UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

Gene regulation and chromatin interactions by the DREAM and E2F transcription factors

Permalink

https://escholarship.org/uc/item/5dm727f1

Author

Asthana, Anushweta

Publication Date

2022

Peer reviewed|Thesis/dissertation

University of California
Santa Cruz

**Gene regulation and chromatin interactions by the DREAM and E2F
transcription factors.**

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Anushweta Asthana

June 2022

This dissertation of Anushweta Asthana is approved:

_____
Professor Seth Rubin, chair

_____
Professor Susan Strome

_____
Professor Michael Stone

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Gene regulation and chromatin interactions by the DREAM and E2F

transcription factors.

Anushweta Asthana

The cell cycle is essential for the growth and development of all multi-cellular organisms and is temporally regulated to coordinate processes such as cell size growth, organelle expansion, DNA replication, expression of mitotic proteins, chromatin condensation, and mitosis. Cell cycle proliferation is tightly regulated by a number of mechanisms including the transcription of G1/S and G2/M genes. In mammals, DREAM and Rb-E2F complexes are considered master regulators of the cell cycle and differentiation yet the molecular mechanisms underlying how these complexes activate and repress genes are still emerging. DREAM and Rb-E2F complexes are promoter dominant transcription factors that bind to target genes near their transcription start sites and modulate their expression. Evidence suggests that these transcription factors interact with chromatin and chromatin-modifying enzymes to alter the chromatin state in promoters. This work explores how these transcription factors engage with chromatin to modulate gene expression.

Acknowledgments

I could not have completed my dissertation research in isolation. I have been extremely fortunate to have had the support of great mentors, a loving family, supportive friends, and a vibrant research community throughout the years. For as far back as I can remember, my parents, Alka and Shivraj, and my older brother Abhishek Asthana, have always taught me to be curious about the world and have encouraged my interest in science. From their lessons, I have not only learned to think analytically but also to listen to good advice, stay open-minded, and surround myself with creative and inquisitive people. I would also like to thank my patient fiancé Shaun Kelsey for encouraging and supporting my journey throughout graduate school. His unwavering confidence in my abilities and his genuine desire to see me achieve my goals have largely made my work possible. I would also like to thank my former manager and great friend Dr. Arunashree Bhamidipati for taking me on and training me as a young college graduate and showing me that I have the capacity to ask big questions.

I am extremely grateful to the wonderful research community here on Science Hill at UCSC which has allowed me to collaborate with and learn from many amazing scientists and students. I would like to thank every rotation student that I have worked with over the last six years for giving me the opportunity to grow as a mentor. In addition, I have had the pleasure of working with two amazing undergraduate researchers Haritha Narasimhan and Akshar Ramkumar who will undoubtedly have a huge impact on medicine and technology. I would like to thank my cohort, the chemistry department, and all the members of the Rubin Lab for their thoughts and advice, and for helping me grow as a better scientist. Finally, I would like to thank my advisor Dr. Seth Rubin for mentoring me throughout my PhD. Seth has always

encouraged me to ask big questions and present my work with confidence. He is genuinely interested in helping his students grow as independent scientists and as compassionate members of the research community. I am very grateful to have had the opportunity to work with him for my dissertation research.

Chapter 1. Introduction

The cell cycle is highly regulated and is comprised of two waves of transcription

The cell cycle is essential for the growth and development of all multi-cellular organisms and is temporally regulated to coordinate processes such as cell size growth, organelle expansion, DNA replication, expression of mitotic proteins, chromatin condensation, and mitosis. Cell cycle-dependent gene expression directs proliferation, cell cycle arrest, and cell cycle exit – processes that are necessary for cellular homeostasis, differentiation, and proper immune function [1]. Two distinct sets of genes exhibit peak expression in either late G1 and S phase (G1/S genes), or in G2 and mitosis (G2/M genes), respectively. G1/S genes largely encode for proteins controlling processes related to DNA synthesis and S phase progression, while the products of G2/M genes regulate mitosis and cytokinesis [2]. The timely synthesis and degradation of these proteins are controlled by several layers of regulation: (I) mRNA transcription in G1/S or G2/M, (II) post-translational modifications, and (III) degradation through ubiquitin-dependent pathways. These tightly concerted processes ensure that cells can enter the cell cycle upon sensing mitogenic stimuli, progress unidirectionally through G1, S, G2, and M, and halt or exit the cell cycle in response to growth-limiting signals [3]. Defects in the networks controlling the cell cycle can have dramatic effects on the survival of cells and organisms. Aberrant expression of cell cycle genes caused by loss of repressors or hyperactivity of activators can stimulate uncontrolled proliferation and can compromise cell cycle arrest or exit [4,5]. Such defects represent an important step in oncogenic

transformation.

Figure 1.1: G1/S and G2/M gene regulation by Rb-E2F and MuvB complexes. In G0 and early G1, the MuvB core binds to the pocket proteins p130/p107, E2F4/5 and DP1/2 proteins forming the DREAM complex. This represses G1/S or G2/M genes which are controlled by *E2F* or *CHR* promoter elements. At this stage, G1/S genes are also repressed by the pocket protein Rb which binds to the E2F family of transcription factors. When a cell receives mitogenic stimuli, D-type cyclins accumulate and form complexes with CDK4/6. These activated CDKs phosphorylate the pocket proteins disassembling Rb-E2F and DREAM which triggers the expression of the G1/S genes including *CDK2, CCNE1/2, MYBL2*. CDK2/Cyclin-E complexes sustain pocket protein phosphorylation in S phase. G2/M genes remain repressed by the MuvB core complex until B-MYB (the gene product of *MYBL2*) and later, FOXM1, bind to the complex and activate G2/M genes.

The transcription of G1/S and G2/M genes mainly depends on four groups of proteins (Figure 1.1) [6,7]: (I) retinoblastoma family pocket proteins (Rb, p107, p130), (II) E2F transcription factors (the activator E2Fs 1-3a, the repressor E2Fs 3b-5, the non-canonical E2Fs 6-8, and the dimerization partners DPs 1/2), (III) proteins forming the MuvB complex (LIN9, LIN37, LIN52, LIN54, and RBAP48), and (IV) the transcriptional activators B-MYB and FOXM1. These proteins are often misregulated

in cancers and are attractive targets for cancer therapeutics [8, 9, 10]. Based on various Chromatin Immunoprecipitation (ChIP) with sequencing and ChIP-on-ChIP experiments, the DNA binding location of these transcription factor complexes have been previously identified. These transcription factors bind to cell cycle-dependent gene promoters via the E2F or LIN54 DNA binding subunits, which recognize the E2F and CHR binding motifs near the transcription start sites of target genes [11, 12, 13].

Repression of G1/S genes by the Rb-E2F complex

Expression of G1/S genes is suppressed by both Rb-E2F and the MuvB containing DREAM (DP, RB-like, E2F, and MuvB) complex. Rb is a tumor-suppressor protein that binds to the E2F family of transcription factors and represses E2F-dependent gene transcription.

There are two well-studied mechanisms by which Rb represses target genes. Under conditions of gene repression in G0 and early G1, Rb binds and inactivates the function of activator E2Fs. By binding to E2Fs, Rb blocks the E2F transactivation domain (TAD) which disables E2F-dependent recruitment of co-activators to target gene promoters [14]. In addition, Rb recruits co-repressors to modify chromatin on target genes. The Rb pocket domain contains an LXCXE binding cleft through which Rb can interact with a large network of proteins. Through this binding site, Rb recruits' factors such as histone deacetylases (HDACs), HDAC-associated SIN3 complexes, as well as chromatin remodelers including the mammalian SWI/SNF complex to repress target genes [15, 16, 17, 18]. Brg1, the catalytic ATPase subunit of mammalian SWI/SNF binds the LXCXE cleft and has been shown to be required for Rb mediated gene repression of G1/S genes in cell culture [16, 19]. Rb-E2F also

forms complexes with the DNA methylase DMNT1 to repress target genes [18, 20]. Viral oncoproteins, including the human papillomavirus E7 protein, the adenoviral protein E1a, and the simian virus 40 T-antigen utilize an LXCXE sequence to bind Rb and inhibit Rb mediated proliferation suppression by antagonizing the recruitment of chromatin remodelers and chromatin modifiers. Indeed, a large majority of human tumors exhibit the phenotype of a defective or inactivated Rb protein. In addition, in some contexts such as permanent cell cycle exit, Rb may recruit histone methylases to cell cycle genes; in terminally differentiated myoblasts and myotubes Rb depletion results in loss of H3K27me2/3 marks at cell cycle genes as well their upregulation [21]. In this context, Rb is thought to be required for the maintenance of H3K27 methylation. The interaction of Rb with Suv39H1, a histone methylase has been reported to be disrupted by the SV40 viral peptide suggesting that Suv39H1 is recruited to Rb via LXCXE binding cleft and may also be linked to Rb functions in the repression of cell cycle genes [22]. Beyond cell cycle control, Rb-E2F complexes may have roles in genomic maintenance such as in response to DNA damage and chromosome stability [23].

Repression of G1/S and G2/M genes by the MuvB containing DREAM complex

DREAM represses both G1/S and G2/M genes in quiescent cells. DREAM binds E2F promoter sites in G1/S genes and CHR promoter elements in G2/M genes. Like Rb, DREAM suppresses proliferation, is necessary for maintaining quiescence and cell cycle arrest, and is subject to inactivation by viral oncoproteins [24, 25]. Both Rb and DREAM cooperate with the tumor suppressor protein p53 to arrest cells under conditions of DNA damage and oxidative stress. In response to genotoxic stressors, p53 upregulates the expression of p21/Cip1 an inhibitor of CDK/cyclin complexes.

This ultimately results in downregulation of cell cycle genes that are controlled by both DREAM and Rb [2]. While the mechanisms involved in Rb-mediated gene repression are well characterized, the DREAM mechanism in gene repression have been largely elusive.

DREAM-mediated gene repression is largely dependent on the MuvB core [26, 27]. In G0 and early G1, the MuvB core complex, which consists of the protein subunits LIN9, LIN37, LIN52, LIN54, and RBAP48/RBBP4, binds to the Rb paralogs p130 or p107 and the canonical repressor E2Fs (E2F4 or E2F5) forming the DREAM repressor complex (Figure 1.1). The roles of the individual MuvB subunits and how the complex assembles have been studied for some time and will be discussed in greater detail in Chapter 2. In brief, the MuvB complex is evolutionarily conserved throughout animals and has been extensively studied in flies and worms where it was first discovered [28, 29, 30]. The LIN52 subunit of the MuvB complex binds to p130 and p107 through the LXCXE resembling LXSXE sequence at the N-terminus of the protein, facilitating the formation of the DREAM complex [31]. In S/G2 LIN52 also binds to the oncoprotein protein B-MYB, forming the MMB complex, which initiates the activation of G2/M genes [32]. LIN54 is the largest MuvB subunit and is predicted to be disordered outside of its C-terminus which contains DNA a binding domain consisting of two tandem cysteine-rich CxC sequences required for CHR binding [11]. LIN52 and LIN54 play an important role in maintaining quiescence as well as promoting proliferation depending on the cell cycle stage. Loss of LIN52 results in sterility in worms, causes larval lethality in flies, and knockdown of this component results in cell cycle arrest in human ESCs [33, 29, 34]. Similarly, loss of LIN54 is linked with sterility in flies and LIN54 knockdown causes G2/M arrest in human ESCs [34].

The functions of the other components LIN37, RBAP48 and LIN9 have remained elusive however they result in compromised cell cycle gene repression and will be discussed in greater detail in Chapter 2.

To date, mass spectrometric and proteomic experiments have not robustly identified co-repressors that bind to DREAM and the complex itself does not have enzymatic activity [35]. In one recent study the SIN3B-HDAC complex was shown to co-repress MuvB target genes in T98G cells; Bainor et al. show SIN3B co-precipitates with DREAM components and knockout of SIN3B moderately depresses DREAM controlled genes [36]. How this complex is recruited to DREAM and whether these results are cell type specific remains unclear. Although p130 and p107 contain the LXCXE binding cleft found in Rb, which recruits utilizes this interface to recruit HDACs, this binding site on p107 and p130 is used instead to bind LIN52 and therefore may not be available to recruit additional factors [31, 33].

One possible hypothesis that explains this complexes' intrinsic repressor activity involves the binding of nucleosomes on gene promoters by the histone binding protein RBAP48, a component of the MuvB core. Recent experiments aimed at uncovering MuvB's repressive mechanism are just now beginning to emerge and will be covered in greater detail in Chapter 2 of this dissertation. Interestingly, the MuvB core is necessary for the expression of G2/M genes by sequentially recruiting the proto-oncoproteins B-MYB and FOXM1 to target promoters [37]. The binding of these activator proteins is thought to alleviate MuvB mediated gene repression and induce gene activation by recruiting additional factors to the promoter.

Activation of the G1/S and G2/M genes

When a cell receives mitogenic stimuli, G1 cyclins are expressed and accumulate to activate cyclin-dependent kinases (CDKs). In G1, the activated CDKs are CDK4/6 which form complexes with Cyclin D and later CDK2/Cyclin E. These CDK/Cyclin complexes phosphorylate the pocket proteins, which disrupts repressor Rb-E2F and DREAM complexes. At this stage, G1/S promoters are induced by activator E2Fs (E2F1-3a) while G2/M genes are thought to remain largely repressed by the intact MuvB core, although loss of pocket protein association has been shown to impede some MuvB repression of G2/M genes [33].

There are three known canonical activator E2Fs: E2F1, E2F2 and E2F3a. Of these activators, E2F1 was the first to be discovered and remains the most well characterized member of the family. While all the members within this family regulate distinct genes and may possess cell-type specific functions [38], they share overlapping functions in gene activation and cell proliferation. All of the canonical E2F members, including the repressor proteins E2F4 and E2F5 contain a C-terminal transactivation domain that is critical for E2F dependent gene expression and cell cycle entry. Activator E2Fs have an extensive network of known binding partners and likely employ several mechanisms to induce gene expression including the acetylation and methylation of histones, recruitment of co-activators to target gene promoters, and interactions with the basal transcription machinery [39,40]. Activator E2Fs bind the histone acetylation machinery including p300/CBP, PCAF/GCN5 as well Tip60 HAT on target gene promoters [41, 14]. In addition, E2Fs are linked with the deposition of the active histone marks such H3K4me3 through cooperation with MLL histone methyltransferase complex [42, 14]. Whether these marks are a

7

prerequisite to or a consequence of activator E2F binding is an open question and remains to be thoroughly explored, although studies in cell culture have shown that E2F1 is required for histone acetylation on target gene promoters [41]. This finding raises the possibility that activator E2Fs could be recruited by or function as a pioneer factor to bind chromatin and facilitate gene activation by enabling interactions with chromatin modifiers [42]. In Chapter 3 of this dissertation, I will present some preliminary data that explores E2F1 and E2F2 interactions with histone proteins on target gene promoters.

Upon expression of *MYBL2*, an S phase gene induced by activator E2Fs, the B-MYB gene product accumulates and binds to the MuvB core forming the MMB complex (B-MYB-MuvB). MMB alleviates MuvB-mediated gene repression and later recruits FOXM1 forming the FOXM1-MuvB activator complex. The MMB complex is required for proliferation and is essential for activation of the mitotic transcription program during G2/M [35, 43, 44, 37, 45, 46]. The specific biochemical activities of B-MYB and MuvB in this complex are still being understood. Like all members of the MYB protein family (A-MYB, B-MYB, C-MYB), B-MYB contains a DNA binding domain (DBD). *In vitro* DNA-binding assays have shown that all MYB proteins can contact DNA via MYB binding sites (MBS), which contain the minimal sequence (C/T)AAC(G/T)G, and overexpression can activate the same reporter constructs [47]. However, the groups of genes regulated by the three members are largely different [48]. Several key amino acid sequences in the MuvB-binding domain of B-MYB are not conserved in A-MYB or C-MYB, and only B-MYB makes a high-affinity interaction with MuvB *in vitro* [32]. B-MYB, but not A-MYB or C-MYB, could be co-immunoprecipitated with LIN9 [44], although proteomic analysis by mass

spectrometry detected minor populations of A-MYB-MuvB complexes [35]. Chromatin immunoprecipitation experiments demonstrate that B-MYB binding to G2/M gene promoters correlates with the presence of the CHR element and that B-MYB association with these genes requires MuvB [12, 37]. Furthermore, the CHR element is sufficient for MMB binding to DNA probes, and evolutionary conserved CHRs are highly enriched in MMB target genes [12]. These data suggest that MuvB recruits B-MYB to CHR sites within cell cycle gene promoters, although some of these promoters do contain B-MYB recognition elements [49]. Conversely, B-MYB knockdown led to loss of MuvB from two G2/M gene promoters [37], which supports a model that binding of MMB to CHR sites may be stabilized by an interaction of B-MYB with MBSs or non-sequence-specific DNA.

Interactions with additional transcription factors have an important role in gene activation by the MMB complex. It was recently shown that the transcriptional co-activator YAP cooperates with MMB to activate a set of mitotic genes [50]. YAP together with TEAD binds to distant enhancer sites and through interaction via DNA looping, increases the stability of the MMB complex on promoters [50]. Furthermore, YAP stimulates the expression of both B-MYB and another critical interaction partner of MMB: the oncogenic transcription factor FOXM1 [50].

FOXM1 binding to MuvB is essential for expression of late cell cycle genes during G2/M [37]. The structure, function, and regulation of FOXM1 is remarkably similar to B-MYB. Both proteins regulate the expression of a similar set of mitotic genes. FOXM1 contains a transactivation domain and a regulatory domain that like B-MYB is activated through CDK and PLK1 phosphorylation [51]. Also similar to B-MYB, FOXM1 contains its own DNA binding domain, yet binding to G2/M gene promoters

highly correlates with CHR sequences and not with canonical forkhead binding sites [37, 52]. These data suggest that like B-MYB, FOXM1 is recruited to promoters through MuvB. However, a subset of G2/M genes showed reduced promoter occupancy and gene activity when FOXM1 mutants with a non-functional DBD were expressed, suggesting that FOXM1-DNA interactions may stabilize FOXM1-MuvB binding to CHR sites in a non-sequence specific manner. Structural details regarding how FOXM1 binds MuvB are not known, and there has been no reported reconstitution of the complex *in vitro* from purified proteins. In one report, FOXM1 binds LIN9 directly [32], but others have found that B-MYB is required for association of FOXM1 with MuvB and binding of all the complex components to DNA [37]. Both B-MYB and FOXM1 are known to bind the CBP/p300 acetylation machinery suggesting that histone acetylation by these co-activators are likely linked to activation of G2/M genes [53, 30].

Conclusions

The transcriptional regulators that drive cell cycle gene expression are localized to the promoter regions of target genes and involve interactions with histones, chromatin modifying enzymes and remodelers near the transcription start sites of target genes. Therefore, a thorough understanding of how these transcription factors assemble and function in the context of chromatin and the underlying mechanisms that facilitate chromatin interactions will inform their function.

In **Chapter 2** of this dissertation, I examine the assembly and repressive function of the MuvB complex and determine the structure of subcomplex that is critical for MuvB function in gene repression. I present both in-vitro and genomic data

exploring how the MuvB complex binds and interacts with chromatin under conditions of cellular arrest and gene repression. In **Chapter 3** of this thesis, I explore the chromatin association of E2F1 and E2F2 in cycling HCT116 cells. I show that both the activator E2Fs can bind histone proteins and they associate with chromatin on promoters of a diverse set of genes.

Acknowledgements

References

1. Laphanuwat, P. & Jirawatnotai, S. Immunomodulatory Roles of Cell Cycle Regulators. *Front. Cell Dev. Biol.* **7**, 1–8 (2019).

2. Fischer M, Grossmann P, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res*, **44**, 6070-6086 (2016)

3. Morgan DO. *The cell cycle: principles of control.* Published by New Science Press in association with Oxford University Press; Distributed inside North America by Sinauer Associates, Publishers: London Sunderland, MA, 2007.

4. Sherr CJ. The Pezcoller lecture: cancer cell cycles revisited. *Cancer Res* 2000; 60: 3689-3695.

5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646-674.

6. Fischer M, Muller GA. Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Crit Rev Biochem Mol Biol* 2017; 52: 638-662.

