UNDERSTANDING EPO-DEPENDENT ENHANCER-PROMOTER INTERACTIONS IN THE REGULATION OF ERYTHROID GENE EXPRESSION

By

Andrea Anne Perreault

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

May 8, 2020

Nashville, Tennessee

Approved:

Richard O'Brien, Ph.D.

Jonathan Brown, MD

Stephen Brandt, MD

Glenn Webb, Ph.D.

I dedicate this doctoral dissertation to my parents-- Marc and Madeleine. Thank you for the constant support, encouragement, and confidence that I needed to pursue this challenge.

ACKNOWLEDGEMENTS

The work presented in this dissertation was possible due to the training, collaboration, and support of many people during my time at Vanderbilt University. Thus, I would like to take the time to thank those people.

First and foremost, thank you to my research advisor and mentor, Dr. Bryan Venters. His training and support have made me the scientist I am today. I would also like to thank the professors on my committee, whom have been extremely helpful in guiding me and my project. Dr. Richard O'Brien provided valuable expertise due to his research in gene regulation, years of expertise as a professor being on innumerable committees, and his insights as a director of graduate studies. Dr. Jonathan Brown always contributed to discussions with excitement, as well as provided excellent guidance in the completion of my dissertation work. Dr. Stephen Brandt shared his expertise in erythroid biology and a medical perspective for the potential impact of my work. Dr. Glenn Webb added a unique perspective and was a constant reminder to explain my work in such a way to assure the understanding by a diverse audience.

I have had the pleasure of working with, training, and mentoring a wonderful group of students during my time in the Venters lab. Thank you to Spencer Waddle, Tyler Hansen, Doug Shaw, Zenab Mchaourab, Cody Heiser, and Joe Breeyear for the opportunity to share my passion for science and to hone my skills in teaching and mentoring.

I would like to thank the administrative staff at Vanderbilt that has been there to answer all of my many questions and provide support outside of the lab. To Carolyn Berry-- thank you for always having an open door, a comfy chair, a genuine laugh, and endless advice. To Beth Bowman-- thank you for your dedication to your students and for reading a hundred versions of

various documents. To Bruce Damon and Patty Mueller-- thank you for running the CPB program so smoothly and creating a supportive community.

Finally, I would like to thank my family and friends. To Mom, Dad, and Julien-- thank you for the endless support and encouragement, despite not knowing a single detail about what I've been doing for the last 5 years. To Kayla-- thank you for being my first friend in Nashville, for sharing your love of cooking, wine, and Grey's, and for always knowing what to say. To Mary Lauren-- thank you for showing me the magic of wigs, sharing my addiction to books and lipstick, and seeing any and all concerts with me. To Corey and Ian-- thank you for helping me survive QCB and grad school with Marvel movies and meat cookies. To Justine and Rachel-- thank you for never saying no to a happy hour, new restaurant, or adventure, for being the best 50% of the power quad, and for lifting me up when I needed it most. To all of the friends I've met during my time at Vanderbilt and in Nashville-- thank you for being a part of my life. I couldn't have done it without you all.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Gene regulation

The Central Dogma of Biology

Cells contain the entirety of the organism's genetic material in the form of deoxyribose nucleic acid (DNA). It is comprised of three main components- the nucleotide, the deoxyribose sugar, and the phosphate group. Although DNA had been studied for a hundred years, James Watson and Frances Crick are credited with discovering the double helix structure of DNA in 1953[1,2]. In this structure, the deoxyribose sugar and phosphate group form the backbone of the helix and nucleotide pairs form the rungs of the so-called twisted ladder. Crick is also considered the father of the Central Dogma, a foundational theory in biology[3]. The Central Dogma states that DNA is transcribed into ribonucleic acid (RNA), which is translated into protein. DNA contains the instructions necessary for cell, and therefore organism, survival. For this reason, research has been focused on understanding the steps and mechanisms by which the Central Dogma drives cellular function.

Transcription is the process by which DNA is converted into the different type of nucleic acid called RNA. The function of DNA and RNA vary dramatically. DNA is the long-term storage of genetic information for cells and organisms as a whole. RNA comes in a variety of forms to appropriately read and interpret this information, such as messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). mRNA, as its name implies, acts as a

messenger between DNA and the cell. Since transcription is a vital part of cellular function, it has been and will continue to be the subject of extensive study in the field of gene regulation.

mRNA is then converted into protein through the process of translation. Translation takes place within the ribosome, which is composed of rRNA. The ribosome is made of two subunits that surround the mRNA molecule. Within the ribosome are tRNA, which serve as a bridge between mRNA and the end product of amino acids. mRNA is translated into amino acids using codons, a sequence of three amino acids that encode for a specific amino acid. The end of the tRNA has the anticodon, which is used to bind its complementary codon in the mRNA. The anticodon on the tRNA dictates which amino acid it carries and will contribute to the polypeptide chain. This polypeptide chain may require additional processing or translocation to another area of the cell for proper function as a protein. Translation is not as extensively studied as transcription and therefore is an area of gene regulation that remains to be completely understood.

RNA Polymerase II and the transcriptional machinery

Eukaryotes employ a variety of RNA polymerase protein complexes to synthesize distinct types of RNA. RNA polymerase II (Pol II) is necessary for the transcription of mRNA from the template strand of DNA. Pol II's structure, function, and mechanism of action has been extensively studied, resulting in a deep understanding of its role in transcription.

Transcription is a three step process- initiation, elongation, and termination. Each phase has been studied in isolation and in conjunction with one another. Studies have revealed that signals from a variety of proteins and factors are required for proper procession from one stage to the next. The preinitiation complex (PIC) is a large complex that can include over 100 proteins

that function together to facilitate transcription. The necessary components of the PIC are Pol II and the general transcription factors (GTFs)[4]. As the name suggests, GTFs are broadly used by the cell, as opposed to binding to sequence-specific regions of DNA (**Figure 1.1**)[5]. The GTFs bind to the TATA box, a region of repeating thymines and adenines within the promoter region. This facilitates the recruitment of Pol II to the PIC directly upstream of the transcription start site of a gene. The transcription start site (TSS) is the 5' most end of the gene, indicating the beginning of the gene sequence. At this point, Pol II is bound to the promoter but is not synthesizing RNA[6].

The PIC then denatures the DNA, breaking the hydrogen bonds between nucleotide pairs and effectively unwinding the double helix. This unwinding creates what is called the transcription bubble, allowing access to the now unpaired nucleotides and proceed to the elongation step. Pol II adds complementary RNA nucleotides to the DNA nucleotides present (**Figure 1.1**). Pol II often pauses after transcribing a short span of RNA, about 20-60 nucleotides. This is known as promoter-proximal pausing and is an important mechanism of gene regulation (**Figure 1.1**)[6-9]. There are several factors that induce and release promoter-proximal pausing. DRB Sensitivity Inducing Factor (DSIF) is composed of two subunits and has both transcriptional activation and repressing functions. Negative elongation factor (NELF) is comprised of four subunits and when it interacts with DSIF, the two proteins decrease elongation rates and stabilize the transcriptional machinery in the paused state[6].

Promoter-proximal pausing is released by the recruitment of positive transcription elongation factor (P-TEFb). P-TEFb is a cyclin dependent kinase that phosphorylates DSIF, NELF, and the C-terminal tail of Pol II. This phosphorylation causes release of the DSIF-NELF complex, switching DSIF's function to an activator of transcription. The transcriptional

machinery can then proceed into productive elongation[6]. Studies have investigated the early and late stages of elongation to complement the extensive work conducted to understand the transition from paused Pol II to elongating Pol II (**Figure 1.1**)[10].

Chromatin: DNA, nucleosomes, and histones

As stated previously, DNA holds the information for cells to survive. DNA is compacted into the nucleus of cells using proteins called nucleosomes, the basic unit of chromatin. Each nucleosome is comprised of eight histone proteins. A 147 base pair segment of DNA wraps itself around a nucleosome 1.65 times. Nucleosomes fold together to form the 30 nanometer fiber, which is compressed and folded in various higher order structures to ultimately form the chromatid of a chromosome[11-13]. This DNA compaction allows for the storage of DNA within the nucleus in a space and energy favorable manner.

The question then becomes- how does the transcriptional machinery access regions of DNA to be transcribed? This can occur via two mechanisms- histone modification and histone displacement. Briefly, histone modification consists of the addition of acetyl, methyl, or phosphate groups to specific histone proteins in the nucleosome[12,14,15]. This topic will be covered in detail in the *Epigenetics* section.

Alternatively, the nucleosome itself can be moved through the recruitment of chromatin remodeling complexes. The movement of the nucleosome protein exposes the underlying DNA sequences to the transcriptional machinery, as well as other proteins necessary for transcription and regulation[12,16]. There are four main groups of ATP-dependent, meaning they require energy input, chromatin remodeling complexes- imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and INO80[5,17,18].

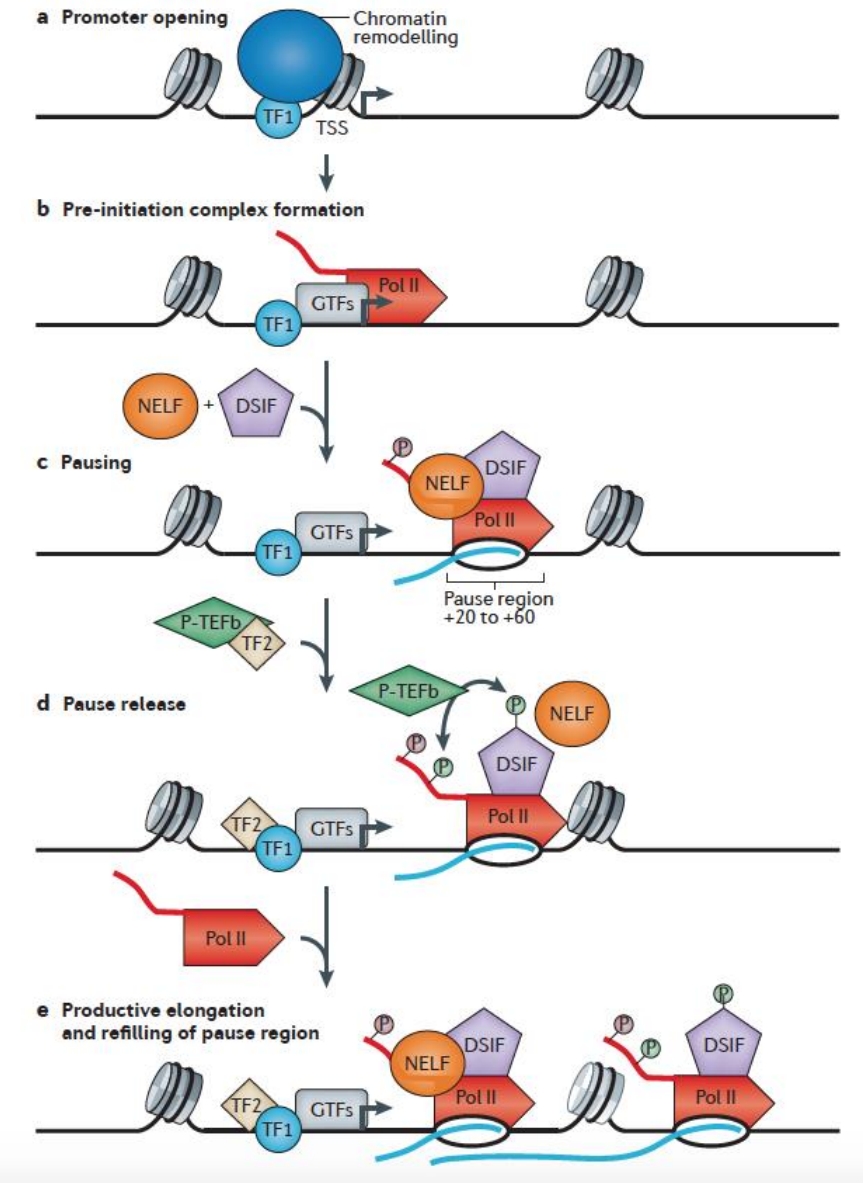

Figure 1.1. Establishment and release of paused Pol II. (**A**) Promoter opening involves the binding of sequence specific TFs that recruit chromatin remodeling enzymes that provide accessible DNA for the transcription machinery. (**B**) PIC formation involves the recruitment of GTFs, Pol II, and TFs. (**C**) Pol II pausing occurs after initiation and involves DSIF and NELF (**D**) Pause release is triggered by the recruitment of p-TEFb. (**E**) Escape of paused Pol II results in productive elongation. Another Pol II is rapidly recruited to the pause site, allowing for efficient RNA production. [Figure adapted from Adelman and Lis (2012) Nat Gen Rev[6] and used in accordance with Copyright Clearance Center's RightsLink service.]

The ISWI subfamily of chromatin remodelers assemble nucleosomes and regularly space them to facilitate the compaction of DNA. This inherently creates a closed chromatin conformation, inhibiting transcription and gene expression[5]. The CHD subfamily is comprised of a wide variety of proteins in metazoans. Despite this diversity, these remodelers function to space nucleosomes across DNA, rearrange the position of nucleosomes on DNA, and edit the composition of nucleosomes[5,17]. The SWI/SNF subfamily primarily function in nucleosome readjusting, sliding or removing nucleosomes to facilitate open or closed chromatin[17]. This therefore plays an important role in gene activation or repression, respectively[5]. The INO80 subfamily has been primarily studied in *Saccharomyces cerevisiae*, or budding yeast, and comprises of two main members, Ino80 and Swr1[17]. These remodelers have unique editing functions, such that they replace canonical histone proteins in the nucleosome with histone variants[5]. Mammalian orthologs to the components of this subfamily were discovered in 2005 and shown to exhibit similar functions as the yeast[19].

The assembly, movement, and rearrangement of nucleosomes and associated histone proteins is an integral part of gene regulation. Therefore, the proteins that act upon nucleosomes and their function have been the subject of expansive research, in both normal and diseased states. For more information, see the *Epigenetics* section.

Epigenetics: how DNA is interpreted for gene expression

Epigenetics

Epigenetics is the heritable transmission of gene expression patterns that are a result of DNA modification without changing the underlying DNA sequence. This field within gene regulation has been the subject of recent extensive study. There are three modes by which

epigenetics can contribute to the modification of DNA, thereby influencing the genes that are expressed- DNA methylation, histone modifications, and nucleosome positioning[12,18].

DNA methylation

DNA methylation is studied in the context of epigenetics at cytosines, specifically at CpG dinucleotides. CpG dinucleotides are are regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence and these are usually found in CpG islands, which are regions of more than 200 nucleotides with a cytosine and guanine content of at least 50%. DNA methylation is canonically associated with gene silencing[18]. For example, methylated CpG islands blocks transcription of genes surrounding this region. Additionally, DNA methylation in the gene body can be used to control transcription in an exon-specific manner. Repetitive sequences in the genome are methylated for genomic stability. DNA methylation is controlled by DNA methyltransferase enzymes, known as the DNMT family. Members of this family are recruited to chromatin by a variety of factors and are responsible for methylating specific regions of the genome (**Figure 1.2**).

