Title
Regulation of enhancer dynamics by MLL3/4 in embryonic stem cells

Author
Boileau, Ryan Michael

Publication Date
2023

Peer reviewed|Thesis/dissertation

The regulation of enhancer dynamics by MLL3/4 in embryonic stem cells

by
Ryan Boileau

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Developmental and Stem Cell Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Daniel Lim
FC7A2B6D45EF428...

Daniel Lim
Chair

DocuSigned by:
Benoit Bruneau
DocuSigned by:

Benoit Bruneau

Robert Blelloch
91DFF4C4A2FF4FC...

Robert Blelloch

Committee Members

Dedicated to my parents Michael and LeeAnn and big brother Danny

Acknowledgements

Throughout my career I feel extremely fortunate to have met extraordinary people of all kinds that I can now call mentors, colleagues, and friends. This thesis would not be possible without their help to overcome the challenges inherent to a life in the sciences. Their influence, both direct and indirect, pushed me to become a better scientist, mentor, and human being. I thank many for providing the opportunities that built a future where I can continue being a life-long student for myself and a life-long mentor for others.

Firstly, I thank my graduate advisor, Robert Blelloch. In Robert's lab had the freedom and independence to succeed (and fail) in the ways I decided and had Robert's support the entire time. Among many examples, this support can be much needed encouragement to dig deeper into the data until a clear picture of the biology emerges. His support was not limited to the lab space, outside of which he has also kindly helped me out of a variety of difficult situations. One of the most useful skills I've acquired from working with Robert is constructive scientific debate, which appears to be one of his favorite activities. I admire his eagerness to hear about new results and learn about the smallest technical details of an experiment. Robert's relentless scientific curiosity, optimism, and championing of trainees set incredible examples for me that I intend to carry forward in my career.

I would also like to acknowledge other faculty that have played major roles in my graduate career. I thank Dan Lim and Benoit Bruneau for serving on my thesis committee. As I had hoped, they provided excellent and honest feedback on my projects. Similarly, they provided a necessary, measured perspective that accelerated my career as a scientist. I thank Licia Selleri for her unending support throughout my graduate years

including valuable critique during my supergroup presentations, career wisdom, and frequent laughter during our coffee chats. I thank Elphège Nora for insightful thoughts on my project, my career, and the opportunity to work with the talented students in his lab. I would also like to specifically thank Yin Shen and Andrew Brack who have played significant roles in supporting my graduate career.

My fellow lab members have been a constant source of help throughout graduate school. One consistent support pillar in the lab for me has been Cassie Belair who has always greeted my questions with patience and a smile. I have the utmost appreciation for how often Cassie not only takes care of lab matters but also takes care of the people working in it. I would also like to thank Brian DeVeale for being generous in his feedback, his kindness, and being a genuinely good human. It was an honor to mentor Kevin Chen and I am very proud to see his hard work pay off when he was admitted to his top choice for graduate programs. Kevin and I both learned a lot from each other during our work together and I've become a better mentor and scientist because of Kevin. As well, I have been lucky to interact with plenty of other amazing people in my lab over the years including but not limited to: Carolyn Sangokoya, Jake Freimer, Kayla Lenshoek, Hannah Driks, Li Wang, Hauigeng Xu, Chisato Kamei, Jiuling Yang, Jason Soriano, Tayma Handal, and Brandon Chu.

Two people from the Blelloch Lab deserve special mention for playing defining roles in my graduate schooling. First, I would like to thank Deniz Gökbuget for hosting me as a rotation student and continuing to play the role of an in-lab mentor for me ever since. I have learned an incalculable amount from Deniz in many areas: experimentalism, following scientific instincts, and effort with intentionality. I will miss our debates and

sharing of American internet culture. I'd like to thank Bryan Marsh having been my closest colleague in the lab. We joined the same program, the same lab, and sat back-to-back in the lab for almost the entirety of graduate school. I treasure many moments that I've had with Bryan, especially casual conversations skirting the lines of absurdity. I've learned a lot by following Bryan in his successes throughout grad school.

I had the honor of conducting my graduate school with the Developmental and Stem Cell Biology program at UCSF. Our small, focused program led me to befriend many amazing, talented, and diverse young scientists. As well, program events were often a much need reprieve from the bench. I do not have the space to list all of the awesome people. Special gratitude to Ryan Samuel, Matthew Schmitz, Karissa Hansen, and Eliza Gaylord. As well as those in other programs: Ramiro Patiño, Alexis Leigh Krup, and many more. Thank you for being you and for all the laughs.

Lastly, I thank my family who have all provided unconditional love throughout my life and gave me opportunities to choose my own adventures. I thank my grandmother for exposure to the joys of learning so early on and inspiring me to become a teacher. I thank my brother for letting me tag along with his friends back in the day and exposing me to the Wu Tang Clan. I thank my sister-in-law, Emily, who has been an amazing, welcoming, and warm addition to our family. I thank my mother for her unending patience throughout my teenage years and making our home a safe place to develop my own identity. I am thankful for, and proud of, my father for working so diligently to give me and my brother a childhood free of the hardships that he experienced in his own.

Contributions

The research that comprises this thesis was conducted under the supervision of Dr. Robert Blelloch at the University of California, San Francisco. This work was supported by funding from Achievement Rewards for College Scientists (ARCS) Foundation, the UCSF Discovery Fellowship, NIH T32HD007470, and the NIH R01GM122439.

Chapter 2 describes work published in Genome Biology in 2023 with the following authors: Ryan M. Boileau, Kevin X. Chen, Robert Blelloch. Kevin and I performed the experiments and I analyzed all of the data. Robert and I wrote the manuscript.

Chapter 3 describes unpublished, ongoing work. Current contributors to the work are: Ryan M. Boileau, Kevin X. Chen, and Robert Blelloch. Kevin and I conducted experiments. I analyzed all of the data.

I wrote this thesis with input from Robert Blelloch.

Regulation of enhancer dynamics by MLL3/4 in embryonic stem cells

Ryan M. Boileau

Abstract

The acquisition of cell fate is dependent on gene regulatory networks that are regulated spatiotemporally by cell type specific transcription factors (TFs). In binding to a class of cis-regulatory elements known as enhancers, TFs stimulate transcription at target genes. Studies on enhancers have revealed that enhancers exist in multiple different states and interconvert between them. The establishment of an active, transcriptionally promoting enhancer state from an inactive state is a stepwise process initiated by TFs and then facilitated by chromatin regulators. Here, we investigate the molecular processes required to establish an active, transcriptionally promoting enhancer state. To study enhancer activation, we utilize an in vitro embryonic stem cell differentiation system known as the naive to formative transition that recapitulates gene regulatory events that occur in vivo during mouse early embryogenesis. We first investigate the generalizability of current molecular models of enhancer activation in Chapter 2. Using the naive to formative transition and genetic deletions we find that the homologous enzymes MLL3/4 (KMT2C/D) are required for activation of some but not all enhancers as previously thought. Moreover, surprisingly, there is an underwhelming impact on gene expression changes during the transition despite loss of MLL3/4 and the loss of “active” molecular signatures at many enhancers. This work demonstrates the existence of multiple modes of enhancer activation and suggests a more cell-context specific role for the key chromatin regulators MLL3/4 than known before. In ongoing work in Chapter 3, we are investigating the current prevailing molecular model of enhancer activation more deeply.

We built a new synthetic system using the TF Grhl2 for rapid and conditional induction of enhancer activation. We are currently employing this system to identify the mechanistic relationships between Grhl2, MLL3/4, other key chromatin regulators, and transcription. Our discoveries on the molecular fundamentals of enhancer state dynamics serve as a foundation on which to better understand drivers of human development and disease.

Table of Contents

CHAPTER 1 : INTRODUCTION 1

EMBRYONIC STEM CELLS AS A MODEL FOR STUDIES OF GENE REGULATION 3

MOLECULAR FEATURES OF ENHANCERS AND METHODS OF THEIR DETECTION 6

REGULATING ENHANCER FUNCTION 9

MLL3 and MLL4: Enhancer histone methyltransferases 9

Enhancer H3K4 methylation 11

CBP and P300: Enhancer acetyltransferases 13

Enhancer histone acetylation 14

THE CANONICAL MODEL OF ENHANCER ACTIVATION 15

THE INTERDEPENDENCE OF CHROMATIN REGULATORS IN ENHANCER ACTIVATION 17

PREMISE OF STUDIES 18

FIGURES 20

REFERENCES 23

CHAPTER 2 : LOSS OF MLL3/4 DECOUPLES ENHANCER H3K4 MONOMETHYLATION, H3K27 ACETYLATION, AND GENE ACTIVATION DURING EMBRYONIC STEM CELL DIFFERENTIATION 40

SUMMARY 40

INTRODUCTION 41

RESULTS 43

DISCUSSION 59

CONCLUSIONS 65

METHODS 66

FIGURES 77

CHAPTER 3 : THE ROLE OF MLL3/4 IN GRHL2 MEDIATED ENHANCER ACTIVATION 109

INTRODUCTION 109

RESULTS 112

DISCUSSION 119

MATERIALS AND METHODS 121

FIGURES 128

REFERENCES 132

CHAPTER 4 : CONCLUSIONS 137

CONCLUDING REMARKS 144

REFERENCES 145

List of Figures

FIGURE 1.1 – ACTIVE ENHANCERS DRIVE TRANSCRIPTION OF TARGET GENES..... 20

FIGURE 1.2 – ACTIVE ENHANCER HISTONE MODIFICATIONS ARE MEDIATED BY COMPASS-LIKE COMPLEX AND P300/CBP. 21

FIGURE 1.3 - THE CANONICAL MODEL OF ENHANCER ACTIVATION FACILITATED BY MLL3/4. 22

FIGURE 2.1: MLL3/4 IS DISPENSABLE FOR TRANSCRIPTIONAL ACTIVATION OF MUCH OF THE FORMATIVE PROGRAM. 77

FIGURE 2.2 – SUPPLEMENT FOR MLL3/4 IS DISPENSABLE FOR TRANSCRIPTIONAL ACTIVATION OF MUCH OF THE FORMATIVE PROGRAM..... 79

FIGURE 2.3 - MLL3/4 IS REQUIRED FOR ALL DYNAMIC H3K4ME1 DEPOSITION DURING PLURIPOTENT TRANSITION. 81

FIGURE 2.4 – SUPPLEMENT FOR MLL3/4 IS REQUIRED FOR ALL DYNAMIC H3K4ME1 DEPOSITION DURING PLURIPOTENT TRANSITION. 83

FIGURE 2.5 - MLL3/4 DEPENDENT AND INDEPENDENT DISTAL H3K27AC DEPOSITION. 85

FIGURE 2.6 – SUPPLEMENT FOR MLL3/4 DEPENDENT AND INDEPENDENT DISTAL H3K27AC DEPOSITION. 87

FIGURE 2.7 - ENHANCER ACTIVATION CAN OCCUR INDEPENDENTLY OF MLL3/4. 89

FIGURE 2.8 – SUPPLEMENT 1 FOR ENHANCER ACTIVATION CAN OCCUR INDEPENDENTLY OF MLL3/4. 90

FIGURE 2.9 – SUPPLEMENT 2 FOR ENHANCER ACTIVATION CAN OCCUR INDEPENDENTLY OF MLL3/4. 92

FIGURE 2.10 - DISTAL H3K4ME1 AND H3K27AC ARE NOT FUNCTIONALLY COUPLED WITH FORMATIVE TRANSCRIPTIONAL ACTIVATION. 94

FIGURE 2.11 – SUPPLEMENT FOR DISTAL H3K4ME1 AND H3K27AC ARE NOT FUNCTIONALLY COUPLED WITH FORMATIVE TRANSCRIPTIONAL ACTIVATION. 96
FIGURE 2.12 - GENE-CENTRIC ANALYSIS REVEALS A SUBSET OF DISTAL LOCI ASSOCIATE WITH MLL3/4 DEPENDENT FORMATIVE GENES. 97
FIGURE 2.13 – SUPPLEMENT FOR GENE-CENTRIC ANALYSIS REVEALS A SUBSET OF DISTAL LOCI ASSOCIATE WITH MLL3/4 DEPENDENT FORMATIVE GENES. 98
FIGURE 3.1 - MLL3/4 IS REQUIRED FOR H3K4ME1 AND H3K27AC DEPOSITION AT GRHL2 SITES IN THE FORMATIVE STATE. 128
FIGURE 3.2 - A SYNTHETIC SYSTEM FOR INDUCTION OF GRHL2 ACHIEVES RAPID ENHANCER BINDING AND GENE ACTIVATION 129
FIGURE 3.3 - MLL3/4 IS NOT REQUIRED FOR TRANSCRIPTIONAL ACTIVATION UPON ACUTE GRHL2 INDUCTION 130
FIGURE 3.4 - MLL3/4 IS REQUIRED FOR GRHL2 MEDIATED H3K4ME1 AND H3K27AC DEPOSITION. 131

Chapter 1 : Introduction

During mammalian fertilization, a sperm cell and an oocyte fuse to form a single cell known as the zygote. This single cell and its resulting progeny divide and give rise to successive generations of cells. In the process these early cell types lose their old identities and gain new ones, transitioning into diversified cell types from the early embryonic to adult stages. All cells and cell types necessary for the adult mammal are generated in this process of development. Completion of development, however, requires that the time and place in which cell identity transitions occur is carefully regulated. Without an appropriate number and variety of functional cells, the tissues they make up may otherwise be severely compromised. Errors in the control of development can have catastrophic consequences in many cases including embryonic lethality, congenital birth defects, and diseases like autoimmunity or cancer (Deciphering Developmental Disorders Study 2017; Frank and Nowak 2003; Heward and Gough 1997; Spielmann et al. 2022). Consequently, by studying development and the fundamentals of cell identity acquisition we vastly improve our understanding of the molecular underpinnings of normal physiology and pathology.

All cells of the mammalian organism have the same genome with few exceptions. How diversified cell identities or “fates” arise from identical genetic information during development is due to differential gene regulation including control of transcription. A class of proteins that directly coordinate transcription are known as transcription factors (TFs). These proteins bind to sequence specific DNA motifs and recruit additional protein complexes that act to regulate transcription of specific gene targets positively or

negatively. Depending on cellular context, a single active TF can drive the expression of a network of genes which direct a particular cell identity. The complexity of TF function is furthered by the presence or absence of cofactor proteins which may alter the target genes of a TF. Collectively, the spatiotemporal regulation of gene transcription is achieved through cell-type specific expression and activity of multiple TFs and the coordinated and layered activity resulting from the ensemble of their cofactors. These molecular events in concert underlie cell fate and function.

A necessary component for TF activity is the DNA elements they bind to perform their function. At least two major types of cis regulatory elements are recognized and bound by TFs. The first type is the promoter region of a gene. The promoter is the site of initiation for transcription of a gene; it resides directly upstream of the transcript sequence and dictates the directionality of transcription through the orientation of DNA motifs recognized by general transcription factors. The second type is known as an enhancer. Canonically, an enhancer upregulates transcription of a gene by targeting the promoter from a distance regardless of its orientation to the target gene. Enhancers operate as binding platforms for sequence specific TFs which recruit cofactors and RNA polymerase to the target gene promoter (Figure 1.1). While most genes only have one promoter, a single gene can have multiple enhancers. Also, while the common view is that an enhancer has one specific target promoter a single enhancer region can switch and or target additional promoters. Collectively, through the myriad ways in which enhancers operate, they play critical roles in facilitating proper gene expression.

The following introduction will cover the biological systems used in this thesis and relevant molecular biology of enhancers. First, we will describe embryonic stem cells and

how we use them as systems to study embryonic development and gene regulation. Afterwards, we will discuss the molecular features associated with enhancers, the protein machinery that facilitates their implementation at enhancers, and their functional interdependence.

Embryonic stem cells as a model for studies of gene regulation

Embryonic stem cells (ESCs) are derived from the *in vitro* culture of the inner cell mass of early-stage embryos. In mice, these cells arise at embryonic day 3.5 (E3.5) post-fertilization and before implantation in the uterus. The defining features of an ESC are that they self-renew in culture and display pluripotency - the capacity to transition into all other cell types of the adult animal. This latter feature of pluripotency is often demonstrated by *in vitro* differentiation protocols or by production of teratomas when injected into animal subjects. The ultimate test of pluripotency, however, is the production of chimeras and transmission to the germ line when ESCs are injected into embryos.

Most cultured cell lines are derived from cancer tissue or from cells that are transformed by oncogenic viruses, where chromosomal aberrations may be selected for and fluctuate throughout time in culture. In contrast, ESCs offer a physiologically relevant, genetically stable, and karyo-normal alternative to these other commonly used cell lines. However, it should be noted that ESCs have unique cellular processes relative to other cell types. For instance, mouse ESCs (mESCs) have a uniquely structured cell cycle and also display increased DNA damage markers associated with high levels of transcription relative to somatic cell types(Percharde, Bulut-Karslioglu, and Ramalho-Santos 2017;

Efroni et al. 2008). Nonetheless, ESC culture provides the opportunity to ask many fundamental questions about biology.

Historically, mESCs have been cultured in media that includes Leukemia Inhibitory Factor (LIF)(A. G. Smith et al. 1988). Activating STAT signaling through LIF binding to LIF receptor is required to maintain mESCs in a proliferative, pluripotent state(Niwa et al. 1998). However, studies since have demonstrated that mESCs cultured with LIF alone are heterogeneous in cell identity showing signs of early differentiation and thus have been described as metastable(Marks et al. 2012; Singer et al. 2014). Currently, mESCs are commonly cultured with LIF and the addition of two inhibitors (2i) that agonize WNT signaling and antagonize MEK signaling(Mulas et al. 2019). In LIF+2i culture conditions, mESCs homogeneously adopt a “naive” state which is highly similar to the early epiblast cells of an E4.5 embryo(Marks et al. 2012).

Starting from the naive state, removal of LIF+2i from the culture, causes mESCs to initiate a cell state transition toward differentiation. Over approximately two days these cells will homogeneously transition from the naive state to a “formative” state. This pluripotent transition highly recapitulates cellular processes and gene regulatory events that occur between preimplantation (E4.5) and post-implantation epiblast cells (E5.5). The transition is marked by decreased expression of Rex1, Klf4, and Esrrb, and upregulation of Otx2, Oct6, and Grhl2, which, respectively mark the dissolution of naïve pluripotency and activation of the formative transcriptional program(P. Yang et al. 2019; A. Smith 2017). Correspondingly, enhancers and factors that regulate these genes are turned off and on. Among other critical processes X chromosome inactivation and global gains in DNA methylation are also initiated in the formative state and are features inherent to

somatic cell types. These events together support the hypothesis made by Austin Smith that the formative state represents an essential executive phase in pluripotency which is required for all future lineage specification events in development.

In addition to developmental relevance, the naive to formative transition has emerged as a commonly used *in vitro* system to study gene regulation because of the experimental feasibility of mESCs and the rapid and homogeneous nature of the transition. The epigenetic basis of transcriptional regulation has been one area investigated using the naïve to formative transition. As in other systems, TFs during the naïve to formative transition perform important functions to prepare for the changes in transcriptional programs that occur in differentiation. Several TF drivers of the formative state have been identified. For example, the formative state requires OTX2-mediated rearrangement of enhancer binding by OCT4 to new sites such as the enhancers for Fgf5(Buecker et al. 2014; S.-H. Yang et al. 2014). Premature expression of Otx2 is sufficient to drive the activation of many formative genes. We found that in the formative state the TF GRHL2 is also upregulated and activates several hundred enhancers and target genes(A. F. Chen et al. 2018). In doing so, GRHL2 maintains gene expression for a subset of gene targets for Klf2/Klf4 thereby maintaining expression for genes that would otherwise be deactivated in the absence of these naïve TFs. Interestingly, a transient increase in the TF Zic3 during the transition also activates a part of the formative transcriptional program, including Grhl2 and Fgf5(S.-H. Yang et al. 2019). Additionally, the transcription factor Foxd3 plays dual roles during the naïve to formative transition by both maintaining enhancer activity and priming new enhancers for later function (Krishnakumar et al. 2016). By using the naive to formative transition as a system to

understand gene regulation, these studies not only provide the opportunity to better understand the biology of early development but also allow for a focus on fundamental biological questions, including those related to enhancer-based gene regulation.

Molecular features of enhancers and methods of their detection

Cis-regulatory elements are generally classified and assessed based on their molecular features. One of the most studied and utilized features historically to demarcate the genome are the post translational modifications of histones, such as methylation, acetylation, and phosphorylation. The “histone code” hypothesis proposes different combinations of histone modifications distinguish types of cis-regulatory elements(Jenuwein and Allis 2001). For enhancer regions these include mono/di-methylation of Histone 3 (H3) at lysine 4 (H3K4me1/2) and acetylation of H3 at lysine 27 (H3K27ac)(see Figure 1.1)(Heintzman et al. 2007). Besides histone modifications, specific histone variants have also been associated with active enhancer regions. For instance, a variant of H3 known as H3.3 is also deposited at specific genomic regions including active enhancers and promoters(Ahmad and Henikoff 2002; P. Chen et al. 2013).

The characterization of histone modifications, histone variants, and chromatin binding proteins has been achieved through the use of chromatin immunoprecipitation and sequencing (ChIP-seq). In this approach, proteins are first chemically fixed to DNA using paraformaldehyde. Then, chromatin is sheared using sonication down to small fragments of ~200bp in length. Sonicated chromatin bound to a target protein is isolated using an antibody specific to the target. When these isolated fragments of interest are

sequenced and aligned to a genome their enrichment indicates regions bound by the target protein or harboring histone modifications of interest. Recent alternative methods to ChIP-seq involve permeabilizing cells without fixation and then using antibody directed enzymatic digestion or transposition. These approaches, called CUT&RUN and CUT&Tag respectively, offer several advantages to ChIP-seq including the absence of artifacts due to fixation conditions, higher signal to noise, and reduced experimental cost(Kaya-Okur et al. 2020; 2019).

Chromatin accessibility is also a prominent physical feature of enhancers. Accessible chromatin represents gaps in the histone packaging of chromatin. Because the default state of DNA is to be packaged by histones into nucleosomes, the absence of nucleosomes in a particular region is believed to represent ongoing biological activity such as transcription or the binding of a transcription factor to its cognate motif.

The measurement of accessibility is often conducted using the Assay for Transposase-Accessible Chromatin and sequencing (ATAC-seq)(Buenrostro et al. 2013). In this method, nuclei are extracted from cells and incubated with a Tn5 hyperactive transposase. The transposition activity of Tn5 is heavily biased towards places where it can penetrate through chromatin DNA packaging and transpose on accessible DNA. Transposition fragments the DNA and ligates oligo DNA tags on the target DNA of interest. Purification and amplification by PCR for these tags allows for the sequencing and identification of regions of accessible chromatin. Ease of use and high sensitivity has made ATAC-seq a leading method for accessibility measurements over other methods such as MNase-seq and DNAse-I hypersensitivity assays.

The genome is compartmentalized in non-random ways, and proper packaging of chromosomes is required to maintain cellular functions(de Wit and Nora 2023; Nora et al. 2012). In the 3D space of the nucleus, chromatin looping between enhancers and promoters occurs. This physical proximity allows enhancers to regulate target promoters at extreme linear distances on a chromosome – in some cases, up to 1 megabase or more away(Lettice et al. 2003; Schoenfelder and Fraser 2019). Therefore, contacts detected between enhancers and promoters is predictive of enhancer function.

The most common approach for elucidating chromosome structure is through chromatin conformation capture assays. In this approach, chromatin-chromatin interactions are cross-linked using formaldehyde and digested using restriction enzymes. Cross-linked and digested chromatin is then ligated under low DNA concentrations or within intact nuclei to favor ligation between nearby interacting chromatin fragments. Amplifying, sequencing, and mapping the resulting ligation junctions enriches for foci of interacting chromatin. The enzyme used during chromatin digestion can greatly affect the resolution at which chromatin interactions can be detected by allowing the isolation of smaller interacting fragments. Indeed, a recent advance in chromatin conformation capture approaches called Micro-C XL has even utilized MNase to achieve nucleosome-level resolution(Hsieh et al. 2016). Interestingly, results from these approaches suggest that contact between a distal region and promoter regions is not a guarantee of enhancer based increases in transcription(Bantignies et al. 2011). Moreover, enhancer function does not depend on stable looping events(Alexander et al. 2019). The resolution of interacting regions and signal to noise with chromatin conformation capture assays is still

improving. It will be interesting to revisit the necessity and sufficiency of enhancer-promoter contacts with further advances in chromatin conformation capture approaches.

Regulating enhancer function

Molecular features of an enhancer are mediated by chromatin regulating complexes which themselves play instrumental roles at enhancers. Originally in reference to histone modifications, chromatin regulators broadly fall into classes that read, write, and erase molecular features of chromatin. Two pairs of homologous enzymes, MLL3/MLL4 methyltransferases and CBP/P300 acetyltransferases, are key regulators that promote enhancer activity and catalyze enhancer H3K4me1 and H3K27ac respectively. In the following sections, we describe further the functions proposed for MLL3/4, H3K4me1, CBP/P300, and finally H3K27ac.

MLL3 and MLL4: Enhancer histone methyltransferases

The COMPASS and COMPASS-like complexes form around the core subunits MLL1/MLL2 (KMT2A/B) and or MLL3/MLL4 (KMT2C/KMT2D) respectively. All MLL proteins contain a SET domain with H3K4 methyltransferase activity but have distinct function. The COMPASS complex deposits H3K4me2/3 primarily at promoters while COMPASS-like complex catalyzes H3K4me1/2 at distal enhancers(Hu et al. 2013). MLL3 and MLL4 are both large ~550kDa enzymes with a high degree of protein homology. The degree of similarity between MLL3 and MLL4 suggest, consistent with studies, that MLL3/4 function with at least partial redundancy in gene regulation (Lee et al. 2013). As

key regulators of enhancers, MLL3 and MLL4 function has been shown to be essential for a number of cell fate transitions including those in early development, adipogenesis, myogenesis, and cardiogenesis(Lee et al. 2013; C. Wang et al. 2016; Ashokkumar et al. 2020; Ang et al. 2016). These proteins are also important in disease as loss of function of MLL3 is associated with Kleefstra's syndrome(Kleefstra et al. 2012) and MLL4 loss causes Kabuki Syndrome(Ng et al. 2010). Additionally, Heterozygous loss of the MLL3 or MLL4 proteins is one of the most common mutations associated with cancer(Mendiratta et al. 2021; Sze and Shilatifard 2016).

The structure of the intact mammalian COMPASS-like complex with all subunits is incompletely known but domain interactions between MLL3/4 and the main COMPASS subunits have been characterized. The WIN motif at the C-terminal end of MLL3/4 physically interacts with WDR5 which then mediates assembly with RBBP5, ASH2L and indirectly with DPY30(Sze and Shilatifard 2016). These proteins, known together as "WRAD", are core subunits of both the COMPASS and COMPASS-like complexes, and are required for efficient catalytic activity(Avdic et al. 2011). However, unlike the SET domain of MLL1 the MLL4 SET domain retains a minor level of methyltransferase activity in the absence of WRAD(Y. Zhang et al. 2015). Other subunits associated with the COMPASS-like complex are PAGR1A, PTIP, NCOA6, and UTX (KDM6A). Of note, the subunit UTX demethylates for H3K27me3 which is a repressive mark at enhancers and promoters. Consequently, UTX in complex with MLL3/4 complements the stimulatory role in transcription of the COMPASS-like complex(Hong et al. 2007).

Recent studies have begun to uncouple the roles of MLL3/4 catalytic activity and their function as a protein scaffold with surprising results. Loss of MLL3/4 catalytic activity

either by point mutation of key catalytic residues or deletion of the SET domains does not result in loss of viability or major transcriptional changes in steady state ESCs(Rickels et al. 2017; Dorighi et al. 2017; Cao et al. 2018). In Drosophila, development into adulthood is unimpacted by the loss of catalytic activity of Trr, the ortholog of Mll3/4(Rickels et al. 2017). These findings are in contrast with those found with genetic deletion of MLL3/4 protein which results in substantial transcriptional effects. Therefore, it has been suggested that MLL3/4 and not H3K4me1 at enhancers are functionally important. Interestingly, human mutations in the SET domain are moderately enriched in Kabuki Syndrome and malignancies(Bögershausen et al. 2016; Tate et al. 2019). Further, defective H3K4 methylation activity in MLL4 or MLL3/4 leads to mouse developmental abnormalities in neural crest tissue and extraembryonic tissue(Bjornsson et al. 2014; Xie et al. 2022). However, catalytic activity also seems to be important for MLL4 protein stability(Y. Jang et al. 2017). Whether the developmental and disease phenotypes from SET domain mutations are due to loss of catalytic activity or a partial loss in MLL4 protein levels is not yet known.

Enhancer H3K4 methylation

Though evidence favors a dispensability for catalytic activity from MLL3/4, H3K4me1 at enhancers may still have important functions. Importantly, studies assessing the role of H3K4me1 at enhancers do so by perturbing catalytic activity of MLL3/4 which is not the sole source of intergenic H3K4me1. For example, in ESCs mass spectrometry suggests at least 50% of H3K4me1 persists in the absence of MLL3/4 activity(Dorighi et

al. 2017) with ChIP-seq suggesting H3K4me1 remaining at intergenic sites. The alternate source for H3K4 methylation is unclear. One possibility is that WRAD in the absence of MLL proteins possesses a low level of intrinsic mono-methyltransferase activity(Patel et al. 2009) which could be an alternative source. Other SET domain containing proteins such as the SMYD family and their potential compensatory roles for MLL3/4 activity are much less characterized. Consequently, while current evidence suggests MLL3/4 catalytic activity may be dispensable the role of H3K4me1 in enhancer activity at large remains unclear.

Several studies have demonstrated important functions for H3K4me1. First, dCas9-MLL3SET targeting of H3K4 methylation activity to the human beta globin locus synergizes with H3K27ac catalysis to maximally stimulate transcription(S.-P. Wang et al. 2017). Using a similar approach with dCas9 targeted methylation Bing Ren and colleagues have also shown that H3K4me1 strengthens enhancer-promoter contacts and transcription at the Sox2 locus(Yan et al. 2018). H3K4me1 is positively correlated with BAF complex localization at the genomic level and the BAF complex preferentially remodels H3K4me1 modified nucleosomes over those modified with H3K4me3 *in vitro*(Local et al. 2018). Interestingly, a separate study using *in vitro* experiments found that H3K4me1 inhibits BAF complex activity(Mashtalir et al. 2021). It is unclear how different subunit compositions of the BAF complexes used in each of these studies may contribute to the results reported. For example, purified complexes used in Mashtalir et al. do not appear to include the BAF subunit DPF2/3 (SMARCG3), which was shown to recognize H3K4me1 in Local et al. Either way, the positive or negative regulation by BAF complex in the native chromatin context represent important potential functions of

H3K4me1. Finally, it has been proposed that H3K4me1 prevents silencing of regions by DNA methylation in early germline development(Bleckwehl et al. 2021). Because many studies that are addressing H3K4me1 function are using MLL3/4 perturbations in steady-state ESCs, which have a uniquely DNA hypomethylated genome(Leitch et al. 2013), the role of H3K4me1 including that catalyzed by MLL3/4 may be underappreciated. Indeed, a recent study found MLL3/4 catalytic activity to be important *in vivo* for extraembryonic development but not gastrulation(Xie et al. 2022). It remains to be determined how much developmental context may play a role in the functions of H3K4me1 and or MLL3/4 catalytic activity.

CBP and P300: Enhancer acetyltransferases

The most common histone acetyltransferases associated with enhancers are CBP (KAT3A) and P300 (KAT3B). CBP/P300 are highly homologous and acetylate multiple residues on each histone subunit including H3K27. Complete genetic deletion or heterozygous loss of CBP or P300 leads to lethality during development(Yao et al. 1998). Gene-dosage of CBP/P300 together is also critical for development, suggesting at least a partial functional redundance(Xu et al. 2006; Yao et al. 1998). Accordingly, mutations in either CBP or P300 are frequently associated with cancer and the neurocristopathy Rubinstein-Taybi syndrome(Mendiratta et al. 2021; Petrij et al. 1995; Roelfsema et al. 2005).

A stable complex with specific subunits does not appear to be required for CBP or P300 to perform their function. Instead, CBP/P300 interact with a variety of coactivating

proteins including another histone acetyltransferase PCAF (KAT2B)(X. J. Yang et al. 1996; Wallberg et al. 2002). By default, the catalytic cores of CBP/P300 are blocked intramolecularly by a pseudosubstrate domain known as the autoinhibitory loop (Karanam et al. 2007; Thompson et al. 2004). Once bound by coactivators, the autoinhibitory loop is autoacetylated and displaced from the catalytic core to enable target substrate access. Acetylation of the loop is also able to occur in trans(Ortega et al. 2019). Notably, histone acetylation is not the only target of CBP/P300. Shortly after uncovering histone acetyltransferase activity a variety of non-histone target substrates for CBP/P300 were also discovered including p53, E2F, Rb, and hundreds of additional targets(Gu and Roeder 1997; Martínez-Balbás et al. 2000; Chan et al. 2001; Weinert et al. 2018). Collectively these results show CBP and P300 play major roles in promoting transcription through catalytic activity and scaffolding coactivators.

Enhancer histone acetylation

Histone acetylation, catalyzed by CBP/P300 and other acetyltransferase families, occurs at actively transcribed genes and active enhancer elements. Histone acetylation is proposed to serve at least two functions: First, neutralizing the positive charge of lysine by acetylation weakens the interaction between the nucleosome and the negatively charged DNA phosphate backbone(Shvedunova and Akhtar 2022). With more acetylation, it may be easier for TFs to bind or for nucleosome rearrangement by chromatin regulators. For example, H4K16ac reduces inter-nucleosomal interactions(R. Zhang, Erler, and Langowski 2017). Second, many chromatin regulators have domains

which recognize specific patterns of histone modifications including acetylation. Domains known to bind acetylated histones include bromodomains, DPF, and YEATS(Khan, Bridgers, and Strahl 2017). Chromatin regulators recruited by acetylation perform critical functions such as BRD4 recruitment promoting transcription by mediating p-TEFb phosphorylation and RNA Pol II recruitment(M. K. Jang et al. 2005).