7. Sadasivam S, DeCaprio JA. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer* 2013; 13: 585-595.

8. Cicirò, Y. & Sala, A. MYB oncoproteins: emerging players and potential therapeutic targets in human cancer. *Oncogenesis* **10**, (2021).

9. Johnson, J., Thijssen, B., McDermott, U. *et al.* Targeting the RB-E2F pathway in breast cancer. *Oncogene* **35,** 4829–4835 (2016).

10. Wiseman EF, Chen X, Han N, Webber A, Ji Z, Sharrocks AD *et al.* Deregulation of the FOXM1 target gene network and its coregulatory partners in oesophageal adenocarcinoma. *Mol Cancer.* **14**, 69 (2015).

11. Marceau, A. H. *et al.* Structural basis for LIN54 recognition of CHR elements in cell cycle-regulated promoters. *Nat. Commun.* **7**, 12301 (2016).

12. Muller GA, Wintsche A, Stangner K, Prohaska SJ, Stadler PF, Engeland K. The CHR site: definition and genome-wide identification of a cell cycle transcriptional element. *Nucleic Acids Res.* **42,** 10331-10350 (2014).

13. Rabinovich, A., Jin, V. X., Rabinovich, R., Xu, X. & Farnham, P. J. E2F in vivo binding specificity: Comparison of consensus versus nonconsensus binding sites. *Genome Res.* **18**, 1763–1777 (2008).

14. Frolov, M. V & Dyson, N. J. Molecular mechanisms of E2F-dependent activation and pRB-mediated repression. *Journal of Cell Science* **117**, 2173–2181 (2004).

15. Dahiya, A., Gavin, M. R., Luo, R. X. & Dean, D. C. Role of the LXCXE Binding Site in Rb Function. *Mol. Cell. Biol.* **20**, 6799–6805 (2000).

16. Dunaief, J. L. *et al.* The Retinoblastoma Protein and BRG1 Form a Complex and Coorperate to Induce Cell Cycle Arrest. **79**, 1–12 (2003).

17. David J Cantor & Gregory David. The potential of targeting Sin3B and its associated complexes for cancer therapy, *Expert Opinion on Therapeutic Targets*, **21**, 1051-1061 (2017)

18. Zhang, H. S. & Dean, D. C. Rb-mediated chromatin structure regulation and transcriptional repression. *Oncogene* **20**, 3134–3138 (2001).

19. Strobeck, M. W. *et al.* BRG-1 is required for RB-mediated cell cycle arrest. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7748–7753 (2000).

20. Robertson, K. D. *et al.* DNMT1 forms a complex with RB, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat. Genet.* **25**, 338–342 (2000).

21. Blais, A., Van Oevelen, C. J. C., Margueron, R., Acosta-Alvear, D. & Dynlacht, B. D. Retinoblastoma tumor suppressor protein-dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. *J. Cell Biol.* **179**, 1399–1412 (2007).

22. Vandel, L., Nicolas, E., Vaute, O. & Ferreira, R. Transcriptional Repression by the Retinoblastoma Protein through the Recruitment of a Histone Methyltransferase. *Mol Cell Biol.* **21**, 6484–6494 (2001).

23. Dick, F.A., Goodrich, D.W., Sage, J. *et al.* Non-canonical functions of the RB protein in cancer. *Nat Rev Cancer* **18,** 442–451 (2018).

24. Fischer M, Uxa S, Stanko C, Magin TM, Engeland K. Human papilloma virus E7 oncoprotein abrogates the p53-p21-DREAM pathway. *Sci Rep* 2017; 7: 2603.

25. Nor Rashid N, Yusof R, Watson RJ. Disruption of repressive p130-DREAM complexes by human papillomavirus 16 E6/E7 oncoproteins is required for cell cycle progression in cervical cancer cells. *J Gen Virol* 2011; 92: 2620-2627.

26. Goetsch, P. D., Garrigues, J. M. & Strome, S. Loss of the Caenorhabditis elegans pocket protein LIN-35 reveals MuvB's innate function as the repressor of DREAM target genes. *PLoS Genet.* **13**, 1–25 (2017).

27. Mages, C.F., Wintsche, A., Bernhart, S.H. & Müller, G.A. The DREAM complex through its subunit Lin37 cooperates with Rb to initiate quiescence. *Elife* **6**(2017).

28. Fay, D. S. & Yochem, J. The SynMuv genes of Caenorhabditis elegans in vulval development and beyond. *Dev. Biol.* **306**, 1–9 (2007).

29. Lewis, P. W. *et al.* Drosophila Lin-52 Acts in Opposition to Repressive Components of the Myb-MuvB / dREAM Complex. **32**, 3218–3227 (2012).

30. Harrison, M. M., Ceol, C. J., Lu, X. & Horvitz, H. R. Some C . elegans class B synthetic multivulva proteins encode a conserved LIN-35 Rb-containing complex distinct from a NuRD-like complex. **103**, (2006).

31. Guiley, K. Z. *et al.* Structural mechanisms of DREAM complex assembly and regulation. *Genes Dev.* **29**, 961–974 (2015).

32. Guiley, K. Z. *et al.* Structural mechanism of Myb–MuvB assembly. *Proc. Natl. Acad. Sci.* **115**, 10016–10021 (2018).

33. Goetsch, P. D., & Strome, S., DREAM Interrupted: Severing LIN-35-MuvB association in *Caenorhabditis elegans* impairs DREAM function but not its chromatin localization, *Genetics* (2022).

34. Wang, C. *et al.* The MuvB complex safeguards embryonic stem cell identity through regulation of the cell cycle machinery. *J. Biol. Chem.* **298**, 101701 (2022).

35. Litovchick, L. *et al.* Evolutionarily Conserved Multisubunit RBL2/p130 and E2F4 Protein Complex Represses Human Cell Cycle-Dependent Genes in Quiescence. *Mol. Cell* **26**, 539–551 (2007).

36. Bainor, A.J. et al. The HDAC-Associated Sin3B Protein Represses DREAM Complex Targets and Cooperates with APC/C to Promote Quiescence. *Cell Rep* **25**, 2797-2807 e8 (2018).

37. Sadasivam S, Duan S, DeCaprio JA. The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev* 2012; 26: 474-489.

38. DeGregori, J. & Johnson, D. G. Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis. *Curr. Mol. Med.* **6**, 739–748 (2012).

39. Tyagi, S., Chabes, A. L., Wysocka, J. & Herr, W. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Mol. Cell* **27**, 107–119 (2007).

40. Pearson, A. & Greenblatt, J. Modular organization of the E2F1 activation domain and its interaction with general transcription factors TBP and TFIIH. *Oncogene* **15**, 2643–2658 (1997).

41. Taubert, S. *et al.* E2F-Dependent Histone Acetylation and Recruitment of the Tip60 Acetyltransferase Complex to Chromatin in Late G 1 . *Mol. Cell. Biol.* **24**, 4546–4556 (2004).

42. Revenko, A. S., Kalashnikova, E. V., Gemo, A. T., Zou, J. X. & Chen, H.-W. Chromatin Loading of E2F-MLL Complex by Cancer-Associated Coregulator ANCCA via Reading a Specific Histone Mark. *Mol. Cell. Biol.* **30**, 5260–5272 (2010).

43. Schmit F, Korenjak M, Mannefeld M, Schmitt K, Franke C, von Eyss B et al. LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. *Cell Cycle,* **6**, 1903-1913 (2007).

44. Pilkinton M, Sandoval R, Colamonici OR. Mammalian Mip/LIN-9 interacts with either the p107, p130/E2F4 repressor complex or B-Myb in a cell cycle-phase-dependent context distinct from the Drosophila dREAM complex. *Oncogene,* **26**, 7535-7543 (2007).

45. Iltzsche F, Simon K, Stopp S, Pattschull G, Francke S, Wolter P *et al.* An important role for Myb-MuvB and its target gene KIF23 in a mouse model of lung adenocarcinoma. *Oncogene* 2017; 36: 110-121.

46. Wolter P, Hanselmann S, Pattschull G, Schruf E, Gaubatz S. Central spindle proteins and mitotic kinesins are direct transcriptional targets of MuvB, B-MYB and FOXM1 in breast cancer cell lines and are potential targets for therapy. *Oncotarget* 2017; 8: 11160-11172.

47. Rushton JJ, Ness SA. The conserved DNA binding domain mediates similar regulatory interactions for A-Myb, B-Myb, and c-Myb transcription factors. *Blood Cells Mol Dis*, **27**, 459-463 (2001).

48. Rushton JJ, Davis LM, Lei W, Mo X, Leutz A, Ness SA. Distinct changes in gene expression induced by A-Myb, B-Myb and c-Myb proteins. *Oncogene* , **22**, 308-313 (2003).

49. Knight AS, Notaridou M, Watson RJ. A Lin-9 complex is recruited by B-Myb to activate transcription of G2/M genes in undifferentiated embryonal carcinoma cells. *Oncogene*. **28**, 1737-1747 (2009).

50. Pattschull, G. *et al.* The Myb-MuvB Complex Is Required for YAP-Dependent Transcription of Mitotic Genes. *Cell Rep.* **27**, 3533-3546.e7 (2019).

51. Marceau AH, Brison CM, Nerli S, Arsenault HE, McShan AC, Chen E *et al.* An order-to-disorder structural switch activates the FoxM1 transcription factor. *Elife.* (2019)

52. Chen X, Muller GA, Quaas M, Fischer M, Han N, Stutchbury B *et al.* The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. *Mol Cell Biol*, **33**, 227-236 (2013).

53. Major, M. L., Lepe, R. & Costa, R. H. Forkhead Box M1B Transcriptional Activity Requires Binding of Cdk-Cyclin Complexes for Phosphorylation-Dependent Recruitment of p300/CBP Coactivators. *Mol. Cell. Biol.* **24**, 2649–2661 (2004).

Chapter 2. Structure and chromatin association of MuvB

Introduction

Chromatin architecture and the position of nucleosomes influence DNA-mediated processes including the transcription of genes[1]. Transcription by RNA polymerase results in significant changes to nucleosome positioning, as the basal transcription machinery must overcome the energetic barriers presented by the placement of nucleosomes along promoters and the gene body[2-4]. RNA polymerase with the aid of elongation factors and histone chaperones can bind and evict octamer proteins or reposition nucleosomes present in the gene body to access the gene for transcription. In addition, chromatin remodelers and histone modifying enzymes are thought to facilitate or inhibit transcription by arranging or displacing nucleosomes near transcription start sites, by altering the packing of nucleosomes, and by modulating the affinity of histone proteins for the DNA backbone. Less is known about how the nucleosomal architecture is influenced by the activity of transcription factors (TFs). While recent evidence shows that pioneer TFs can bind target DNA sites within the nucleosome wrap and recruit remodelers to alter chromatin architecture, other TFs compete with nucleosomes for access to their DNA consensus sequence[5-7]. A thorough molecular description of how many regulatory TFs cooperate and engage with nucleosomes to modulate gene expression remains elusive.

The MuvB TF complex binds to target gene promoters and regulates a large set of cell cycle genes. MuvB temporally coordinates the expression of genes necessary for DNA synthesis, centromere construction, mitotic division, and cell cycle exit[8-10]. In mammals, cell cycle-dependent gene expression occurs primarily in two

waves of transcription which take place around the G1/S and G2/M transitions and depend on the activity of MuvB and the other TFs E2F, B-MYB, and FOXM1[11-15]. These TFs and their regulators are commonly deregulated in cancer[16-18].

The MuvB complex, components of which are evolutionarily conserved throughout animals and ciliates, plays a key role in development and differentiation and is an essential regulator of cell cycle-dependent gene expression programs[9,10,19-23]. During quiescence and in early G1, MuvB binds to the retinoblastoma protein (RB) paralogs p130 or p107 (p130/p107) and E2F4-DP. This complex, known as DREAM, represses S phase genes and late cell cycle genes[22,24,25]. Upon entry into the cell cycle, cyclin-dependent kinases along with their cyclin partners phosphorylate and release p130/p107 from the MuvB core, disassembling DREAM but keeping the core MuvB intact[22,25-27]. During S phase, the MuvB core binds to the onco-protein B-MYB and forms the MYB-MuvB (MMB) complex, which in concert with FOXM1 functions as a transcriptional activator of G2/M genes[22,23,28]. While the cellular imbalance of activating and repressive MuvB complexes is associated with several cancers[29,30], the molecular details of MuvB assembly and function are poorly understood.

The core MuvB complex is composed of the five proteins LIN9, LIN37, LIN52, LIN54, and RBAP48 (or RBBP4). MuvB is localized to its target cell cycle genes through LIN54, which binds target promoters directly at a consensus DNA sequence[31-33]. The short sequence motif, known as the cell cycle genes homology region (CHR), is found in close proximity to the transcription start site (TSS) and is often located just downstream of a truncated E2F binding site, known as the cell cycle-dependent element (CDE)[34]. LIN52 is a transcription factor adaptor protein that recruits either B-MYB or p130/p107, depending on cell cycle phase[35,36]. RBAP48 is a

histone binding chaperone protein that is found in several complexes that interact with chromatin, including CAF-1, NuRD, PRC2, and SIN3-HDAC[37-40]. In mammals, RBAP48 has a highly similar (89% sequence identity) paralog named RBAP46 (or RBBP7), which has not been identified in complexes with MuvB components[22]. Both proteins are found in chromatin remodeler complexes, sometimes together. Less is known regarding the structure and biochemical function of LIN9 and LIN37, although a LIN9 sequence near its C-terminus co-folds with LIN52 to create the B-MYB binding site[35].

Genetic evidence suggests that MuvB core proteins are essential in regulating cell cycle-dependent gene expression. In flies and worms, knockout of MuvB components contributes to inappropriate derepression of developmental gene programs[19-21]. In mammals, LIN9 is essential for the expression of G2/M genes; loss of LIN9 causes mitotic defects and is embryonically lethal in mice[41,42]. On the other hand, knockdown of LIN9 in cell culture results in compromised repression of DREAM target genes upon induced cell cycle exit[22]. Knockout of the MuvB subunit LIN37 results in loss of MuvB-mediated gene repression in G0 and G1, but it does not lead to any observable changes in MYB-MuvB (MMB) mediated gene expression in G2/M[24]. Similarly, RNAi depletion of the *Drosophila* ortholog of RBAP48 specifically results in a derepression of dE2F2 target genes but does not result in defects in proliferation or gene expression[43]. These findings implicate MuvB core subunits in both positively and negatively modulating gene-expression, yet the biochemical mechanism behind their function remains unknown.

Here we investigated how MuvB represses gene expression, with emphasis on characterizing the structure and function of LIN9, LIN37, and RBAP48. We

demonstrate that LIN9 and LIN37 together form an essential scaffold that holds together the core complex, and we determined a crystal structure that reveals how they together recruit RBAP48. We show that through RBAP48, MuvB binds directly to nucleosomes, either by interacting with H3 tails or the core particle. Using single-molecule electron microscopy, we found that MuvB increases nucleosome occupancy in a reconstituted cell cycle gene promoter. These data indicate that MuvB associates with and stabilizes nucleosomes in the absence of other factors. Finally, we implemented a protocol that applies micrococcal nuclease digestion of chromatin and co-precipitation (MNase-ChIP) to study interactions of MuvB with nucleosomes in HCT116 cells. Our results support a model that MuvB binds to nucleosomes near the transcription start sites of target genes and stabilizes nucleosomes to repress cell cycle-dependent gene expression.

LIN9 and LIN37 are together required for assembly of MuvB

Beyond the role of LIN9 in binding B-MYB, the structure and biochemical function of LIN9 and LIN37 have not been previously characterized. Human LIN37 is a 246 amino acid protein that has no homology to any known structures. Sequence analysis suggests the presence of several short, structured regions (1-43, 95-126, 203-246) that are interspersed with sequences that are likely disordered (Figure 2.2.1a). The segment 95-126, which we call the CRAW domain for the presence of a CRAW amino acid sequence, is highly conserved among animal orthologs and is necessary for LIN37 assembly into MuvB and for its activity in gene repression[24]. Human LIN9 contains 558 amino acids, and beyond the presence of a Tudor domain, it also exhibits no homology to known structures (Figure 2.1a). The N-terminal ~95 amino acids of LIN9 are poorly conserved and have no predicted structure. The next

segment from 94-278 (previously called the domain in RB-related pathway or DIRP; Pfam 06584) contains the Tudor domain and is conserved between MuvB and the related tMAC complex[44]. The helical segment between 349-466 forms the MYB-binding domain (MBD) together with LIN52[35], while the C-terminus (residues 473-538) also has predicted helical structure.

Considering previous observations that LIN9 binds directly to multiple core MuvB and MuvB-interacting proteins[23,27,35] and that LIN9 knockdown results in DREAM complex assembly defects in T98G cells[22], we hypothesized that LIN9 is a scaffold onto which the other proteins assemble (Figure 2.1b). To probe MuvB complex assembly in a reconstituted system, we performed co-precipitation experiments by expressing human proteins with different affinity tags in Sf9 insect cells (Figure 2.1c). We expressed full-length RBAP48, LIN52, and LIN37 and the relatively conserved and structured regions of LIN9 (residues 92-538, called LIN9[92-538]) and LIN54 (residues 504-749, LIN54[504-749]). When the three MuvB components RBAP48, LIN52, and LIN54 were co-expressed, we did not see co-precipitation (Figure 2.1c, lanes 1-3, 6). In contrast, we were able to reconstitute the MuvB complex when all five components were co-expressed (Figure 2.1c, lanes 4 and 7), and we could demonstrate co-elution as a single complex by performing successive precipitations of different affinity tags (Figure 2.1c, lane 8). In our baculovirus system, we were unable to express LIN9 in the absence of LIN37, so we could not test whether LIN9 alone is required in our reconstitution. However, it has previously been reported that DREAM and MuvB complexes are able to assemble in the absence of LIN37[19,24,45]. Taken together, these results suggest that the LIN9 subunit of MuvB coordinates RBAP48, LIN52 and LIN54 to assemble the complex.

To further probe how LIN9 interactions with the other MuvB subunits organize the overall architecture of the complex, we expressed Flag-tagged LIN9 constructs in HCT116 cells and analyzed binding by co-immunoprecipitation (Figure 2.2.1d). We observed differences in the interactions made by LIN9$^{1-300}$, which contains the DIRP domain, and the interactions made by LIN9$^{300-558}$, which contains the MYB-binding domain and C-terminus. Only LIN9$^{300-558}$ co-precipitated p130. This observation is consistent with the known direct association of LIN9^{MBD} with LIN52 and the direct association of the LIN52 N-terminus with p130[26,35]. LIN9$^{300-558}$ also associates with LIN54, whereas LIN9$^{1-300}$ does not immunoprecipitate LIN54 above background in our experiment. In contrast, only LIN9$^{1-300}$ co-precipitated RBAP48 and LIN37. We conclude that the LIN9 N-terminus is necessary and sufficient for binding RBAP48 and LIN37, while the C-terminus binds LIN52 and LIN54 (Figure 2.1b). We found that co-expression of RBAP48 with the DIRP region of LIN9 (LIN9$^{94-274}$) and the conserved CRAW domain of LIN37 (LIN37$^{92-130}$) in Sf9 cells yielded a MuvB subcomplex that was stable through affinity purification and size-exclusion chromatography (Supplementary Figure 2.1). We call this subcomplex MuvBN, as it contains sequences toward the N-termini of LIN9 and LIN37.

Figure 2.1: LIN9 and LIN37 scaffold the MuvB complex. (A) Domain architecture of LIN9 and LIN37, with regions of predicted and validated structure shown as blocks. The conserved LIN37 CRAW domain and LIN9 DIRP domain structures are determined here. MBD is the Myb-binding domain. (B) Schematic model for subunit interactions within MuvB. (C) The indicated tagged subunits or MuvB complex (GST-LIN52, Strep-RBAP48, His-LIN54$^{504-749}$, GST-LIN37, and GST-LIN9$^{92-538}$) were expressed in Sf9 cells and extracts were precipitated with resin capturing the indicated tag. Proteins were visualized with coomassie staining. *Indicates impurities or degradation observed in some RBAP48 expressions. These bands are not pulled out from the tandem purification. (D) HCT116 cells were transfected with plasmids encoding the indicated FLAG-tagged protein. FLAG-tagged proteins were precipitated from extracts using anti-FLAG antibody and visualized with anti-FLAG immunoblotting and immunoblotting with antibodies that recognize RBAP48, LIN37, LIN54, and p130.