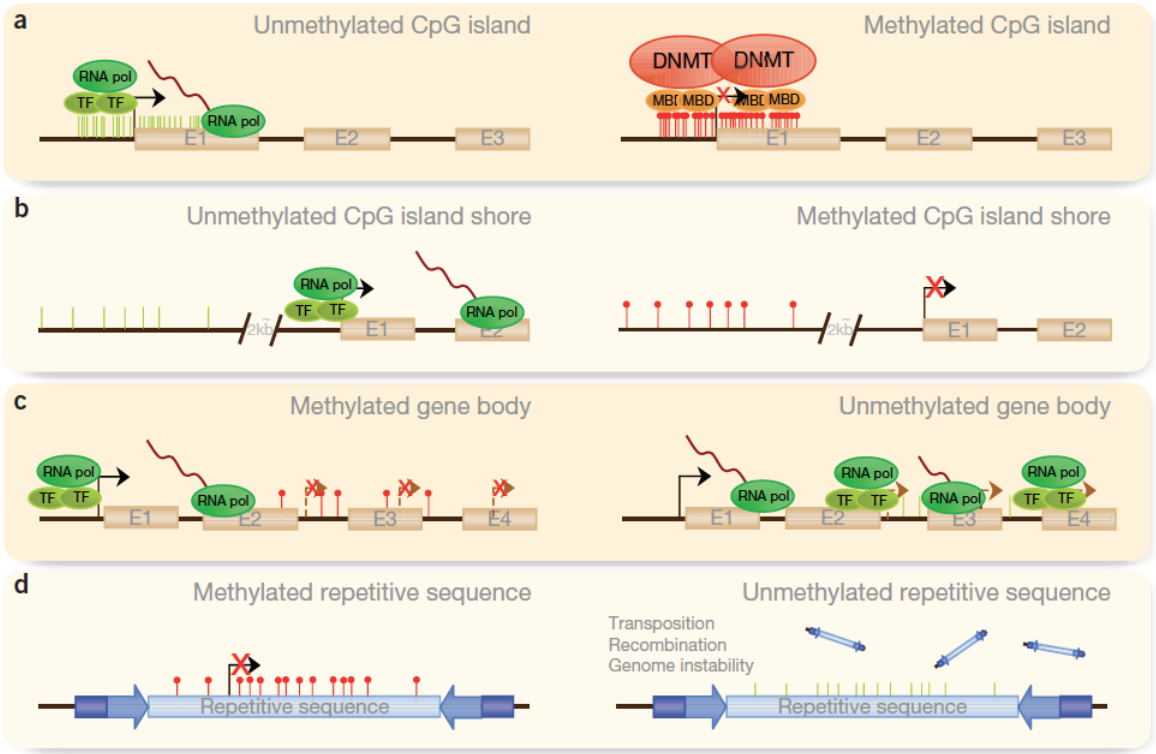

Figure 1.2. DNA methylation patterns in normal and disease scenarios. (**A**) CpG islands at promoters are usually unmethylated, which facilitates active transcription. (**B**) CpG island shores are also normally unmethylated. (**C**) Methylation of the gene body allows for proper transcription, as opposed to the disease state where demethylated gene bodies initiate the transcription at incorrect sites. (**D**) Repetitive sequences are hypermethylated to promote genome stability and prevent translocations. [Figure adapted from Portela and Esteller (2010) Nat Biotech Rev[18] and used in accordance with Copyright Clearance Center's RightsLink service.]

Histone modifications

As mentioned previously, DNA is compacted into chromatin by wrapping itself around nucleosomes. Nucleosomes are comprised of 4 core histones: H2A, H2B, H3, and H4. These histones have tails that extend from the nucleosome and are modified in a multitude of ways by a variety of enzymes. The most common histone modifications are acetylation, methylation, phosphorylation, and ubiquitination[20]. These modifications can be deposited and removed at specific locations on the tails of distinct core histones. The genome can be divided into actively transcribed and inactive chromatin through patterns of histone modifications. Euchromatin,

which is active or open to transcription, is canonically associated with acetylation and trimethylation at H3K4, H3K36, and H3K79. Heterochromatin, which is closed and therefore transcriptionally repressed, is canonically associated with H3K9, H3K27, and H4K20 methylation[21]. These patterns have spurred the study and creation of a so-called "histone code", which aims to generate a map of chromatin states and define their functional significance in relation to gene expression[22-24]. Histone writers and erasers add and remove particular modifications. For example, DNMT can also methylate residues on histone tails in addition to methylating the DNA itself. Demethylases will remove methylation marks. These enzymes are usually specific to histones or residues. Alternatively, histone acetyltransferases (HATs) and histone deacetylases (HDACs) will add and remove acetylation marks more broadly, respectively (**Figure 1.3**)[20].

Chromatin remodeling

The machinery involved in the movement, removal, and rearrangement of nucleosomes in the genome to facilitate transcription was discussed in the *Chromatin: DNA, nucleosomes, and histones* section. Here, histone variants will be discussed. Histone variants differ from the core histones in that they are not used by DNA during replication and are the primary regulators of nucleosome position in regulating gene expression. The most extensively studied histone variant is H2A.Z, which replaces the H2A histone in the H2A-H2B dimer. H2A.Z has been found to be protective against DNA methylation, thereby keeping chromatin open for transcription and gene expression. This histone variant can be modified like the core histones, such as acetylation and ubiquitination. More recently, H2A.Z has been found to be incorporated near TSSs of actively transcribed genes, as well as enhancers[25].

Figure 1.3. A small selection of histone modifications. All histones are subject to post-translational modification of the histone tails. The gray numbers represent the position of the amino acids that is being modified. The most common modifications are depicted above: acetylation (blue), methylation (red), phosphorylation (yellow), and ubiquitination (green). [Figure adapted from Portela and Esteller (2010) Nat Biotech Rev[18] and used in accordance with Copyright Clearance Center's RightsLink service.]

Enhancers: regulatory regions that control gene expression

Enhancers are regions of the genome that regulate the transcription of target genes in a position and orientation independent manner, facilitating the different gene expression programs of different cell and tissue types (**Figure 1.4**)[13,26-28]. Enhancers have a set of defining characteristics, including being hypersensitive to DNase treatment, marked by specific histone modifications, and containing binding sites for TFs. The identification, study, and dissection of enhancers has been the subject of intense research in the fields of gene regulation, developmental biology, and disease treatment.

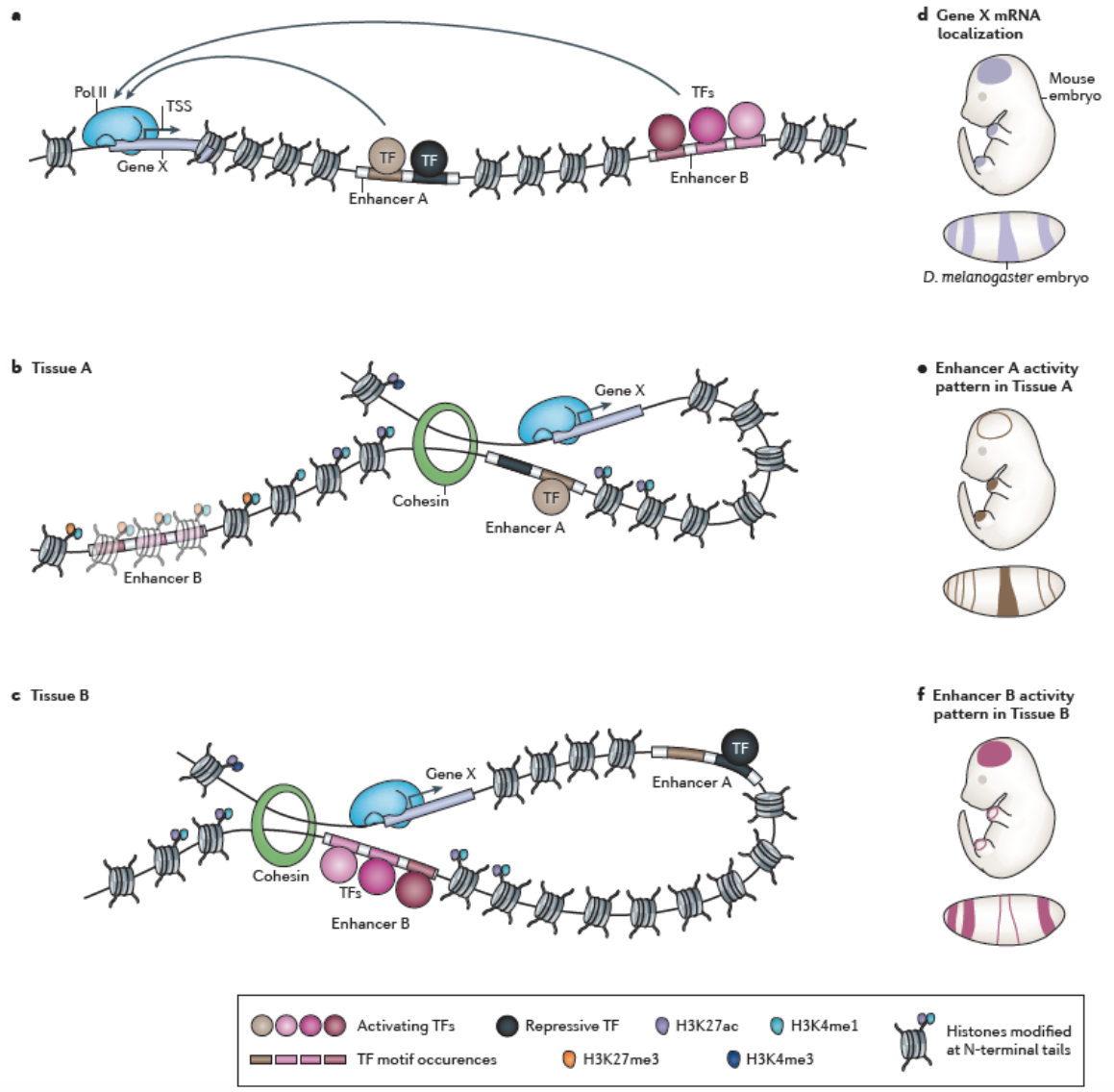

Figure 1.4. Enhancers and their genomic features. (**A**) Enhancers are regions of the genome that regulate the transcription of a target gene. Enhancers also contain binding sequences for TFs. (**B**, **C**) In a specific tissue (or cell type, not shown here), active enhancers are bound by TFs and are brought into genomic proximity to their target gene promoters through chromatin looping. Nucleosomes flanking enhancer regions are marked with specific histone modifications, namely H3K4me1 and H3K27ac. Enhancers can be silenced through repressive chromatin marks and TF binding. (**D-F**) Complex gene expression patterns are the result of cell- and tissue-specific enhancer activities. [Figure adapted from Shlyueva et al (2014) Nat Gen Rev[13] and used in accordance with Copyright Clearance Center's RightsLink service.]

Deoxyribonuclease (DNase) is an enzyme that catalyzes the breakdown of the bonds in the DNA backbone, therefore degrading DNA. Regions that are sensitive to DNase treatment indicates that these are open regions of chromatin. As discussed in previous sections (see *Chromatin: DNA, nucleosomes, and histones* and *Chromatin remodeling*), nucleosome position and composition can be altered as a mechanism of gene regulation. These open regions of the genome often contain enhancers, which along with the promoters and body of target genes, must be accessible to the transcriptional machinery and other proteins.

The development of a "histone code" has been used to decipher patterns in histone modification presence and their role in gene regulation[29-32]. Enhancer regions are open, but the histones flanking these regions have been found to be marked by a specific set of modifications. Namely, H3K4me1 and H3K27ac have been identified as markers of active enhancer regions, with the presence of H3K27ac separating active from poised enhancers[33-35]. Alternatively, H3K9me3 is canonically associated with repressed regions of the genome, suggesting genes within this region are not transcribed[36]. These marks, among others, have been the subject of extensive study by individual labs and have sparked the creation of consortia, aimed at annotating the functional regulatory elements found in the genome[37-39].

The presence of TF consensus binding motifs identified computationally or TF occupancy established through ChIP-based assays has been used to locate and validate enhancer regions. Transcription factors recognize and bind short sequences that often include a combination of stable and degenerate nucleotides. TF motif clusters are often found in enhancer regions, suggesting regulatory roles for the TFs in enhancer function. Studies have identified TF binding locations through ChIP-seq and similar assays (See *High-throughput sequencing and the development of –seq assays* section for more information). Although this addresses one aspect of

simply finding consensus motifs in the genome, a bound TF does not necessarily imply that it is regulating transcription through the enhancer it is bound to[40]. Instead, studies have shown that occupancy of important cell type TFs, such as TAL1 in erythroid cells, in conjunction with transcriptional data is more suggestive of enhancer function than investigating histone modifications[41]. Understanding these discrepancies and identifying functional significance of TFs in enhancer function is an ongoing area of research.

With the abundance of data and new methodologies, laboratories studying enhancers have been able to identify classes and categories of enhancers (**Figure 1.5**). These are usually studied in the context of differentiation or development, such cellular differentiation[42] and development from fetus to adult[43,44]. Poised enhancers, as mentioned briefly, have the canonical H3K4me1 mark but not the activating H3K27ac mark[35]. Additionally, a class of enhancers termed "latent" acquire both H3K4me1 and H3K27ac in response to external stimulus[45]. Identifying enhancers during differentiation or development has expanded our knowledge of gene regulation, providing enhancer repertoires that are activated at specific stages.

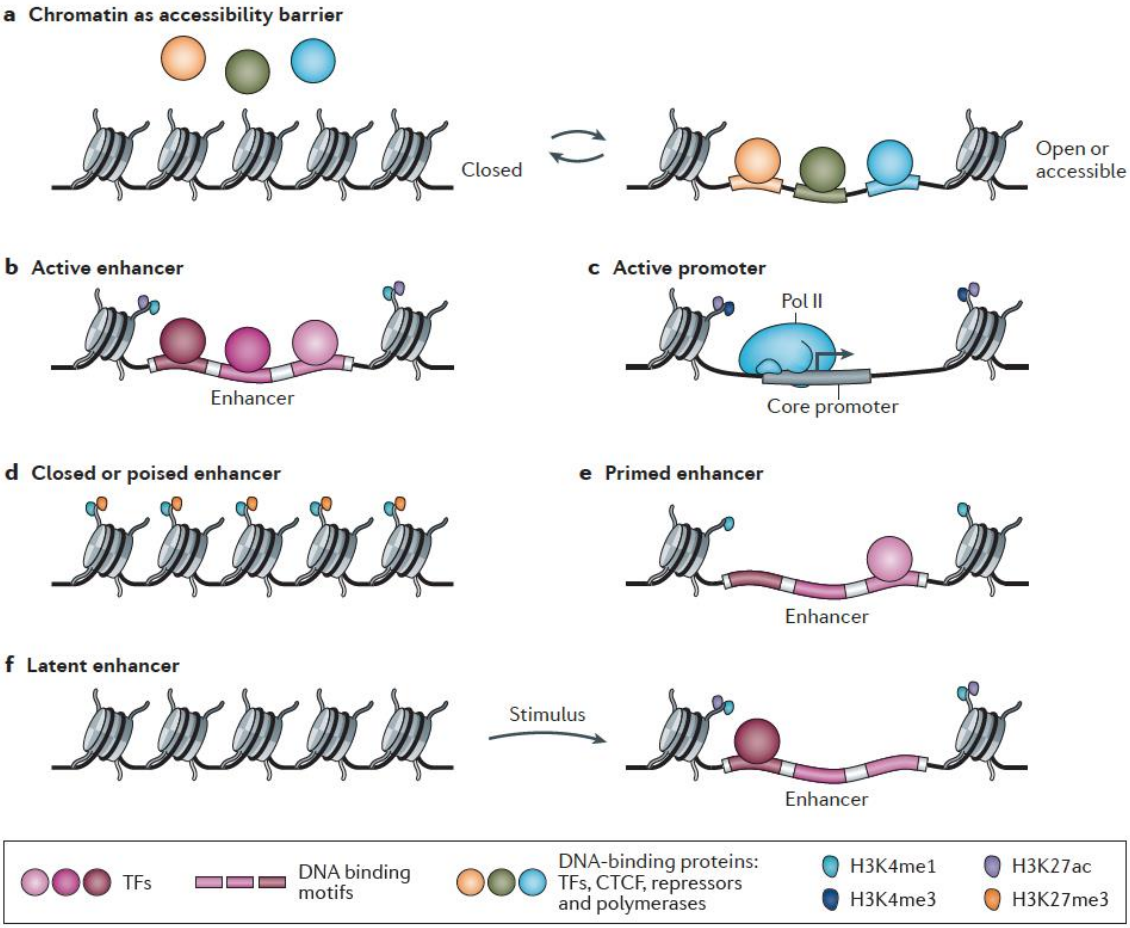

Figure 1.5. Chromatin accessibility and histone marks as mechanisms of epigenetics. (**A**) Chromatin can be open or closed to TFs, CTCF, Pol II, and other proteins and transitions between these phases to control gene expression. (**B**) Nucleosomes flanking active enhancers are marked by H3K4me1 and H3K27ac. (**C**) Active promoters are found near nucleosomes bearing H3K4me3 and H3K27ac histone modifications. (**D**) Poised enhancers are marked by the H3K4me1 mark associated with enhancers, but have the H3K27me3 repressive mark or lack the H3K27ac activation mark. (**E**) Primed enhancers are not yet active but are ready for activation at a later developmental point or in response to stimuli and thus are pre-marked with H3K4me1. (**F**) Latent enhancer are found in inaccessible regions of the genome. However, upon external stimuli the DNA becomes accessible and the histones acquire the enhancer-associated histone modifications. [Figure adapted from Shlyueva et al (2014) Nat Gen Rev[13] and used in accordance with Copyright Clearance Center's RightsLink service.]