Acetylation of H3K27 or H3.3K27 has been widely associated with active enhancers and promoters. Studies are emerging, however, that demonstrate H3K27ac is not essential for gene regulation. Point mutations replacing lysine 27 with arginine in both H3F3A and H3F3B (H3.3 homologs) have considerable loss in enhancer H3K27ac(Shvedunova and Akhtar 2022). However, transcription is largely undisturbed. One study took this approach a step farther and generated point mutants of lysine 27 for all H3 and H3.3 genes(Sankar et al. 2022). Despite complete loss of H3K27ac the transcriptome remained relatively unperturbed in steady state cell culture and during differentiation. Most defects were associated with loss of H3K27me3 rather than acetylation. Loss of CBP/P300 catalytic activity, in contrast, has dramatic transcriptional impacts(Lasko et al. 2017; Narita et al. 2021). Thus, even if H3K27ac does not appear to have an essential function at enhancers, other histone and non-histone acetylation by CBP/P300 may still play critical functional roles.

The canonical model of enhancer activation

Enhancers can exist in several regulatory states and the coordination of enhancer state dynamics are an important layer part of enhancer-based gene regulation. The conversion of a latent, inactive enhancer to an active enhancer involves acquiring

chromatin accessibility, H3K4me1, and H3K27ac. Enhancer activation is proposed to occur in a stepwise fashion(Lee et al. 2013; C. Wang et al. 2016; Lai et al. 2017). First, the initiation for de novo enhancer activation is the binding of a TF to its sequence specific motif and there are multiple proposed ways in which a TF can penetrate the nucleosome occluded chromatin of an inactive enhancer. For example, a TF like FoxA1 or Grhl2 may have “pioneering” activity that allows binding and remodeling of nucleosome wrapped DNA alone or in conjunction with cofactors(Jacobs et al. 2018; Zaret and Carroll 2011). On the other hand, some TFs may initiate binding during a stochastic opening in the chromatin. The TF, when enhancer-bound, then recruits chromatin regulating complexes such as MLL3/4 and CBP/P300. At this step, the specificity of TF activity at an enhancer is transduced through chromatin regulators with more generalizable functions in regulating transcription.

One of the earliest proposed steps in enhancer activation is the recruitment of MLL3/4 and H3K4 monomethylation by an enhancer bound TF. It is suggested that the deposition of H3K4me1 by MLL3/4 in the absence of H3K27ac is a signature for “primed” enhancers, an intermediate regulatory state between an inactive and active enhancer. Indeed the priming of enhancers by MLL3/4 is a critical function of MLL3/4 in early differentiation of mouse ESCs(C. Wang et al. 2016). During de novo enhancer activation MLL3/4 is required to recruit CBP/P300 which acetylate a variety of histone residues including H3K27ac(Lee et al. 2013; Lai et al. 2017; Lee et al. 2017). Recruited CBP/P300 then stimulates transcription. These steps reflect the prevailing, canonical model of enhancer activation (Figure 1.3).

The interdependence of chromatin regulators in enhancer activation

Other factors have been associated with enhancer function but the mechanistic relationship between them and steps in the canonical model are incompletely known. The COMPASS-like subunit UTX is thought to scaffold the interaction between MLL4 and P300, and it has been shown recently that BRG1/BAF complex may be recruited by MLL4 through UTX as well(S.-P. Wang et al. 2017; Park et al. 2021). MLL3/4 also promotes eRNA production from enhancers in a catalytic activity-independent manner(Dorighi et al. 2017). However, catalytic activity of MLL3/4 has been shown to play a role in Cohesin complex recruitment and enhancer-promoter contacts(Yan et al. 2018). MLL3/4 has not yet been directly related to other chromatin regulating complexes such as the HIRA complex which deposit H3.3 at enhancers.

The dependence between chromatin regulating complexes and CBP/P300 has been extensively studied. The BAF complex (BRG1) interacts with P300 and has been shown to be required for P300 activity(Alexander et al. 2015; Alver et al. 2017). This interaction may be mediated by UTX which in conjunction with MLL4, has been shown to bind to and activate CBP with Brm1/BAF(Tie et al. 2012). Together with BAF, CBP and P300 are required for NIPBL recruitment which is the loader for the Cohesin complex. This suggests Cohesin is downstream of P300/CBP activity. However, whether BAF or CBP/P300 are in parallel or sequential functional pathways is unclear. Phosphorylation of H3.3 stimulates P300 acetylation *in trans* suggesting H3.3 deposition by HIRA is upstream of P300 activity(Martire et al. 2019). Once acetylated, histones at enhancers recruit BRD4 which promotes elongation of RNApol II, stimulating eRNA

production(Kanno et al. 2014). Ultimately, between MLL3/4 and CBP/P300, the latter appears to play the largest role in translating enhancer function into promoter activity.

In summary, it is not clear how other chromatin regulators might be involved before MLL3/4 is recruited to enhancers. However, enhancer localized MLL3/4 is important for recruiting CBP/P300 and the BAF complex. The recruitment of CBP/P300 leads to eRNA transcription, Cohesin accumulation, and target promoter transcription. The HIRA complex appears to play a role upstream of CBP/P300 activity, but it remains to be known whether HIRA function relies on or is independent of MLL3/4. Therefore, many questions remain to be addressed about enhancer activation mechanisms including those facilitated by MLL3/4.

Premise of studies

The canonical model of enhancer activation suggests a central role for MLL3/4 in recruiting CBP/P300. Exceptions to the model have been observed but not directly explored. Several challenges make it difficult to formulate a general model of enhancer activation from existing studies. The functional redundancy between related members of the same complex such as MLL3/MLL4 and P300/CBP requires multi-factor perturbations which may be hard to perform. Further, loss-of-function of one or more core members of a complex often has dramatic effects on gene regulation which make it difficult to determine primary effects. Finally, chromatin regulators likely have different, context dependent roles depending on the recruiting TF, the chromatin environment, and or cellular signaling. For these reasons, the generalizability of the canonical model is not clear. Identification of meaningful exceptions to the model would play an important role in

expanding our concept of enhancer-based gene regulation and therefore drivers of cell identity. **In Chapter 2**, I use genetic deletions of MLL3/4 and the naive to formative transition in ESCs to rigorously demonstrate exceptions to the canonical model of enhancer activation.

It is also important to advance our understanding of the canonical model of enhancer activation. In addition to challenges mentioned previously, our conceptual framework is also complicated by trying to link mechanisms from studies utilizing diverse biological systems, such as various transcription factors and cell identities. Therefore, there is an urgent need to evaluate the mechanistic relationships between chromatin regulators for the canonical model, which remains incompletely understood, in a single system. **In Chapter 3**, we build on our work using MLL3/4 deficient ESCs and previous studies in our lab to create a system to directly study the canonical model of enhancer activation for a single, well-described pioneer TF Grhl2. Collectively, our efforts expand on the scope of possible models of enhancer activation and deepen our mechanistic understanding of the prevailing model in the field.

Figures

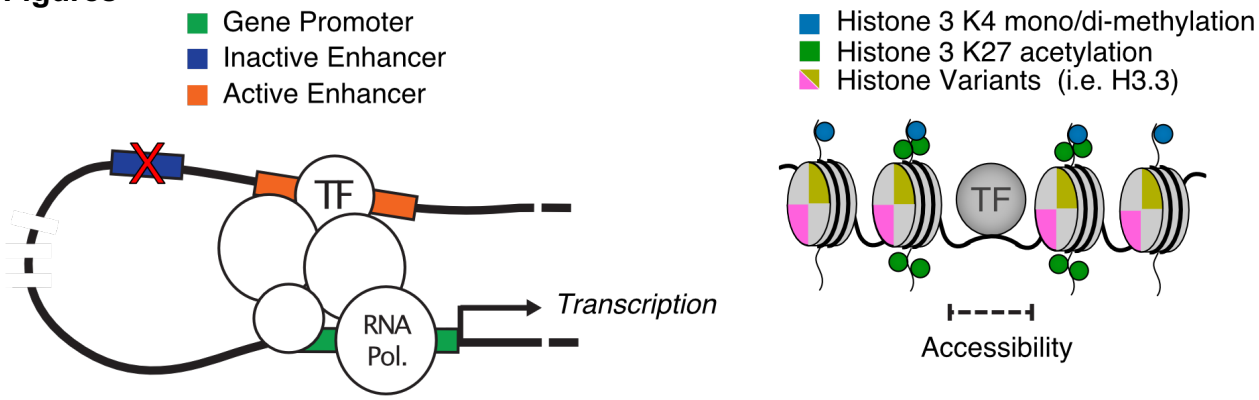

Figure 1.1 – Active enhancers drive transcription of target genes.

Active enhancers drive transcription of target gene promoters (Left) and are characterized by several “active’ molecular features including histone modifications and histone variants (Right).

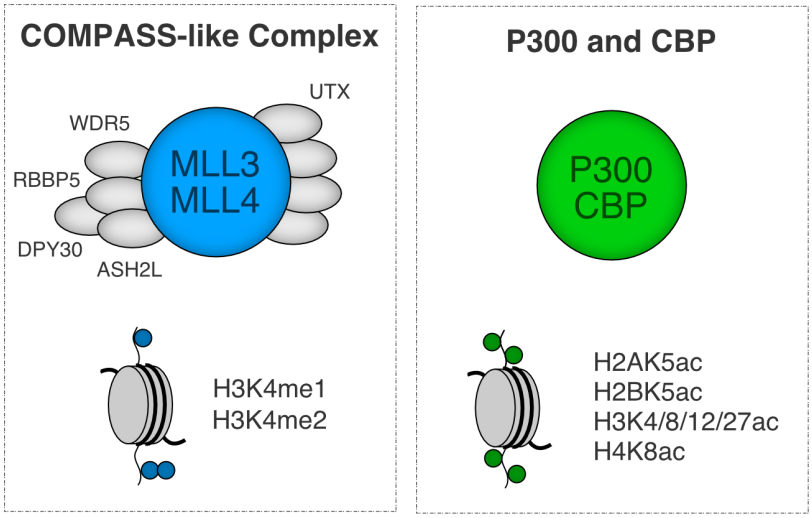

Figure 1.2 – Active enhancer histone modifications are mediated by COMPASS-like complex and P300/CBP.

Enhancer H3K4me1 and H3K27ac are deposited by homologous pairs of enzymes known as MLL3/MLL4 and CBP/P300 respectively.

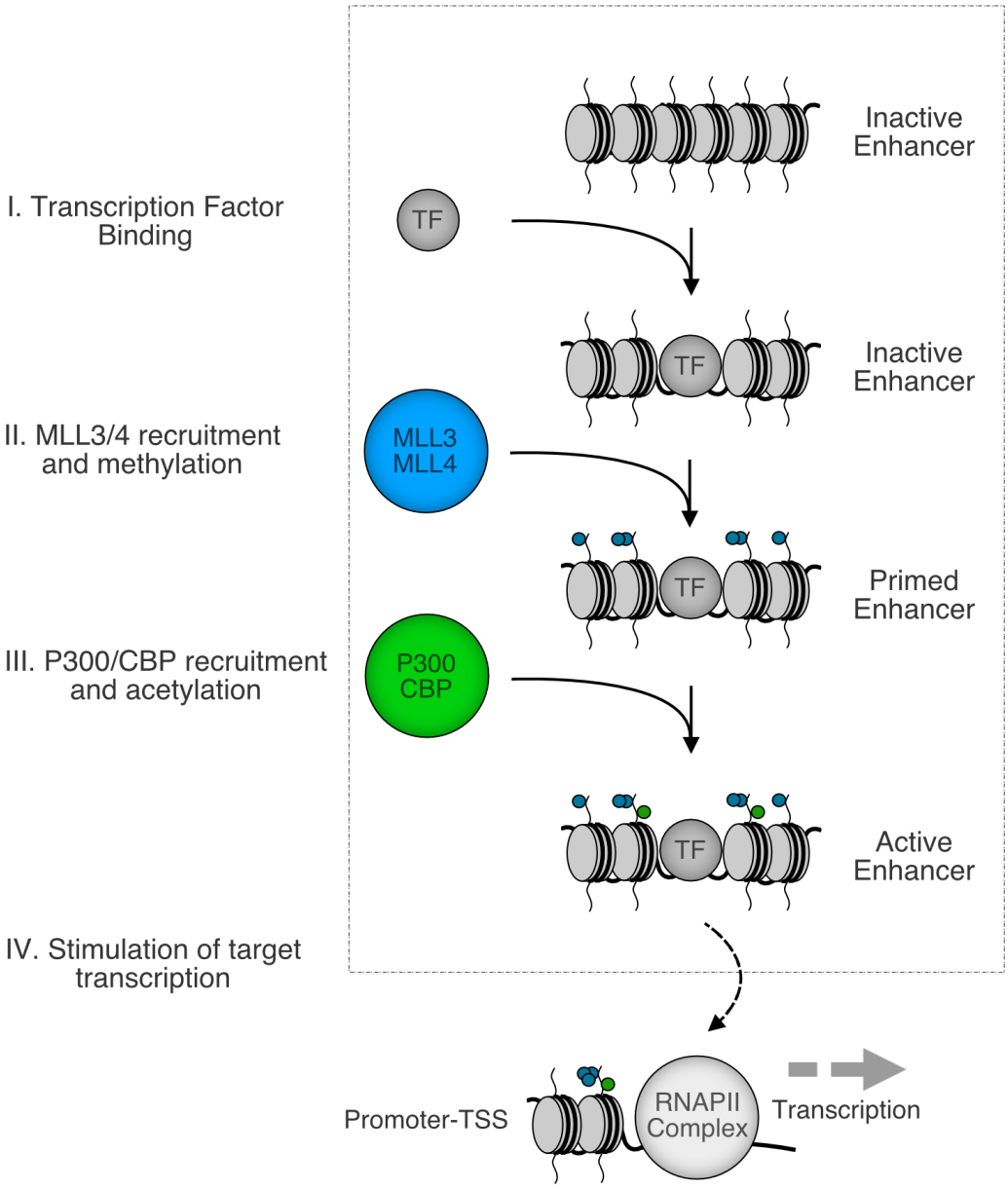

Figure 1.3 - The canonical model of enhancer activation facilitated by MLL3/4.

References

Ahmad, Kami, and Steven Henikoff. 2002. “The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly.” *Molecular Cell* 9 (6): 1191–1200. https://doi.org/10.1016/s1097-2765(02)00542-7.

Alexander, Jeffrey M, Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D Weiner. 2019. “Live-Cell Imaging Reveals Enhancer-Dependent Sox2 Transcription in the Absence of Enhancer Proximity.” Edited by Robert H Singer, Kevin Struhl, and Zhe Liu. *ELife* 8 (May): e41769. https://doi.org/10.7554/eLife.41769.

Alexander, Jeffrey M., Swetansu K. Hota, Daniel He, Sean Thomas, Lena Ho, Len A. Pennacchio, and Benoit G. Bruneau. 2015. “Brg1 Modulates Enhancer Activation in Mesoderm Lineage Commitment.” *Development (Cambridge, England)* 142 (8): 1418–30. https://doi.org/10.1242/dev.109496.

Alver, Burak H., Kimberly H. Kim, Ping Lu, Xiaofeng Wang, Haley E. Manchester, Weishan Wang, Jeffrey R. Haswell, Peter J. Park, and Charles W. M. Roberts. 2017. “The SWI/SNF Chromatin Remodelling Complex Is Required for Maintenance of Lineage Specific Enhancers.” *Nature Communications* 8 (1): 14648. https://doi.org/10.1038/ncomms14648.

Ang, Siang-Yun, Alec Uebersohn, C. Ian Spencer, Yu Huang, Ji-Eun Lee, Kai Ge, and Benoit G. Bruneau. 2016. “KMT2D Regulates Specific Programs in Heart Development via Histone H3 Lysine 4 Di-Methylation.” *Development (Cambridge, England)* 143 (5): 810–21. https://doi.org/10.1242/dev.132688.

Ashokkumar, Deepthi, Qinyu Zhang, Christian Much, Anita S. Bledau, Ronald Naumann, Dimitra Alexopoulou, Andreas Dahl, et al. 2020. “MLL4 Is Required after Implantation, Whereas MLL3 Becomes Essential during Late Gestation.” *Development (Cambridge, England)* 147 (12): dev186999. https://doi.org/10.1242/dev.186999.

Avdic, Vanja, Pamela Zhang, Sylvain Lanouette, Adam Groulx, Véronique Tremblay, Joseph Brunzelle, and Jean-François Couture. 2011. “Structural and Biochemical Insights into MLL1 Core Complex Assembly.” *Structure (London, England: 1993)* 19 (1): 101–8. https://doi.org/10.1016/j.str.2010.09.022.

Bantignies, Frédéric, Virginie Roure, Itys Comet, Benjamin Leblanc, Bernd Schuettengruber, Jérôme Bonnet, Vanessa Tixier, André Mas, and Giacomo Cavalli. 2011. “Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila.” *Cell* 144 (2): 214–26. https://doi.org/10.1016/j.cell.2010.12.026.

Bleckwehl, Tore, Giuliano Crispatzu, Kaitlin Schaaf, Patricia Respuela, Michaela Bartusel, Laura Benson, Stephen J. Clark, et al. 2021. “Enhancer-Associated H3K4 Methylation Safeguards in Vitro Germline Competence.” *Nature Communications* 12 (October): 5771. https://doi.org/10.1038/s41467-021-26065-6.

Bögershausen, Nina, Vincent Gatinois, Vera Riehmer, Hülya Kayserili, Jutta Becker, Michaela Thoenes, Pelin Özlem Simsek-Kiper, et al. 2016. “Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-

Linked Kabuki Syndrome Subtype 2." *Human Mutation* 37 (9): 847–64. https://doi.org/10.1002/humu.23026.

Buecker, Christa, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. 2014. "Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency." *Cell Stem Cell* 14 (6): 838–53. https://doi.org/10.1016/j.stem.2014.04.003.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nature Methods* 10 (12): 1213–18. https://doi.org/10.1038/nmeth.2688.

Cao, Kaixiang, Clayton K. Collings, Marc A. Morgan, Stacy A. Marshall, Emily J. Rendleman, Patrick A. Ozark, Edwin R. Smith, and Ali Shilatifard. 2018. "An Mll4/COMPASS-Lsd1 Epigenetic Axis Governs Enhancer Function and Pluripotency Transition in Embryonic Stem Cells." *Science Advances* 4 (1): eaap8747. https://doi.org/10.1126/sciadv.aap8747.

Chan, H. M., M. Krstic-Demonacos, L. Smith, C. Demonacos, and N. B. La Thangue. 2001. "Acetylation Control of the Retinoblastoma Tumour-Suppressor Protein." *Nature Cell Biology* 3 (7): 667–74. https://doi.org/10.1038/35083062.

Chen, Amy F., Arthur J. Liu, Raga Krishnakumar, Jake W. Freimer, Brian DeVeale, and Robert Blelloch. 2018. "GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naïve

Pluripotency." *Cell Stem Cell* 23 (2): 226-238.e4. https://doi.org/10.1016/j.stem.2018.06.005.

Chen, Ping, Jicheng Zhao, Yan Wang, Min Wang, Haizhen Long, Dan Liang, Li Huang, et al. 2013. "H3.3 Actively Marks Enhancers and Primes Gene Transcription via Opening Higher-Ordered Chromatin." *Genes & Development* 27 (19): 2109–24. https://doi.org/10.1101/gad.222174.113.

Deciphering Developmental Disorders Study. 2017. "Prevalence and Architecture of de Novo Mutations in Developmental Disorders." *Nature* 542 (7642): 433–38. https://doi.org/10.1038/nature21062.

Dorighi, Kristel M., Tomek Swigut, Telmo Henriques, Natarajan V. Bhanu, Benjamin S. Scruggs, Nataliya Nady, Christopher D. Still, Benjamin A. Garcia, Karen Adelman, and Joanna Wysocka. 2017. "Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation." *Molecular Cell* 66 (4): 568-576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.

Efroni, Sol, Radharani Duttagupta, Jill Cheng, Hesam Dehghani, Daniel J. Hoeppner, Chandravanu Dash, David P. Bazett-Jones, et al. 2008. "Global Transcription in Pluripotent Embryonic Stem Cells." *Cell Stem Cell* 2 (5): 437–47. https://doi.org/10.1016/j.stem.2008.03.021.

Ernst, Patricia, and Christopher R. Vakoc. 2012. "WRAD: Enabler of the SET1-Family of H3K4 Methyltransferases." *Briefings in Functional Genomics* 11 (3): 217–26. https://doi.org/10.1093/bfgp/els017.

Frank, Steven A., and Martin A. Nowak. 2003. “Developmental Predisposition to Cancer.” *Nature* 422 (6931): 494–494. https://doi.org/10.1038/422494a.

Gu, W., and R. G. Roeder. 1997. “Activation of P53 Sequence-Specific DNA Binding by Acetylation of the P53 C-Terminal Domain.” *Cell* 90 (4): 595–606. https://doi.org/10.1016/s0092-8674(00)80521-8.

Heintzman, Nathaniel D., Rhona K. Stuart, Gary Hon, Yutao Fu, Christina W. Ching, R. David Hawkins, Leah O. Barrera, et al. 2007. “Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome.” *Nature Genetics* 39 (3): 311–18. https://doi.org/10.1038/ng1966.

Heward, J., and S. C. Gough. 1997. “Genetic Susceptibility to the Development of Autoimmune Disease.” *Clinical Science (London, England: 1979)* 93 (6): 479–91. https://doi.org/10.1042/cs0930479.

Hong, Sunhwa, Young-Wook Cho, Li-Rong Yu, Hong Yu, Timothy D. Veenstra, and Kai Ge. 2007. “Identification of JmjC Domain-Containing UTX and JMJD3 as Histone H3 Lysine 27 Demethylases.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (47): 18439–44. https://doi.org/10.1073/pnas.0707292104.

Hsieh, Tsung-Han S., Geoffrey Fudenberg, Anton Goloborodko, and Oliver J. Rando. 2016. “Micro-C XL: Assaying Chromosome Conformation from the Nucleosome to the Entire Genome.” *Nature Methods* 13 (12): 1009–11. https://doi.org/10.1038/nmeth.4025.

Hu, Deqing, Xin Gao, Marc A. Morgan, Hans-Martin Herz, Edwin R. Smith, and Ali Shilatifard. 2013. “The MLL3/MLL4 Branches of the COMPASS Family Function

as Major Histone H3K4 Monomethylases at Enhancers." *Molecular and Cellular Biology* 33 (23): 4745–54. https://doi.org/10.1128/MCB.01181-13.

Jacobs, Jelle, Mardelle Atkins, Kristofer Davie, Hana Imrichova, Lucia Romanelli, Valerie Christiaens, Gert Hulselmans, et al. 2018. "The Transcription Factor Grainyhead Primes Epithelial Enhancers for Spatiotemporal Activation by Displacing Nucleosomes." *Nature Genetics* 50 (7): 1011–20. https://doi.org/10.1038/s41588-018-0140-x.

Jang, Moon Kyoo, Kazuki Mochizuki, Meisheng Zhou, Ho-Sang Jeong, John N. Brady, and Keiko Ozato. 2005. "The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription." *Molecular Cell* 19 (4): 523–34. https://doi.org/10.1016/j.molcel.2005.06.027.

Jang, Younghoon, Chaochen Wang, Lenan Zhuang, Chengyu Liu, and Kai Ge. 2017. "H3K4 Methyltransferase Activity Is Required for MLL4 Protein Stability." *Journal of Molecular Biology*, Deciphering Histone Modifications in Development and Disease, 429 (13): 2046–54. https://doi.org/10.1016/j.jmb.2016.12.016.

Jenuwein, T., and C. D. Allis. 2001. "Translating the Histone Code." *Science (New York, N.Y.)* 293 (5532): 1074–80. https://doi.org/10.1126/science.1063127.

Kanno, Tomohiko, Yuka Kanno, Gary LeRoy, Eric Campos, Hong-Wei Sun, Stephen R. Brooks, Golnaz Vahedi, et al. 2014. "BRD4 Assists Elongation of Both Coding and Enhancer RNAs by Interacting with Acetylated Histones." *Nature Structural & Molecular Biology* 21 (12): 1047–57. https://doi.org/10.1038/nsmb.2912.

Karanam, Balasubramanyam, Ling Wang, Dongxia Wang, Xin Liu, Ronen Marmorstein, Robert Cotter, and Philip A. Cole. 2007. “Multiple Roles for Acetylation in the Interaction of P300 HAT with ATF-2.” *Biochemistry* 46 (28): 8207–16. https://doi.org/10.1021/bi7000054.

Kaya-Okur, Hatice S., Derek H. Janssens, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2020. “Efficient Low-Cost Chromatin Profiling with CUT&Tag.” *Nature Protocols* 15 (10): 3264–83. https://doi.org/10.1038/s41596-020-0373-x.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. “CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells.” *Nature Communications* 10 (1): 1930. https://doi.org/10.1038/s41467-019-09982-5.

Khan, Abid, Joseph B. Bridgers, and Brian D. Strahl. 2017. “Expanding the Reader Landscape of Histone Acylation.” *Structure (London, England: 1993)* 25 (4): 571–73. https://doi.org/10.1016/j.str.2017.03.010.

Kleefstra, Tjitske, Jamie M. Kramer, Kornelia Neveling, Marjolein H. Willemsen, Tom S. Koemans, Lisenka E. L. M. Vissers, Willemijn Wissink-Lindhout, et al. 2012. “Disruption of an EHMT1-Associated Chromatin-Modification Module Causes Intellectual Disability.” *American Journal of Human Genetics* 91 (1): 73–82. https://doi.org/10.1016/j.ajhg.2012.05.003.

Krishnakumar, Raga, Amy F. Chen, Marisol G. Pantovich, Muhammad Danial, Ronald J. Parchem, Patricia A. Labosky, and Robert Blelloch. 2016. “FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and Repressing

Enhancer Activity." *Cell Stem Cell* 18 (1): 104–17. https://doi.org/10.1016/j.stem.2015.10.003.

Lai, Binbin, Ji-Eun Lee, Younghoon Jang, Lifeng Wang, Weiqun Peng, and Kai Ge. 2017. "MLL3/MLL4 Are Required for CBP/P300 Binding on Enhancers and Super-Enhancer Formation in Brown Adipogenesis." *Nucleic Acids Research* 45 (11): 6388–6403. https://doi.org/10.1093/nar/gkx234.

Lasko, Loren M., Clarissa G. Jakob, Rohinton P. Edalji, Wei Qiu, Debra Montgomery, Enrico L. Digiammarino, T. Matt Hansen, et al. 2017. "Discovery of a Selective Catalytic P300/CBP Inhibitor That Targets Lineage-Specific Tumours." *Nature* 550 (7674): 128–32. https://doi.org/10.1038/nature24028.

Lee, Ji-Eun, Young-Kwon Park, Sarah Park, Younghoon Jang, Nicholas Waring, Anup Dey, Keiko Ozato, Binbin Lai, Weiqun Peng, and Kai Ge. 2017. "Brd4 Binds to Active Enhancers to Control Cell Identity Gene Induction in Adipogenesis and Myogenesis." *Nature Communications* 8 (1): 2217. https://doi.org/10.1038/s41467-017-02403-5.

Lee, Ji-Eun, Chaochen Wang, Shiliyang Xu, Young-Wook Cho, Lifeng Wang, Xuesong Feng, Anne Baldridge, et al. 2013. "H3K4 Mono- and Di-Methyltransferase MLL4 Is Required for Enhancer Activation during Cell Differentiation." *ELife* 2 (December): e01503. https://doi.org/10.7554/eLife.01503.

Leitch, Harry G., Kirsten R. McEwen, Aleksandra Turp, Vesela Encheva, Tom Carroll, Nils Grabole, William Mansfield, et al. 2013. "Naïve Pluripotency Is Associated with Global DNA Hypomethylation." *Nature Structural & Molecular Biology* 20 (3): 311–16. https://doi.org/10.1038/nsmb.2510.

Lettice, Laura A., Simon J. H. Heaney, Lorna A. Purdie, Li Li, Philippe de Beer, Ben A. Oostra, Debbie Goode, Greg Elgar, Robert E. Hill, and Esther de Graaff. 2003. "A Long-Range Shh Enhancer Regulates Expression in the Developing Limb and Fin and Is Associated with Preaxial Polydactyly." *Human Molecular Genetics* 12 (14): 1725–35. https://doi.org/10.1093/hmg/ddg180.

Local, Andrea, Hui Huang, Claudio P. Albuquerque, Namit Singh, Ah Young Lee, Wei Wang, Chaochen Wang, et al. 2018. "Identification of H3K4me1-Associated Proteins at Mammalian Enhancers." *Nature Genetics* 50 (1): 73–82. https://doi.org/10.1038/s41588-017-0015-6.

Marks, Hendrik, Tüzer Kalkan, Roberta Menafra, Sergey Denissov, Kenneth Jones, Helmut Hofemeister, Jennifer Nichols, et al. 2012. "The Transcriptional and Epigenomic Foundations of Ground State Pluripotency." *Cell* 149 (3): 590–604. https://doi.org/10.1016/j.cell.2012.03.026.

Martínez-Balbás, Marian A., Uta-Maria Bauer, Søren J. Nielsen, Alexander Brehm, and Tony Kouzarides. 2000. "Regulation of E2F1 Activity by Acetylation." *The EMBO Journal* 19 (4): 662–71. https://doi.org/10.1093/emboj/19.4.662.

Martire, Sara, Aishwarya A. Gogate, Amanda Whitmill, Amanuel Tafessu, Jennifer Nguyen, Yu-Ching Teng, Melodi Tastemel, and Laura A. Banaszynski. 2019. "Phosphorylation of Histone H3.3 at Serine 31 Promotes P300 Activity and Enhancer Acetylation." *Nature Genetics* 51 (6): 941–46. https://doi.org/10.1038/s41588-019-0428-5.

Mashtalir, Nazar, Hai T. Dao, Akshay Sankar, Hengyuan Liu, Aaron J. Corin, John D. Bagert, Eva J. Ge, et al. 2021. "Chromatin Landscape Signals Differentially

Dictate the Activities of MSWI/SNF Family Complexes." *Science (New York, N.Y.)* 373 (6552): 306–15. https://doi.org/10.1126/science.abf8705.

Mendiratta, Gaurav, Eugene Ke, Meraj Aziz, David Liarakos, Melinda Tong, and Edward C. Stites. 2021. "Cancer Gene Mutation Frequencies for the U.S. Population." *Nature Communications* 12 (1): 5961. https://doi.org/10.1038/s41467-021-26213-y.

Mulas, Carla, Tüzer Kalkan, Ferdinand von Meyenn, Harry G. Leitch, Jennifer Nichols, and Austin Smith. 2019. "Defined Conditions for Propagation and Manipulation of Mouse Embryonic Stem Cells." *Development (Cambridge, England)* 146 (6): dev173146. https://doi.org/10.1242/dev.173146.

Narita, Takeo, Shinsuke Ito, Yoshiki Higashijima, Wai Kit Chu, Katrin Neumann, Jonas Walter, Shankha Satpathy, et al. 2021. "Enhancers Are Activated by P300/CBP Activity-Dependent PIC Assembly, RNAPII Recruitment, and Pause Release." *Molecular Cell* 81 (10): 2166-2182.e6. https://doi.org/10.1016/j.molcel.2021.03.008.

Ng, Sarah B., Abigail W. Bigham, Kati J. Buckingham, Mark C. Hannibal, Margaret J. McMillin, Heidi I. Gildersleeve, Anita E. Beck, et al. 2010. "Exome Sequencing Identifies MLL2 Mutations as a Cause of Kabuki Syndrome." *Nature Genetics* 42 (9): 790–93. https://doi.org/10.1038/ng.646.

Niwa, H., T. Burdon, I. Chambers, and A. Smith. 1998. "Self-Renewal of Pluripotent Embryonic Stem Cells Is Mediated via Activation of STAT3." *Genes & Development* 12 (13): 2048–60. https://doi.org/10.1101/gad.12.13.2048.

Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. “Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre.” *Nature* 485 (7398): 381–85. https://doi.org/10.1038/nature11049.

Ortega, Esther, Srinivasan Rengachari, Ziad Ibrahim, Naghmeh Hoghoughi, Jonathan Gaucher, Alex S. Holehouse, Saadi Khochbin, and Daniel Panne. 2019. “Transcription Factor Dimerization Activates the P300 Acetyltransferase.” *Nature* 562 (7728): 538–44. https://doi.org/10.1038/s41586-018-0621-1.

Park, Young-Kwon, Ji-Eun Lee, Zhijiang Yan, Kaitlin McKernan, Tommy O’Haren, Weidong Wang, Weiqun Peng, and Kai Ge. 2021. “Interplay of BAF and MLL4 Promotes Cell Type-Specific Enhancer Activation.” *Nature Communications* 12 (1): 1630. https://doi.org/10.1038/s41467-021-21893-y.

Patel, Anamika, Venkatasubramanian Dharmarajan, Valarie E. Vought, and Michael S. Cosgrove. 2009. “On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex.” *The Journal of Biological Chemistry* 284 (36): 24242–56. https://doi.org/10.1074/jbc.M109.014498.

Percharde, Michelle, Aydan Bulut-Karslioglu, and Miguel Ramalho-Santos. 2017. “Hypertranscription in Development, Stem Cells, and Regeneration.” *Developmental Cell* 40 (1): 9–21. https://doi.org/10.1016/j.devcel.2016.11.010.

Petrij, F., R. H. Giles, H. G. Dauwerse, J. J. Saris, R. C. Hennekam, M. Masuno, N. Tommerup, G. J. van Ommen, R. H. Goodman, and D. J. Peters. 1995.

"Rubinstein-Taybi Syndrome Caused by Mutations in the Transcriptional Co-Activator CBP." *Nature* 376 (6538): 348–51. https://doi.org/10.1038/376348a0.

Rickels, Ryan, Hans-Martin Herz, Christie C Sze, Kaixiang Cao, Marc A Morgan, Clayton K Collings, Maria Gause, et al. 2017. "Histone H3K4 Monomethylation Catalyzed by Trr and Mammalian COMPASS-like Proteins at Enhancers Is Dispensable for Development and Viability." *Nature Genetics* 49 (11): 1647–53. https://doi.org/10.1038/ng.3965.

Roelfsema, Jeroen H., Stefan J. White, Yavuz Ariyürek, Deborah Bartholdi, Dunja Niedrist, Francesco Papadia, Carlos A. Bacino, et al. 2005. "Genetic Heterogeneity in Rubinstein-Taybi Syndrome: Mutations in Both the CBP and EP300 Genes Cause Disease." *American Journal of Human Genetics* 76 (4): 572–80. https://doi.org/10.1086/429130.

Sankar, Aditya, Faizaan Mohammad, Arun Kumar Sundaramurthy, Hua Wang, Mads Lerdrup, Tulin Tatar, and Kristian Helin. 2022. "Histone Editing Elucidates the Functional Roles of H3K27 Methylation and Acetylation in Mammals." *Nature Genetics* 54 (6): 754–60. https://doi.org/10.1038/s41588-022-01091-2.