Overall structure of LIN9-LIN37-RBAP48 subcomplex

We were able to crystalize the MuvBN subcomplex, and we determined the structure to 2.55 Å by molecular replacement using the known RBAP48 subunit structure as an initial model (PDB: 3GFC) (Supplementary Table 1)[46]. The crystal structure contained one complex in the asymmetric unit, and we built the LIN9 and LIN37 fragments into the unmodeled electron density. The final refined MuvBN model contains one copy of each protein (Figure 2.2). As previously described, RBAP48 has a b-propeller domain fold, consisting of seven small α-sheets, along with a single N-terminal helix[46-48]. The atomic structure of RBAP48 in MuvBN aligns well with other structures of the protein in other complexes with RMSDs ~0.3-0.6 Å (Supplementary Figure 2.2). The LIN9[94-274] sequence is almost entirely visible in the electron density and contains six alpha helices and the Tudor domain. The helices are N-terminal to the Tudor domain and do not appear to form a globular structure. Instead, they wrap around and from extensive contacts with RBAP48, create a binding site for LIN37, and anchor the Tudor domain to the rest of the complex. The LIN37[92-130] CRAW domain is also nearly all visible in the electron density. This continuous LIN37 sequence forms two small β-strands and a short a helix. Our recombinant LIN9 was unstable without co-expression of this highly conserved fragment of LIN37[92-130], which interacts with both RBAP48 and LIN9 in the subcomplex.

Structure of the LIN9-RBAP48 interface

LIN9 and RBAP48 associate across a broad interface focused around the N-terminal helix (αN) of RBAP48 and the adjacent side of the β-propeller domain (Figure

2.2). All six LIN9 helices contact RBAP48, and five of them (α2-α6) surround and make interactions with the RBAP48 αN helix (Figure 2.2a,b). Numerous hydrophobic, polar, and electrostatic contacts are observable between the proteins (Figure 2.2b), and we highlight a few specific examples here that are relevant for the mutagenesis experiments described below. For example, a cluster of two arginines (R174 and R175) and two phenylalanines (F180 and F181) in LIN9 anchor α6 and the preceding loop against the RBAP48 αN helix and so-called PP-loop, which is an insertion in the sixth propeller β-sheet (Figure 2.2c). The sidechains of R174 and R175 make a series of electrostatic interactions with side chain and main chain atoms in RBAP48 residues Q354, D358, P361, and G362, while F180 and F181 pack against RBAP48 residues I23 and W24.

Figure 2.2: Structure of the MuvBN subcomplex. (A) Overall structural model. (B) Alignment of LIN9 sequences from *H. sapiens*, *C. japonica*, *D. rerio*, *D. melanogaster*, *S. purpuratus*, and *C. intestinalis*. The * marks residues that contact RBAP48, the . marks residues that contact LIN37, and the @ marks residues that contact both. (C) Close-up view of one interface between LIN9 and RBAP48. (D) Location of histone H3 and histone H4 peptide binding sites on RBAP48. When bound to LIN9, the H4 sites is blocked while the H3 site is mostly accessible. The model was generated from PDB IDs: 2YBA and 3CFV.

Structures of RBAP48 and RBAP46 bound with various peptides depict how they are assembled into diverse complexes. A survey of known structures reveals two common peptide binding sites on the b-propeller domains (Figure 2.2D and Supplementary Figure 2.2). One site is across the face of the β-propeller and is found

26

occupied by histone H3, Fog1, and PHF6. The second site is along the side of the propeller between αN and the PP loop; it is found occupied by histone H4, Mta1, and Suz12. In the MuvBN structure, the H3 site is for the most part accessible, although the α1 and α2 helices of LIN9 pack against the edge of the propeller where the H3 site-binding peptides exit the propeller face. In contrast, the H4 site is bound by the sequence in LIN9 between α5 and α6 and is not accessible in the MuvBN complex. It was recently reported that the proliferating cell nuclear antigen (PCNA)-associated factor (PAF) binds RBAP48 through a sequence in RBAP48 (residues 346-352) that in our structure is near the H4 site but somewhat solvent exposed[30]. It Is feasible that PAF could access this extruded part of RBAP48 in the MuvB complex; however, how PAF binding to RBAP48 competes with p130 binding, as suggested[30], is unclear considering our result that MuvBN components are not required for p130 association (Figure 2.1d). Several structures of RBAP48 in complex with one or more proteins or larger protein fragments have also been previously determined. For example, RBAP48 is present in the polycomb complex PRC2[37,49]. As observed in MuvBN, RBAP48 is bound in these other complexes at multiple sites and on both sides of αN. One striking difference in how LIN9 and LIN37 bind RBAP48 compared to how proteins bind in other complexes is the extensive interactions with a glycine rich loop in RBAP48 (residues 88-115) (Figure 2.3a). This RBAP48 loop, which is an insertion between two strands in the first complete propeller blade, is disordered in almost all the structures with peptides and partially ordered when binding Mta1 or the polycomb complex protein Suz12 (Supplementary Figure 2.2). In contrast, the interactions of the insertion loop with LIN9 and LIN37 are much more extensive, and the entire loop appears ordered in the MuvBN structure. With respect to histone

Figure 2.3: LIN37 CRAW domain binds both LIN9 and RBAP48. (A) Interactions of LIN9 and LIN37 with the RBAP48 insertion loop. (B) Alignment of LIN37 sequences from organisms as in Figure 2.2B. Residues that contact LIN9 (.), RBAP48 (*), and both LIN9 and RBAP48 (@) are indicated. (C) Close-up view of the LIN9-LIN37 interface. (D) The indicated FLAG-GFP control and FLAG-LIN9 WT and mutants were expressed by transient transfection in arrested HCT116 cells. Proteins were immunoprecipitated using anti-FLAG beads and the indicated proteins visualized by Western blot. The 3x LIN9 mutant is E125A/W126A/F127A and the 4x LIN9 mutant is R174A/R175A/F180A/F181A. (E) Same as (D) but expressing the indicated RBAP46 and RBAP48 WT and mutant proteins.

binding, the H4 binding site in RBAP48 is occluded and the H3 site is more accessible

in both the PRC2 and MuvBN complexes.

The LIN9 Tudor domain has a non-canonical aromatic cage

LIN9 additionally contains a conserved Tudor domain that is visible in the subcomplex (residues 223-273). Tudor domains are protein interaction modules that are found in many chromatin-binding proteins. In several cases, they recognize methylated lysines and arginines and function as readers of modified histones[50-52]. The Tudor structure is defined by five anti-parallel β-strands that fold into a barrel. Target peptides are bound by an aromatic cage at one end of the barrel. The cage typically surrounds the modified basic sidechain and makes stabilizing p-cation interactions. We aligned the LIN9 Tudor domain with structures of the PHF1 (PDB: 2M0O, RMSD 1.0 Å) and the SMN (PDB: 4A4E, RMSD 0.9Å) Tudor domains in complex with their target peptides (Figure 2.4)[50,52]. The alignments suggest that the LIN9 cage contains fewer aromatics and is relatively inaccessible, as it makes an interaction with a loop that adjoins β3 and β4 at L261 (Figure 2.4). We note that we have not been able to detect binding of the LIN9 Tudor domain to several unmodified and modified histone peptides or modified lysine and arginine at high concentrations. While we do not rule out the possibility that the LIN9 Tudor domain binds histone or other proteins, we conclude that the structural features of the cage that mediate the interactions of other Tudor domains are not obviously present in LIN9.

LIN37 structure and interface with LIN9 and RBAP48

Previous functional domain mapping studies demonstrated that two highly conserved sequences in LIN37 were critical for LIN37 binding to other DREAM components and for DREAM repression of cell cycle genes[24]. These sequences in LIN37 correspond with the CRAW domain of LIN37 that appears structured in our crystals of MuvBN, and they play a critical role in interacting with LIN9 and RBAP48.

This observation firmly implicates the MuvBN subcomplex as the structural subunit of DREAM responsible for gene repression.

Figure 2.4: Structure of the LIN9 Tudor domain and comparison to other Tudor domains.
(A) Alignment of LIN9 Tudor domain with PHF1, which binds a trimethylated peptide in histone 3 (H3K36Me3), and SMN, which recognizes a dimethylated arginine. Pairwise RMSD for C_α atoms are 1.0 Å between LIN9 and PHF1 and 0.9 Å between LIN9 and SMN. (B) View of LIN9 residues L230, F238, F256 and H264, which correspond to residues in the other structures that form the aromatic cage. The key modified interacting residue with the PHF1 and SMN structures are modeled in the view. The "cage" in LIN9 contains fewer aromatic residues (alignment shown in (C)), and the position of the interacting residue is occupied by L261 in LIN9. We conclude from the structure that the Tudor domain of LIN9 is non-canonical, and it may not interact with modified histone species, or it may interact in a different manner as previously observed.

The small structured LIN37 CRAW domain is bound between LIN9 α1 and α2 (Figure 2.3 A-C). The two LIN9 helices form a V-shape that straddles one face of the LIN37 structure. Sidechains along one hydrophobic face of the LIN9 α1 helix (I104, L108, L111, and L112) are inserted into a groove formed by hydrophobic residues from all the LIN37 secondary structure elements (I97, L99, F100, V104, L106, F109, L115, I118, and W122). The LIN9 α2 helix binds the opposite face of the LIN37 helix

from the LIN9 α1 helix with LIN9 W125 packing against the LIN37 backbone and interacting with LIN37 Y116. The LIN37 helix forms the primary interface between LIN37 and RBAP48 (Figure 2.3A). Y116 and R120, both of which are highly conserved among LIN37 orthologs, make several interactions with the glycine rich insertion loop in RBAP48. The nearby LIN9 α2 helix also contributes to this interface such that E124 from LIN9, Y116 from LIN37, and Y98 from RBAP48 all interact through a network of hydrogen bonds.

We tested the importance of several interface contacts observed in the structure on assembly of MuvB in HCT116 cells (Figure 2.3D). We expressed either FLAG-tagged WT LIN9 or two FLAG-tagged LIN9 mutants and performed anti-FLAG immunoprecipitation to assay association with other MuvB proteins. A triple mutant (E125A/W126A/F127A) that contains mutations in the LIN9 α1 helix (LIN9[3X]) failed to co-precipitate LIN37, whereas a quadruple mutant (R174A/R175A/F180A/F181A, LIN9[4X]) with mutations in LIN9 α6 and the preceding linker (Figure 2.2c) failed to co-precipitate both LIN37 and RBAP48. It is notable that LIN37 was lost in the LIN9[4X] co-precipitation even though the mutated residues are not directly at the LIN37 interface. These results indicate that despite the extensive interface, RBAP48 association with LIN9 can be disrupted through a few key mutations. The results of the LIN9[4x] mutant experiment also suggest that LIN37 association with LIN9 is likely stabilized by the presence of RBAP48 in the complex.

Analysis of the interactions at the LIN9-LIN37-RBAP48 interface reveals the structural mechanism for the specificity of RBAP48 in the MuvB complex. Previous analysis of MuvB components using mass spectrometry did not identify the presence of the RBAP48 paralog RBAP46[22]. In our co-immunoprecipitation experiments, we

also did not observe association of RBAP46 with components of the complex (Figure 2.3E). The two human homologs are 89% identical, but notably RBAP46 contains a cysteine at position Y98 in RBAP48. In the MuvBN structure, Y98 is in the RBAP48 insertion loop and is involved in a network of hydrogen bonds at the interface with both LIN9 and LIN37 (Figure 2.3a). We found that while Flag-tagged wild-type mouse RBAP48 could co-precipitate MuvB components in HCT116 cells extracts, mouse RBAP48 with an RBAP46-mimicking Y98C mutation does not co-precipitate MuvB components (Figure 2.3E). Conversely, a mouse RBAP48-mimicking C97Y mutation in mouse RBAP46 results in some additional affinity, although we note association still appears weaker than with WT RBAP48. We conclude that the MuvB complex has specificity for RBAP48 and that this specificity arises through this unique insertion loop association with LIN9 and LIN37.

MuvB binds histone H3 tails and reconstituted nucleosomes lacking a CHR

The MuvB complex contains two domains that have potential histone binding properties: the Tudor domain of LIN9 and the b-propeller domain of RBAP48. We wanted to test whether these domains, within the context of MuvB, are able to engage with histone peptides and nucleosomes. We first tested whether our recombinant purified MuvB complexes bind histone peptides that are known to form complexes with RBAP48. We tested binding of both MuvB (Figure 2.5a) and the MuvBN subcomplex (Supplementary Figure 2.2a) to fluorescein-labeled H3 (1-21) and H4 (21-41) peptides by fluorescence polarization. We found that MuvB and MuvBN bound the H3 tail but that they did not bind the H4 peptide. This observation is consistent with the MuvBN structure, which shows that the H3 site in RBAP48 is accessible while the H4 site is occluded by LIN9 (Figure 2.2d). Using the fluorescence

Figure 2.5: MuvB binds to histone peptides and nucleosomes. Fluoresence polarization measurements of association between recombinant MuvB and dye-labeled histone peptides. (A) or nucleosomes (B). FP is normalized to dye-labeled probe alone. Data are fit assuming a single binding constant in each case. (C) Average affinities from three replicates.
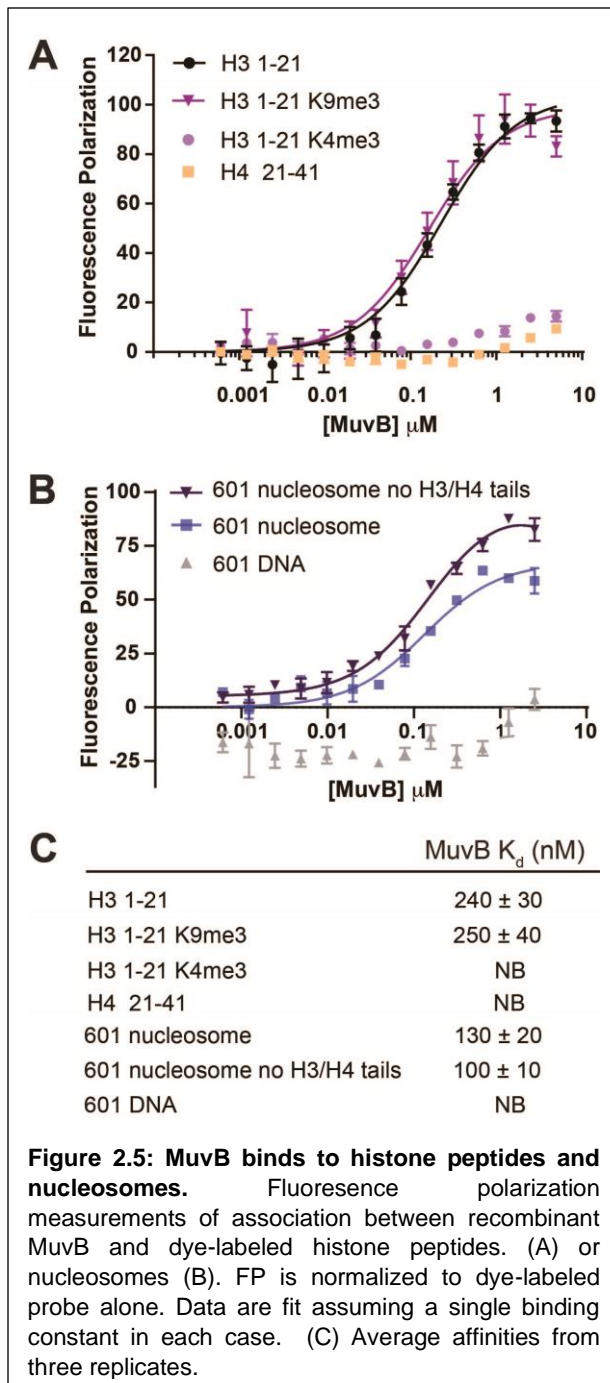
polarization assay, we found that MuvBN binds H3 peptide with similar but slightly weaker affinity as the full MuvB complex, suggesting that the MuvBN complex is sufficient to make the most significant contacts with the H3 peptide (Supplementary Figure 2.3a). Isothermal titration calorimetry measurements also demonstrate binding of MuvBN to H3 but not H4 tails and suggest that H3 binding is mediated through RBAP48 as previously described (Supplementary Figure 2.3b) [47-48].

To probe whether post-translational modifications on H3 tails influence MuvB binding, we tested two H3 marks that are associated with active transcription (H3K4me) or transcriptionally silent heterochromatin (H3K9me3). We found that MuvB bound H3 tails when methylated at K9 but failed to bind with H3 tails methylated at K4 (Figure 2.5a). This result is consistent with available structural data demonstrating that K4 methylation inhibits H3

tail binding to RBAP48[47]. We note that our observations contrast with experiments performed with purified *Drosophila* dREAM complex, which appeared to bind non-acetylated H4 peptides[20]. However, the Drosophila complex contains additional histone-interacting proteins (L3MBT and an HDAC ortholog) not present in the mammalian complex.

We then asked whether MuvB could bind reconstituted nucleosomes and whether nucleosome binding was conferred by H3 tails alone. We reconstituted nucleosomes with full-length histones and the Widom 601 strong positioning sequence containing a fluorescein label. MuvB bound to these nucleosomes with slightly greater affinity than to the tails (Figure 2.5b,c). The 601 DNA sequence lacks a CHR sequence, and we found that MuvB did not bind fluorescein-labeled free 601 DNA, indicating that nucleosome association occurs independently of DNA consensus motif binding. In the FP assay, we found that MuvB binds to the nucleosomes lacking histone H3 and H4 N-terminal tails (H3:39-136; H4:19-103) with a similar affinity compared to nucleosomes with tails (Figure 2.5b). This observation is consistent with a known association of RBAP48 with tailless histone H3-H4 dimers, although we detect here association in the context of a reconstituted nucleosome[53]. Our data indicate that MuvB can bind nucleosomes through the H3 tails but that H3-tail binding is not necessary for MuvB-nucleosome association. To rule out any potential binding of the LIN9 Tudor domain, we reconstituted a mutant MuvB complex harboring LIN9 Tudor aromatic cage mutations (L230A/F238A/F256A/H264A) and found this mutant engages with Widom nucleosomes similar to the wild-type complex in the FP assay (Supplementary Figure 2.3c). This result suggests that the LIN9 aromatic cage is not necessary for binding nucleosomes. Considering these results

together, we propose that MuvB engages with H3 tails and/or the folded octamer to bind nucleosomes and that this association is primarily mediated by the MuvBN subcomplex including RBAP48.

MuvB binds and stabilizes nucleosome occupancy on a reconstituted and chromatinized cell cycle gene promoter

We previously analyzed late cell cycle genes in available ENCODE data sets and found that DREAM target gene promoters show a higher nucleosome density within the few hundred bases downstream from the transcription start site relative to genes that lack a CHR site and relative to constitutively expressed genes[31]. Following our observation here that MuvB binds nucleosomes in the absence of additional factors, we tested whether MuvB directly increases nucleosome occupancy on cell cycle gene promoters. We cloned and amplified a minimal promoter from the human *TTK* gene, which is a late cell cycle gene regulated by MuvB[54]. We folded a purified TTK-derived 461bp DNA fragment with recombinant histone octamer in the presence and absence of MuvB. The DNA contains a single CHR located 187 bp from the 5' end (Figure 2.5a and Supplementary Figure 2.4a) An electromobility shift assay demonstrated that MuvB was able to associate with the chromatinized promoter (Figure 2.6a).

