, which is fundamental to the work presented in all the chapters of this dissertation.

Bibliography

Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).

Baldi, S. et al. Genome-wide Rules of Nucleosome Phasing in Drosophila. *Mol Cell* **72**, 661-672 e4 (2018).

Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).

Bowman, G.D. & McKnight, J.N. Sequence-specific targeting of chromatin remodelers organizes precisely positioned nucleosomes throughout the genome. *Bioessays* **39**, 1-8 (2017).

Breeden, L. Start-specific transcription in yeast. *Curr Top Microbiol Immunol* **208**, 95-127 (1996).

Clapier, C.R. & Cairns, B.R. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* **492**, 280-4 (2012).

Clapier, C.R., Iwasa, J., Cairns, B.R. & Peterson, C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422 (2017).

Cunningham, F. et al. Ensembl 2015. *Nucleic Acids Res* **43**, D662-9 (2015).

Dang, W. & Bartholomew, B. Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. *Mol Cell Biol* **27**, 8306-17 (2007).

Dang, W., Kagalwala, M.N. & Bartholomew, B. Regulation of ISW2 by concerted action of histone H4 tail and extranucleosomal DNA. *Mol Cell Biol* **26**, 7388-96 (2006).

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

Donovan, D.A. et al. Engineered Chromatin Remodeling Proteins for Precise Nucleosome Positioning. *Cell Rep* **29**, 2520-2535 e4 (2019).

Fazzio, T.G. et al. Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol Cell Biol* **21**, 6450-60 (2001).

Foord, R., Taylor, I.A., Sedgwick, S.G. & Smerdon, S.J. X- ray structural analysis of the yeast cell cycle regulator Swi6 reveals variations of the ankyrin fold and has implications for Swi6 function. *Nat Struct Biol* **6**, 157-65 (1999).

Freese, N.H., Norris, D.C. & Loraine, A.E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089-95 (2016).

Fyodorov, D.V. & Kadonaga, J.T. Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly. *Mol Cell Biol* **22**, 6344-53 (2002).

Gelbart, M.E., Bachman, N., Delrow, J., Boeke, J.D. & Tsukiyama, T. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev* **19**, 942-54 (2005).

Gkikopoulos, T. et al. A role for Snf2-related nucleosome- spacing enzymes in genome-wide nucleosome organization. *Science* **333**, 1758-60 (2011).

Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M. & Tsukiyama, T. The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* **103**, 423-33 (2000).

Hauk, G., McKnight, J.N., Nodelman, I.M. & Bowman, G.D. The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Mol Cell* **39**, 711- 23 (2010).

Hota, S.K. et al. Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat Struct Mol Biol* **20**, 222-9 (2013).

Hwang, W.L., Deindl, S., Harada, B.T. & Zhuang, X. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature* **512**, 213-7 (2014).

Ito, T. et al. ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* **13**, 1529-39 (1999).

Kagalwala, M.N., Glaus, B.J., Dang, W., Zofall, M. & Bartholomew, B. Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *EMBO J* **23**, 2092-104 (2004).

Kassabov, S.R., Henry, N.M., Zofall, M., Tsukiyama, T. & Bartholomew, B. High-resolution mapping of changes in histone-DNA contacts of nucleosomes remodeled by ISW2. *Mol Cell Biol* **22**, 7524-34 (2002).

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-58 (2015).

Koch, C., Moll, T., Neuberg, M., Ahorn, H. & Nasmyth, K. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* **261**, 1551-7 (1993).

Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-71 (1974).

Krietenstein, N. et al. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167**, 709-721 e12 (2016).

Kubik, S. et al. Opposing chromatin remodelers control transcription initiation frequency and start site selection. *Nat Struct Mol Biol* **26**, 744-754 (2019).

Lai, W.K.M. & Pugh, B.F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* **18**, 548-562 (2017).

Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).

Lee, W. et al. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**, 1235-44 (2007).

Li, M. et al. Dynamic regulation of transcription factors by nucleosome remodeling. *Elife* **4**(2015).

Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

Lowary, P.T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42 (1998).

Ludwigsen, J. et al. Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. *Elife* **6**(2017).

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60 (1997).

Luger, K., Rechsteiner, T.J. & Richmond, T.J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16 (1999).

Lusser, A., Urwin, D.L. & Kadonaga, J.T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol* **12**, 160-6 (2005).

Mavrich, T.N. et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**, 1073-83 (2008).

Mavrich, T.N. et al. Nucleosome organization in the Drosophila genome. *Nature* **453**, 358-62 (2008).

McKnight, J.N. & Tsukiyama, T. The conserved HDAC Rpd3 drives transcriptional quiescence in S. cerevisiae. *Genom Data* **6**, 245-8 (2015).

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T. & Bowman, G.D. Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Mol Cell Biol* **31**, 4746-59 (2011).

McKnight, J.N., Tsukiyama, T. & Bowman, G.D. Sequence- targeted nucleosome sliding in vivo by a hybrid Chd1 chromatin remodeler. *Genome Res* **26**, 693-704 (2016).

Nair, M. et al. NMR structure of the DNA-binding domain of the cell cycle protein Mbp1 from Saccharomyces cerevisiae. *Biochemistry* **42**, 1266-73 (2003).

Ocampo, J., Chereji, R.V., Eriksson, P.R. & Clark, D.J. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Res* **44**, 4625-35 (2016).

Pointner, J. et al. CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. *EMBO J* **31**, 4388-403 (2012).

Rhee, H.S. & Pugh, B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301 (2012).

Rodriguez, J., McKnight, J.N. & Tsukiyama, T. Genome- Wide Analysis of Nucleosome Positions, Occupancy, and Accessibility in Yeast: Nucleosome Mapping, High- Resolution Histone ChIP, and NCAM. *Curr Protoc Mol Biol* **108**, 21 28 1-16 (2014).

Stockdale, C., Flaus, A., Ferreira, H. & Owen-Hughes, T. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J Biol Chem* **281**, 16279-88 (2006).

Tsukiyama, T., Palmer, J., Landel, C.C., Shiloach, J. & Wu, C. Characterization of the imitation switch subfamily of ATP- dependent chromatin-remodeling factors in Saccharomyces cerevisiae. *Genes Dev* **13**, 686-97 (1999).

Valouev, A. et al. Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516-20 (2011).

Washburn, B.K. & Esposito, R.E. Identification of the Sin3- binding site in Ume6 defines a two-step process for conversion of Ume6 from a transcriptional repressor to an activator in yeast. *Mol Cell Biol* **21**, 2057-69 (2001).

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag* (2016).

Wiechens, N. et al. The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genet* **12**, e1005940 (2016).

Yadon, A.N., Singh, B.N., Hampsey, M. & Tsukiyama, T. DNA looping facilitates targeting of a chromatin remodeling enzyme. *Mol Cell* **50**, 93-103 (2013).

Yan, C., Chen, H. & Bai, L. Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. *Mol Cell* **71**, 294-305 e4 (2018).

Yan, L., Wang, L., Tian, Y., Xia, X. & Chen, Z. Structure and regulation of the chromatin remodeller ISWI. *Nature* **540**, 466-469 (2016).

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T. & Pugh, B.F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461-73 (2012).

Zakeri, B. et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* **109**, E690-7 (2012).

Zhang, Z. et al. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332**, 977-80 (2011).

Zhou, C.Y., Johnson, S.L., Gamarra, N.I. & Narlikar, G.J. Mechanisms of ATP Dependent Chromatin Remodeling Motors. *Annu Rev Biophys* **45**, 153-81 (2016).

Zofall, M., Persinger, J. & Bartholomew, B. Functional role of extranucleosomal DNA and the entry site of the nucleosome in chromatin remodeling by ISW2. *Mol Cell Biol* **24**, 10047-57 (2004).

Zofall, M., Persinger, J., Kassabov, S.R. & Bartholomew, B. Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. *Nat Struct Mol Biol* **13**, 339-46 (2006).

CHAPTER IV

RAPID AND INEXPENSIVE PREPARATION OF GENOME-WIDE NUCLEOSOME FOOTPRINTS FROM MODEL AND NON-MODEL ORGANISMS

*This chapter contains previously published co-authored material

McKnight LE, Crandall JG, Bailey TB, **Banks OGB**, Orlandi KN, Truong VN, Donovan DA, Waddell GL, Wiles ET, Hansen SD, Selker EU, McKnight JN. Rapid and inexpensive preparation of genome-wide nucleosome footprints from model and non-model organisms. *STAR Protoc*. 2021 May 18;2(2):100486. doi: 10.1016/j.xpro.2021.100486.

Conceptualization, J.G.C. and J.N.M.; methodology, L.E.M., J.G.C., T.B.B., V.N.T., and J.N.M.; investigation, L.E.M., T.B.B., O.G.B.B., K.N.O., V.N.T., E.T.W., G.L.W., and J.N.M.; resources, G.L.W., E.T.W., D.A.D., S.D.H., and E.U.S.; writing – original draft, L.E.M. and J.N.M.; writing – reviewing and editing, L.E.M., J.G.C., T.B.B., O.G.B.B., K.N.O., V.N.T., E.T.W., G.L.W., D.A.D., S.D.H., E.U.S., and J.N.M.; visualization, J.N.M.; supervision, S.D.H., E.U.S., and J.N.M.; project administration, L.E.M. and J.N.M.; funding acquisition, J.N.M.

Introduction

MNase-seq (micrococcal nuclease sequencing) is used to map nucleosome positions in eukaryotic genomes to study the relationship between chromatin structure and DNA-dependent processes. Current protocols require at least two days to isolate nucleosome-protected DNA fragments. We have developed a streamlined protocol for *S. cerevisiae* and other fungi which takes only three hours. Modified protocols were developed for wild fungi and mammalian cells. This method for rapidly producing

sequencing-ready nucleosome footprints from several organisms makes MNase-seq faster and easier, with less chemical waste.

Methods

Before you begin

Timing: [0.5–2 h]

Here, we describe the protocol for use in budding yeast in liquid culture, which was optimized from a previously published protocol (Rodriguez et al, 2014). We have also adapted this streamlined protocol for use with *S. cerevisiae* growing on a plate, *S. cerevisiae* quiescent cells, *S. pombe*, *N. crassa*, wild mushrooms, *C. elegans, D. rerio* fin cells, and mammalian cells. The key changes to adapt the protocol to other cell stages or organisms are noted in the relevant step of the protocol.

1. If needed, prepare buffers (see materials and equipment for recipes).

    a. The following stock solutions can be prepared ahead of time and stored for an extended period at 25°C: 2M sorbitol, 1M Tris pH 7.5, 2.5M glycine, STOP buffer.

    b. Stock solutions of Proteinase K (20 mg/mL) and RNase A (10 mg/mL) can be prepared ahead of time and stored for an extended time at −20°C.

    c. MNase (20 units/μL) and MNase digestion buffer can be aliquoted and stored long-term at −80°C. We have observed that MNase aliquots are stable for at least a year, but have not tested longer storage.

2. Ensure that the formaldehyde being used is less than 3 months old.

3. 2 days before the start of the experiment (day -2), streak the strain(s) of interest onto appropriate selection plates and incubate at 30°C. Sick strains, or strains grown in synthetic medium, may take longer to grow.

Note: If isolating quiescent (Q) cells, you will need an additional 3 days before starting the experiment, so cells need to be streaked out at least a week prior.

4. The day before the experiment (day -1), start an overnight culture of yeast in YPD or other appropriate media; grow at 30°C with shaking for 18–20 h.

Note: If isolating Q cells, start an overnight culture on day -4 or earlier, as the culture must grow for at least 3 days to reach quiescence.

Note: If working from yeast on a plate, instead of an overnight culture, make a large patch (1cm × 3 cm) on a plate with appropriate selection if needed; sick strains may take up to 48 h to grow.

Preparing cell pellets

Timing: [3–5 h]

5. On Day 0 (or Day 1, if proceeding directly to crosslinking) dilute an overnight culture of yeast to an OD_{600} = 0.2–0.3 in 25 mL of YPD, or other appropriate media. Grow at 30°C with shaking, to an OD_{600} = 0.8–1.

Note: For more slowly growing strains, a larger volume may be used and grown to a lower OD_{600}; ensure at least one doubling takes place before pelleting. The goal is to have 20 OD∗mL of cells (i.e., 20ml of OD_{600} = 1, or 25ml of OD_{600} = 0.8).

6. Pellet 20 OD∗mL of cells in a 50-mL conical tube at 3000×*g* for 5 min.

Note: See Table 2 for information on input of different cell types.

7. Carefully remove supernatant. Resuspend pellet in 1 mL of deionized water and transfer to a 1.5-mL microcentrifuge tube.

Optional: If time permits, you can proceed directly to the crosslinking protocol at this point instead of pelleting, flash freezing and storing the cells ahead of time; there is a Pause Point after crosslinking.

8. Pellet at 20,000 × *g* for 1 minute in a microcentrifuge. Remove supernatant.

9. Flash freeze the pellet in liquid nitrogen and store it at −80°C. Samples are stable for at least 2 or 3 months at −80°C; we have not tested longer storage times.

Step-by-step method details

Prepare cross-linked cells

Timing: [20−**30 min]**

In this step, formaldehyde is used to capture interactions between DNA and nucleosomes. The crosslinking reaction is then quenched by glycine. For yeast, cells are generally crosslinked with formaldehyde to maintain nucleosome positions throughout the subsequent steps, though it has been debated whether these crosslinks can efficiently

trap nucleosome positions without introducing artifacts (Henikoff et al., 2011; Cole et al., 2012). It may be preferable to crosslink cells on the same day as growth, though we have achieved good results with the method below. Due to natural variation in cells, we recommend performing the experiment on three biological replicates.

MNase-seq of mammalian cells does not require crosslinking. We performed experiments on the cell line PLB-895 (Tucker et al., 1987) both with and without crosslinking and there was no noticeable difference in visualization of nucleosome footprints (Figure 6), though we recommend doing a side-by-side comparison when working with a new cell type for the first time, if possible.

1. Thaw MNase digestion buffer on ice; you will need 100 µL per sample plus another 10 µL per 10 samples to dilute the ExoIII.

2. Resuspend cell pellet in 1 mL of deionized water.

Note: Crosslinking is optional for mammalian cells; if skipping this step, proceed directly to step 8

3. Add 27 µL of 37% formaldehyde to a final concentration of 1%. Rotate end-over-end for 15 min at 25°C.

    a. While the cells are cross-linking, prepare spheroplast buffer; you will need 1 mL per sample with 2 mg zymolyase per mL.

4. To quench the crosslinking reaction, add 50 µL of 2.5M glycine for a final concentration of 125 mM.

5. Pellet at 20,000 × *g* for 1 minute in a microcentrifuge and remove supernatant carefully.

Pause Point: At this point, cells can be flash frozen in liquid nitrogen and stored at −80°C. Samples are stable for at least 2 or 3 months at −80°C; we have not tested longer storage times.

Make spheroplasts and digest chromatin

Timing: [3−**4 h]**

In this step, the cell walls of fungi are broken and nuclei are permeabilized to allow micrococcal nuclease (MNase) to access and digest extranucleosomal DNA. Permeabilization is required for fungi but not for mammalian cells, *C. elegans*, or *D. rerio*; for non-fungal samples proceed directly to step 8. [Troubleshooting Problem 1:]

After permeabilization, accessible regions of genomic DNA are digested by MNase, and the ends of the resulting nucleosome fragments are trimmed by Exonuclease III to reduce variation in size, which enables consistent, high-resolution identification of the nucleosome dyad at the center of the core particle. [Troubleshooting Problem 2:]

Cellular RNA is removed, formaldehyde crosslinks are reversed, and protein is digested prior to DNA purification. Finally, the residual 3’ phosphates (from MNase cleavage activity) are removed prior to genomic library construction.

6. Resuspend the pellet of crosslinked cells (fresh or previously frozen) in 1 mL spheroplast buffer plus 2 mg zymolyase to permeabilize the cell walls. Rotate at 25°C for 15 min.

a. While the sample incubates, dilute ExoIII 1:10 (from 100 U/µL stock) in MNase digestion buffer; you will need 3 µL of diluted ExoIII per sample. Keep on ice.

b. Thaw MNase (3 µL per sample), RNase (3 µL per sample), and Proteinase K (10 µL per sample) on ice.

Note: For yeast taken directly from a plate, increase spheroplasting incubation time to 30 min. For purified Q cells, spheroplast for 60 min and increase the amount of zymolyase to 10 mg.

7. Pulse-pellet (15 seconds at 20,000 × *g*); carefully remove supernatant.

8. Add 100 µL digestion buffer + 3 µL MNase (20 U/µL; 60U total) + 3 µL ExoIII (10 U/µL). Incubate at 37°C for 10 min to digest chromatin.

Note: For yeast from a plate, use 1.5µl MNase. For Q cells, use 0.2µl MNase and incubate for 2–10 min (see troubleshooting 3 for more information on optimizing this protocol for Q cells).

Critical: Exact timing of digestion across samples is essential for consistency. To achieve an identical start time for digest, we recommend pipetting the MNase into the open lid of the microcentrifuge tube, and then briefly spinning all the tubes down so that the enzyme reaches each sample at precisely the same time.

9. Stop digestion with 12.5 µL STOP buffer.

Note: We have eliminated SDS from the STOP buffer and switched the order of the Proteinase K and RNAse A steps in order to streamline the protocol. Traditionally, SDS was included in the STOP buffer and then Proteinase K was added (Rodriguez et al, 2014); both of these components are used to break down protein in the sample, but they also inactivate RNAse, necessitating an additional purification step before proceeding to the RNAse step. In our version, EDTA and EGTA are sufficient to halt crosslinking, and no extra purification is needed.

10. Add 3 µL RNase A (10 mg/mL stock). Incubate at 42°C for 30 min.

11. Add 12.5 µL 10% SDS + 10 µL Proteinase K (20 mg/mL stock) to digest protein and reverse crosslinks. Incubate at 65°C for 45 min.

Note: For uncrosslinked mammalian cells, the incubation time is reduced to 15 min

12. Add 700 µL Buffer PB (from Qiagen kit). Purify using a single Qiagen MinElute PCR column. Follow protocol instructions, with the exception of the elution step. Elute with 12 µL 1× CutSmart buffer (NEB; diluted from 10× stock)

Note: For wild mushrooms, after adding Buffer PB centrifuge the sample for 2 min at 20,000 × *g* to further clear remaining cellular debris. Apply the supernatant to the MinElute column.

13. Add 1 µL Quick CIP. Incubate for 10 min at 37°C.

14. Add 7 µL Orange Loading Dye (NEB) and run the entire sample on a 2% agarose gel.

Critical: The use of Orange Loading Dye and a 2% gel is required for visualization and separation of the digested bands.

Note: Using a 7cm × 10cm or 15cm × 10cm TAE-agarose gel in TAE, it takes approximately 40 min at 100V to get proper separation of bands

15. Cut out the mononucleosome band, gel extract with a Qiagen MinElute column. Follow protocol instructions, with the exception of the elution step. Elute with 12 µL EB.

16. Quantify with PicoGreen or other preferred method; store at −80°C. DNA yield typically ranges from 10-50 ng total. This sample is now ready for preparing libraries; we use 5 ng of input with the Ovation Ultralow Library System V2 (from NuGEN), though we have successfully sequenced libraries prepared from as little as 0.5 ng input. There are a variety of options for preparation of libraries, and this protocol can likely feed into whatever pipeline you are currently using. We have not tested how long samples are stable at −80°C, but we usually prepare library samples within a month.

Expected outcomes

The first step toward validating the nucleosome footprints obtained using this protocol is assessing the quality of the digests on agarose gel (Step 14). Figure 9 shows an example of well-digested DNA, where bands are evident for the tri-, di-, and mono-nucleosomal fragments. Some variation is acceptable, but over- or under-digested DNA may yield poor quality data (see troubleshooting, Problem 2).

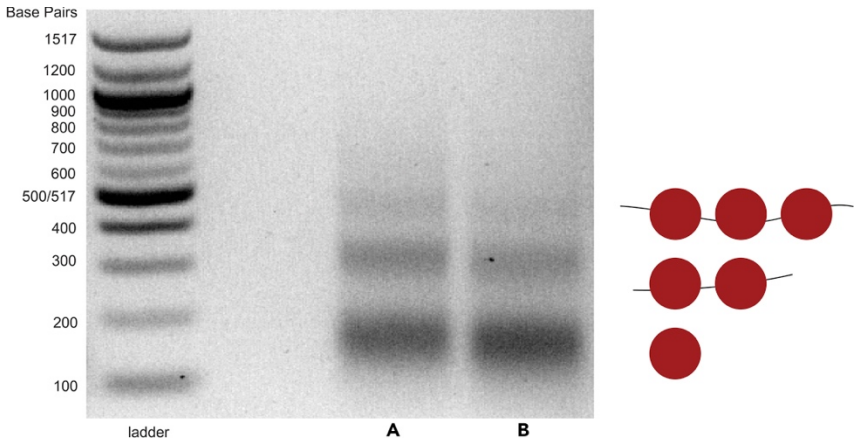


Figure 9. Sample gel showing properly digested nucleosome footprints Agarose gel showing nucleosome footprints recovered from liquid culture of S. cerevisiae using the rapid MNase protocol. (A) and (B) are replicates from the same WT yeast strain.

After preparing libraries from the S. cerevisiae samples obtained using our new rapid protocol and a previously published protocol (Rodriguez et al., 2014), we sequenced them and analyzed the data using our standard protocol (McKnight and Tsukiyama 2015; Donovan et al., 2019). We compared our data to previously published datasets (Donovan et al., 2019; McKnight et al., 2016) for wild type and isw2 yeast (Figure 10). We chose Isw2-deficient yeast because Isw2 is required for nucleosome shifts at specific target loci, which allows us to determine if our protocol can recapitulate Isw2-specific nucleosome positions at transcription start sites. For all samples, nucleosome organization at transcription start sites (TSSs) displayed the stereotypical structure, with a nucleosome-depleted region flanked by packed nucleosome arrays (Figure 10A). Comparison of nucleosome positions at Isw2 targets showed that nucleosomes were detected at strain-specific but not protocol-specific locations (Figure 10B). Both the rapid protocol and the standard protocol recovered strain-specific nucleosome positioning at Ume6 binding sites, a known Isw2-recruitment protein (Goldmark et al., 2000) across the genome (Figure 10C and 10D).