Schoenfelder, Stefan, and Peter Fraser. 2019. "Long-Range Enhancer–Promoter Contacts in Gene Expression Control." *Nature Reviews Genetics* 20 (8): 437–55. https://doi.org/10.1038/s41576-019-0128-0.

Shvedunova, Maria, and Asifa Akhtar. 2022. "Modulation of Cellular Processes by Histone and Non-Histone Protein Acetylation." *Nature Reviews Molecular Cell Biology* 23 (5): 329–49. https://doi.org/10.1038/s41580-021-00441-y.

Singer, Zakary S., John Yong, Julia Tischler, Jamie A. Hackett, Alphan Altinok, M. Azim Surani, Long Cai, and Michael B. Elowitz. 2014. “Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells.” *Molecular Cell* 55 (2): 319–31. https://doi.org/10.1016/j.molcel.2014.06.029.

Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers. 1988. “Inhibition of Pluripotential Embryonic Stem Cell Differentiation by Purified Polypeptides.” *Nature* 336 (6200): 688–90. https://doi.org/10.1038/336688a0.

Smith, Austin. 2017. “Formative Pluripotency: The Executive Phase in a Developmental Continuum.” *Development (Cambridge, England)* 144 (3): 365–73. https://doi.org/10.1242/dev.142679.

Spielmann, Nadine, Gregor Miller, Tudor I. Oprea, Chih-Wei Hsu, Gisela Fobo, Goar Frishman, Corinna Montrone, et al. 2022. “Extensive Identification of Genes Involved in Congenital and Structural Heart Disorders and Cardiomyopathy.” *Nature Cardiovascular Research* 1 (2): 157–73. https://doi.org/10.1038/s44161-022-00018-8.

Sze, Christie C., and Ali Shilatifard. 2016. “MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer.” *Cold Spring Harbor Perspectives in Medicine* 6 (11): a026427. https://doi.org/10.1101/cshperspect.a026427.

Tate, John G., Sally Bamford, Harry C. Jubb, Zbyslaw Sondka, David M. Beare, Nidhi Bindal, Harry Boutselakis, et al. 2019. “COSMIC: The Catalogue Of Somatic

Mutations In Cancer." *Nucleic Acids Research* 47 (D1): D941–47. https://doi.org/10.1093/nar/gky1015.

Thompson, Paul R., Dongxia Wang, Ling Wang, Marcella Fulco, Natalia Pediconi, Dianzheng Zhang, Woojin An, et al. 2004. "Regulation of the P300 HAT Domain via a Novel Activation Loop." *Nature Structural & Molecular Biology* 11 (4): 308–15. https://doi.org/10.1038/nsmb740.

Tie, Feng, Rakhee Banerjee, Patricia A. Conrad, Peter C. Scacheri, and Peter J. Harte. 2012. "Histone Demethylase UTX and Chromatin Remodeler BRM Bind Directly to CBP and Modulate Acetylation of Histone H3 Lysine 27." *Molecular and Cellular Biology* 32 (12): 2323–34. https://doi.org/10.1128/MCB.06392-11.

Wallberg, Annika E., Kia Pedersen, Urban Lendahl, and Robert G. Roeder. 2002. "P300 and PCAF Act Cooperatively to Mediate Transcriptional Activation from Chromatin Templates by Notch Intracellular Domains in Vitro." *Molecular and Cellular Biology* 22 (22): 7812–19. https://doi.org/10.1128/MCB.22.22.7812-7819.2002.

Wang, Chaochen, Ji-Eun Lee, Binbin Lai, Todd S. Macfarlan, Shiliyang Xu, Lenan Zhuang, Chengyu Liu, Weiqun Peng, and Kai Ge. 2016. "Enhancer Priming by H3K4 Methyltransferase MLL4 Controls Cell Fate Transition." *Proceedings of the National Academy of Sciences of the United States of America* 113 (42): 11871–76. https://doi.org/10.1073/pnas.1606857113.

Wang, Shu-Ping, Zhanyun Tang, Chun-Wei Chen, Miho Shimada, Richard P. Koche, Lan-Hsin Wang, Tomoyoshi Nakadai, et al. 2017. "A UTX-MLL4-P300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer

Landscapes for Eliciting Transcription." *Molecular Cell* 67 (2): 308-321.e6. https://doi.org/10.1016/j.molcel.2017.06.028.

Weinert, Brian T., Takeo Narita, Shankha Satpathy, Balaji Srinivasan, Bogi K. Hansen, Christian Schölz, William B. Hamilton, et al. 2018. "Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/P300 Acetylome." *Cell* 174 (1): 231-244.e12. https://doi.org/10.1016/j.cell.2018.04.033.

Wit, Elzo de, and Elphège P. Nora. 2023. "New Insights into Genome Folding by Loop Extrusion from Inducible Degron Technologies." *Nature Reviews Genetics* 24 (2): 73–85. https://doi.org/10.1038/s41576-022-00530-4.

Xie, Guojia, Ji-Eun Lee, Anna D. Senft, Young-Kwon Park, Shreeta Chakraborty, Joyce J. Thompson, Chengyu Liu, et al. 2022. "MLL3/MLL4 Methyltransferase Activities Control Early Embryonic Development and Embryonic Stem Cell Differentiation in a Lineage-Selective Manner." *BioRxiv*, January, 2020.09.14.296558. https://doi.org/10.1101/2020.09.14.296558.

Xu, Wu, Tomofusa Fukuyama, Paul A. Ney, Demin Wang, Jerold Rehg, Kelli Boyd, Jan M. A. van Deursen, and Paul K. Brindle. 2006. "Global Transcriptional Coactivators CREB-Binding Protein and P300 Are Highly Essential Collectively but Not Individually in Peripheral B Cells." *Blood* 107 (11): 4407–16. https://doi.org/10.1182/blood-2005-08-3263.

Yan, Jian, Shi-An A. Chen, Andrea Local, Tristin Liu, Yunjiang Qiu, Kristel M. Dorighi, Sebastian Preissl, et al. 2018. "Histone H3 Lysine 4 Monomethylation Modulates Long-Range Chromatin Interactions at Enhancers." *Cell Research* 28 (2): 204–20. https://doi.org/10.1038/cr.2018.1.

Yang, Pengyi, Sean J. Humphrey, Senthilkumar Cinghu, Rajneesh Pathania, Andrew J. Oldfield, Dhirendra Kumar, Dinuka Perera, et al. 2019. “Multi-Omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency.” *Cell Systems* 8 (5): 427-445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

Yang, Shen-Hsi, Munazah Andrabi, Rebecca Biss, Syed Murtuza Baker, Mudassar Iqbal, and Andrew D. Sharrocks. 2019. “ZIC3 Controls the Transition from Naive to Primed Pluripotency.” *Cell Reports* 27 (11): 3215-3227.e6. https://doi.org/10.1016/j.celrep.2019.05.026.

Yang, Shen-Hsi, Tüzer Kalkan, Claire Morissroe, Hendrik Marks, Hendrik Stunnenberg, Austin Smith, and Andrew D. Sharrocks. 2014. “Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency.” *Cell Reports* 7 (6): 1968–81. https://doi.org/10.1016/j.celrep.2014.05.037.

Yang, X. J., V. V. Ogryzko, J. Nishikawa, B. H. Howard, and Y. Nakatani. 1996. “A P300/CBP-Associated Factor That Competes with the Adenoviral Oncoprotein E1A.” *Nature* 382 (6589): 319–24. https://doi.org/10.1038/382319a0.

Yao, T. P., S. P. Oh, M. Fuchs, N. D. Zhou, L. E. Ch’ng, D. Newsome, R. T. Bronson, E. Li, D. M. Livingston, and R. Eckner. 1998. “Gene Dosage-Dependent Embryonic Development and Proliferation Defects in Mice Lacking the Transcriptional Integrator P300.” *Cell* 93 (3): 361–72. https://doi.org/10.1016/s0092-8674(00)81165-4.

Zaret, Kenneth S., and Jason S. Carroll. 2011. “Pioneer Transcription Factors: Establishing Competence for Gene Expression.” *Genes & Development* 25 (21): 2227–41. https://doi.org/10.1101/gad.176826.111.

Zhang, Ruihan, Jochen Erler, and Jörg Langowski. 2017. “Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-Nucleosome Interaction.” *Biophysical Journal* 112 (3): 450–59. https://doi.org/10.1016/j.bpj.2016.11.015.

Zhang, Tiantian, Zhuqiang Zhang, Qiang Dong, Jun Xiong, and Bing Zhu. 2020. “Histone H3K27 Acetylation Is Dispensable for Enhancer Activity in Mouse Embryonic Stem Cells.” *Genome Biology* 21 (1): 45. https://doi.org/10.1186/s13059-020-01957-w.

Zhang, Ying, Anshumali Mittal, James Reid, Stephanie Reich, Steven J. Gamblin, and Jon R. Wilson. 2015. “Evolving Catalytic Properties of the MLL Family SET Domain.” *Structure(London, England:1993)* 23 (10): 1921–33. https://doi.org/10.1016/j.str.2015.07.018.

Chapter 2 : Loss of MLL3/4 decouples enhancer H3K4 monomethylation, H3K27 acetylation, and gene activation during embryonic stem cell differentiation

Summary

Enhancers are essential in defining cell fates through the control of cell-type-specific gene expression. Enhancer activation is a multi-step process involving chromatin remodelers and histone modifiers including the monomethylation of H3K4 (H3K4me1) by MLL3 (KMT2C) and MLL4 (KMT2D). MLL3/4 are thought to be critical for enhancer activation and cognate gene expression including through the recruitment of acetyltransferases for H3K27. Here we test this model by evaluating the impact of MLL3/4 loss on chromatin and transcription during early differentiation of mouse embryonic stem cells. We find that MLL3/4 activity is required at most if not all sites that gain or lose H3K4me1 but is largely dispensable at sites that remain stably methylated during this transition. This requirement extends to H3K27 acetylation (H3K27ac) at most transitional sites. However, many sites gain H3K27ac independent of MLL3/4 or H3K4me1 including enhancers regulating key factors in early differentiation. Furthermore, despite the failure to gain active histone marks at thousands of enhancers, transcriptional activation of nearby genes is largely unaffected, thus uncoupling the regulation of these chromatin events from transcriptional changes during this transition. These data challenge current models of enhancer activation and imply distinct mechanisms between stable and dynamically changing enhancers. Collectively, our study highlights gaps in knowledge about the steps and

epistatic relationships of enzymes necessary for enhancer activation and cognate gene transcription.

This study was published in Genome Biology in 2023

Introduction

Gene regulatory networks that drive cell fate are regulated spatiotemporally by cell type specific transcription factors (TFs). The critical functions of TFs in development are coupled to their target genes through TF binding of *cis*-regulatory elements such as enhancers. Target gene regulation is mediated by general chromatin regulators which are recruited to enhancers bound by TFs. As such, chromatin regulators are an essential part of enhancer function and their dysregulation has severe consequences in development and disease. For example, MLL3 and MLL4 are functionally redundant histone methyltransferases within the COMPASS complex which deposit the histone modification H3K4me1 primarily at enhancers(Hu et al. 2013, 4; Sze and Shilatifard 2016). Moreover, mutations in MLL3 and MLL4 cause developmental defects such as Kabuki Disease and are in the top ten of frequently mutated genes in cancer(Ng et al. 2010; Lavery et al. 2020; Mendiratta et al. 2021). Despite the importance of chromatin regulators to human health, we lack fundamental insight into how general chromatin regulators such as MLL3/4 coordinate enhancer function.

The interdependence of different histone modifications and their depositing enzymes at enhancers have been widely studied, although their mechanistic roles remain controversial. A prevailing model of enhancer activation is that DNA bound TFs recruit

MLL3/4 which deposit H3K4me1(Lee et al. 2013; Lai et al. 2017; S.-P. Wang et al. 2017; Park et al. 2021). MLL3/4 in turn recruits the acetyltransferases P300/CBP which deposit H3K27ac. An activated enhancer then promotes expression of its cognate gene, most often one that is a nearby neighbor(Moore et al. 2020; Nasser et al. 2021). However, it is unclear how generalizable this model is. Indeed, recent studies have begun to challenge the importance of histone modifications and even chromatin regulators in gene regulation by enhancers(Dorighi et al. 2017; Rickels et al. 2017; T. Zhang et al. 2020; Sankar et al. 2022; Martire et al. 2020).

Therefore, we decided to revisit the prevailing model including the epistatic relationship between MLL3/4 to H3K4me1, H3K27ac, and gene transcription in the context of ESCs transitioning from the naive to formative pluripotent states. The naive to formative transition is a rapid, homogenous, and well-characterized transition that faithfully recapitulates the early embryonic cell fate transition in epiblast cells of the mouse between E4.5 and E5.5 at the transcriptional and epigenomic level(Fig1A)(Smith 2017; Gökbuget and Blelloch 2019; Yang et al. 2019; Krishnakumar et al. 2016). Importantly, this developmental window immediately precedes previously described cell fate specification and migrational phenotypes seen in gastrulation with the loss of MLL3/4 *in vivo*(Lee et al. 2013; Ashokkumar et al. 2020, 3). Therefore, the naive to formative transition provides an ideal model to study relationships in enhancer mechanics and transcriptional regulation.

Using state-of-the-art epigenomics techniques to profile genetic deletion of both MLL3 and MLL4 in the naive to formative transition we make several surprising discoveries. First, MLL3/4 is only required at a small subset of sites that maintain stable

H3K4me1 levels during the transition but is essential for all sites that gain or lose H3K4me1. This demonstrates the existence of distinct mechanisms of H3K4me1 deposition at most stable vs dynamic sites. Second, loss of MLL3/4 reduces H3K27ac levels at most dynamic sites, but often does so independent of underlying H3K4me1 levels. Finally, despite dramatic loss of active histone modifications at many enhancers in the formative state, few loci can be linked to changes in nearby gene expression with these changes being relatively minor. Taken together, these results demonstrate that MLL3/4 protein and H3K4me1 deposition can be uncoupled from H3K27ac deposition and also from changes in gene expression at most enhancer sites during early ESC differentiation. Therefore, our data challenge the prevailing model of enhancer activation and suggests that the functional role of MLL3/4 protein in cell fate transitions is more context dependent than previously appreciated.

Results

MLL3/4 is dispensable for transcriptional activation of much of the formative program

To study the role of MLL3 and MLL4 in stem cell self-renewal and differentiation we acquired MLL3$^{-/-}$; MLL4$^{fl/fl}$ ESCs (hereafter called MLL3KO)(C. Wang et al. 2016). We integrated Cre-ERT2 into the *Rosa26* locus and induced Cre-recombination with tamoxifen to generate MLL3$^{-/-}$; MLL4$^{-/-}$ double knockout clones (DKO)(Figure 2.2A). Western blotting confirmed loss of MLL3 protein alone in the parental line and both MLL3 and MLL4 in the DKO cells compared to a wildtype (WT) control (Figure 2.1B). Under

naive pluripotency culture conditions (Serum+LIF+2i), the DKO cells formed colonies that were less compact and grew at a slower rate compared to their WT or MLL3KO counterparts (Figure 2.1C, Figure 2.2B). Next, we removed LIF+2i from the media to induce the formative state after ~63 hours in WT cells (Chen et al. 2018, 2). As expected, WT ESC colonies flattened out during this transition. Similar phenotypic changes were seen in MLL3KO and DKO, although in both cases cells appeared more dispersed. Proliferation was reduced in MLL3KO and DKO cells relative to WT, with the greatest impact on transitioning DKO (Figure 2.2B). All cell lines expressed OCT4 and NANOG protein in naive culture conditions and similarly downregulated NANOG in formative conditions as indicated by Western blot (Figure 2.2C). These data show that while there are morphological and proliferative differences, the knockout cells self-renew under naive culture conditions, have a pluripotent-like cell identity, and show evidence of a cell state transition in formative conditions.

To determine the transcriptional effects of loss of MLL3/4 we performed RNA-seq on WT, MLL3KO, and DKO lines in naive and formative conditions. Differential gene expression analysis between WT naive and formative samples uncovered 887 downregulated genes (naive enriched genes) and 1602 upregulated genes (formative enriched genes) (Figure 2.1D). After subsetting for all naive and formative enriched genes we conducted UMAP analysis of both the WT and mutant cells under both conditions and found samples separated predominately by developmental state rather than by genotype (Figure 2.1E). Correlation analysis of the gene expression changes seen in MLL3KO vs WT and DKO vs. WT showed Pearson correlations of 0.8 and 0.6 respectively showing that transitional changes in gene regulation still occur in the absence of MLL3/4 (Figure

2.1F,G). The dynamic range of gene expression changes was reduced in the MLL3/4 DKO as evidenced by a reduced slope in the linear model. This reduction was largely driven by diminished expression of naïve enriched genes in the naïve state (Figure 2.2D). Upregulation of the formative program was largely unimpacted by MLL3/4 loss (Figure 2.2D). qPCR analysis validated the successful upregulation of known formative markers (Fgf5, Otx2, Dnmt3b)(Figure 2.2E). These results suggested a relatively normal overall transition in gene expression changes between naive and formative even in the absence of both MLL3 and MLL4.

In contrast, direct comparison of MLL3/4 DKO vs. WT identified hundreds of misregulated genes in both the naive and formative states. In the naive state, DKO cells had 783 upregulated genes and 929 downregulated genes (P.adj < 0.05 and Log2CPM > 1)(Figure 2.1H). We categorized these genes that were up or down by whether they are normally naive enriched, formative enriched, or unchanged between naive and formative (Figure 2.1I). This analysis showed a similar proportion of formative genes to be prematurely expressed (275 of 783, 35%) as naive enriched genes that were reduced in the naive state (348 of 928, 39%). In the formative state we found 646 genes were upregulated and 461 genes were downregulated in DKO cells compared to WT(Figure 2.1J). Surprisingly, fewer genes were down in the DKOs in the formative state than the naive state (461 vs. 929). Focusing on genes that normally change during the transition, 83 naive enriched genes were abnormally up and 283 formative enriched genes were abnormally down in the formative state (Figure 2.1K). These numbers were overall low relative to normal gene expression changes seen with naive to formative transition (887 down, 1602 up) showing that MLL3/4 only regulates a minor subset of genes among those

that are normally gained with the transition from naive to formative. Given that the WT and DKO ESCs represent slightly different strain backgrounds (see methods), we also compared MLL3/4DKO to its MLL3KO parental line. This comparison showed even fewer impacted genes (Figure 2.2F,G).

MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent transition

Though expression of a small number of genes were impacted it was unclear if this was due to direct or indirect effects of MLL3/4 loss. To gain a better understanding of the direct effects of MLL3/4 loss on transcription changes during the naive to formative transition we evaluated the monomethylation of H3K4, the primary known role of MLL3/4 and a surrogate marker of their binding. We performed CUT&RUN(Skene, Henikoff, and Henikoff 2018) to profile H3K4me1 in naive and formative samples for WT, MLL3KO, and DKO cells. Differential peak analysis using DiffBind(Stark and Brown 2022) on WT naive and formative samples with peaks called by SEACR(Meers, Tenenbaum, and Henikoff 2019) uncovered approximately ~13000 peaks each that significantly gained (formative H3K4me1 peaks) or lost H3K4me1 (naive H3K4me1 peaks) during the normal pluripotent transition (FDR < 0.01, Log2 FC > 1). An additional ~94000 peaks did not significantly change between cell states (shared H3K4me1 peaks, FDR > 0.1, Log2 FC < 0.7)(Figure 2.3A,B). Naïve and formative enriched peaks were almost completely lost in the DKO cells, but only slightly reduced in MLL3KO cells (Figure 2.3C,D). Reanalysis by Diffbind uncovered 0 total peaks that showed a significant change between the two states in the

DKO background while we found 29088 significantly changing peaks in MLL3KO compared to 27458 total peaks in WT (Figure 2.3E, Figure 2.4A). In contrast, shared H3K4me1 peaks (shown in gray) were not obviously affected. To test whether the loss of dynamic H3K4me1 was a direct effect of MLL3/4 enzymatic activity, we performed and evaluated RNA-seq and H3K4me1 CUT&RUN data using ESCs with point mutations in the SET domains of both MLL3 and 4 that render them catalytically dead (dCD cells)(Dorighi et al. 2017). Consistent with previous reports(Dorighi et al. 2017; Bleckwehl et al. 2021), we found a comparable reduction in global H3K4me1 in westerns of acid-extracted histones between DKO and dCDs when compared to WT while seeing few transcriptional differences between dCD and WT cells in the naive and formative states (Figure 2.4B-E). Moreover, H3K4me1 CUT&RUN on dCD cells largely recapitulated the H3K4me1 defects observed in DKOs (Figure 2.3F). Diffbind uncovered only 272 total differential peaks between naive and formative in the catalytic mutant background (Figure 2.3G).

The majority of H3K4me1 sites, whether shared, naive-enriched, or formative-enriched were located at distal elements (intergenic or intronic) and not promoters (Figure 2.3H). To quantify dependency of MLL3/4 H3K4me1 at the various sites, we set cutoffs for dependent (Log2 FC naive DKO/WT < -1 or Log2 FC formative DKO/WT < -1) and independent H3K4me1 (Log2 FC naive DKO/WT > 0.7 & Log2 FC form DKO/WT < 0.7). Of the distal H3K4me1 sites, only 15% of shared sites (8632 of 64494) showed dependency on MLL3/4 for H3K4me1, while all naive and formative specific sites were dependent. Heatmap visualization validated these categories and similar behavior at dynamic sites during the transition was observed with analysis of published naive and

formative H3K4me1 ChIP-seq datasets (Figure 2.3J and Figure 2.4F)(Buecker et al. 2014; Chen et al. 2018). Furthermore, evaluation of published naive MLL3/4 ChIP-seq data showed enrichment of the proteins at H3K4me1 sites whether naïve-enriched or shared (Dorighi et al. 2017). Together, these data show that MLL3/4 are essential for a subset of H3K4me1 peaks including all dynamic sites and a small fraction of shared sites.

Next, we asked about the impact of MLL3/4 loss on the transcription of genes closest to the sites representing the different MLL3/4 dependent categories (shared, naive, and formative H3K4me1 sites) (Figure 2.3K). Genes near MLL3/4 dependent shared H3K4me1 sites showed a clear upregulation in expression in the transition from naive to formative in WT cells, suggesting these enhancers may be primed in the naive state and became activated in the formative state. The loss of MLL3/4 had no impact on the upregulation of these genes with the transition to the formative state, although there was a slight reduction in expression in the naïve state. This later finding is consistent with previous findings showing downregulation of genes near MLL3/4 peaks in steady state naïve DKO ESCs(Dorighi et al. 2017), something we also saw with our expression data (Figure 2.4G). In contrast, genes nearby MLL3/4 dependent naive-specific H3K4me1 sites were not impacted in the naïve DKO ESCs. Most surprising, while genes nearby MLL3/4 dependent formative-specific H3K4me1 showed a strong gain in expression during the naive to formative transition, this gain was not diminished by the loss of MLL3/4. Together, these results show that while MLL3/4 is required for all dynamic H3K4me1 and a subset of shared H3K4me1 between the naive and formative state, this requirement has minimal impact on gain of nearby gene transcription with the transition to the formative state.

MLL3/4 dependent and independent distal H3K27ac deposition

MLL3/4 is thought to be important for the recruitment of histone acetyltransferase P300/CBP and thus histone acetylation including at H3K27 at enhancers. To test this general model at the genomic level, we conducted CUT&TAG(Kaya-Okur et al. 2019) for H3K27ac in WT and DKO cells in the naive and formative states. We performed Diffbind at all peaks called by SEACR, intersected these with ATAC-seq peaks from naive and formative WT cells, and subset for distal sites to further enrich for active enhancers. A total of 12802 H3K27ac peaks were significantly down (naive H3K27ac peaks) and 8924 peaks were significantly up (formative H3K27ac peaks) during the transition (Figure 2.5A,B). Another 12205 peaks did not show significant change in H3K27ac between the two cell states (shared H3K27ac peaks). Loss of MLL3/4 resulted in a dramatic decrease in the dynamic changes in H3K27ac normally seen in the WT cells (Figure 2.5C). However, unlike with H3K4me1, many sites still showed some change including sites that appeared to be unique to each pluripotent state. Diffbind using DKO samples uncovered 4295 naive-enriched, 1935 formative-enriched, and 23454 shared peaks, a reduction compared to WT, but still suggestive of cell state dependent dynamic acetylation (Figure 2.5D). In both WT or DKO cells, the majority of H3K27ac was located at distal sites, including both shared and naive/formative-enriched peaks. Most of the cell-state specific H3K27ac sites were dependent on MLL3/4 (Log2 FC DKO/WT < -1 or Log2 FC DKO/WT > -0.7 respectively) with MLL3/4 dependent sites accounting for 22% of shared (2661 of 12205), 57% at naive-enriched (5717 of 9963), and 62% at formative-enriched sites (4495 of 7205) (Figure 2.5F). Heatmap visualization confirmed these findings (Figure 2.5G,H,

Figure 2.6A). Interestingly though, metagene analysis identified a low-level persistent signal and premature H3K27ac signal at MLL3/4 independent naive and formative distal sites respectively. Analysis of published ChIP-seq for H3K27ac and the acetyltransferase P300 in naive and formative cells, showed commensurate binding at the sites with correlated increases and decreases in signal with H3K27ac signal in WT cells (Figure 2.6B)(Buecker et al. 2014; Chen et al. 2018, 2). These data uncover both MLL3/4 dependent and independent mechanisms of P300 acetylation of H3K27 at cell-type specific enhancer sites.

Given that H3K27ac is thought to be a marker of active enhancers, we next asked whether sites with MLL3/4 dependent H3K27ac had impacted transcription at neighboring genes upon loss of MLL3/4. We identified the nearest neighbor transcription start site (TSS) to all distal sites where H3K27ac was dependent on MLL3/4 and separated these sites by whether they were normally shared, naive-enriched, or formative-enriched. We then evaluated expression of genes associated with the TSSs (Figure 2.5I). Shared peaks were associated with genes that normally increased in expression during the transition; this increase was not impacted by MLL3/4 loss. Naive-enriched peaks were associated with genes that show a slight reduction during the transition and that reduction occurred prematurely in the MLL3/4 DKO cells. Strikingly, there was a strong upregulation in gene expression nearby sites that normally show MLL3/4 dependent increases in H3K27ac, but this upregulation was not impacted by MLL3/4 loss. These data show that the upregulation of gene expression associated with a gain of H3K27ac at nearby enhancer sites is largely independent of that acetylation, while maintenance of gene expression

associated with naive enriched H3K27ac sites is at least in part dependent on maintenance of that acetylation.

Enhancer Activation Can Occur Independently of MLL3/4

The data above suggested similar trends in terms of the effects of MLL3/4 on H3K4me1, H3K27ac and nearby gene expression, consistent with a partial dependency on MLL3/4 recruitment for H3K27ac deposition. To investigate this dependency, we focused on all distal sites that show changes in H3K27ac in the context of their H3K4me1 status. We first separated all chromatin accessible H3K4me1 sites (ATAC+) into those that lose, gain, or have shared methylation during the naive to formative transition (Figure 2.8A). Reanalyzing nearest neighbor transcription with these sites did not appreciably change our results obtained with all H3K4me1 sites (Figure 2.8B vs. Figure 2.3K). Next, we subdivided the H3K4me1 categories into ones that lose, gain, maintain, or never have coincident H3K27ac (Figure 2.7A). In general, most H3K4me1 sites never showed H3K27ac. However, among the 5960 sites that lose H3K4me1 during the transition, 1370 also lose H3K27ac, and 913 maintain acetylation. Among 5583 sites that gain H3K4me1, 1172 sites also gain H3K27ac. We identified few sites (<150) for the remaining categories of H3K27ac at sites that either gain or lose H3K4me1. Among 37589 sites maintaining H3K4me1, 2264 sites gain acetylation, 3498 sites lose acetylation, 4353 sites show maintained acetylation. This suggests, although sites with dynamic H3K4me1 tended to have dynamic H3K27ac in the same direction (gain or loss), most H3K27ac dynamics during the naive to formative transition typically occur at sites with preexisting H3K4me1.

To understand how these subgroups were impacted by the loss of MLL3/4, we visualized the data as heatmaps and filtered for MLL3/4 independent or dependent H3K27ac (Log2FC DKO/WT > -0.7 or Log2FC DKO/WT < -1 respectively). Several important findings were uncovered by these efforts. Among the sites that are enriched for both H3K4me1 and H3K27ac (H3K4me1+/H3K27ac+) in the naive state, the loss of MLL3/4 resulted in a consistent loss of monomethylation as seen across all naive enriched H3K4me1 sites (Figure 2.7B). However, only 61% of these sites lost H3K27ac (891 of 1452 sites). Among sites that gain H3K4me1 and H3K27ac during the transition from naive to formative, again all H3K4me1 is dependent on MLL3/4, while 77% also lost H3K27ac (901 out of 1172 sites)(Figure 2.7C). To confirm that the naive and formative H3K4me1+/H3K27ac+ sites were direct MLL3/4 targets, we analyzed published naive ChIP-seq of MLL3/4 and generated CUT&RUN for MLL4 in the formative state (Figure 2.7B,C). Naive MLL3/4 ChIP-seq signal was enriched at naive sites and MLL4 CUT&RUN was enriched specifically at formative sites. Together these analyses show H3K27ac maintenance and de novo deposition can occur at enhancer sites independent of MLL3/4 including at sites requiring MLL3/4 for H3K4me1 deposition.

Next, we analyzed the sites that normally retain monomethylation during the transition (H3K4me1 shared sites), which included a mix of sites that did or did not require MLL3/4 for H3K4me1 (Figure 2.3J). Among the MLL3/4 independent H3K4me1 sites, 2039 sites normally gained H3K27ac, 2357 sites lost H3K27ac, and 3611 sites had H3K27ac that did not change. An additional 18135 sites never showed H3K27ac. For sites that gained H3K27ac, 48% required MLL3/4 to do so even though the same enzymes were not required for the monomethylation at those sites (Figure 2.7D). The

remaining 52% showed strong H3K27ac in formative cells independent of MLL3/4 and most showed premature H3K27ac deposition in naive DKO cells. MLL3/4 signal was low at all these sites consistent with the unaffected H3K4me1 which suggested the loss of acetylation at these sites was due to indirect effects. Similar to the formative H3K27ac sites, naive H3K27ac at shared H3K4me1 sites also demonstrated mixed dependency on MLL3/4 with only 844 of the sites losing H3K27ac (Figure 2.8B). Among the remaining 1320 sites, which retained high levels of naive H3K27ac, the reduction in acetylation upon differentiation was incomplete in DKOs. In contrast to the formative sites, MLL3/4 signal was strong at the naive dependent sites. In MLL3/4 dependent shared H3K4me1 sites (2113 sites) few sites gained or had shared H3K27ac (225 and 742, respectively). Despite losing H3K4me1 these sites showed mixed dependency of H3K27ac on MLL3/4. Notably, sites with naive enriched H3K27ac comprised the largest group (1141 sites) and most of these sites (973 of 1141 sites, 85%) were dependent on MLL3/4 for H3K27ac. Finally, H3K4me1 independent shared sites showed no appreciable MLL3/4 binding despite a loss of H3K27ac in DKO cells at 992 of 3297 sites, implying indirect effects (Figure 2.8D). Taken together, these data show that H3K27ac deposition at cell-type specific sites is not necessarily coupled with binding or activity of MLL3/4. Indeed, we identified almost as many exceptions as examples for the canonical model coupling MLL3/4 and H3K27ac deposition at enhancers.

Next, we asked how the loss of MLL3/4 impacted H3K4me1 and H3K27ac at well-known functionally validated enhancers regulating genes important for pluripotent cell identity including Sox2, Nanog, Fgf5, and Oct6(Buecker et al. 2014; Zhou et al. 2014, 2; Blinka et al. 2016; Thomas et al. 2021; Romeike et al. 2022; Agrawal et al. 2021). Sox2

and Nanog expression normally decrease, while Fgf5 and Oct6 expression increase during the naive to formative transition. In the MLL3/4 DKOs, Sox2 expression failed to decrease during the transition, while the other genes showed normal patterns of expression (Figure 2.1G, Figure 2.8E). However, the impact of MLL3/4 loss on H3K4me1 and H3K27ac at associated enhancers for each of these genes varied (Figure 2.7E). The Sox2 enhancer cluster showed stable H3K4me1 and naive specific H3K27ac, both of which were dependent on MLL3/4. Nanog has at least two known enhancers, all normally showing stable H3K4me1 and a loss of H3K27ac during the transition; none of which were impacted by MLL3/4 loss. Fgf5 has 4 known enhancers, all normally showing increases in H3K4me1 and H3K27ac during the transition. In the DKO cells, all four enhancers showed reduced H3K4me1, with two showing normal upregulation of H3K27ac and two showing reduced H3K27ac in the formative state. Oct6 has one known enhancer, which normally shows a gain in H3K4me1 and H3K27ac during the transition; H3K4me1 was reduced, but H3K27ac was unchanged in DKO cells. These examples of well-known enhancers demonstrate the variable nature of dependency on MLL3/4 for H3K4me1 and H3K27ac, including a surprising dispensability for de novo activation of enhancers for Fgf5 and Oct6 expression.

We hypothesized that dependency of H3K4me1/H3K27ac on MLL3/4 might vary based on the TFs bound at each enhancer. Therefore, we performed differential motif analysis using Gimmemotifs(Bruse and Heeringen 2018) to determine TF binding motifs enriched at enhancer sites dependent on MLL3/4 for H3K4me1 and/or H3K27ac (Figure 2.9A). This analysis identified strong enrichment for the binding motifs for TCF7L2, GRHL, and ZIC1 at sites requiring MLL3/4 for the gain of both H3K4me1 and H3K27ac during

the naive to formative transition. Expression of each of these TFs is maintained or increased during the transition, and GRHL2 is essential for the gain of H3K4me1 and H3K27ac at hundreds of formative specific enhancers(Chen et al. 2018). At MLL3/4 dependent naive peaks, we identified a strong enrichment for multiple TF binding motifs including RXRA, ZBTB20, ZNF257, MYC, ESR2, and NR6A1, all of which are either expressed or have closely related TFs that are expressed in naive cells (Figure 2.9B). Interestingly, the motifs for the critical TF regulators of the naive and/or formative states such as Oct4, Sox2, Nanog, KLFs, Oct6, and Otx2, were generally not uncovered. These analyses suggest that TFs differ in their dependency on MLL3/4 to maintain or establish H3K27ac.