Shadow enhancers are a category of enhancer that have been studied and discussed recently in the literature[46]. The idea of a shadow, or secondary, enhancer was brought forth by the Levine lab in 2008[47]. The authors suggest that shadow enhancers for two *Drosophila* genes are further from the target gene than its primary enhancer, but have overlapping activity as an evolutionary source of gene expression stability. Subsequently, the Cervera lab identified a secondary enhancers necessary for the proper transcription and expression of the Troponin 1 genes[48]. The Stern lab also identified a secondary enhancer that controlled the *Drosophila* TF *shavenbaby,* showing that the two enhancers induced similar expression patterns[49]. This report suggested that shadow, or secondary, enhancers were an evolutionary adaption to provide robust gene expression of important genes. The Furlong lab took the idea of shadow enhancers and investigated the genome-wide distribution and presence in *Drosophila*[50]. This subset of enhancer research contributed to the idea that two, or even more, enhancers control the expression of target genes, as opposed to a singular enhancer[51,52].

Super enhancers are another type of enhancer that have emerged from studying epigenetics. In 2013, Whyte and colleagues suggested the presences of super enhancers, which they defined as clusters of enhancers that are densely occupied by master regulator TFs and Mediator[53]. Using genome-wide data for 18 TFs, histone modifications, chromatin regulators, and chromatin accessibility, the authors found that super enhancers regulated key cell-type specific genes and functionally validated candidate enhancers using luciferase reporter assays. The identification of super enhancers led to the dissection of super enhancer components to address the questions of cooperativity and hierarchy of control[43,54,55]. These studies, among others, have raised questions of the actual difference between typical and super enhancers, which has been the subject of research and discussion in the field.

It is important to note that the majority of enhancers studied in the literature and annotated by various symposia are merely predictions. These predictions are made using a variety of programs and assumptions. Therefore, there has been a recent push to functionally validate predicted enhancer regions. However, this brings about another set of problems. Functional validation of enhancers can be investigated using a variety of methods. One method is to construct reporters, such as luciferase or *lacZ*. This provides a visual result of whether the enhancer influenced transcription of the target gene and luciferase gene, which was the goal of many studies. However, this methodology cannot be done in a high-throughput manner and it removes the regulatory region and gene from their native chromatin context. These caveats have spurred the development of high-throughput or parallel enhancer screening assays to functionally validate enhancers on a genome-wide scale[56-58]. Although these recently technologies have improved the scalability of functional enhancer testing, there is still much to be understood about the way enhancer regulate the transcription of their target genes.

Chromatin interactions: bringing enhancers and promoters together

Dramatic advances have been made in understanding enhancer biology- from identification to function to validation. However, a missing piece of information was how enhancers come in contact with their target gene's promoters to influence transcription. Thus, the field began investigating chromatin folding and the high order structure of chromatin that facilitates proximity of enhancers and promoters.

This area of study began with the identification of chromatin compartments (**Figure 1.6**). The genome is divided into compartment A and B, each with unique characteristics. Compartment A is comprised of genes that are transcribed, marked by active histone marks. In

contrast, compartment B contains inactive genes that are marked by repressive marks. Lieberman-Aiden and colleagues first described this distinction through the use of their novel technique HiC[59]. This finding instigated the further dissection of chromatin architecture and organization. In 2012, Dixon and colleagues defined topologically associated domains (TADs)[60]. TADs are defined by CTCF boundaries, contain regions of the genome that interact preferentially with one another, and include transcribed genes (**Figure 1.6**). The characterization of TADs was feasible through the generation and analysis of higher resolution HiC data.

CTCF, or CCCTC-binding factor, is a sequence specific zinc-finger TF that is highly conserved across vertebrate species[61]. Extensive research has been conducted to understand CTCF's unique characteristics. Interestingly, the direction of CTCF consensus binding motifs have been implicated in CTCF's function[62,63]. Furthermore, the knockdown of CTCF does not affect domain boundaries, but does increase inter-domain contacts that would not ordinarily occur[64]. CTCF can be found at regions of the genome with other structural protein complexes that are important for transcriptional regulation. Indeed, CTCF-defined domains that are correlated with transcriptional activation have been shown to strengthen during development[65]. CTCF has been shown to co-localize frequently with cohesin[66-68], and together they have been implicated in the loop extrusion model of transcription[62,69,70], which incorporates the theories of chromatin scanning and chromatin looping[71-73]. The loss of cohesin has been shown to effectively remove loop domains, while keeping compartment domains intact[74]. Interestingly, the loss of loop domains does not lead to widespread ectopic gene activation, but does affect some active genes.

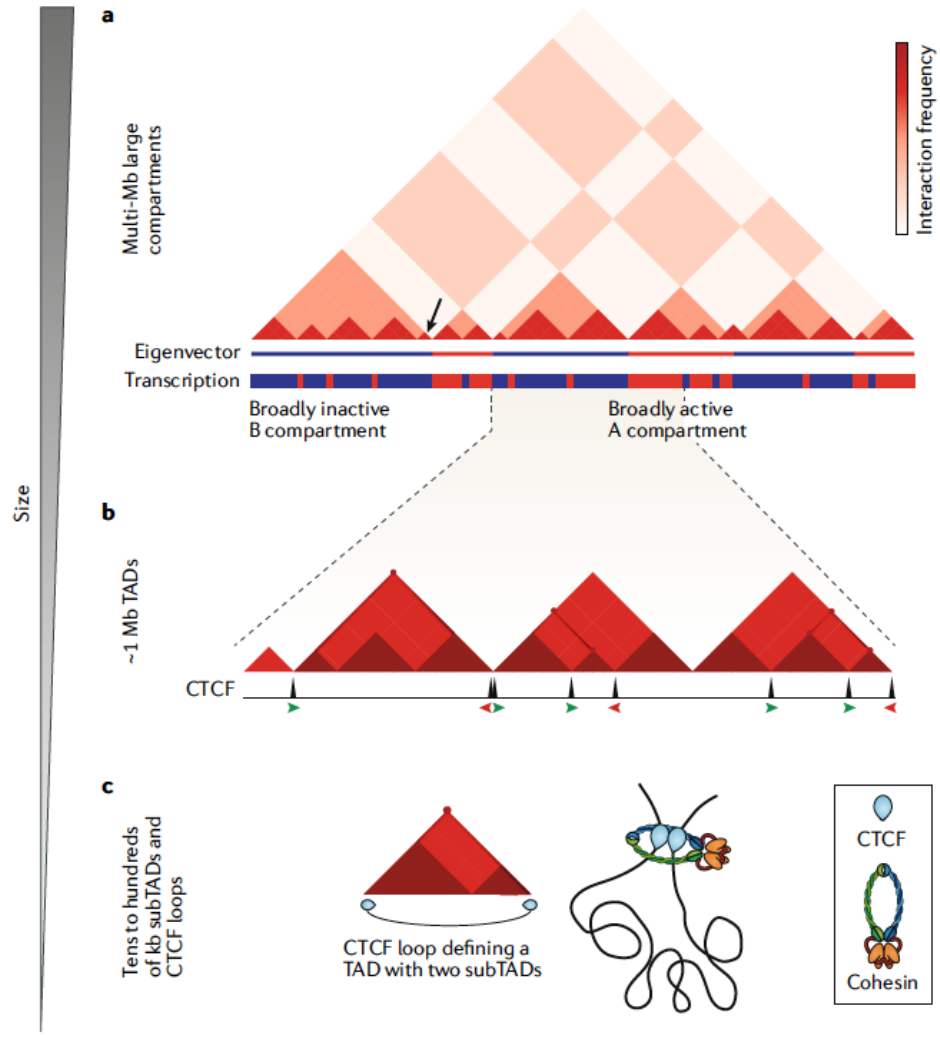

Figure 1.6. The current model of chromatin organization. (**A**) Compartments A (red) and B (blue) mostly correlate with transcriptionally active and inactive regions of the genome, respectively. These compartments are identified using HiC data. (**B**) TADs are smaller regions of the genome using higher resolution HiC data. TADs contain smaller subTADs characterized by higher interaction frequencies. The orientation of CTCF sites are denoted by the red and green arrows. (**C**) The structure of a TAD, containing two subTADs and flanked by CTCF-cohesin sites forming a chromatin loop. [Figure adapted from Rowley and Corces (2018) Nat Gen Rev[75] and used in accordance with Copyright Clearance Center's RightsLink service.]

Cohesin can also be found co-localized with Mediator, a multiprotein complex that acts as a transcriptional activator by interacting with Pol II and TFs[76,77]. This interaction physically and functionally connects components necessary for transcriptional regulation of genes through enhancers.

As technology has advanced and sequencing costs decrease, the field has been able to investigate the chromatin architecture in more detail. High resolution HiC enabled the definition of subTADs, known more frequently as insulated neighborhoods (**Figure 1.6**)[78,79]. These are characterized by a higher interaction frequency between chromatin regions within the larger TAD. The insulating effect of CTCF was studied explicitly at the α-globin locus, showing that the deletion of a conserved CTCF-cohesin boundary extends the subTAD, allowing the α-globin enhancers to upregulate genes outside of the original domain[80].

Within subTADs are specific enhancer-promoter (E-P) interactions. These chromatin regions are brought together through loop extrusion and create a chromatin interactome that is cell-type specific[81,82]. A variety of chromosome conformation capture assays have been employed to study the chromatin organization that facilitates these regions interacting. An important result from these studies revealed that E-P interactions were not limited to a one-to-one interaction[83-85]. In fact, the majority of enhancers were associated with several genes and vice versa. Research has also been conducted to understand the arrangement and complexity of these multiple interactions. Schoenfelder and colleagues found that co-regulated genes were spatial organized nonrandomly into interaction networks correlated with their biological function and expression level[86].

Recently, YY1 has been the subject of study in the context of E-P interactions. YY1 is a ubiquitously expressed zinc-finger TF that plays an important role in cellular differentiation[87,88].

It has also been found to be more specifically involved in chromatin architecture around E-P loops, as opposed to the larger loops established by CTCF[88,89]. When the YY1 binding motif was deleted in a locus-specific manner using CRISPR-Cas9, Weintraub and colleagues found there was decreased YY1 binding at the promoter, reduced contact frequency between the enhancer and promoter, and a decrease in mRNA levels[89]. These findings support the essential role of YY1 in controlling gene expression through facilitating E-P interactions. YY1 binding locations in erythroid cells have yet to be determined, resulting in a gap of important data for understanding E-P interactions during erythropoiesis.

The stability of the chromatin architecture in relation to transcriptional regulation has also been investigated. Jin and colleagues used HiC data at 5-10kb resolution to establish general principles of chromatin organization[82]. Most interestingly, this study revealed that chromatin loops connecting enhancers and promoters, as well as the larger chromatin context, were pre-established, suggesting that the chromatin interactome is established before exposure to stimuli that would include cell-type specific signaling and the activation of cell-specific transcriptional programs. This finding has been reproduced in a variety of cellular contexts, including glucocorticoid stimulation[90], heat shock[91], and serum-induced differentiation of mouse embryonic stem cells (ESCs)[92]. However, the impact of the hormone Epo on the chromatin interactome has yet to be investigated.

Epigenetics and disease

Transcriptional regulation is responsible for proper gene expression. Many diseases stem from transcriptional dysregulation[93]. For example, mutations in key TFs can cause cancer, such as the oncogenic TAL1, which is overexpressed in half T cell acute lymphoblastic leukemia

cases. More broadly, most malignant tumors rely on MYC, an overexpressed oncogene, for uncontrolled growth and proliferation. Chromatin modifiers, as discussed previously, are responsible for rearranging the accessibility of DNA and have been implicated in a variety of cancer types. Mutations that affect TF binding and nucleosome positioning therefore impact gene expression in disease, such as cancer, developmental disorders, diabetes, and others.

Since enhancers are intricately involved in the regulation of gene expression, one can imagine that dysregulation of genes, such as in the diseased context discussed above, could be linked to mutations in enhancer regions. The β-globin LCR is one of the most extensively and detailed studied enhancers (See *The β-globin locus control region* section). It has been linked to several hemoglobinopathies and has been leveraged to develop treatments for such diseases[94,95].

A variety of diseases have been attributed to misregulation of other epigenetic mechanisms. For example, Cornelia de Lange syndrome has been attributed to a mutation in the cohesin complex, potentially resulting in faulty E-P interactions[96]. Additionally, Kabuki syndrome is thought to result from a mutation in MLL4, a gene that functions as a histone methyltransferase, impacting the methylation of nucleosomes and therefore proper gene expression[97].

Erythropoiesis: an ideal developmental system to study gene regulation

Erythropoiesis

Erythropoiesis is the process by which the trillions of red blood cells (RBCs) that circulate the body are replenished. RBCs serve to deliver oxygen to and remove carbon dioxide from organs and tissues. There are three developmental waves of erythropoiesis during embryonic development marked by the presence of specific cell types- primitive erythroblasts

during the first wave, erythroid-myeloid progenitors during the second wave, and hematopoietic stem cells (HSCs) during the third wave[98]. After birth, HSCs in the bone marrow mature to erythrocytes through a series of differentiation stages. The first committed cell types in the erythroid lineage are the burst-forming unit erythroid (BFU-E) and the colony-forming unit erythroid (CFU-E) cells. Their morphology is indistinct from other early stage cells from other hematopoietic lineages. The first distinct erythroid progenitor is the proerythroblast. In the presence of erythropoietin, proerythroblasts will differentiate into basophilic, polychromatophilic, and orthochromatic erythroblasts, enucleate to become a reticulocyte, and finally a terminally differentiated RBC[98]. The process of differentiation is marked by a reduction in cell size and shift in cell surface markers. Different methods, such as flow cytometry, can be used to distinguish and separate cells in the successive stages of differentiation by leveraging the unique membrane proteins expressed[99].

Erythropoietin and its receptor

The hormone erythropoietin (Epo) is produced in low levels in the kidney[100]. Koury and colleagues conducted extensive work showing that Epo is necessary and sufficient for terminal erythroid differentiation building on work by Friend and others[101-103]. In 1957, Charlotte Friend isolated a virus that resulted in rapid erythroblastosis, enlargement of both the liver and spleen, and anemia when injected in mice[104,105]. The Friend virus is a retrovirus with a truncated form of the stem-cell kinase receptor, which is found in only certain mouse strains and confers susceptibility to the virus. As the Friend virus was studied in various laboratories, scientists found that certain strains caused polycythemia instead of anemia, leading to the classification of viral strains as anemia-inducing (FVA) or polycythemia-inducing (FVP)[106]. The FVA and FVP

strains have different forms of the pg55 envelop glycoprotein, resulting in the varied cellular response. Friend's work built a foundation for future viral studies, becoming a predecessor for isolating the human immunodeficiency virus (HIV) and the idea on the oncovirus.