We then cross-linked our chromatinized samples and several control samples with trimethyl psoralen, digested protein, and performed metal-shadowing electron microscopy to assess nucleosome occupancy along the DNA molecules across our conditions (Figure 2.6b,c)[55,56]. In these experiments, the presence of nucleosomes is inferred from the appearance of nucleosome-sized bubbles in the micrograph

(Supplementary Figure 2.4b). As expected, we observed nucleosomes in the samples folded with histone octamers prior to cross-linking but not in the sample that only contained free TTK DNA. When MuvB is present in the reconstitution reaction, we observe more molecules containing nucleosomes and an increase in the average number of nucleosomes per molecule (Figure 2.6c). Furthermore, the distribution of inferred nucleosomes titrated with MuvB concentration. When a lower concentration of MuvB was present, we observed fewer nucleosomes per molecule relative to the high-concentration condition. We conclude that MuvB stabilizes nucleosomes in the synthetic TTK promoter, as MuvB increased nucleosome occupancy in the equilibrium established by the reconstitution reaction. We did not observe a significant change in the nucleosome distribution with the inclusion of LIN54[504-709] alone, suggesting that binding of the CHR by the LIN54 DBD is not sufficient to increase nucleosome occupancy. When we mutated the CHR site in the TTK promoter, we observed a significant decrease in the average nucleosomes per molecule (Figure 2.5c), which is consistent with weaker MuvB binding to the mutated CHR site in the DNA (Supplemental Figure 2.4c). However, this average is still greater than the average in the absence of MuvB. We propose that MuvB binds and stabilizes nucleosome occupancy in the DNA even when LIN54 is not bound to the CHR site (i.e. in *trans* association with nucleosomes) but that simultaneous engagement of both the CHR and nucleosome (i.e. in *cis* association) results in increased stability. A histone chaperone-like activity has been reported for RBAP48 in other chromatin-bound complexes, and RBAP48 binds histone octamer intermediates[38,53]. Because our experiment probes the equilibrium established by the reconstitution reaction beginning with a folded octamer, we cannot rule out a role for MuvB in facilitating the

assembly of nucleosome intermediates (Supplementary Figure 2.5b). However, we favor the interpretation that, by binding the CHR and the nucleosome (Figure 2.4b), MuvB stabilizes fully assembled nucleosomes in the promoter.
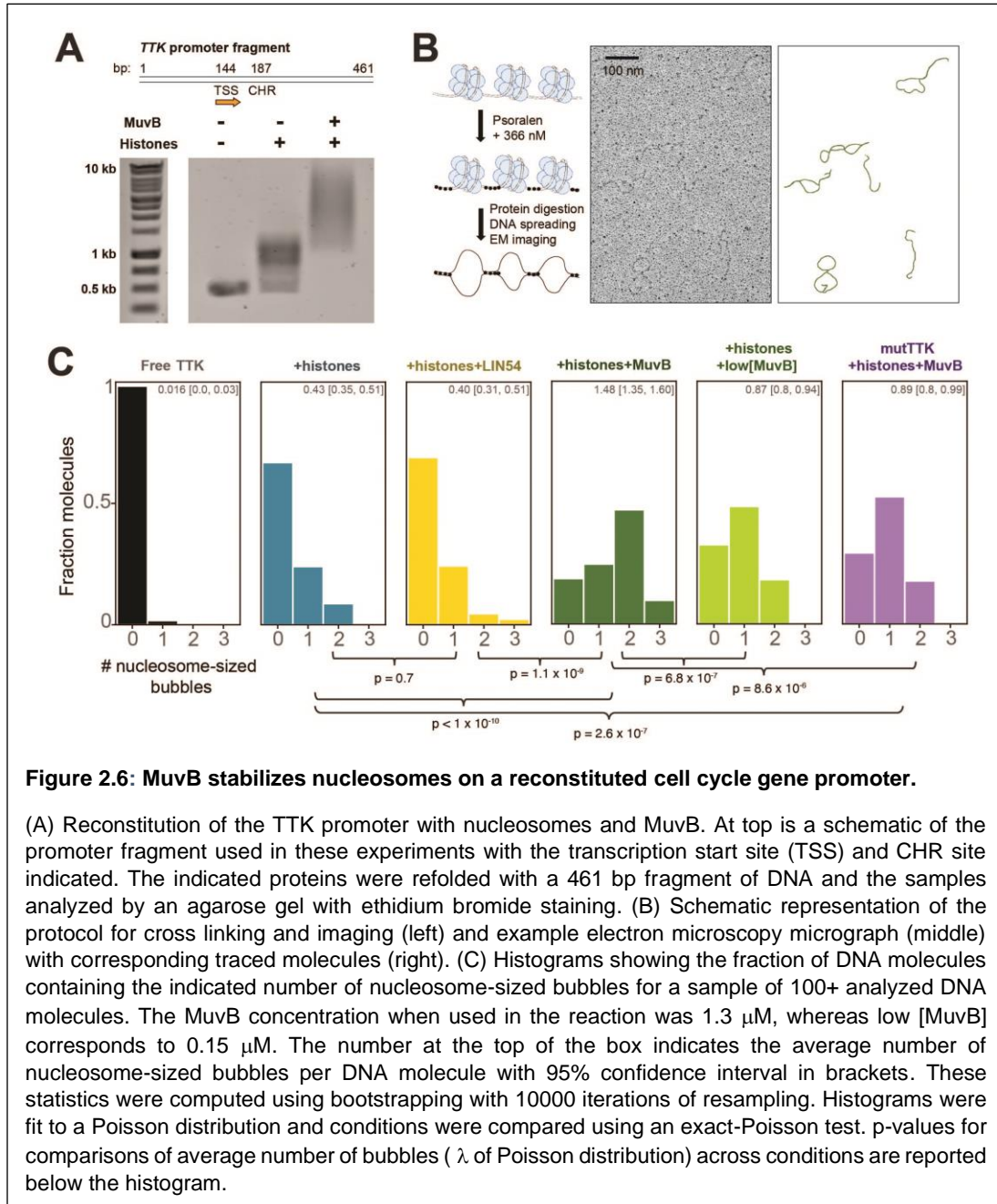
Figure 2.6: MuvB stabilizes nucleosomes on a reconstituted cell cycle gene promoter.

(A) Reconstitution of the TTK promoter with nucleosomes and MuvB. At top is a schematic of the promoter fragment used in these experiments with the transcription start site (TSS) and CHR site indicated. The indicated proteins were refolded with a 461 bp fragment of DNA and the samples analyzed by an agarose gel with ethidium bromide staining. (B) Schematic representation of the protocol for cross linking and imaging (left) and example electron microscopy micrograph (middle) with corresponding traced molecules (right). (C) Histograms showing the fraction of DNA molecules containing the indicated number of nucleosome-sized bubbles for a sample of 100+ analyzed DNA molecules. The MuvB concentration when used in the reaction was 1.3 μM, whereas low [MuvB] corresponds to 0.15 μM. The number at the top of the box indicates the average number of nucleosome-sized bubbles per DNA molecule with 95% confidence interval in brackets. These statistics were computed using bootstrapping with 10000 iterations of resampling. Histograms were fit to a Poisson distribution and conditions were compared using an exact-Poisson test. p-values for comparisons of average number of bubbles (λ of Poisson distribution) across conditions are reported below the histogram.

MuvB associates with the +1 nucleosome in cell cycle gene promoters

We next used an MNase-ChIP approach to detect MuvB association with nucleosomes in cells (Figure 2.7a)[57,58]. Chromatin preparations from HCT116 cells expressing Strep-tagged LIN9 were cross-linked and MNase digested. Samples were precipitated with Strep-Tactin, and after crosslinks were reversed and protein was digested, DNA fragments were purified, ligated with barcoded adapters, and sequenced. In contrast to traditional ChIP experiments that identify transcription factor binding motifs, we aimed to purify nucleosomal-DNA fragments that associate with our transcription factor [57]. To enrich for these longer fragments (>100 bp), we ligated adapters after a SPRI-bead DNA purification. Compiled DNA sequences were aligned to the human genome, and we used MACS2 to locate enriched peaks corresponding to LIN9-interacting sequences.

We first analyzed precipitated DNA sequences from HCT116 cells that were treated with Nutlin-3a, which induces DREAM-mediated repression of both S phase and M phase genes through the p53 pathway (Supplementary Figure 2.6)[45,59]. By comparing Strep-LIN9-precipitated samples to control samples in which cells were transfected with empty vector, we identified 253 genes with MACS peaks having greater than 4.7-fold enrichment in sequencing reads. Gene ontology analysis of this data set reveals enrichment in genes related to cell cycle, mitotic division, and response to DNA damage (Figure 2.7b). We found that 177 (70%) of the 253 most enriched genes have previously been identified as DREAM regulated genes based on LIN9 and E2F4/p130 ChIP, RNA expression, and promoter analysis[15,22,34,45] (Figure 2.7c). Moreover, these DREAM genes tended to show higher enrichment than the other identified genes among the top hits. We performed two additional replicate

experiments, one technical replicate with a different MNase concentration for the digestion and one biological replicate, and we found that the enrichment of many DREAM genes was reproducible (Supplementary Figure 2.7b). We performed an analogous experiment in which we expressed Strep-LIN9[4X]. The LIN9[4x] mutant fails to associate with LIN37 and RBAP48 but still associates with CHR consensus sites (Figs. 3d, 6c, and Supplementary Figure 2.7a,c). Considering the same 4.7-fold threshold, this data set contained fewer genes overall and only four DREAM genes containing enriched sequences (Figure 2.7c). We also performed an experiment precipitating Strep-LIN9 from extracts of cycling HCT116 cells and found enrichment of fewer genes compared to arrested cells (Figure 2.7c). We conclude that our experimental protocol successfully enriches LIN9-bound DNA sequences at expected cell cycle genes in arrested HCT116 cells and that enrichment depends on intact MuvBN.

Inspection of the WT LIN9-immunoprecipitated sequence reads aligned to the human genome reveals enrichment of DNA corresponding to nucleosome-sized fragments (~150 base pairs) near the transcription start site and E2F or CHR consensus sites in the DREAM-regulated genes (Figure 2.7d and Supplementary Figure 2.7d). For example, in the *CCNB2* promoter, which contains a canonical CHR DREAM-binding site, the strongest enriched peak is located just downstream of the closely spaced TSS and CHR site. This nucleosome corresponds to the +1 nucleosome, which has been previously identified as being well-positioned in repressed genes[60-63]. We observed secondary sites of enrichment, which correspond to nucleosomes (e.g. +2 and +3 nucleosomes) further downstream of the TSS. The enrichment decreases with increasing distance from the CHR site. In *FOXM1* and

ORC6, which contain CDE-CHR and E2F binding sites for DREAM respectively, we observed a similar pattern, with the +1 nucleosome showing the strongest enrichment, followed by a weaker coverage of the distal nucleosomes. Multiple lines of evidence suggest that these enriched fragments correspond to sequences in MuvB-bound nucleosomes rather than sequences protected simply by MuvB binding. First, the peaks are centered adjacent to MuvB-binding sites in DNA as opposed to centered on them, suggesting the read sequences are not protected from steric occlusion of MNAse by MuvB itself. Second, we do not see enrichment in the LIN9[4x] mutant or cycling cells experiments (Supplementary Figure 2.6d), which probe conditions in which MuvB is still bound to chromatin and would still offer MNase protection. Third, secondary sites of enrichment even more distal to the CHR sites are also nucleosome-sized. Fourth, we commonly observed sequence reads of sizes corresponding to integral numbers of nucleosomes (Supplementary Figure 2.6e).

We aligned the promoter regions of the 177 enriched DREAM genes according to their transcription start sites (TSS) to identify more broadly the structural signature of LIN9-associated nucleosomes. Most of these genes show a sharply positioned nucleosome within 150 bases downstream of the TSS (Figure 2.8a). We conclude that LIN9 primarily precipitated the +1 nucleosome in these promoters. Considering that expressed wild-type LIN9 forms complexes with other endogenous MuvB components (Figure 2.1d) but that LIN9[4X] does not associate with LIN37 and RBAP48 (Figure 2.3d), we further conclude that these nucleosomes are bound by MuvB complexes and that these interactions are mediated by MuvB[N].

We emphasize that the enriched nucleosomes in the set of DREAM genes do not overlap with the E2F and CHR consensus binding sites, suggesting that DREAM

binds these DNA elements in linker DNA and not in the nucleosome core particle. Nucleosomes are positioned next to the DREAM-binding site, which is typically in close proximity to the TSS (Figure 2.7d), and do not necessarily contain the E2F and CHR DNA sequence motifs. We also note that the primary and secondary peaks in the sequence coverage persist when the reads are filtered for exclusively mononucleosome-sized inserts (Supplementary Figure 2.6e). This observation that MuvB precipitated both proximal but not overlapping and distal nucleosomes to its consensus binding sequence further suggests that MuvB makes direct contact with the nucleosome core. This interpretation is consistent with our biochemical observations that interactions with nucleosomes are facilitated through protein-protein binding rather than through proximal DNA interactions (Figure 2.6). Importantly, the peak corresponding to the +1 nucleosome is stronger and more tightly positioned in the precipitated sequencing data compared to the input data (Figs. 2.7d and 2.8a). This enrichment of a strongly positioned nucleosome is consistent with a role for MuvB in binding and stabilizing the +1 nucleosome in DREAM promoters.

MuvB association with a tightly positioned +1 nucleosome correlates with gene repression

We next performed a similar MNase-ChIP experiment using LIN37 knock-out HCT116 cells (HCT116-LIN37$^{-/-}$, Figs. 6c, 7a,b, and Supplementary Figure 2.6a,d). In these cells, the MuvB complex assembles on CHR promoters, but cell cycle genes are no longer fully repressed by DREAM when cells are arrested[24,45]. In fact, we observed in the input MNase data from the set of DREAM genes the nucleosome phasing pattern that is characteristic of genes undergoing transcription (Figure 2.8b)[61,63]. We still observed enrichment of known DREAM genes in the pool of Strep-

LIN9 precipitated DNA reads (Figure 2.7c), and many of the enriched genes overlap between the data sets from wild-type and LIN37 knock-out cells (Supplementary Figure 2.7b). We note that this result from precipitating Strep-LIN9 from knock-out cells is distinct from what we observed precipitating Strep-LIN9[4x] from wild-type cells. In the former experiment, LIN37 is missing from MuvB complexes, while in the latter both LIN37 and RBAP48 are missing; however, both complexes can associate with DNA (Figure 2.3D and Supplementary Figure 2.6C). From this comparison, we conclude that RBAP48 is necessary for nucleosome association.

Analysis of nucleosome occupancy generated for enriched genes in the data set from knock-out cells suggests that MuvB still associates with +1 nucleosomes (Figure 2.8a,b, and Supplementary Figure 2.6). However, the bound nucleosomes are distributed over a broader region of DNA, i.e. the boundaries of the positioned nucleosome are more poorly defined, and the position more typically encroaches on the TSS. We observe a significant ($p < 0.05$) difference in nucleosome occupancy comparing the wild-type and LIN37 knockout data sets in regions 100-200 bp both downstream and upstream of the TSS in DREAM genes (Supplementary Figure 2.6b). We cannot determine that this broader distribution of nucleosome positions is directly a result of the absence of LIN37 from the complex or is a signature of expressed DREAM genes in the KO cells. Still, these results, together with our observation that Strep-Lin9 does not robustly precipitate nucleosomes from cycling cells, demonstrate that the sharply positioned MuvB-associated nucleosomes correlate with gene repression.
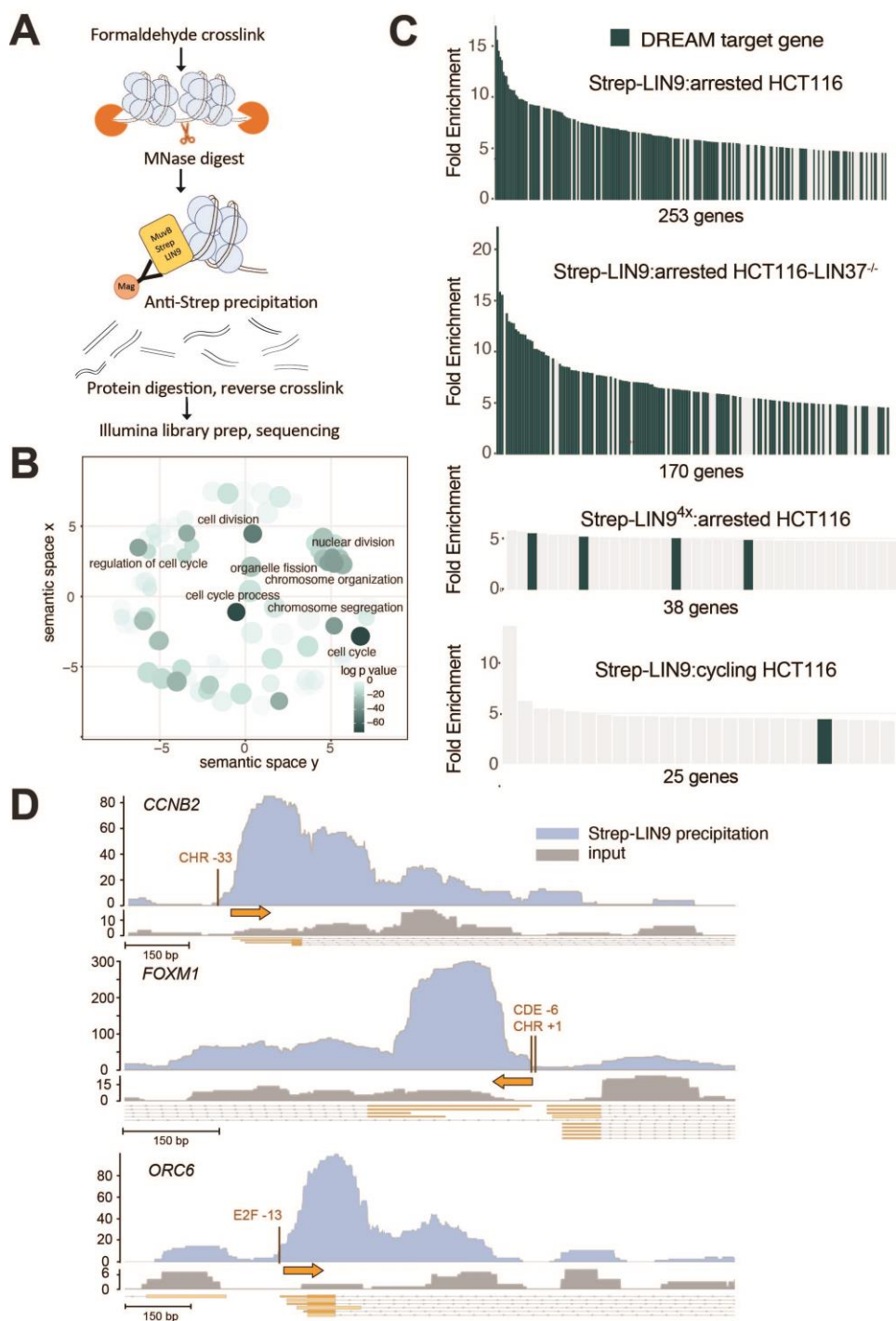
Figure 2.7: MuvB associates with nucleosomes in DREAM-regulated gene promoters in HCT116 cells. (A) Diagram of the MNase-ChIP experiment designed to enrich nucleosome-sized DNA fragments that interact with MuvB complexes. (B) Gene ontology analysis of the enriched DNA sequences following MNase digestion and precipitation from arrested cell extracts. Gene groups with p values less than 1 x 10^{-30} are labeled C) Top enriched genes with DNA sequences co-precipitated with Strep-LIN9. Experiments were performed using WT or LIN37-/- HCT116 cells, WT LIN9 or LIN94x, and in arrested or cycling cells.

43

(The number of genes with an enrichment greater that 4.7-fold are indicated for each experiment. (D) Genome browser tracks corresponding to the CCNB2, FOXM1, and ORC6 promoters. The number of DNA sequence reads is plotted for the input (grey) and Strep-LIN9 precipitated DNA samples. These data were for one replicate performed in arrested HCT116 cells. Data for other replicates and experiments are shown in Supplementary Figure. 6. The transcription start site (TSS) in each gene (base of orange arrow) along with the position of the DREAM-binding DNA motif relative to the TSS are indicated.