Distal H3K4me1 and H3K27ac are not functionally coupled with formative transcriptional activation

We next asked whether focusing on sites of H3K27ac change in the context of dynamic H3K4me1 peaks might enrich for sites near genes whose expression was impacted by MLL3/4. Similar to analysis of H3K27ac alone, genes nearby naive H3K4me1+/H3K27ac+ sites with MLL3/4 dependent H3K27ac showed reduced expression in naive cells (Figure 2.7B,Figure 2.11A) while genes nearby MLL3/4 independent H3K27ac were unchanged, supporting a role for acetylation in maintaining expression at these sites. In contrast, in the formative state there was little impact on transcription of genes nearby either MLL3/4 independent or dependent H3K27ac (Figure

2.7C,Figure 2.11B). This suggested gain of both H3K4me1 and H3K27ac at formative sites was not associated with formative transcriptional activation.

To expand on this surprising finding, we used ChromHMM as an unbiased approach to look at all potential combinations of H3K4me1, H3K27ac in WT and DKO cells in the naive and formative states. We modeled 16 chromatin states to segment all possible combinations of these modifications, and subset for distal peaks overlapping WT ATAC peaks separately in naive and formative cells (Figure 2.10A,C). The sites were then mapped to nearby TSS, and the impact of MLL3/4 on expression of these genes was measured (Figure 2.10B,D). In the naive state, genes near sites that lost H3K27ac showed highly significant downregulation whether or not they also lost H3K4me1 (States 1,15). These sites also showed strong MLL3/4 signal supporting a direct role for the enzymes in coordinating both acetylation and transcription. In the formative state, only genes near distal sites that lost both H3K4me1 and H3K27ac showed a significant reduction in gene expression (State 2) (Figure 2.10D). This state represented a relatively small number of enhancers (1040 sites) linked to only 154 genes. Therefore, this unbiased approach at defining chromatin states again showed that even though MLL3/4 loss dramatically alters the landscape of enhancer activity states, the impact on gene expression changes that occur during the naive to formative transition is relatively minor.

Genes can be regulated by multiple enhancers(Moorthy et al. 2017; Osterwalder et al. 2018). Thus, we hypothesized that MLL3/4 independent enhancers may compensate for loss of MLL3/4 dependent enhancers explaining the underwhelming impact on expression. To test this hypothesis, we compared the change in expression (naive or formative DKO/WT) relative to the number of enhancers impacted by loss of

MLL3/4 for each gene and their associated enhancers. In naive cells, this approach showed a correlation with the change in nearby gene expression (Figure 2.10E). In formative cells, only a very small effect was seen and only for genes with the greatest numbers of impacted enhancers.

We obtained similar results by analyzing the proportion of enhancers that were lost per gene in each pluripotent state (Figure 2.10F). The contrast between naive and formative was even more striking when only considering sites within 10kb of the TSS (Figure 2.11C,D). Therefore, compensation by non-MLL3/4 dependent enhancers does not appear to be a major factor rescuing gene expression in formative cells.

Gene-centric analysis reveals a subset of distal loci associate with MLL3/4 dependent formative genes

Our data showed a requirement for MLL3/4 for H3K4 monomethylation at all sites that normally gain or lose the mark during the naive to formative transition (Figure 2.3D,J,Figure 2.12A). They also showed a requirement for MLL3/4 for H3K27ac at 66% of sites that gain and 66% that lose acetylation during the transition (Figure 2.5G,H, Figure 2.12A, Figure 2.13A). However, only 13% of formative genes that normally increase in expression in the transition and only 39% naive genes that are normally downregulated were impacted by loss of MLL3/4. Among the naive genes that were impacted, there was a strong enrichment for cytokine pathways (Figure 2.13B). Among the formative genes that were impacted, there was an enrichment for development related pathways including regionalization, migration, and morphogenesis (Figure 2.12B). This later finding

suggested a role for initiating transcription of genes involved in a later stage of development, specifically gastrulation.

Our previous analyses utilized an enhancer-centric approach to identify changes in nearby genes. We next asked if impacted genes could be connected to nearby enhancers that showed loss of H3K27ac. To take this gene-centric approach, we categorized all genes that were normally upregulated during the transition to the formative state into those that were not impacted (independent) vs. impacted (dependent) by MLL3/4 loss. Among formative genes, this resulted in 820 independent formative genes and 131 dependent formative genes which could be linked to 2693 and 402 nearby H3K4me1/H3K27ac+ enhancers respectively. As a background control, we linked 12506 expressed genes from our RNA-seq to an associated 37362 H3K4me1/H3K27ac+ peaks. Both the dependent and independent formative genes normally showed an increase in H3K27ac at nearby enhancer sites during the naive to formative transition (Figure 2.12C). This increase in acetylation was mostly lost in DKO cells, whether or not the nearby genes failed to increase gene expression. Similarly, both dependent and independent naive genes (normally downregulated during transition) showed a decrease in nearby enhancer H3K27ac during the naive to transition. However, H3K27ac at enhancers nearby dependent naive genes was more impacted than those nearby independent genes in the DKO cells (Figure 2.13C). Therefore, the gene-centric approach was consistent with enhancer-centric approach showing a more prevalent role of MLL3/4 regulated chromatin changes on the maintenance of naïve gene expression than the gain of formative gene expression.

To look deeper, we investigated the chromatin states nearby the small number of genes that were dependent on MLL3/4 for upregulation of transcript expression during the naive to formative transition by creating genome tracks (Figure 2.12D). These dependent genes included important gastrulation genes such as *T*/*Brachyury*, *Fgf8*, and *Lefty1*, the latter of which is a previously identified target of MLL3/4 (Wang et al. 2016). All three of these genes had one or more nearby enhancers that showed a loss of H3K4me1 and H3K27ac in MLL3/4 DKO formative cells. Thus, the expression of these key regulators relies on enhancers which require MLL3/4 for proper spatiotemporal regulation. However, they represent a very small subset of the thousands of enhancers that lose active marks upon the loss of MLL3/4. All together, these data show that from both a gene-centric or enhancer-centric perspective, H3K4me1 and/or H3K27ac at distal sites are largely uncoupled from target gene activation during the naïve to formative transition with the loss of MLL3/4.

Discussion

Enhancers are thought to play a central role in gene regulation during cell fate transitions. Yet, the molecular epistasis of enhancer activation is incompletely understood. The prevailing model proposes that H3K4me1 deposited by MLL3/4 precedes H3K27ac deposition by P300/CBP and these events stimulate transcription. By combining state-of-the-art CUT&RUN/CUT&TAG technologies and knockouts of both MLL3 and MLL4, we evaluated this model during the well-defined transition from naive to formative pluripotency, where hundreds of genes are both silenced and activated. Surprisingly, the majority of these gene expression changes including de novo expression

of many genes were unaffected by MLL3/4 loss. In contrast, loss of MLL3/4 led to large effects on the active histone modification signatures at thousands of enhancers. These results suggest that MLL3/4 and its orchestration of post-translational histone marks, play a relatively minor role in gene regulation during early ESC differentiation.

MLL3/4 are thought to be the major H3K4 monomethyltransferases in mammals(Hu et al. 2013, 4; Sze and Shilatifard 2016). However, we find that the majority of H3K4me1 is unaffected by MLL3/4 loss in both the naive and formative states. Consistent with our findings, similar results have been previously described in naive ESCs(Dorighi et al. 2017). By investigating the transition to the formative state, we were able to make several novel discoveries though. We discovered that all distal sites that normally show dynamic changes in H3K4me1, either gained or lost, were fully dependent on MLL3/4 for monomethylation. In contrast, only a small fraction of the sites that normally do not change during the transition showed any requirement for enzymes in maintaining monomethylation levels. Therefore, there appear to be at least two distinct types of cis-regulatory elements with regard to MLL3/4 dependency. This raises important follow-up questions such as what enzymes are responsible for the predominant group of H3K4me1 sites that do not require MLL3/4 for maintenance? MLL2 has been suggested as an alternative monomethyltransferase in naive ESCs, but only at a small number of sites that co-exist with H3K27m3, a repressive mark(Dorighi et al. 2017; Mas et al. 2018). MLL1 represents another potential candidate. Interestingly, MLL1 has been shown to inhibit the reprogramming of primed pluripotent (EpiSCs) to naive ESCs(H. Zhang et al. 2016, 1). However, its role in regulating H3K4me1 in naive cells during differentiation has not been evaluated. Other candidates include SMYD1 and SMYD2 which are SET domain

containing proteins that have previously been shown to be required for H3K4me1 and or H3K4m2 at de novo enhancers during macrophage activation(Kaikkonen et al. 2013). Additionally, whether there are distinct roles for the MLL3/4 dependent vs. independent distal sites remains unclear. Given most enhancer activation in the absence of MLL3/4 occurs at these independent H3K4me1 sites they may represent a collection of primed sites distinct from MLL3/4. Surprisingly though, previous studies have shown that at least in the case of MLL3/4, the proteins, but not their enzymatic activity are important in CBP/P300 recruitment and H3K27ac(Dorighi et al. 2017). Therefore, it will be important to understand the link between the factors resulting in methylation of these MLL3/4 independent sites and those that deposit H3K27ac at these sites.

MLL3/4 are thought to be important recruiters of CBP/P300, whose catalytic byproduct H3K27ac is a key marker of enhancer activation. In our data, we identify that more than half of H3K27ac distal sites lose acetylation with MLL3/4 loss, with a greater proportion at dynamic sites than unchanging sites. Many sites that lose H3K27ac do so independent of H3K4me1 loss. Whether this represents an indirect role or a direct role for MLL3/4 in the recruitment of CBP/P300 at sites that also retain H3K4me1 is unclear. However, the presence of MLL3/4 binding at many of these sites supports a direct role. In addition to the MLL3/4 dependent sites, there is also a large number of MLL3/4 independent H3K27ac sites, including many occurring at regions that do require MLL3/4 for H3K4me1. This indicates multiple mechanisms for CBP/P300 recruitment to enhancers that are differentially used depending on the specific enhancer and chromatin state. How different enhancers recruit CBP/P300 using distinct mechanisms is unclear but is likely in part due to the underlying transcription factors binding to those enhancers.

Enhancer subcategories consisting of varying types of TFs have recently been shown to require specific chromatin complexes for enhancer activity(Neumayr et al. 2022). Consistent with the role for different TFs underlying these requirements, TF motif enrichment analysis uncovered several TFs whose motifs were highly enriched at each distinct category of MLL3/4 dependent and independent enhancers. CBP/P300 could either be directly recruited by these TFs and/or be recruited through other factors compensating for loss of MLL3/4. Future studies are necessary to fully elucidate the requirements for recruitment of specific epigenetic complexes by distinct TFs.

Ultimately, it is thought that enhancer activation as defined by the gain of H3K27ac drives increased expression of its target gene. Surprisingly though, while we observe major impacts on the chromatin state of sites with dynamic H3K4me1 and H3K27ac during the naive to formative transition, there are relatively modest changes in the gain and loss of gene expression. The activation of the formative transcriptional program is largely unaffected, with only a few exceptions such as Lefty1 and Fgf8. The key markers of the formative state including Otx2, Oct6, and Fgf5 are all normally upregulated. This unexpected finding raised the question whether there was any specific MLL3/4 dependent enhancer phenotype in formative cells that could be linked to a failure to activate gene expression. Analysis of genes either nearby MLL3/4 dependent formative H3K4me1 sites or MLL3/4 dependent formative H3K27ac sites during the transition did not identify a reduction in gene activation. Most surprising, interrogation of genes nearby enhancers that fail to gain both H3K4me1 and H3K27ac also showed normal increases in gene expression. Only when considering the small number of sites where both H3K4me1 and H3K27ac were depleted, independent of their dynamics, was there a small but significant

effect for a handful of genes. Consistent with our data, recent work mutating H3 and/or H3.3 Lysine 27 sites to Arginine, which cannot be acetylated did not impact either naive gene expression or the activation of the formative program in differentiating ESCs(T. Zhang et al. 2020, 27; Sankar et al. 2022). However, these findings are interpreted as a lack of a requirement for H3K27ac rather than an inability to recruit enhancer CBP/P300 at large where we also expect a loss in other active enhancer histone acetylations deposited by CBP/P300 such as H2B acetylation(Narita et al. 2022). Either way, it is unclear whether many of the sites dependent on MLL3/4 for active histone marks are dispensable for gene expression changes and or these sites are functional enhancers without an active histone signature. Of note, loss of MLL3/4 can also impact the recruitment of chromatin remodelers (i.e. the BAF complex) and enhancer-promoter looping, making the minimal impact MLL3/4 loss has on gene transcription during this transition even more surprising (Park et al. 2021; Local et al. 2018; Yan et al. 2018).

MLL4 KO embryos die shortly following gastrulation but show defects starting in early gastrulation(Lee et al. 2013; Ashokkumar et al. 2020, 3). In contrast, MLL3 KO embryos die perinatally due to developmental defects in the lung. MLL3/4 double KO embryos phenocopy the MLL4 KO consistent with the MLL4 being able to compensate for MLL3 loss in early development. The direct molecular basis for these phenotypes is poorly understood. However, it is important to note that even though MLL3/4 loss had little impact on gene activation during the ESC transition from naive to formative pluripotency, many genes were dysregulated in steady state naive ESCs and remained so in the formative state. In the case of persisting naive genes (Figure 2.1K), this is consistent with previous reports that MLL4 KOs fail to repress naive genes in LIF cultured

cells after withdrawal of 2i from naive (LIF+2i) culture conditions(Cao et al. 2018, 1). Some misregulated genes are likely direct targets given the downregulation of genes nearby MLL3/4 dependent enhancers upon loss of the enzymes. In addition, there are likely indirect effects potentially including additional unknown non-transcription related functions for MLL3/4. It is also important to note that other studies have shown important roles for MLL3/4 in the upregulation of gastrulation markers in the setting of embryoid body differentiation, a model for a later stage in early development than studied here(C. Wang et al. 2016). Even later in development, MLL3/4 is essential for developmental programs in a variety of tissues including cardiac, neural crest, muscle, and adipose specific programs(Lee et al. 2013; Ang et al. 2016; Lai et al. 2017; Shpargel et al. 2020). The differences between these studies and ours suggest a cell-context specific function of MLL3/4. A potential reason is the compendium of TFs that require or do not require MLL3/4 is different for each transition. For instance, PPARγ and CEBPα are major adipogenic factors that require MLL3/4 to function, while at least some of the pluripotency factors and the drivers of formative state are MLL3/4 independent based on our data. One possibility may be that transcriptional roles of MLL3/4 proteins are limited to determinants of differentiation pathways post-pluripotency and is most important in specifying somatic cell fates. Interestingly, a recent preprint suggests that MLL3/4 enzymatic activity is also important for a subset of cell fates such as extraembryonic endoderm and trophectoderm but not germ layer formation(Xie et al. 2022). Taken together, these data suggest fundamental and currently underappreciated differences in transcriptional control between pluripotent and somatic cell types.

Conclusions

In sum, our data demonstrates that cell-type specific enhancer activation is not uniquely performed by MLL3 and MLL4. Instead, gene regulation appears to be partitioned during the naive to formative transition. MLL3/4 is required for a small subset of genes critical for gastrulation and later developmental processes, but formative transcriptional activation is mediated by other mechanisms including MLL3/4 independent enhancers (Figure 2.12E). By identifying rules and exceptions to the current model of enhancer activation we highlight current gaps in knowledge concerning the molecular steps regulating these processes. Further investigation into the molecular interplay and steps governing enhancer function will advance our understanding of gene regulation in health and gene dysregulation in disease.

Methods

Cell Culture

Mouse ESCs were cultured in Knockout DMEM (Thermo Fisher, CAT#10829018) supplemented with 15% FBS, L-Glutamine, Penicillin/Streptamycin, NEAA, LIF (1000U/mL), and 2i (1uM MEK inhibitor PD0325901 and 3uM GSK3 inhibitor CHIR99021). Wildtype and dCD cells were R1 ESCs from a 129 strain background. MLL3$^{-/-}$; MLL4$^{fl/fl}$ ESCs were a mixed background of C57BL/6J and 129 strains. Formative cells were generated by removal of LIF and 2i. Briefly, 5e4 ESCs were plated per well of a 6 well plate on day −1 in LIF+2i media. To initiate differentiation, LIF and 2i were removed 24 hours after seeding (Day 0). Formative cells were collected on day 3 of differentiation, 63 hours after removal of LIF and 2i. To overcome proliferation defects 7.5e4 MLL3/4 DKO cells were plated per well of a 6 well. Naive cells were passaged and staged appropriately for simultaneous harvest. The expected amino acid substitutions in MLL3/4 dCD cells were validated using Sanger sequencing on amplicons generated from genomic PCR. To generate DKOs from MLL3$^{-/-}$; MLL4$^{fl/fl}$ ESCs we transfected with Rosa26-CRE-ERT2 plasmid linearized by EcoRI, selected with puromycin, and genotyped using genomic PCR. We then added tamoxifen for 3 days and picked clones to validate knockout of MLL4 floxed allele. Primers for genotyping listed in the supplemental table. Lines consistently tested negative for mycoplasm.

Crystal Violet Assay

To assess proliferation cells were plated in 24 well plates at approximately 12.5e4 cells per well. After 24, 48, and 72 hours, cells were then washed with PBS and 200ul Crystal Violet (0.2% Crystal Violet, 2% ethanol in dH2O) was added to the plate for 10 minutes at room temperature. Cells were then washed twice by gently submerging plate in tap water before 600ul 1% SDS was added to the well to solubilize the stain. Plate was placed on a shaker until color was uniform in the well and 200ul was transferred to a 96 well plate for reading absorbance at 570nm with a plate reader (SpectraMax M5). Absorbance for each time point was normalized to a blank well.

qPCR and analysis

To perform qPCR we first extracted RNA by adding Trizol directly to plates. After adding chloroform, an isopropanol precipitation with GlycoBlue was performed followed by ethanol washes. RNA pellets were resuspended in RNAse-free water and quantified using a NanoDrop. 200ug of RNA was used for cDNA synthesis using Maxima First Strand Synthesis Kit (Thermo Fisher, CAT#K1672) with half reactions according to the manufacturer's protocol. qPCR on cDNA was performed using SYBRgreen master mix (Applied Biosystems, CAT#A25742) using 6ul final volume on a QuantStudio5 qPCR machine (Applied Biosystems). qPCR primers for targets are listed in the supplemental table. Target Ct values were normalized to GAPDH internally for each sample and then set relative to the naive WT negative control.

Protein extraction and Westerns

To harvest protein for western assays, cells were trypsinized, washed once with ice cold PBS before adding RIPA buffer with protease inhibitors (Sigma CAT#P8340). After 15 minutes on ice, lysed cells were centrifuged at 16000g for 10 minutes and the supernatants representing whole cell fractions were collected and snap frozen in liquid nitrogen. Protein quantification was conducted using a Micro BCA protein assay kit (Thermo CAT#23235). 40ug of protein was loaded per well of SDS PAGE gels for western. For histone purification we utilized the acid extraction protocol exactly as described in Shechter et al. 2007 using 5e6 cells as input(Shechter et al. 2007).

Westerns were typically conducted using a Bio-Rad system with Tris-Glycine gels purchased from Bio-Rad and transferred to methanol activated PVDF membranes. Membranes were blocked, and stained with primaries and secondaries using Li-Cor Odyssey Blocking Buffer (Li-Cor, CAT#927-60001) mixed 1:1 with TBS. Primary antibodies for western were Rabbit anti-MLL3 (provided by Kai Ge), Rabbit anti-MLL4 (provided by Kai Ge), Mouse anti-OCT4 (BD Biosciences, CAT#611202), Rabbit anti-NANOG (Cell Signaling, CAT#61419), Rabbit anti-H3K4me1 (Abcam, CAT#ab8895), and Mouse anti-H3 (Cell Signaling, CAT#3738). Westerns for histones were conducted similarly to other targets except gels were transferred to PVDF membranes using CAPS buffer (500mM CAPS, adjusted to pH 10.5 with NaOH).

High molecular weight westerns for MLL3 and MLL4 were conducted using a NuPage SDS-PAGE and Transfer system with an XCell electrophoresis unit according to

manufacturer's protocols with several modifications (Thermo Fisher, CAT#E10002). In brief, 40ug of whole cell protein lysate was incubated with LDS loading buffer (Thermo Fisher, CAT#NP0007, added BME to 1% final concentration) and incubated at 70C for 10min. Samples were loaded into 3-8% Tris-acetate gels (Thermo Fisher, CAT#EA03752BOX) and ran in sample running buffer (Thermo Fisher, CAT#LA0041) supplemented with NuPage Antioxidant (Thermo Fisher, CAT#NP0005). Samples were run at 80V for 30min and then 120V for 120min. Samples were then transferred to PVDF membranes using NuPage Transfer Buffer (Thermo Fisher, CAT#NP0006, with added 10% Methanol, NuPage Antioxidant, 0.01% SDS) at 30V in a cold room for either 2 hours or overnight.

RNA sequencing

Total RNA was extracted and purified from cells using Trizol followed by ethanol precipitation. RNA-seq libraries were generated using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, CAT#A01172) according to their protocol using 200ng of total RNA for input. We utilized the PCR Add-on kit for Illumina (Lexogen, CAT#M02096) to determine an appropriate number of PCR cycles to amplify libraries. Amplified libraries were quantified using Agilent Tapestation 4200. Libraries were pooled and sequenced using a HiSeq 4000 to obtain single end 50bp reads. At least 10 million mapped reads or more per sample were obtained.

RNA sequencing processing and analysis

Adapters for sample FastQ files were trimmed using cutadapt v2.5 followed by alignment to the mm10 genome using STAR align 2.7.2b. Featurecounts from Subread v1.6.4 was used to generate a counts matrix of reads per gene. FastQC v0.11.8 and MultiQC were used to validate quality sequencing and mapping.

The gene count matrix was converted to CPM and filtered for genes greater than 1 cpm in at least two total samples. Samples were normalized using TMM. Next, Log2 CPM averages were calculated for replicates of each sample for scatterplot visualization and nearest neighbor TSS analysis. All transcriptomics analyses were conducted using CPM values from TMM normalization of all samples except for nearest neighbor analysis where WT and DKO were TMM normalized together. To conduct differential gene expression we performed DESeq2 v1.34.0 analysis using the raw gene count matrix as input for each desired comparison. Gene ontology was performed using ClusterProfiler 4.2.2. Custom R code for other downstream transcriptomics analyses and visualization provided on Github.

CUT&RUN sequencing and processing

CUT&RUN was conducted using the protocol from Skene et al. 2018 with the following modifications: Freshly trypsinized cells were bound to activated Concanavalin A beads (Bang Laboratories, #BP531) at a ratio of 2e5 cells/10ul beads in CR Wash buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine with protease inhibitors added) at room temp. An input of 2e5 cells were used per target. Bead-bound cells were then incubated rotating

overnight at 4C in CR Antibody buffer (CR Wash with 0.05% Digitonin, 2mM EDTA) containing primary antibody. We used the following antibodies for CUT&RUN: 1:100 Rabbit anti-H3K4me1 (Abcam, ab4729), Rabbit anti-MLL4 1:100 (provided by Kai Ge), and 1:100 Rabbit IgG isotype control (Abcam, ab171870). After primary, we washed 3 times 5 minutes each with cold CR Dig-wash buffer (CR Wash with 0.05% Digitonin) and incubated with pA-MNase (1:100 of 143ug/mL provided by Steve Henikoff) for 1 hour rotating at 4C. After MNase binding, we washed 3 times 5 each with cold CR Dig-wash buffer, and chilled cells down to 0C using a metal tube rack partially submerged in an ice water slurry. MNase digestion was induced by adding CaCl2 at a final concentration of 2mM. After 30 minutes of digestion, the reaction was quenched using Stop Buffer containing 340mM NaCl, 20mM EDTA, 4mM EGTA, 0.05% Digitonin, 100ug/mL RNAse A, 50ug/mL Glycogen, and approximately 2pg/mL Yeast spike-in DNA (provided by Steve Henikoff). The digested fragments for each sample were then extracted using a phenol chloroform extraction. Library preparation on samples was conducted using manufacturer's protocols for NEBNext Ultra II Dna Library Prep Kit (New England BioLabs, CAT#E7645) and NEB Multiplex Dual Index oligos (New England BioLabs, CAT#E7600, #E7780) with the following modifications. We input approximately 10ng of sample for half reactions, we diluted the NEBNext Illumina adaptor 1:25, we used the following PCR cycling conditions: 1 cycle of Initial Denaturation at 98C for 30 seconds, 12+ cycles of Denaturation at 98C for 10 seconds then Annealing/Extension at 65C for 10 seconds, and 1 final cycle of extension at 65C for 5 minutes. Following library preparation, double size selection was performed using Ampure beads and quality and

concentration of libraries were determined by an Agilent 4200 Tapestation with High-Sensitivity D1000 reagents before pooling for sequencing.

Fastq files for CUT&RUN samples were processed using Nextflow(Ewels et al. 2020) and the nf-core CUT&RUN pipeline v1.0.0 beta. In brief, adapters were trimmed using Trim Galore. Paired-end alignment was performed using Bowtie2 and peaks were called using SEACR with a peak threshold of 0.05 using spike in calibration performed using the E.coli genome K12. However, all downstream analysis was performed using CPM normalized samples. To reduce high background observed with MLL4 antibody we subtracted MLL4 CUT&RUN in DKO cells away from WT to generate formative MLL4 CUT&RUN bigWigs for use in heatmaps.

CUT&TAG sequencing and processing

CUT&TAG was conducted using the protocol from Kaya-Okur et al. 2020 with the following modifications: freshly trypsinized cells were bound to Concanavalin A beads at a ratio of 2e5 cells/7ul beads in CR wash at room temp. We used 2e5 cells as input per sample. Bead-bound cells were then incubated rotating overnight at 4C in CT Antibody buffer (CR Wash with 0.05% Digitonin, 2mM EDTA, 1mg/mL BSA) containing primary antibody. We used the following primary antibodies for CUT&TAG: 1:100 Rabbit anti-H3K27ac (Abcam, ab4729), 1:100 Rabbit anti-H3K4m3 (Abcam, ab8580), 1:100 Rabbit IgG isotype control (Abcam, ab171870). After primary, samples were washed 3 times for 5 minutes each using CR Dig-wash buffer and resuspended in 1:100 secondary antibody (Guinea pig anti-rabbit, Antibodies Online #ABIN101961) in CR Dig-wash buffer at 4C for

1 hour rotating at 4C. Samples were then incubated for 1 hour at 4C with 50ul of approximately 25nM homemade pA-Tn5 in CT Dig300 wash buffer (20mM HEPES, 300mM NaCl, 0.01% Digitonin, 0.5mM Spermidine with Roche cOmplete protease inhibitors added). Recombinant Tn5 was purified and loaded with adapters as previously described (Kaya-Okur et al. 2019). After Tn5 incubation, samples were washed 3 times for 5 minutes each with CT Dig300 wash buffer. Tagmentation was then initiated for 1hr at 37C in a thermocycler by adding MgCl2 to 10mM final concentration in 50uL volume. The tagmentation reaction was quenched immediately afterwards by adding 1.6ul of 0.5M EDTA, 1ul of 10mg/mL Proteinase K, and 1ul of 5% SDS. Samples were then incubated at 55C for 2 hours in a thermocycler to denature Tn5 and solubilize tagmented chromatin. After incubation, samples were magnetized and the supernatant was transferred to new wells where SPRI bead purification was performed using homemade beads to select all DNA fragment lengths larger than 100bp. Samples were eluted in 0.1X TE and approximately half of each sample was used for library preparation using NEBNext HIFI Polymerase with custom indices synthesized by IDT. An appropriate number of cycles for each target was chosen to prevent overamplification bias. After amplification, libraries were purified with 1.2x homemade SPRI beads to select for fragments >250bp and eluted in 0.1X TE. Quality and concentration of libraries were determined by an Agilent 4200 Tapestation with D1000 reagents before pooling for sequencing.

CUT&TAG samples were processed similarly to CUT&RUN samples using the same Nextflow pipeline.

ATAC-seq sequencing and processing

ATAC-seq libraries were generated using the Active Motif commercial kit following the provided protocol with kit components (Active Motif, CAT#53150): Specifically, freshly trypsinized cells were washed with ice cold PBS using 75K cells as input for each sample. Cells were lysed using ice cold ATAC Lysis Buffer and we added 50ul of Tagmentation Master Mix to each sample. We incubated tagmentation reactions at 37C for 30min in a thermomixer set to 800 rpm. We immediately transferred samples to a new tube and performed DNA column purification. After purification we amplified the libraries using the following PCR conditions: 1 cycle at 72C for 5 minutes, 1 cycle at 98C for 30 seconds, and finally 10 cycles of 98C for 10 seconds, 63C for 30 seconds, 72C for 1 minute. Amplified libraries were purified using 1.2x SPRI bead solution. Quality and concentration of libraries were determined by an Agilent 4200 Tapestation with D1000 reagents before pooling for sequencing.

Fastq files for ATAC-seq samples were processed using Nextflow and the nf-core ATAC-seq pipeline v1.2.1. In brief, adapters were trimmed using Trim Galore. Paired-end alignment was performed using BWA and peaks were called using MACS2 in broad peak mode with a cutoff value of 0.1. BAM files were converted to bigWigs. Using read counts from consensus peaks of WT and DKO samples we performed TMM normalization using EdgeR to generate scale factors used during bigWig generation of samples.

ChIP-seq processing

Fastq files for ChIP-seq from previously published studies were processed using Nextflow and the nf-core ChIP-seq pipeline v1.2.1. In brief, adapters were trimmed using Trim Galore. Paired-end alignment was performed using BWA, duplicates were removed, and peaks were called using MACS2 in broad peak mode with a cutoff value of 0.1. We subtracted the MLL3/4 ChIP-seq in DKO cells away from WT samples to generate naive MLL3/4 ChIP-seq bigWigs for use in heatmaps.

Peak analysis

We performed differential signal enrichment analysis using Diffbind on SEACR peaks derived from CUT&RUN or CUT&TAG samples. We then used the output from Diffbind, containing the union of input SEACR peaks, for two packages: 1) We used Deeptools multiBigwigsummary to count reads at each peak for the different targets/samples and 2) we used HOMER annotatePeaks to identify nearest gene and annotate genomic features. These files were then imported into R and joined together by peak coordinates to make a dataframe which we used to filter for phenotypes and calculate fold changes. After genomic feature annotation by Homer we classify intergenic and intronic peaks as “Distal”, TSS as “Promoter”, and everything else as “Other” including: exon, transcriptional termination sites (TTS), UTR regions, and those without assigned features. To calculate Log2CPM peak densities we took counts for each peak from multibigwigsummary multiplied them by a constant (1e6), added 1, and then performed Log2(x). Finally, we divided by the width in basepairs of the peak. Fold changes were conducted using these Log2CPM peak density values. H3K27ac CUT&TAG data was

subset by H3K27ac peaks that overlapped at least 75% of an WT ATAC-seq peak using bedops. We also used bedops on ATAC subset H3K27ac sites to identify H3K27ac peaks that had overlapping H3K4me1 peaks to generate our H3K4me1/H3K27ac+ peak lists. ChromHMM segmentations were similarly filtered for chromatin states that overlapped at least 75% of an WT ATAC-seq peak.

Acknowledgements

We thank Kai Ge for providing the MLL3KO;MLL4 floxed ESC line and antibodies for MLL3 and MLL4. We thank Joanna Wysocka for providing the MLL3/4 catalytic dead cell line. We thank Steve Henikoff for sending purified pA-MNase. We would like to thank the Nextflow community for helpful advice on processing of sequencing samples including Chris Cheshire and Charlotte West for providing access to a beta for nf-core CUT&RUN pipeline. We would like to thank Bryan Marsh, Deniz Gökbuget, and Brian Deveale for helpful discussions throughout the project including proofreading the original manuscript. We would like to thank Dan Lim, Yin Shen, and Elphege Nora for their thoughtful comments and constructive feedback on the manuscript.

Figures

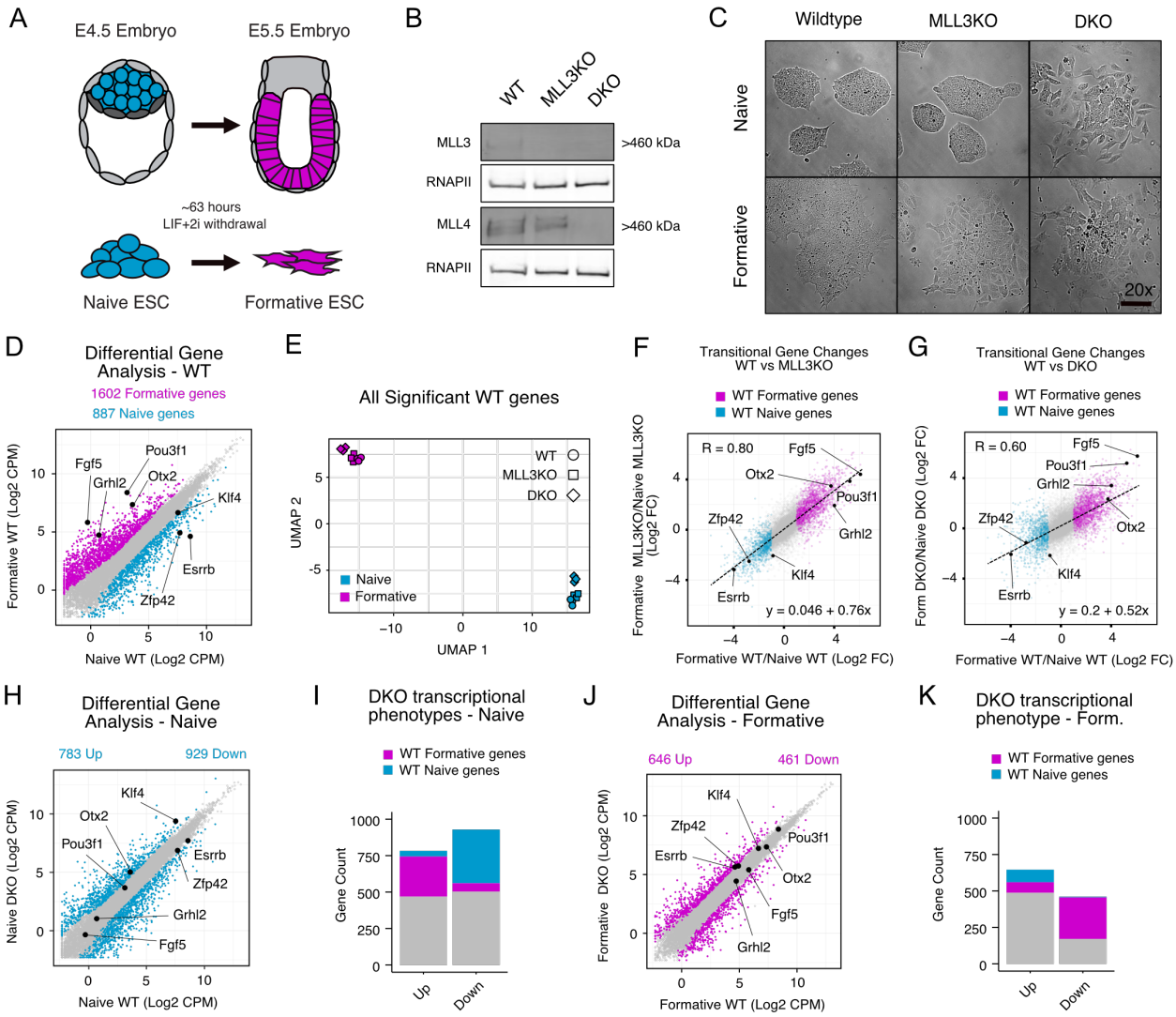

Figure 2.1: MLL3/4 is dispensable for transcriptional activation of much of the formative program.