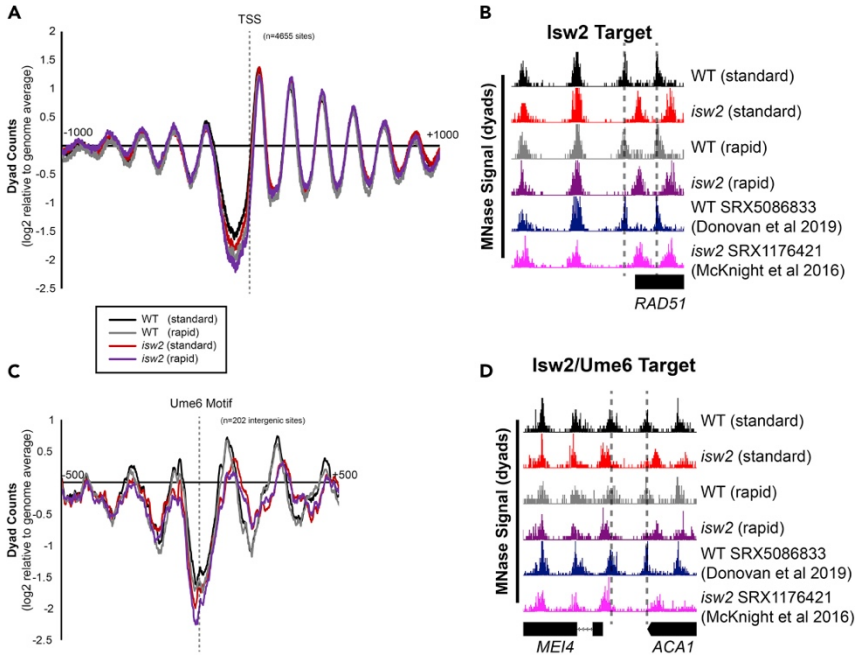


Figure 10. Rapid MNase can accurately map nucleosome positions in S. cerevisiae cells (A) Nucleosome dyad signal at 4655 yeast transcription start sites (TSSs) are plotted for WT and isw2 nucleosomes harvested using a standard or rapid protocol. **(B)** Example Genome Browser image showing the standard method and rapid method can map Isw2-directed nucleosome positions similar to published data sets at the RAD51 locus. Dashed lines indicate Isw2-positioned nucleosomes. **(C)** Nucleosome dyad signal at 202 intergenic Ume6 target sites showing rapid and standard MNase methods can accurately identify global changes in nucleosome structure at an Isw2 recruitment motif. **(D)** Genome Browser image showing all methods correctly identify Isw2-positioned nucleosomes at the MEI4-ACA1 locus, a Ume6 target site. Dashed lines indicate Ume6- and Isw2- positioned nucleosomes.

We also applied the rapid MNase protocol to purified quiescent cells (Allen et al., 2006), or Q cells. As previously observed, quiescent yeast required longer spheroplasting with an increased amount of zymolyase (McKnight et al., 2015) to compensate for a highly fortified cell wall (Li et al., 2015). We also tested different amounts of time for the MNase digest (Figure 11A), as we found our Q cell protocol required more adjustments to get acceptable footprints, which were not entirely consistent across samples of Q cells prepared on different days. We hypothesize that this is because isolation of Q cells involves several steps and therefore produces more heterogeneous cell samples. For this

reason we recommend testing different MNase digest times and/or amounts of MNase when processing Q cells to obtain properly-digested nucleosomes.

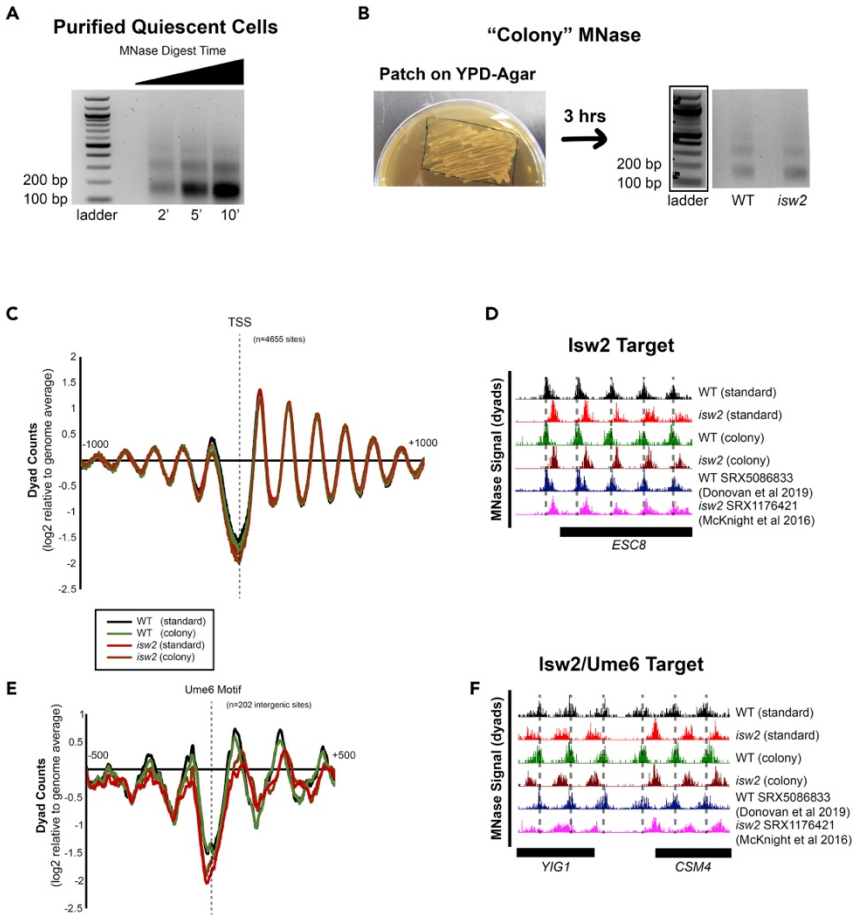


Figure 11. Rapid MNase can recover nucleosome footprints from isolated quiescent cells and yeast patches (A) Representative gel showing nucleosome footprints recovered from purified quiescent cells using the rapid MNase protocol. **(B)** Representative gel (right) showing nucleosome footprints recovered from a fresh patch of yeast collected from YPD-Agar (left). **(C)** Nucleosome dyad signal at transcription start sites (TSSs) comparing standard and patch- recovered "colony" MNase footprints. **(D)** Example Genome Browser image showing "colony" MNase footprints can accurately identify Isw2-directed nucleosome positions at the ESC8 locus compared to the standard MNase protocol and published data sets. **(E)** Nucleosome dyad signal at 202 intergenic Ume6 target sites showing "colony" MNase can accurately identify global changes at an Isw2 recruitment motif. **(F)** Genome Browser image showing colony MNase can similarly identify Isw2-positioned nucleosomes at the YIG1-CSM4 locus, an example Ume6 target site. Dashed lines indicate Ume6- and Isw2-positioned nucleosomes.

We found that we could also reproducibly recover nucleosome footprints from patches of yeast grown on agar plates, or "colony MNase" (Figure 11B) and the captured nucleosome positions accurately reflect nucleosome positioning across the yeast genome (Figures 11C and 11D). Similar to liquid culture, the rapid colony MNase protocol can also detect Isw2-specific nucleosome events genome-wide (Figures 11E and 11F). It is important to note that colony MNase is not appropriate for all types of experiments, and researchers should consider whether stationary or log-phase cells are more appropriate for their needs.

Without any protocol modifications, we were able to recover a well-defined and appropriately digested nucleosome ladder from wild type S. pombe cells (Figure 12A). We compared our sequencing data set with previously published MNase-seq data sets from S. pombe (DeGennaro et al., 2013; Steglich et al., 2015). Genomic nucleosome dyad positions from samples prepared by the rapid MNase protocol were the same as seen previously (Figure 12B). In addition, global nucleosome positioning at S. pombe TSSs was nearly identical across data sets (Figure 12C). We also performed the rapid MNase protocol on crosslinked, germinated N. crassa (Figure 12D). Importantly, we verified that the genomic nucleosome positions we obtained were similar to previously published N. crassa data sets (Seymour et al., 2016; Klocko et al., 2019) (Figure 12E).

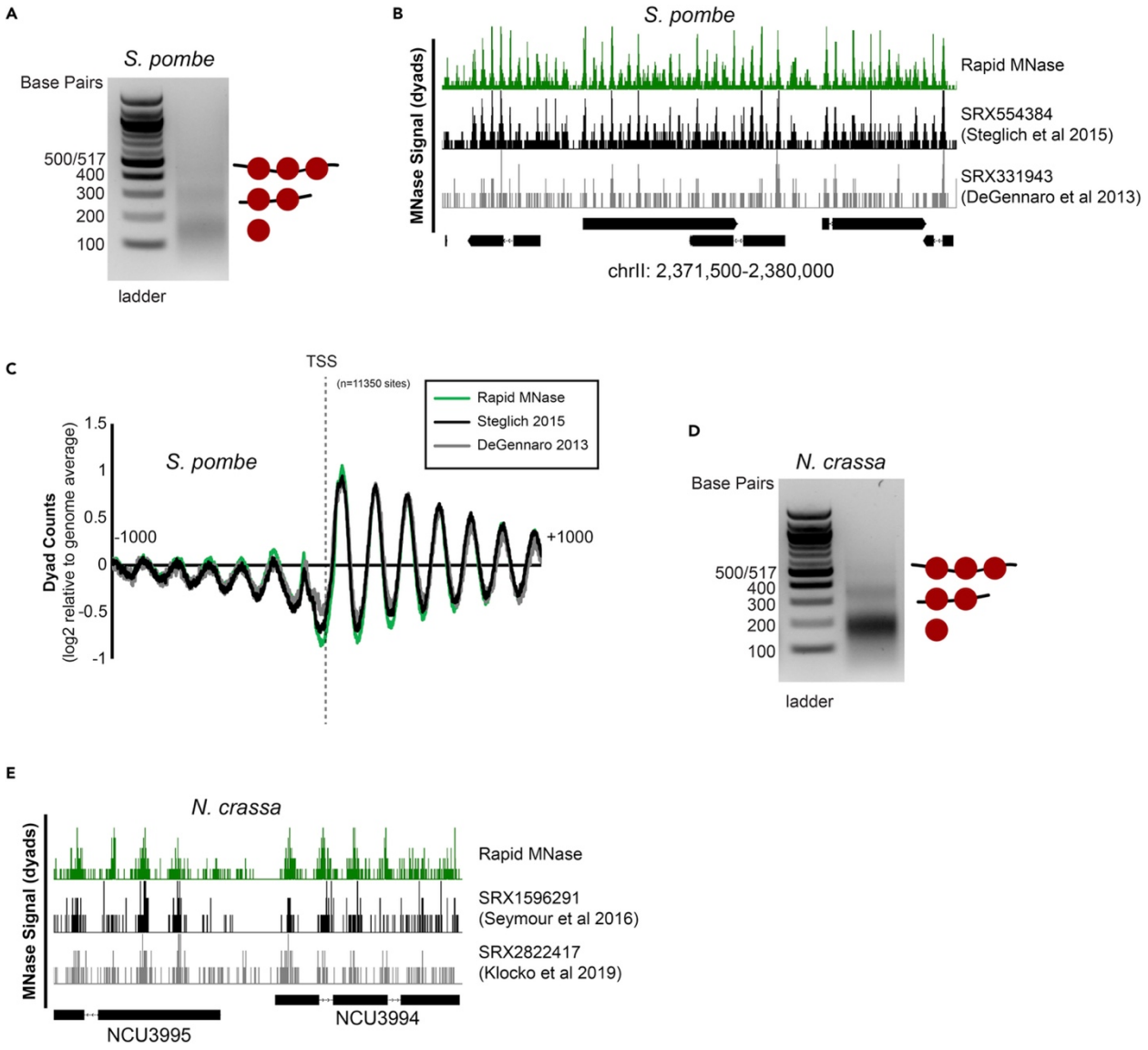


Figure 12. Rapid MNase can accurately map nucleosome positions in S. pombe and N. crassa (A) Agarose gel showing example nucleosome footprints recovered from S. pombe using the rapid MNase protocol. **(B)** Genome Browser image comparing nucleosome dyad positions on S. pombe chrII recovered for the rapid MNase protocol (top) and previously- published data sets. **(C)** Alignment of nucleosome dyads at 11,350 transcription start sites (TSSs) for rapid MNase-recovered nucleosome footprints and previously-published data sets. **(D)** Agarose gel showing example nucleosome footprints recovered from N. crassa using the rapid MNase protocol. **(E)** Genome Browser image comparing nucleosome dyad positions at the N. crassa NCU3995-NCU3994 locus for the rapid MNase protocol (top) and previously-published data sets.

We wished to design a rapid protocol that was standardized across species, so we performed experiments on a variety of locally foraged wild fungi and were able to obtain well-spaced nucleosome footprints (Figure 13). It is important to note that the size of the mononucleosome band is usually between 100 and 200 bp, but this product might run slightly larger in different organisms, as in lane 2 of Figure 13A. Because we did not sequence this sample, we cannot be absolutely certain that this band reflects mononucleosomal fragments, but we are reasonable sure. We also tested C. elegans, D. rerio fin cells, and the cell line PLB-895 (Tucker et al., 1987); while we do not have gel images for the *C. elegans* or *D. rerio* samples, we saw a prominent mononucleosome band for both of these organisms after a single test using the same protocol as for mammalian cells (with crosslinking), and we are confident that this protocol can be easily adapted for sequencing-quality nucleosome footprints. For PLB-895 cells, we performed the protocol with crosslinking in order to better compare it to our other samples (Figure 14A), but we also prepared samples without crosslinking as this is the more standard practice for MNase of mammalian cells (Figure 14B); there is no obvious difference visually. A 1-day MNase protocol already exists for mammalian cells (Ramani et al., 2019), but ours is significantly shorter and provides a promising standardized alternative to previously published protocols across organisms. We did not sequence these nucleosome fragments in the interest of time and cost, along with the lack of annotated genomes for wild fungi, but we encourage users to test the rapid protocol in their model organism of choice.

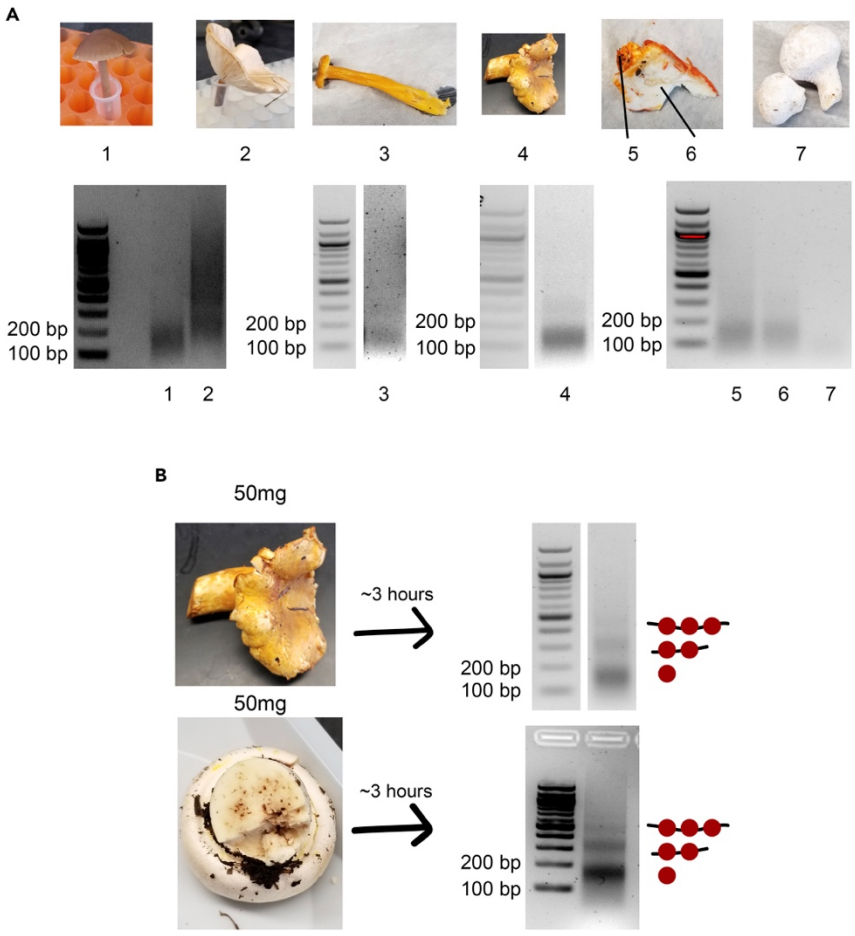


Figure 13. Nucleosome footprints can be rapidly recovered from wild mushroom samples (A) Images of locally-foraged wild mushrooms that were subjected to the rapid MNase protocol (top). Sample 5-6 consists of a distinct surface fungal specimen (5) growing on a host specimen (6). Recovered nucleosome footprints for corresponding mushrooms are shown (bottom). Speculative identities for these samples are (1) Panaeolus foenisceii, (2) unknown, (3) Craterellus tubaeformis, (4) Cantharellus formosus, (5) Hypomyces lactifluorum, (6) Russula brevines, (7) Lycoperdon perlatum. **(B)** Optimized rapid MNase for specimen 4 (chanterelle) was achieved using 50 mg of tissue leading to well-digested nucleosome footprints (top). The optimized protocol was performed on 50 mg of a previously-untested specimen leading to well-digested nucleosome footprints (bottom). Speculative identity of these samples are Cantharellus formosus (top) and Agricus xanthodermus (bottom).

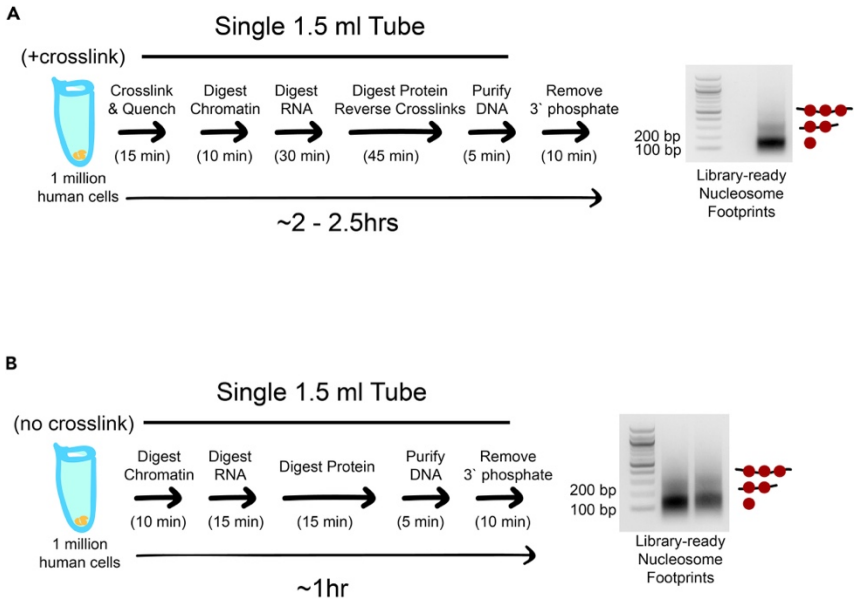


Figure 14. The rapid MNase protocol can be performed on human cells with or without crosslinking (A) Cartoon schematic showing rapid MNase protocol and associated nucleosome footprints for 1 million human cells from the diploid myeloid leukemia cell line PLB-895. The protocol is identical to the yeast liquid culture protocol except that there is no zymolyase treatment. **(B)** Cartoon schematic showing rapid MNase protocol for PLB-895 cells and associated nucleosome footprints when the crosslinking and crosslinking reversal steps are omitted and other steps are shortened. The crosslink-free protocol can provide nucleosome footprints that are ready for library construction in less than an hour. Left lane: 1.5 million cells input; right lane: 1 million cells input.

Quantification and statistical analysis

Libraries were sequenced at the University of Oregon's Genomics and Cell Characterization Core Facility on an Illumina NextSeq500 on the 37 cycle, paired-end, High Output setting, yielding approximately 10–20 million paired reads per sample. MNase sequencing data were analyzed as described previously (McKnight and Tsukiyama 2015). Paired-end reads were aligned to the S. cerevisiae sacCer3 (Yates et al, 2019), S. pombe (Wood et al., 2002), or N. crassa (Galagan et al., 2003) reference genome with Bowtie 2 (Langmead and Salzberg 2012), and filtered computationally for unique fragments between 100 and 200 bp. Dyad positions were calculated as the

midpoint of alignment coordinates, then dyad coverage was normalized across the genome for an average read/bp of 1.0. Nucleosome alignment to the Ume6 binding site, WNGGCGGCWW, was performed by taking average dyad signal at each position relative to all intergenic instances of a motif center. Intergenic instances of the Ume6 motif were found using the Saccharomyces Genome Database Pattern Matching tool (https://yeastgenome.org/nph-patmatch). Transcription start sites were obtained from published datasets for S. cerevisiae (Nagalakshmi et al., 2008) and S. pombe (Thodberg et al., 2019). Previously-published S. cerevisiae data (SRX5086833, (Donovan et al., 2019); SRX1176421, (McKnight et al., 2016)), S. pombe data (SRX554384 (Steglich et al., 2015); SRX331943, (DeGennaro et al 2013)), and N. crassa data (SRX1596291, (Seymour et al., 2016); SRX2822417, (Klocko et al., 2019)) were downloaded from the Sequence Read Archive (SRA) and analyzed using our computational pipeline to identify nucleosome dyad positions. Data were visualized using Integrated Genome Browser (Freese et al., 2016). Sequencing data from this work can be accessed at the GEO database under accession code GSE141676.

Troubleshooting

Problem 1

Genomic DNA is not being digested, as evidenced by a very high molecular weight band on the agarose gel and little or no signal for nucleosome footprints, as in Figure 15.

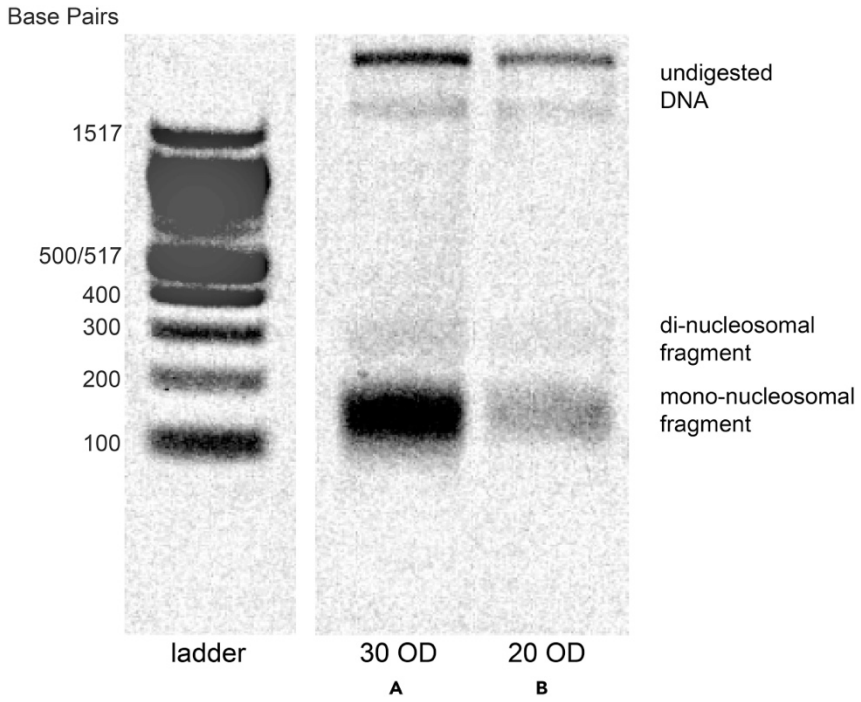


Figure 15. Sample gel showing intact genomic DNA and partial nucleosome footprints from non-permeabilized cells Agarose gel showing example nucleosome footprints recovered from patches of S. cerevisiae using unoptimized "rapid colony MNase" protocol (2 mg of zymolyase for 15 mi) using an input of **(A)** 30 OD pellet of yeast, or **(B)** 20 OD pellet of yeast.

Potential solution

This most likely is the result of cells not being permeabilized. We suggest extending the incubation time for the spheroplasting step (Step 6) or increasing the

amount of zymolyase per mL of buffer in 1-mg increments. Another option is to perform this step at 37°C rather than at 25°C. After spheroplasting, the solution should be much less opaque than it was prior to this step, so it is possible to troubleshoot this step without completing the entire protocol. Furthermore, there are likely organisms that may require additional steps to help permeabilize cells, particularly if the cells possess zymolyase-resistant cell walls. Previous work has demonstrated success using cryogrinding as the cell-breaking step (Gonzalez and Scazzocchio 1997; Givens et al., 2011).