In 1980, Hankins and Troxler developed an *in vitro* system to study erythropoiesis and compare FVA and FVP, as well as other strains[107]. Mainly, their study investigated the ability to form BFU-E, finding that bone marrow cells injected with FVA produced BFU-E in the absence of Epo, but that had significantly less hemoglobin compared to cells injected with FVP[107]. Koury et al continued these *in vitro* studies by investigating the role of Epo in FVA-infected cell proliferation. They found that progenitor cells proliferated to the same extend with or without the addition of Epo. However, Epo was necessary for complete differentiation of the FVA-infected bone marrow cells that are the source of erythroid progenitors[108]. This established that Epo is not required for proliferation, but is essential for differentiation.

Koury et al then developed an in vivo system to study FVA-infected splenic erythroblasts and the effect of Epo on this population of cells[109]. This methodology will be described in detail below (See *Large-scale procurement of erythropoietin-responsive erythroid cells*). Using the FVA system, Koury and Bondurant were able to investigate the Epo requirements of erythroid cells for viability, proliferation, and differentiation. In this study, they found that there is a minimum concentration of Epo required for proliferation and maturation, but an increased concentration does not linearly increase these measures[101]. They then tested the time requirements of Epo in cell culture, showing that increased delay in adding Epo to culture resulted in a decrease in cell viability and developmental capacity and that withdrawal of Epo also significantly affected cell viability[101]. Another study specifically investigating the role of Epo in protecting DNA breakdown and cell death quickly followed. In 1990, Koury and

Bondurant published a study that revealed Epo slows the DNA cleavage that occurs in late erythroid progenitor stages, effectively steering erythroid cells away from programmed cell death[102].

Epo stimulates RBC production by binding its receptor, the erythropoietin receptor (EpoR). EpoR is expressed on several cell types, such as brain[110], heart[111], and fat tissue[112], but is predominantly on the surface of erythroid precursors[113]. EpoR is a type I cytokine receptor and in a dimer form when not bound by Epo[100]. When Epo binds the extracellular portion of the receptor, dimerization occurs and associated intracellular Janus Kinase (JAK) molecules phosphorylate each other. Activated JAK molecules phosphorylate tyrosine kinases on the receptor, creating binding sites for the Signal Transducer and Activators of Transcription (STAT) molecules. JAK then phosphorylates STAT, causing it to disassociate from the receptor and dimerize with another STAT molecule. This facilitates the translocation of STAT into the nucleus of the cell. The JAK-STAT signaling pathway is essential for erythropoiesis and is initiated by the hormone Epo, which is necessary and sufficient for terminal erythroid differentiation.

The master regulators of erythropoiesis

Erythropoiesis is primarily regulated by three transcription factors known as the master regulators. Gata1, Klf1, and Tal1 are lineage restricted, meaning they are expressed in erythroid cells, and are essential for cell survival (**Figure 1.7**)[114,115]. Gata binding protein 1 (Gata1) was the first master regulator identified and binds the sequence motif WGATAR. It is a transcription factor forms by two zinc fingers and its N-terminal domain interacts with other transcription factors, co-activators, and co-repressors[114,115]. It activates both early and late erythroid genes,

such as the EpoR and hemoglobin genes respectively. Weiss and Orkin conducted the first experiments characterizing and investigating the role of Gata1 in erythroid differentiation[116,117], including the development of a Gata1-null erythroid cell line that can be induced to terminally differentiate when Gata1 is reexpressed[118]. Their respective labs have gone on to further the basic and clinical knowledge related to Gata1 function in erythroid cells.

Erythroid Kruppel-like factor (Klf1) is the founding member of the Klf family, of which its 17 members are expressed and function in a variety of cell and tissue types. Klf1 is exclusively expressed in the erythroid lineage was discovered in 1993 by Miller and Bieker[119]. Its structure is comprised of three zinc fingers at the C-terminal, which is shared across the Klf family, and an N-terminal transactivation domain that is specific to Klf1[114]. Klf1 binds NCNCNCCCN, a motif found in the adult β-globin promoter. This, among other studies, has implicated Klf1 as an important regulator of the switch from fetal to adult globin expression[114,115,120].

Tal1/Scl was discovered through its role in the chromosomal translocation associated with T-cell acute lymphoblastic leukemia[114,115]. It is in the basic helix-loop-helix (bHLH) transcription factor family and requires heterodimerization with E-box proteins, another member of the bHLH family, for proper DNA binding[114,115]. The E-box binding motif of CANNTG is found in regulatory regions with Gata binding motifs, suggesting these transcriptional factors are involved together in regulating erythropoiesis. Research in the field has proposed a model where Tal1/E-box and Gata1 are bridged by a complex formed by two LIM domain containing TFs, Lmo2 and Ldb1[115,121].

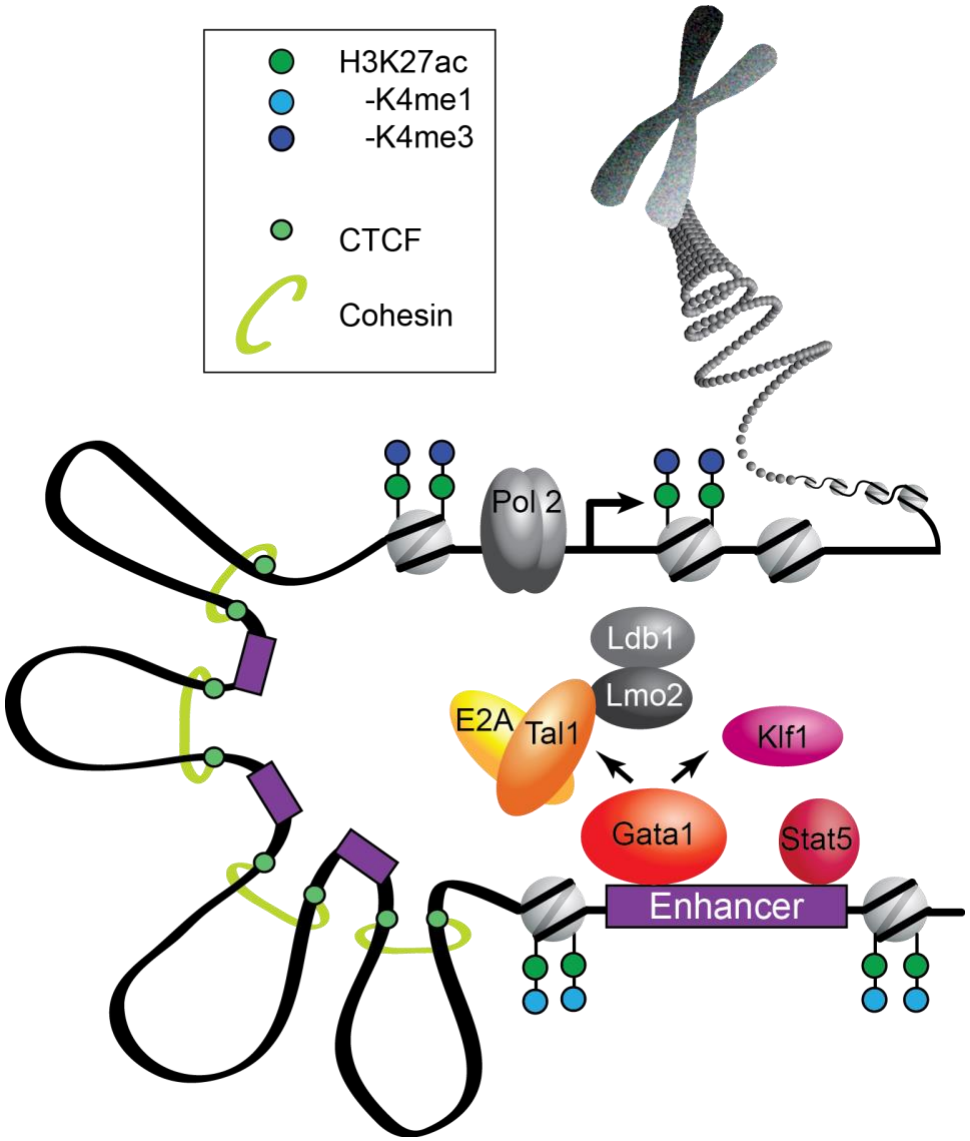

Figure 1.7. Integrative model illustrating how TFs, enhancers, and chromatin looping work together to recruit Pol II to promoters during erythropoiesis. The erythroid master regulators predominantly bind to enhancers in conjunction with accessory proteins. TFs and enhancers interact within the confines of CTCF and cohesin defined domains. [Figure modified from Perreault and Venters (2018) Curr Opin Hematol[121] and used in accordance with Copyright Clearance Center's RightsLink service]

The β-globin locus control region

A unique characteristic of erythroid cells is the expression of the α-like and β-like globin genes, which encode components that form the larger hemoglobin protein. The genes encoding the β-like subunit are conserved across several species[122]. The β-like globin gene cluster is organized in the same order in which the genes are expressed during development- ε, γ, δ, β[123]. However, not all genes are expressed at the same time, undergoing a switch between embryonic, fetal, and adult stages of development[124,125]. During the first trimester of human embryonic development, the ε-globin is highly expressed in the primitive erythroid cells in the yolk sac. Once definitive erythrocytes develop from stem cells in the fetal liver, the γ-globin is primarily expressed. γ-globin is encoded by two genes that are found within the globin gene cluster and is a subunit of fetal hemoglobin. Shortly after birth, there is a switch from fetal to adult hemoglobin, which is mediated by a switch from γ- to β-globin transcription in erythroid cells[124]. This switch has been extensively studied in the erythroid field and is an example of fine-tuned regulation of gene expression through a locus control region (**Figure 1.8**).

In 1987, Grosveld and colleagues showed position-independent regulation of the β-globin gene in transgenic mice[126]. This finding was supported by subsequent research, establishing an important characteristic that LCRs and enhancers regulate transcription of target genes in a position- and orientation-independent manner[127]. The LCR has been systematically dissected, resulting in the discovery of five critical DNase-I hypersensitive sites. These hypersensitive sites (HSs) were tested in transgenic mice, revealing that the sites had varying effects on β-globin gene expression when individually deleted[123].

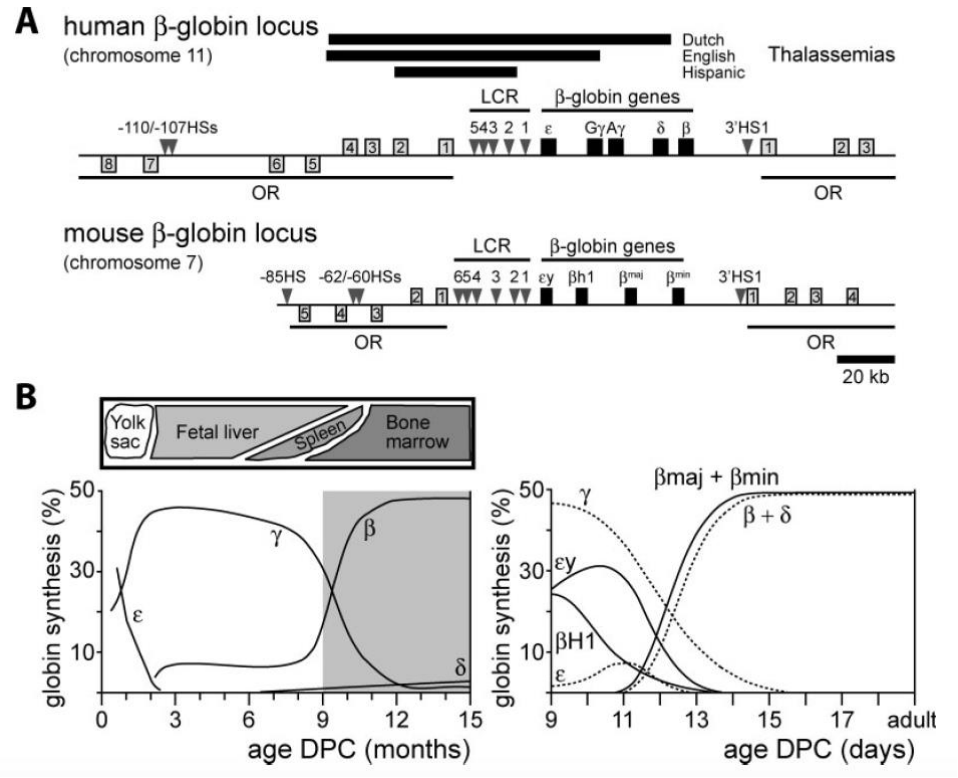

Figure 1.8. The erythroid-specific beta-globin LCR. (**A**) The human and mouse beta-globin locus. (**B**) Globin protein expression during development in human (left) and mouse (right). [Figure adapted from Noordermeer and de Laat (2008) IUBMB Life[125] and used in accordance with Copyright Clearance Center's RightsLink service.]

Grosveld and colleagues also investigated the orientation of globin genes relative to the LCR, finding that the developmental expression pattern of the genes is dependent upon their position relative to the LCR[123]. The Groudine lab also studied the HSs using chicken/human hybrid cell lines and supported these findings with data that shows the region between HS2 and HS5 is essential for adult β-globin gene expression[128]. This study also found that the region between HS2 and HS4 is essential for embryonic and fetal globin gene expression[128]. Investigation of binding motifs in the HSs shows Gata1, Klf1, Tal1/E-box motifs, as well as others, indicating the importance of the master regulators in the regulation of globin gene expression through the LCR[120]. The evolutionary conservation of the β-globin LCR has been

studied using the DNA sequences of human, rabbit, goat, and mouse LCRs. Hardison and colleagues found that there are several LCR characteristics that are shared across these species, namely the number and order of HSs and the sequence within the LCR[122].

Large-scale procurement of erythropoietin-responsive erythroid cells

Koury and colleagues extensively studied the role of Epo on erythroid progenitors using both in vitro and in vivo systems. Importantly, they developed a method to isolate a large population of pure, stage-synchronous proerythroblasts[109,129]. BALB/c female mice 8-12 weeks' old are injected with 10^4 spleen focus-forming units of FVA. Approximately 2 weeks (13 to 15 days) later, mice are sacrificed and the spleens harvested. Spleens are homogenized into a single cell suspension and cells are separated using sedimentation in a bovine serum albumin (BSA) gradient. Proerythroblasts isolated during sedimentation are then cultured in an ex vivo system, importantly adding Epo to culture medium. In developing this system, Koury, Sawyer, and Bondurant measured several important characteristics of these cultured cells. They found that over 95% of the cells isolated from FVA-infected spleens were erythroblasts[109]. Cell morphology, appearance, benzidine staining, and heme synthesis kinetics were measured for a time course of FVA cells in culture. Cells cultured in Epo became smaller and nuclei condensed. Eventually cells became enucleated, marking differentiation into a reticulocyte[109]. Benzidine staining assesses the amount of hemoglobin in a cell, which is indicative of a fully terminated erythroid cells because RBCs express hemoglobin, which facilitates oxygen transport through the body. Heme is an iron-containing compound, which forms the nonprotein part of hemoglobin. Heme synthesis is critical for hemoglobin function. Using these measures, Koury and colleagues

were able to show that a large population of pure proerythroblasts isolated from spleens of FVA-infected mice can fully differentiate in culture in the presence of Epo[109,129].