A) Schematic of naïve to formative transition in vivo and in vitro. B) Western validation of MLL3 and MLL4 knockout cell lines. C) Brightfield microscopy of cell lines in naive and formative conditions, 20x, scale bar = 100um. D) Differential gene expression analysis (DGE) on WT naive and formative RNA-seq samples (significant genes colored, P.adj < 0.05, Log2 Foldchange (Log2FC) >1). E) UMAP analysis on all RNA-seq samples subset by all naive and formative genes from WT DGE. F,G) Fold change (formative/naive) of all genes for either MLL3KO and DKO compared to WT. WT formative and naive genes from WT DGE analysis colored. R, Pearson's coefficient. Dashed line and linear equation represent linear model of all genes. H) DGE analysis on WT naive and DKO naive samples (significant genes colored, P.adj < 0.05 and

Log2FC > 1). I) Naive state DKO misexpressed genes categorized by whether they are also WT naive or formative genes. J,K) Same as H,I respectively but comparing within the formative state.

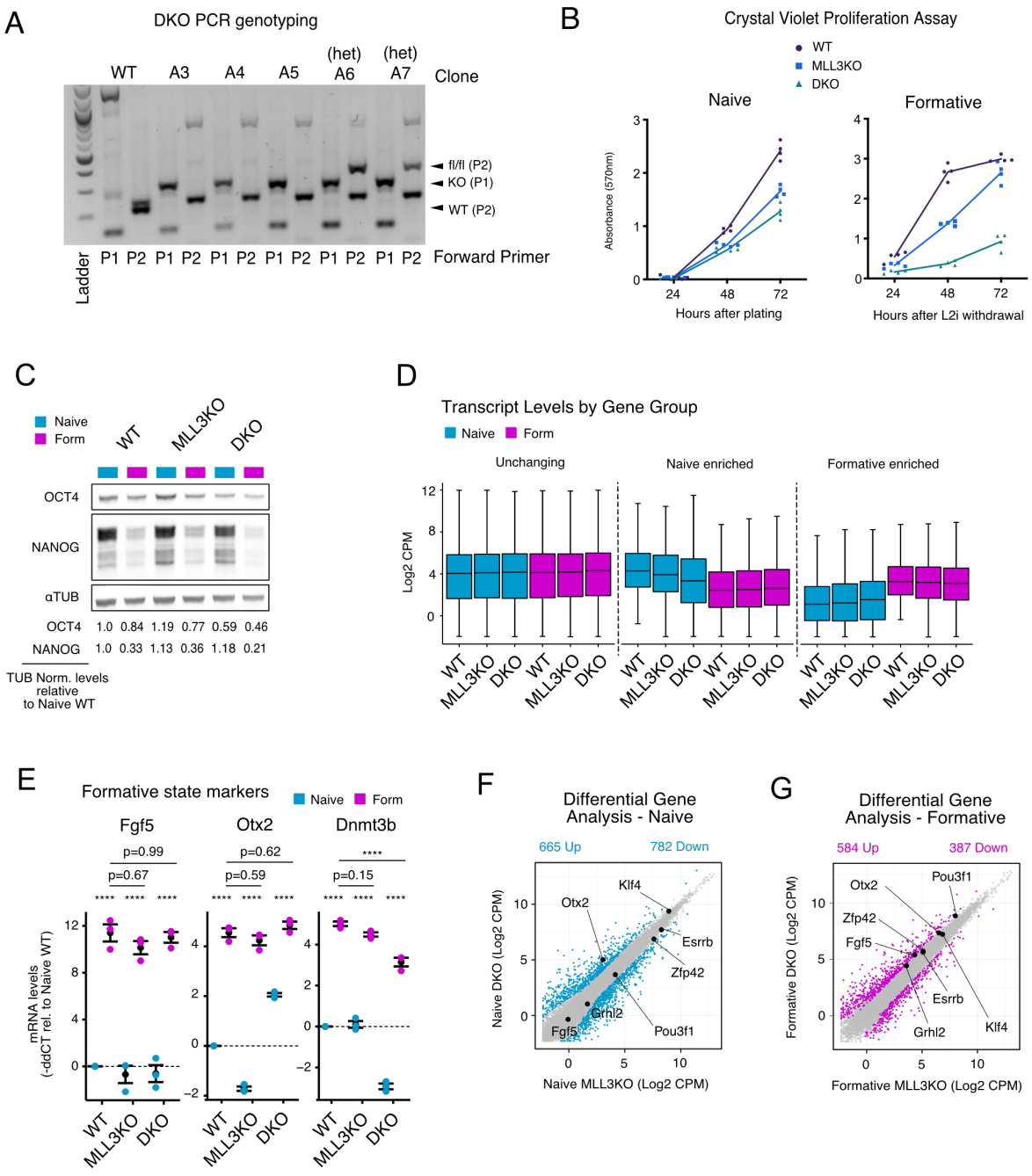

Figure 2.2 – Supplement for MLL3/4 is dispensable for transcriptional activation of much of the formative program.

A) Confirmation of multiple DKO clones using genomic PCR following tamoxifen treatment. B) Proliferation measurements using crystal violet assays. C) Westerns on whole cell protein fractions of OCT4, NANOG and TUBULIN loading control. Quantifications below are fold changes relative to naÔve WT after normalizing OCT4 or NANOG for TUBULIN levels in each lane. D) qPCR of formative enriched markers Fgf5, Otx2, Dnmt3b. For each marker selected comparisons are shown resulting from a Two-

way ANOVA, Tukey's multiple comparison test for all samples. Mean and S.E.M. in black. Additional statistics provided in supplemental table. E) Boxplots of transcript levels in Log2CPM for unchanging, naive enriched or formative enriched genes. F) DGE analysis on MLL3KO naive and DKO naive samples (significant genes colored, P.adj < 0.05 and Log2FC > 1). G) Same as F but comparing the formative state. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

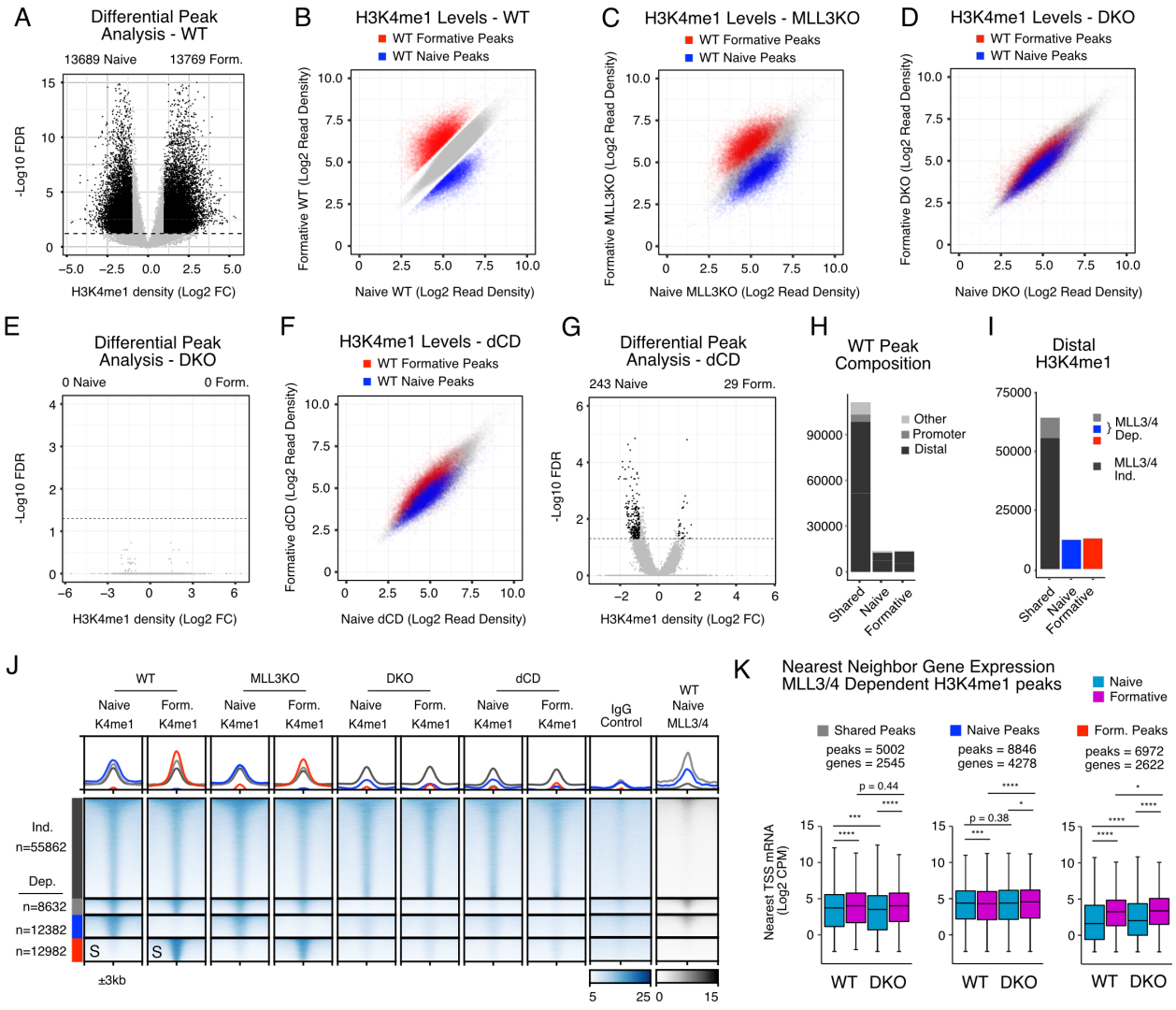

Figure 2.3 - MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent transition.

A) Differential signal enrichment between naive and formative WT H3K4me1 peaks identifies significant peaks (Black, FDR <0.05, Log2 Foldchange (Log2FC) > 1). B) Scatterplot of H3K4me1 peak intensities between naive and formative WT samples. (gray, shared peaks, FDR > 0.1 & Log2FC < 0.7, red or blue, FDR < 0.05 & Log2FC > 1). C,D) H3K4me1 signal in naive and formative MLL3KO or DKO samples at peaks called from WT. Colors correspond to peak categories derived from WT. E) Differential signal enrichment of naive and formative DKO samples. F) Same as C,D but with naive and formative dCD cells. G) Differential signal enrichment of naive and formative dCD samples. H) Feature annotation of WT peaks stratified by peak category including intergenic/intronic (Distal), Promoter, and all else (Others). I) Distal H3K4me1 peak categories further stratified by MLL3/4 independent (DKO/WT Log2FC > -0.7) or dependent (DKO/WT Log2FC < -1). J) Heatmap of MLL3/4 independent and dependent H3K4me1 peak categories, rows sorted on “S” columns. All heatmap values and range

are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. K) nearest neighbor TSS analysis of expression levels in Log2CPM for each RNA-seq dataset near H3K4me1 peak categories. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

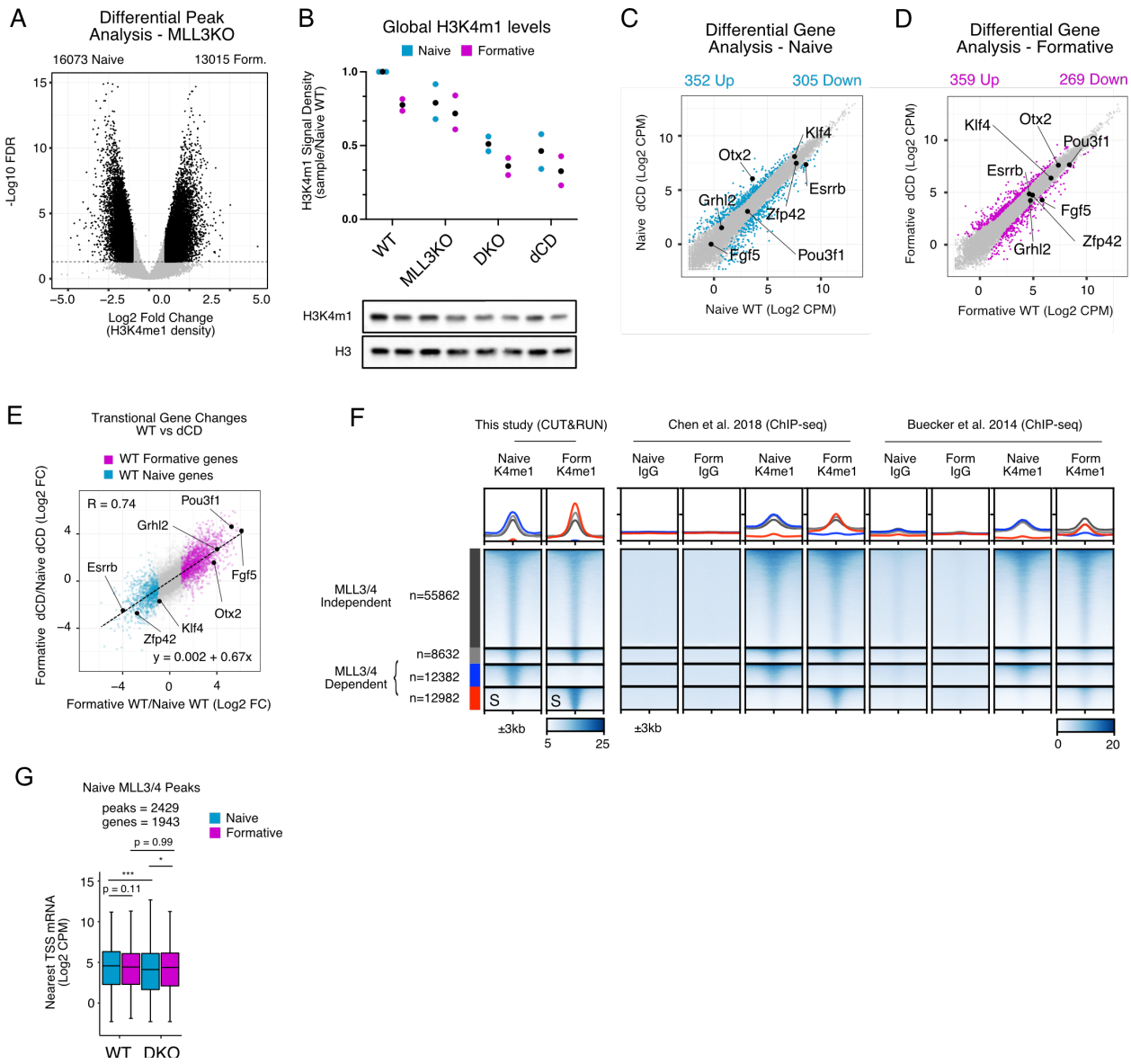

Figure 2.4 – Supplement for MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent transition.

A) Diffbind analysis of H3K4me1 signal in MLL3KO cells at WT peaks. B) Quantifications of westerns on histones purified by acid-extraction. H3K4me1 levels are relative to naÔve WT levels after first normalizing by H3 signal in each lane. Black dot represents mean (n=2). C) mRNA Log2CPM of naive MLL3/4 dCD cells vs. WT cells. Significant genes highlighted. D) same as B with formative samples. E) Foldchange (formative/naive) of all genes for MLL3/4 dCD compared to WT. WT formative and naive genes from WT DGE analysis colored. R, Pearson's coefficient. Dashed line and linear equation represent linear model of all genes. F) Heatmap of ChIP-seq for H3K4me1 in naive and formative cells from two published studies. All heatmap values and range are

in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. G) Nearest neighbor TSS analysis of expression levels in Log2CPM for each RNA-seq dataset near naive MLL3/4 ChIP-seq peak categories. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

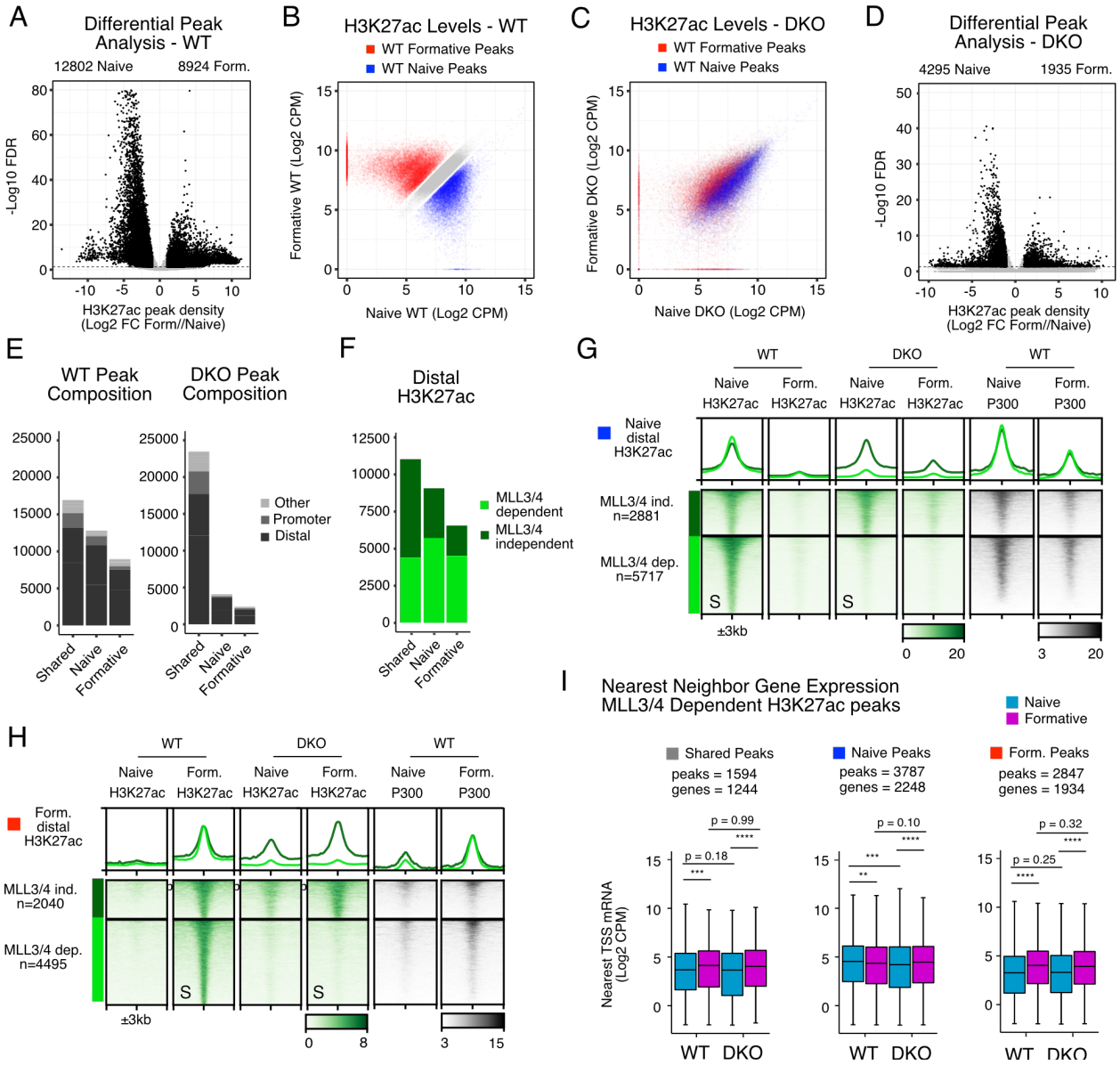

Figure 2.5 - MLL3/4 dependent and independent distal H3K27ac deposition.

A) Differential signal enrichment between naive and formative WT H3K27ac peaks identifies significant peaks (Black, FDR <0.05, Log2 Foldchange (Log2FC) > 1). B) Scatterplot of H3K27ac peak intensities between naive and formative WT samples. (gray, shared peaks, FDR > 0.1 & Log2FC < 0.7, red or blue, FDR < 0.05 & Log2FC > 1). C) H3K27ac signal in naive and formative DKO samples at peaks called from WT. Colors correspond to peak categories derived from WT. D) Feature annotation of WT or DKO peaks stratified by peak category including intergenic/intronic (Distal), Promoter, and all else (Others). E) Distal H3K27ac peak categories further stratified by MLL3/4 independent (DKO/WT Log2FC > -0.7) or dependent (DKO/WT Log2FC < -1). G,H) Heatmap of MLL3/4 independent and dependent H3K27ac peak categories for naive or

formative specific peaks, rows sorted on “S” columns. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. I) nearest neighbor TSS analysis of expression levels Log2CPM near H3K27ac peak categories. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

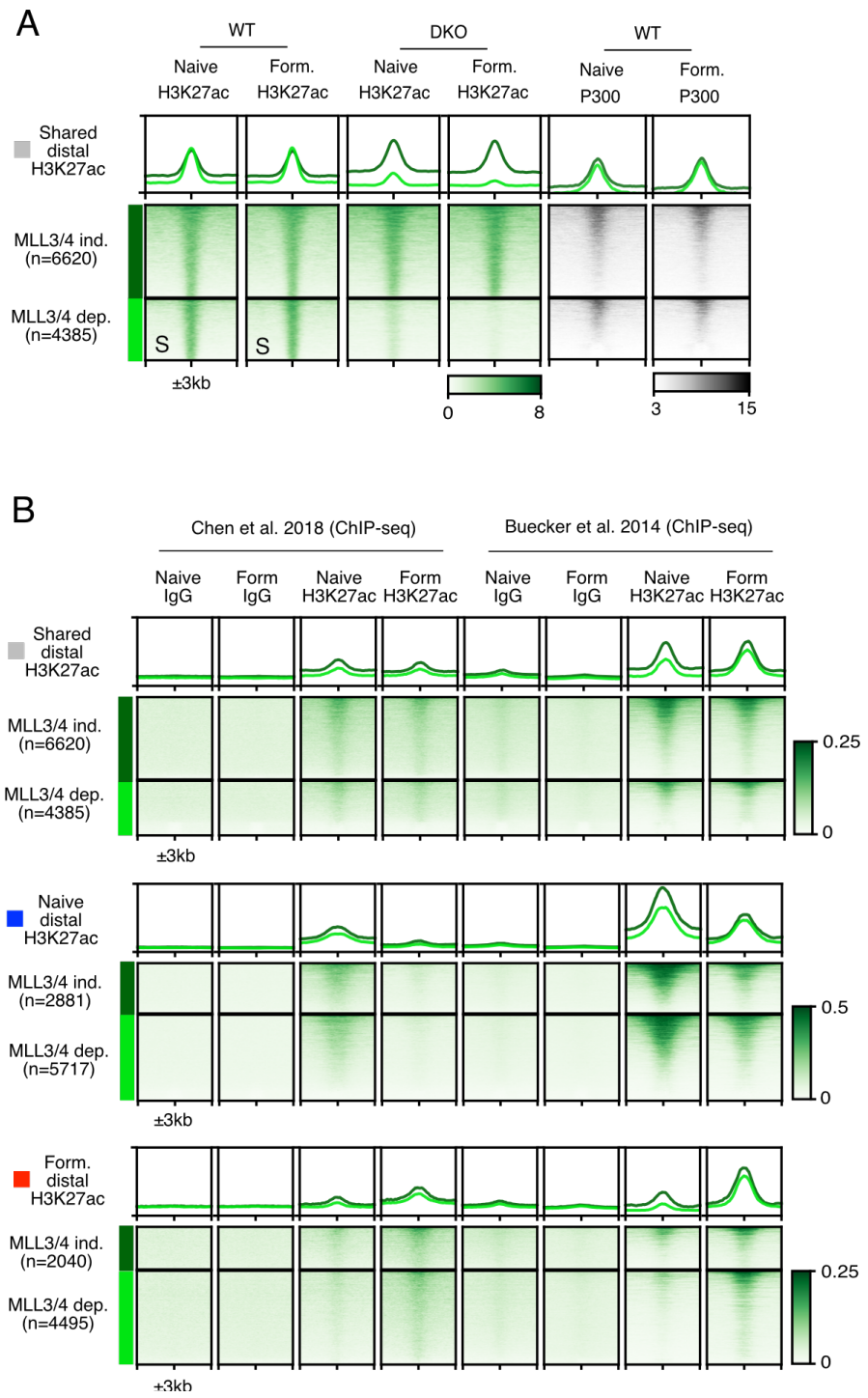

Figure 2.6 – Supplement for MLL3/4 dependent and independent distal H3K27ac deposition.

A) Heatmap of shared H3K27ac sites clustered by MLL3/4 independent or dependent H3K27ac. B) Heatmaps of ChIP-seq for H3K27ac in naive and formative cells from two

published studies. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor.

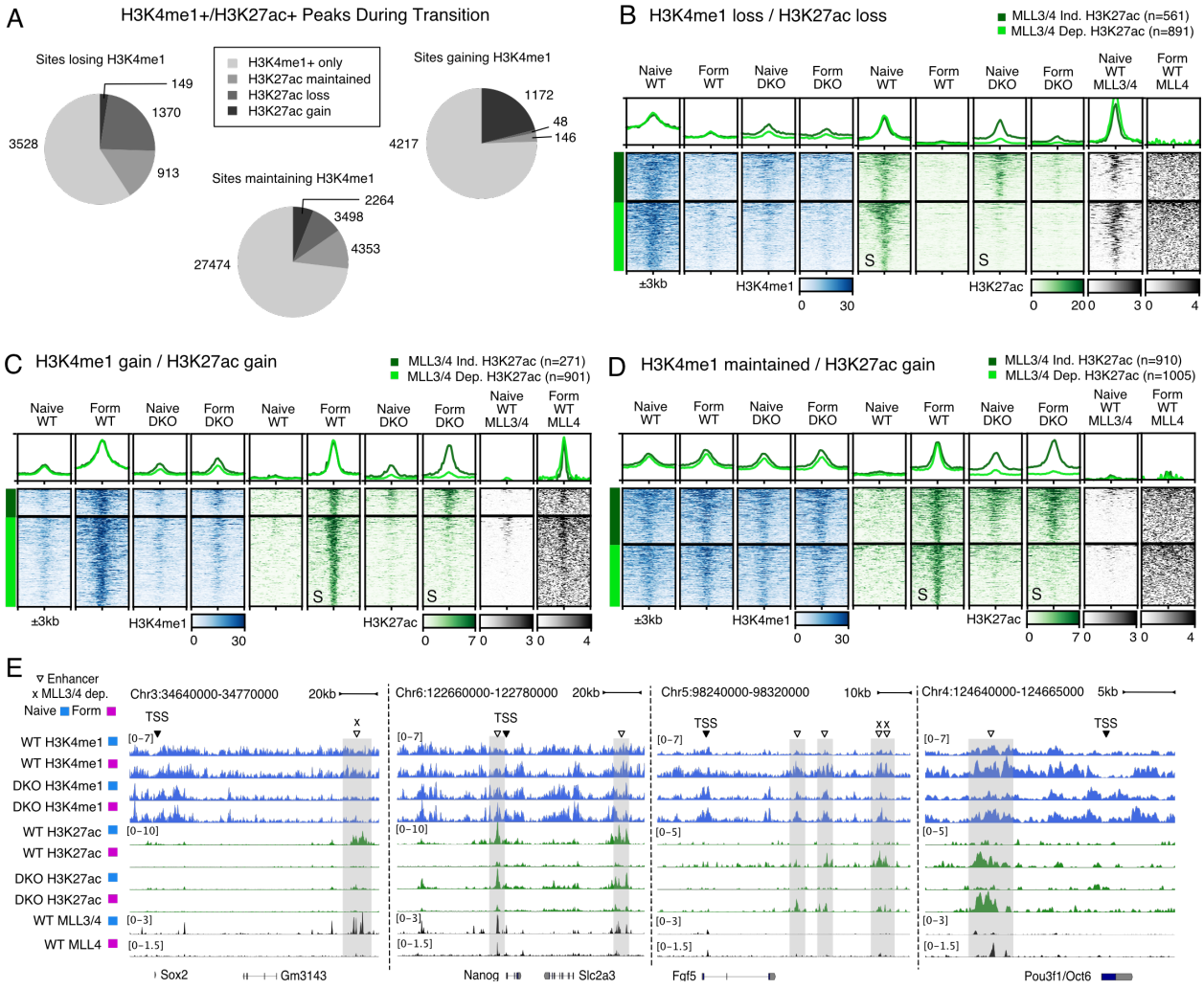

Figure 2.7 - Enhancer Activation Can Occur Independently of MLL3/4.

A) Pie charts showing fraction of coincident changes in H3K27ac within the different categories of changing H3K4me1 following MLL3/4. B) Heatmap of sites that normally lose H3K4me1 and H3K27ac during naive to formative transition clustered by either MLL3/4 dependent or independent H3K27ac. Rows sorted on “S” columns. Y-axis scale of metagene profiles above is same as signal ranges of heatmap, values in CPM. C) Same as B but using sites that normally gain H3K4me1 and H3K27ac during naive to formative transition clustered by either MLL3/4 dependent or independent H3K27ac. D) Same as B but using sites that have unchanging H3K4me1 but gain H3K27ac during naive to formative transition clustered by either MLL3/4 dependent or independent H3K27ac. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. E) Naive and formative examples of functionally validated enhancers that are either MLL3/4 dependent or independent. Naive MLL3/4 dependent Sox2 enhancer cluster, Naive MLL3/4 independent Nanog enhancers, Formative independent and dependent enhancers of Fgf5, formative independent enhancer of Pou3f1/Oct6.

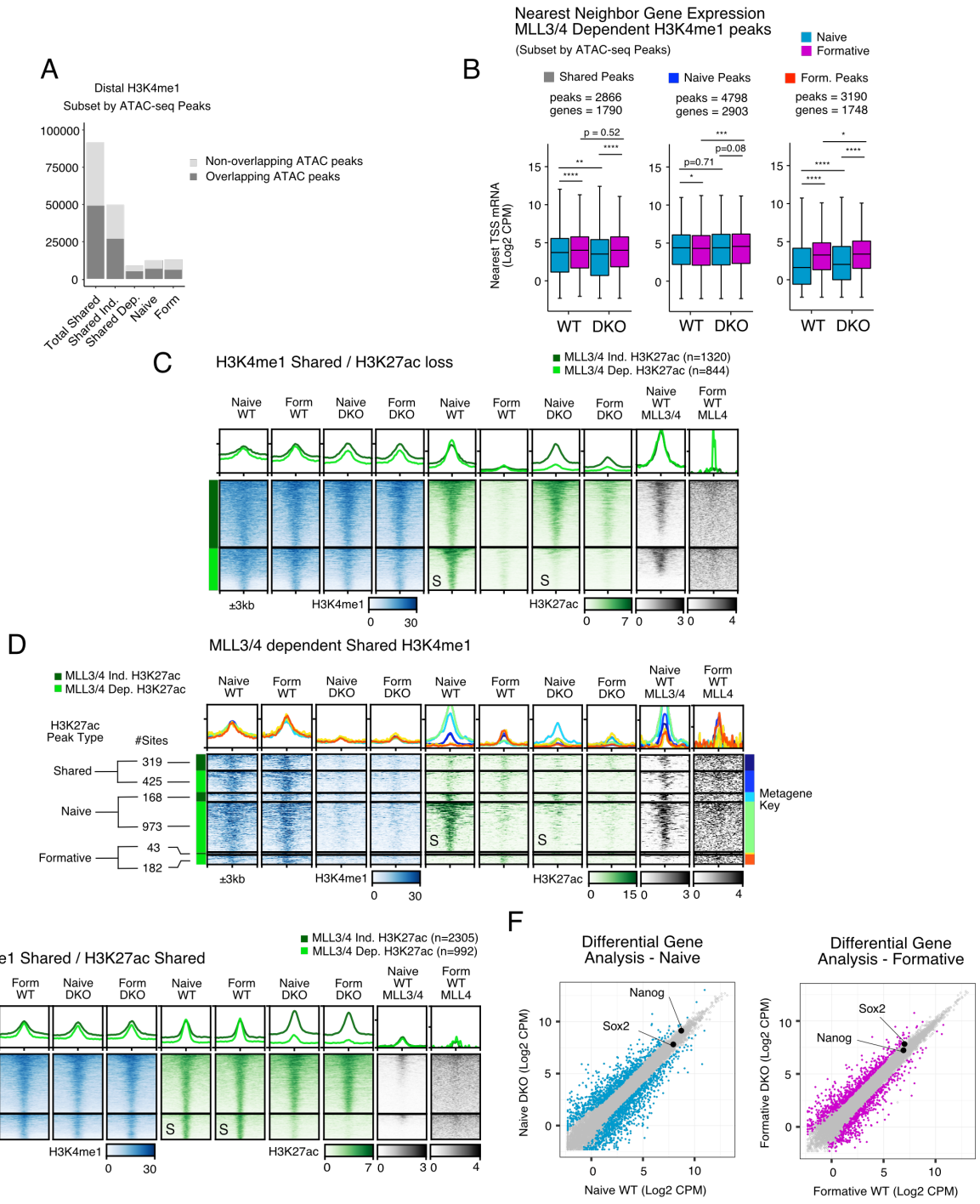

Figure 2.8 – Supplement 1 for Enhancer Activation Can Occur Independently of MLL3/4.