Figure 2.8: The sharp positioning of the MuvB-bound +1 nucleosome correlates with gene repression. (A) Heat maps showing relative read density across all genes that show greater than 4.7-fold enrichment in the Strep-LIN9 precipitant and correspond to a known DREAM gene. Both input and immunoprecipitant sequencing data are shown for the same gene set and in the same order. At the bottom are the aggregated and normalized read intensities across all genes shown in the heat map for the input and precipitated DNA data sets. Data from HCT116 (blue) and HCT116-LIN37$^{-/-}$ (pink) are shown. The observation of more peaks with periodicity corresponding to nucleosomes (i.e. the +2 and +3 nucleosome peaks) in the HCT116-LIN37$^{-/-}$ input data set is consistent with DREAM genes being active during Nutlin-3a-induced quiescence of those cells[45]. (B) Comparison of three example gene tracks showing the number of DNA sequence reads in Strep-LIN9 precipitated samples from HCT116 (blue, left axis) and HCT116-LIN37$^{-/-}$ (pink, right axis) cells. The TSS (orange arrow) and DREAM-binding DNA motifs are indicated. (C) Overall model showing organization of DREAM and the association between MuvBN and the +1 nucleosome, which we propose mediates gene repression.

Discussion

Genetic studies across model organisms all point to the function of the MuvB core as an intrinsically repressive complex that interacts with other TFs to modulate gene expression. In *C.elegans*, even when the p107/p130 ortholog is knocked out such that DREAM does not form on promoters, the MuvB core retains the ability to repress target genes[64]. In *Drosophila,* the lethal *myb-null* phenotype can be rescued by the loss of function of the fly orthologs of LIN9 and LIN37, which restores expression of MYB target genes[65-67]. In mammalian cells, LIN37 knockout and LIN9 knockdown leads to specific loss of repression of cell cycle genes upon driving cell cycle exit[22,24,42], and a similar defect is observed upon loss of the RBAP48 ortholog in *Drosophila*[43]. While mammalian LIN9 loss also fails to activate mitotic genes, this activation defect may be linked to the requirement of LIN9 for recruiting B-MYB[22,35,41,42]. Together, these results demonstrate that LIN9, LIN37 and RBAP48 contribute to a repressive MuvB function.

Our results implicate the MuvBN subcomplex as the structural unit in MuvB responsible for this intrinsic repressive function and link repression to nucleosome binding. The structure and biochemical data demonstrate that LIN9 and LIN37 form a scaffold for the MuvB core in that they bind and assemble LIN52, LIN54, and RBAP48 (Figure 2.8C). The MuvBN structure contains RBAP48 and the conserved sequences of LIN37 that are both required specifically for cell cycle gene repression. Our data demonstrate that MuvB binds and stabilizes nucleosomes and that MuvBN, which contains the repressor subunits, is sufficient for this interaction. We observed the association of LIN9-containing MuvB complexes with the +1 nucleosome in the promoters of repressed cell cycle genes in arrested cells, but this association is lost

in cycling cells or with a LIN9 mutant that does not assemble the MuvBN components LIN37 and RBAP48. While we still see association of MuvB with nucleosomes in active gene promoters in arrested LIN37 knock-out cells, the associated nucleosome appears more strongly positioned under conditions of repression. We propose that MuvB stabilizes nucleosome position by making a bipartite interaction with the nucleosome and DNA. LIN54 binds the CHR sequence while MuvBN tethers the nucleosome near the CHR through a direct association with the histone tails or additional nucleosome contacts. We further propose that this association with +1 nucleosomes contributes to repression by inhibition of remodeling, polymerase activity, or posttranslational histone modification required for transcription (Figure 2.8c). For example, association of MuvB with the histone H3 tail may sequester the tail from other chromatin-binding proteins and remodelers. Another possibility is that by tightly binding and positioning the +1 nucleosome, MuvB may increase the energy barrier that stalls RNA polymerase activity, resulting in uninitiated or aborted transcripts. By binding through multiple modes, i.e. histone tails and core, MuvB could prevent the unwrapping and movement of the +1 nucleosome. Although initial mass spectrometry analysis revealed few binding partners to human MuvB that could explain its repressive role, more recent studies have found MuvB can in certain contexts recruit proteins such as PAF and SIN3B[30,68]. Additional factors may also function to enhance repression in addition to the nucleosome binding activity of MuvB. Our result that MuvB can bind nucleosomes even in the absence of an H3 tail interaction suggests that the H3 site in RBAP48, which our structure shows is accessible in MuvB, might be used by MuvB to recruit other repressor complexes.

Research on the structure of chromatin has revealed important factors that determine the nucleosome position in the genome, including intrinsic properties of DNA sequence, chromatin remodeling complexes, the polymerase machinery, and sequence specific TFs[1,3,4]. The role of TFs has focused on their potential for maintaining the nucleosome depleted region around the TSS and for establishing the +1 nucleosome. Evidence supports a model in which the mutually exclusive interaction between TFs and histones for DNA allows TFs to act as a barrier for nucleosome deposition such that the +1 and other proximal nucleosomes form at the closest accessible sites. Our data support a more direct function for TFs in establishing the +1 position through physical association and correlate this association in cells with a more tightly positioned nucleosome at repressed genes. Under conditions when MuvB is not actively repressing (LIN37 KO cells in quiescence), we observe more variability in the nucleosome position. The extent to which these observations result from the dynamics of RNA polymerase during the transition from repressed to active genes remains uncertain.

Several important questions remain about this MuvB repressive function including the structural mechanism of nucleosome recognition and the role of LIN37. RBAP48 in many studies is sufficient for nucleosome binding, and it is still present in MuvB complexes that lack LIN37 yet cannot repress gene expression[24]. We speculate that this non-functional complex may be unstable or improperly structured such that it cannot enact repression or bind co-repressors. The extensive interaction interface and co-dependence of their association in our mutagenesis study support the hypothesis that the core subunits of MuvBN, (LIN9, LIN37 and RBAP48) co-fold to form a stable complex. Another important remaining question is how the structure and

function of MuvB changes such that it switches from a repressor to an activator of gene expression once cells enter the cell cycle. MuvB components are still present on the promoter and are required for recruiting B-MYB and FOXM1. One possibility is that MuvB repression activity is relieved, for example by the binding of B-MYB, which is consistent with observations in *Drosophila* that the MuvB-binding sequence in B-MYB is alone sufficient to rescue a B-MYB deletion phenotype[69]. Another possibility is that Cdk phosphorylation, detected on all the MuvB subunits, plays a role in modulating MuvB function[70]. A third possibility is that in addition to its repressive function, MuvB positions the +1 nucleosome to prime genes for expression upon the binding of the activator transcription factors B-MYB and FOXM1. In this mechanism, MuvB may facilitate the acetylation of histones by the p300 acetylation machinery, which is recruited by the activator TFs. Through TF and p300 association, MuvB may also help recruit the basal transcription machinery. Finally, it will be important to understand how widespread interactions of TFs with the +1 nucleosome are and how these interactions regulate chromatin and gene expression.

Methods

Mammalian expression plasmids

The LIN9 and RBAP48 ORFs were amplified from cDNA derived from mouse NIH3T3 cells by standard PCR. The EGFP ORF was amplified from pEGFP-N1 (Clontech). The ORFs were cloned into pcDNA3.1(+) and fused either with an N-terminal 3xFlag tag (LIN9) or an N-terminal 1xFlag tag (RBAP48, EGFP). Site-directed mutagenesis was performed following the QuikChange protocol (Stratagene). For MNase-Seq

experiments, the LIN9 ORFs were subcloned into pcDNA3.1(+) containing an N-terminal Twin-StrepII tag.

Recombinant Protein expression

To assemble the entire MuvB complex, proteins (GST- or Strep-LIN9$^{94-538}$, GST-LIN37, GST-LIN52, His- or GST-LIN54$^{504-749}$, and Strep-RBAP48) were co-expressed in Sf9 cells via baculovirus infection. Cell pellets were harvested after 72 hours of growth in suspension at 27°C, and complexes were purified using GST-affinity purification followed by Strep-affinity purification. After removal of affinity tags through TEV protease cleavage, purified complexes were isolated through size-exclusion chromatography using a Superdex200 column. The final buffer contained 200 mM NaCl, 25 mM Tris HCl and 1 mM DTT at pH 8.0. The MuvBN subcomplex was assembled by co-expressing GST-LIN9$^{94-274}$, LIN37$^{92-130}$, and full-length RBAP48 in Sf9 cells as described for the full complex. The subcomplex was purified using GST affinity purification followed by anion exchange. Affinity tags were then removed with TEV protease and the complex was isolated with a Superdex 200 column. The final buffer contained 150 mM NaCl, 25 mM Tris HCl and 1 mM DTT at pH 8.0.

X-ray crystallography

The MuvBN subcomplex was crystallized in a sitting drop at 4°C containing 0.2 M sodium tartrate tetrahydrate, 0.1 M bis-tris propane pH 6.5, and 20% PEG 3350. Crystals were harvested and directly frozen in liquid nitrogen. Data were collected at λ = 1.0332Å and 100 K on Beamline 23-ID-B at the Advanced Photon Source, Argonne National Laboratory. Diffraction spots were integrated with Mosflm[71]. Phases were solved by molecular replacement with PHASER[72] and using RBAP48 (PDB:

[3GFC](#)) as a search model. The initial model was rebuilt with Coot[73], and LIN9 and LIN37 were added to the unmodeled electron density. The resulting model was refined with Phenix[74]. Several rounds of position refinement with simulated annealing and individual temperature-factor refinement with default restraints were applied. The final refined model was deposited in the Protein Data Bank under Accession Code PDB ID: [7N40](#).

Co-immunoprecipitation and DNA affinity experiments

Human HCT116 colon carcinoma cells were cultivated in DMEM supplemented with 10% FBS. Transfections were performed in 10cm plates using 7 ug plasmid and 35 ul PEI per plate. To stimulate DREAM formation, cells were treated with 10uM Nutlin-3a for 24 hours. Cells were harvested 48h after transfection. Whole cell extracts were prepared by lysing the cells in IP lysis buffer (50 mM TRIS-HCl pH 8.0, 0.5% Triton-X 100, 0.5 mM EDTA, 150 mM NaCl, 1mM DTT, and protease inhibitors) for 10 min on ice followed by 5x 1s direct sonication. Flag-tagged proteins were immunoprecipitated from 2-3 mg cellular extracts with Pierce Anti-DYKDDDDK Magnetic Agarose (Invitrogen). Beads were washed 5x with 1ml IP lysis buffer end eluted with 50µl 1xLaemmli buffer. 12 µg of input samples and 12 µl IP samples were analyzed by SDS-PAGE and western blot following standard protocols. The following antibodies were applied for protein detection: FLAG-HRP (RRID:AB_2017593, Santa Cruz Biotechnology; dilution 1:2000), p130/RBL2 (D9T7M) (RRID:AB_2798274, Cell Signaling; dilution 1:1000), LIN54 A303-799A (RRID:AB_11218173, Bethyl Laboratories; dilution 1:1000), LIN9 ab62329 (RRID:AB_1269309, Abcam; dilution 1:1000), RBBP4 A301-206A (RRID:AB_890631, Bethyl Laboratories; dilution

1:5000), LIN37-T3 (custom-made at Pineda Antikörper-Service, Berlin, Germany; dilution 1:1000)[54].

For DNA affinity purifications, HCT116 cells were cultivated in 15 cm plates and transfected with 70 µl PEI and 15µg plasmids expressing wild-type and mutant LIN9 fused with an N-terminal 3xFlag tag. 24 hours after the transfection cells were treated with 5 µM Nutlin-3a for 48 hours. Affinity purifications were performed as described earlier[75]. Biotinylated DNA probes were either amplified from the pGL4.10 empty vector or from pGL4.10 containing the mouse *Ccnb2* CDE/CHR MuvB-binding site[34]. The following antibodies were applied for protein detection: FLAG-M2 (RRID:AB_262044, Sigma-Aldrich; dilution 1:1000), p130/RBL2 (D9T7M) (RRID:AB_2798274, Cell Signaling; dilution 1:1000), LIN37-T3 (custom-made at Pineda Antikörper-Service, Berlin, Germany)[54], Histone H3 (RRID:AB_331563, Cell Signaling Technology; dilution 1:1000).

Nucleosome reconstitution

Xenopus histones as well as their tailless counterparts were expressed and purified in *E. coli* as inclusion body preparations as previously[76,77]. Octamer reconstitution was completed by mixing equimolar amounts of purified histones in a buffer containing 7 M guanidinium HCl, 20 mM Tris pH 7.5 and 10mM *β-mercaptoethanol*, followed by dialysis into 2 M NaCl, 10 mM Tris pH7.5, 1 mM EDTA and 5 mM *β-mercaptoethanol*. Folded octamers were purified using size-exclusion chromatography on a Superdex 200 column. Nucleosome reconstitution was performed by mixing purified histone octamers with the Widom 601 positioning sequence and de-salting by gradient dialysis [76]. For Widom nucleosomes, we used a 1.1:1 ratio of octamers:DNA

molecules. At a salt concentration of 50 mM, nucleosome samples were collected in a buffer containing 50 mM Tris pH 7.5 and 1 mM DTT.

Fluorescence polarization assay

Histone peptides were synthesized with fluorescein. For experiments with Widom nucleosomes, the 601 sequences were PCR amplified with a primer containing fluorescein and reconstituted with octamer as described above. 20 nM peptide was mixed with varying concentrations of MuvB protein complex in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT, and 0.1% (v/v) Tween-20. Twenty microliters of the reaction were used for the measurement in a 384-well plate. Fluorescence polarization (FP) measurements were made in triplicate, using a Perkin-Elmer EnVision plate reader. The K_d values were calculated using global fitting in Prism 8 (Version 8.2.1).

Electron microscopy on reconstituted promoters

The minimal region of the human *TTK* promoter (461bp) was cloned, amplified by PCR, and purified by agarose gel extraction. Histone octamers were folded with the *TTK* DNA as described above for Widom nucleosomes, but we used an octamer to DNA ratio of 3.1:1 to allow for the formation of di- and tri-nucleosome species. For the relevant conditions, purified MuvB complex or LIN54 was added to the nucleosome folding reaction during the de-salting process at a NaCl concentration of ~800mM. Cross-linking of gene promoters and electron micrograph preparation was performed as previously described[55]. In brief, samples were treated with trimethylpsoralen and UV radiation to allow double-stranded DNA crosslinks to form at unprotected, octamer-free regions. Following crosslinking, proteins were digested by Proteinase K,

and DNA molecules were purified, denatured, and spread across the surface of a copper transmission electron microscopy grid. Electron micrographs of all samples were prepared by rotary metal shadowing, and grids were visualized and collected on a JEOL 1230 TEM at the UC Santa Cruz IBSC Microscopy facility and a Tecnai 12 TEM at the UC Berkeley ELM lab. DNA molecules were traced, and molecules coordinates were saved using Fiji tools in the ImageJ software package as previously described[55]. The resulting traces were analyzed using custom python tools. Each DNA strand was traced such that an "end" of the molecule could be identified. Thus, every coordinate in one strand can be aligned to its complement by closest distance. Coordinates are assigned to base positions using a scale based on the physical distance between coordinates within each strand based on the known length of the TTK promoter (461 bp). A base pair is labeled single-stranded if the distance between strands exceeds a threshold distance, determined empirically. Once all base pairs are labeled, "bubbles" are determined by contiguous single-stranded stretches. Finally, a single-stranded "bubble" is labeled a nucleosome if its length is >90 basepairs (Supplementary Figure 2.5). Bubble fusions occur such that two or more adjacent nucleosomes form one contiguous bubble; for this analysis, bubble fusions were labeled as a single nucleosome. This estimate was used because the number of total bubble fusions observed within the dataset was small <5%.

MNase-ChIP

HCT116 cells transfected with Strep-LIN9 constructs or an empty Strep expression plasmid. 24h after transfection, cells were treated with 10μM Nutlin-3a (Selleckchem) and harvested after 48h. Cells were crosslinked with 1% formaldehyde for 15 minutes. The crosslinking reaction was stopped with glycine, cells were washed twice with

PBS, and pellets were collected. Cell lysis and MNase digestion (1x or 5x) were performed as described earlier, and following digestion, LIN9-bound samples were precipitated using Streptactin-XT magnetic beads (IBA Lifesciences)[58]. Both input and IP samples were subject to RNAse treatment and proteinase K digestion and were reverse crosslinked by incubation at 65 °C for 16 hours. DNA was purified by 2x SPRI bead clean-up. Library prep was carried out using NEB Next Ultra II kits, and paired-end sequencing was carried out on the NovaSeq 6000 platform with 150bp paired end mode for Illumina at Novogene Biotech, Co., LTD.

Sequencing reads were aligned against hg38 using the bwa-mem aligner[78,79]. Samtools and bedtools were used to convert data into bam and bed files respectively. Peak calling for the precipitated samples was performed using the MACS2 *-bampe* algorithm and using the empty Strep-IP conditions as the control. To retrieve gene names for MACS2 peaks, coordinates were intersected with known genes using the Table Browser tool provided by the UCSC genome browser. Gene ontology analysis was performed on MACS2 peaks showing a greater than 4.7-fold enrichment using the web-based tools GeneOntology.org and Revigo[80]. We generated coverage plots of our reads using Gviz and rtracklayer and other opensource R tools.

We utilized NucTools in paired-end mode to analyze nucleosome occupancy on input and Strep-LIN9 precipitated reads with single base-pair resolution (bin width=1bp)[81]. We restricted our analysis to genes that showed a MACS enrichment of greater than 4.7-fold and were previously annotated to bind DREAM, to respond to p53 stimulation, and to become derepressed in LIN37 knockout cells[15,22,34,45]. We retrieved the TSS for this set of genes, either from those annotations or using bioMart[82]. As needed, the

TSS sites were mapped on to hg38 using liftover. We oriented the output to center on the TSS and maintain a uniform direction of transcription. We then utilized the Cluster Map Builder feature of NucTools to generate aggregate plots and heatmaps of our genes.

Data Availability

X-ray diffraction data and model coordinates for the MuvBN structure in this study have been deposited in the Protein Data Bank (RCSB.org) under accession code 7N40. Previously determined structures that were used for comparison are also available in the Protein Data Bank under accession codes 2YBA, 3CFV, 3GFC, 2M0O, 4A4E, 2XU7, 4R7A, 4PC0, 2YB8, 5FXY, 5WAI. MNase-ChIP data have been deposited in the NCBI Gene Explression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under accession code GSE189435.

Acknowledgements

Spplementary Table 2.1: **Data collection and Refinement statistics**.
Data were collected on a single crystal. Values in parentheses are for the highest resolution shell.

Data collection	
Beam line	APS(23IDB)
Resolution Range (Å)	60.0-2.55 (2.69 - 2.55)
Space group	C 1 2 1
Unit cell dimensions(a,b,c) (Å), (α, β, γ (⁰)	133.58 77.81 64.56 90 114.71 90
Wavelength (Å)	1.03
Total observations	43599 (6521)
Unique reflections	19131 (2805)
Completeness (%)	97.4 (97.8)
R_{merge}	13.5 (52.1)
$<I/\sigma>$	7.9 (3.5)
$CC_{1/2}$	0.97 (0.47)
Redundancy (highest shell)	2.3 (2.3)
Refinement	
R_{work} %/ R_{free} %	16.1/ 25.0
Number of non-hydrogen atoms	4862
Protein	4720
Water	142
Wilson B-factor (Å^2)	32.79
RMSD Bond length (Å)	0.008
RMSD Bond angle (°)	0.94
Ramachandran favored(%)/Ramachandran outliers (%)	96.0/0.0
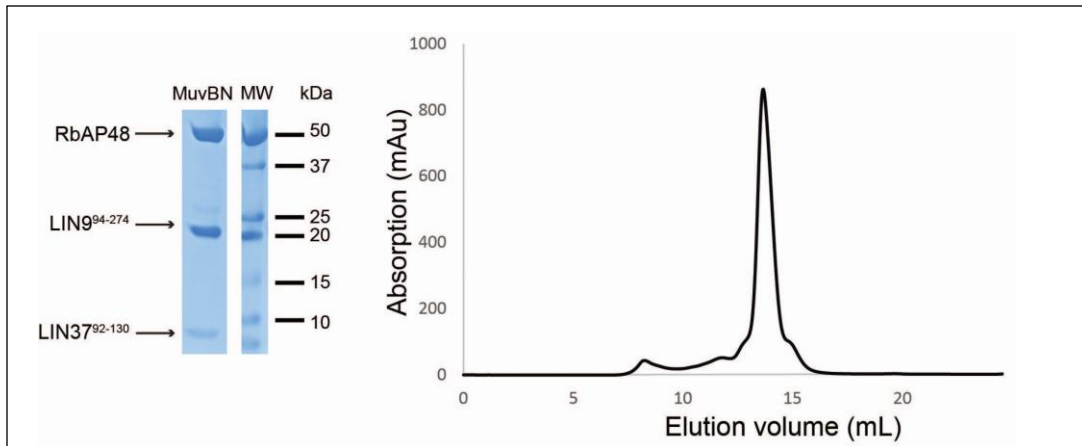
Supplementary Figure 2.1: MuvBN reconstitution for crystallization. Stable complex of RBAP48 and the indicated LIN9 and LIN37 constructs eluted on Superdex 200. (Left) Coomassie stained gel of three subunits following purification. (Right) UV absorption trace from size-exclusion chromatography.