Problem 2

The DNA is over- or under-digested, as evidenced by too much or too little signal for the mononucleosome band. Examples of under-digested samples are shown in lanes B, C, and E of Figure 16. Examples of over-digested samples are shown in lanes D and F. Lane A shows a sample where fewer than 20 OD of cells were pelleted initially.

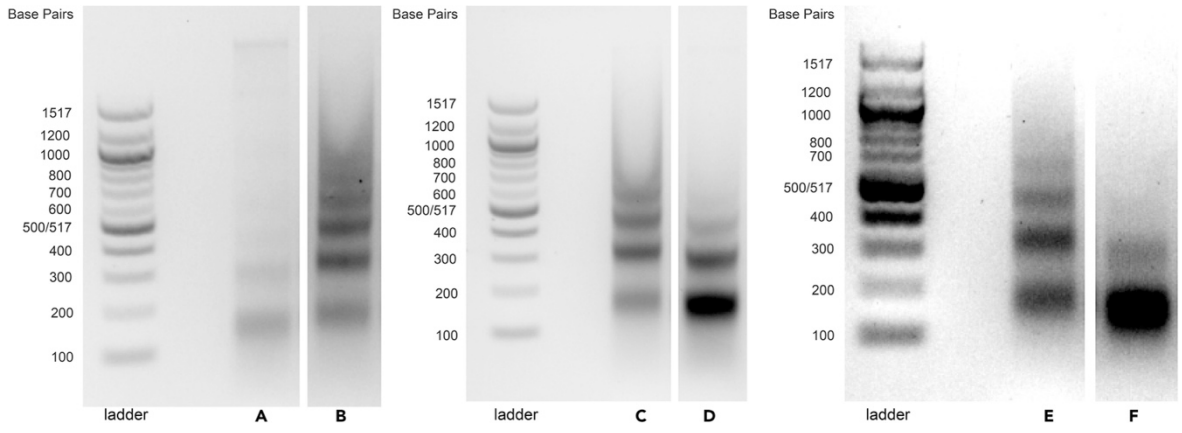


Figure 16. Agarose gel showing example nucleosome footprints recovered from *S. cerevisiae* liquid culture using the rapid MNase protocol (A) Only 15 OD of cells were used as input. **(B)** Under-digested sample. **(C)** Under-digested sample. **(D)** Over-digested sample. **(E)** Under-digested sample. **(F)** Over-digested sample.

Potential solution

MNase activity can vary across lots and vendors, so it is critical to calibrate the MNase concentration to give the desired extent of digestion. We recommend performing a test of 3 samples with either 2 µL, 3 µL, or 4 µL of MNase (20 U/µL stock) in Step 6 for S. cerevisiae; different amounts of MNase should be chosen for other types of samples based on Table 2. When running these tests, it is not necessary to treat samples with CIP, so Step 13 of the protocol can be skipped. An ideal digest will result in signal for mono-, di-, and tri-nucleosome fragments, with the most signal for the mono-nucleosome band, as in Figure 9.

Sometimes it is necessary to adjust the amount of MNase used in the protocol based on your specific enzyme activity, particularly when working with an organism that we have not tested. It is worth noting that often it is still possible to excise the mononucleosome band and obtain high-quality data of over-digested samples, but it is preferable to work with appropriately digested footprints, and the extent of digestion should be consistent across samples that are being compared in a particular experiment.

If the ratio of mono- to di-nucleosome bands appears to be correct, the issue may actually be that too few cells were used, as in lane A of Figure 16. Within an organism, the number of cells used could vary depending on the growth conditions or media composition, as evidenced by the adjustments needed for colony MNase. It therefore possible that specific conditions or mutant strains may require subtle changes to digestion or input amount, particularly if it is challenging to accurately quantify the initial number of cells.

Problem 3

Quiescent cells produce improperly-digested footprints. Similar to Problem 2, if DNA appears under- or over-digested for quiescent cells, there are additional steps we took to optimize these samples.

Potential solution

Assuming that the MNase activity has been calibrated as in Problem 2, there are additional considerations when processing quiescent cells. We found that quiescent cells isolated on different days were not consistently digested, so we recommend preparing multiple MNase samples as standard practice, and testing different incubation times for MNase digest in Step 8, as shown in Figure 11A. For the sample shown in Figure 11A, the mononucleosome band from the 5-minute sample was excised, but for other Q cells the 5-minute time point was over- or under-digested. As indicated in the protocol, using multiple different quiescent cells, we observed a range of 2–10 min to yield appropriate footprints, but due to the variability we have seen, it is conceivable that the optimum range might be shifted for quiescent samples prepared in another lab, as the technique seems sensitive to small changes. In our hands, once we optimized the amount of MNase, the same amount of MNase worked for a variety of Q cell samples.

Problem 4

Unexpectedly low DNA yield, as evidenced by faint bands or no signal on the agarose gel.

Potential solution

Increase the input amount when preparing the sample. Assuming calculations for input were correct, we recommend doubling or even tripling the amount of sample as a

test. Yeast cells in log phase versus quiescence, for example, give slightly different OD readings, and sometimes sick strains, certain mutants, or yeast grown in different medium vary in size, so the OD reading and pellet size do not accurately reflect the number of cells and therefore amount of DNA in the sample. Furthermore, the amount of input required can vary drastically for different organisms, so for a sample type that we have not tested, optimization of the amount of input will be necessary.

Problem 5

Instead of nucleosome footprints, only a large smear is visible on the agarose gel.

Potential solution

This is likely due to RNA and/or residual protein in the sample. We most often see this with improperly digested RNA and suggest first increasing the incubation time with RNAse A in Step 10. If that is not effective, increase the incubation time with Proteinase K in Step 11. Another possibility is that the RNAse A and/or Proteinase K stocks have lost activity, and these may need to be re-made.

Bridge from Chapter IV to V

Following the experience working on rapid MNase, we turned our attention to another outstanding concern in the lab, centered around detecting protein-DNA interactions without the need for crosslinking or immunoprecipitation. While originally intended as a method to monitor Isw2 localization in the yeast genome, the results we obtained from Spycatcher Linked Targeting of Chromatin Endogenous Cleavage show it produces quality results that may be generally useful, even for proteins that can be studied with standard factor mapping techniques.

For this project, I generated the GAL-inducible MNase-SpyCatcher002 plasmid and the SpyTag002 tagging plasmid using existing vectors. I engineered the yeast strains, collected each of the SpLiT-ChEC samples, prepared the NGS libraries, and designed the computational pipeline to process NGS data from FASTQs to visualization. I created several tools and additional pipelines for this analysis, all of which are available on GitHub (Bankso/SCAR and Bankso/SEAPE). Last, I prepared the figures, wrote the first draft of the manuscript, and was involved in editing.

References Cited

Allen, C., Buttner, S., Aragon, A.D., Thomas, J.A., Meirelles, O., Jaetao, J.E., Benn, D., Ruby, S.W., Veenhuis, M., Madeo, F., et al. (2006). Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. J. Cell Biol. *174*, 89-100.

Cam, H.P. and Whitehall, S. (2016). Micrococcal nuclease digestion of *schizosaccharomyces pombe chromatin*. Cold Spring Harb. Protoc. *2016*, 996-1000.

Cole, H.A., Howard, B.H., and Clark, D.J. (2012). Genome-wide mapping of nucleosomes in yeast using paired-end sequencing. Methods Enzymol. *513*, 145-168.

DeGennaro, C.M., Alver, B.H., Marguerat, S., Stepanova, E., Davis, C.P., Bähler, J., Park, P.J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol. Cell Biol. *33*, 4779-4792.

Donovan, D.A., Crandall, J.G., Banks, O.G.B., Jensvold, Z.D., Truong, V., Dinwiddie, D., McKnight, L.E., and McKnight, J.N. (2019). Engineered chromatin remodeling proteins for precise nucleosome positioning. Cell Rep. *29*, 2520-2535.e4.

Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. Bioinformatics *32*, 2089-2095.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., et al. (2003). The genome sequence of the filamentous fungus Neurospora crassa. Nature *422*, 859-868.

Givens, R.M., Mesner, L.D., Hamlin, J.L., Buck, M.J., and Huberman, J.A. (2011). Integrity of chromatin and replicating DNA in nuclei released from fission yeast by semi-automated grinding in liquid nitrogen. BMC Res. Notes *4*, 499.

Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M., and Tsukiyama, T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell *103*, 423-433.

Gonzalez, R. and Scazzocchio, C. (1997). A rapid method for chromatin structure analysis in the filamentous fungus Aspergillus nidulans. Nucleic Acids Res. *25*, 3955-3956.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. U S A *108*, 18318-18323.

Klocko, A.D., Uesaka, M., Ormsby, T., Rountree, M.R., Wiles, E.T., Adhvaryu, K.K., Honda, S., and Selker, E.U. (2019). Nucleosome positioning by an evolutionarily conserved chromatin remodeler prevents aberrant DNA methylation in neurospora. Genetics *211*, 563-578.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359.

Li, L., Miles, S., and Breeden, L.L. (2015). A genetic screen for saccharomyces cerevisiae mutants that fail to enter quiescence. G3 (Bethesda) *5*, 1783-1795.

McKnight, J.N., Boerma, J.W., Breeden, L.L., and Tsukiyama, T. (2015). Global promoter targeting of a conserved lysine deacetylase for transcriptional shutoff during quiescence entry. Mol. Cell *59*, 732-743.

McKnight, J.N. and Tsukiyama, T. (2015). The conserved HDAC Rpd3 drives transcriptional quiescence in S. cerevisiae. Genom. Data *6*, 245-248.

McKnight, J.N., Tsukiyama, T., and Bowman, G.D. (2016). Sequence- targeted nucleosome sliding in vivo by a hybrid Chd1 chromatin remodeler. Genome Res. *26*, 693-704.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Ramani, V., Qiu, R., and Shendure, J. (2019). High sensitivity profiling of chromatin structure by MNase-SSP. Cell Rep. *26*, 2465-2476.e4.

Rodriguez, J., McKnight, J.N., and Tsukiyama, T. (2014). Genome-wide analysis of nucleosome positions, occupancy, and accessibility in yeast: nucleosome mapping, high-resolution histone ChIP, and NCAM. Curr. Protoc. Mol. Biol. *108*, 21.28.1-21.28.16.

Seymour, M., Ji, L., Santos, A.M., Kamei, M., Sasaki, T., Basenko, E.Y., Schmitz, R.J., Zhang, X., and Lewis, Z.A. (2016). Histone H1 limits DNA methylation in neurospora crassa. G3 (Bethesda) *6*, 1879-1889.

Steglich, B., Stralfors, A., Khorosjutina, O., Persson, J., Smialowska, A., Javerzat, J.P., and Ekwall, K. (2015). The Fun30 chromatin remodeler Fft3 controls nuclear organization and chromatin structure of insulators and subtelomeres in fission yeast. PLoS Genet. *11*, e1005101.

Thodberg, M., Thieffry, A., Bornholdt, J., Boyd, M., Holmberg, C., Azad, A., Workman, C.T., Chen, Y., Ekwall, K., Nielsen, O., et al. (2019). Comprehensive profiling of the fission yeast transcription start site activity during stress and media response. Nucleic Acids Res. *47*, 1671-1691.

Tucker, K.A., Lilly, M.B., Heck, L., and Rado, T.A. (1987). Characterization of a new human diploid myeloid leukemia cell line (PLB-985) with granulocytic and monocytic differentiating capacity. Blood *70*, 372-378.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2019). Ensembl 2020. Nucleic Acids Res. *48*, D682-D688.

CHAPTER V

MAPPING DNA-ASSOCIATED PROTEINS AND PROXIMAL CHROMATIN ELEMENTS IN THE BUDDING YEAST GENOME USING SPYCATCHER-LINKED TARGETING OF CHROMATIN ENDOGENOUS CLEAVAGE (SPLIT-CHEC)

Introduction

DNA-binding proteins (DBPs) have well-documented roles in replication (Bell et al., 2001), metabolic adaptation to environmental inputs (Miles et al., 2013), and tissue differentiation (Ferri et al., 2007). Many DBPs, including transcription factors (TFs) (Ju et al., 1990; Rhode et al., 1992), chromatin remodeling proteins (Cairns et al., 1996; Flaus et al., 2006; Clapier and Cairns, 2009), and histones (Huang and Bonner, 1962; Kornberg, 1974; Luger et al., 1997), form detectable complexes with DNA and bind throughout the genome of eukaryotic organisms. Protein-DNA interactions *in vivo* are critical for targeting functional outputs of DBPs and any interacting proteins (Dangkulwanich et al., 2014); a hallmark example are TFs, which bind to target sites in the genome and influence gene expression (Lee et al., 2002). Protein-DNA interactions in the context of TFs generally occur between specific DNA motifs and the DNA-binding domain of the TF (Zamanighomi et al., 2017; Jolma et al., 2013). TFs often interact with transcription-related machinery to target RNA polymerase to transcription start sites (TSSs), which encourages expression of genes, but can also contribute to DNA accessibility by helping to establish elements of chromatin structure (Ganapathi et al., 2011).

The organization of chromatin, the condensed complex of DNA and protein found in the nuclei of eukaryotic cells (Kornberg, 1974), plays a critical role in determining the relative accessibility of DNA sequences to DNA-interacting proteins (Adams and Workman, 1995). Chromatin is often regarded as a binary control system that moves between "open" and "closed" accessibility states, though chromatin accessibility can be a poor indicator of TF binding (Tsompana and Buck, 2014; Coux et al., 2020). Nucleosomes, the fundamental unit of chromatin, consist of roughly 150bp of DNA wrapped around an octamer of histone proteins (Luger et al., 1997). Histones are deposited by several chromatin-interacting proteins to form nucleosomes (Tachiwana et al., 2021) which are organized into evenly spaced "arrays" throughout the genome (Singh et al., 2021). Organization of nucleosomes is observable from yeast (Lee et al., 2007) through human tissue (Ozsolak et al., 2007; Schones 2008) and often occurs with respect to TF binding sites (Kasinathan et al., 2014), TSSs (Lai et al., 2018), and autonomous replication sequences (Fennessy and Owen-Hughes, 2016). Nucleosomes can interfere with DNA binding by other proteins, which helps define DNA accessibility during transcription (Adams and Workman, 1995; Venkatesh and Workman, 2015).

The mechanisms that position nucleosomes have been studied for several decades, but dissecting the relative contributions of active nucleosome positioning and DNA sequence-based positioning is challenging. There is likely significant interplay between mechanisms, where some elements of nucleosome positions are defined by DNA sequence preference (Thastrom et al., 2004; Kaplan et al., 2008; Krietenstein et al., 2016), TF binding and recruitment of co-factors (Svaren et al., 1994; Ganapathi et al., 2011; Voss and Hager, 2013), and general and site-specific chromatin remodeling

activity (Henikoff, 2016; Singh et al., 2021). Therefore, it is advantageous to examine both DNA sequences and nucleosome localization when considering DBP targeting, as many protein-DNA interactions are influenced by and are capable of directing chromatin organization.

Techniques like ChIP (Johnson et al., 2007) and ORGANIC (Kasinathan et al., 2014) are used to recover regions of DNA associated with a DBP, termed a consensus motif, that serves as a favorable site of interaction between DNA and the DBP. By combining ChIP with exonuclease digestion, a technique called ChIP-exo (Rhee and Pugh, 2011), it is possible to further probe the interaction of a DBP with its consensus motif. ChIP-exo has been used to precisely determine the binding site footprints associated with many of the DBPs found in budding yeast, and also confirmed that factors make direct contact with their consensus sequences when binding to a target region. These results aligned with observations that consensus motifs often predict where a protein will localize in the genome (Zamanighomi et al., 2017; Jolma et al., 2013), but fail to reveal information about the local chromatin environment that enables DBP localization to the target sequence.

Current methods of measuring chromatin accessibility and protein localization simultaneously, including CUT&RUN (Skene and Henikoff, 2017) and CUT&Tag (Kaya-Okur et al., 2019), are attractive because DNA fragments are generated via epitope recognition by a fused antibody, which distributes the protein fusion to target sites throughout the genome. By directing micrococcal nuclease (MNase) digestion or a sequencing adaptor-loaded Tn5 transposase, high-resolution maps of DNA fragments corresponding to a combination of protected DNA sequences can be obtained by next-

generation sequencing (NGS). Fragment size can be used to discern between nucleosome information and small DBP binding to reveal the target site and proximal nucleosome positions (Henikoff et al., 2011; Meers et al., 2019), which has been applied to several experimental systems. Antibody-free systems that capture DNA-protein interactions have also been successful (Policastro and Zentner, 2018), including chromatin endogenous cleavage and sequencing (ChEC-Seq) (Zentner et al., 2015). This system employs a direct fusion of MNase to a protein of interest and maps cleavage sites adjacent to predicted consensus motifs. By relying on the inherent targeting capacity of TFs and transcription machinery, MNase digestion can be performed *in situ* without the need to select fragmentation sites via antibody recognition. The results of ChEC-Seq are somewhat limited in their application, as the analysis of cleavage sites was performed using pre-annotated target sequences from ORGANIC (Kasinathan et al., 2014). However, the results demonstrated that self-targeted nuclease digestion is highly desirable in terms of detecting differences in binding site accessibility and the orientation of DNA interaction within target sequences for yeast TFs.

In yeast, a subset of TFs are deemed general regulatory factors (GRFs). GRF binding in the genome is often correlated with the formation of a nucleosome free region (NFR) in the DNA sequences proximal to the consensus sequence (Ganapathi et al., 2011; Rossi et al., 2018; Badis et al., 2008; van Bakel et al., 2013; Barnes and Korber, 2021). Some GRFs are involved in recruiting ATP-dependent chromatin remodelers to shift or dissociate nucleosomes from DNA, which is thought to help establish and maintain NFRs (Badis et al., 2008; van Bakel et al., 2013; Barnes and Korber, 2021). ARS-binding factor 1 (Abf1) is a well-studied GRF that stimulates specific nucleosome

positions surrounding Abf1-bound DNA (Rhode et al., 1992; Ganapathi et al., 2011). This activity is not anticipated to be through chromatin remodeling activity of Abf1, but rather binding site competition with nucleosomes (Yarragudi et al., 2004) and recruitment of the RSC remodeling complex (Kubik et al., 2018), which forces nucleosomes out of NFRs and into canonical positions (Yen et al., 2012; Ganguli et al., 2014).

We envisioned a system that would allow detection of protein-DNA interactions and nucleosome-derived signal, with no requirement for crosslinking or antibody recognition, by expressing a targetable MNase fusion in the yeast genome. MNase is suitable for this task, as it requires millimolar concentrations of Ca^{2+} to be activated, which is several orders of magnitude higher than the typical Ca^{2+} concentration found in yeast cells (Cui et al., 2009). To target nuclease digestion, we fused MNase to Spycatcher via a flexible linker (Chen et al., 2013) and placed it under an inducible promoter in the genome of *S. cerevisiae*. SpyCatcher is able to join with SpyTag, a 13 amino acid sequence, which is selectively targeted by SpyCatcher to form a covalent isopeptide bond, fusing together any protein attached to either component of the system (Zakeri et al., 2012; Keeble et al., 2017). This technology has been used in a wide range of applications (Hatlem et al., 2019), including to target chromatin remodeling activity to TF-bound genome loci (Donovan et al., 2018). By incorporating SpyTag at endogenous loci of TFs and other DBPs, we can capture high-resolution maps of both protein binding and local nucleosome patterning, achieved by *in situ* DNA digestion by the MNase-SpyCatcher (MNSC) fusion targeted to TF-bound sites. Further, by limiting expression of MNSC to a brief period prior to cell harvest, we observed low background digestion, presumably due to the difference in digestion rates that can be achieved by tethered vs

untethered MNSC. We anticipate that SpyTagged proteins exhibit native behavior, considering the diminished potential for misfolding or reduced binding capability imparted by a small, unstructured tag. Here, we present SpyCatcher Linked Targeting of Chromatin Endogenous Cleavage and deep sequencing (SpLiT-ChEC-Seq), which enables simultaneous mapping of protein binding signal and proximal nucleosome patterning at genomic sequences bound by endogenously expressed factors.

Results

SpLiT-ChEC produces well-defined signal with multiple fragment sources

As a proof of concept, we tagged Abf1 and performed SpLiT-ChEC in time course post-Ca^{2+} addition (Figure 17a). Abf1 is an important regulator of chromatin structure that helps establish NFRs at loci throughout the yeast genome. Importantly, SpyTagged Abf1 is expressed under the control of its endogenous promoter and only becomes conjugated with the MNSC fusion (Figure 17b) after expression is induced by addition of galactose to cultures.

After purifying DNA and enriching samples for <200bp fragments, we performed paired-end sequencing genome-wide to determine the complement of DNA sequences recovered from the experiments. Alignment to the budding yeast genome and subsequent visualization of the sequenced fragment sizes showed a bimodal distribution in our SpLiT-ChEC samples at each timepoint (Figure 17c), suggesting that fragmentation is influenced by protection of DNA associated with proteins and nucleosomes. We used 1x normalized genome coverage from Abf1-tagged and non-tagged ("No-tag" or "NT") samples at each timepoint to calculate the ratio of signal over background that occurs when MNSC is targeted to Abf1-bound loci.

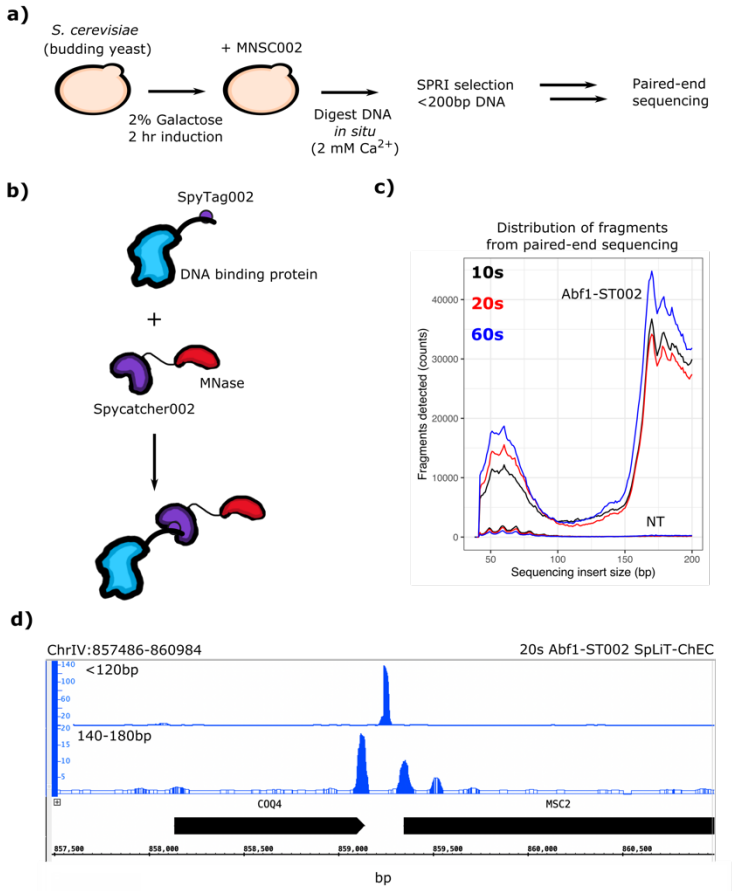


Figure 17. SpLiT-ChEC-seq captures protein binding and nucleosome localization signal from sites of targeted nuclease digestion. a) outline of the basic workflow to express MNase-Spycatcher and perform in situ digestion of DNA, which is purified and subjected to next-generation sequencing. **b)** cartoon representation of the Spycatcher/Spytag targeting system employed to localize MNase to protein-bound loci in the yeast genome. **c)** fragment size distributions observed across three timepoints of both untargeted (No-tag, NT) and Abf1-targeted SpLiT-ChEC-seq. Abf1 was C-terminally tagged with Spytag at the endogenous locus in yeast harboring the GAL-inducible MNase-Spycatcher construct. **d)** Integrated Genome Browser (IGB) screen capture showing the signal over background for small (<120bp) and large (140-180bp) fragment protection observed in Abf1-targeted SpLiT-ChEC-seq after 20 sec. of calcium-activated DNA digestion.