A) Peak count of H3K4me1 categories from Fig2. after filtering for distal H3K4me1 overlapping WT ATAC peaks. B) Nearest neighbor TSS analysis of expression levels in Log2CPM for each RNA-seq dataset near H3K4me1 peak categories after subsetting for ATAC peak overlap. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. C) Heatmaps for shared H3K4me1 that overlaps with naive

enriched H3K27ac, clustered by MLL3/4 independent or dependent H3K27ac. D) Heatmaps for MLL3/4 dependent H3K4me1 overlapped with sites that have shared, naive, or formative H3K27ac. Clustered additionally by MLL3/4 independent or dependent H3K27ac. E) Heatmaps for shared H3K4me1 that overlaps with shared H3K27ac, clustered by MLL3/4 independent or dependent H3K27ac. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. F) Log2CPM of Sox2 and Nanog in naive and formative samples comparing WT and DKO cells. Sox2 is significantly up in the formative state. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

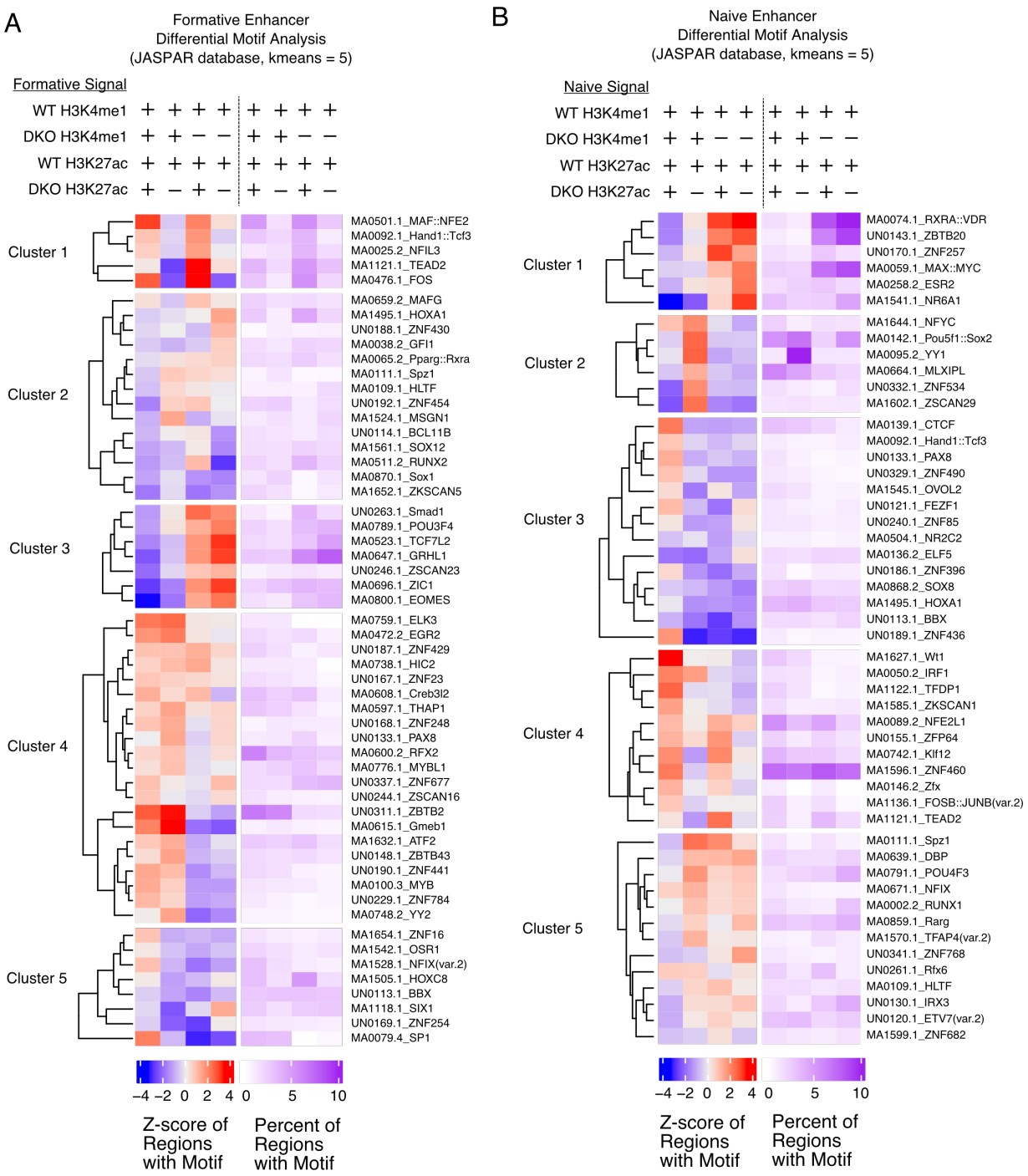

Figure 2.9 – Supplement 2 for Enhancer Activation Can Occur Independently of MLL3/4.

A) Clusters from differential motif analysis using Gimmemotifs on formative enriched MLL3/4 independent and dependent enhancers marked by both H3K4me1 and H3K27ac (Sites from Figure 4C and 4D). Motifs derived from JASPAR 2020 database, k-means

clustering used. B) Same as A but with naive enriched independent and dependent enhancers (using sites from Figure 4B and Figure S5B).

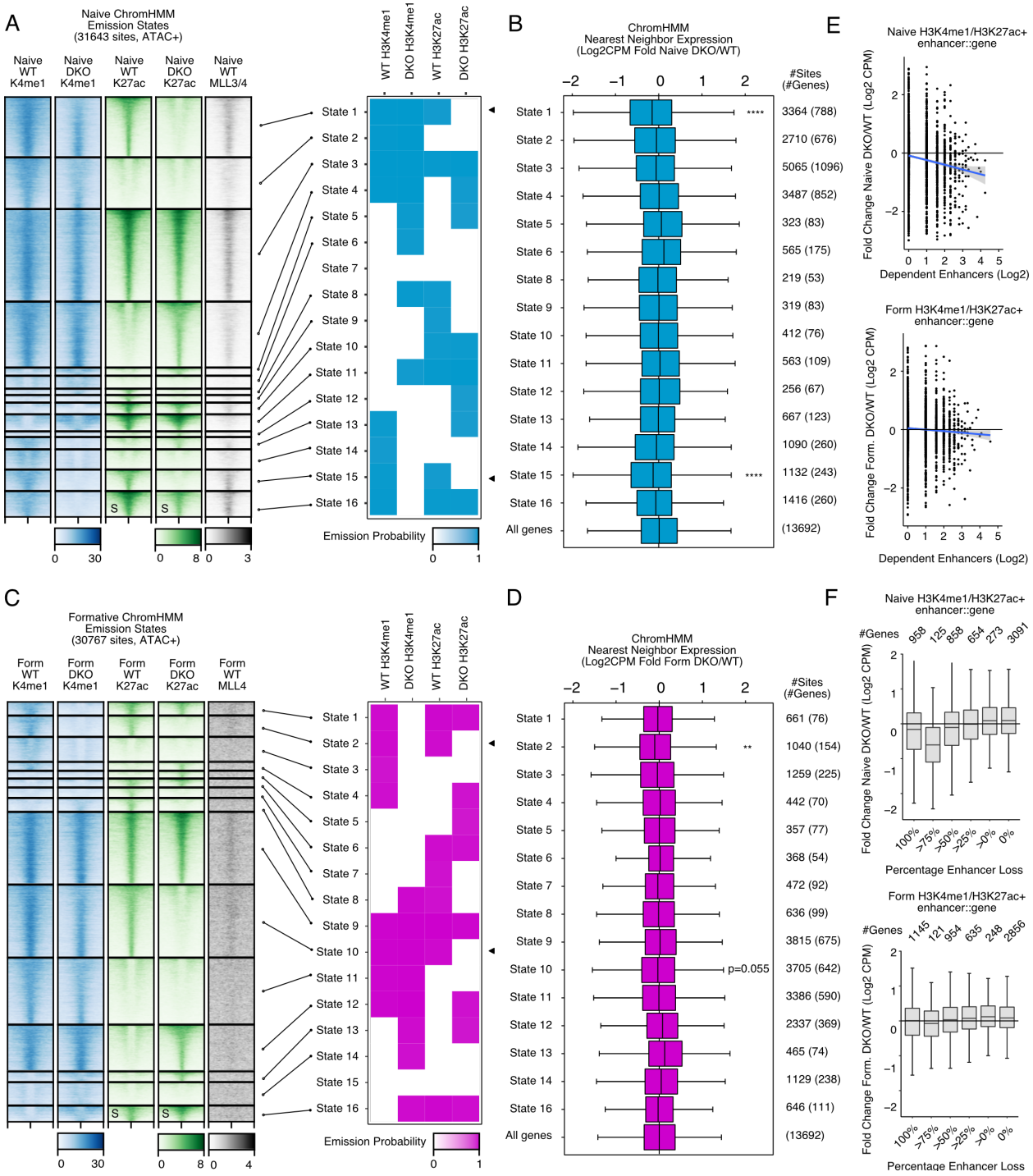

Figure 2.10 - Distal H3K4me1 and H3K27ac are not functionally coupled with formative transcriptional activation.

A) Naive ChromHMM emissions for 16 states yields all combinations of H3K4me1 and H3K27ac in WT and DKO. Heatmaps are clustered by each chromatin state. Black arrows denote states with MLL3/4 dependent H3K27ac. B) Gene expression changes of nearest TSS for all sites in each given state. Monte Carlo permutation sampling of “all

genes control” to perform Mann-Whitney U Test for each state. Benjamini-Hochberg corrected. Selected statistics shown, additional statistics provided in Additional File 2. C) Same as A with Formative samples. D) Same as B with Formative RNA-seq. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. ChromHMM states without any H3K4me1 or H3K27ac emission probability are excluded. E) For all H3K4me1/H3K27ac+ sites in either Naive (top panel) or Formative (bottom panel), the total number of MLL3/4 dependent enhancers for a gene compared with the fold change of RNA levels DKO/WT in Log2CPM. Each dot represents one gene. Blue line represents generalized linear model, gray 95% confidence interval. F) Boxplots of relative expression DKO/WT of RNA levels for all genes associated with any H3K4me1/H3K27ac+ peak in either the naïve (top panel) or formative state (bottom panel). Each gene is binned by the percentage of their associated enhancer loss in DKOs. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

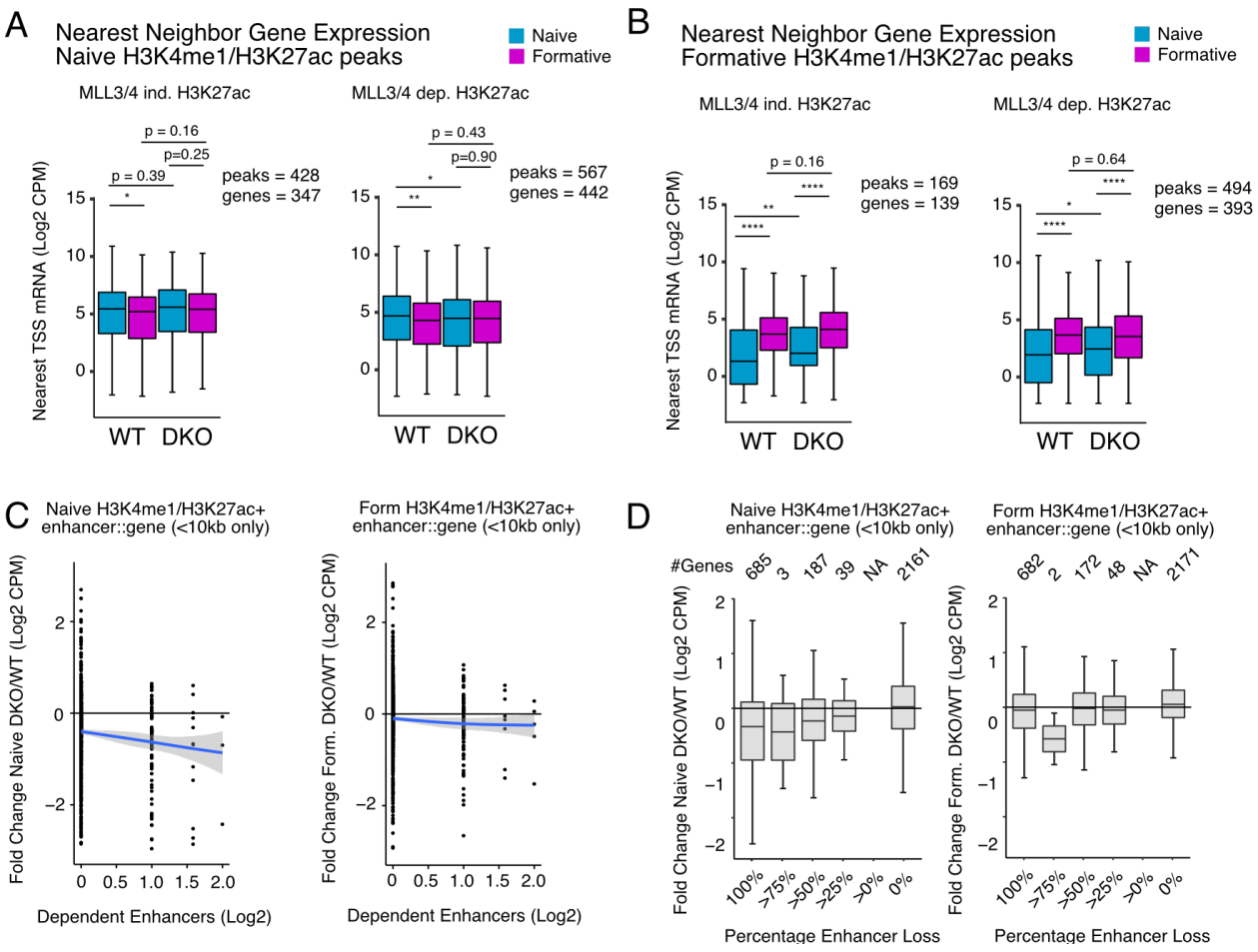

Figure 2.11 – Supplement for Distal H3K4me1 and H3K27ac are not functionally coupled with formative transcriptional activation.

A) Nearest neighbor TSS analysis of expression levels in Log2CPM near naive H3K4me1 dependent sites that are either have MLL3/4 independent (left panel) or dependent H3K27ac (right panel)(clusters from Figure 2.7B). Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. B) Same as A but at formative H3K4me1 dependent sites (clusters from Figure 2.7C). C) The relative expression DKO/WT of RNA levels for all genes associated with any H3K4me1/H3K27ac+ peak in either naive or formative state compared with the number of dependent enhancers for each gene. Only dependent enhancers within 10kb of the nearest TSS are considered. Each dot represents one gene. Blue line represents generalized linear model, gray 95% confidence interval. D) Boxplots of relative expression DKO/WT of RNA levels for all genes associated with any H3K4me1/H3K27ac+ peak in either the naÔve (left) or formative state (right). Each gene is binned by the percentage of their associated enhancer loss in DKOs using only enhancers within 10kb of the nearest TSS. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

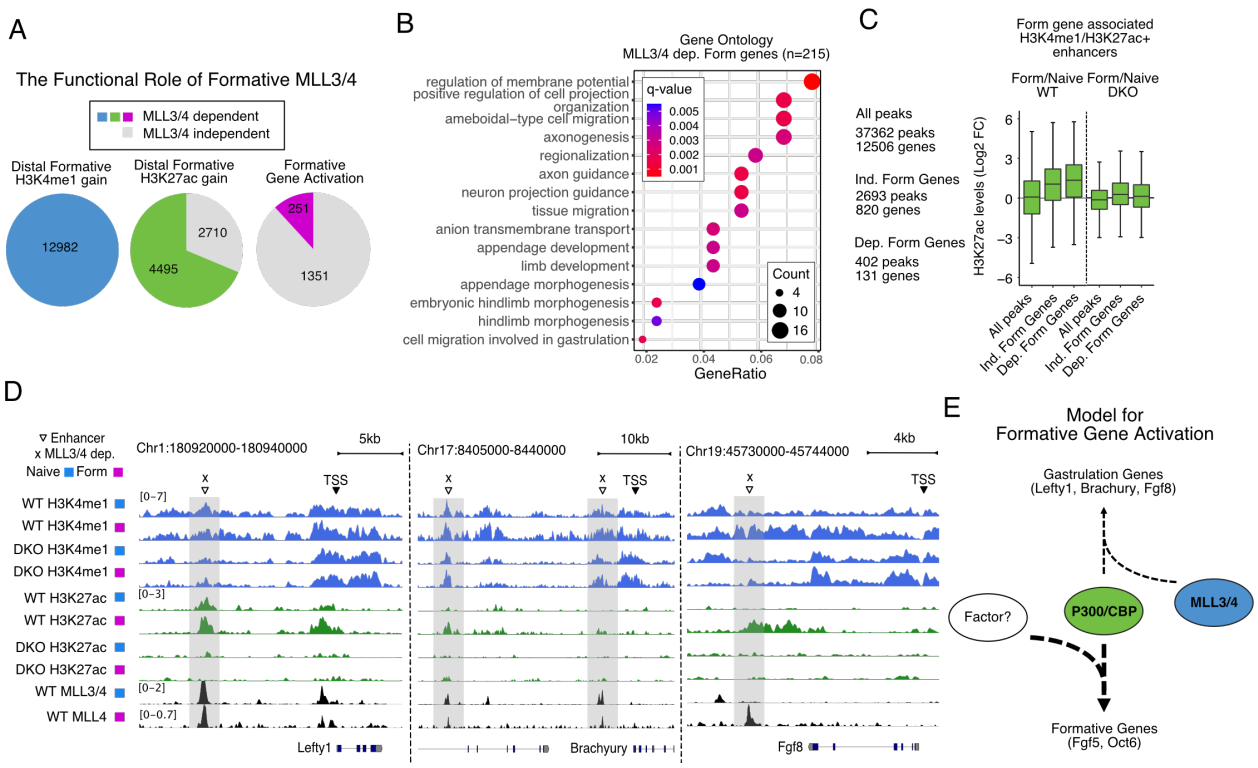

Figure 2.12 - Gene-centric analysis reveals a subset of distal loci associate with MLL3/4 dependent formative genes.

A) Pie charts showing fraction of normally gained H3K4me1, H3K27ac, or gene expression that are dependent on MLL3/4 or not. 215 formative genes represent those of the 283 formative down genes from Figure 2.1K without preexisting naive expression defects (Log2CPM naive DKO/WT > -1). B) Clusterprofile analysis of Biological Processes Gene Ontology for 215 MLL3/4 dependent formative genes. C) Fold change Log2CPM of H3K27ac density for H3K4me1/H3K27ac+ enhancers associated with formative genes that gained expression during transition in an either MLL3/4 independent or dependent fashion. Only 131 out of 215 dependent genes and 820 out of 1387 independent genes were able to be associated with an enhancer. D) Genome tracks for known and predicted formative targets of MLL3/4 Lefty1, Brachyury/T, and Fgf8. E) Model for the role of MLL3/4 independent and dependent mechanisms in the regulation of gained gene expression during the naïve to formative transition.

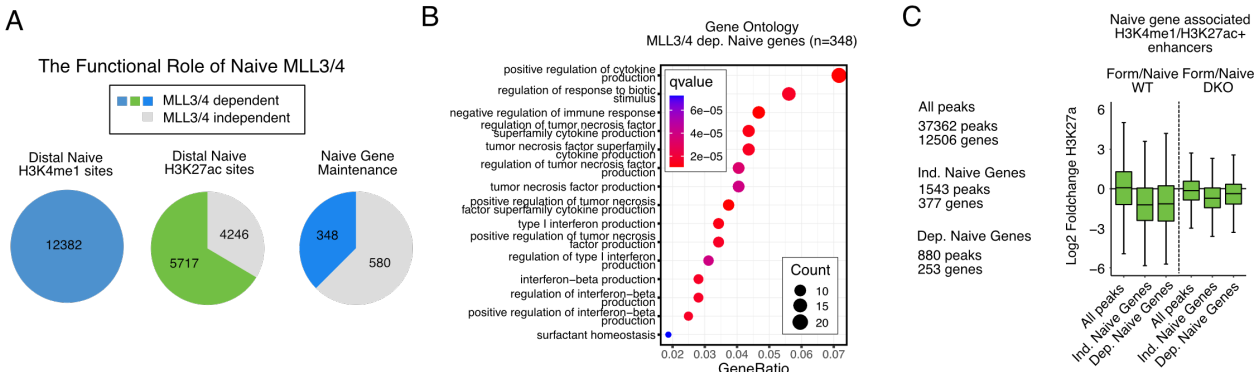

Figure 2.13 – Supplement for Gene-centric analysis reveals a subset of distal loci associate with MLL3/4 dependent formative genes.

A) A major percentage of naive enriched peaks fail to maintain distal H3K4me1 and H3K27ac with moderate consequences on naive gene expression. B) Clusterprofile analysis of Biological Processes Gene Ontology for 348 MLL3/4 dependent naive genes. C) Fold change Log2CPM of H3K27ac density for H3K4me1/H3K27ac+ enhancers that are associated with naive genes whose expression in the naive state is either MLL3/4 independent or dependent. Only a subset of genes in each category were able to be associated with an H3K4me1/H3K27ac+ enhancer.

References

Agrawal, Puja, Steven Blinka, Kirthi Pulakanti, Michael H. Reimer, Cary Stelloh, Alison E. Meyer, and Sridhar Rao. 2021. “Genome Editing Demonstrates That the -5 Kb Nanog Enhancer Regulates Nanog Expression by Modulating RNAPII Initiation and/or Recruitment.” *The Journal of Biological Chemistry* 296 (June): 100189. https://doi.org/10.1074/jbc.RA120.015152.

Ang, Siang-Yun, Alec Uebersohn, C. Ian Spencer, Yu Huang, Ji-Eun Lee, Kai Ge, and Benoit G. Bruneau. 2016. “KMT2D Regulates Specific Programs in Heart Development via Histone H3 Lysine 4 Di-Methylation.” *Development (Cambridge, England)* 143 (5): 810–21. https://doi.org/10.1242/dev.132688.

Ashokkumar, Deepthi, Qinyu Zhang, Christian Much, Anita S. Bledau, Ronald Naumann, Dimitra Alexopoulou, Andreas Dahl, et al. 2020. “MLL4 Is Required after Implantation, Whereas MLL3 Becomes Essential during Late Gestation.” *Development (Cambridge, England)* 147 (12): dev186999. https://doi.org/10.1242/dev.186999.

Bleckwehl, Tore, Giuliano Crispatzu, Kaitlin Schaaf, Patricia Respuela, Michaela Bartusel, Laura Benson, Stephen J. Clark, et al. 2021. “Enhancer-Associated H3K4 Methylation Safeguards in Vitro Germline Competence.” *Nature Communications* 12 (October): 5771. https://doi.org/10.1038/s41467-021-26065-6.

Blinka, Steven, Michael H. Reimer, Kirthi Pulakanti, and Sridhar Rao. 2016. “Super-Enhancers at the Nanog Locus Differentially Regulate Neighboring Pluripotency-

Associated Genes." *Cell Reports* 17 (1): 19–28. https://doi.org/10.1016/j.celrep.2016.09.002.

Bruse, Niklas, and Simon J. van Heeringen. 2018. "GimmeMotifs: An Analysis Framework for Transcription Factor Motif Analysis." bioRxiv. https://doi.org/10.1101/474403.

Buecker, Christa, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. 2014. "Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency." *Cell Stem Cell* 14 (6): 838–53. https://doi.org/10.1016/j.stem.2014.04.003.

Cao, Kaixiang, Clayton K. Collings, Marc A. Morgan, Stacy A. Marshall, Emily J. Rendleman, Patrick A. Ozark, Edwin R. Smith, and Ali Shilatifard. 2018. "An Mll4/COMPASS-Lsd1 Epigenetic Axis Governs Enhancer Function and Pluripotency Transition in Embryonic Stem Cells." *Science Advances* 4 (1): eaap8747. https://doi.org/10.1126/sciadv.aap8747.

Chen, Amy F., Arthur J. Liu, Raga Krishnakumar, Jake W. Freimer, Brian DeVeale, and Robert Blelloch. 2018. "GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naïve Pluripotency." *Cell Stem Cell* 23 (2): 226-238.e4. https://doi.org/10.1016/j.stem.2018.06.005.

Dorighi, Kristel M., Tomek Swigut, Telmo Henriques, Natarajan V. Bhanu, Benjamin S. Scruggs, Nataliya Nady, Christopher D. Still, Benjamin A. Garcia, Karen Adelman, and Joanna Wysocka. 2017. "Mll3 and Mll4 Facilitate Enhancer RNA

Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation." *Molecular Cell* 66 (4): 568-576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.

Ewels, Philip A., Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso, and Sven Nahnsen. 2020. "The Nf-Core Framework for Community-Curated Bioinformatics Pipelines." *Nature Biotechnology* 38 (3): 276–78. https://doi.org/10.1038/s41587-020-0439-x.

Gökbuget, Deniz, and Robert Blelloch. 2019. "Epigenetic Control of Transcriptional Regulation in Pluripotency and Early Differentiation." *Development (Cambridge, England)* 146 (19): dev164772. https://doi.org/10.1242/dev.164772.

Hu, Deqing, Xin Gao, Marc A. Morgan, Hans-Martin Herz, Edwin R. Smith, and Ali Shilatifard. 2013. "The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers." *Molecular and Cellular Biology* 33 (23): 4745–54. https://doi.org/10.1128/MCB.01181-13.

Kaikkonen, Minna U, Nathanael Spann, Sven Heinz, Casey E. Romanoski, Karmel A. Allison, Joshua D. Stender, Hyun B. Chun, et al. 2013. "Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription." *Molecular Cell* 51 (3): 310–25. https://doi.org/10.1016/j.molcel.2013.07.010.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. "CUT&Tag

for Efficient Epigenomic Profiling of Small Samples and Single Cells." *Nature Communications* 10 (1): 1930. https://doi.org/10.1038/s41467-019-09982-5.

Krishnakumar, Raga, Amy F. Chen, Marisol G. Pantovich, Muhammad Danial, Ronald J. Parchem, Patricia A. Labosky, and Robert Blelloch. 2016. "FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and Repressing Enhancer Activity." *Cell Stem Cell* 18 (1): 104–17. https://doi.org/10.1016/j.stem.2015.10.003.

Lai, Binbin, Ji-Eun Lee, Younghoon Jang, Lifeng Wang, Weiqun Peng, and Kai Ge. 2017. "MLL3/MLL4 Are Required for CBP/P300 Binding on Enhancers and Super-Enhancer Formation in Brown Adipogenesis." *Nucleic Acids Research* 45 (11): 6388–6403. https://doi.org/10.1093/nar/gkx234.

Lavery, William J., Artem Barski, Susan Wiley, Elizabeth K. Schorry, and Andrew W. Lindsley. 2020. "KMT2C/D COMPASS Complex-Associated Diseases [KCDCOM-ADs]: An Emerging Class of Congenital Regulopathies." *Clinical Epigenetics* 12 (1): 10. https://doi.org/10.1186/s13148-019-0802-2.

Lee, Ji-Eun, Chaochen Wang, Shiliyang Xu, Young-Wook Cho, Lifeng Wang, Xuesong Feng, Anne Baldridge, et al. 2013. "H3K4 Mono- and Di-Methyltransferase MLL4 Is Required for Enhancer Activation during Cell Differentiation." *ELife* 2 (December): e01503. https://doi.org/10.7554/eLife.01503.

Local, Andrea, Hui Huang, Claudio P. Albuquerque, Namit Singh, Ah Young Lee, Wei Wang, Chaochen Wang, et al. 2018. "Identification of H3K4me1-Associated Proteins at Mammalian Enhancers." *Nature Genetics* 50 (1): 73–82. https://doi.org/10.1038/s41588-017-0015-6.

Martire, Sara, Jennifer Nguyen, Aishwarya Sundaresan, and Laura A. Banaszynski. 2020. “Differential Contribution of P300 and CBP to Regulatory Element Acetylation in MESCs.” *BMC Molecular and Cell Biology* 21 (1): 55. https://doi.org/10.1186/s12860-020-00296-9.

Mas, Glòria, Enrique Blanco, Cecilia Ballaré, Miriam Sansó, Yannick G. Spill, Deqing Hu, Yuki Aoi, et al. 2018. “Promoter Bivalency Favors an Open Chromatin Architecture in Embryonic Stem Cells.” *Nature Genetics* 50 (10): 1452–62. https://doi.org/10.1038/s41588-018-0218-5.

Meers, Michael P., Dan Tenenbaum, and Steven Henikoff. 2019. “Peak Calling by Sparse Enrichment Analysis for CUT&RUN Chromatin Profiling.” *Epigenetics & Chromatin* 12 (1): 42. https://doi.org/10.1186/s13072-019-0287-4.

Mendiratta, Gaurav, Eugene Ke, Meraj Aziz, David Liarakos, Melinda Tong, and Edward C. Stites. 2021. “Cancer Gene Mutation Frequencies for the U.S. Population.” *Nature Communications* 12 (1): 5961. https://doi.org/10.1038/s41467-021-26213-y.

Moore, Jill E., Henry E. Pratt, Michael J. Purcaro, and Zhiping Weng. 2020. “A Curated Benchmark of Enhancer-Gene Interactions for Evaluating Enhancer-Target Gene Prediction Methods.” *Genome Biology* 21 (1): 17. https://doi.org/10.1186/s13059-019-1924-8.

Moorthy, Sakthi D., Scott Davidson, Virlana M. Shchuka, Gurdeep Singh, Nakisa Malek-Gilani, Lida Langroudi, Alexandre Martchenko, Vincent So, Neil N. Macpherson, and Jennifer A. Mitchell. 2017. “Enhancers and Super-Enhancers Have an Equivalent Regulatory Role in Embryonic Stem Cells through Regulation of

Single or Multiple Genes." *Genome Research* 27 (2): 246–58. https://doi.org/10.1101/gr.210930.116.

Narita, Takeo, Yoshiki Higashijima, Sinan Kilic, Tim Liebner, Jonas Walter, and Chunaram Choudhary. 2022. "A Unique H2B Acetylation Signature Marks Active Enhancers and Predicts Their Target Genes." *BioRxiv*, January, 2022.07.18.500459. https://doi.org/10.1101/2022.07.18.500459.

Nasser, Joseph, Drew T. Bergman, Charles P. Fulco, Philine Guckelberger, Benjamin R. Doughty, Tejal A. Patwardhan, Thouis R. Jones, et al. 2021. "Genome-Wide Enhancer Maps Link Risk Variants to Disease Genes." *Nature* 593 (7858): 238–43. https://doi.org/10.1038/s41586-021-03446-x.

Neumayr, Christoph, Vanja Haberle, Leonid Serebreni, Katharina Karner, Oliver Hendy, Ann Boija, Jonathan E. Henninger, et al. 2022. "Differential Cofactor Dependencies Define Distinct Types of Human Enhancers." *Nature* 606 (7913): 406–13. https://doi.org/10.1038/s41586-022-04779-x.

Ng, Sarah B., Abigail W. Bigham, Kati J. Buckingham, Mark C. Hannibal, Margaret J. McMillin, Heidi I. Gildersleeve, Anita E. Beck, et al. 2010. "Exome Sequencing Identifies MLL2 Mutations as a Cause of Kabuki Syndrome." *Nature Genetics* 42 (9): 790–93. https://doi.org/10.1038/ng.646.

Osterwalder, Marco, Iros Barozzi, Virginie Tissières, Yoko Fukuda-Yuzawa, Brandon J. Mannion, Sarah Y. Afzal, Elizabeth A. Lee, et al. 2018. "Enhancer Redundancy Provides Phenotypic Robustness in Mammalian Development." *Nature* 554 (7691): 239–43. https://doi.org/10.1038/nature25461.

Park, Young-Kwon, Ji-Eun Lee, Zhijiang Yan, Kaitlin McKernan, Tommy O'Haren, Weidong Wang, Weiqun Peng, and Kai Ge. 2021. "Interplay of BAF and MLL4 Promotes Cell Type-Specific Enhancer Activation." *Nature Communications* 12 (1): 1630. https://doi.org/10.1038/s41467-021-21893-y.

Rickels, Ryan, Hans-Martin Herz, Christie C Sze, Kaixiang Cao, Marc A Morgan, Clayton K Collings, Maria Gause, et al. 2017. "Histone H3K4 Monomethylation Catalyzed by Trr and Mammalian COMPASS-like Proteins at Enhancers Is Dispensable for Development and Viability." *Nature Genetics* 49 (11): 1647–53. https://doi.org/10.1038/ng.3965.

Romeike, Merrit, Stephanie Spach, Marie Huber, Songjie Feng, Gintautas Vainorius, Ulrich Elling, Gjis A. Versteeg, and Christa Buecker. 2022. "Transient Upregulation of IRF1 during Exit from Naive Pluripotency Confers Viral Protection." *EMBO Reports*, July, e55375. https://doi.org/10.15252/embr.202255375.

Sankar, Aditya, Faizaan Mohammad, Arun Kumar Sundaramurthy, Hua Wang, Mads Lerdrup, Tulin Tatar, and Kristian Helin. 2022. "Histone Editing Elucidates the Functional Roles of H3K27 Methylation and Acetylation in Mammals." *Nature Genetics* 54 (6): 754–60. https://doi.org/10.1038/s41588-022-01091-2.

Shechter, David, Holger L. Dormann, C. David Allis, and Sandra B. Hake. 2007. "Extraction, Purification and Analysis of Histones." *Nature Protocols* 2 (6): 1445–57. https://doi.org/10.1038/nprot.2007.202.

Shpargel, Karl B., Cassidy L. Mangini, Guojia Xie, Kai Ge, and Terry Magnuson. 2020. "The KMT2D Kabuki Syndrome Histone Methylase Controls Neural Crest Cell

Differentiation and Facial Morphology." *Development (Cambridge, England)* 147 (21): dev187997. https://doi.org/10.1242/dev.187997.

Skene, Peter J., Jorja G. Henikoff, and Steven Henikoff. 2018. "Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers." *Nature Protocols* 13 (5): 1006–19. https://doi.org/10.1038/nprot.2018.015.

Smith, Austin. 2017. "Formative Pluripotency: The Executive Phase in a Developmental Continuum." *Development (Cambridge, England)* 144 (3): 365–73. https://doi.org/10.1242/dev.142679.

Stark, Rory, and Gord Brown. 2022. "DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data." Bioconductor version: Release (3.15). https://doi.org/10.18129/B9.bioc.DiffBind.

Sze, Christie C., and Ali Shilatifard. 2016. "MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer." *Cold Spring Harbor Perspectives in Medicine* 6 (11): a026427. https://doi.org/10.1101/cshperspect.a026427.

Thomas, Henry F., Elena Kotova, Swathi Jayaram, Axel Pilz, Merrit Romeike, Andreas Lackner, Thomas Penz, et al. 2021. "Temporal Dissection of an Enhancer Cluster Reveals Distinct Temporal and Functional Contributions of Individual Elements." *Molecular Cell* 81 (5): 969-982.e13. https://doi.org/10.1016/j.molcel.2020.12.047.