Supplementary Figure 2.2: Comparison of RBAP48 structures and interactions with different peptides and proteins. The MuvBN complex structure is in shown at the top. Pairwise RMSD for C$_\alpha$ atoms in RBAP48 are indicated under each structure. Structures with peptides (top two rows) show two major binding sites. The H3 binding site across the face of the β-propeller domain (top row) is accessible in MuvBN, while the H4 binding site (middle row) is not. The RBAP48 insertion loop (residues 88-115) is not ordered and modeled in nearly all the structures of RBAP48 with small

peptides. It appears in the structure with the Mta1 peptide, although in that crystal it is involved in packing. In the structures of larger complexes (bottom row), part of the insertion loop is ordered and forms a strand and packs against an added strand from an RBAP48-interacting partner. In the MuvBN structure, the loop is completely ordered and makes extensive interactions with both LIN9 and LIN37.

Supplementary Figure. 2.3: Supporting data characterizing MuvB association with nucleosomes. (A) Fluorescence polarization (FP) assay as in Figure. 4 but using the MuvBN subcomplex. FP is normalized to dye-labeled probe alone. Data are fit assuming a single binding constant. Average affinities from three replicates are shown. (B) Isothermal titration calorimetry measurements of H3 and H4 peptide binding to RBAP48 and MuvB. These data demonstrate MuvB binds H3 similar to RBAP48 alone, and the affinity of H3 for RBAP48 is similar to that previously reported[47]. We suspect that the affinity of MuvBN for H3 is slightly tighter when measured by FP because of the hydrophobic dye on the probe. ITC measurements were made with a Microcal, LLC calorimeter. 1 mM histone was titrated into 50 μM of the indicated protein in a buffer containing 10 mM HEPES pH 7.0, 200 mM NaCl, 5 MM DTT. The data were fit to a one site binding model and the stoichiometry parameter in each case fit to ~1. (C) Fluorescence polarization (FP) assay as in Figure. 4 but using MuvB assembled with the LIN9 Tudor cage mutant (L230A, F238A, F256A, and H264A). FP is normalized to dye-labeled probe alone. Data are fit assuming a single binding constant. Average affinities from three replicates are shown.

Supplementary Figure. 2.4: Supporting data for nucleosome crosslinking experiment with reconstituted *TTK* promoter. (A) DNA fragment from *TTK* promoter used in these experiments. The boxed sequence corresponds to the fragment used in the binding experiment in panel C. (B) Estimated probability density functions for bubble size for each of the experimental conditions: blue (*TTK* promoter with histones), gold (+LIN54), dark green (+MuvB), light green (+ low[MuvB]), and purple (mutant *TTK* with histones and MuvB). Considering the resolution of the experiment and allowing for the possibility of partially wrapped octamers, we interpreted bubbles larger than 90 bp as nucleosomes. We have previously shown that a partially unwrapped pre-nucleosomal structure can be distinguished from a fully wrapped nucleosome with this psoralen cross-linking assay[56]. We note that probability density around 90 bp decreases when MuvB but not LIN54 is added to the sample. This shift in the probability density away from pre-nucleosomal to nucleosomal size suggests that MuvB stabilizes the fully wrapped nucleosomes. (C) Fluorescence polarization assay measuring the affinity of MuvB for a 40-bp fragment from the TTK promoter (boxed in Panel A). 20 nM of fluorescein-labeled fragment was titrated with increasing concentrations of purified MuvB in a buffer containing 150mM NaCl, 20mM Tris pH 8.0, and 0.1%Tween-20, and the fluorescence polarization recorded in a 384-well plate (volume 20 μL). Data were fit to a single site model. Three replicates were performed. The error bars around each point correspond to the standard deviation of each measurement across replicates, and the K_d error is the error estimate after global fitting of all data across three replicates. The K_d value measured here is similar to that previously reported for LIN54 binding to a similar CHR motif[31]. The scrambled mutTTK is the same sequence but with the CHR sequence relaced with TGCATA.

Supplementary Figure. 2.5: Data supporting the MNase-ChIP experiment. (A) Western blot demonstrates the expression of Strep-LIN9 and Strep-LIN9[4x], the absence of LIN37 in the HCT116-LIN37[-/-] cells, and the induction of p53 and the p53 target p21 upon Nutlin-3a treatment. * Indicates a nonspecific band in the blot. (B) Overlap of enriched genes between replicates of the MNase-ChIP experiment. The top two Venn diagrams show overlap of all genes (left) and DREAM genes (right) that were enriched >4.7-fold in the three replicate experiments in which Strep-LIN9 was expressed in HCT116 cells. The bottom Venn diagram shows overlap of genes that were enriched in experiments

expressing Strep-LIN9 in HCT116 and HCT116-LIN37$^{-/-}$ cells. This comparison uses the total DREAM gene list compiled across all three replicates for each condition. (C) LIN9 mutants that do not assemble LIN37 and/or RBAP48 still bind DNA containing the consensus CDE-CHR site. HCT116 cells were transfected with plasmids expressing 3xFLAG-tagged wild-type LIN9, LIN9^{3x}, and LIN9^{4x}. Cells were arrested with Nutlin-3a and MuvB complex components were purified. Purification was performed with a fragment of the pGL4.10 vector containing the mouse Ccnb2 CDE/CHR MuvB-binding site. Protein binding to a pGL4.10 fragment without this element was analyzed to control for background binding (CTRL). Binding of Flag-LIN9 and endogenous LIN37 and p130 was tested by western blotting. Histone H3 was probed as a control for DNA affinity purification. (D) Number of DNA sequence reads across the *CCNB2* gene track for all eight indicated experiments. The plots show DNA reads from the Strep-precipitated samples. The TSS (orange arrow) and position of the CHR element are shown. The number of reads from the Strep-LIN9^{4x} experiments are much lower and do not consistently show peaks corresponding to the +1 nucleosome. (E) DNA sequence reads from the Strep-LIN9:HCT116 precipitation are plotted for the *CCNB2* gene track as in Supplementary Figure. 2.6D. The alignment tracks of the sequence fragments are diagrammed below. In these tracks, explicitly sequenced DNA is shown as a block and inferred sequence from the paired end analysis is shown as a line. The grey tracks and coverage plot represent the full data as shown in Figure. 6D and Supplementary Figure. 2.6D. The blue tracks and coverage plot represent a subset of that data, in which the alignment tracks are filtered to include only sequence reads that are 130-200 base pairs, corresponding to mononucleosome-size fragments.

Supplementary Figure. 2.6: Comparison of MNase-ChIP replicate experiments in HCT116 and HCT116-LIN37⁻/⁻ cells. Replicates 1 and 3 are biological replicates. Replicate 2 is a technical replicate of Replicate 1 but using a 5-fold lower MNase concentration. (**A**) As in Figure. 7B, heat maps for each experiment show the normalized DNA sequence read density across all DREAM genes that were detected with greater than 4.7-fold enrichment compared to a Strep-precipitation control. Replicates 1 to 3 are shown top to bottom in each cell type. –next page—

(**B**) As in Figure. 7B, overlay of the aggregated and normalized read signal density from the same set of genes as shown in the heat maps. The right graph shows results of a two-tailed student's t-test comparing the three replicate datasets from wild-type and the three replicate datasets from LIN37[-/-] HCT116 cells. The p-value comparing the normalized signal values in the two groups is reported for each base pair in the aggregated peaks. The pink line shows p = 0.05. **(C)** For every gene that is detected at greater than 4.7-fold enrichment in Replicate 1 of both the WT and LIN37[-/-] datasets, the ratio of the integrated normalized intensity of DNA reads 200 bp downstream to 200 bp upstream of the TSS is plotted (top).As seen in this plot and in panel B, in almost all cases the integrated intensity is greater downstream of the TSS (ratio greater than 1); however, the ratio of intensity is less for the experiment in the LIN37[-/-] cells, consistent with the conclusion that the nucleosome is less well positioned. To perform a statistical analysis, we calculated the ratio of these intensity ratios, comparing WT to LIN37[-/-] cells for each gene (plotted in bottom panel). The average ratio across the genes for this replicate (Replicate 1) was 3.5 and the 95% confidence interval, treating each gene as an individual experiment, is 2.5 to 4.4. We conclude that in LIN37[-/-] HCT116 cells, there is significantly less difference in nucleosome density between upstream and downstream of the TSS relative to WT HCT116 cells. We note that we obtained similar statistical significance when performing this analysis using data from the other two replicates of the experiment in WT and LIN37[-/-] HCT116 cells.

References

1.   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).

2.   Lorch, Y., LaPointe, J.W. & Kornberg, R.D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203-10 (1987).

3.   Teves, S.S., Weber, C.M. & Henikoff, S. Transcribing through the nucleosome. *Trends Biochem Sci* **39**, 577-86 (2014).

4.   Kujirai, T. & Kurumizaka, H. Transcription through the nucleosome. *Curr Opin Struct Biol* **61**, 42-49 (2020).

5.   Michael, A.K. et al. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460-1465 (2020).

6.   Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568 (2015).

7.   Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76-81 (2018).

8.	Dynlacht, B.D. Regulation of transcription by proteins that control the cell cycle. *Nature* **389**, 149-52 (1997).

9.	Fischer, M. & Müller, G.A. Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Crit Rev Biochem Mol Biol*, 1-25 (2017).

10.	Sadasivam, S. & DeCaprio, J.A. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer* **13**, 585-95 (2013).

11.	Bar-Joseph, Z. et al. Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc Natl Acad Sci U S A* **105**, 955-60 (2008).

12.	Grant, G.D. et al. Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol Biol Cell* **24**, 3634-50 (2013).

13.	Liu, Y. et al. Transcriptional landscape of the human cell cycle. *Proc Natl Acad Sci U S A* **114**, 3473-3478 (2017).

14.	Whitfield, M.L. et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000 (2002).

15.	Fischer, M., Grossmann, P., Padi, M. & DeCaprio, J.A. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res* **44**, 6070-86 (2016).

16.	Kent, L.N. & Leone, G. The broken cycle: E2F dysfunction in cancer. *Nat Rev Cancer* **19**, 326-338 (2019).

17.	Musa, J., Aynaud, M.M., Mirabeau, O., Delattre, O. & Grunewald, T.G. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis* **8**, e2895 (2017).

18.	Myatt, S.S. & Lam, E.W. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer* **7**, 847-59 (2007).

19.	Harrison, M.M., Ceol, C.J., Lu, X. & Horvitz, H.R. Some C. elegans class B synthetic multivulva proteins encode a conserved LIN-35 Rb-containing complex distinct from a NuRD-like complex. *Proc Natl Acad Sci U S A* **103**, 16782-7 (2006).

20.	Korenjak, M. et al. Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell* **119**, 181-93 (2004).

21.   Lewis, P.W. et al. Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev* **18**, 2929-40 (2004).

22.   Litovchick, L. et al. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell* **26**, 539-51 (2007).

23.   Schmit, F. et al. LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. *Cell Cycle* **6**, 1903-13 (2007).

24.   Mages, C.F., Wintsche, A., Bernhart, S.H. & Müller, G.A. The DREAM complex through its subunit Lin37 cooperates with Rb to initiate quiescence. *Elife* **6**(2017).

25.   Schade, A.E., Oser, M.G., Nicholson, H.E. & DeCaprio, J.A. Cyclin D-CDK4 relieves cooperative repression of proliferation and cell cycle gene expression by DREAM and RB. *Oncogene* **38**, 4962-4976 (2019).

26.   Guiley, K.Z. et al. Structural mechanisms of DREAM complex assembly and regulation. *Genes Dev* **29**, 961-974 (2015).

27.   Sandoval, R., Pilkinton, M. & Colamonici, O.R. Deletion of the p107/p130-binding domain of Mip130/LIN-9 bypasses the requirement for CDK4 activity for the dissociation of Mip130/LIN-9 from p107/p130-E2F4 complex. *Exp Cell Res* **315**, 2914-20 (2009).

28.   Sadasivam, S., Duan, S. & DeCaprio, J.A. The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev* **26**, 474-89 (2012).

29.   Iness, A.N. et al. The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. *Oncogene* **38**, 1080-1092 (2019).

30.   Kim, M.J. et al. PAF remodels the DREAM complex to bypass cell quiescence and promote lung tumorigenesis. *Mol Cell* **81**, 1698-1714 e6 (2021).

31.   Marceau, A.H. et al. Structural basis for LIN54 recognition of CHR elements in cell cycle-regulated promoters. *Nat Commun* **7**, 12301 (2016).

32.   Müller, G.A. et al. The CHR promoter element controls cell cycle-dependent gene transcription and binds the DREAM and MMB complexes. *Nucleic Acids Res* **40**, 1561-78 (2012).

33.   Schmit, F., Cremer, S. & Gaubatz, S. LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence-specific manner. *FEBS J* **276**, 5703-16 (2009).

34.   Müller, G.A. et al. The CHR site: definition and genome-wide identification of a cell cycle transcriptional element. *Nucleic Acids Res* (2014).

35.   Guiley, K.Z. et al. Structural mechanism of Myb-MuvB assembly. *Proc Natl Acad Sci U S A* **115**, 10016-10021 (2018).

36.   Litovchick, L., Florens, L.A., Swanson, S.K., Washburn, M.P. & DeCaprio, J.A. DYRK1A protein kinase promotes quiescence and senescence through DREAM complex assembly. *Genes Dev* **25**, 801-13 (2011).

37.   Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735-45 (2007).

38.   Verreault, A., Kaufman, P.D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95-104 (1996).

39.   Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**, 357-64 (1997).

40.   Zhang, Y. et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**, 1924-35 (1999).

41.   Osterloh, L. et al. The human synMuv-like protein LIN-9 is required for transcription of G2/M genes and for entry into mitosis. *EMBO J* **26**, 144-57 (2007).

42.   Reichert, N. et al. Lin9, a subunit of the mammalian DREAM complex, is essential for embryonic development, for survival of adult mice, and for tumor suppression. *Mol Cell Biol* **30**, 2896-908 (2010).

43.   Taylor-Harding, B., Binne, U.K., Korenjak, M., Brehm, A. & Dyson, N.J. p55, the Drosophila ortholog of RbAp46/RbAp48, is required for the repression of dE2F2/RBF-regulated genes. *Mol Cell Biol* **24**, 9124-36 (2004).

44.   White-Cooper, H., Leroy, D., MacQueen, A. & Fuller, M.T. Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* **127**, 5463-73 (2000).

45.   Uxa, S. et al. DREAM and RB cooperate to induce gene repression and cell cycle arrest in response to p53 activation. *Nucleic Acids Res* **47**, 9087-9103 (2019).

46. Xu, C. & Min, J. Structure and function of WD40 domain proteins. *Protein Cell* **2**, 202-14 (2011).

47. Schmitges, F.W. et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-41 (2011).

48. Song, J.J., Garlick, J.D. & Kingston, R.E. Structural basis of histone H4 recognition by p55. *Genes Dev* **22**, 1313-8 (2008).

49. Chen, S., Jiao, L., Shubbar, M., Yang, X. & Liu, X. Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol Cell* **69**, 840-852 e5 (2018).

50. Cai, L. et al. An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol Cell* **49**, 571-82 (2013).

51. Lu, R. & Wang, G.G. Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem Sci* **38**, 546-55 (2013).

52. Tripsianes, K. et al. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat Struct Mol Biol* **18**, 1414-20 (2011).

53. Zhang, W. et al. Structural plasticity of histones H3-H4 facilitates their allosteric exchange between RbAp48 and ASF1. *Nat Struct Mol Biol* **20**, 29-35 (2013).

54. Müller, G.A., Stangner, K., Schmitt, T., Wintsche, A. & Engeland, K. Timing of transcription during the cell cycle: Protein complexes binding to E2F, E2F/CLE, CDE/CHR, or CHR promoter elements define early and late cell cycle gene expression. *Oncotarget* **8**, 97736-97748 (2017).

55. Brown, C.R. et al. Chromatin structure analysis of single gene molecules by psoralen cross-linking and electron microscopy. *Methods Mol Biol* **1228**, 93-121 (2015).

56. Fei, J. et al. The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev* **29**, 2563-75 (2015).

57. Tolstorukov, M.Y. et al. Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proc Natl Acad Sci U S A* **110**, 10165-70 (2013).

58. Gutin, J. et al. Fine-Resolution Mapping of TF Binding and Chromatin Interactions. *Cell Rep* **22**, 2797-2807 (2018).

59. Schade, A.E., Fischer, M. & DeCaprio, J.A. RB, p130 and p107 differentially repress G1/S and G2/M genes after p53 activation. *Nucleic Acids Res* **47**, 11197-11208 (2019).

60. Hughes, A.L. & Rando, O.J. Mechanisms underlying nucleosome positioning in vivo. *Annu Rev Biophys* **43**, 41-63 (2014).

61. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-98 (2008).

62. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267-73 (2013).

63. Oruba, A., Saccani, S. & van Essen, D. Role of cell-type specific nucleosome positioning in inducible activation of mammalian promoters. *Nat Commun* **11**, 1075 (2020).

64. Goetsch, P.D., Garrigues, J.M. & Strome, S. Loss of the Caenorhabditis elegans pocket protein LIN-35 reveals MuvB's innate function as the repressor of DREAM target genes. *PLoS Genet* **13**, e1007088 (2017).

65. Beall, E.L., Bell, M., Georlette, D. & Botchan, M.R. Dm-myb mutant lethality in Drosophila is dependent upon mip130: positive and negative regulation of DNA replication. *Genes Dev* **18**, 1667-80 (2004).

66. Wen, H., Andrejka, L., Ashton, J., Karess, R. & Lipsick, J.S. Epigenetic regulation of gene expression by Drosophila Myb and E2F2-RBF via the Myb-MuvB/dREAM complex. *Genes Dev* **22**, 601-14 (2008).

67. Beall, E.L. et al. Discovery of tMAC: a Drosophila testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev* **21**, 904-19 (2007).

68. Bainor, A.J. et al. The HDAC-Associated Sin3B Protein Represses DREAM Complex Targets and Cooperates with APC/C to Promote Quiescence. *Cell Rep* **25**, 2797-2807 e8 (2018).

69. Andrejka, L. et al. Animal-specific C-terminal domain links myeloblastosis oncoprotein (Myb) to an ancient repressor complex. *Proc Natl Acad Sci U S A* **108**, 17438-43 (2011).

70. Odajima, J. et al. Proteomic Landscape of Tissue-Specific Cyclin E Functions in Vivo. *PLoS Genet* **12**, e1006429 (2016).

71. Leslie, A.G. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57 (2006).

72.   McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

73.   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

74.   Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

75.   Müller, G.A. & Engeland, K. DNA Affinity Purification: A Pulldown Assay for Identifying and Analyzing Proteins Binding to Nucleic Acids. *Methods Mol Biol* **2267**, 81-90 (2021).

76.   Luger, K., Rechsteiner, T.J. & Richmond, T.J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* **304**, 3-19 (1999).

77.   Yang, J.G. & Narlikar, G.J. FRET-based methods to study ATP-dependent changes in chromatin structure. *Methods* **41**, 291-5 (2007).

78.   Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493-6 (2004).

79.   Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

80.   Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).

81.   Vainshtein, Y., Rippe, K. & Teif, V.B. NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data. *BMC Genomics* **18**, 158 (2017).