While the entire complement (<200bp) of fragments can be visualized (Figure S1), we found that separately aligning fragments from two size ranges allowed TF-derived small factor protection (SFP) (<120bp) to be viewed separately from the large fragment protection (LFP) (140-180bp), which possibly results from nucleosomes bound

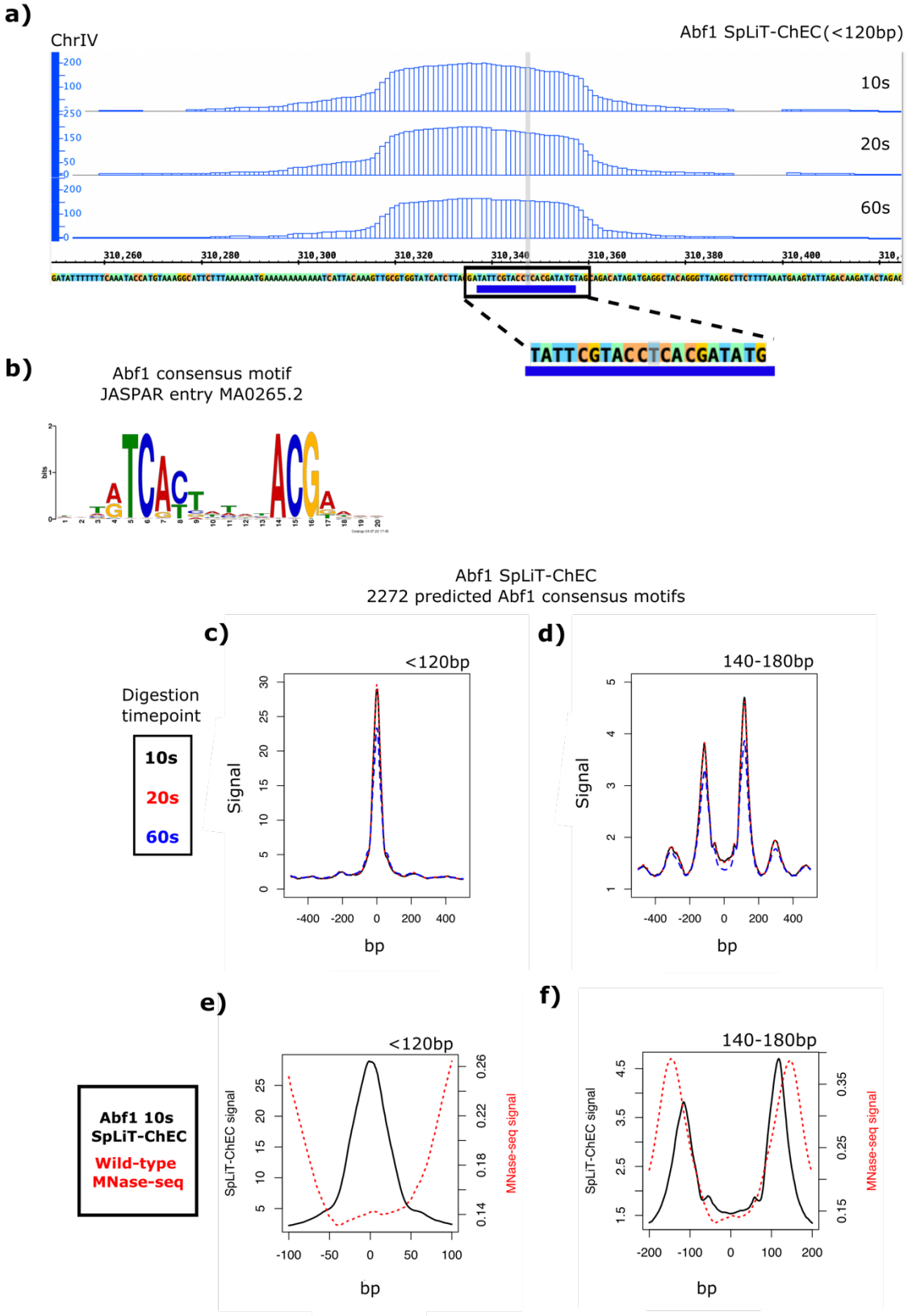
near Abf1-targets (Figure 17d). Both signal types display high-intensity peaks, primarily found in promoter and terminator regions of genes throughout the genome. When comparing the two signal sets, most often SFP signal is found in sequences lacking LFP signal, with LFP signal directly proximal to SFP sites at many targets, as is the case in Figure 17d.

Canonical Abf1 binding and chromatin organization is captured in SpLiT-ChEC

Since Abf1 localization is associated with the consensus motif 5'-WHWTCGTRTAWAGTGAYAND-3' as determined by ChIP-exo (Rhee and Pugh, 2011), we examined SpLiT-ChEC datasets and found many examples of Abf1 SFP directly situated over DNA sequences matching this motif (Figure 18a). In order to visualize SpLiT-ChEC signal at predicted Abf1 target regions in the yeast genome, we used FIMO (Grant et al., 2011) to locate Abf1 consensus sequences based on the JASPAR (Castro-Mondragon et al., 2022) entry corresponding to the Abf1 ChIP-exo footprint (MA0265.2, Figure 18b). Plotting the average SpLiT-ChEC signal at all 2272 predicted Abf1 consensus sequences shows Abf1-targeted digestion occurring at these sites, with high intensity SFP signal (Figure 18c) directly over the predicted Abf1 consensus sequences. The average Abf1 SpLiT-ChEC LFP at consensus sequences (Figure 18d) shows that LFP signal is not typically found in the regions marked by SFP over consensus sequences. Overlaid SpLiT-ChEC signal from three sequenced timepoints (10s, 20s, 60s) shows that the shape and location of averaged SFP and LFP signal is similar across all timepoints. Time-dependent changes in signal intensity are present in both signal sets, most obviously seen as decreases in height of SFP and LFP dominant peaks in 60s timepoint traces (Figure 18c,d). This observation is consistent with data

from ChEC-seq, which can discern between high and low scoring pre-annotated binding sites based on time-dependent variability in cleavage intensity (Zentner et al., 2015). Overlaying SpLiT-ChEC SFP and LFP with nucleosome positions derived from MNase-seq (McKnight et al., 2021, GEO: GSE141676), which is used to map nucleosome positions in genomic DNA (Figure S2a), shows a striking similarity between the NFRs surrounding Abf1 consensus sequences. Abf1 SFP is present within the bounds of the NFR defined by MNase-seq data (Figure 18e). Nucleosome positions are not necessarily recapitulated exactly by SpLiT-ChEC LFP, which is clear when examining the nucleosome-derived fragment centers in MNase-Seq relative to LFP high-intensity regions (Figure 18f). Nevertheless, LFP signal provides information on nucleosome protection, similar to that derived from MNase-Seq, which could be used to examine the relationship between protein binding and nucleosome protection at targets throughout the yeast genome. In total, the agreement between MNase-seq data and the signal derived from Abf1-targeted SpLiT-ChEC suggests that we are capturing signal at true NFRs associated with Abf1 localization.

Figure 18. (next page) Abf1 SpLiT-ChEC signal is present at Abf1 consensus sequences and recapitulates chromatin organization elements observed in nucleosome positioning data. a) IGB screen capture of Abf1 SpLiT-ChEC 10s, 20s, and 60s digestion timepoints showing small fragment signal located over an instance of the 20bp Abf1 consensus sequence (blue underline). **b)** JASPAR entry MA0265.2, the 20 bp Abf1 consensus sequence derived from ChIP-exo. **c)** Overlay of the average Abf1 SpLiT-ChEC small fragment signal from 10s, 20s, and 60s timepoints found at all 2272 Abf1 ChIP-exo consensus sequences in the yeast genome predicted by FIMO based on the motif found in (b). **d)** Overlay of the average Abf1 SpLiT-ChEC large fragment signal from 10s, 20s, and 60s timepoints found at all 2272 predicted Abf1 ChIP-exo consensus sequences in the yeast genome. **e)** Average traces of Abf1 10s SpLiT-ChEC small fragment signal and MNase-seq signal found at all 2272 predicted Abf1 ChIP-exo consensus sequences in the yeast genome. **f)** Average traces of Abf1 10s SpLiT-ChEC large fragment signal and MNase-seq signal found at all 2272 predicted Abf1 ChIP-exo consensus sequences in the yeast genome.


a)
ChrIV
Abf1 SpLiT-ChEC(<120bp)
10s
20s
60s
310,260
310,280
310,300
310,320
310,340
310,360
310,380
310,400
TATTCGTACCTCACGATATG
b)
Abf1 consensus motif
JASPAR entry MA0265.2
bits
Abf1 SpLiT-ChEC
2272 predicted Abf1 consensus motifs
c)
<120bp
d)
140-180bp
Digestion timepoint
10s
20s
60s
Signal
bp
e)
<120bp
f)
140-180bp
Abf1 10s SpLiT-ChEC
Wild-type MNase-seq
SpLiT-ChEC signal
MNase-seq signal
bp

Abf1 SFP can be used to find peaks with MACS3

Given the intensity of Abf1 SFP signal at known consensus sequences, we imagined our data could be used to identify bound targets throughout the yeast genome using a peak calling algorithm. Analysis of aligned reads with MACS3 (https://macs3-project.github.io/MACS/) generated three sets of SFP-derived peaks, one from each timepoint in Abf1-targeted SpLiT-ChEC. Variable numbers of peaks were called from each timepoint, with the highest number of peaks coming from the 20s signal (Figure S2b).

Rather than handling each timepoint peak set independently, we designed a computational process to identify target regions in Abf1 SFP signal according to peak groupings from across timepoints. We took the central two basepairs (dyads) from all peaks identified by MACS3 in each timepoint sample (n=4322); merging this set of dyads without extension resulted in 3476 independent entries. Next, the bounds of each target site were symmetrically extended by increasing values and any regions overlapping as a result of the expanded region size were merged. By visualizing the region counts, we anticipated that an optimal merge distance could be identified as the extension value where the total peak counts settled into a local minimum. A bar chart of the number of regions after each cycle showed substantial decreases at low merge distances, demonstrating that peaks were detected in similar regions across all timepoints (Figure S2c). When we applied this methodology to Abf1-derived SFP peaks, our analysis suggested Abf1-targeted signal occurs within 90bp regions (n = 1807), which is consistent with the approximate size of the NFR present in MNase-seq and SpLiT-ChEC LFP (Figure 18f). Plotting SpLiT-ChEC signal using the calculated target sequences

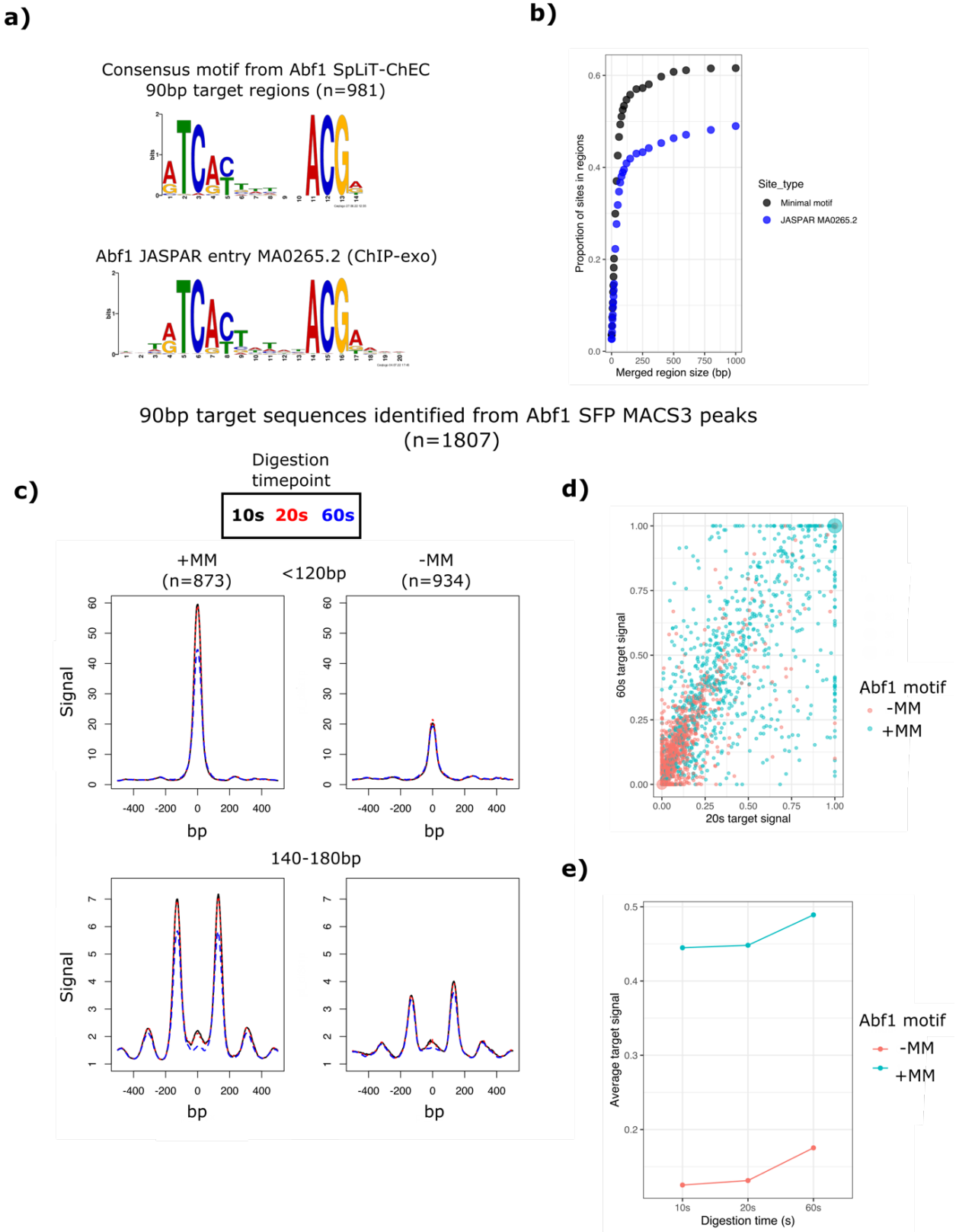
displays center-aligned SFP (Figure S2d) and LFP (Figure S2e), similar to that observed for signal at consensus sequences, suggesting our merge process identified the centers of Abf1-targeted regions from across all timepoints.

A minimal Abf1 motif marks canonical targets in SpLiT-ChEC

With Abf1 target regions in hand, we were curious if the Abf1 consensus motif was enriched in our recovered DNA sequences. Motif enrichment analysis showed that, of the 1627 sequences randomly selected from inputs for SEA (Simple enrichment analysis, Bailey and Grant, 2021), 46% (Q value = 5.69E-224) contained a motif matching JASPAR entry MA0265.2. We also performed de novo motif discovery via XSTREME (Grant and Bailey, 2021) on 90bp sequences with Abf1 signal, which revealed a "minimal" version of the Abf1 motif (MM, 14bp vs 20bp from ChIP-exo), found by SEA in 53% (Q value = 1.17E-250) of the sites (Figure 19a). Next, we calculated the number of Abf1 consensus motifs overlapped by Abf1 targets based on SpLiT-ChEC peaks, as determined by the extend-and-merge process. This analysis suggested Abf1-targeted SpLiT-ChEC digestion is found within 90bp regions. Plotting the fraction of FIMO-predicted motifs found in target regions from each merged peak set showed that the SpLiT-ChEC derived minimal motifs more frequently found in target centers (Figure 19b). Based on this result, we elected to use the FIMO-predicted MM sites as a filter to separate the 90bp target regions into +MM and -MM sets. Consistent with previous observations of lower perceivable levels of Abf1 binding at non-canonical sites (Kasinathan et al., 2014), -MM target sequences showed overall lower average SFP and LFP (Figure 19c). Since the LFP shape does not appear to be disrupted at -MM sites, we presume that the lower LFP intensity observed at -MM sites is related to lower levels

of Abf1-targeted digestion occurring in these regions, not less defined nucleosome patterning in the region. Based on the observation that SpLiT-ChEC signal varies across time, we were curious if examining the per-basepair signal intensity within 90bp merge regions could be used to visualize the progress of MNase digestion as it occurs post-Ca^{2+} addition. Initially, we plotted two timepoints for each target region by treating the scaled intensity values as Cartesian coordinates (Figure 19d). Coloring the points based on the presence of Abf1 minimal motif (+/- MM) shows that there is separation between the two groups, with -MM generally showing lower intensity at both the 20s and 60s timepoints. Plotting average per-basepair intensities for each timepoint, separated by +/-MM, allowed us to view the trajectory of binding intensity over time (Figure 19e).

Figure 19. (next page) Abf1 SpLiT-ChEC allows consensus motifs to be identified and shows higher signal intensity within target regions containing consensus motifs. **a)** Comparison of motifs identified for Abf1 from SpLiT-ChEC (top, 14bp) and ChIP-exo (20 bp, bottom). **b)** Plot showing the proportions of 14bp SpLiT-ChEC minimal motifs (black) and 20bp ChIP-exo motifs (blue) found in regions identified by merging MACS3 peaks derived from Abf1 SpLiT-ChEC 10s, 20s, and 60s signal. **c)** Overlay of average Abf1 SpLiT-ChEC signal centered on 90bp target sequences based on MACS3 peaks from 10s, 20s, and 60s timepoints. Signal is separated into sites containing a predicted 14bp minimal motif (+MM) and sites without a motif (-MM). **d)** Per basepair Abf1 small fragment signal intensity found in 90bp target regions for 20s and 60s digestion timepoints, plotted as Cartesian coordinates and colored by sites with a minimal motif (+MM, blue) and without a minimal motif (-MM, orange). Scaling was achieved using the 5^{th} and 95^{th} percentile values for each timepoint as minimum and maximum values, respectively. Any values < 0 after scaling were assigned to 0 and values >1 were assigned to 1. **e)** Abf1 small fragment signal intensity at all 90bp target sequences from all timepoints calculated as in (d), separated into +MM and -MM groups and averaged for each group and each timepoint.


a)
Consensus motif from Abf1 SpLiT-ChEC
90bp target regions (n=981)
Abf1 JASPAR entry MA0265.2 (ChIP-exo)
b)
Proportion of sites in regions
Merged region size (bp)
Site_type
Minimal motif
JASPAR MA0265.2
90bp target sequences identified from Abf1 SFP MACS3 peaks
(n=1807)
c)
Digestion
timepoint
10s 20s 60s
+MM
(n=873)
<120bp
-MM
(n=934)
140-180bp
Signal
bp
d)
60s target signal
20s target signal
Abf1 motif
-MM
+MM
e)
Average target signal
Digestion time (s)
10s
20s
60s
Abf1 motif
-MM
+MM

Although the trends in SFP intensity appear similar, the overall intensity of signal without a minimal motif appears lower at all timepoints, consistent with observations made from plotting the raw SFP signal. While the height of the peaks in SFP appear to decrease at 60s, the intensity per-basepair increases, which we anticipate is due to signal broadening as digestion progresses and DNA fragments are shortened (Henikoff et al., 2011). In summary, we observe a difference in the average Abf1 SpLiT-ChEC SFP signal intensity in target regions separated by the presence or absence of a minimal motif at each timepoint, where Abf1 targets without a minimal motif show lower average signal compared to sites that contain a minimal motif.

SpLiT-ChEC signal can be clustered based on intensity and shape

Next, we wanted to see if unsupervised clustering of signal derived from SpLiT-ChEC would allow subsets of Abf1-targeted sequences to be identified. We applied k-means (k = 3) clustering to raw LFP signal from the 60s timepoint at 90bp target regions (Figure S3a). The signal patterns observed in each LFP cluster were surprisingly distinct, with obvious signal bias towards either side of the Abf1 target center (clusters 1 and 2), and the 3rd cluster showing targets with LFP signal in the center of sites. Focusing on the central 200bp of the 90bp target regions (Figure S3b) shows the asymmetry in LFP protection, with the most substantial LFP at the edges of the central 90bp in each target. When we plotted the SFP average signal trajectories for each LFP cluster (Figure S3c), we observed similar signal evolution patterns for clusters 1 and 2, which remained mostly constant across timepoints. Interestingly, the cluster 3 average trajectory shows lower intensities at each timepoint when compared to cluster 1 and 2 patterns, with a consistent upward trend in signal evolution.

When k-means (k = 3) clustering was applied to SFP per-basepair intensity measured across the three timepoints, plotting the values for each site at 20s and 60s post-Ca^{2+} addition shows distinct groups of points (Figure S3d). The values of average intensity per bases from each cluster suggest that k-means clustering separates sites primarily by signal intensity at each timepoint (Figure S3e). Plotting average LFP signal according to SFP cluster annotations (Figure S3f) revealed signal shapes that are distinct from those found in LFP-based clusters (Figure S3a). Unlike the LFP clusters, signal intensity on either side of the center within each cluster was roughly equal with no clear preference for sites with or without LFP signal in the central NFR between groups. Instead, LFP patterns associated with SFP clusters differ primarily by intensity, suggesting that the per-base SFP calculations are limited to the relative amount of Abf1-directed digestion occurring within the 90bp targets.