High-throughput sequencing and the development of –seq assays

First generation sequencing

Fred Sanger and colleagues were instrumental in bringing sequencing research to the forefront of science. Sanger's work established the amino acid sequence of insulin, a protein secreted by the pancreas[130]. This project was followed by sequencing ribosomal RNA from Escherichia coli[131,132]. Finally, Sanger and his lab approached the challenge of sequencing DNA using DNA polymerase with radiolabeled nucleotides that he called the "Plus and Minus" technique[133]. Using this methodology, which could sequence 80 nucleotides at a time, the Sanger lab sequenced the genome of the bacteriophage ΦX174. In 1977, Sanger and colleagues published a new technique that would become known as the Sanger method or Sanger sequencing[134,135]. This chain-termination method requires a single-stranded DNA template, a DNA primer, DNA polymerase, normal deoxynucleotidetriphosphates (dNTPs), and modified di-deoxynucleotidetriphosphates (ddNTPs). This technique was used to sequence human mitochondrial DNA[136], bacteriophage λ[137], and ultimately the human genome[138-140]. The Human Genome Project was proposed by the United States government in 1984 and was a collaborative scientific research project with the goal of determining the base pairs that make up human DNA. The program officially began in 1990 and with the help of advances in sequencing and computing technologies, the project was completed in 2013[141]. This landmark project initiated a burst of sequence-based research, which resulted in faster and cheaper techniques, known collectively as next generation sequencing.

Next generation sequencing to identify protein-DNA interactions

Chromatin immunoprecipitation is an assay that aims to identify protein-DNA interactions. It has been most widely applied to finding transcription factor binding locations and regions of the genome with histone modifications. To accomplish this, DNA interactions are frozen by crosslinking the cells. The DNA-protein complexes are sheared to between 200 and 500 basepairs using sonication or restriction enzyme digestion. DNA fragments associated with the DNA-protein complexes are immunoprecipitated, or selected for, using an antibody specific to the protein of interest. These fragments are then purified and sequenced. The resulting fragments represent regions where the protein of interest interacts with DNA, such as cell-type specific transcription factors binding to promoters of genes. This technology was first published by Lis and Gilmour in 1984, where they investigated the RNA polymerase II (RNA Pol II) abundance at specific bacterial genes[142]. This was quickly followed by another paper by the Lis lab investigating Pol II at heat shock genes in Drosophila, finding that the abundance of Pol II associated with several heat shock genes increased dramatically in response to heat shock, indicating that the induction of heat shock gene expression by heat shock occur at the transcriptional level[143]. The ChIP technique has been modified and refined to investigate DNA-protein interactions at a genome-wide level.

ChIP-on-chip is a version of ChIP where the experiment is done on a DNA microarray, or chip. The first ChIP-chip experiment was conducted by Kleckner and Blat in 1999, where they determined the distribution of cohesin Mcd1/Sccl and Smc1 along yeast chromosome III[144]. Richard Young's lab published the first genome-wide use of ChIP-chip, establishing the binding

sites of 106 transcription factors in yeast[145,146]. This was followed by work from the Farnham lab, which published the first instance of ChIP-chip in mammalian cells[147].

The work conducted using ChIP-chip was quickly followed by ChIP-seq[148], which was developed in the wake of enhanced sequencing technologies. Oligonucleotide adaptors are added to the size-selected fragments of DNA that were bound to the protein of interest, facilitating sequencing of these fragments. This technology was developed concordantly by three separate labs, which published findings on histone methylation[149], transcription factor binding[150], and chromatin states[151] in human and mouse. Since 2007, there have been over 4,000 scientific publications that use ChIP-seq, with over 400 of those in 2019 alone.

With ChIP-seq becoming a commonplace experiment, scientists aimed to refine this technology. ChIP-exo was developed in the Pugh lab in 2011 and uses lambda exonuclease digestion to achieve basepair resolution of protein-DNA interactions (**Figure 1.9**)[152,153]. Rhee and Pugh used the ChIP-exo methodology to identify genome-wide binding of yeast transcription factors Reb1, Gal4, Phd1, Rap1, and human CTCF. The development of CHIP-exo was followed by ChIP-nexus[154], ChIP-tag-exo[155], ChIP-SSL-exo[155], ChIP-exo 5.0[155], and Cleavage Under Targets and Release Using Nuclease (CUT&RUN)[156]. The development of genome-wide sequencing assays will continue as sequencing becomes less expensive and more efficient.

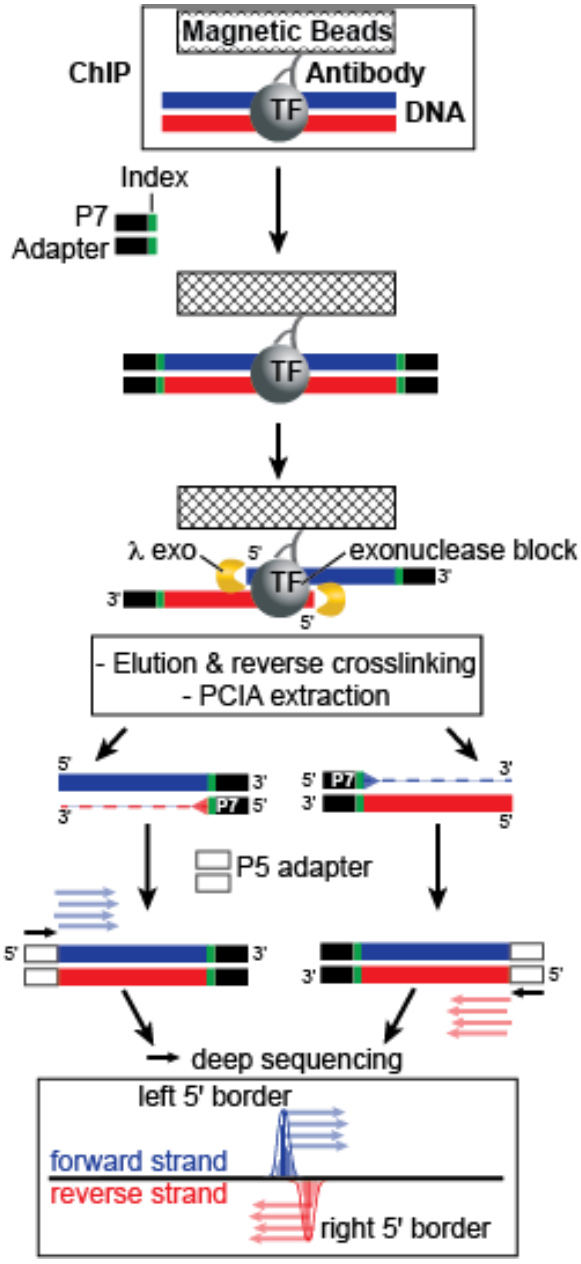

Figure 1.9. Schematic for the ChIP-exo workflow. After ChIP, the P7 adapter is ligated to the sonication borders. Lambda exonuclease then trims DNA 5' to 3' to the crosslink point, thereby footprinting the protein-DNA interaction. After elution and crosslink reversal, primer extension synthesizes duplex DNA. Lastly, ligation of the P5 adapter marks the left and right exonuclease borders and the resulting library is subjected to high-throughput sequencing. Mapping the 5' ends of the sequence tags to the reference genome demarcates the exonuclease barrier and thus the precise site of protein-DNA crosslinking. [Figure modified from Perreault and Venters (2016) J Vis Exp[153] and used in accordance with Copyright, *Journal of Visualized Experiments*]

Next generation sequencing to investigate transcription and gene expression

To complement the NGS assays used to investigate protein-DNA interactions, scientists developed RNA-seq to study the quantity and sequence of RNA molecules in a sample. RNA-seq is key to connecting the information on our genome, which can be studied using ChIP-based assays, with its protein expression. It also provides information that is not encoded in the DNA, such as post-translational modifications, alternative splicing, and gene function.

The earliest RNA-seq libraries leveraged Sanger sequencing. However, these were inefficient and usually inaccurate. In 2008, the Wold[157], Snyder[158], and Ecker[159] labs all published research using RNA-seq technology. This research studied transcripts in adult mouse brain, liver and skeletal muscle tissues[157], in the yeast genome[158], and in *Arabidopsis*[159], respectively. The development of RNA-seq technologies has increased knowledge of non-coding regions of the genome, transcriptional structures of genes, and the changes in gene expression that occur during development and different conditions.

Although RNA-seq has been instrumental in gaining deeper understanding of the transcriptome, it measures bulk RNA populations that are stable enough to be converted to cDNA. Due to the broad nature of the assay, it does not investigate native transcripts. To address this gap, global run-on sequencing (GRO-seq) was developed by the Lis lab. GRO-seq maps the position, amount, and orientation of transcriptionally engaged RNA polymerases genome-wide[160]. This research provided detailed information about the orientation of polymerase relative to its associated promoter and mechanisms of gene regulation. This assay was followed by the development of precision nuclear run-on sequencing (PRO-seq) to map the genome-wide distribution of transcriptionally engaged Pol II at base pair resolution[161]. These revolutionary

methodologies have allowed for the explicit study of nascent RNA, advancing the understanding of transcriptional dynamics and gene regulation at a basepair resolution.

Next generation sequencing to establish chromatin contacts

Establishing protein-DNA interactions and transcriptome data necessitated the need to also understand the chromatin architecture of DNA. The first foray into investigating DNA-DNA interactions was through the chromosome conformation capture (3C) assay developed in the Kleckner lab[162]. Using this method, one can identify the interaction frequency of two specific DNA segments in the genome. Dekker et al employed 3C to determine the chromosomal structure of yeast chromosomes (**Figure 1.10**)[162]. There are several limitations to this method, most importantly that it can only identify the frequency of interaction between two pre-determined regions. Just as with the ChIP methodology, the 3C technique has many derivatives to address limitations.

In 2006, three 4C methodologies were published concordantly that extended the on-to-one interactions found with 3C to one-to-all interaction identification by using a "bait" segment of DNA (**Figure 1.10**). Chromosome conformation capture on ChIP developed by the de Laat lab leverages ChIP ideology and was used to study both active and inactive gene chromosomal interactions[163]. Wurtele and Chartrand developed an open-ended 3C technique and investigated the dynamics of the spatial environment of the HoxB1 gene before and after the induction of its expression in mouse embryonic stem cells[164]. Circular chromosome conformation capture developed by Zhao et al creates circularized DNA and was first used to identify 114 unique sequences, several of which interact primarily with the maternally inherited H19 imprinting

control region[165]. Two limitations with these 4C assays are sequencing depth and the cis to trans ratio of interactions.

The use of 3C and 4C techniques increased the information known about specific regions' interactions, whether it was with another single region or an abundance of regions. However, these methods did not address the many-to-many interactions that could occur, such as the rosette structure that can form when multiple enhancers come in contact with a target gene promoter to regulate transcription. This gap was filled with the chromosome conformation capture carbon copy assay developed by the Dekker lab. The 5C methodology begins much like a 3C experiment, except there can be up to several hundred primers designed to probe potential interactions within a large genomic region (**Figure 1.10**). Dostie et al employed the 5C technique to confirm chromatin interactions at the β-globin locus control region, as well as identified new looping interaction in K562 cells between it and the gamma-beta-globin intergenic region[166]. Sanyal et al employed the 5C method to identify thousands of long-range interactions between promoters and enhancer, promoters, and CTCF-bound sites in GM12878, K562 and HeLa-S3 cell lines[84]. Although the 5C method allows for the investigation of complex multiple interactions, it necessitates identifying a region of interest and it is not genome-wide.

The next advancement in 3C methodology addressed the issue of finding genome-wide interactions. HiC again begins like a classic 3C experiment, except the post-digested fragments are labeled with biotin. These fragments are then isolated using streptavidin, identifying chromosomal interactions genome-wide in an unbiased way. This allows for the investigation of all-to-all interactions. Lieberman-Aiden et al. developed and used the HiC assay to investigate chromosomal structure, confirming the presence of chromosome territories and the spatial proximity of small, gene-rich chromosomes in the human genome (**Figure 1.10**)[59].

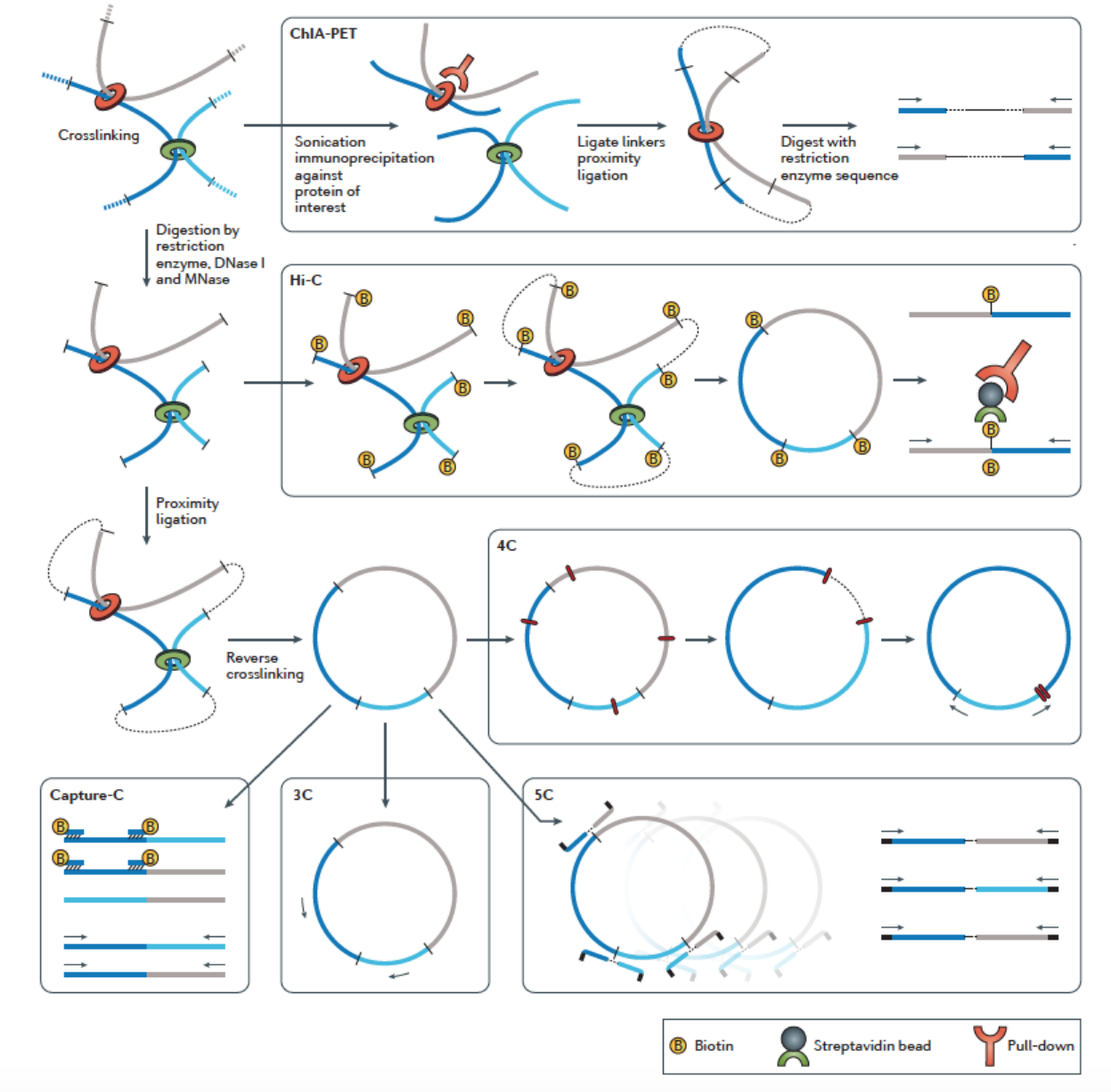

Figure 1.10. 3C-based approaches to study chromatin architecture. The first step of most 3C-based methods involves the formaldehyde crosslinking of cells. In most downstream protocols this is followed by fragmentation of the chromatin by digestion with a restriction enzyme or by sonication. In ChIA-PET, the next steps involve enrichment for interactions mediated by a protein of interest by immunoprecipitation, ligation of adaptors to the restriction fragment ends followed by proximity ligation, fragmentation by restriction enzyme digestion, isolation of paired-end tags containing adaptors and paired-end sequencing. In standard 3C-based protocols the digestion by restriction enzymes is then followed by proximity-based ligation of adjacent DNA ends and determination of pair-wise interactions. In the 4C protocol a second round of digestion and ligation is used to increase resolution, followed by inverse PCR with locus-specific primers to detect genome-wide interactions involving the locus of interest. In the 5C approach, primer sequences overlapping restriction fragment ends are ligated only when the

two ends are immediately adjacent. In Capture-C methodology, enrichment for interacting pairs is accomplished using biotin-labelled probes complementary to restriction fragment ends of interest. In the HiC method the restriction fragment ends are labelled using biotin, ligated products are enriched using streptavidin pull-down after sonication and interactions are interrogated in a genome-wide all-versus-all unbiased manner. [Figure adapted from Bonev and Cavalli (2016) Nat Gen Rev[167] and used in accordance with Copyright Clearance Center's RightsLink service.]