Wang, Chaochen, Ji-Eun Lee, Binbin Lai, Todd S. Macfarlan, Shiliyang Xu, Lenan Zhuang, Chengyu Liu, Weiqun Peng, and Kai Ge. 2016. "Enhancer Priming by H3K4 Methyltransferase MLL4 Controls Cell Fate Transition." *Proceedings of the*

National Academy of Sciences of the United States of America 113 (42): 11871–76. https://doi.org/10.1073/pnas.1606857113.

Wang, Shu-Ping, Zhanyun Tang, Chun-Wei Chen, Miho Shimada, Richard P. Koche, Lan-Hsin Wang, Tomoyoshi Nakadai, et al. 2017. “A UTX-MLL4-P300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription.” *Molecular Cell* 67 (2): 308-321.e6. https://doi.org/10.1016/j.molcel.2017.06.028.

Xie, Guojia, Ji-Eun Lee, Anna D. Senft, Young-Kwon Park, Shreeta Chakraborty, Joyce J. Thompson, Chengyu Liu, et al. 2022. “MLL3/MLL4 Methyltransferase Activities Control Early Embryonic Development and Embryonic Stem Cell Differentiation in a Lineage-Selective Manner.” *BioRxiv*, January, 2020.09.14.296558. https://doi.org/10.1101/2020.09.14.296558.

Yan, Jian, Shi-An A. Chen, Andrea Local, Tristin Liu, Yunjiang Qiu, Kristel M. Dorighi, Sebastian Preissl, et al. 2018. “Histone H3 Lysine 4 Monomethylation Modulates Long-Range Chromatin Interactions at Enhancers.” *Cell Research* 28 (2): 204–20. https://doi.org/10.1038/cr.2018.1.

Yang, Pengyi, Sean J. Humphrey, Senthilkumar Cinghu, Rajneesh Pathania, Andrew J. Oldfield, Dhirendra Kumar, Dinuka Perera, et al. 2019. “Multi-Omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency.” *Cell Systems* 8 (5): 427-445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

Zhang, Hui, Srimonta Gayen, Jie Xiong, Bo Zhou, Avinash K Shanmugam, Yuqing Sun, Hacer Karatas, et al. 2016. “MLL1 Inhibition Reprograms Epiblast Stem Cells to

Naïve Pluripotency.” *Cell Stem Cell* 18 (4): 481–94. https://doi.org/10.1016/j.stem.2016.02.004.

Zhang, Tiantian, Zhuqiang Zhang, Qiang Dong, Jun Xiong, and Bing Zhu. 2020. “Histone H3K27 Acetylation Is Dispensable for Enhancer Activity in Mouse Embryonic Stem Cells.” *Genome Biology* 21 (1): 45. https://doi.org/10.1186/s13059-020-01957-w.

Zhou, Harry Y., Yulia Katsman, Navroop K. Dhaliwal, Scott Davidson, Neil N. Macpherson, Moorthy Sakthidevi, Felicia Collura, and Jennifer A. Mitchell. 2014. “A Sox2 Distal Enhancer Cluster Regulates Embryonic Stem Cell Differentiation Potential.” *Genes & Development* 28 (24): 2699–2711. https://doi.org/10.1101/gad.248526.114.

Chapter 3 : The role of MLL3/4 in GRHL2 mediated enhancer activation

Summary

This study is a follow up to our previous study in Chapter 2 which described the global impacts of loss of MLL3/4 during the naive to formative ESC transition. Here, we examine the role of MLL3/4 at enhancers more closely using the transcription factor (TF) Grhl2, a known interactor with MLL3/4 and driver of formative gene expression. Using genetic deletions we find MLL3/4 is required for GRHL2 mediated H3K4me1 and H3K27ac deposition at GRHL2 enhancers in the formative state but does not appear to impact nearby transcription. We then developed and applied a synthetic system to induce GRHL2 activity in the absence of MLL3/4. Our preliminary data suggests MLL3/4 is only partially required GRHL2 mediated induction of transcription but is required for enhancer H3K4me1 and H3K27ac. Further experiments in this study will serve to distinguish between models of TF mediated enhancer activation where MLL3/4 is essential or plays a transcriptional amplifying function at enhancers.

The following results are preliminary data and analyses from an unpublished study

Introduction

The acquisition of cell fate is dependent on gene regulatory networks that are regulated spatiotemporally by cell type specific transcription factors (TFs) binding to sequence specific motifs on DNA. One major class of TF binding elements that regulate transcription are cis-regulatory elements known as enhancers which operate distal to a

target promoter. An active, transcriptionally stimulating enhancer is established when a TF binds and recruits general chromatin regulating factors. Chromatin regulators are key factors that mediate enhancer activity and their dysregulation has severe developmental and disease-related consequences.

Several key chromatin regulators are associated with enhancers including the histone methyltransferases MLL3/4 and histone acetyltransferases CBP/P300 which deposit H3K4me1/2 and H3K27ac respectively(Sze and Shilatifard 2016; Hm and Nb 2001; Bannister and Kouzarides 1996; Ogryzko et al. 1996). These histone modifications are considered to be active marks of enhancers and their deposition is predictive of enhancer activation(Creyghton et al. 2010; Heintzman et al. 2007; Visel et al. 2009). A molecular epistasis for the early stages of enhancer activation has been proposed. Enhancer-bound TFs recruit MLL3/4; MLL3/4 deposit H3K4me1/2 and recruit CBP/P300; CBP/P300 deposit H3K27ac, recruit additional factors, and promote transcription(Lee et al. 2013; Lai et al. 2017). Additional chromatin regulators have been demonstrated to be critical for enhancer activity but the relationships among all the factors within this epistasis is unresolved.

To investigate the epistasis of enhancer activation further we utilized an embryonic stem cell (ESC) differentiation known as the naive to formative transition. The naive to formative transition highly recapitulates gene regulatory events that occur during mouse early embryogenesis between E4.5 and E5.5 including enhancer activation(Smith 2017; Krishnakumar et al. 2016; Yang et al. 2014; Buecker et al. 2014). Previously, we found that, while MLL3/4 are required for all H3K4me1 and most H3K27ac changes between naive and formative ESC states, MLL3/4 were largely dispensable for transcriptional

activation of the formative program(Boileau, Chen, and Blelloch 2023). Defects in active mark deposition were observed at enhancer sites for Grhl2, a TF that is transcriptionally upregulated during the transition, is required for formative enhancer activation, and a direct interactor of MLL3 and MLL4. (Chen et al. 2018; MacFawn et al. 2019). Using GRHL2 as a model TF in the naive to formative transition we sought to examine why loss of enhancer active marks were not closely associated with transcriptional defects.

Between the naive and formative ESCs we find MLL3/4 is required for active chromatin state acquisition at GRHL2 binding sites but is not required for GRHL2 expression or binding. However, interestingly, we were unable to identify transcriptional defects nearby GRHL2 sites in the absence of MLL3/4. To determine whether GRHL2 may activate transcription independent of MLL3/4 we developed a system for rapid and conditional GRHL2 activity and applied this system to ESCs lacking MLL3/4. Surprisingly, acute activation of GRHL2 activity led to gene activation in the absence of MLL3/4 albeit reduced relative to controls. Using high throughput epigenomics assays for histone modifications we also found that acute activation of GRHL2 in MLL3/4 knockouts led to GRHL2 binding but detected reduced gains in H3K4me1 and H3K27ac. Our current interpretations position MLL3 and MLL4 as factors that amplify but are not required for GRHL2 TF activity and which contrasts with the proposed essential role of MLL3/4 in enhancer activation.

Results

MLL3/4 is required for GRHL2 mediated deposition of H3K4me1 and H3K27ac

To study the relationship between GRHL2 and MLL3/4 we used previously published cell lines whereby Cre-ERT2 mediated recombination of MLL3 knockout ESC lines with floxed alleles for MLL4 ($MLL3^{-/-}$; $MLL4^{fl/fl}$, henceforth MLL3KO) created a double knockout of MLL3/4 ($MLL3^{-/-}$;$MLL4^{-/-}$, henceforth DKO). Thereafter, we maintained DKO and wildtype (WT) control lines in naive pluripotency conditions (LIF+2i+Serum) and differentiated them into the formative state by removing LIF+2i for 63 hours (Figure 3.1A). We first performed westerns for GRHL2 on naive and formative DKO cells and found GRHL2 protein levels were increased in DKO cells in the formative state (Figure 3.1B). Next, we profiled binding of GRHL2 by performing CUT&RUN(Skene, Henikoff, and Henikoff 2018) on WT and DKO cells in the naive and formative state. Since GRHL2 is not expressed in the naive state we used naive samples as a background control. We called GRHL2 peaks using SEACR(Meers, Tenenbaum, and Henikoff 2019) and used Diffbind analysis(Stark and Brown 2022) to contrast formative vs. naive samples for GRHL2 signal in WT cells. Over 60000 peaks were called for GRHL2 CUT&RUN for each naive and formative state. However, we identified only 493 peaks that significantly gained GRHL2 signal in the formative state ($p.adj < 0.05$, $Log2FC > 1$). The GRHL motif was detected by HOMER at 403 out of 493 CUT&RUN peaks, but only 176 of 403 sites overlapped with 332 formative GRHL2 ChIP-seq peaks published which were identified using GRHL2 knockout controls (Data not shown)(Chen et al. 2018). We reasoned

knockout validation is a better control for this antibody and, going forward, conducted our analyses using the 332 sites from the GRHL2 ChIP-seq peak set (GRHL2 sites).

To evaluate GRHL2 binding in DKO cells we generated heatmaps for naive/formative WT and DKO cells at GRHL2 sites. We included naive and formative data for GRHL2 CUT&RUN, H3K4me1 CUT&RUN, and H3K27ac CUT&TAG(Kaya-Okur et al. 2020; 2019). We also plotted published formative MLL4 CUT&RUN (Figure 3.1C). We found a gain in GRHL2 signal in both WT and DKO cells upon formative transition consistent with GRHL2 binding. However, while these sites gained H3K4me1 signal in WT formative cells we did not observe a gain in H3K4me1 in DKO formative cells. Similarly, H3K27ac was gained in WT cells, but the average gain in H3K27ac signal was dramatically reduced in DKO cells. Notably, these GRHL2 sites are also normally enriched for MLL4 signal in the formative state. Together, this data suggests that GRHL2 is able to bind enhancers in the absence of MLL3/4, but MLL3/4 is largely required to deposit active marks H3K4me1 and H3K27ac.

We expected that the loss of MLL3/4 and aberrant H3K4me1/H3K27ac deposition would impact nearby transcriptional targets. In our reanalysis of microarray expression data generated for formative WT and GRHL2 formative cells we find 400 genes that are significantly downregulated in GRHL2KO vs. WT cells (adj.p < 0.05, Log FC < -0.25, Data not shown). When comparing RNA-seq data between formative WT and DKO cells for these 400 genes we did not observe a decrease in transcript levels for DKO cells (Figure 3.1D). We also examined transcript levels of the 213 nearby genes associated with GRHL2 binding sites. For these genes, microarray data showed a strong and significant decrease in nearby expression in GRHL2KO cells compared to WT (Figure 3.1E).

However, formative DKO cells did not have reduced transcript levels compared to WT controls (Figure 3.1F). Finally, we calculated fold changes in levels of formative transcripts in both DKO cells and the parental MLL3KO line relative to WT (Figure 3.1G). Here again, we did not observe a significant reduction in nearby transcript levels when comparing DKO to parental MLL3KO controls. Together these data suggest that MLL3/4 is required to facilitate H3K4me1/H3K27ac by GRHL2 at enhancers in the formative state, but that nearby transcription in the formative state is not widely perturbed.

Conditional induction of GRHL2 activity by tamoxifen rapidly binds enhancers and activates target genes

GRHL2 target gene expression is not impacted by loss of MLL3/4 in formative cells. GRHL2 enhancer activity helps maintain the expression of a subset of KLF2/4 enhancer regulated genes as KLFs are downregulated during the naive to formative transition. In DKO cells, the naive network including KLF4 is not able to be fully repressed during differentiation(Chen et al. 2018). Is GRHL2 sufficient to drive transcription in the absence of MLL3/4 and enhancer remodeling or are their mechanisms compensating for loss of GRHL2 activity in DKO cells? To address this question, we created a synthetic system to turn on GRHL2 activity without differentiation.

Premature expression of GRHL2 in the naive state is sufficient for activation of enhancers including a gain in levels of H3K4me1, H3K27ac, and Cohesin recruitment(Chen et al. 2018). We hypothesized that fusing ERT to GRHL2 could be an improved system to study enhancer activation by providing rapid and conditional activity

through tamoxifen treatment, working independently of transcriptional perturbations, and overcoming the leakiness associated our previous TET-inducible GRHL2 system (data not shown). For proof of principle experiments, we generated a plasmid that constitutively expressed GRHL2 fused to an ERT domain and an HA epitope tag (Fig. 3.1A). We used PiggyBac transposition to integrate our plasmid into V6.5 wildtype (WT) ES cells and selected a clone for further characterization.

To validate nuclear translocation of GRHL2-ERT we treated cells with tamoxifen (Tam) for 0, 1, or 8hrs and performed subcellular fractionation for western blots (Fig. 3.1B). Within 1hr we found GRHL2-ERT levels decrease in the cytoplasm with a concomitant increase in the nucleus and chromatin accumulation. GRHL2-ERT was detectable in the nucleus in the absence of Tam, though appreciable levels of GRHL2-ERT in the chromatin fraction was only detected with tamoxifen treatment. As an initial test of tamoxifen dependent activity of the GRHL2-ERT system we performed qPCR for Grhl2 and Cldn6, a transcriptional target of GRHL2. The GRHL2-ERT cell line expressed high levels of GRHL2 compared to the WT parent line but only the GRHL2-ERT line induced Cldn6 upon Tam treatment after 8hrs. We did not detect meaningful differences between baseline levels of Cldn6 suggesting, at least for Cldn6, induction of GRHL2 activity is Tam dependent. To determine if GRHL2-ERT binds enhancers in a Tam dependent fashion, we performed ChIP-qPCR for two validated enhancers of GRHL2 (Cdh1, Dsp) and two predicted enhancer candidates for GRHL2 (Cldn6, Wnt7b). With Tam treatment after 8hrs we found significant enrichment of GRHL2-ERT at all four distal regions. We found no significant enrichment at these four sites without Tam treatment, nor did we identify enrichment in our negative ethanol (Eth) treated control. This suggests,

despite detectable levels of GRHL2-ERT in the nucleus, enhancer binding is highly Tam dependent.

Although we observed induction of Cldn6 at 8hrs with Tam, we suspected transcriptional activation by GRHL2-ERT could be more rapid. To resolve the transcriptional activation kinetics of the system we performed a qPCR time course for GRHL2 targets Cldn6 and Wnt7b. Relative to Cdh1 and Dsp, these genes have a low baseline expression level and a higher detectable fold change upon induction (Data not shown). We included measurements using exon:exon primers and intron:exon primers which reflect mRNA and unspliced (nascent) RNA levels respectively. We found either exon:exon and or intron:exon levels of Cldn6 and Wnt7b to be noticeably upregulated by 1-2 hours after tamoxifen. As well, for Cldn6, we observed upregulation of intron:exon levels preceding exon:exon upregulation. We noted an increase in levels of Wnt7b and Cldn6 continued across 24 hours of Tam treatment though the rate in upregulation appeared to begin plateauing around 24 hours. Together, these data suggest that, at least for select genes, Tam treatment allows GRHL2-ERT to rapidly bind and transcriptionally activate target genes.

MLL3/4 are not required for transcriptional activation by GRHL2

To further test the GRHL2-ERT system as well as the role of MLL3/4 in GRHL2 enhancer activation we generated MLL3KO and DKO clones that expressed GRHL2-ERT. Using westerns, we identified two clones with similar levels of GRHL2 that we used for our experiments (Figure 3.3A). In order to determine whether MLL3/4 was required for

GRHL2 target gene activation we performed RNA-sequencing. In particular, we treated MLL3KO-ERT and DKO-ERT cells with either Tam or Eth for 8 hours and performed RNA-sequencing. As a control for Tam treatment, we also included samples for MLL3KO parental line without the ERT system but treated with Tam. In our preliminary analysis, we first performed differential gene expression analysis using DESeq2 and compared both MLL3KO-ERT Tam vs. Eth and DKO-ERT Tam vs. Eth (Data not shown). We identified 167 genes differentially expressed in MLL3KO-ERT and 27 in DKO-ERT (p.adj < 0.05, Log2 FC > 1). Combined, we identified 180 significant induced genes suggesting many differential genes are unique to MLL3KO Tam treated samples (167 out of 180 total significant genes).

Next, we analyzed the fold changes for these unique genes. We started by comparing MLL3KO-ERT Tam with our samples for MLL3KO Tam (without the ERT system) and set them both relative to MLL3KO-ERT Eth (Figure 3.3B). We observed some genes that were upregulated by Tam treatment itself, but notable induction of more genes in the MLL3KO-ERT line. This suggests that our GRHL2-ERT system is activating transcription in MLL3KO-ERT lines. Surprisingly, when comparing fold changes Tam/Eth for MLL3KO-ERT and DKO-ERT for all significantly induced genes (e.g., 180) we found gene induction was highly correlated (Pearson = 0.7). We also calculated a higher slope (m=0.37) than Tam treatment controls (Figure 3.3C, m=0.17). This suggests gene induction is more likely due to GRHL2 activity rather than Tam treatment. However, while DKO-ERT cells were able to induce transcription similarly to MLL3KO-ERT, it was relatively reduced. This data suggests, at least for some genes, GRHL2 is able to induce transcription in the absence of MLL3/4 albeit to a reduced degree.

MLL3/4 is required for GRHL2 mediated H3K4me1 and H3K27ac deposition

The activation of transcription by GRHL2 without MLL3/4 could be driven by other factors that remodel enhancers. To test the role of MLL3/4 in GRHL2 driven activation we induced GRHL2-ERT with Tam or Eth controls for 8 hours in MLL3KO and DKO-ERT lines and then conducted CUT&TAG for HA, H3K4me1, H3K27ac, and IgG controls. Using CPM normalization, we observed an increase in IgG signal in both MLL3KO and DKO-ERT lines when GRHL2 was induced (Data not shown). This suggests there is some level of increased background, likely due to increased accessibility from GRHL2 binding. However, using spike-in normalization this background was dramatically reduced (Figure 3.4A). Comparing spike-in normalized HA we find that GRHL-ERT2 binding is highly dependent on Tam, with binding in both MLL3KO and DKO-ERT cells. We did detect slightly lower average signal in DKO cells compared to MLL3KO-ERT cells. For active histone marks H3K4me1 and H3K27ac we find dramatic increases in their levels in MLL3KO-ERT cells upon Tam treatment. Consistent with the role of MLL3/4 mediating their deposition, DKO-ERT cells were unable to acquire meaningful signal for both marks. Some increased signal was detected in Tam treated DKO-ERT cells though it did not appear to be much different than average levels of IgG signal. These data are consistent with the idea that GRHL2 recruits MLL3/4 in order to deposit H3K4me1 and mediate H3K27ac deposition. Together our data, suggests MLL3/4 is required for coordinating active mark deposition but may only be partially required for transcriptional upregulation by GRHL2.

Discussion

We have previously found that key chromatin regulators MLL3 and MLL4 were primarily responsible for active chromatin changes at distal sites between the naive and formative ESC states(Boileau, Chen, and Blelloch 2023). Unexpectedly, loss of MLL3/4 resulted in underwhelming effects on transcriptional upregulation of the formative transcriptional program. Are MLL3/4 dispensable for transcriptional regulation or are other mechanisms were compensating for loss of MLL3/4? In this follow up study we use the formative TF Grhl2 as a model to determine whether MLL3/4 is required for enhancer-based TF regulation of gene expression. Our preliminary analyses suggest an unexpected discovery. Consistent with our previous data, enhancers bound by GRHL2 require MLL3/4 for active mark deposition including H3K4me1 and H3K27ac. However, while reduced, we were able to detect induction of gene expression in the absence of MLL3/4 as well as increased active mark deposition. As a result, our working hypothesis is that GRHL2 has the capacity to induce transcription and that the role of MLL3/4 is to amplify GRHL2 activity during the naïve to formative transition.

Further work is necessary to address current limitations of this study. Namely, the effect sizes of GRHL2 target gene induction for both MLL3KO and DKO lines are relatively small and with mediocre levels of upregulation. This may be consistent with the fact that few enhancers and genes are regulated by GRHL2 normally. This data was collected from an RNA-seq experiment with 8 hours of GRHL2 induction, but our qPCR time course suggests that Cldn6/Wnt7b induction does not peak for at least 24 hours. A longer induction may yield a greater effect size at the expense of identifying primary targets. A

small effect size may also impact our future analysis associating ERT system gene expression with GRHL2 enhancers and epigenomics data. In our epigenomics experiments using the ERT system we also find minor increases in H3K4me1 and H3K27ac signal without MLL3/4. While we cannot exclude the possibility that GRHL2 can recruit other factors that catalyze these modifications it seems more likely this is a technical artifact from Tn5 transposition/CUT&TAG at sites with gains in chromatin accessibility. Generating additional CUT&TAG datasets, such as P300 or H3K4me2 could both complement and reinforce our findings. Additionally, performing an induction time course analysis of enhancer modification versus IgG background may help resolve whether we are observing small but real increases in signal for active histone modifications.

If this conclusion persists, that GRHL2 can induce transcription independent of MLL3/4, it will have important implications for our understanding of the role of general chromatin regulators at enhancers. Namely, the role of chromatin regulators such as MLL3/4 may be to perform an amplification of enhancer-bound TF activity. Using H3K27ac as a surrogate for CBP/P300 activity, our data also suggests CBP/P300 recruitment/activity may be partially required for enhancer-bound TF activity. Cooperativity of chromatin regulators in regulating enhancer activity has been demonstrated, such as between MLL4 and P300 or MLL4 and BRG1(Wang et al. 2017; Park et al. 2021). Our results could be consistent with an additional synergy between TF activity and chromatin regulators. When a TF binds an enhancer, chromatin regulators and their remodeling of enhancers may function to stabilize transcription factor binding or enhancer-promoter interactions which could further stimulate target gene transcription.

Either way, these findings may lead to a revision of our conceptual models of enhancer activation from a stepwise recruitment model with essential factors to a model based on equilibriums between collections of TFs and chromatin regulators. The diversity of enhancer activation mechanisms is also important to investigate, and the extent to which this phenomenon may exist for TFs other than GRHL2 would be an intriguing avenue of study.

Further experiments for this study, and other studies, on the relationship between enhancers, TFs, and chromatin regulators will be important steps towards understanding mechanisms of enhancer regulation of transcription. Establishing the rules and exceptions to principles in gene regulation will surely inform future efforts to understand development and the manifestation of disease.

Materials and Methods

ESC culture and line generation

Mouse ESCs were cultured in Knockout DMEM (Thermo Fisher, CAT#10829018) supplemented with 15% FBS, L-Glutamine, Penicillin/Streptamycin, NEAA, LIF (1000U/mL), and 2i (1uM MEK inhibitor PD0325901 and 3uM GSK3 inhibitor CHIR99021). WT cells used for GRHL2-ERT proof of principle tests are V6.5 ESCs. Clones used to conduct RNA-seq are derived from WT R1 ESCs. MLL3KO ESCs ($MLL3^{-/-}$; $MLL4^{fl/fl}$) and MLL3/4 DKO ($MLL3^{-/-}$; $MLL4^{-/-}$) are derived from a mixed background of C57BL/6J and 129 strains. Formative cells were generated by removal of LIF and 2i. Briefly, 5e4 ESCs were plated per well of a 6 well plate on day −1 in LIF+2i media. To initiate differentiation, LIF and 2i were removed 24 hours after seeding (Day 0). Formative

cells were collected on day 3 of differentiation, 63 hours after removal of LIF and 2i. To overcome proliferation defects 7.5e4 MLL3/4 DKO cells were plated per well of a 6 well. Naive cells were passaged and staged appropriately for simultaneous harvest. Lines consistently tested negative for mycoplasm.

Protein extraction and Westerns

To harvest protein for western assays, cells were trypsinized, washed once with ice cold PBS before adding RIPA buffer with protease inhibitors (Sigma CAT#P8340). After 15 minutes on ice, lysed cells were centrifuged at 16000g for 10 minutes and the supernatants representing whole cell fractions were collected and snap-frozen in liquid nitrogen. Protein quantification was conducted using a Micro BCA protein assay kit (Thermo CAT#23235). 40ug of protein was loaded per well of SDS PAGE gels for western. For subcellular fraction we used the Pierce Subcellular Fractionation Kit (Thermo Fisher, CAT#78840) according to the manufacturer's protocol. For SDS PAGE of cellular fractions we loaded 20ug per well for each cytoplasmic and nuclear fraction and 10ug for the chromatin fraction.

Westerns were typically conducted using a Bio-Rad system with Tris-Glycine gels purchased from Bio-Rad and transferred to methanol activated PVDF membranes. Membranes were blocked, and stained with primaries and secondaries using Li-Cor Odyssey Blocking Buffer (Li-Cor, CAT#927-60001) mixed 1:1 with TBS. Primary antibodies for western were Rabbit anti-H3 (Cell Signaling, CAT#4493), Mouse anti-alpha Tubulin (Sigma Aldrich, CAT#T-6074), Mouse anti-TBP (Thermo Fisher, CAT#MA1-

21516), Rabbit anti-HA (Abcam, CAT#ab9110), and Rabbit anti-GRHL2 1:100 (Sigma, CAT#HPA004820)

qPCR and analysis

To perform qPCR we first extracted RNA by adding Trizol directly to plates. After adding chloroform, an isopropanol precipitation with GlycoBlue was performed followed by ethanol washes. RNA pellets were resuspended in RNAse-free water and quantified using a NanoDrop. 200ug of RNA was used for cDNA synthesis using Maxima First Strand Synthesis Kit (Thermo Fisher, CAT#K1672) with half reactions according to the manufacturer's protocol. qPCR on cDNA was performed using SYBRgreen master mix (Applied Biosystems, CAT#A25742) using 6ul final volume on a QuantStudio5 qPCR machine (Applied Biosystems). qPCR primers for targets are listed in Additional File 3. Target Ct values were normalized to GAPDH internally for each sample and then set relative to the naive WT negative control.

RNA sequencing

Total RNA was extracted and purified from cells using Trizol followed by ethanol precipitation. RNA-seq libraries were generated using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, CAT#A01172) according to their protocol using 200ng of total RNA for input. We utilized the PCR Add-on kit for Illumina (Lexogen, CAT#M02096) to determine an appropriate number of PCR cycles to amplify libraries. Amplified libraries were quantified using Agilent Tapestation 4200. Libraries were pooled

and sequenced using a HiSeq 4000 to obtain single end 50bp reads. At least 10 million mapped reads or more per sample were obtained.

RNA sequencing processing and analysis

Fastq files for RNA-seq samples were processed using Nextflow(Ewels et al. 2020) and the nf-core RNA-seq pipeline v3.9 with default settings. The gene count output from the pipeline was filtered for genes greater than 1 cpm in at least two total samples. Samples were normalized using TMM. Next, Log2 CPM averages were calculated for replicates of each sample for scatterplot visualization and nearest neighbor TSS analysis. All transcriptomics analyses were conducted using CPM values from TMM normalization of all samples except for nearest neighbor analysis where WT and DKO were TMM normalized together. To conduct differential gene expression we performed DESeq2 v1.34.0 analysis using the raw gene count matrix as input for each desired comparison. Gene ontology was performed using ClusterProfiler 4.2.2. Custom R code for other downstream transcriptomics analyses and visualization provided on Github.

CUT&RUN sequencing and processing

CUT&RUN was conducted using the protocol from Skene et al. 2018(Skene, Henikoff, and Henikoff 2018) with the following modifications: Freshly trypsinized cells were bound to activated Concanavalin A beads (Bang Laboratories, #BP531) at a ratio of 2e5 cells/10ul beads in CR Wash buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine with protease inhibitors added) at room temp. An input of 2e5 cells were used per target. Bead-bound cells were then incubated rotating overnight at 4C in CR Antibody buffer (CR

Wash with 0.05% Digitonin, 2mM EDTA) containing primary antibody. We used the following antibodies for CUT&RUN: Rabbit anti-GRHL2 1:100 (Sigma, CAT#HPA004820) and 1:100 Rabbit IgG isotype control (Abcam, ab171870). After primary, we washed 3 times 5 minutes each with cold CR Dig-wash buffer (CR Wash with 0.05% Digitonin) and incubated with pA-MNase (1:100 of 143ug/mL provided by Steve Henikoff) for 1 hour rotating at 4C. After MNase binding, we washed 3 times 5 each with cold CR Dig-wash buffer, and chilled cells down to 0C using a metal tube rack partially submerged in an ice water slurry. MNase digestion was induced by adding CaCl2 at a final concentration of 2mM. After 30 minutes of digestion, the reaction was quenched using Stop Buffer containing 340mM NaCl, 20mM EDTA, 4mM EGTA, 0.05% Digitonin, 100ug/mL RNAse A, 50ug/mL Glycogen, and approximately 2pg/mL Yeast spike-in DNA (provided by Steve Henikoff). The digested fragments for each sample were then extracted using a phenol chloroform extraction. Library preparation on samples was conducted using manufacturer's protocols for NEBNext Ultra II Dna Library Prep Kit (New England BioLabs, CAT#E7645) and NEB Multiplex Dual Index oligos (New England BioLabs, CAT#E7600, #E7780) with the following modifications. We input approximately 10ng of sample for half reactions, we diluted the NEBNext Illumina adaptor 1:25, we used the following PCR cycling conditions: 1 cycle of Initial Denaturation at 98C for 30 seconds, 12+ cycles of Denaturation at 98C for 10 seconds then Annealing/Extension at 65C for 10 seconds, and 1 final cycle of extension at 65C for 5 minutes. Following library preparation, double size selection was performed using Ampure beads and quality and concentration of libraries were determined by an Agilent 4200 Tapestation with High-Sensitivity D1000 reagents before pooling for sequencing.

Fastq files for CUT&RUN samples were processed using Nextflow and the nf-core CUT&RUN pipeline v3.1. In brief, adapters were trimmed using Trim Galore. Paired-end alignment was performed using Bowtie2 and peaks were called using SEACR with a peak threshold of 0.05 using spike in calibration performed using the E.coli genome K12.

CUT&TAG sequencing and processing

CUT&TAG was conducted using the protocol from Kaya-Okur et al. 2020 with the following modifications: freshly trypsinized cells were bound to Concanavalin A beads at a ratio of 2e5 cells/7ul beads in CR wash at room temp. We used 2e5 cells as input per sample. Bead-bound cells were then incubated rotating overnight at 4C in CT Antibody buffer (CR Wash with 0.05% Digitonin, 2mM EDTA, 1mg/mL BSA) containing primary antibody. We used the following primary antibodies for CUT&TAG: 1:100 Rabbit anti-H3K4me1 (Abcam, ab8895), 1:100 Rabbit anti-H3K27ac (Abcam, ab4729), 1:100 Rabbit IgG isotype control (Abcam, ab171870). After primary, samples were washed 3 times for 5 minutes each using CR Dig-wash buffer and resuspended in 1:100 secondary antibody (Guinea pig anti-rabbit, Antibodies Online #ABIN101961) in CR Dig-wash buffer at 4C for 1 hour rotating at 4C. Samples were then incubated for 1 hour at 4C with 50ul of approximately 25nM homemade pA-Tn5 in CT Dig300 wash buffer (20mM HEPES, 300mM NaCl, 0.01% Digitonin, 0.5mM Spermidine with Roche cOmplete protease inhibitors added). Recombinant Tn5 was purified and loaded with adapters as previously described (Kaya-Okur et al. 2019). After Tn5 incubation, samples were washed 3 times for 5 minutes each with CT Dig300 wash buffer. Tagmentation was then initiated for 1hr

at 37C in a thermocycler by adding MgCl2 to 10mM final concentration in 50uL volume. The tagmentation reaction was quenched immediately afterwards by adding 1.6ul of 0.5M EDTA, 1ul of 10mg/mL Proteinase K, and 1ul of 5% SDS. Samples were then incubated at 55C for 2 hours in a thermocycler to denature Tn5 and solubilize tagmented chromatin. After incubation, samples were magnetized and the supernatant was transferred to new wells where SPRI bead purification was performed using homemade beads to select all DNA fragment lengths larger than 100bp. Samples were eluted in 0.1X TE and approximately half of each sample was used for library preparation using NEBNext HIFI Polymerase with custom indices synthesized by IDT. An appropriate number of cycles for each target was chosen to prevent overamplification bias. After amplification, libraries were purified with 1.2x homemade SPRI beads to select for fragments >250bp and eluted in 0.1X TE. Quality and concentration of libraries were determined by an Agilent 4200 Tapestation with D1000 reagents before pooling for sequencing.

CUT&TAG samples were processed similarly to CUT&RUN samples using the same Nextflow pipeline.

Figures

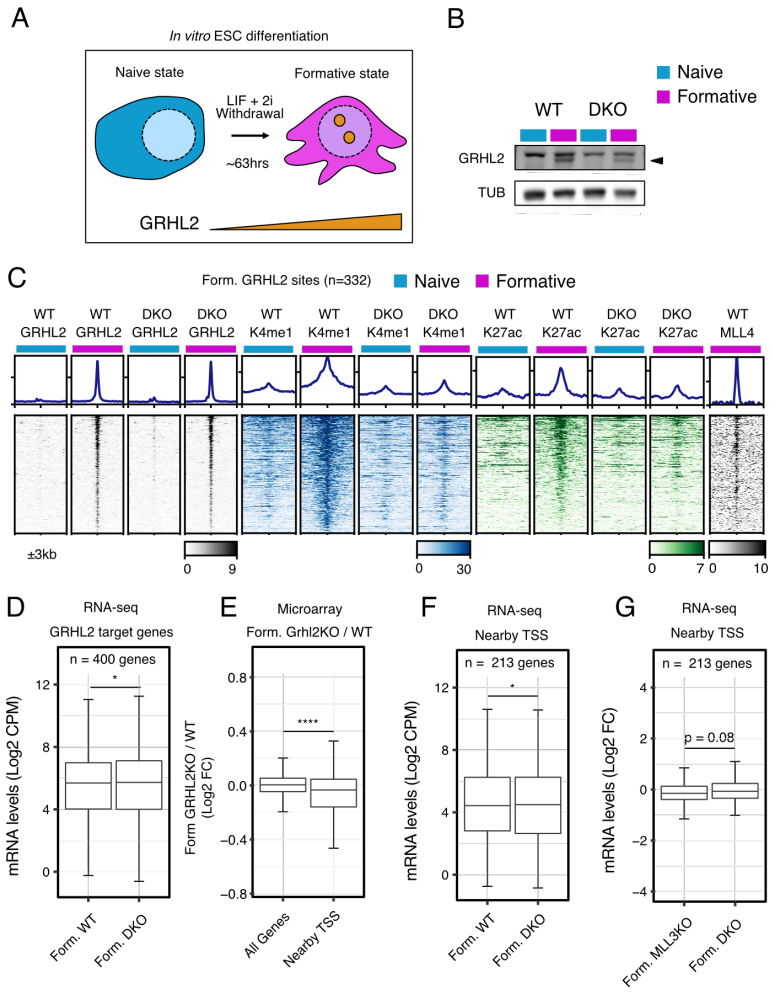

Figure 3.1 - MLL3/4 is required for H3K4me1 and H3K27ac deposition at GRHL2 sites in the formative state.