Abf1 SpLiT-ChEC signal near TSSs is directional and highly variable

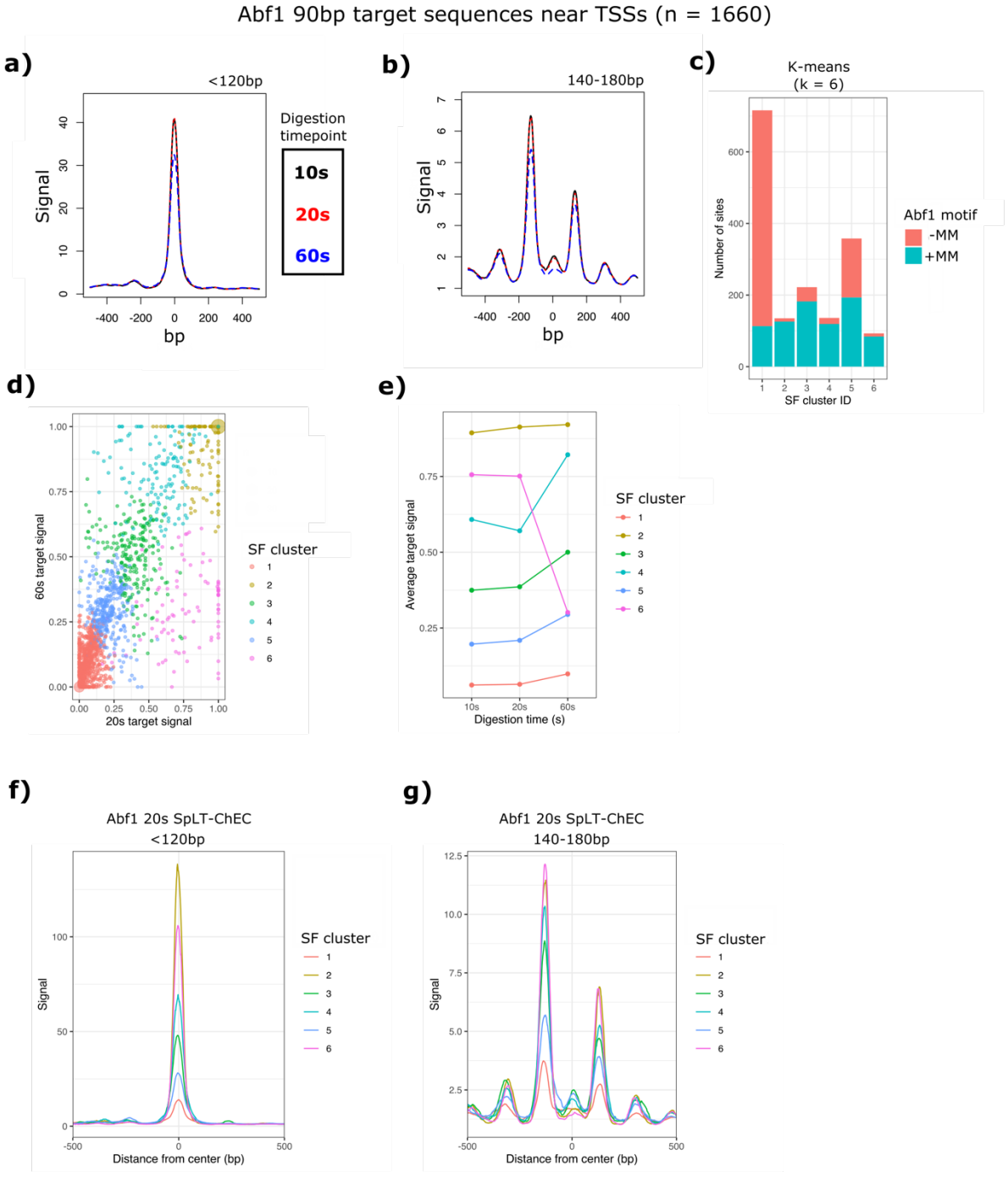
Given the directionality we observed in Abf1 LFP signal clusters, we applied strand information derived from TSSs near 90bp targets to see whether signal is orientated with respect to the direction of transcription. Using T-Gene (O'Connor et al., 2020), we located TSSs in the budding yeast genome nearest to each 90bp target, limited to within +/- 1000bp (n = 1660). After applying the strand designation for the identified TSSs, we plotted the average SFP signal at target centers across all timepoints (Figure 20a). The average traces of Abf1 SFP near TSSs show the expected high intensity central peak, but also a more prominent low-intensity protection peak roughly 200-250bp upstream of Abf1 target centers, which was present on either side of the central peak before the addition of strand information. When we examined LFP signal at the TSS-

proximal targets (Figure 20b), we noticed a prominent bias in peak intensity opposite from the direction of transcription at all timepoints. The relative peak intensity is in opposite orientation compared to +1/-1 nucleosome positions observed by MNase-Seq at NFRs near TSSs. While we cannot rule this out as a physical limitation of the SpLiT-ChEC system, this could also represent the presence of co-bound factors at Abf1 targets that help define precise nucleosome positions (Yen et al., 2012; Ganguli et al., 2014).

We were curious if Abf1 SFP per-base intensity across timepoints for TSS-proximal targets would reveal subgroups based on optimized k-means clustering. We applied the WSS (within sum of squares) calculation to create an elbow plot (Figure S4) and identified k = 6 as an appropriate number of k-means clusters. When examining the number of sites in each cluster, we were surprised by the lack of separation between sites with or without an example of the Abf1 minimal motif (Figure 20c). While cluster 1 is seemingly dominated by -MM targets, no group appears to be completely composed of -MM or +MM sites. Plotting the per-basepair SFP signal at 20s and 60s for each target site shows visually distinct groups of points throughout the coordinate plane when colored by cluster (Figure 20d). Using the SF cluster annotations, we plotted the average per base intensities at each timepoint (Figure 20e), revealing that the clusters are mostly separated by signal intensity, which was expected based on our preliminary cluster analysis (Figure S3). However, some degree of the change in signal between timepoints appears to be captured within the k=6 clusters, particularly when examining clusters 4 and 6, which are graphically similar in intensity and trend for 10s and 20s, but display opposite trends when examining the 60s average per base signal.

To examine the average signal identified within the groups of Abf1 targets, we plotted SFP (Figure 20f) and LFP (Figure 20g) according to SFP cluster labels, using only the average signal derived from the 20s Abf1 SpLiT-ChEC timepoint. The SFP signal traces differ based on central peak intensity, with some small differences in proximal signal, though no clear differences are readily identifiable. No obvious LFP patterning trends are evident in the average signal within the SF clusters, though each trace differs by overall intensity, matching expectations set by our pre-TSS cluster analysis. We anticipate this analysis could be more insightful for proteins directly involved in nucleosome positioning, which may show distinct LFP signal patterns associated with targeted chromatin remodeling.

Figure 20. (next page) Abf1 SpLiT-ChEC small fragment signal near transcription start sites is directional with respect to gene orientation and can be used to identify similarly protected target sequences. a) Overlay plot of 10s, 20s, and 60s Abf1 SpLiT-ChEC small fragment signal centered on 90bp targets within +/- 1000 bp of a TSS. TSS strand designation was applied to plotting calculations. **b)** equivalent to (a), except using long fragment signal. **c)** Counts of the number of Abf1 90bp target sequences in each k-means cluster (k = 6) colored by sites with and without a minimal motif. **d)** Per basepair Abf1 small fragment signal intensity found in 90bp target regions for 20s and 60s digestion timepoints, plotted as Cartesian coordinates and colored by k-means SF cluster designation. **e)** Abf1 small fragment signal intensity at TSS-proximal 90bp target sequences from all timepoints calculated as in (d), separated into SF cluster groups and averaged for each group at each timepoint. **f)** Average traces of 20s Abf1 SpLiT-ChEC small fragment signal separated by k-means cluster annotations. **g)** equivalent to f), except applied to 20s Abf1 long fragment signal.


Abf1 90bp target sequences near TSSs (n = 1660)
a)
<120bp
Signal
bp
Digestion timepoint
10s
20s
60s
b)
140-180bp
Signal
bp
c)
K-means (k = 6)
Number of sites
SF cluster ID
Abf1 motif
-MM
+MM
d)
60s target signal
20s target signal
SF cluster
e)
Average target signal
Digestion time (s)
SF cluster
f)
Abf1 20s SpLT-ChEC
<120bp
Signal
Distance from center (bp)
SF cluster
g)
Abf1 20s SpLT-ChEC
140-180bp
Signal
Distance from center (bp)
SF cluster

Abf1 SpLiT-ChEC LFP near TSSs reveals distinct clusters of chromatin organization

We wanted to visualize the differences in LFP patterning among 90bp targets near TSSs, considering the known relationship between Abf1 localization, nucleosome positioning, and NFR formation at Abf1-targeted genome loci (Rhode et al., 1992, Ganapathi et al., 2011). We selected k = 8 as the optimal number of clusters, based on an elbow plot using the WSS method (Figure S5). Applying k-means clustering to the signal from a single LFP timepoint, 60s, shows clusters that contain various proportions of sites both with and without an Abf1 minimal motif (Figure 21a); as with the SFP signal (Figure 20c), +MM and -MM sites do not separate into distinct clusters based on the presence or absence of a minimal motif. When we plotted the average 60s SpLiT-ChEC LFP signal based on long fragment (LF) cluster annotations, we observed differences in signal intensity, directionality, and the locations of the most intense LFP signal (Figure 21b,c). Similar to our preliminary clustering analysis, we observe a subset of sites with LFP signal in the predicted NFR. LF clusters 4 and 7 both display this feature, with cluster 7 showing a clear LFP peak in the center of the average signal traces (Figure 21c). The average LFP signal associated with cluster 6 shows a narrowing of the NFR with respect to other signal traces, an interesting feature which is often associated with decreased transcriptional activity for nearby target genes (Weiner et al., 2010). We were curious about the relationship, if any, between the differences we observed in average LFP signal among LF clusters and the corresponding SFP signal qualities. We were surprised to find that SFP signal patterns in these clusters vary substantially in signal intensity and shape, with various degrees of SFP outside of the expected 90bp central

region (Figure 21d). While every average SFP signal trace appears to show some amount of protection proximal to the central high-intensity signal, cluster 3 provides the most striking example. Cluster 3 displays a defined peak 200-250bp upstream of the central SFP and contains a high intensity central SFP peak. The upstream signal feature originally became more pronounced when strand information was applied to TSS-proximal 90bp target sequences (Figure 20a). Interestingly, when examining the average LFP signal for cluster 3, we observed a high level of signal directly between the region defined by the two SFP average signal features. Additionally, we observed cluster 7 as having the most visually distinct average SFP signal (Figure 21e), which shows a broad shoulder in place of the high intensity, tightly defined signal we observe in other clusters.

When compared to the average LFP trace for cluster 7 (Figure 21c), this suggests that centralized LFP signal is not an effective predictor of Abf1 localization, but rather that Abf1 is capable of targeting digestion to these sites regardless of LFP within the expected target regions. Further, when we examined the average per-base SFP intensity at each timepoint for the LF-derived clusters (Figure 21f), we found several interesting patterns, most notably in clusters 5 and 2. Cluster 5 displays the lowest overall intensities among this set of clusters, which we anticipate could be related to the relatively high number of target sequences in this group that lack an Abf1 consensus motif (Figure 21a). Cluster 2 essentially displays the opposite effect – the relative number of -MM sites is lower for this group of Abf1 targets when compared to cluster 5 and average per base SFP intensity appears to be higher at all timepoints.

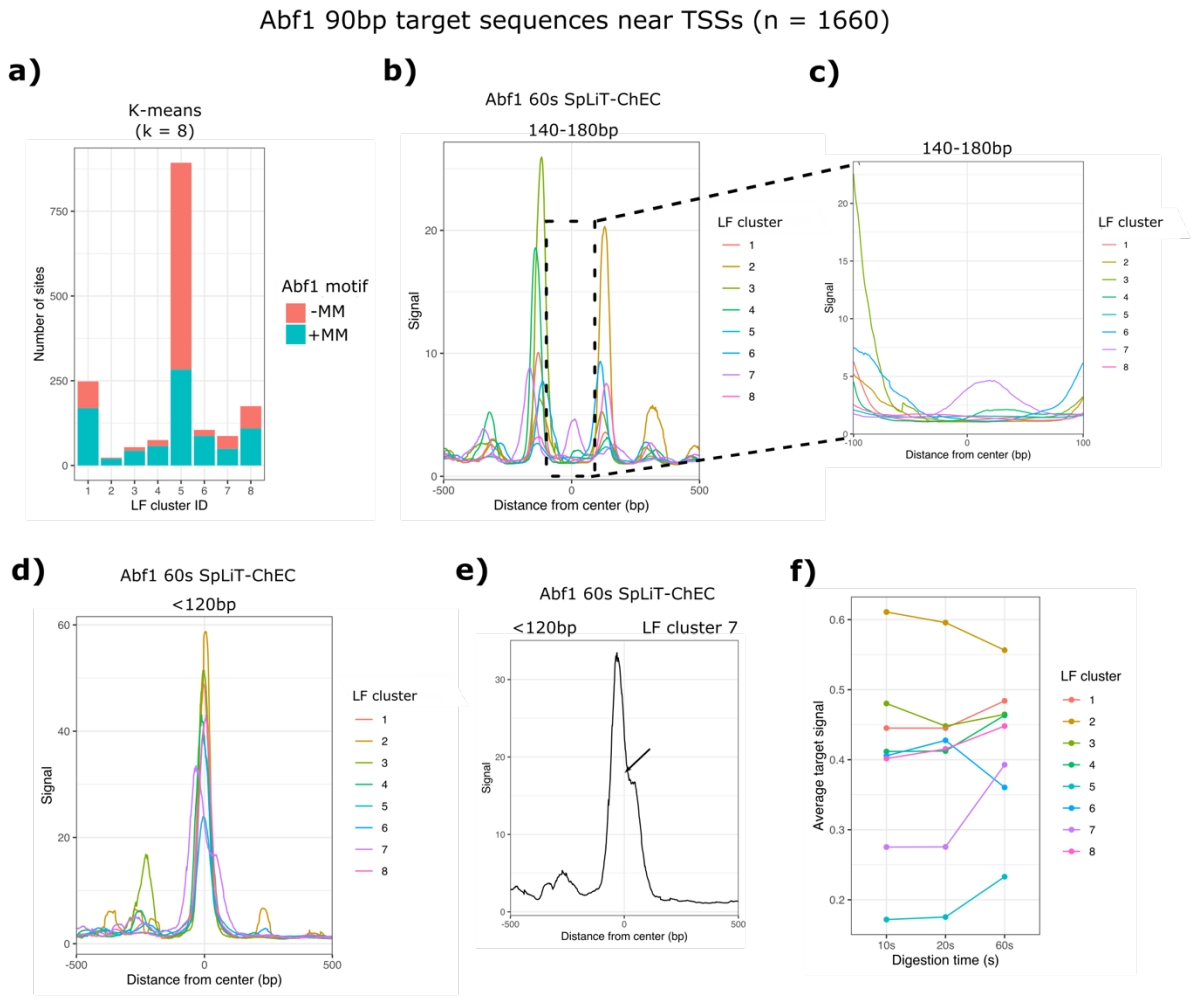


Figure 21. Abf1 SpLiT-ChEC large fragment signal near transcription start sites can be used to identify chromatin organization subgroups with distinct protection characteristics. a) Counts of the number of Abf1 90bp target sequences in each k-means cluster (k = 8) colored by sites with and without a minimal motif. **b)** Average traces of 60s Abf1 SpLiT-ChEC large fragment signal separated by k-means cluster annotations. **c)** Close up view of the signal within the central 200bp of traces found in (b). **d)** equivalent to b), except plotting 60s Abf1 SpLiT-ChEC small fragment signal. **e)** trace showing the average small fragment signal of target sites found in LF cluster 7, which contains a centralized region of long fragment protection. **f)** Abf1 small fragment signal intensity at TSS-proximal 90bp target sequences from all timepoints, separated into LF cluster groups and averaged for each group at each timepoint.

Discussion

While many methods exist that allow protein-DNA interactions to be detected, few allow transcription factor binding sites and chromatin organizational elements to be measured simultaneously. Generally, multiple experiments must be performed if information on both protein-DNA interactions and nucleosome positions is required to make conclusions about the behavior of proteins related to chromatin organization. Further, current methods for producing genomic DNA fragments from multiple sources are either non-targeted (ATAC-Seq, MNase-Seq) or require preparation of functionalized antibodies to target enzymatic digestion (CUT&RUN) or transposase activity (CUT&Tag) to loci of interest. SpLiT-ChEC provides an alternative to these systems, as it relies on the inherent capacity for SpyCatcher and SpyTag to self-associate in a biological environment. By SpyTagging proteins in the yeast genome, expression is controlled by an endogenous promoter, which avoids the issue of mismatched expression associated with the use of a plasmid and/or exogenous promoter. Since the SpyTag is small (13 amino acids), it is straightforward and inexpensive to add to an endogenous protein in a variety of organisms and cell types.

Integrating MNSC into the yeast genome with an inducible promoter provides two benefits: first, a common base strain can be used as a platform for SpyTagging proteins of interest and second, inducing MNSC expression for only a short period before harvesting cultures allows the SpyTagged target protein to be expressed in its near-native state and bind to chromatin without interference from a large fusion protein. After expressing the MNSC construct, MNase is covalently tethered to the SpyTagged target and DNA digestion is activated by adding high concentrations of $Ca^{2+}$to permeabilized

yeast cells. This provides better control of MNase activity, which reduces the level digestion by free MNSC and provides high signal-over-background in our calculated coverage. By separating the sequenced DNA fragments into predefined size ranges, we assess the levels of protection from DNA digestion corresponding to DNA binding proteins (<120bp fragments) and nucleosomes (140-180bp) from a single paired-end sequencing run. Our results suggest that SpLiT-ChEC can serve as an effective tool for monitoring protein localization and proximal nucleosome patterning at DNA sequences that are bound by SpyTagged proteins.

Abf1 presents an ideal model system to evaluate SpLiT-ChEC because of its known roles in chromatin organization, its activity as a GRF, and the well-defined consensus motif associated with its localization (Rhode et al., 1992; Yarragudi et al., 2004; Ganapathi et al., 2011; Rhee and Pugh, 2011). NFRs are associated with GRF binding and are frequently found surrounding consensus motifs (Yarragudi et al., 2004), which was readily observed in the LFP signal we detected with SpLiT-ChEC. When compared to SFP signal, LFP signal is generally found outside of regions associated with protein binding, like those marked by Abf1 consensus sequences. This aligns with the observation that Abf1 can nucleate NFRs by competing with nucleosomes to bind DNA, similar to the activity observed for pioneering factors in higher eukaryotes (Rhode et al., 1992; Balsalobre and Drouin, 2022). While Abf1 is not capable of repositioning nucleosomes, GRF-mediated targeting of chromatin remodeling proteins, like RSC, is a known mechanism that allows nucleosomes to be evicted from NFRs throughout the genome (Kubik et al., 2018; Yen et al., 2012; Ganguli et al., 2014). Consistent with this activity is the appearance of well-defined LFP signal in many of the regions directly

proximal to the Abf1 target sites, which we defined using peak centers from multiple timepoints of SpLiT-ChEC SFP signal. We predict that analyzing clusters of targets defined by SpLiT-ChEC could identify enriched sequence repeats, subsets of sequence motifs, or protein co-localization, any of which may correspond to biologically relevant mechanisms that define precise, site-specific nucleosome positions or enable association of factors with nucleosome-bound DNA. Based on our results, we anticipate SpLiT-ChEC will be a valuable tool for studying proteins that are directly involved in dynamic processes such as chromatin remodeling and transcription. We have also collected SpLiT-ChEC data for Reb1, Med14, Ume6, and Gal4 (unpublished), demonstrating that this technique can be applied to a wide variety of chromatin-associated factors. Full analysis of this data was beyond the scope of this manuscript, but preliminary results suggest that SpLiT-ChEC is a flexible tool for detecting DNA binding and nucleosome patterning in each of these contexts.

When separating the Abf1 SpLiT-ChEC signal by the presence of a minimal consensus motif, we were surprised to see the relative amount of LFP definition was not obviously perturbed in sites lacking a predicted MM. While we cannot rule out that Abf1 is able to stimulate NFR formation at these sites, though with seemingly lower binding intensity, we imagine these interactions represent either co-targeting or binding to NFRs established by other GRFs found in yeast (Yarragudi et al., 2004; Kasinathan et al., 2014; Barnes and Korber, 2021). We noted that, of the 1807 90bp sequences identified as Abf1 targets in this study, 296 also contain a consensus motif (JASPAR entry MA0363.2) associated with binding of the well-studied GRF Reb1 (Ju et al., 1990), based on sites located in the yeast genome with FIMO pattern matching. However, after separating

Abf1 targets into -MM and +MM groups, we found that 241 of those 296 Abf1 target sites do not contain an Abf1 consensus sequence. While a more thorough analysis is warranted to substantiate our hypothesis, we believe this highlights a valuable insight from SpLiT-ChEC data: Abf1 binding throughout the genome can occur independently of its consensus motif, but is not necessarily non-specific, as it appears to be associated with gaps in LFP signal, and, theoretically, specific chromatin organizational states that allow access to a subset of loci.

Since SpLiT-ChEC sequencing data is relatively complex to process and interpret, we invested substantial effort into designing computational pipelines and tools that simplify the analysis of our datasets, which is freely available on GitHub (Bankso/SCAR and Bankso/SEAPE). For example, the process we designed to merge peaks into high-confidence regions of Abf1 binding could be applied to SpLiT-ChEC data recovered from other factors, regardless of their association with NFRs. In the future, we predict that timepoints collected earlier after that addition of Ca^{2+} may be better suited to this task, as we observed a time-dependent shift in the population of large fragments, associated with extended MNase digestion, which could overrepresent SFP signal found in the experiments. As more SpLiT-ChEC data is generated, we hope to refine the computational processes supporting this analysis, with the intention to make the technique generalizable and easily accessible for studying chromatin organization.

Methods

Strain construction

Stock strains stored at -80°C were streaked onto selective media. From overnight starter cultures, yeast were diluted and grown at 30°C in yeast peptone media containing 2% glucose (YPD), then harvested by centrifugation at 3428 RCF for 5 minutes. Pellets were resuspended in sterile water and aliquoted to sterile tubes, then washed twice with 1 mL of sterile water. Washed pellets were resuspended in transformation buffer (240 uL PEG 50%, 36 uL LiOAc, 50 uL salmon sperm DNA) with a DNA cassette to be integrated into the genome via homologous recombination. Samples were placed at 42°C for a minimum of 1 hr before plating on YPD. Replica plates of transformants were made onto selective media after sufficient growth.

Plasmid cloning

Gibson cloning (Gibson et al., 2009) was used to subclone SpyCatcher002 from Addgene 102827. Sticky-end PCR (Walker et al., 2008) was performed as previously described to insert MNase-GGSx5-SpyCatcher002 into an HO-pGAL plasmid derived from Addgene 51664. The SpyTag002 plasmid was made by site-directed mutagenesis of a SpyTag001 plasmid, which was constructed via Gibson cloning using gBlocks from IDT. A list of plasmids can be found in Appendix B.

Yeast strain verification

Selection plates were grown at 30°C and screened for colonies containing integrated DNA at the required locus via colony PCR. With a toothpick, a small sample

was taken from each candidate colony on the selective plate and stirred into 100 uL of lysis buffer, then heated at 65°C for 5 minutes. Next, 300 uL of EtOH were added and the samples were centrifuged for 5 minutes at 21130 RCF. After removing the supernatant, the pellets were washed with 300 uL 70% EtOH and allowed to dry at room temperature. Dried pellets were resuspended in 100 uL sterile water and centrifuged for 5 minutes at 21130 RCF, after which 50 uL of supernatant containing gDNA was transferred to new tubes. Genomic DNA samples were analyzed by PCR with sample-specific primers designed to amplify target regions of the genome where the integration was expected to occur. PCR reactions were analyzed by agarose gel electrophoresis. Positive transformants were streaked onto selection media and verified on YPG before use in experiments. A list of yeast strains can be found in Appendix A.

Growth, harvest, and preparation of yeast for SpLiT-ChEC

Adapted from Zentner et al., 2015. From overnight starter cultures, yeast were diluted, then grown in yeast peptone media containing 2% raffinose (YPR) at 30°C before adding 20% galactose to 2% final concentration. Cultures were returned to 30°C for 2 hours to express MNase-SpyCatcher. Next, cultures were centrifuged at 1500 RCF for 1 minute at room temperature. Pellets were resuspended in 1 mL buffer A (15mM Tris, pH 7.5, 80 mM KCl, 0.1 mM EGTA, 0.5 mM spermidine, 1x Proteoloc protease inhibitors, 1x PMSF/leupeptin/benzamidine) and transferred to 1.5 mL tubes. Samples were pelleted at 1500 RCF for 30 seconds and the supernatant was removed. This was repeated twice more with 1 mL buffer A. Washed pellets were resuspended in 570 uL of buffer A, then combined with 30 uL of 2% digitonin. Samples were briefly mixed and

allowed to incubate at 30°C for 5 minutes to permeabilize cells. Permeabilized yeast were mixed again, then a 100 uL pre-digestion sample was removed to one of 6 tubes containing 10 uL 10% SDS and 90 uL of 2x ChEC stop buffer (400 mM NaCl, 20 mM EDTA, 4 mM EGTA).