The sequencing depth and number of biological replicates influence the resolution of the interaction frequency maps. The resolution needed depends highly on the biological question at hand.

The above discussed techniques focus on DNA-DNA interactions. Going back to the investigation of protein-DNA interactions, researchers developed techniques to study chromatin structure mediated by specific proteins. Chromatin interaction analysis using paired-end tag sequencing (ChIA-PET) developed by Fullwood and Ruan uses a linker sequence to connect DNA fragments that are in contact with each other by protein factors (**Figure 1.10**)[168,169]. This methodology was used to map the chromatin interaction network bound by estrogen receptor alpha (ER-alpha) in the human genome, finding that ER-alpha-binding sites are anchored at gene promoters through long-range chromatin interactions[85].

Although ChIA-PET refines the all-to-all interactions identified with HiC by focusing on proteins that connect these DNA interactions, it still requires hundreds of millions of cells. Thus, HiChIP was developed by Mumbach et al. to identify chromatin contacts mediated by a protein of interest with as few as one million cells[170]. HiChIP effectively reveals chromatin contacts with lower cell number and greater signal-to-noise (**Figure 1.11**). This technique was used to generate H3K27ac HiChIP contact maps, which revealed interactions between active enhancers and target genes in rare primary human T cell subtypes and coronary artery smooth muscle cells[171].

The findings discussed above have identified regulatory regions called enhancers, established mechanisms of regulatory control, and evaluated the chromatin architecture that contain enhancers and their target gene promoters. Specifically, the finding that up to 40% of enhancers do not control the transcription of the nearest neighboring gene confounded the ability to accurately establish EP interactions merely using 1D genome characteristics[83,84].

The findings that a gene can be associated with multiple enhancers has added an additional layer of complexity to enhancer biology[47,48,50,51]. Furthermore, studies found that enhancers can work through various mechanisms to regulate transcription in multiple model systems and cell types. These qualities pose a challenge in assigning functional significance to candidate enhancers and motivate a need for more precise and comprehensive understanding of enhancer relationships in the regulation of gene expression programs.

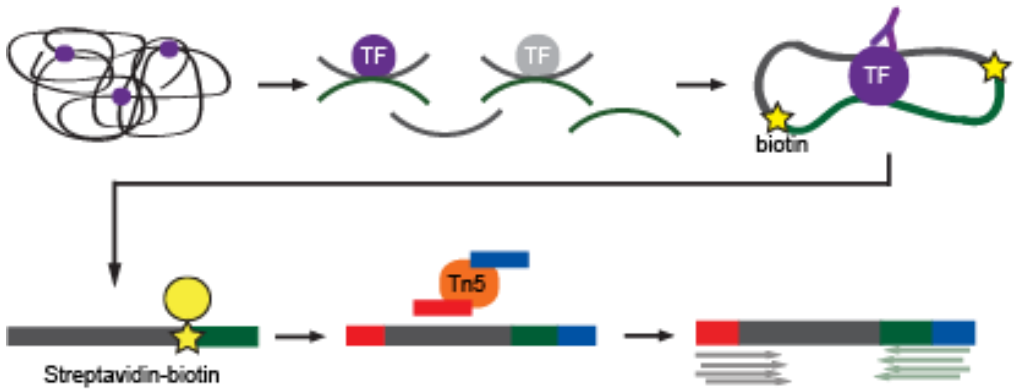

Figure 1.11. Schematic of HiChIP workflow. Long range DNA contacts are established in situ before lysis through formaldehyde crosslinking. Classic ChIP is performed, capturing the interactions associated with a protein of interest. Streptavidin beads are used to isolate fragments that have undergone proximity ligation. Paired-end sequencing identified two distant regions of the genome that are involved in the chromatin contact.

Specific aims of dissertation

The *objective* of this thesis was to understand the epigenetic regulation of gene expression during early erythroid differentiation compared to the erythroleukemic state. The *central hypothesis* is the erythropoietin (Epo) initiates epigenetic modifications genome-wide that alter transcriptional programs. This hypothesis is based on 1) findings by Koury et al. using the FVA system that reveal Epo is necessary and sufficient for terminal erythroid differentiation, which is accompanied by morphological changes and differential protein expression, 2) a body of work that identifies enhancer regions supports the gene regulatory impact of these regions on gene expression, and 3) recent literature that investigates the chromatin organization necessary to bring enhancers in genomic proximity to their target genes. To test the central hypothesis, the following specific aims were developed:

Aim 1: Establish enhancer locations in FVA cells

The *working hypothesis* was that enhancer regions in FVA proerythroblasts would be present and identifiable through specific chromatin marks, and would be associated with genes involved in erythroid physiology and cell survival. To test the hypothesis, I proposed to identify enhancers pre and post Epo stimulation in FVA-derived proerythroblasts and link these regulatory regions to the gene promoters they regulate.

Aim 2: Identify enhancer-promoter interactions in FVA cells

The *working hypothesis* was that the use of chromosome conformation capture assays would reveal chromatin interactions involved in regulation of key erythroid genes. Additionally, the use of higher order chromatin architecture data would refine the enhancer-promoter links previously identified. To test the hypothesis, I proposed to employ a novel chromosome

conformation capture assay to locate interacting regions of the genome in FVA-derived proerythroblasts pre and post Epo stimulation.

Aim 3: Investigate the epigenetic landscape of erythroleukemic cells

The *working hypothesis* was that profiling the epigenome at high resolution of a widely used cell line (K562) would provide important insights into epigenetic regulation of erythroid cells. To test the hypothesis, I proposed to compare ChIP-seq and ChIP-exo to investigate the important of high resolution data in identifying Pol II occupancy and the presence of histone variants in nucleosomes.

CHAPTER 2

EPO REPROGRAMS THE EPIGENOME OF ERYTHROID CELLS[1]

Introduction

Erythropoiesis has long served as an experimental paradigm for understanding gene regulatory mechanisms governing cell identity and development. The hormone erythropoietin (Epo) is physiologically necessary and sufficient to trigger the initiation of terminal erythroid differentiation of proerythroblasts. Epo initiates erythroid differentiation and gene expression patterns through binding to its cognate receptor, which activates the Jak2-Stat5 signaling axis[172-174]. Erythroid expression patterns are highly dynamic and have been extensively studied by genome-wide expression profiling[175-180], providing numerous insights into the molecular pathways that control red blood cell development.

Cell identity is established and propagated primarily through epigenetic mechanisms, including the cell-type specific repertoire of enhancers[12,13,181]. Recent reports tracking histone modifications described the erythroid enhancer landscape in human and murine erythroid cells[43,182,183]. Interestingly, despite the dramatic transcriptional changes that accompany erythropoiesis, previous work found that broad features of chromatin states remain largely unchanged during Gata1-induced differentiation in the murine G1E-ER4 cell line[182]. This study suggested that erythroid enhancers are established in erythroid precursor cells, but precisely when this occurs remains unclear. While the locations of erythroid enhancers have been determined in murine and human cell systems[43,182,183], how Epo influences the enhancer

[1] Portions have been published as Perreault et al, (2017). *Experimental Hematology*

landscape is currently unknown. Defining the comprehensive set of Epo-modulated enhancers will better illuminate the molecular and epigenetic pathways that are controlled by Epo.

To investigate the interplay of Epo signaling and the chromatin landscape during erythroid development, we used highly purified proerythroblasts derived from mice injected with the anemia-inducing strain of the Friend virus (FVA)[102]. FVA-derived proerythroblast proliferation depends on the truncated form of the stem cell-derived tyrosine kinase receptor (sf-Stk), but not Jak2-Stat5 activity[184,185]. Importantly, FVA proerythroblasts remain sensitive to physiological levels of Epo in an ex vivo culture system, and upon Epo stimulation will activate Jak2-Stat5 signaling and synchronously differentiate into fully mature, enucleated erythrocytes. Thus, FVA-derived proerythroblasts represents an ideal system to delineate the molecular action of Epo on a genomic scale and to examine to epigenetic effects of Epo in a physiological context.

Here, we report the use of chromatin immunoprecipitation-exonuclease analyses (ChIP-exo) to profile the enhancer landscape during Epo-initiated erythropoiesis. Genome-wide locations for H3K4me1, H3K27ac, and H3K4me3 were examined in the FVA model system at baseline and at 1 hour after Epo-initiated synchronous, terminal erythroid differentiation. This model cell system uniquely allowed us to define the immediate impact of Epo signaling on the epigenetic landscape during erythropoiesis.

Our work illustrates several new aspects for the role of Epo in erythroid chromatin biology. Remarkably, Epo stimulation alters the histone mark signatures across several thousand enhancer locations, highlighting an underappreciated role for Epo in reprogramming the epigenome. Enhancer marks were highly dynamic within the first hour of Epo stimulation, enabling classification of enhancer behavior based on the Epo-induced modulation of these histone marks. Upon Epo stimulation, a similar number of enhancers displayed increased

(n=1,589) and decreased (n=1,529) acetylation. For example, loss of enhancer acetylation was linked to genes known to be down regulated during erythropoiesis, such as CD44. Likewise, enhancers displaying increased acetylation levels in response to Epo were linked to genes involved in regulating the survival pathways, TFs, and chromatin regulators. Nevertheless, our finding that the vast majority of enhancers, including super enhancers, were unaffected within the first hour of Epo stimulation fits with the notion that chromatin states remain largely unchanged during erythropoiesis[182].

Bioinformatic analyses of motif enrichment and H3K27ac enhancer-promoter correlation validated many of the enhancer locations identified in this study. Enhancer overlap with published data further validated enhancer locations, and revealed the location of evolutionarily conserved cis-regulatory modules across mouse and human erythroid cell genomes. Given that Tal1 is one of the master TF regulators of erythropoiesis, we also show that Tal1 binds to thousands of enhancers, adding a layer of validation to our enhancer maps and suggesting a mechanistic link between Epo signaling, Tal1 occupancy, and chromatin dynamics. Together, these aspects of the presented study identify cis-regulatory enhancer circuits that are controlled by Epo signaling pathways. Importantly, the epigenetic profiles from the current study provide the framework for generating testable hypotheses for how Epo, enhancers, and TFs coordinately control erythroid expression programs.

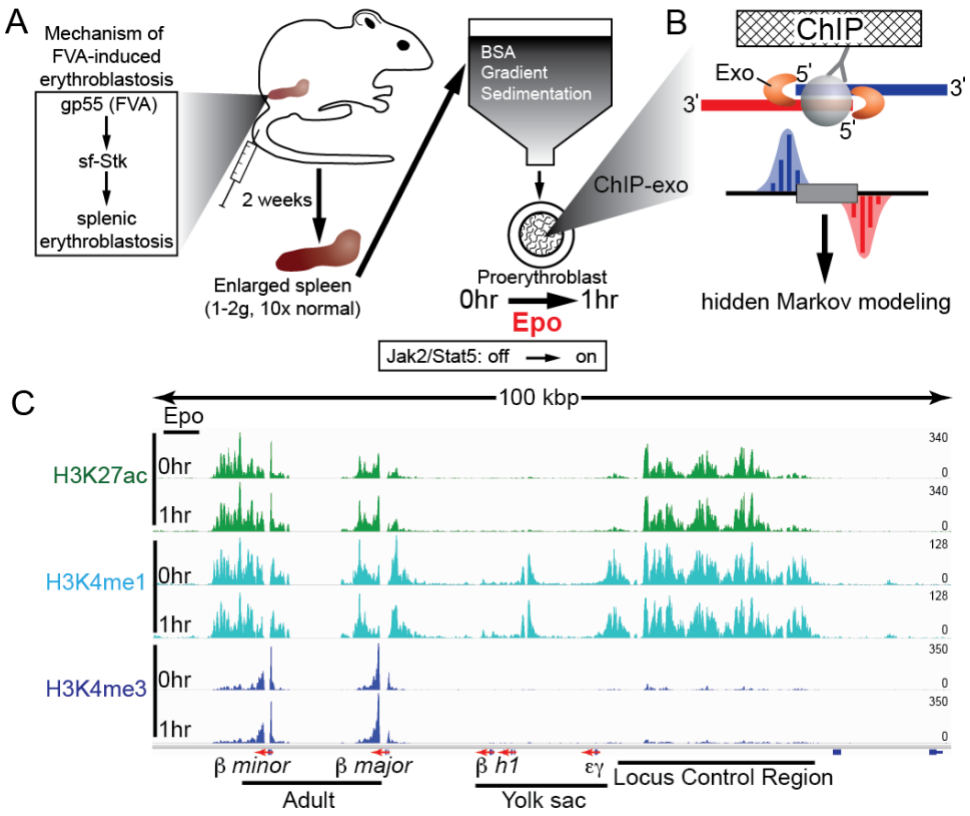

Figure 2.1. Experimental overview and distribution of histone marks at the β-globin locus control region. (**A**) Shown is the workflow for generating and isolating highly purified Epo-responsive proerythroblasts from a mouse injected with the Friend Virus that induces Anemia (FVA). (**B**) ChIP-exo cartoon illustrates how the 5′ to 3′ lambda exonuclease trims DNA to the crosslink point, effectively footprinting the protein-DNA interaction. Mapping the 5′ ends of the sequence tags to the reference genome demarcates the exonuclease barrier and thus the precise site of protein-DNA crosslinking. (**C**) Genome browser view of H3K4me1, H3K27ac, and H3K4me3 ChIP-exo signal in murine proerythroblasts shown at the □-globin (HBB) locus control region (LCR). [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Materials & methods

Isolation of murine proerythroblasts

Highly purified proerythroblasts were obtained from spleens of mice infected with the Friend virus as previously described[109,129], with the following modifications. All animal procedures were performed in compliance with and approval from the Vanderbilt Division of

Animal Care (DAC) and Institutional Animal Care and Use Committee (IACUC). Female BALB/cJ mice (12 weeks old, Jackson Laboratories) were infected via intraperitoneal injection of ~10^4 spleen focus-forming units of Anemia-inducing strain of the Friend virus (FVA). At 13 to 15 days' post-infection, the mice were sacrificed and spleens removed. The spleens were homogenized to a single cell suspension by passing the minced spleens through a sterile 100 micron nylon mesh filter into sterile solution of 0.2% bovine serum albumin (BSA) in 1x PBS. The filtrate was then repeatedly pipetted to ensure a single cell suspension. The homogenized spleen cells were size-separated by gravity sedimentation for 4 hours at 4°C in a continuous gradient of 1% to 2% deionized BSA. The sedimentation apparatus consisted of a 25cm diameter sedimentation chamber containing a 2.4L BSA gradient, two BSA gradient chambers containing 1.2L 1% and 2% deionized BSA in 1x PBS, and a cell loading chamber (ProScience Inc.) containing the 50ml cell suspension. After a 4 hour sedimentation, cells were collected in 50ml fractions, with proerythroblasts typically enriched in fractions 5-20 of 24 total fractions. Typically, about 10^9 proerythroblasts were obtained from the separation of 10^{10} nucleated spleen cells (6-7g spleen weight) across three 25cm sedimentation chambers.