82.   Smedley, D. et al. BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).

Chapter 3. Genomic characterization of E2F transcription factors in the context of nucleosomes

Introduction

E2F proteins play a critical role in modulating cellular proliferation both positively and negatively through the function of activator (E2F1, E2F2, E2F3a), repressor (E2F3b, E2F4, E2F5), and the non-canonical repressor (E2F6, E2F7, E2F8) members of the family [1]. Increases in activator E2F abundance or activity through copy number amplifications, loss of Rb or INK4 and CIP/KIP family CDK inhibitors and increases in CDK/Cyclins contribute to unregulated proliferation [1]. There is considerable complexity in deconvolving the specific role for each member of the E2F family since these transcription factors tend to have some redundant and overlapping functions in cell proliferation and their expression is often regulated by other members within the family. Furthermore, different classes of E2F proteins can have various and distinctive functions within specific contexts. For example, in mouse embryonic stem cells, the repressor E2F4, which is classically associated with gene repression, was shown to upregulate proliferative genes [2].

Beyond their well-established role in the cell cycle gene activation, activator E2Fs also play a role in a wide range of cellular activities including metabolism, DNA damage repair, and chromatin organization [1,3]. ChIP studies of E2Fs have shown that these proteins localize to promoters that are not considered S phase genes and E2Fs have been shown to regulate processes such as glucose and fatty acid metabolism in non-proliferating conditions [3, 4]. It has been estimated that E2Fs bind

to 25% of human promoters [4]. How and why E2Fs are recruited to large number of promoters, outside of those that are under cell cycle control is not well understood.

In the context of proliferative genes, E2F target promoters are subject to histone modification and chromatin remodeling [5] (Figure 3.1). Through the transactivation domain (TAD), E2Fs recruit the histone acetyltransferases Tip60 and GCN5 as well as the co-activators CBP/p300 to target gene promoters [6]. The activities of these factors are thought to promote chromatin opening and allow the basal transcription machinery to transcribe target genes. Under conditions of gene-repression, activator E2Fs are directly bound by the Rb protein which blocks the E2F-TAD and prevents E2F-dependent recruitment factors that activate gene expression [6]. Additionally, Rb recruits' chromatin remodelers and histone deacetylases to target gene promoters; this activity is correlated with chromatin compaction [7].

Given that E2F target promoters are subject to several histone modifications as well as chromatin remodeling, we wondered: (I) whether E2Fs could bind nucleosomes (II) whether E2Fs are localized to regions containing the active H3K4me3 mark or (III) whether E2Fs could act as pioneer transcription factors that bind condensed chromatin and serve as adaptors for histone modifiers and chromatin remodelers to open chromatin and allow gene activation. In this chapter I present some preliminary work on uncovering how E2Fs interact with chromatin.

Figure 3.1: General domain architecture of canonical E2Fs. E2F target gene promoters are subject to histone modification and chromatin remodeling. Canonical E2Fs (1-5) have a relatively conserved domain architecture featuring at the N-terminus, a winged-helix DNA-binding domain (DBD) followed by coiled-coil and marked box domain (CM) and a C-terminal transactivation domain (TAD). The CM domain of E2Fs dimerizes with DP proteins and binds the C-terminal domain of the Rb protein. Rb, through its pocket domain, also binds the E2F TAD blocking recruitment of E2F co-activators: Such co-activators include: histone acetyltransferases (HATs), CREB-binding protein (CBP)/p300 both of which have enzymatic acetyltransferase activity, MLL-associated histone methyltransferase (HMT), Set1-HMT, TBP (Tata-box binding protein) and TFIIH a helicase complex implicated in transcription and nucleotide excision repair. Rb also recruits chromatin modifying factors to target genes. Such factors include Histone Deacetylases (HDAC1/2/3), DNA methyltransferase (DMNT1), the mammalian SWI/SNF chromatin remodeler (mSWI/SNF), Sin3B and the methyltransferase Suv39H1.

The E2F family of proteins ChIP with H3 and H4 in asynchronous cycling conditions

We wanted to test whether E2Fs could bind to histone proteins. Previous electromobility shift assays performed by our laboratory and by collaborators have shown that the DNA binding domain of E2Fs (E2F1-DBD/DP1) are able to bind nucleosomes reconstituted on the WIDOM positioning sequence (Data not shown). We wanted to perform an MNase-ChIP experiment combined with immunoblotting to assess whether E2Fs could precipitate histone proteins. We transiently transfected Strep-E2F4, Strep-E2F1, Strep-E2F2, and Strep-LIN9 in asynchronously cycling

HCT116 cells along with Strep-GFP. We then crosslinked chromatin, digested with MNase, performed a StrepTactin pull-down of the soluble fraction and immunoblotted for H3 and H4 (Figure 3.2).

Figure 3.2: Schematic of E2F MNase-ChIP seq and MNase-ChIP blot for histones (A) Strep-GFP, E2F4, E2F1, E2F2, and LIN9 constructs were transiently transfected into HCT116 cells which were asynchronously cycling. After 48 hours cells were harvested and cross-linked and subjected to deep MNase-digestion to isolate mono-nucleosomes. Input samples were taken at this stage. Following digestion, a StrepTactin precipitation was performed to isolate nucleosome bound TFs. This was then used for either immunoblotting or library prep for sequencing. (B) MNase-ChIP blots for Strep tagged E2F constructs along with Strep-GFP and Strep-LIN9 in asynchronous cycling HCT116 cells. Input and Immunoprecipitated panels are indicated at the top and the immunoblotting antibodies are described on the right.

In this experiment, we saw robust histone precipitation with E2F2 as well as histone precipitation with E2F1 and E2F4 that was above the GFP background. We note here the histones could have been pulled down through DNA tethers of long chromatin fragments or the presence of additional binding partners that we did not deplete or blot for. We then wanted to probe the genomic locations of these E2F-precipitated histones. To this end we performed MNase-ChIP seq experiments on E2F1 and E2F2 under asynchronous cycling conditions.

E2F1 and E2F2 MNase-ChIP peaks are distributed throughout promoters in the human genome

We performed an MNase-ChIP experiment after transiently transfecting Strep-E2F1 and Strep-E2F2 constructs in asynchronously cycling HCT116 cells. As a negative control, we transfected cells with the empty strep vector. At the time of harvesting, we treated cells with 1% formaldehyde to crosslink chromatin and performed deep digestion with MNase to digest chromatin down to mono-nucleosomes. After digestion we performed a StrepTactin precipitation to enrich and isolate E2F-bound nucleosome sized fragments. After DNA purification and Illumina sequencing, we aligned the reads to hg38 (bwa-mem) and performed peak calling using MACS2 against the empty-Strep vector pull down as a negative control. We restricted our analysis to fragments that were mono-nucleosome sized (120-200bp, filtered using bamtools) and performed gene ontology analysis using the CISTROME-GO and PANTHER databases [10]. Both PANTHER and CISTROME-GO ontology results agreed; we present the CISTROME-GO analysis of the Strep-E2F1 and Strep-E2F2 experiment in Figure 3.3. E2F1 and E2F2 MNase-ChIP peaks are enriched at genes that broadly belong to similar biological processes with the top terms being regulation of RNA biosynthetic processes, regulation of biosynthesis, regulation of nucleobases, regulation of gene expression, and chromatin organization which agrees with the large overlap of target genes regulated by the two E2Fs.

In order to assess whether the distribution of E2F MNase peaks belonging to different biological processes is an artifact resulting from overexpression of E2Fs in our experiment, we compared our MNase-ChIP peaks to ChIP-seq experiments performed for various cell-lines published and deposited into the ENCODE database

by genome browser visualization. We also compared our MNase-ChIP results to CISTROME-GO analysis of a published E2F1-ChIP dataset performed using an E2F1 antibody precipitation and nuclei sonication in LNCaP cells published by Ramos-Montoya et. al [9, 10]. As shown in Figure 3.4, there are many common GO terms between the ChIP and MNase-ChIP experiments including the top GO biological processes related to DNA metabolism, regulation of RNA biosynthesis, chromatin organization and regulation of gene expression. This agrees with reports that E2F1 is regulator of metabolism and chromatin stability [3, 11]. This result also suggests that peaks precipitated from the MNase-ChIP experiment are localized to genes with known E2F binding sites. Unexpectedly, we did not find an enrichment of cell cycle related processes like S phase regulation, DNA damage, and DNA repair, which I will discuss in the next section.

Under cycling conditions, classical cell cycle genes are not strongly enriched by MNase-ChIP

Interestingly, cell cycle-related processes were not distinctively prominent among the top hits from the E2F MNase-ChIP experiment as was previously observed for MNase-ChIP experiments that probed LIN9-bound nucleosomes in arrested HCT116 cells (discussed in Chapter 2 of this dissertation). While some of the E2F MNase-ChIP peaks did enrich for processes such as regulation of nucleobases which does occur in S phase, biological processes generally belonging to cell cycle proliferation were missing (Figure 3.3). The terms highlighted in green from the ChIP-seq experiment in Figure 3.4A contain genes that are widely considered classical cell cycle genes [12]. This is in sharp contrast to E2F1 MNase-ChIP: while cell cycle genes are enriched when we probe E2F1 binding through ChIP-seq methods, they

appear to be missing in our MNase-ChIP experiment which is designed to precipitate

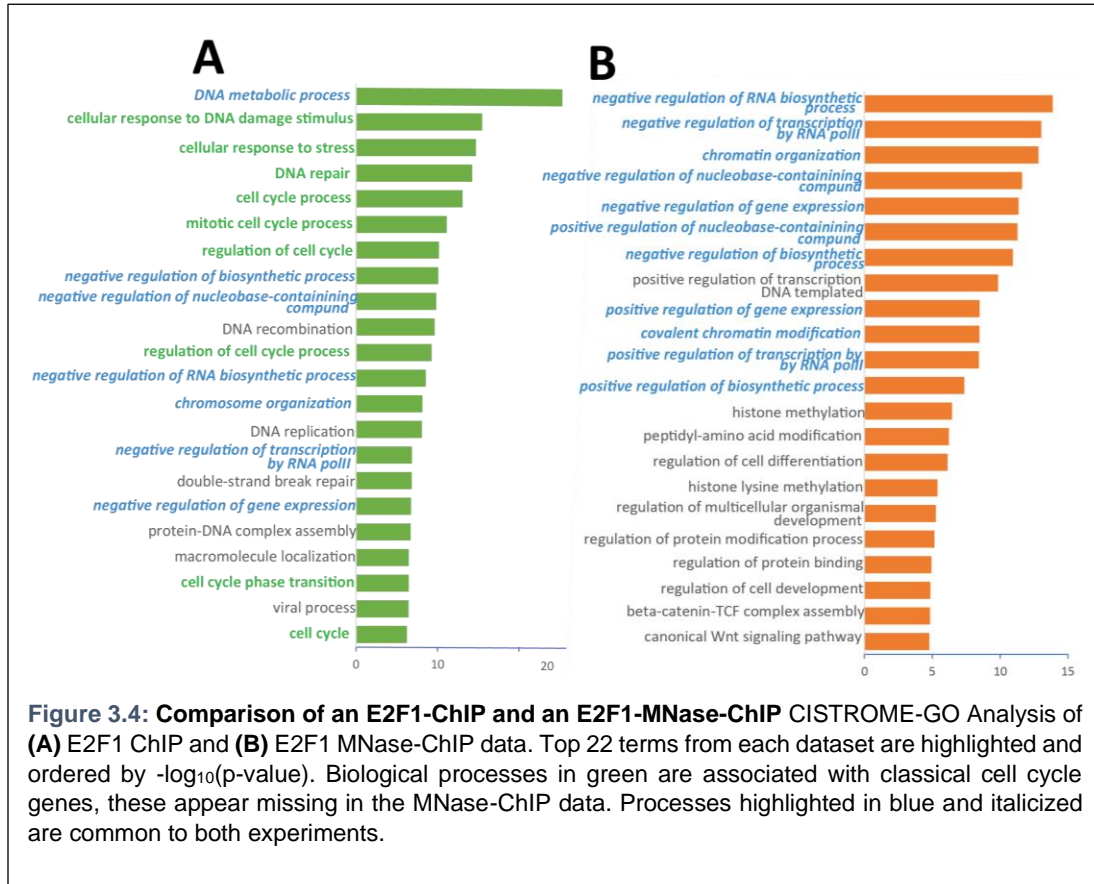E2F-bound nucleosomes rather than DNA association alone.
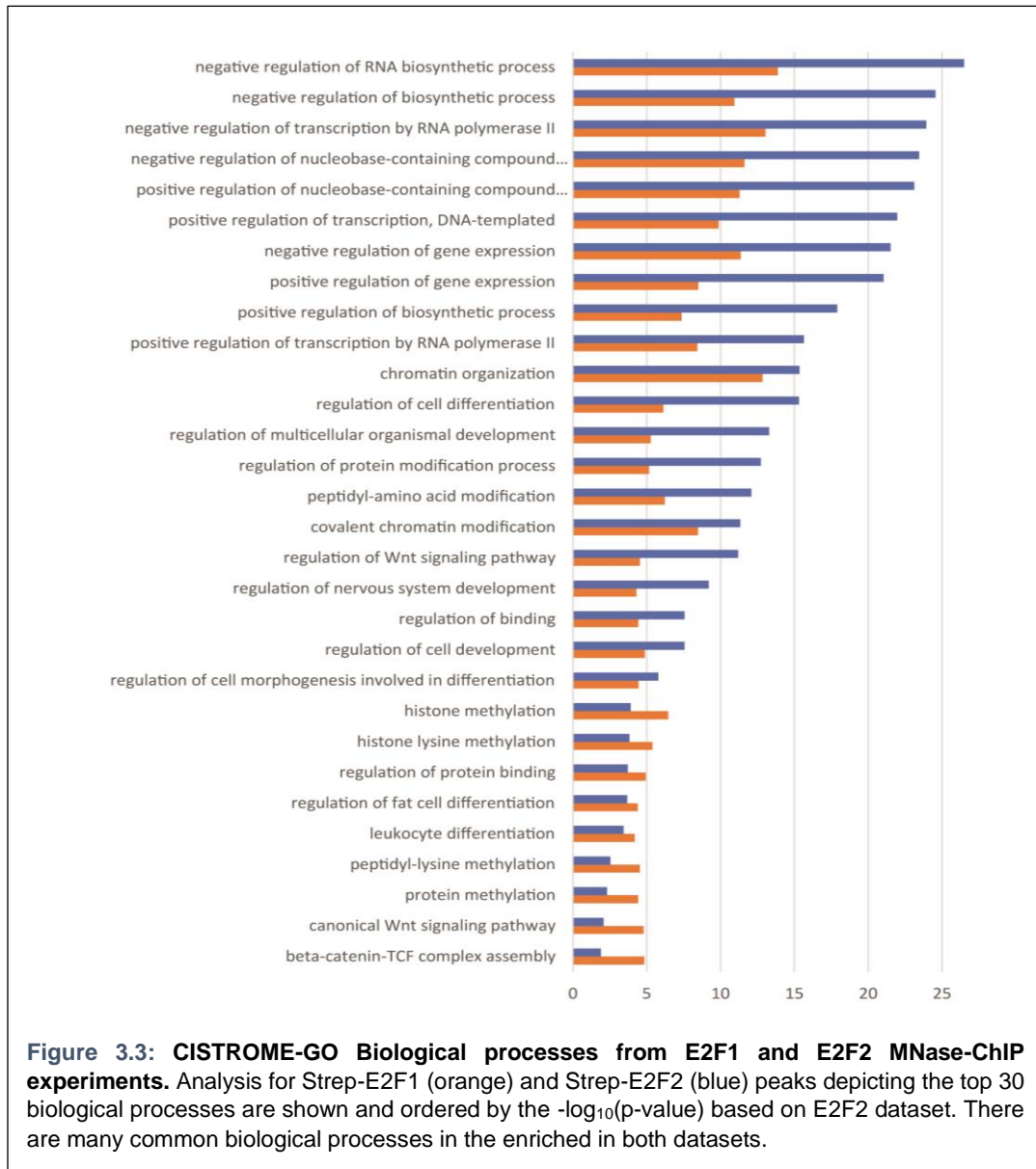
Interestingly, genes belonging to these biological processes were enriched in the

LIN9 MNase-ChIP experiment (Figure 2.7B) but were overall not significantly

enriched in both the E2F MNase-ChIP experiments (Figure 3.4B). The overlap of both

E2F and LIN9 MNase-ChIP experiments, described in Chapter 2 of this dissertation with known cell cycle genes (both G1/S and G2/M genes that contain a well conserved CHR or E2F binding site and respond to p53 stimulation) published by Fischer et al, are shown in Supplementary Figure 3.1A [12]. Out of ~1300 high-confidence, known cell cycle genes, about ~200 and ~45 genes were enriched in the E2F2 and E2F1 MNase-ChIP experiments respectively. In contrast to E2F MNase-ChIP analysis, gene ontology revealed that almost all LIN9 MNase-ChIP peaks belong to biological processes associated with DNA damage and cell cycle regulation.

This result could suggest that perhaps E2F1 and E2F2 are not bound to nucleosomes but rather to non-nucleosomal DNA when target cell cycle genes are actively undergoing transcription–although more detailed analysis and comparisons with a dataset generated under conditions of cellular arrest and under conditions of gene repression would be needed to corroborate this. One possibility that could resolve this puzzling observation is in this class of genes, E2F-nucleosome interactions may be transient and could diminish after potential TAD recruited co-activators have performed some functions on these genes. It could also be possible that classical cell cycle genes are instead bound by other E2Fs such as E2F3 or E2F7/8 under the conditions assayed. Although unlikely, overexpression of E2F1 could cause p53 accumulation which may upregulate E2F7 expression and binding to classical G1/S genes. To fully understand E2F chromatin binding as well as nucleosome interactions, it would be important to perform MNase-ChIP experiments with these additional E2Fs in synchronized and phase-sorted cell populations, preferably with endogenous TFs rather than by overexpression. Finally, we add that due to the wide number of genes that E2Fs bind to, perhaps classical cell cycle genes

may be diluted in our experiment warranting additional replicates and deeper sequencing reads.

Figure 3.4: Comparison of an E2F1-ChIP and an E2F1-MNase-ChIP CISTROME-GO Analysis of **(A)** E2F1 ChIP and **(B)** E2F1 MNase-ChIP data. Top 22 terms from each dataset are highlighted and ordered by -$\log_{10}$(p-value). Biological processes in green are associated with classical cell cycle genes, these appear missing in the MNase-ChIP data. Processes highlighted in blue and italicized are common to both experiments.

E2F1 and E2F2 MNase-ChIP peaks contain nucleosome-sized read densities near TSSs and are adjacent to nucleosomes harboring an H3K4me3 mark

We wanted to probe where E2F1 and E2F2 MNase-ChIP peaks were located in the genome to examine the locations of E2F-bound nucleosomes relative to the TSSs. Preliminary CISTROME-GO analysis revealed that MNase-ChIP peaks were generally confined to promoter regions of genes with 74% and 80.4% of all E2F1 and E2F2 peaks (mono-nucleosome size filtered and above a 5-fold threshold of enrichment) occurring within 1kb of known transcription start sites. We decided to

further probe the location of these MNase-ChIP peaks using NucTools. To probe the relative position of enriched peaks we restricted our analysis to peaks associated with genes that were greater than or equal to 5-fold enriched relative to the empty-Strep immunoprecipitated control. We found that in aggregate both Strep-E2F1 and Strep-E2F2 conditions largely showed enriched sequences near the transcription start sites of occupied genes (Figure 3. A and B top panel). We observed a larger number of genes enriched for the E2F2-precipitated (2027) relative to the E2F1-precipitated (373) data. While we may need to perform additional replicates to corroborate whether this discrepancy is reproducible, the observation seems consistent with our MNase-ChIP immunoblot results from Figure 3.2 which depicted relatively stronger enrichment of both H3 and H4 from Strep-E2F2 relative to Strep-E2F1. Further analysis of the normalized aggregated profiles shows that the E2F1 and E2F2 MNase-ChIP profile is distinct from the Strep-LIN9 ChIPs discussed in Chapter 2 (Figure 2.8 & Supp Figure 2.6). Under conditions of G1-arrest by Nutlin-3a, Strep-LIN9 precipitated samples with the same fold-enrichment cutoff showed (I) a higher normalized occupancy signal (at ~0.8 vs 0.45 for E2Fs) and (II) a distinct enrichment of nucleosome-sized reads immediately downstream of the TSS whereas in the Strep-E2F1 and Strep-E2F2 MNase-ChIP experiments this directionality downstream of the TSS is not obvious.