SpLiT-ChEC targeted DNA digestion

Adapted from Zentner et al., 2015. To perform DNA digestion, 1 uL of 1 M $CaCl_2$ (2 mM final) was added to the remaining permeabilized yeast, followed by briefly vortexing and placing the sample at 30°C, at which point a timer was started. Samples were collected at five time points post-Ca^{2+} addition by removing 100 uL of sample to pre-filled and labeled tubes with ChEC stop buffer and 10% SDS. All tubes were mixed immediately after adding a sample aliquot to ensure MNase digestion was deactivated. After collecting the final timepoint sample, 4 uL of 20 mg/mL proteinase K was added to each tube, followed by incubation at 55°C for a minimum of 30 minutes. At this point, samples were stored at -20°C or directly purified.

DNA purification and size selection

Adapted from Zentner et al., 2015. SpLiT-ChEC DNA fragments were isolated from proteinase K-treated samples by first adding 200 uL of phenol-chloroform-isoamyl alcohol (25:24:1) and vortexing thoroughly. Samples were centrifuged at 21130 RCF for 5 minutes, after which ~180 uL of the aqueous (top) layer was removed to new 1.5 mL tubes. To precipitate DNA, 500 uL EtOH, 15 uL 3M NaOAc pH 5.3, and 1.5 uL (30 ug) of 10 mg/mL glycogen were added to each sample and placed at -80°C for at least 20

minutes. DNA pellets were resuspended in RNase solution (1 uL RNase A 20 mg/mL, 29 uL 1x Cutsmart buffer [New England Biolabs] or low-TE buffer) and digested at 37°C for at least 30 minutes. 5 uL of each timepoint sample was analyzed by agarose gel electrophoresis to confirm time-dependent DNA streaking. Three samples from across the timepoint range were used in size-selection with a 3:1 ratio of SPRI beads (AmpureXP, Beckman Coulter). Supernatants from size selection were retained for each sample, primarily containing fragments <200 bp in size. The size-selected samples were PCI extracted and precipitated as above. The final DNA pellet was solubilized in 12 uL low-TE buffer and the concentration determined by Pico Green analysis (Thermo Fisher Scientific).

SpLiT-ChEC NGS library preparation

For library preparation (via Ovation Ultralow V2 kit, NuGen), at most 10 ng of DNA was used from each sample as input and all steps were performed per the manufacturer's protocol, except for the noted exceptions here. Library SPRI purification steps were performed with 1.8x vol of beads to increase the small fragment content of the final libraries and 15 uL of low-TE buffer were used to elute final libraries, collecting 12 uL as a final sample for sequencing. Amplification cycles were scaled as necessary depending on DNA available, with 13 cycles used for 10ng inputs. Libraries were sequenced in paired-end mode for 37 cycles on an Illumina HiSeq 5000 system in high output mode, yielding 10-20 million paired end reads per sample.

Data analysis

Paired-end reads received in FASTQ format were run through FastQC (Andrews 2010), then aligned to the Ensembl R64 (Howe et al., 2021) reference genome using bowtie2 (Langmead and Salzberg 2012) with the "--no-unal", "--no-mixed" and "--no-discordant" flags. Fragment filtering to analyze small (0-120 bp), large (140-180 bp) and full (0-200 bp) fragment sets was accomplished with bowtie2 flags '-I' and '-X'. Aligned reads were sorted, filtered (MAPQ > 30), and indexed with samtools (Li et al., 2009). Deeptools (Ramirez et al., 2016) 'bamCoverage' was used to determine sequencing coverage from filtered BAM files and output bigWig-formatted files were normalized to 1x coverage (RPGC, genome size of 12.1 megabases). Coverage files were processed with deeptools 'bigWigCompare' to calculate the ratio of sample coverage over background. Peaks were identified from sample BAM files by MACS3 (https://macs3-project.github.io/MACS/) 'callpeak' using the corresponding timepoint control BAM ("No tag") as background with the flags '--keep-dup -all' and '-f BAMPE'. Bedtools (Quinlan and Hall, 2010) 'intersect' and 'merge' were used to subset/combine BED file entries. A pipeline was created to locate groups of peak centers from across timepoints using custom tools in python and functions from bedtools. Deeptools 'computeMatrix', 'plotHeatmap', and 'plotProfile' were used to generate matrices, heatmaps and average profiles according to input BED regions, respectively.

Predicted binding sites in the Ensembl reference genome for *S. cerevisiae* were located using FIMO (Grant et al., 2011) with the standard cutoff score of 10^{-4}. Published motifs were acquired from JASPAR 2022 database (Castro-Mondragon et al., 2022)

To perform motif analysis, BED files containing target regions were converted to FASTA using bedtools ‘getfasta’. Output FASTA files containing target sequences, now fragments of the yeast genome, were processed with MEME suite XSTREME (Grant and Bailey 2021) using standard input parameters and the JASPAR 2022 non-redundant database, except the motif range was expanded to 5-20bp. This also produces a list of motifs found in the submitted FASTA fragments, which can be viewed in Tables 1 and 2. The most enriched motifs that grouped with JASPAR entries for tagged DBPs were used with FIMO (standard cutoff) to locate sites in the yeast genome.

All scripts and tools used to analyze SpLiT-ChEC data are available on GitHub (Bankso/SCAR and Bankso/SEAPE). Data will be made available on the Gene Expression Omnibus.

Chapter V Additional Figures

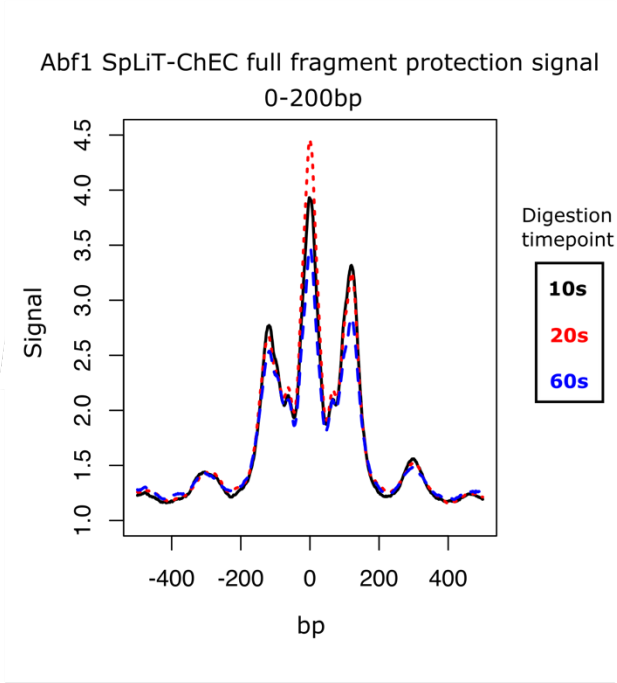


Figure S1 Abf1 SpLiT-ChEC total fragment pattern is a composite of SFP and LFP signal Overlay of signal derived from all DNA fragments less than 200bp collected at three timepoints of Abf1-targeted SpLiT-ChEC

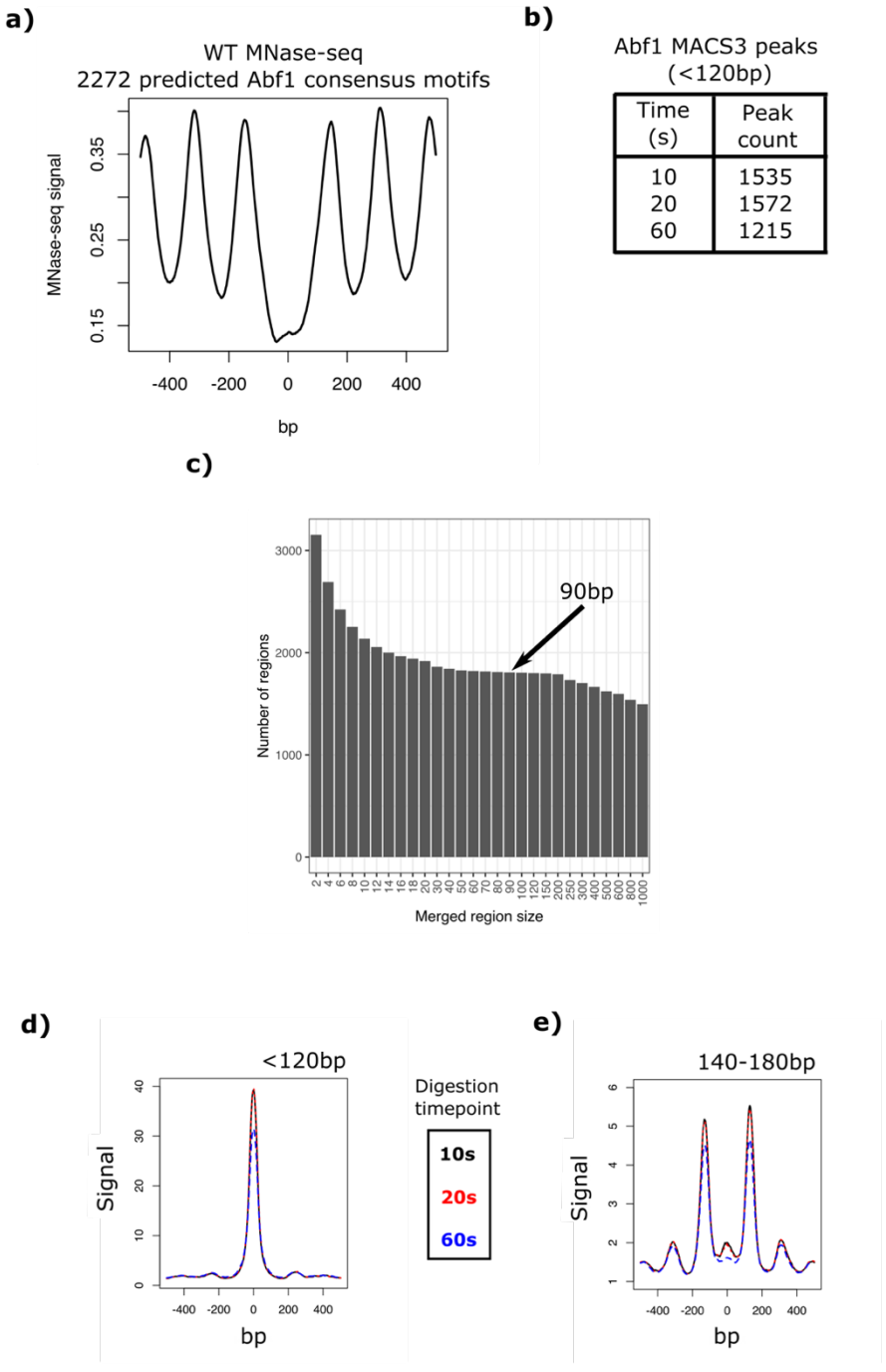


Figure S2 Abf1 SpLiT-ChEC signal is present at predicted sites and is found at similar sites across timepoints a) MNase-Seq signal from wild-type yeast cells at all predicted Abf1 consensus sequences in the yeast genome. b) Counts of the number of peaks found by MACS3 at each timepoint using small-fragment signal only. c) bar graph showing the number of peak regions reported after merging MACS3 peaks from all timepoints at increasing distances, noted on the x-axis. d) Abf1 SpLiT-ChEC SFP plotted at all center aligned 90bp targets for each Abf1 SpLiT-ChEC timepoint. e) equivalent to d, except for Abf1 SpLiT-ChEC LFP.

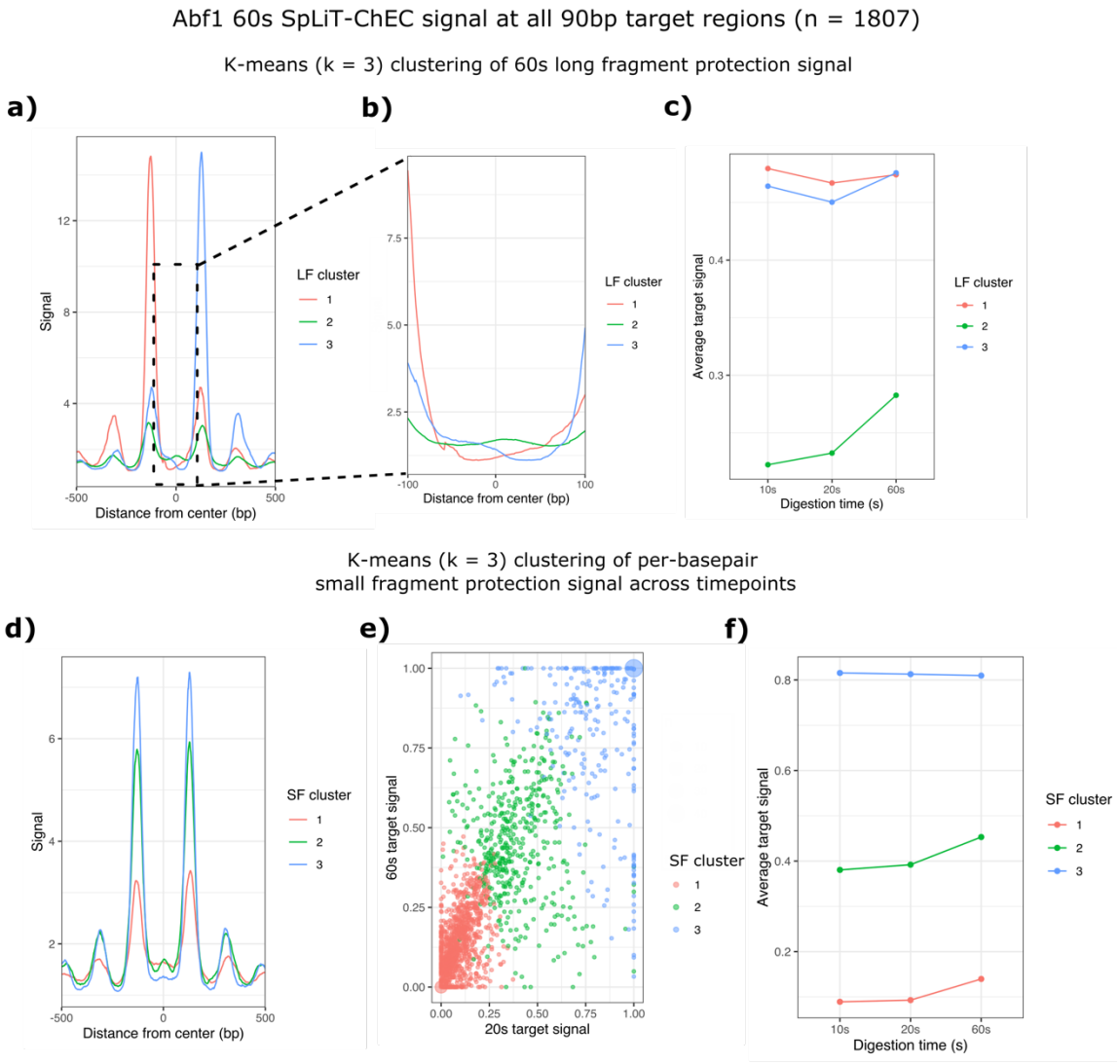


Figure S3 Clustering Abf1 SpLiT-ChEC signal without strand information identifies groups of similar signal a) Average Abf1 SpLiT-ChEC LFP signal found in each cluster (k=3) after applying k-means to only the 60s LFP signal. b) Identical to a), but zoomed in on the central 200bp region of clustered signal. c) Average per-basepair signal intensity of Abf1 SpLiT-ChEC SFP signal separated by LF cluster identified in a). d) Abf1 60s SpLiT-ChEC LFP signal average traces associated with k-means clusters (k=3) obtained from clustering SFP per-basepair signal associated with each 90bp target across all three timepoints. e) Plotting 20s and 60s per-basepair signal as cartesian coordinates for each of the 90bp regions associated with SpLiT-ChEC SFP signal peaks. Coloring by SFP signal k-means cluster shows separated groups of points. f) identical to c), except the signal is grouped by SFP signal k-means clusters.

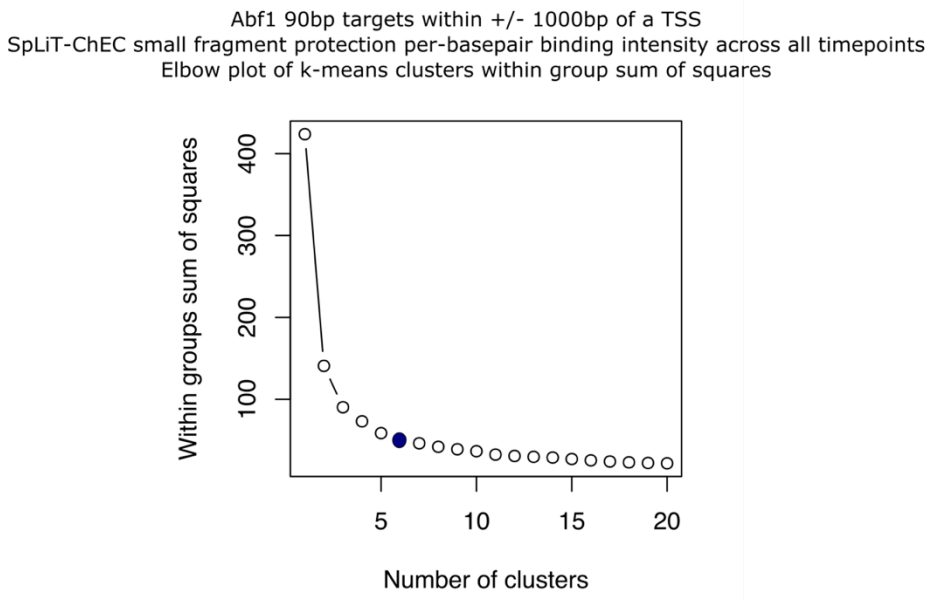


Figure S4 Locating optimal k-means cluster counts for Abf1 SFP per-basepair signal in 90bp targets near TSSs Elbow plot of WSS calculated for each k-means cluster. The blue dot at k=6 indicates the selected cluster number.

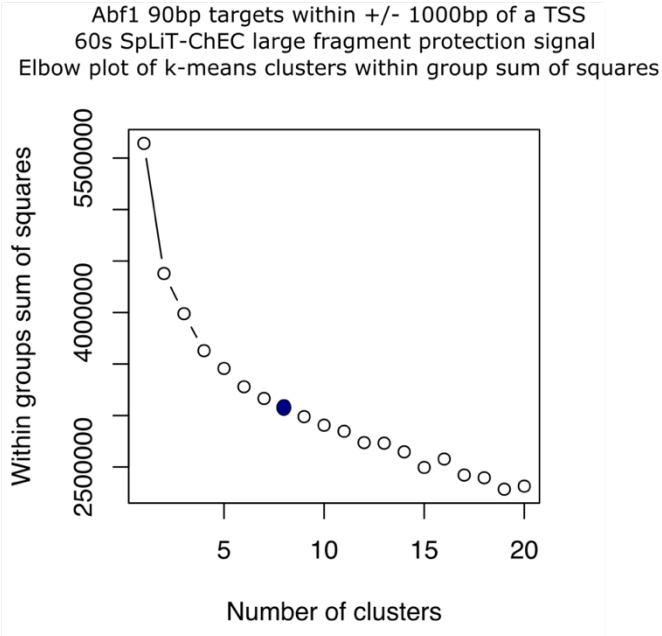


Figure S5 Locating optimal k-means cluster counts for Abf1 LFP signal in 90bp targets near TSSs Elbow plot of WSS calculated for each k-means cluster. The blue dot at k=8 indicates the selected cluster number.