A) GRHL2 becomes expressed as ESCs transition from the naive to formative state. B) Western blots show MLL3/4 is not required for GRHL2 expression in the formative state. C) Heatmap visualization of CUT&RUN and CUT&TAG data for WT/DKO ESCs in naive and formative states. Includes GRHL2, H3K4me1, H3K27ac, and MLL4 signals at 332 GRHL2 binding sites (n = 332). All heatmap values and range are in CPM. For metagene analysis above heatmaps the range in CPM is the same as shown in heatmap for each factor. D) RNA-seq data (Log2 CPM) for 400 GRHL2 target genes, originally detected for GRHL2KO cells, in formative WT and DKO cells. E) Log2 Fold Change values for microarray signal of GRHL2KO vs WT formative cells of 213 genes nearby GRHL2 sites compared to all other detected genes. F) RNA-seq data (Log2 CPM) of transcript levels of 213 genes nearby GRHL2 sites. G) RNA-seq data (Log2 FC) of formative transcript levels for MLL3KO and DKO cells relative WT controls. Paired Wilcoxon Rank Sum Test used for 3.1D,F,G. Mann-Whitney U test used for 3.1E *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

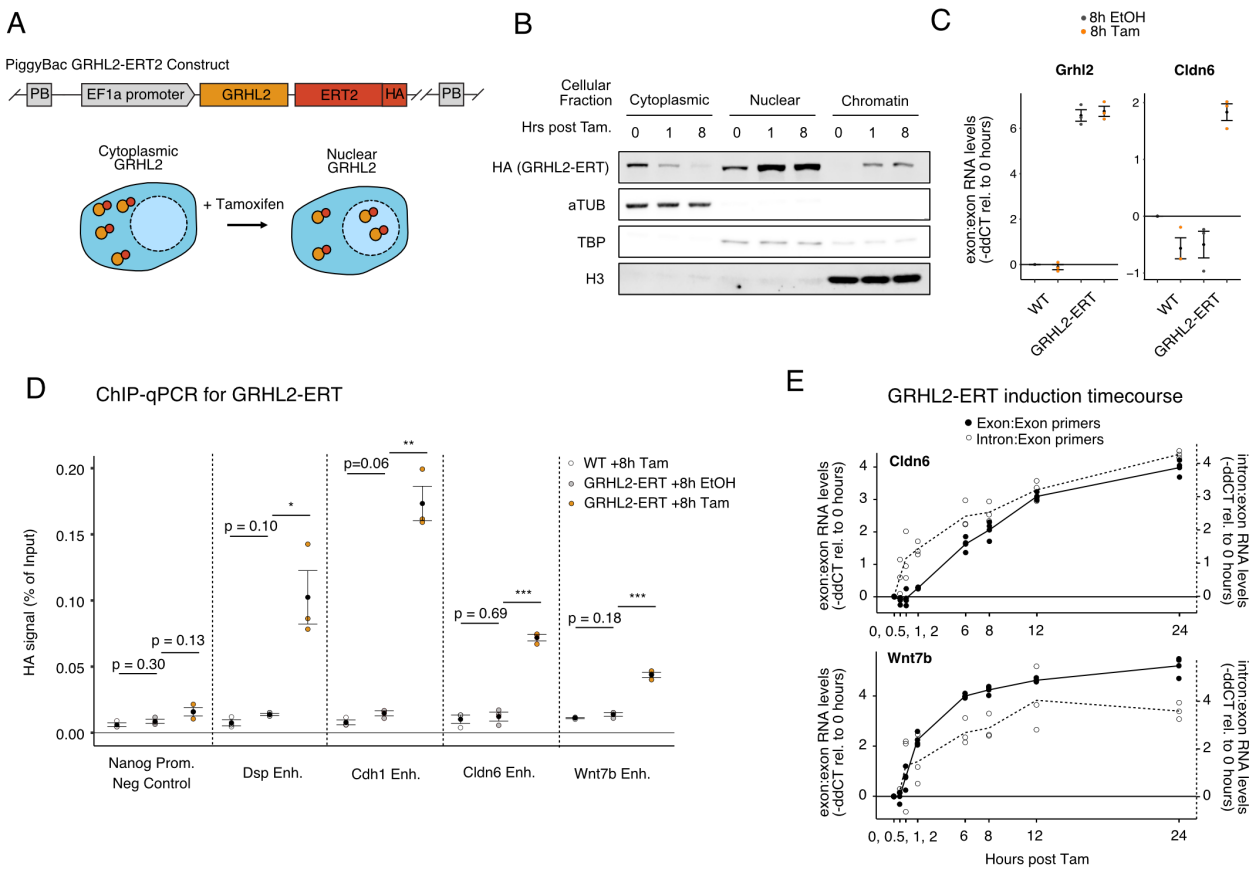

Figure 3.2 - A synthetic system for induction of GRHL2 achieves rapid enhancer binding and gene activation

A) Schematic of PiggyBac plasmid containing GRHL2-ERT transgene and mechanism of induction by tamoxifen (Tam). B) Western blots using subcellular fractions generated from a WT cell line expressing GRHL2-ERT and treated with Tam for 0, 1 or 8hrs. C) qPCR data for Grhl2 and Grhl2 target Cldn6 in WT cells with or without GRHL2-ERT system. Values are -ddCT relative to GAPDH and WT, 8hr EtOH baseline. Mean and SEM shown in black (n=3). D) ChIP-qPCR for GRHL2-ERT at select enhancers after treatment with Tam. Values are normalized to each other by percentage of input. Mean and SEM shown in black (n=3). Paired Wilcoxon Rank Sum Test. E) qPCR measurements over a time course of Tam treatment. For each Cldn6 and Wnt7b measurements were made using exon:exon primers (left Y-axis) and intron:exon primers (right Y-axis). Values are -ddCT relative to 0hr induction control. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

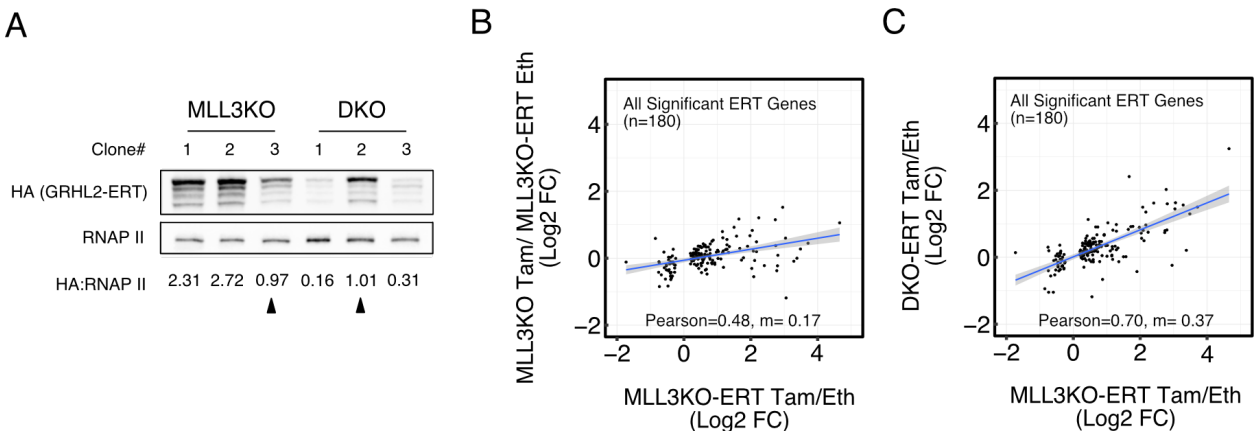

Figure 3.3 - MLL3/4 is not required for transcriptional activation upon acute GRHL2 induction

A) Western blots showing the MLL3KO and DKO clones used for downstream experiments (triangles). B) Log2 fold changes comparing induction of GRHL2 in MLL3KO-ERT lines and MLL3KO WT lines treated with tamoxifen to MLL3KO Eth controls. Only significantly different genes are shown (n=180). Linear model and confidence interval shown by blue line and gray respectively. C) Same as 3.3B but comparing MLL3KO-ERT induction to DKO-ERT induction.

A

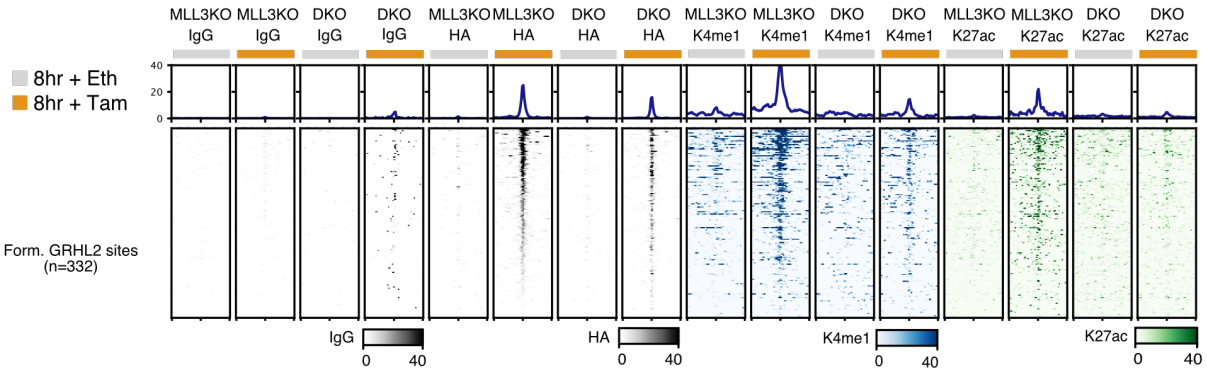

Figure 3.4 - MLL3/4 is required for GRHL2 mediated H3K4me1 and H3K27ac deposition.

A) Heatmap visualization of CUT&TAG data before and after GRHL2 induction in MLL3KO/DKO-ERT cells at 332 GRHL2 sites. All heatmap values and range for IgG, HA, H3K4me1, H3K27ac is spike-in normalized signal. For metagene analysis above heatmaps the range in CPM is the same as shown in heatmap for each factor.

References

Bannister, A. J., and T. Kouzarides. 1996. “The CBP Co-Activator Is a Histone Acetyltransferase.” *Nature* 384 (6610): 641–43. https://doi.org/10.1038/384641a0.

Boileau, Ryan M., Kevin X. Chen, and Robert Blelloch. 2023. “Loss of MLL3/4 Decouples Enhancer H3K4 Monomethylation, H3K27 Acetylation, and Gene Activation during Embryonic Stem Cell Differentiation.” *Genome Biology* 24 (1): 41. https://doi.org/10.1186/s13059-023-02883-3.

Buecker, Christa, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. 2014. “Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency.” *Cell Stem Cell* 14 (6): 838–53. https://doi.org/10.1016/j.stem.2014.04.003.

Chen, Amy F., Arthur J. Liu, Raga Krishnakumar, Jake W. Freimer, Brian DeVeale, and Robert Blelloch. 2018. “GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naïve Pluripotency.” *Cell Stem Cell* 23 (2): 226-238.e4. https://doi.org/10.1016/j.stem.2018.06.005.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21931–36. https://doi.org/10.1073/pnas.1016071107.

Ewels, Philip A., Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso, and Sven Nahnsen. 2020. “The Nf-Core Framework for Community-Curated Bioinformatics Pipelines.” *Nature Biotechnology* 38 (3): 276–78. https://doi.org/10.1038/s41587-020-0439-x.

Heintzman, Nathaniel D., Rhona K. Stuart, Gary Hon, Yutao Fu, Christina W. Ching, R. David Hawkins, Leah O. Barrera, et al. 2007. “Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome.” *Nature Genetics* 39 (3): 311–18. https://doi.org/10.1038/ng1966.

Hm, Chan, and La Thangue Nb. 2001. “P300/CBP Proteins: HATs for Transcriptional Bridges and Scaffolds.” *Journal of Cell Science* 114 (Pt 13). https://doi.org/10.1242/jcs.114.13.2363.

Kaya-Okur, Hatice S., Derek H. Janssens, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2020. “Efficient Low-Cost Chromatin Profiling with CUT&Tag.” *Nature Protocols* 15 (10): 3264–83. https://doi.org/10.1038/s41596-020-0373-x.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. “CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells.” *Nature Communications* 10 (1): 1930. https://doi.org/10.1038/s41467-019-09982-5.

Krishnakumar, Raga, Amy F. Chen, Marisol G. Pantovich, Muhammad Danial, Ronald J. Parchem, Patricia A. Labosky, and Robert Blelloch. 2016. “FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and Repressing

Enhancer Activity." *Cell Stem Cell* 18 (1): 104–17. https://doi.org/10.1016/j.stem.2015.10.003.

Lai, Binbin, Ji-Eun Lee, Younghoon Jang, Lifeng Wang, Weiqun Peng, and Kai Ge. 2017. "MLL3/MLL4 Are Required for CBP/P300 Binding on Enhancers and Super-Enhancer Formation in Brown Adipogenesis." *Nucleic Acids Research* 45 (11): 6388–6403. https://doi.org/10.1093/nar/gkx234.

Lee, Ji-Eun, Chaochen Wang, Shiliyang Xu, Young-Wook Cho, Lifeng Wang, Xuesong Feng, Anne Baldridge, et al. 2013. "H3K4 Mono- and Di-Methyltransferase MLL4 Is Required for Enhancer Activation during Cell Differentiation." *ELife* 2 (December): e01503. https://doi.org/10.7554/eLife.01503.

MacFawn, Ian, Hannah Wilson, Luke A. Selth, Ian Leighton, Ilya Serebriiskii, R. Christopher Bleackley, Osama Elzamzamy, et al. 2019. "GRAINYHEAD-LIKE-2 CONFERS NK-SENSITIVITY THROUGH INTERACTIONS WITH EPIGENETIC MODIFIERS." *Molecular Immunology* 105 (January): 137–49. https://doi.org/10.1016/j.molimm.2018.11.006.

Meers, Michael P., Dan Tenenbaum, and Steven Henikoff. 2019. "Peak Calling by Sparse Enrichment Analysis for CUT&RUN Chromatin Profiling." *Epigenetics & Chromatin* 12 (1): 42. https://doi.org/10.1186/s13072-019-0287-4.

Ogryzko, V. V., R. L. Schiltz, V. Russanova, B. H. Howard, and Y. Nakatani. 1996. "The Transcriptional Coactivators P300 and CBP Are Histone Acetyltransferases." *Cell* 87 (5): 953–59. https://doi.org/10.1016/s0092-8674(00)82001-2.

Park, Young-Kwon, Ji-Eun Lee, Zhijiang Yan, Kaitlin McKernan, Tommy O'Haren, Weidong Wang, Weiqun Peng, and Kai Ge. 2021. "Interplay of BAF and MLL4

Promotes Cell Type-Specific Enhancer Activation." *Nature Communications* 12 (1): 1630. https://doi.org/10.1038/s41467-021-21893-y.

Skene, Peter J., Jorja G. Henikoff, and Steven Henikoff. 2018. "Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers." *Nature Protocols* 13 (5): 1006–19. https://doi.org/10.1038/nprot.2018.015.

Smith, Austin. 2017. "Formative Pluripotency: The Executive Phase in a Developmental Continuum." *Development (Cambridge, England)* 144 (3): 365–73. https://doi.org/10.1242/dev.142679.

Stark, Rory, and Gord Brown. 2022. "DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data." Bioconductor version: Release (3.15). https://doi.org/10.18129/B9.bioc.DiffBind.

Sze, Christie C., and Ali Shilatifard. 2016. "MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer." *Cold Spring Harbor Perspectives in Medicine* 6 (11): a026427. https://doi.org/10.1101/cshperspect.a026427.

Visel, Axel, Matthew J. Blow, Zirong Li, Tao Zhang, Jennifer A. Akiyama, Amy Holt, Ingrid Plajzer-Frick, et al. 2009. "ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers." *Nature* 457 (7231): 854–58. https://doi.org/10.1038/nature07730.

Wang, Shu-Ping, Zhanyun Tang, Chun-Wei Chen, Miho Shimada, Richard P. Koche, Lan-Hsin Wang, Tomoyoshi Nakadai, et al. 2017. "A UTX-MLL4-P300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer

Landscapes for Eliciting Transcription." *Molecular Cell* 67 (2): 308-321.e6. https://doi.org/10.1016/j.molcel.2017.06.028.

Yang, Shen-Hsi, Tüzer Kalkan, Claire Morissroe, Hendrik Marks, Hendrik Stunnenberg, Austin Smith, and Andrew D. Sharrocks. 2014. "Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency." *Cell Reports* 7 (6): 1968–81. https://doi.org/10.1016/j.celrep.2014.05.037.

Chapter 4 : Conclusions

The work presented in this thesis demonstrated the existence of enhancers that do not rely on MLL3/4 for function including during the conversion to an active enhancer from a latent state. In the process, we found that changes in enhancer H3K4me1 and H3K27ac are largely dispensable for transcriptional upregulation of the formative program. Separately, we identified Grhl2 as a TF that requires MLL3/4 for enhancer remodeling - consistent with the canonical model of enhancer activation. We devised a synthetic approach using Grhl2 for rapid inducibility of enhancers and examined more deeply the role of MLL3/4 during GRHL2 enhancer activation. Preliminary results suggested MLL3/4 may not be an essential factor for enhancer-TF mediated transcription.

The results of these studies present new questions in the field and underscore persisting questions which have not yet been addressed. Below, I highlight these questions to address as important future directions in enhancer biology.

What is the function of active enhancer histone modifications?

Histone modifications H3K4me1 and H3K27ac have long been used to classify an active enhancer state. The presence of these modifications correlates with nearby gene expression levels, including a positive correlation between changes in marks and changes in gene expression(Creyghton et al. 2010). Previous efforts have identified biophysical and recruiting roles for H3K4me1/H3K27ac, many using biochemical *in vitro* assays. However, previous studies examining the functional roles of H3K4me1 or

H3K27ac at distal regulatory elements in cellular systems did not identify major roles for these histone modifications at enhancers (Rickels et al. 2017; Dorighi et al. 2017; Sankar et al. 2022). From our study in Chapter 2, we find most sites that gain in active histone modifications between the naive to formative transition are effectively uncoupled from changes in gene expression. This begs the question whether sites that gain active marks represent functioning enhancers that are dispensable for transcriptional upregulation or are enhancers functioning in the absence of active marks.

Studies that demonstrated the dispensability of H3K4me1 and or H3K27ac did so in ways that did not perturb CBP/P300 recruitment to enhancers. Catalytic activity of CBP/P300 targets a variety of substrates - histone and non-histone - and is likely key for enhancer function. Therefore, it may not be surprising that enhancer function is unaffected by the loss of H3K4me1 or H3K27ac individually. In our case, without MLL3/4, we presume CBP/P300 is not recruited to enhancers. While we cannot exclude the possibility that enhancers function without these modifications and CBP/P300 recruitment, I favor the possibility that many distal sites marked by active enhancer marks are not required for transcriptional activation of the formative program. Other gene regulatory mechanisms at the promoter-level or rewiring of the existing enhancer-promoter network may be larger drivers of formative gene upregulation than *de novo* enhancer activation. If active marks indicate a competency for enhancer function rather than function itself, then perhaps sites gaining active marks are becoming licensed in preparation for use later in development. Moreover, these sites may need an additional signaling cue that is not available *in vitro* cell culture. Either possibility is consistent with the proposed role of the formative state acting as a developmental staging period for

multi-lineage specification governed by environmental signaling within the embryo (Smith 2017).

How do mechanisms differ between enhancer maintenance and activation?

Barriers to activation are removed once an enhancer is activated. Several examples of molecular barriers include nucleosome occlusion, repressive histone marks, and DNA methylation. During activation, a functioning enhancer may recruit additional structural proteins to stabilize enhancer-promoter interactions, move the physical location of an enhancer to specific genomic compartments, or alter the biophysical microenvironment through liquid-liquid phase separation. The factors that mediate these molecular features, those removing barriers or those stabilizing enhancer function, may no longer be necessary for enhancer function after activation. In a simple case, a pioneer factor may only be required to rearrange the enhancer landscape by recruiting DNA demethylases and then other TFs bind and maintain enhancer function at unmethylated enhancers. Despite these theoretical differences between maintaining an enhancer and *de novo* enhancer activation, the two processes have not been widely studied as separate entities. These mechanisms may contribute to important and currently poorly grasped biology such as the phenomena of transcriptional memory whereby cells primed with a stimulus show more rapid or stronger transcriptional responses once restimulated (D'Urso and Brickner 2017).

Investigations of enhancer regulators that utilize a steady-state cell culture system are limited in the ability to distinguish between maintenance and activation. In this case,

de novo activation is not occurring. Enhancer induction by differentiation is another common method used in enhancer studies. However, differentiation protocols may require several days and entail the coordination of several gene programs. Consequently, a perturbation before or during differentiation may have outcomes where primary effects are hard to interpret. Animal models of development for enhancer studies have similar limitations, often with an added scarcity of tissue on which to conduct experiments. Approaches inducing single TF function or directly in combination with acute perturbation will be most suitable for comparing enhancer maintenance versus enhancer activation.

Intriguing results have already emerged using acute TF induction. A recent preprint utilized a similar system to ours in Chapter 3 whereby ERT was fused to the pioneer TF Pax7. Once activated, Pax7 was able to recruit H3K4me1/MLL4 activity immediately but required the cell cycle to recruit and deposit P300/H3K27ac(Gouhier et al. 2022). The authors propose division is required for relocation of enhancers away from Lamin associated (repressive) domains and towards euchromatic nuclear compartments. If this mechanism extends to more TFs than Pax7 this would suggest a mechanism based on compartmentalization occurring during the cell cycle for activation that is no longer necessary for other, established enhancers. Further work could focus on how repressive marks and structural factors contribute to the change in compartmentalization of Pax7 enhancers. Additionally, another interesting avenue will be how enhancer activation occurs in non-dividing cells.

What drives different models of enhancer activation?

Our study in Chapter 2 revealed that a subset of enhancer regions does not require MLL3/4 for enhancer activation including enhancers critical for key formative genes. There are a few potential reasons for the differences between enhancer subtypes. With respect to chromatin context, contributing factors may include DNA sequence, CpG content, 3D chromatin structure, and distance to target promoter.

In our analysis of the DNA sequence underlying enhancer subtypes, we identify differentially enriched TF motifs between sites that do or do not require MLL3/4 for active marks. While we cannot exclude other factors related to chromatin context, I hypothesize that the underlying mechanism is a diversity of TFs which function to drive other modes of enhancer activation. In support of this hypothesis, TFs have been shown to directly interact with CBP or P300 which may bypass the need for MLL3/4 during enhancer activation (Dancy and Cole 2015). Secondarily, pluripotency TFs still appear to be functioning in the absence of MLL3/4 such as those controlling Nanog and Oct4 expression to maintain a naïve-like ESC state (C. Wang et al. 2016). Considering Otx2 is a required TF for the formative transcriptional program including activation of MLL3/4 independent Fgf5 enhancers I also speculate MLL3/4 is not obligatory for Otx2 TF activity (Buecker et al. 2014).

Recent work suggests enhancers are diverse in their functional dependency on chromatin regulators including different members of the Mediator complex and CBP/P300 (Neumayr et al. 2022; Narita et al. 2022). If TFs are driving different modes of enhancer activation the mechanism would presumably be through recruitment of a diverse range of

chromatin regulators. Co-bound TFs that mediate a variety of different mechanisms at the same enhancer site could have additive or synergistic effects in regulating transcription. These effects in concert could more finely tune developmental gene regulation and or buffer against the loss in function of a single TF. The ways in which enhancer subtypes arise and differ mechanistically should be a priority for continued investigation.

Where do chromatin regulators fit in the canonical model of enhancer activation?

Our studies focused on a set of chromatin regulating complexes MLL3/4 and their relationship with CBP/P300. However, the functional relationship between MLL3/4 and other factors has been explored. Knockout of H3.3 genes or acute depletion of Cohesin subunits do not lead to notable decreases in H3K4me1(Martire et al. 2019; Rao et al. 2017). Further, a recent study found that PAXIP1, a subunit of the complex containing MLL3/4, was able to interact and recruit Cohesin independently of MLL4 at GR-bound genomic sites (Mayayo-Peralta et al. 2023). A similar finding has been made in a preprint demonstrating a direct interaction between PAXIP1-PAGR1 and Cohesin(Schie et al. 2022). These observations together suggest HIRA/H3.3 and Cohesin accumulation at enhancers operates independently of MLL3/4. Whether these events occur in parallel or upstream of both MLL3 and MLL4 is not clear. The relationship between TFs, Cohesin, and MLL3/4 may be particularly interesting to delineate in order to better understand the connection between 3D genomic structure and transcriptional regulation.

Cooperative function between key chromatin regulators has been proposed, albeit not a focus in our studies. Synergism in enhancer recruitment and function has been

suggested between MLL4 and P300 as well as MLL4 and BRG1 (S.-P. Wang et al. 2017; Park et al. 2021). However, compensation due to redundant function between the remaining homologous family members were not fully considered for these studies (i.e., MLL3, CBP, or BRM). Nonetheless, one cannot exclude cooperativity as a possibility in existing models of enhancer activation. It may be true that some mechanisms of enhancer activation are essential steps that occur sequentially while other mechanisms function as an equilibrium among a collection of factors. Either way, it remains unclear the extent to which cooperativity may play a role and this concept needs to be revisited.

Many additional factors may play essential roles at enhancers which have not been described with respect to the molecular steps of activation. As examples, SRCAP/P400 are chaperones that deposit H2AZ in euchromatic regions including enhancers (Martire and Banaszynski 2020) and PCAF/GCN5 are histone acetyltransferases which are coactivators of CBP/P300 and may acetylate histones at enhancers(Dancy and Cole 2015). How the composition of multi-subunit complexes affects enhancer function is also unknown. For instance, the BAF complex is subcategorized into several forms cBAF, ncBAF, gBAF, and esBAF which contain exclusive subunits and have varied functions (Kadoch and Crabtree 2015). Finally, how post-translational modifications to chromatin regulators affect enhancer function is underexplored. On MLL4 alone, over 30 residues are differentially phosphorylated throughout the naive to formative transition (Yang et al. 2019). Individual modifications may have function and combinations of phosphorylated residues may further alter chromatin regulator function at enhancers. Further work should construct a more holistic picture of the molecular events required for enhancer activation including additional factors, complex composition, and post-translational modifications.

Concluding Remarks

Recent developments in genomic technologies have revolutionized our ability to study enhancer biology with an unprecedented level of detail. As a result, models of enhancer function are being refined and expanded. The studies in this thesis point to an underappreciated diversity of mechanisms for enhancer activation which reshape previous, generalized molecular models. Factors and processes that need to be accounted for in revised models of enhancer biology include subtypes of cis-regulatory elements, 3D genomic structure, cell cycle dynamics, and the biophysics of the chromatin environment. These areas are currently undergoing significant transformation in the enhancer field as we build our knowledge of their molecular underpinnings. Developing diverse models that take these processes, factors, and their regulators into account is necessary to build our understanding of enhancer function. Further discoveries on the fundamentals of enhancer biology will, ultimately, prove extraordinarily useful for understanding the manifestation of many diseases.

References

Buecker, Christa, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. 2014. “Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency.” *Cell Stem Cell* 14 (6): 838–53. https://doi.org/10.1016/j.stem.2014.04.003.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21931–36. https://doi.org/10.1073/pnas.1016071107.

Dancy, Beverley M., and Philip A. Cole. 2015. “Protein Lysine Acetylation by P300/CBP.” *Chemical Reviews* 115 (6): 2419–52. https://doi.org/10.1021/cr500452k.

Dorighi, Kristel M., Tomek Swigut, Telmo Henriques, Natarajan V. Bhanu, Benjamin S. Scruggs, Nataliya Nady, Christopher D. Still, Benjamin A. Garcia, Karen Adelman, and Joanna Wysocka. 2017. “Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation.” *Molecular Cell* 66 (4): 568-576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.

D’Urso, Agustina, and Jason H. Brickner. 2017. “Epigenetic Transcriptional Memory.” *Current Genetics* 63 (3): 435–39. https://doi.org/10.1007/s00294-016-0661-8.

Gouhier, Arthur, Justine Dumoulin-Gagnon, Vincent Lapointe-Roberge, Aurelio Balsalobre, and Jacques Drouin. 2022. “Pioneer Factor Pax7 Initiates Two-Step Cell-Cycle Dependent Chromatin Opening.” bioRxiv. https://doi.org/10.1101/2022.11.16.516735.

Kadoch, Cigall, and Gerald R. Crabtree. 2015. “Mammalian SWI/SNF Chromatin Remodeling Complexes and Cancer: Mechanistic Insights Gained from Human Genomics.” *Science Advances* 1 (5): e1500447. https://doi.org/10.1126/sciadv.1500447.

Martire, Sara, and Laura A. Banaszynski. 2020. “The Roles of Histone Variants in Fine-Tuning Chromatin Organization and Function.” *Nature Reviews. Molecular Cell Biology* 21 (9): 522–41. https://doi.org/10.1038/s41580-020-0262-8.

Martire, Sara, Aishwarya A. Gogate, Amanda Whitmill, Amanuel Tafessu, Jennifer Nguyen, Yu-Ching Teng, Melodi Tastemel, and Laura A. Banaszynski. 2019. “Phosphorylation of Histone H3.3 at Serine 31 Promotes P300 Activity and Enhancer Acetylation.” *Nature Genetics* 51 (6): 941–46. https://doi.org/10.1038/s41588-019-0428-5.

Mayayo-Peralta, Isabel, Sebastian Gregoricchio, Karianne Schuurman, Selçuk Yavuz, Anniek Zaalberg, Aleksandar Kojic, Nina Abbott, et al. 2023. “PAXIP1 and STAG2 Converge to Maintain 3D Genome Architecture and Facilitate Promoter/Enhancer Contacts to Enable Stress Hormone-Dependent Transcription.” *Nucleic Acids Research*, April, gkad267. https://doi.org/10.1093/nar/gkad267.

Narita, Takeo, Yoshiki Higashijima, Sinan Kilic, Tim Liebner, Jonas Walter, and Chunaram Choudhary. 2022. “A Unique H2B Acetylation Signature Marks Active Enhancers and Predicts Their Target Genes.” *BioRxiv*, January, 2022.07.18.500459. https://doi.org/10.1101/2022.07.18.500459.

Neumayr, Christoph, Vanja Haberle, Leonid Serebreni, Katharina Karner, Oliver Hendy, Ann Boija, Jonathan E. Henninger, et al. 2022. “Differential Cofactor Dependencies Define Distinct Types of Human Enhancers.” *Nature* 606 (7913): 406–13. https://doi.org/10.1038/s41586-022-04779-x.

Park, Young-Kwon, Ji-Eun Lee, Zhijiang Yan, Kaitlin McKernan, Tommy O’Haren, Weidong Wang, Weiqun Peng, and Kai Ge. 2021. “Interplay of BAF and MLL4 Promotes Cell Type-Specific Enhancer Activation.” *Nature Communications* 12 (1): 1630. https://doi.org/10.1038/s41467-021-21893-y.

Rao, Suhas S. P., Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, et al. 2017. “Cohesin Loss Eliminates All Loop Domains.” *Cell* 171 (2): 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026.

Rickels, Ryan, Hans-Martin Herz, Christie C Sze, Kaixiang Cao, Marc A Morgan, Clayton K Collings, Maria Gause, et al. 2017. “Histone H3K4 Monomethylation Catalyzed by Trr and Mammalian COMPASS-like Proteins at Enhancers Is Dispensable for Development and Viability.” *Nature Genetics* 49 (11): 1647–53. https://doi.org/10.1038/ng.3965.

Sankar, Aditya, Faizaan Mohammad, Arun Kumar Sundaramurthy, Hua Wang, Mads Lerdrup, Tulin Tatar, and Kristian Helin. 2022. “Histone Editing Elucidates the

Functional Roles of H3K27 Methylation and Acetylation in Mammals." *Nature Genetics* 54 (6): 754–60. https://doi.org/10.1038/s41588-022-01091-2.

Schie, Janne J. M. van, Klaas de Lint, Thom M. Molenaar, Macarena Moronta Gines, Jesper A. Balk, Martin A. Rooimans, Khashayar Roohollahi, et al. 2022. "CRISPR Screens in Sister Chromatid Cohesion Defective Cells Reveal PAXIP1-PAGR1 as Regulator of Chromatin Association of Cohesin." bioRxiv. https://doi.org/10.1101/2022.12.23.521474.

Smith, Austin. 2017. "Formative Pluripotency: The Executive Phase in a Developmental Continuum." *Development (Cambridge, England)* 144 (3): 365–73. https://doi.org/10.1242/dev.142679.

Wang, Chaochen, Ji-Eun Lee, Binbin Lai, Todd S. Macfarlan, Shiliyang Xu, Lenan Zhuang, Chengyu Liu, Weiqun Peng, and Kai Ge. 2016. "Enhancer Priming by H3K4 Methyltransferase MLL4 Controls Cell Fate Transition." *Proceedings of the National Academy of Sciences of the United States of America* 113 (42): 11871–76. https://doi.org/10.1073/pnas.1606857113.

Wang, Shu-Ping, Zhanyun Tang, Chun-Wei Chen, Miho Shimada, Richard P. Koche, Lan-Hsin Wang, Tomoyoshi Nakadai, et al. 2017. "A UTX-MLL4-P300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription." *Molecular Cell* 67 (2): 308-321.e6. https://doi.org/10.1016/j.molcel.2017.06.028.

Yang, Pengyi, Sean J. Humphrey, Senthilkumar Cinghu, Rajneesh Pathania, Andrew J. Oldfield, Dhirendra Kumar, Dinuka Perera, et al. 2019. “Multi-Omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency.” *Cell Systems* 8 (5): 427-445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
8CFE90A10E2A4BE...
Author Signature

5/16/2023
Date