Figure 3.5 Aggregate profiles and heatmaps of E2F1 and E2F2 MNase-ChIP experiments (A) Normalized aggregate profile of E2F1 (orange) and E2F2 (blue) from MNase-ChIPs. Data were restricted to mononucleosome-sized sequences that were above the 5-fold enrichment cutoff from MACS2 peak caller. **(B)** Heatmaps of Strep-E2F1 precipitated sequences featuring 373 genes. Top panel represents unclustered profile while the bottom panel depicts data clustered (k-means) into 4 bins. **(C)** Same as **(B)** for Strep-E2F2 dataset with 2027 genes.

Based on the aggregate profiles of these experiments, we wondered whether the MNase-ChIP peaks were overlapping with TSS or simply clustered both up and downstream of the TSS. To this end, we performed k-means clustering of both datasets using NucTools. We elected to cluster data into four bins, sorted within 800bp of the TSS and generated heatmaps of the MNase-ChIP peaks (Figure 3.5 B and C, bottom panels). Clustering results indicated that from both experiments, MNase-ChIP reads were generally dispersed immediately up or down-stream of the TSS, or further, between 200-500 bases away. This is again in sharp contrast with the LIN9 MNase-ChIP data in which peaks were uniformly distributed downstream of the TSS under conditions of gene repression.

We also wanted to compare the location of these MNase-ChIP peaks relative to the occurrence of active H3K4me3 mark by performing a CUT&RUN experiment with the same cell line under cycling conditions. While it is known that the presence of the H3K4me3 mark is a signature of active gene transcription and that activator E2Fs are known to associate with the MLL-complex that methylates lysine 4 on histone 3, we wondered whether the activator E2Fs are themselves localized to nucleosomes harboring this mark. We prepared CUT&RUN samples of asynchronously cycling HCT116 cells using an H3K4me3 antibody (cell-signaling) and made libraries using the NEBNext Ultra II DNA library kit. We generated 150bp paired-end sequencing data and utilized bowtie to align reads to the hg38. We then utilized genome browsers to visualize E2F MNase-ChIP reads relative to H3K4me3 occupancy along the human genome. We find that in general E2F1 and E2F2 MNase-ChIP peaks tend to occur in regions free off the H3K4me3 mark in both known cell cycle and non-cell cycle genes (Figure 3.6 and Figure 3.7) suggesting that these activator E2Fs are not likely recruited by the modified histone mark. Instead, the H3K4me3 mark appears to adjacent to most of the E2F MNaseChIP peaks, in both the up and downstream directions. This result supports a model in which E2Fs could bind to DNA or nucleosomes adjacent to the transcription start sites and may then recruit methyltransferases to nearby nucleosomes although we note that several additional experiments will be needed to test this (See Discussion).
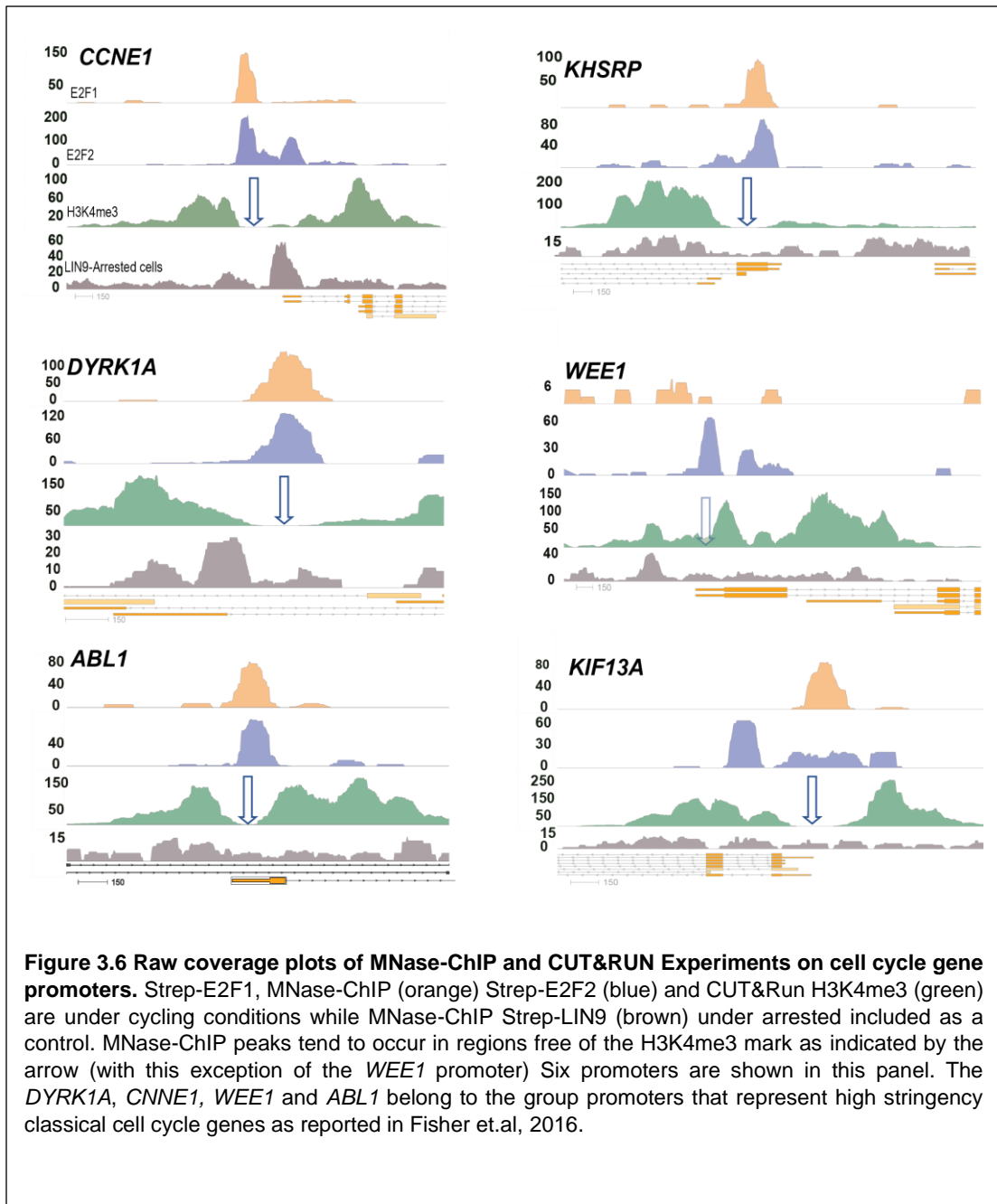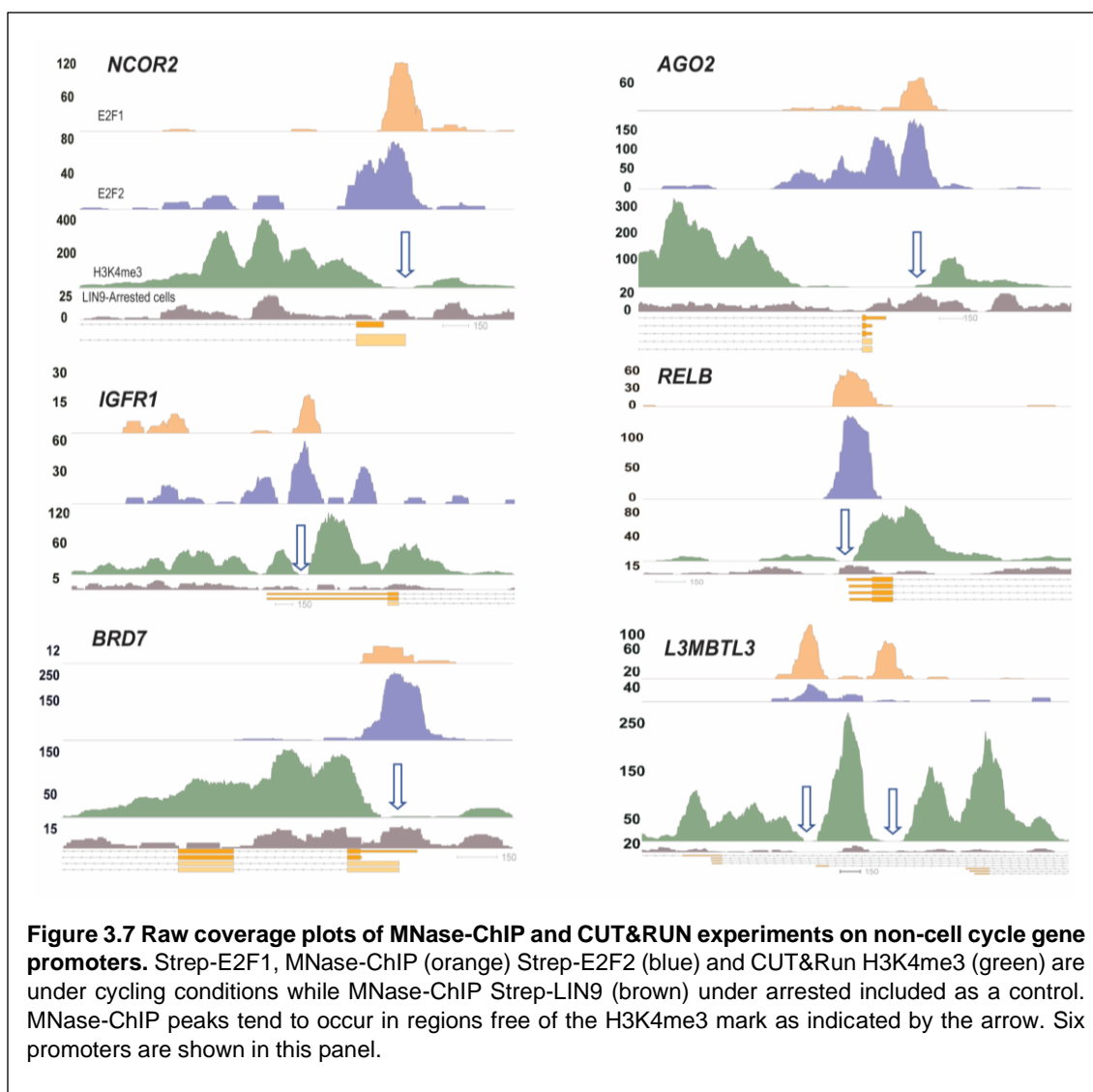
Discussion

The heterogenous nature of our E2F MNase-ChIP profiles, which feature peaks distributed both up and downstream of the TSS could represent the diversity of genes bound by E2F1 and E2F2. E2Fs have been shown to bind many genes
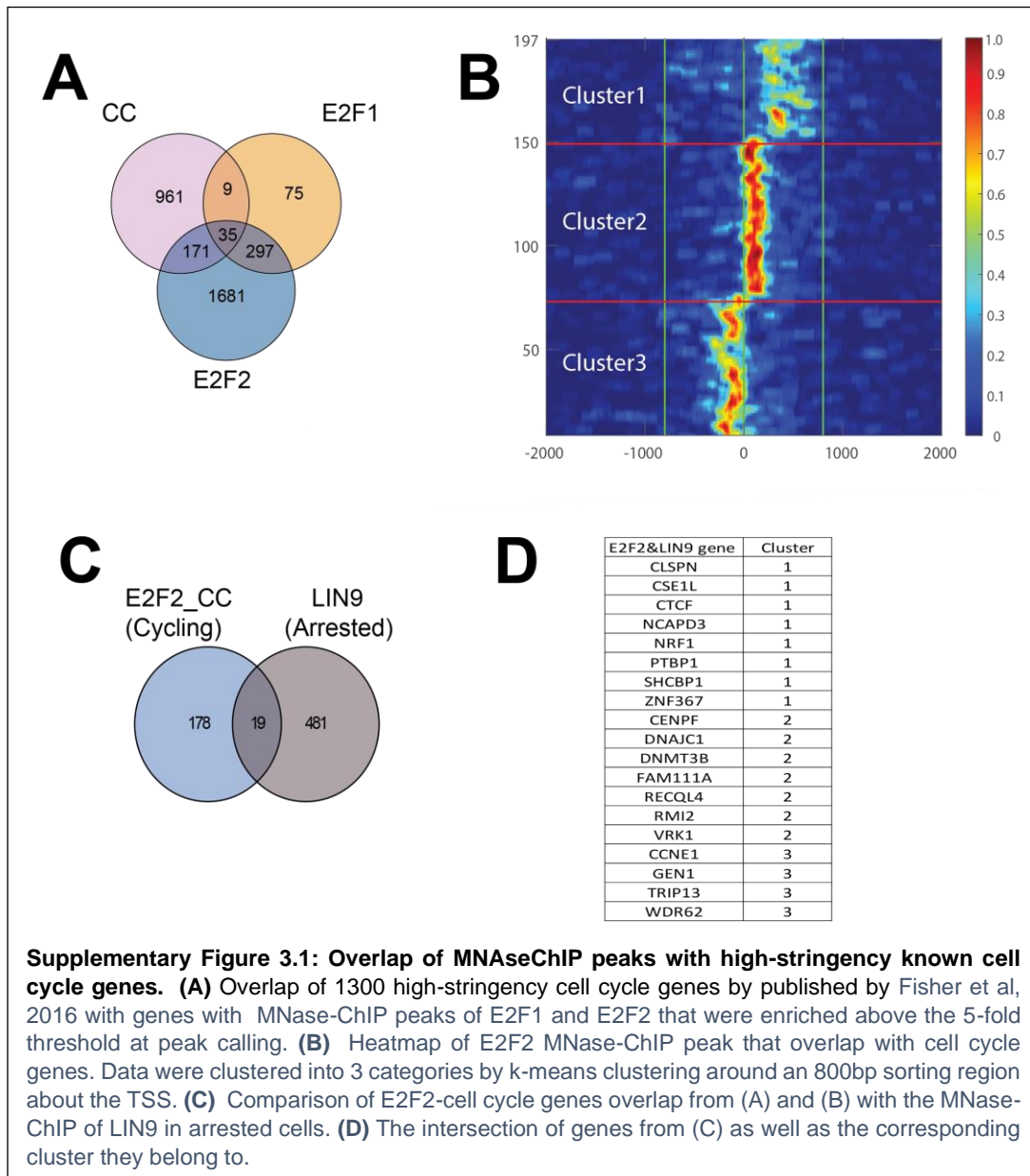
(>2000 promoters) and, as such, the regulation of these genes may entail different mechanisms [4,13]. In addition, activator E2Fs have been reported to bind TFIIH and the transcription initiation complex [10]. ChIP results and analysis from the Farnham lab have shown that E2Fs bind to genes in three separate classes: (I) utilizing the consensus DNA binds motif (TTTSSCGC), (II) a weaker consensus motif (BKTSSCGS), as well as (III) promoter regions lacking this recognition motif [11, 4]. Classical cell cycle genes as well as genes associated with DNA repair fall into the first category where E2Fs directly engage with their consensus recognition sequence [4]. E2F association at promoters falling into the next two categories are speculated to be facilitated by binding of other protein factors that recruit E2F to the partial consensus motif (II) or recruitment of E2F transcription factors by TFIIH and other modules within the transcription initiation complex (III). Characterizing the different modes of how E2Fs engage with chromatin to bind different classes of genes may be useful for cancer therapeutics and will inform how E2Fs activate genes.

It remains an open question as to whether E2Fs bind DNA or nucleosomal-DNA at target genes and then act as pioneer factors to recruit chromatin modifiers. While metagene analysis of the CUT&RUN experiment discussed above is still in progress at the time of this writing, comparing the E2F MNase-ChIP peaks with CUT&RUN experiments using different histone antibodies (i.e H3, H4, H3K9me, H3K9ac, H3K27me2/3) that are found on target gene promoters would be a powerful tool to query the chromatin modification state and architecture at different stages of the cell cycle with high resolution. In addition, comparing MNase-ChIP and CUT&RUN results with RNA-seq at different stages of the cell cycle will further inform how chromatin architecture is tied to E2F-dependent gene activation. Finally, uncovering

86

the structural mechanisms of how E2F associates with nucleosomes will be useful in understanding E2F function.

Figure 3.6 Raw coverage plots of MNase-ChIP and CUT&RUN Experiments on cell cycle gene promoters. Strep-E2F1, MNase-ChIP (orange) Strep-E2F2 (blue) and CUT&Run H3K4me3 (green) are under cycling conditions while MNase-ChIP Strep-LIN9 (brown) under arrested included as a control. MNase-ChIP peaks tend to occur in regions free of the H3K4me3 mark as indicated by the arrow (with this exception of the *WEE1* promoter) Six promoters are shown in this panel. The *DYRK1A*, *CNNE1, WEE1* and *ABL1* belong to the group promoters that represent high stringency classical cell cycle genes as reported in Fisher et.al, 2016.

Figure 3.7 Raw coverage plots of MNase-ChIP and CUT&RUN experiments on non-cell cycle gene promoters. Strep-E2F1, MNase-ChIP (orange) Strep-E2F2 (blue) and CUT&Run H3K4me3 (green) are under cycling conditions while MNase-ChIP Strep-LIN9 (brown) under arrested included as a control. MNase-ChIP peaks tend to occur in regions free of the H3K4me3 mark as indicated by the arrow. Six promoters are shown in this panel.

Supplementary Figure 3.1: Overlap of MNAseChIP peaks with high-stringency known cell cycle genes. **(A)** Overlap of 1300 high-stringency cell cycle genes by published by Fisher et al, 2016 with genes with MNase-ChIP peaks of E2F1 and E2F2 that were enriched above the 5-fold threshold at peak calling. **(B)** Heatmap of E2F2 MNase-ChIP peak that overlap with cell cycle genes. Data were clustered into 3 categories by k-means clustering around an 800bp sorting region about the TSS. **(C)** Comparison of E2F2-cell cycle genes overlap from (A) and (B) with the MNase-ChIP of LIN9 in arrested cells. **(D)** The intersection of genes from (C) as well as the corresponding cluster they belong to.

References

1. Kent, L. N. & Leone, G. The broken cycle: E2F dysfunction in cancer. *Nat. Rev. Cancer* **19**, 326–338 (2019).

2. Hsu, J. *et al.* E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. *Nat. Commun.* **10**, (2019).

3. Denechaud, P. D., Fajas, L. & Giralt, A. E2F1, a novel regulator of metabolism. *Front. Endocrinol. (Lausanne).* **8**, 1–8 (2017).

4. Bieda, M., Xu, X., Singer, M. A., Green, R. & Farnham, P. J. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* **16**, 595–605 (2006).

5. Blais, A. & Dynlacht, B. D. E2F-associated chromatin modifiers and cell cycle control. *Curr. Opin. Cell Biol.* **19**, 658–662 (2007).

6. Frolov, M. V & Dyson, N. J. Molecular mechanisms of E2F-dependent activation and pRB-mediated repression. *Journal of Cell Science* **117**, 2173–2181 (2004)

7. Dick, F. A. & Rubin, S. M. Molecular mechanisms underlying RB protein function. *Nat. Rev. Mol. Cell Biol.* **14**, 297–306 (2013).

8. Pearson, A. & Greenblatt, J. Modular organization of the E2F1 activation domain and its interaction with general transcription factors TBP and TFIIH. *Oncogene* **15**, 2643–2658 (1997).

9. Ramos-Montoya A., Lamb A.D., Russell R., Carroll T., Jurmeister S., Galeano-Dalmau N., Massie C.E., Boren J., Bon H., Theodorou V. et al. . HES6 drives a critical AR transcriptional programme to induce castration-resistant prostate cancer through activation of an E2F1-mediated cell cycle network. *EMBO Mol. Med.* **6**, 651–661 (2014)

10. Li, S. *et al.* Cistrome-GO : a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. *Nucleic Acid Research* **47**, 206–211 (2019)

11. Rabinovich, A., Jin, V. X., Rabinovich, R., Xu, X. & Farnham, P. J. E2F in vivo binding specificity: Comparison of consensus versus nonconsensus binding sites. *Genome Res.* **18**, 1763–1777 (2008).

12. Fischer M, Grossmann P, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Research* **44**, 6070-6086 (2016)

13. Xu, X. *et al.* A comprehensive ChIP – chip analysis of E2F1 , E2F4 , and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res.* 1550–1561 (2007).