Table 1 Simple Enrichment Analysis results for 1807 Abf1 90bp target sequences using motifs identified by XSTREME

ID	CONSENSUS	TP%	FP%	ENR_ RATIO	SCORE_ THR	PVALUE	EVALUE	QVALUE
MEME-2	ATCAYTWTNNACGR	52.74	3.07	16.8	8.3	8.23E-253	1.32E-251	1.32E-251
STREME-1	ATCAYTWTNNACGR	52.12	3.13	16.3	7.8	3.66E-247	5.86E-246	2.93E-246
STREME-2	RTTACCCGGMHD	15.55	1.29	11.5	11	6.12E-56	9.79E-55	3.26E-55
MEME-3	NYRTTACCCGV	14.94	1.72	8.41	12	1.35E-47	2.16E-46	5.39E-47
MEME-1	AAAAAAAAA ARAAAAA	29.01	14.87	1.95	6.6	7.70E-23	1.23E-21	2.46E-22
STREME-5	GAAAAAAAAA AAAAAAAAAA	18.5	7.19	2.56	7.5	1.12E-22	1.79E-21	2.98E-22
STREME-3	AAATTTTTCAB	6.95	0.98	6.71	10	4.08E-20	6.53E-19	9.32E-20
STREME-6	GAAGAAAAAA	29.87	18.32	1.63	6.4	7.12E-15	1.14E-13	1.42E-14
MEME-7	AGCTCATCGCW	3.13	0.12	17.3	13	1.10E-13	1.76E-12	1.96E-13
MEME-4	VAAAATTTTTCA	8.91	3.01	2.92	8.6	3.59E-13	5.75E-12	5.75E-13
STREME-4	CTCATCGC	2.64	0.06	22	15	1.96E-12	3.13E-11	2.84E-12
STREME-7	CARCAATAWK	3.32	0.8	3.93	12	1.72E-07	2.75E-06	2.29E-07
MEME-5	TTTGCCACC	4.92	2.09	2.31	2.1	6.98E-06	1.12E-04	8.59E-06
MEME-6	CCACTAAAATCACCTA	0.37	0	7	31	1.56E-02	2.49E-01	1.78E-02
MEME-8	GTATTATTGTTGAAGRRTAG	0.12	0	3	37	2.50E-01	4.00E+00	2.67E-01
MEME-9	GTWATATTGRCATAAGTGTA	0	0	1	37	1.00E+00	1.60E+01	1.00E+00

Table 2 Motifs identified in 1807 Abf1 90bp target sequences by XSTREME

ID	ALT_ID	CONSENSUS	SITES	SEA_ PVALUE	EVALUE	SIM_ MOTIF
ATCAYTWTNNACGR 1-ATCAYTWTNNACGR	MEME-2 STREME-1	ATCAYTWTNNACGR ATCAYTWTNNACGR	981 899	8.22E-253 3.66E-247	3.50e-587 9.29E-26	MA0265.2 (ABF1) MA0265.2 (ABF1)
MA0265.2 2-RTTACCCGGMHD	ABF1 STREME-2	WHWTCGTRT AWAGTGAYAND RTTACCCGGMHD	754 279	1.21E-225 6.12E-56	2.35E-223 5.92E-07	MA0265.2 MA0363.2 (REB1)
MA0363.2 NYRTTACCCGV	REB1 MEME-3	WNYRTTACCCGGMHD NYRTTACCCGV	313 342	1.33E-48 1.35E-47	2.59E-46 3.60E-122	MA0363.2 MA0363.2 (REB1)
AAAAAAAA AARAAAAA 5-GAAAAAAAAA AAAAAAAAAA	MEME-1 STREME-5	AAAAAAAA AARAAAAA GAAAAAAAA AAAAAAAAAAA	526 216	7.70E-23 1.12E-22	2.30E-268 1.02E-01	MA0277.1 (AZF1) MA0277.1 (AZF1)
3-AAATTTTTCAB MA0421.1	STREME-3 NSI1	AAATTTTTCAB TTTACCCGGM	120 119	4.08E-20 2.94E-17	6.83E-04 5.74E-15	MA0390.1 (STB3) MA0421.1
6-GAAGAAAAAA	STREME-6	GAAGAAAAAA	202	7.12E-15	6.34E-01	MA0377.1 (SFL1)

ID	ALT_ID	CONSENSUS	SITES	SEA_ PVALUE	EVALUE	SIM_ MOTIF
AGCTCATCGCW	MEME-7	AGCTCATCGCW	51	1.10E-13	1.30E-08	MA0350.1 (TOD6)
VAAAATTTTTCA	MEME-4	VAAAATTTTTCA	129	3.59E-13	9.40E-43	MA0390.1 (STB3)
4-CTCATCGC	STREME-4	CTCATCGC	125	1.96E-12	6.50E-02	MA0350.1 (TOD6)
MA0350.1	TOD6	NBBNDHASCT CATCGCSHYND	185	3.69E-11	7.20E-09	MA0350.1
MA0351.1	DOT6	DNBBDCWSC TCATCGCNYCYT	169	3.81E-11	7.42E-09	MA0351.1
MA0390.1	STB3	GTYHAAAWT TTTTCACTNHNN	415	1.13E-09	2.21E-07	MA0390.1
MA0365.1	RFX1	SGTTGCYA	406	4.70E-08	9.17E-06	MA0365.1
7-CARCAATAWK	STREME-7	CARCAATAWK	108	1.72E-07	1.18E+00	MA0371.1 (ROX1)
MA0412.2	UME6	WWTTAG CCGCCSANV	114	1.88E-06	3.66E-04	MA0412.2
TTTGCCACC	MEME-5	TTTGCCACC	46	6.98E-06	3.50E-13	MA0373.1 (RPN4)
MA0356.1	PHO2	WTAWTW	394	1.04E-05	2.03E-03	MA0356.1
MA0359.2	RAP1	DRTGTAT GGGTGTWW	51	3.70E-05	7.21E-03	MA0359.2
CCACTAAAATCACCTA	MEME-6	CCACTAAA ATCACCTA	12	1.56E-02	2.90E-11	CCACTAAA ATCACCTA
GTATTATTGT TGAAGRRTAG	MEME-8	GTATTATTGT TGAAGRRTAG	11	2.50E-01	5.30E-07	MA0426.1 (YHP1)
GTWATATTGR CATAAGTGTA	MEME-9	GTWATATTG RCATAAGTGTA	9	1.00E+00	4.30E-06	MA0426.1 (YHP1)

Bridge from Chapter V to VI

As mentioned in the Chapter IV to V bridge, we were interested in the potential for SpLiT-ChEC to capture genome-wide localization information for Isw2, but we were also interested in applying SpLiT-ChEC to other DNA binding proteins. Following our work demonstrating the effectiveness of SpLiT-ChEC with Abf1, we turned attention to SpLiT-ChEC data collected from SpyTagged versions of Isw2, but also Gal4, Med14, Ume6, and Reb1. The results from our work in Chapters V and VI suggest that SpLiT-ChEC can be generally applied to DNA-binding proteins to capture high-resolution information on protein localization in the yeast genome.

As with the previous chapter, I engineered all the yeast strains described, performed all the experiments, prepared the NGS libraries, and processed and visualized the data with my established pipelines.

References Cited

Adams CC, Workman JL. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Mol Cell Biol. 1995;15(3):1405-1421. doi:10.1128/MCB.15.3.1405

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Badis G, Chan ET, van Bakel H, et al.,. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell. 2008;32(6):878-887. doi:10.1016/j.molcel.2008.11.020

Bailey TL and Grant CE. SEA: Simple Enrichment Analysis of motifs. *BioRxiv* (2021).

Balsalobre A, Drouin J. Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol. 2022 Jul;23(7):449-464. doi: 10.1038/s41580-022-00464-z. Epub 2022 Mar 9. PMID: 35264768.

Barnes T, Korber P. The Active Mechanism of Nucleosome Depletion by Poly(dA:dT) Tracts In Vivo. Int J Mol Sci. 2021 Jul 30;22(15):8233. doi: 10.3390/ijms22158233. PMID: 34360997; PMCID: PMC8347975.

Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem. 2002;71:333-74. doi: 10.1146/annurev.biochem.71.110601.135425. Epub 2001 Nov 9. PMID: 12045100.

Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD. RSC, an Essential, Abundant Chromatin-Remodeling Complex. Cell. 1996; 87(7):1249-1260. doi:10.1016/S0092-8674(00)81820-6.

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N, Fornes O, Leung TY, Aguirre A, Hammal F, Schmelter D, Baranasic D, Ballester B, Sandelin A, Lenhard B, Vandepoele K, Wasserman WW, Parcy F, Mathelier A. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022 Jan 7;50(D1):D165-D173. doi: 10.1093/nar/gkab1113. PMID: 34850907; PMCID: PMC8728201.

Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013;65(10):1357-1369. doi:10.1016/j.addr.2012.09.039

Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annu Rev Biochem. 2009;78:273-304. doi: 10.1146/annurev.biochem.77.062706.153223. PMID: 19355820.

Coux RX, Owens NDL, Navarro P. Chromatin accessibility and transcription factor binding through the perspective of mitosis. Transcription. 2020;11(5):236-240. doi:10.1080/21541264.2020.1825907

Cui J, Kaandorp JA, Sloot PM, Lloyd CM, Filatov MV. Calcium homeostasis and signaling in yeast cells and cardiac myocytes. FEMS Yeast Res. 2009 Dec;9(8):1137-47. doi: 10.1111/j.1567-1364.2009.00552.x. Epub 2009 Jul 16. PMID: 19678847.

Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. Chem Rev. 2014 Mar 26;114(6):3203-23. doi: 10.1021/cr400730x. Epub 2014 Feb 6. PMID: 24502198; PMCID: PMC3983126.

Donovan DA, Crandall JG, Banks OGB, et al.,. Engineered Chromatin Remodeling Proteins for Precise Nucleosome Positioning. Cell Rep. 2019;29(8):2520-2535.e4. doi:10.1016/j.celrep.2019.10.046

Fennessy RT, Owen-Hughes T. Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. Nucleic Acids Res. 2016 Sep 6;44(15):7189-203. doi: 10.1093/nar/gkw331. Epub 2016 Apr 22. PMID: 27106059; PMCID: PMC5009725.

Ferri AL, Lin W, Mavromatakis YE, Wang JC, Sasaki H, Whitsett JA, Ang SL. Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. Development. 2007 Aug;134(15):2761-9. doi: 10.1242/dev.000141. Epub 2007 Jun 27. PMID: 17596284.

Flaus A, Martin DM, Barton GJ, Owen-Hughes T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res. 2006 May 31;34(10):2887-905. doi: 10.1093/nar/gkl295. PMID: 16738128; PMCID: PMC1474054.

Ganapathi M, Palumbo MJ, Ansari SA, He Q, Tsui K, Nislow C, Morse RH. Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res. 2011 Mar;39(6):2032-44. doi: 10.1093/nar/gkq1161. Epub 2010 Nov 16. PMID: 21081559; PMCID: PMC3064788.

Ganguli D, Chereji RV, Iben JR, Cole HA, Clark DJ. RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Res. 2014 Oct;24(10):1637-49. doi: 10.1101/gr.177014.114. Epub 2014 Jul 11. PMID: 25015381; PMCID: PMC4199373.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343-5. doi: 10.1038/nmeth.1318. Epub 2009 Apr 12. PMID: 19363495.

Grant CE, Bailey TL and Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017-1018.

Grant CE and Bailey TL. XSTREME: comprehensive motif analysis of biological sequence datasets. *BioRxiv*. 2021.

Hatlem D, Trunk T, Linke D, Leo JC. Catching a SPY: Using the SpyCatcher-SpyTag and Related Systems for Labeling and Localizing Bacterial Proteins. Int J Mol Sci. 2019;20(9):2129. Published 2019 Apr 30. doi:10.3390/ijms20092129

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A. 2011 Nov 8;108(45):18318-23. doi: 10.1073/pnas.1110731108. Epub 2011 Oct 24. PMID: 22025700; PMCID: PMC3215028.

Henikoff S. Mechanisms of Nucleosome Dynamics In Vivo. Cold Spring Harb Perspect Med. 2016 Sep 1;6(9):a026666. doi: 10.1101/cshperspect.a026666. PMID: 27503998; PMCID: PMC5008063.

Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, IIsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P. Ensembl 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D884-D891. doi: 10.1093/nar/gkaa942. PMID: 33137190; PMCID: PMC7778975.

Huang RC, Bonner J. Histone, a suppressor of chromosomal RNA synthesis. Proc Natl Acad Sci U S A. 1962 Jul 15;48(7):1216-22. doi: 10.1073/pnas.48.7.1216. PMID: 14036409; PMCID: PMC220935.

Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007 Jun 8;316(5830):1497-502. doi: 10.1126/science.1141319. Epub 2007 May 31. PMID: 17540862.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J. DNA-binding specificities of human transcription factors. Cell. 2013 Jan 17;152(1-2):327-39. doi: 10.1016/j.cell.2012.12.009. PMID: 23332764.

Ju QD, Morrow BE, Warner JR. REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. Mol Cell Biol. 1990 Oct;10(10):5226-34. doi: 10.1128/mcb.10.10.5226-5234.1990. PMID: 2204808; PMCID: PMC361205.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. 2009 Mar 19;458(7236):362-6. doi: 10.1038/nature07667. Epub 2008 Dec 17. PMID: 19092803; PMCID: PMC2658732.

Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. High-resolution mapping of transcription factor binding sites on native chromatin. Nat Methods. 2014 Feb;11(2):203-9. doi: 10.1038/nmeth.2766. Epub 2013 Dec 15. PMID: 24336359; PMCID: PMC3929178.

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Apr 29;10(1):1930. doi: 10.1038/s41467-019-09982-5. PMID: 31036827; PMCID: PMC6488672.

Keeble AH, Banerjee A, Ferla MP, Reddington SC, Anuar INAK, Howarth M. Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics. Angew Chem Int Ed Engl. 2017 Dec 22;56(52):16521-16525. doi: 10.1002/anie.201707623. Epub 2017 Dec 5. PMID: 29024296; PMCID: PMC5814910.

Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974 May 24;184(4139):868-71. doi: 10.1126/science.184.4139.868. PMID: 4825889.

Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF, Korber P. Genomic Nucleosome Organization Reconstituted with Pure Proteins. Cell. 2016 Oct 20;167(3):709-721.e12. doi: 10.1016/j.cell.2016.09.045. PMID: 27768892; PMCID: PMC5240917.

Kubik S, O'Duibhir E, de Jonge WJ, Mattarocci S, Albert B, Falcone JL, Bruzzone MJ, Holstege FCP, Shore D. Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. Mol Cell. 2018 Jul 5;71(1):89-102.e5. doi: 10.1016/j.molcel.2018.05.030. PMID: 29979971.

Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, Hu G, Ni B, Zhao K. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature. 2018 Oct;562(7726):281-285. doi: 10.1038/s41586-018-0567-3. Epub 2018 Sep 26. Erratum in: Nature. 2018 Nov 6;: PMID: 30258225; PMCID: PMC8353605.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923. PMID: 22388286; PMCID: PMC3322381.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science. 2002 Oct 25;298(5594):799-804. doi: 10.1126/science.1075090. PMID: 12399584.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet. 2007 Oct;39(10):1235-44. doi: 10.1038/ng2117. Epub 2007 Sep 16. PMID: 17873876.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997 Sep 18;389(6648):251-60. doi: 10.1038/38444. PMID: 9305837.

McKnight LE, Crandall JG, Bailey TB, Banks OGB, Orlandi KN, Truong VN, Donovan DA, Waddell GL, Wiles ET, Hansen SD, Selker EU, McKnight JN. Rapid and inexpensive preparation of genome-wide nucleosome footprints from model and non-model organisms. STAR Protocols 2022; 2(2). doi:10.1016/j.xpro.2021.100486. PMID: 34041500; PMCID: PMC8141940.

Meers MP, Janssens DH, Henikoff S. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. Mol Cell. 2019;75(3):562-575.e5. doi:10.1016/j.molcel.2019.05.025

Miles S, Li L, Davison J, Breeden LL (2013) Xbp1 Directs Global Repression of Budding Yeast Transcription during the Transition to Quiescence and Is Important for the Longevity and Reversibility of the Quiescent State. PLOS Genetics 9(10): e1003854. https://doi.org/10.1371/journal.pgen.1003854

O'Connor T, Grant CE, Bodén M, Bailey TL, "T-Gene: Improved target gene prediction", Bioinformatics. 2020;36(12):3902-3904.

Ozsolak F, Song JS, Liu XS, Fisher DE. High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol. 2007 Feb;25(2):244-8. doi: 10.1038/nbt1279. Epub 2007 Jan 14. PMID: 17220878.

Policastro RA, Zentner GE. Enzymatic methods for genome-wide profiling of protein binding sites. Brief Funct Genomics. 2018 Mar 1;17(2):138-145. doi: 10.1093/bfgp/elx030. PMID: 29028882.

Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6): 841–842.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W160-5. doi: 10.1093/nar/gkw257. Epub 2016 Apr 13. PMID: 27079975; PMCID: PMC4987876.

Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell. 2011 Dec 9;147(6):1408-19. doi: 10.1016/j.cell.2011.11.013. PMID: 22153082; PMCID: PMC3243364.

Rhode PR, Elsasser S, Campbell JL. Role of multifunctional autonomously replicating sequence binding factor 1 in the initiation of DNA replication and transcriptional control in Saccharomyces cerevisiae. Mol Cell Biol. 1992;12(3):1064-1077. doi:10.1128/mcb.12.3.1064-1077.1992

Rossi MJ, Lai WKM, Pugh BF. Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. Genome Res. 2018;28(4):497-508. doi:10.1101/gr.229518.117

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. Dynamic regulation of nucleosome positioning in the human genome. Cell. 2008 Mar 7;132(5):887-98. doi: 10.1016/j.cell.2008.02.022. PMID: 18329373.

Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 2017 Jan 16;6:e21856. doi: 10.7554/eLife.21856. PMID: 28079019; PMCID: PMC5310842.

Singh AK, Schauer T, Pfaller L, Straub T, Mueller-Planitz F. The biogenesis and function of nucleosome arrays. Nat Commun. 2021 Dec 1;12(1):7011. doi: 10.1038/s41467-021-27285-6. PMID: 34853297; PMCID: PMC8636622.

Svaren J, Klebanow E, Sealy L, Chalkley R. Analysis of the competition between nucleosome formation and transcription factor binding. J Biol Chem. 1994 Mar 25;269(12):9335-44. PMID: 8132673.

Tachiwana H, Dacher M, Maehara K, Harada A, Seto Y, Katayama R, Ohkawa Y, Kimura H, Kurumizaka H, Saitoh N. Chromatin structure-dependent histone incorporation revealed by a genome-wide deposition assay. Elife. 2021 May 10;10:e66290. doi: 10.7554/eLife.66290. PMID: 33970102; PMCID: PMC8110306.

Thåström A, Bingham LM, Widom J. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. J Mol Biol. 2004 May 7;338(4):695-709. doi: 10.1016/j.jmb.2004.03.032. PMID: 15099738.

Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. Epigenetics Chromatin. 2014 Nov 20;7(1):33. doi: 10.1186/1756-8935-7-33. PMID: 25473421; PMCID: PMC4253006.

van Bakel H, Tsui K, Gebbia M, Mnaimneh S, Hughes TR, Nislow C. A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet. 2013;9(5):e1003479. doi:10.1371/journal.pgen.1003479

Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol. 2015 Mar;16(3):178-89. doi: 10.1038/nrm3941. Epub 2015 Feb 4. PMID: 25650798.

Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet. 2014 Feb;15(2):69-81. doi: 10.1038/nrg3623. Epub 2013 Dec 17. PMID: 24342920; PMCID: PMC6322398.

Walker A, Taylor J, Rowe D, Summers D. A method for generating sticky-end PCR products which facilitates unidirectional cloning and the one-step assembly of complex DNA constructs. Plasmid. 2008 May;59(3):155-62. doi: 10.1016/j.plasmid.2008.02.002. PMID: 18395798.

Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. 2010;20(1):90-100. doi:10.1101/gr.098509.109

Yarragudi A, Miyake T, Li R, Morse RH. Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. Mol Cell Biol. 2004 Oct;24(20):9152-64. doi: 10.1128/MCB.24.20.9152-9164.2004. PMID: 15456886; PMCID: PMC517901.

Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell. 2012;149(7):1461-1473. doi:10.1016/j.cell.2012.04.036

Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A. 2012 Mar 20;109(12):E690-7. doi: 10.1073/pnas.1115485109. Epub 2012 Feb 24. PMID: 22366317; PMCID: PMC3311370.

Zamanighomi M, Lin Z, Wang Y, Jiang R, Wong WH. Predicting transcription factor binding motifs from DNA-binding domains, chromatin accessibility and gene expression data. Nucleic Acids Res. 2017 Jun 2;45(10):5666-5677. doi: 10.1093/nar/gkx358. PMID: 28472398; PMCID: PMC5449588.

Zentner GE, Kasinathan S, Xin B, Rohs R, Henikoff S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat Commun. 2015 Oct 22;6:8733. doi: 10.1038/ncomms9733. Erratum in: Nat Commun. 2015;6:10264. PMID: 26490019; PMCID: PMC4618392.

CHAPTER VI

GENERAL APPLICATION OF SPLIT-CHEC TO DNA-BINDING PROTEINS AND PRELIMINARY ANALYSIS OF ISW2-TARGETED SPLIT-CHEC SIGNAL

Introduction

A significant motivation in developing SpLiT-ChEC was its potential to generate high-resolution localization maps of chromatin remodeling proteins, in particular Isw2, which we envisioned could help address some of the outstanding questions related to Isw2 localization and chromatin remodeling activity in the yeast genome. Isw2 is often considered a poor target for immunoprecipitation (Gelbart et al., 2005). It is possible to map Isw2 targets using catalytically inactive mutants (Gelbart et al., 2005), but this provides an incomplete picture, since the nucleosome remodeling activity of Isw2 is dependent on ATP hydrolysis. To this end, we collected SpLiT-ChEC samples from yeast where MNase-SpyCatcher was targeted using SpyTag fused to Isw2. All samples produced sufficient DNA fragments for sequencing, which we compared to the ratio of signal over background obtained from untargeted MNase digestion in the base SpLiT-ChEC strain at each timepoint.

Results

Visualizing SFP and LFP signal at all TSSs (Figure 22a and 22b) in the yeast genome shows a propensity for Isw2 to localize near actively transcribed genes. The LFP signal shows a shoulder biased towards regions of Isw2 SFP signal, which we anticipate corresponds to cells in the population where Isw2 remodels nucleosomes. As with Abf1, MACS3 was able to identify peaks at each of the three timepoints we selected for

sequencing (5s, 20s, 60s) based on SFP signal. When the extend-and-merge process was applied to this set of peaks (Figure 23A), the optimal merge distance was around 70bp, reporting 1041 sites as the set of optimally merged Isw2 targets. Plotting Isw2 average SFP signal (Figure 23B) shows that signal is located in the central area of the 70bp target regions.

Since some Isw2 targeted remodeling activity is associated with Ume6, we counted Isw2 70bp target regions that contain a Ume6 consensus sequence (JASPAR entry MA0412.2, ChIP-exo, via FIMO) and found that 61 of the sites matched this criterion (Figure 24A). This aligns with our previous observations that 58 Ume6 binding sites are Isw2 remodeling targets (Donovan et al., 2021). Visualizing SFP signal at these sites shows that Isw2 is effectively localized and that well-defined LFP signal can be found proximal to target centers (Figure 24B).

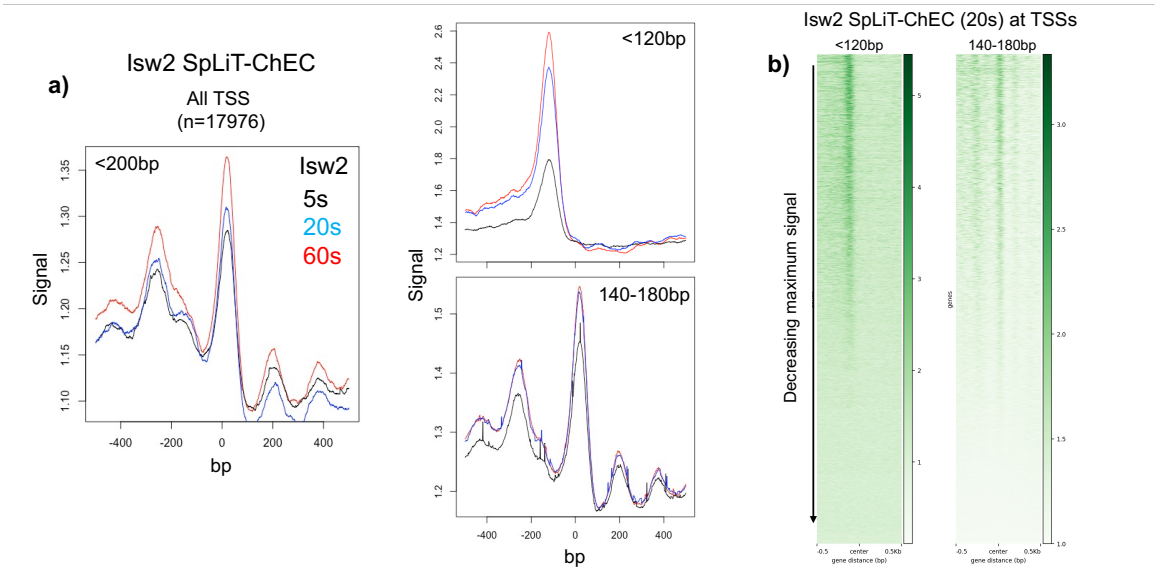


Figure 22. Isw2 SpLiT-ChEC shows high-detail signal at TSSs across the yeast genome a) Overlayed average signal from the three sequenced Isw2 SpLiT-ChEC timepoints at all TSSs in the yeast genome. Fragment size ranges are designated in each plot. **b)** Heatmaps showing 20s Isw2 SpLiT-ChEC signal centered on TSSs for both SFP and LFP signal

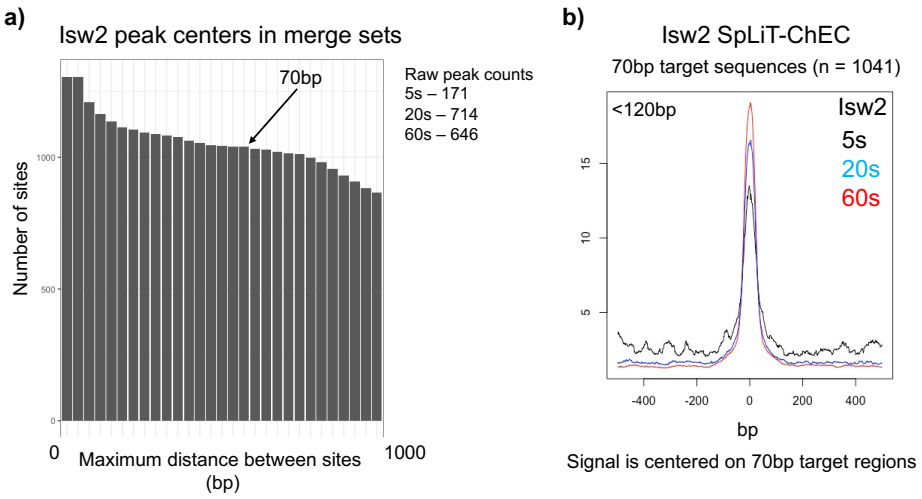


Figure 23. Peaks identified in Isw2 SpLiT-ChEC SFP signal can be merged to locate target sites. **a)** Bar chart showing the number of sites remaining after each extend-and-merge calculation applied to the total set of peaks identified from Isw2 SFP signal at each timepoit. **b)** Overlay of average SpLiT-ChEC SFP signal at each timepoint for the set of 70bp target regions identified in a).