Cell culture

To study the effects of erythropoietin (Epo) on terminal erythroid differentiation, FVA-derived proerythroblasts were cultured at 10^6 cells/ml in Iscove-modified Dulbecco medium (IMDM, Life Technologies #12440043), 30% heat-inactivated fetal bovine serum (Gibco, 26140-079), 1% Penicillin-Streptomycin (Gibco #15140-122), 10% deionized BSA, and 100 μM α-thioglycerol (MP Biomedicals #155723). Terminal erythroid differentiation of purified proerythroblasts was induced by the addition of 0.4 U/ml human recombinant Epo (10kU/ml

Epogen by Amgen, NDC 55513-144-10) to media. At the desired times after the addition of Epo, cells were crosslinked by the addition of 1% formaldehyde for 10 minutes. Crosslinking was then quenched by the addition of 125mM glycine. Crosslinked cells were collected by centrifugation for 5 minutes at 1,000g at 4°C, washed once with 1x PBS, flash frozen in liquid nitrogen, and stored at -80°C until used for ChIP analysis.

ChIP-exo and antibodies

With the following modifications, ChIP-exo was performed as previously described[187] with chromatin extracted from 50 million cells, ProteinG MagSepharose resin (GE Healthcare), and 5 μg of antibody directed against the H3K4me1, H3K4me3, H3K27ac, or Tal1 (Abcam ab8895, ab8580, ab4729, and Santa Cruz Biotech sc-12984, respectively). First, formaldehyde crosslinked cells were lysed with buffer 1 (50 mM HEPES–KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10% Glycerol; 0.5% NP-40; 0.25% Triton X-100), washed once with buffer 2 (10 mM Tris–HCL, pH8; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA), and the nuclei lysed with buffer 3 (10 mM Tris–HCl, pH 8; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na–Deoxycholate; 0.5% *N*-lauroylsarcosine). All cell lysis buffers were supplemented with fresh EDTA-free complete protease inhibitor cocktail (CPI, Roche #11836153001). Purified chromatin was sonicated with a Bioruptor (Diagenode) to obtain fragments with a size range between 100 and 500 bp. Triton X-100 was added to extract at 1% to neutralize sarcosine. Insoluble chromatin debris was removed by centrifugation, and sonication extracts stored at -80°C until used for ChIP analysis.

Illumina sequencing and data pre-processing

Libraries were sequenced using an Illumina NextSeq500 sequencer as single-end reads 75 nucleotides in length on high output mode. The sequence reads were aligned to the mouse mm10 reference genome using BWA-MEM algorithm using default parameters[39]. Since patterns described here were evident among individual biological replicates, and replicates were well correlated, we merged all tags from biological replicate data sets for final analyses.

Data visualization

Raw sequencing tags were smoothed (20bp bin, 100bp sliding window) and normalized to reads per kilobase per million (RPKM) using deepTOOLS[188] and visualized with Integrative Genomics Viewer (IGV)[189]. Chromatin Analysis and Exploration (ChAsE) visualization suite[190] was used to display the distribution of histone marks relative to the TSS. RPKM normalized density plots were generated using Java TreeView[191].

Chromatin state mapping

To identify genomic intervals whose pattern of histone marks were consistent with enhancers, we applied a hidden Markov modeling algorithm using ChromHMM[31]. To obtain a unified set of enhancer intervals, the data for each histone mark was merged across both time points. Briefly, the merged files were binarized using the BinarizeBed function with the signal threshold option set to 100. The minimum number of non-redundant chromatin states was heuristically determined to be six using the LearnModel function with default parameters.

Quantification and annotation of ChromHMM enhancer intervals

Chromatin states 1 and 2 (**Figure 2.4C**), whose chromatin patterns were consistent with enhancers, were annotated and quantified using the Hypergeometric Optimization of Motif EnRichment (HOMER) suite[192]. Briefly, bam files were converted to tag directories using the makeTagDirectory function with the –genome, –checkGC, and –format options. To quantify and normalize tags within enhancer regions to RPKM, the analyzeRepeats function was used with the –rpkm and –d options. The $\log_2$ fold change (from 0 to 1 hour after Epo) was then computed for each histone mark. Finally, enhancer-gene associations were inferred based on the nearest gene paradigm[33,44,193], and each enhancer interval was linked to a gene using the annotatePeaks function with the –noadj and –d options.

Super enhancer analysis

Super enhancers were identified based on the algorithm originally described by Whyte et al.[53], and implemented by HOMER using the findPeaks function with the –style option set to super. For this analysis, the maximum distance to stitch peaks together (-minDist) and the local fold change thresholds (-L) were set to 6,000 and 1, respectively. Data for H3K4me1 and H3K27ac marks were combined to generate a single "enhancer" tag directory for 0 and 1 hours individually. Enhancer intervals whose midpoints were within 1kb of a TSS were filtered out. The remaining enhancer boundaries were trimmed 3kb distal to a TSS if their interval edges overlapped with a TSS. As with the ChromHMM enhancer analysis above, the super enhancer regions were quantified and annotated using the analyzeRepeats and annotatePeaks functions.

Classification of enhancers

Similar to the criteria detailed in Ostuni et al.[45], enhancer intervals were assigned to seven enhancer classes based on H3K27ac and H3K4me1 RPKM values and H3K27ac fold-change for signals in response to Epo stimulation. Briefly, RPKM values within the lowest quartile or the upper three quartiles were designated as "low" or "high", respectively. Occupancy changes in response to Epo were designated as "up" or "down" based on whether RPKM fold changes increased or decreased by more than 2-fold, respectively. RPKM fold changes that were between these values were considered as no change. The criteria for each enhancer class is summarized below: 1) *Constitutive activated* enhancers displayed "high" H3K27ac and "up" in H3K27ac; 2) *Constitutive not-activated* enhancers displayed "high" H3K27ac and no change in H3K27ac; 3) *Poised activated* enhancers displayed "high" H3K4me1 levels, "low" H3K27ac, and "up" in H3K27ac; 4) *Poised not-activated* enhancers displayed "high" H3K4me1 levels, "low" H3K27ac, and no change in H3K27ac; 5) *Latent* enhancers displayed "low" H3K4me1 and H3K27ac levels and "up" in H3K4me1 and H3K27ac levels; 6) *Latent to poised* enhancers displayed "low" H3K4me1 and H3K27ac levels, "up" in H3K4me1, and no change or "down" in H3K27ac levels; 7) *Repressed* enhancers displayed "high" H3K27ac and "down" in H3K27ac. Lastly, 3,279 intervals were excluded from further analysis since they displayed biologically irrelevant characteristics, such as low H3K4me1 and/or H3K27ac levels that further decreased upon Epo stimulation.

Motif discovery

De novo motif discovery for all enhancer regions was conducted using the findMotifsGenome function in the HOMER suite. To comprehensively identify locations for

motifs related to erythroid cell function that were enriched in the de novo search, we conducted a directed search for the following consensus motifs[194] with zero mismatches allowed: Gata1 (WGATAR), Klf1 (YMCDCCCW), Tal1-Ebox (CANNTG), ETS (YWTCCK), and Stat5 (TTCYHDGAA). The scanMotifGenomeWide function in HOMER and intersectBED function in BEDtools [195] were used to find all instances of each motif listed above that resided within enhancer intervals.

Results

Experimental overview and distribution of histone marks at the β-globin locus control region

To study the molecular action of Epo, we leveraged the anemia-inducing strain of the Friend virus (FVA) murine model system that has been shown to recapitulate normal erythropoiesis, as evidenced by Jak-Stat5 signaling, globin expression kinetics, cell morphology, cell surface marker kinetics, and cellular enucleation[99,101,102,185,196-198]. Indeed, the FVA-derived proerythroblasts were recently used as the standard to develop an improved flow-cytometry sorting scheme for bone-marrow derived erythroblasts[99]. This system enables facile large-scale procurement of highly purified murine proerythroblast cell populations that synchronously respond to Epo (**Figure 2.1A**)[129].

Friend virus infection in susceptible mice leads to a multistage disease, characterized initially by erythroblastosis, and then later by erythroleukemia[104]. FVA-derived proerythroblasts used in this study are harvested during the initial cell expansion stage of erythroblastosis, and are thus not erythroleukemic. The two strains of the Friend virus result in distinct phenotypic outcomes due to amino acid differences in the Friend virus glycoprotein, gp55. The polycythemia-inducing gp55$_P$ variant requires both EpoR receptor and sf-Stk, thereby

constitutively activating Jak2-Stat5 signaling and Epo-independent differentiation [199]. Mechanistically, the anemia-inducing gp55$_A$ variant directly interacts with the sf-Stk receptor to enable erythroid precursor proliferation in the absence of Epo (**Figure 2.1A**). Indeed, a previous report showed that neither EpoR nor Stat5 are required for FVA-induced erythroblastosis[185].

To assess the extent to which Epo influences the epigenetic landscape, we performed ChIP-exo on three key histone H3 modifications (H3K4me1, H3K27ac, and H3K4me3) before and after cells were exposed to Epo (0 and 1 hour Epo, **Figure 2.1B**). For each antibody, ChIP-exo signals across biological replicates were highly correlated (R = 0.96 - 0.99), indicating high reproducibility. Importantly, these three histone marks enable complex pattern recognition algorithms, such as hidden Markov modeling, to identify candidate enhancers genome-wide and decipher them from promoter specific patterns[31]. The putative functional nature of enhancers in the present study should be noted since enhancer predictions are based on structural histone modification patterns. In **Figure 2.1C**, the histone modification patterns are illustrated at the well-characterized β-globin locus and nearby locus control region (LCR), which is a cluster of enhancers that control β-globin transcription[200]. Consistent with the marks found at enhancers, H3K4me1 and H3K27ac, but not H3K4me3, were enriched at the LCR. Rather, H3K4me3 was enriched at the promoters of the β-major and -minor globin genes that are expressed in adult erythroid cells. In contrast, the embryonically expressed βh1 and εγ loci lacked both the H3K27ac and H3K4me3, which are associated with active promoters.

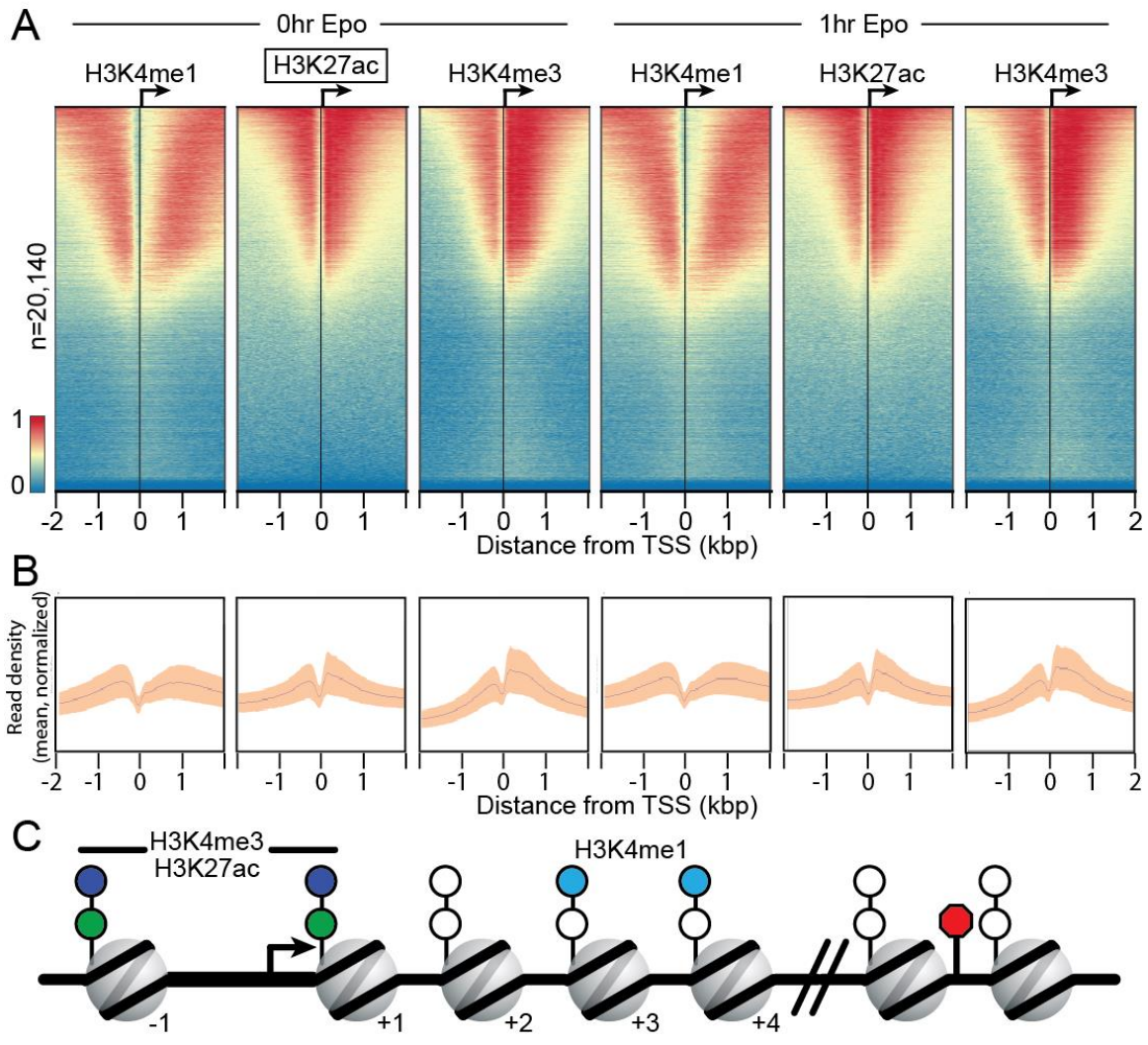

Figure 2.2. ChIP-exo reveals histone modification patterns at gene promoters. (**A**) ChIP-exo libraries were prepared for H3K4me1, H3K27Ac, and H3K4me3 and visualized using ChAsE. Aligned heatmaps show RPKM normalized number of reads across a 4kb genomic interval in 20bp bins relative to the TSS before and after Epo stimulation. Regions are sorted in descending order based on average row tag density for 0 hour H3K27ac. Each row represents one protein-coding gene, with 20,140 lines/genes present. Red and blue reflect high and low read densities, respectively. (**B**) Composite plots below each heatmap quantify the normalized tag density. The central trace denotes the average tag density for each 20bp bin and the orange fill reflects the standard deviation. (**C**) Model illustrating the nucleosome position for enriched histone marks, which are derived from the tag enrichment patterns in panel A. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

ChIP-exo analysis reveals histone modification patterns at gene promoters

Next, we examined the global ChIP-exo patterns for H3K4me1, H3K27ac, and H3K4me3 at promoters of protein coding genes. The ChIP-exo signal for each mark before and after Epo treatment (0 and 1 hour, respectively) was aligned to the coding transcription start site (TSS) and displayed as a heatmap (**Figure 2.2A**, sorted by maximal peak position in **Figure 2.3A**). Each histone mark was detected at nearly half of all protein-coding genes, consistent with previous reports showing that these marks are not found at all genes, but rather, are enriched in the subset of genes that are actively transcribed in erythroid cells[175-177].