While preliminary, we believe this data supports the notion that SpLiT-ChEC could be applied to studying TF and ChRP mutants where disrupted nucleosome positioning is expected or theorized. Mutations in ChRPs are common in human cancers; we expected that SpLiT-ChEC could be used to measure defects in chromatin remodeling imparted by these mutations by allowing ChRP localization and nucleosome-related signal to be measured simultaneously.

Though not yet fully analyzed, we have collected and sequenced SpLiT-ChEC samples for several other DPBs, including Reb1 (three biological replicates), Med14 (two biological replicates), Gal4 (two biological replicates), and Ume6 (three biological replicates). For each protein, three timepoints were selected for sequencing. Preliminary analysis suggests that each sample contains fragments corresponding to SFP and LFP

signal and could be processed in a manner similar to Abf1 and Isw2 SpLiT-ChEC. This data will be made available on the Gene Expression Omnibus.

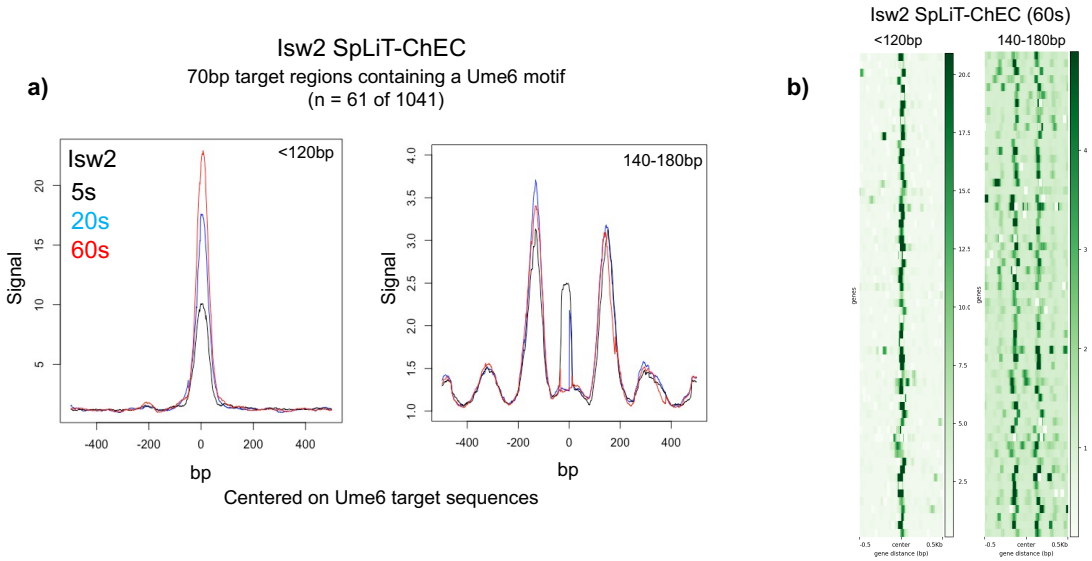


Figure 24. A subset of Isw2 SpLiT-ChEC target sites contain Ume6 consensus sequences. a) Overlay of average signal from Isw2 SFP and LFP at each timepoint for 70bp target regions that also contain a FIMO-predicted Ume6 consensus binding motif. **b)** Heatmap showing 60s SpLiT-ChEC SFP and LFP signal for the sites used to create average traces shown in a).

References Cited

Donovan DA, Crandall JG, Truong VN, Vaaler AL, Bailey TB, Dinwiddie D, Banks OG, McKnight LE, McKnight JN. Basis of specificity for a conserved and promiscuous chromatin remodeling protein. Elife. 2021 Feb 12;10:e64061. doi: 10.7554/eLife.64061. PMID: 33576335; PMCID: PMC7968928.

Gelbart ME, Bachman N, Delrow J, Boeke JD, Tsukiyama T. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. Genes Dev. 2005;19(8):942-954. doi:10.1101/gad.1298905

CHAPTER VII

CONCLUSIONS AND FUTURE DIRECTIONS

The work detailed in this dissertation contributes to our understanding of nucleosome positioning mechanisms, including the methods by which researchers can manipulate and capture information from chromatin organizational patterns. Our work with E-ChRPs, described in Chapter II, demonstrated the utility of targeted chromatin remodeling for inducing ectopic nucleosome positions. In particular, the Chd1-SpyCatcher system provides versatility in target selection, which can be performed by simply appending a SpyTag to a protein of interest. Another important result from our work with Chd1-SpyCatcher is the observation that SpyTagged factors can be evicted from their binding sites by a nucleosome being pulled into an ectopic position. Theoretically, this could be used as a form of transcriptional control and underpins the importance of well-controlled nucleosome positioning. E-ChRPs, including degradable versions, could be useful in higher eukaryotes, where we may be able to observe nucleosome position dynamics in more complex environments. Based on our initial results, budding yeast may not be the best system for studying this phenomenon, as it has relatively low amounts of heterochromatin, which we anticipate underlies nucleosome position inheritance (Escobar et al., 2019), if it occurs. However, transcription start sites producing non-canonical transcripts can be associated with specific nucleosome positions in the yeast genome (Wei et al., 2019). These targets may serve as suitable loci for experiments measuring the transcript response with respect to inheritance of ectopic nucleosomes. Further, it would be interesting to study the implications of ectopic nucleosome position inheritance in the context of a system where cryptic transcripts are

deleterious, which may provide more insight on how nucleosome positions become reset, if possible, to avoid detrimental effects.

With regards to Isw2 and the biochemical basis for its interaction with Ume6 detailed in Chapter III, we found this activity is imparted by the adaptor protein Itc1, which interacts with Ume6 to target Isw2 to Ume6 consensus motifs. Importantly, the short helical element recognized by Itc1 is believed to be present in other TFs, indicating this could be a general mechanism of Isw2 recruitment to position +1 nucleosomes.

Through simplifying and shortening the process for collecting nucleosomes by MNase digestion, as presented in Chapter IV, we are confident this technique will be more accessible to researchers interested in studying nucleosome positioning in their experimental systems. We were also able to demonstrate the applicability of the quick MNase technique to several different biological systems, indicating that it could be adapted to collect nucleosome positions for most organisms of interest.

Based on the data in Chapters V and VI, we anticipate SpLiT-ChEC can be applied to studying chromatin organization, protein-DNA interactions, and protein-protein interactions. By collecting DNA fragments arising from multiple forms of DNA protection, we showed capture of protein binding and chromatin organization elements in proximity to target sites in the yeast genome. Other systems provide this information, though we have not observed a system that achieves DNA digestion through covalent linkage of a nuclease to a target protein. While not always desirable, this is advantageous in the systems we often study, because it allows for DNA to be digested *in situ* without the need for crosslinking or antibody detection. We believe SpLiT-ChEC could be an effective tool to examine the differences in protein localization and nucleosome

positioning that arise from mutations in TFs or ChRPs. One simple way to test this is with ChRP mutants that are known to be deficient in ATPase activity (Gelbart et al., 2005), which is required for nucleosome remodeling, and examine the difference in SpLiT-ChEC signal between mutant and WT samples. We also anticipate SpLiT-ChEC could be modified for use in higher eukaryotes, using CRISPR to introduce SpyTags, and expressing MNase-SpyCatcher from a plasmid or introducing purified recombinant versions of the fusion.

One theoretical extension of SpLiT-ChEC is to pair the technique with nascent chromatin labeling (Vasseur et al., 2016). By incorporating a modifiable nucleotide substitute like ethylene deoxyuridine (EdU) into replicating DNA, nascent chromatin can be labeled with click reactions isolated from bulk DNA for analysis. This could provide insight into Isw2-bound target sites where remodeling was achieved by allowing the pre-remodeled LFP signal to be observed. This would also be an interesting technique to pair with E-ChRP targeted chromatin remodeling, where ectopic nucleosome positions could be induced and SpLiT-ChEC could be used to monitor SFP and LFP signal at target sites, with or without nascent chromatin labeling. In this type of system, SpLiT-ChEC applied to TFs that may be influenced by ectopic nucleosome positions may also provide insight on how mispositioned nucleosomes can influence chromatin structure and protein binding.

Analyzing data from SpLiT-ChEC has suggested that Abf1 targeting and nucleosome positioning near targets are coupled, but no obvious preference for Abf1 targets with or without a minimal motif was present when considering LFP signal patterns. While co-targeting of Abf1 at Reb1 targets is well-described, this is not

sufficient to explain LFP patterning at all Abf1 targets without a consensus motif. As such, it is not clear how DNA sequence plays a direct role in our LFP signal patterns. Based on this observation, we theorize that there are subsets of DNA sequences associated with similar LFP signal patterns independent of transcription factor binding sites. Previous work has determined a clear relationship between DNA sequence and nucleosome positions (Kaplan et al., 2009), but the short repeating sequences that are suitable for nucleosome formation are not easily located by standard motif analysis.

To better investigate the links between target sequences and LFP signal patterns, we propose that DNA sequences recovered from SpLiT-ChEC be analyzed using classical machine (Alvarez-Gonzalez and Erill, 2021) or deep learning techniques (Koo and Ploenzke, 2020). One simple way to test the feasibility of this approach would be to determine if Abf1 target sites can be accurately classified into sites with or without a motif, based on LFP signal patterns or signal intensity measurements from SpLiT-ChEC data. Further, the presence of different TF motifs could be used as a feature in this analysis, to better understand if co-targeting of proteins helps define nucleosome positions in the yeast genome.

A more complex route would be to convert target sequences or groups of target sequences to frequency chaos game representations (FCGRs), which have been useful for researchers that perform alignment-free comparison of DNA sequences (Löchel and Heider, 2021). The base chaos game transformation allows a sequence of basepairs in DNA to be converted to a 2-dimensional plot, which represents the combinations of nucleotides that appear in the original sequence and their order. CGRs can also be used to reconstruct the original sequence of nucleotides by performing the calculation in reverse.

After dividing the plot into 2-dimensional bins and coloring the bins by the relative number of points in each region, we could generate pictures of DNA sequences that are based on the frequency of each group of nucleotides used to build the CGR map. We imagine this framework could be applied to SpLiT-ChEC target sequences or groups of target sequences linked to form a mini-genome, which would be classified based on LFP signal patterns or other biological information.

FCGR maps of DNA sequences could then be used with a convolutional neural network (Vaz and Balaji, 2021), which is able to process pictures and learn about the visual patterns associated with inputs to make predictions on which group a picture belongs to. Through this analysis, we may be able to learn about the sequence features associated with specific nucleosome positioning patterns at TF-bound sites throughout the genome, independent of the need to classify sites using consensus motifs. Further, FCGR could simplify the comparison of DNA sequences between groups, by allowing for difference calculations to be made between FCGRs that describe the similarity between target sequences obtained from SpLiT-ChEC. It may be possible to find differences in the set of sequences recovered from binding sites of WT and mutant proteins, which could provide a relatively simple method for comparing groups of DNA sequences recovered from experiments measuring DBP localization.

In summary, this dissertation includes several studies that further our understanding of the mechanisms that define and maintain nucleosome positions in eukaryotic genomes. The synthetic protein systems and computational processes designed for these studies may serve as source material for other researchers. In time, expanding the knowledge of nucleosome positioning mechanisms and refining the tools available for

controlling these mechanisms could advance treatments for human disease. By combining rich biological datasets with artificial intelligence, it is highly feasible that connections between chromatin organization, DNA sequence, and protein binding will be better understood. Artificial intelligence has already proven useful by supporting predictive tools for human disease treatment and may be furthered by well-defined models of protein binding in the context of nucleosome patterning.

References Cited

Alvarez-Gonzalez, S. & Erill, I. (2021). Design of Machine Learning Models for the Prediction of Transcription Factor Binding Regions in Bacterial DNA. *Engineering Proceedings*, *7*(1), 59. doi:10.3390/engproc2021007059

Escobar, T. M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., & Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. *Cell*, *179*(4), 953–963.e11. doi:10.1016/j.cell.2019.10.009

Gelbart, M. E., Bachman, N., Delrow, J., Boeke, J. D., & Tsukiyama, T. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev.*, *19*(8), 942-954. doi:10.1101/gad.1298905

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, *458*(7236), 362-366. doi: 10.1038/nature07667.

Koo, P. K., & Ploenzke, M. (2020). Deep learning for inferring transcription factor binding sites. *Current opinion in systems biology*, *19*, 16–23. https://doi.org/10.1016/j.coisb.2020.04.001

Löchel, H. F., & Heider, D. (2021). Chaos game representation and its applications in bioinformatics. *Computational and structural biotechnology journal*, *19*, 6263–6271. https://doi.org/10.1016/j.csbj.2021.11.008

Vasseur, P., Tonazzini, S., Ziane, R., Camasses, A., Rando, O. J., & Radman-Livaja, M. (2016). Dynamics of Nucleosome Positioning Maturation following Genomic Replication. *Cell reports*, *16*(10), 2651–2665. https://doi.org/10.1016/j.celrep.2016.07.083

Vaz, J. M., & Balaji, S. (2021). Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. *Molecular diversity*, *25*(3), 1569–1584. https://doi.org/10.1007/s11030-021-10225-3

Wei, W., Hennig, B. P., Wang, J., Zhang, Y., Piazza, I., Pareja Sanchez, Y., Chabbert, C. D., Adjalley, S. H., Steinmetz, L. M., & Pelechano, V. (2019). Chromatin-sensitive cryptic promoters putatively drive expression of alternative protein isoforms in yeast. *Genome research*, *29*(12), 1974–1984. https://doi.org/10.1101/gr.243378.118

APPENDIX A: YEAST STRAINS GENERATED

Strain ID	Relevant genotype
YS356	Ume6-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS357	Ume6-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS358	Hsf1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS359	Hsf1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS360	Abf1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS361	Abf1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS362	Reb1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS363	Reb1-SpyFLAG (NATMX); GPD-Chd1[D513N]-Spycatcher (P196 [+URA])
YS364	HO-pGAL Chd1-GGSx11-SpyCatcher (KANMX); Abf1-SpyFLAG (NATMX)
YS365	HO-pGAL Chd1-GGSx11-SpyCatcher (KANMX); Hsf1-SpyFLAG (NATMX)
YS366	HO-pGAL Chd1-GGSx11-SpyCatcher (KANMX); Reb1-SpyFLAG (NATMX)
YS402	HO-pGAL Chd1-GGSx11-Engrailed (KANMX)
YS403	HO-pGAL Chd1-GGSx11-Engrailed (KANMX)
YS404	HO-pGAL Chd1-GGSx11-GR (KANMX)
YS405	HO-pGAL Chd1-GGSx11-GR (KANMX)
YS406	HO-pGAL Chd1-GGSx11-Res1 (KANMX)
YS407	HO-pGAL Chd1-GGSx11-Res1 (KANMX)
YS409	HO-pGAL Chd1-GGSx11-Ume6 (KANMX); Ume6-SpyFLAG (NATMX); Reb1-SpyFLAG (HPHMX)
YS410	HO-pGAL Chd1-GGSx11-Ume6 (KANMX); Ume6-SpyFLAG (NATMX); Reb1-SpyFLAG (HPHMX)
YS411	HO-pGAL Chd1-GGSx11-SpyCatcher (KANMX); Ume6-SpyFLAG (NATMX); Reb1-SpyFLAG (HPHMX)

Strain ID	Relevant genotype
YS412	HO-pGAL Chd1-GGSx11-SpyCatcher (KANMX); Ume6-SpyFLAG (NATMX); Reb1-SpyFLAG (HPHMX)
YS424	HO-pGAL-Chd1[D513N]-GGSx9-URA3 (KANMX)
YS425	HO-pGAL-Chd1[D513N]-GGSx9-URA3 (KANMX)
YS436	HO-pGAL-Chd1-GGSx9-URA3-SpyFLAG (KANMX) (NATMX)
YS437	HO-pGAL-Chd1-GGSx9-URA3-SpyFLAG (KANMX) (NATMX)
YS533	PV1 (URA3) (KANMX); HO-pGAL-Chd1-Ume6 (HPHMX)
YS534	PV1 (URA3) (KANMX); HO-pGAL-Chd1-Ume6 (HPHMX)
YS535	Dharmacon yTHC Base Cat YSC1210 (KANMX); HO-pGAL-Chd1-Ume6 (HPHMX)
YS536	Dharmacon yTHC Base Cat YSC1210 (KANMX); HO-pGAL-Chd1-Ume6 (HPHMX)
YS576	ARS306::HPHMX
YS577	ARS306::HPHMX
YS655	Dharmacon yTHC Base Cat YSC1210 (KANMX); HO-TetO-3xFLAG-pCyc1-Chd1-Ume6 (HPHMX) (NATMX)
YS656	Dharmacon yTHC Base Cat YSC1210 (KANMX); HO-TetO-3xFLAG-pCyc1-Chd1-Ume6 (HPHMX) (NATMX)
YS657	PV1 (URA3) (KANMX); HO-TetO-3xFLAG-pCyc1-Chd1-Ume6 (HPHMX) (NATMX)
YS658	PV1 (URA3) (KANMX); HO-TetO-3xFLAG-pCyc1-Chd1-Ume6 (HPHMX) (NATMX)
YS659	ARS305::HPHMX
YS660	ARS305::HPHMX
YS661	ARS603.5::HPHMX
YS662	ARS603.5::HPHMX
YS663	ARS1021::HPHMX
YS664	ARS1021::HPHMX
YS741	HO-pGAL-Chd1-GGSx9-Ume6-SpyFLAG (KANMX) (NATMX)
YS742	HO-pGAL-Chd1-GGSx9-Ume6-SpyFLAG (KANMX) (NATMX)
YS743	HO-pGAL-Chd1-GGSx9-Engrailed-SpyFLAG (KANMX) (NATMX)

Strain ID	Relevant genotype
YS744	HO-pGAL-Chd1-GGSx9-Engrailed-SpyFLAG (KANMX) (NATMX)
YS833	PV1 (URA3) (KANMX); ARS306::TRP1
YS834	PV1 (URA3) (KANMX); ARS305::TRP1
YS960	H2A.Z-SNAC-FLAG-Spytag002 (HPHMX)
YS961	H2A.Z-SNAC-FLAG-Spytag002 (HPHMX)
YS962	Ume6-SNAC-FLAG-Spytag002 (HPHMX)
YS963	Ume6-SNAC-FLAG-Spytag002 (HPHMX)
YS964	H2A-SNAC-FLAG-Spytag002 (HPHMX)
YS965	H2A-SNAC-FLAG-Spytag002 (HPHMX)
YS966	ARS305::Trp1
YS967	ARS306::Trp1
YS968	HO-pGAL-Chd1-GGSx9-Ume6-SpyFLAG (KANMX) (NATMX)
YS969	ARS305n (Trp1 counterselection)
YS970	ARS305n (Trp1 counterselection)
YS981	HO-pGAL-Chd1-GGSx9-Ume6 (KANMX); trp1::GPD-OsTIR1-Trp1 (linearized P235)
YS982	PV1 (KANMX) (URA3+); trp1::GPD-OsTIR1-Trp1 (linearized P235)
YS1019	PV1 (KANMX) (URA3+); trp1::GPD-OsTIR1-Trp1 (linearized P235); HO-pGAL-Chd1-Ume6-HO (HPHMX)
YS1020	PV1 (KANMX) (URA3+); trp1::GPD-OsTIR1-Trp1 (linearized P235); HO-pGAL-Chd1-Ume6-HO (HPHMX)
YS1049	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1050	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1051	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1052	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1053	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1054	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1055	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]

Strain ID	Relevant genotype
YS1056	MATa W303 Abf1-3C-FLAG-SpyTag002_hphMX+ P310 [ADH_MNase_GGSx5_SpyCatcher002; URA3+]
YS1063	PV1 (KANMX) (URA3+); trp1::GPD-OsTIR1-Trp1 (linearized P235); HO-pGAL-Chd1-Ume6-IAA7-3XFLAG-HO (NATMX)
YS1064	W303 MATa Ume6::URA3 at N-term; FLAG-KAN at C-term
YS1065	W303 MATa Ume6::URA3 at N-term; FLAG-KAN at C-term
YS1066	W303 MATa Ume6::URA3 at N-term; FLAG-KAN at C-term
YS1067	W303 MATa Itc1-URA (URA *replaces* nt76-210); FLAG-KAN at C-term
YS1068	W303 MATa Itc1-URA (URA *replaces* nt76-210); FLAG-KAN at C-term
YS1069	W303 MATa Itc1-URA (URA *replaces* nt76-210); FLAG-KAN at C-term
YS1079	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO
YS1080	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO
YS1081	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Med14-3C-GGS-SpyTag002-NATMX
YS1082	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Med14-3C-GGS-SpyTag002-NATMX
YS1083	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Med14-3C-GGS-SpyTag002-NATMX
YS1084	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Med14-3C-GGS-SpyTag002-NATMX
YS1085	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Gal4-3C-GGS-SpyTag002-NATMX
YS1086	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Gal4-3C-GGS-SpyTag002-NATMX
YS1087	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Gal4-3C-GGS-SpyTag002-NATMX
YS1088	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Abf1-3C-GGS-SpyTag002-NATMX
YS1089	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Abf1-3C-GGS-SpyTag002-NATMX
YS1090	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Isw2-3C-GGS-SpyTag002-NATMX
YS1091	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Isw2-3C-GGS-SpyTag002-NATMX
YS1092	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Isw2-3C-GGS-SpyTag002-NATMX

Strain ID	Relevant genotype
YS1093	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Isw2-3C-GGS-SpyTag002-NATMX
YS1094	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Reb1-3C-GGS-SpyTag002-NATMX
YS1095	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Reb1-3C-GGS-SpyTag002-NATMX
YS1096	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Reb1-3C-GGS-SpyTag002-NATMX
YS1097	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Reb1-3C-GGS-SpyTag002-NATMX
YS1098	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Ume6-3C-GGS-SpyTag002-NATMX
YS1099	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Ume6-3C-GGS-SpyTag002-NATMX
YS1100	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Ume6-3C-GGS-SpyTag002-NATMX
YS1101	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Abf1-3C-GGS-SpyTag002-NATMX
YS1102	HO_pGAL_MNase_SpyCatcher002_HPHMX_HO; Abf1-3C-GGS-SpyTag002-NATMX

APPENDIX B: PLASMIDS GENERATED FOR SPLIT-CHEC

Plasmid ID	Plasmid contents
P314	HO_pGAL_MNase_SpyCatcher002_hphmx_HO
P315	pFA6a_3C_GGS_SpyTag002_NATMX
P316	pFA6a_3C_GGS_SpyTag002_HIS3MX