It is well established that the genome-wide distribution of histone marks at promoters is not random, but exhibits distinct preferences to specific nucleosomes within promoter regions[27]. Given the mapping precision afforded by ChIP-exo analysis, we examined the spatial distribution of histone marks relative to each other and nucleosome positions relative to the TSS. As expected, we found that H3K4me1 and H3K4me3 tended to be mutually exclusive in their preferred nucleosomes, with H3K4me3 most enriched at the -1 and +1 nucleosomes. In contrast, H3K4me1 marks were enriched at nucleosomes distal to the TSS, particularly the +3 and +4 positions, but decayed rapidly thereafter. Notably, the H3K27ac ChIP-exo signal was enriched in a similar pattern to H3K4me3 at the -1 and +1 nucleosomes, reflecting the expected signature pattern at the nucleosomes flanking core promoter regions[12]. Composite plot distribution peaks (in **Figure 2.2B** and the illustration in **Figure 2.2C**) further highlight the nucleosomal preferences of specific histone marks at promoters of protein-coding genes in proerythroblasts.

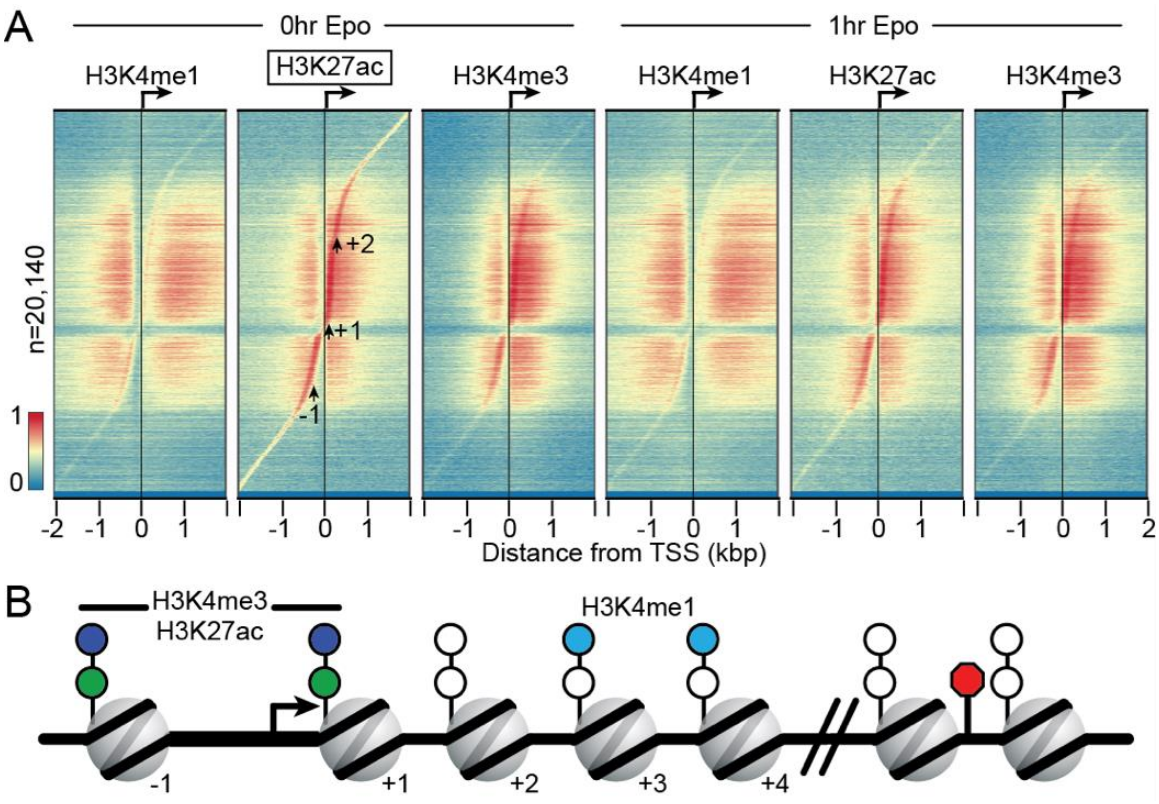

Figure 2.3. Preferred nucleosome position for histone modification patterns at protein coding genes. (**A**) Heatmaps were aligned such that regions are sorted in descending order based on maximum peak position of 0 hour H3K27ac. Sorting in this manner reveals the preferred nucleosome positions for each histone modification. Red and blue reflect high and low read densities, respectively. Nucleosomes in positions -1, +1, and +2 are highlighted with arrows for orientation. (**B**) Model illustrating the nucleosome positions for enriched histone marks. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Hidden Markov modeling of chromatin states reveals unique enhancer signatures

Pioneering studies by the ENCODE consortium suggested that the mammalian genome may harbor hundreds of thousands of enhancer regions, with each of the approximately 200 different cell types containing distinct repertoires of enhancer elements[37,38,42,201]. Given this complexity and despite recent advances in high throughput technologies, it remains a challenge to identify the complete set of bona fide functional enhancers in a given cell type. Nevertheless, this large-scale project and other studies have demonstrated that enhancers and promoters display

distinct epigenetic signatures (**Figure 2.4A**), which can be used to predict enhancer locations in differing cell types with relatively high success rate[29]. In particular, enhancer elements are demarcated by histone 3 lysine 4 mono-methylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac)[27,34,193]. This signature distinguishes enhancers from promoters, which are marked by histone 3 lysine 4 tri-methylation (H3K4me3).

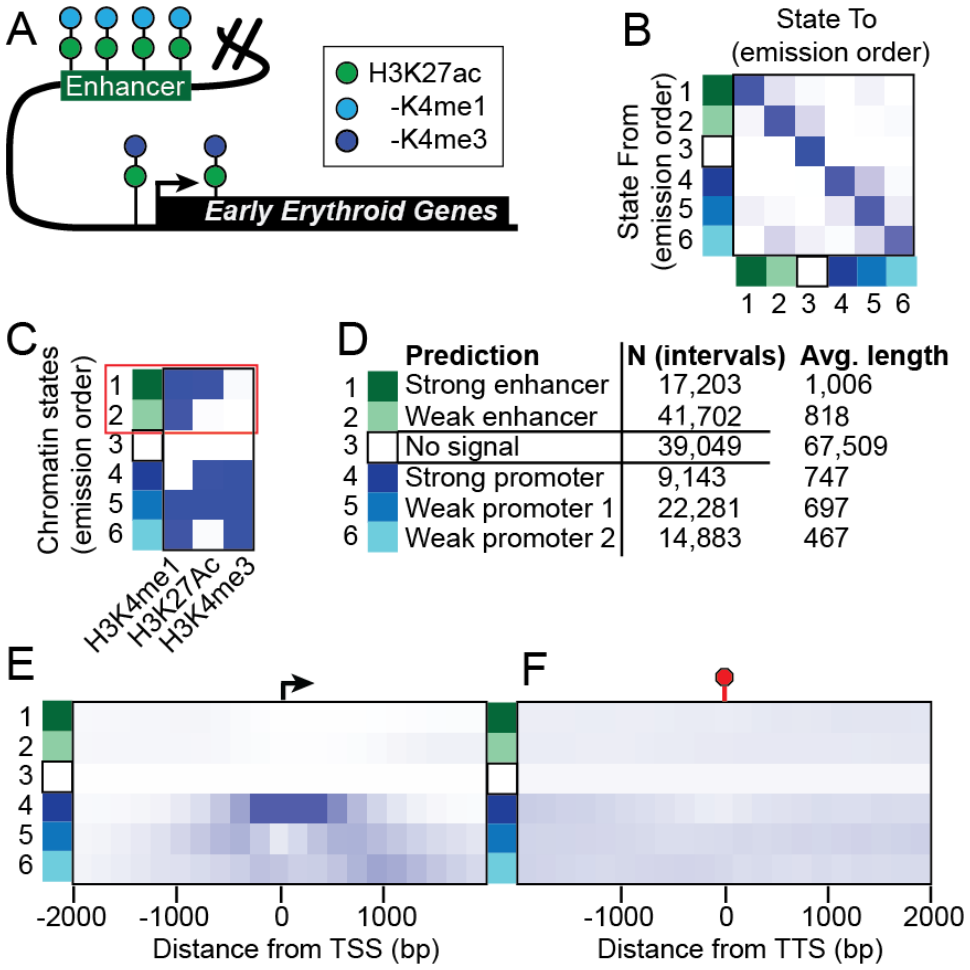

Figure 2.4. Hidden Markov modeling of chromatin states identify unique enhancer signatures. (**A**) The cartoon illustrates distinct chromatin signatures associated with enhancers and promoters. (**B**) The transition matrix shows the probability that the state in the row (State From) will transition to the state in the column (State To) using hidden Markov modeling of chromatin states. (**C**) Heatmap representation of chromatin state emissions from hidden Markov modeling that predicts enhancer intervals from H3K4me1, H3K27ac, and H3K4me3. The red box highlights states 1 and 2 that display chromatin state patterns consistent with enhancers, which is enrichment of H3K4me1 and/or H3K27ac, but not H3K4me3. (**D**) Functional annotation inferences for each chromatin state are listed next to the number of intervals

comprising the state and the average length. (**E-F**) Heatmaps showing the enrichment of each learned chromatin state +/- 2000bp from the TSS and TTS, respectively. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Thus, to systematically identify the genomic locations of chromatin signatures consistent with enhancers in FVA-derived proerythroblasts, we applied hidden Markov modeling (HMM) to ChIP-exo data **(Figure 2.4)**. ChromHMM employs a powerful algorithm for finding unique patterns in complex data[31]. HMM analysis of the ChIP-exo data for H3K4me1, H3K27ac, and H3K4me3 from the 0 and 1 hour Epo time points revealed six non-redundant chromatin states (**Figure 2.4B, C**). Chromatin states 1 and 2, representing nearly 60,000 candidate enhancers, had an average length of ~900bp and displayed signatures consistent with strong and weak enhancers, respectively, that is having H3K4me1 and/or H3K27ac, but lacking H4K4me3, respectively. As expected, these candidate enhancer regions tended to be distal to transcription start and end sites (**Figure 2.4E, F**, quantified in **Figure 2.5**). In contrast, chromatin states 4-6 were strongly associated with promoter regions (**Figure 2.4E**) and were each strongly enriched with H3K4me3 relative to states 1 and 2. Lastly, chromatin state 3 was depleted of ChIP signal for the three histone marks in this study and encompassed the majority of the genome, suggesting low representation of these marks and/or that other histone modifications may be present in the intervals of this state.

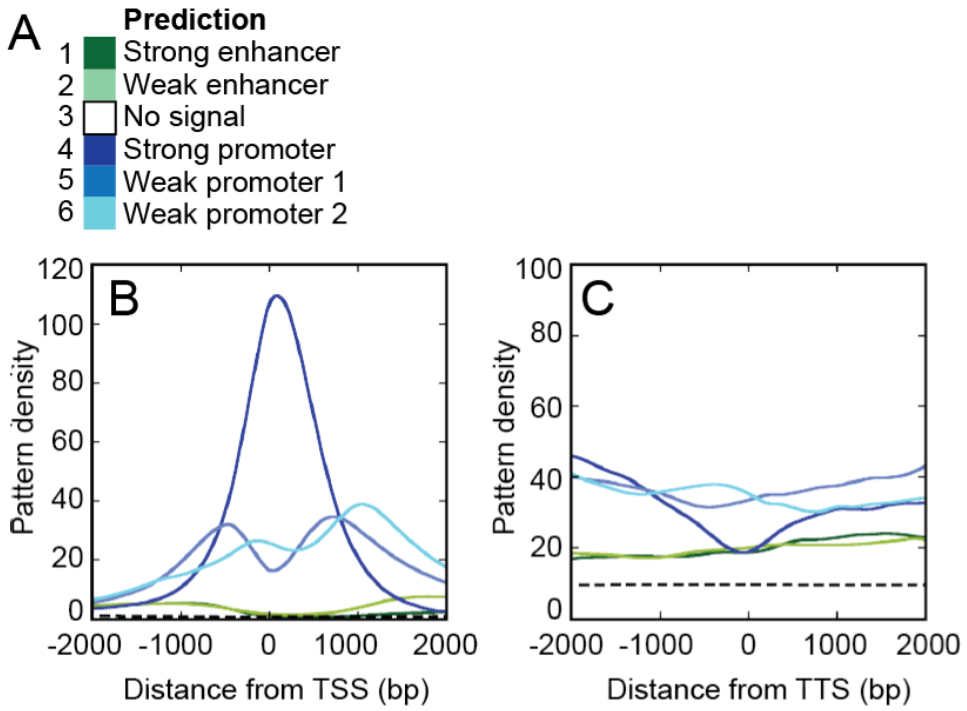

Figure 2.5. Quantification of chromatin state signatures. (**A**) Functional annotation inferences for each state. (**B-C**) Quantification of frequency density of distance from TSS and TTS, respectively, for each ChromHMM predicted chromatin state. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Epo-induced remodeling of the erythroid enhancer landscape

Tens of thousands of enhancers are scattered across mammalian genomes in a cell-type specific manner and are dynamically shaped in response to environmental stimuli[12]. Although a handful of studies have focused on erythroid enhancers[43,182,183], how Epo influences erythroid enhancers remained unclear. We hypothesized that Epo signaling reshapes the epigenetic landscape of a specific set of erythroid enhancers to drive the erythroid transcriptional program. To test this hypothesis, we systematically classified the nearly 60,000 candidate enhancer regions based on their response to Epo stimulation, in a manner similar to that previously described[45]. The presence or absence of H3K27ac is associated with enhancer activation or repression, respectively[35]. Poised enhancers initially lack H3K27ac, but can acquire the acetyl mark upon

cellular stimulation by environmental cues, such as cytokines or hormones[33,45,202,203]. Latent enhancers were recently described as regions of the genome that lack enhancer histone marks, but acquire these marks upon stimulation[45].

Overall, response to Epo stimulation was highly dynamic, enabling stratification of the ~60,000 genomic intervals into four major classes (**Figure 2.6A**). Candidate enhancers across each class shared several features, including commonly residing within intergenic and intronic regions of the genome, averaging about 1kb in length, and varying quite dramatically in their distances to the nearest gene (**Figure 2.7**). First, the class of constitutive enhancers was associated with both H3K4me1 and H3K27ac in unstimulated proerythroblasts (0 hour Epo). While most (45,297) were unaffected after 1 hour of Epo stimulation (constitutive not-activated, **Figure 2.6E**), 855 displayed increased acetylation within the first hour of Epo stimulation (constitutive activated, **Figure 4.6E**). **Figure 2.6B** illustrates the Epo-induced dynamics of a constitutive activated enhancer at the *TNFRSF13C* gene, which, interestingly, encodes a receptor known to promote survival in B-cells[204]. Indeed, transcriptional profiling in murine proerythroblasts found that *TNFRSF13C* was strongly induced in response to Epo treatment[178]. This observation supports the notion that the enhancer we identified upstream of the *TNFRSF13C* locus may be mediating its transcriptional response to Epo to further promote pro-survival pathways that Epo is well known to regulate[102]. As shown in **Figure 2.6C**, the *GATA1* locus, which encodes one of the three master regulators of the erythroid expression program, exemplifies the constitutive not-activated class of enhancers. Consistent with this observation, previous studies have shown that maximal *GATA1* gene expression, and presumably increased enhancer acetylation, occurs in basophilic and orthochromatic erythroblasts[175-177], which are not present in the model system until 12-36 hours after Epo stimulation.

Next, the class of poised enhancers was characterized by H3K4me1 but not H3K27ac in unstimulated proerythroblasts. Of the 7,841 poised enhancers we detected, 688 were Epo-responsive and displayed increased acetylation after 1 hour of Epo stimulation (poised activated, **Figure 2.6E**). Gene ontology analysis revealed that the set of genes proximal to poised activated enhancers were most associated with positive regulation of TF activity (P=10^{-3}), including *KIT*, the gene encoding the receptor tyrosine kinase that activates a number of signaling pathways critical to cell survival during erythropoiesis. Also included in the group of poised activated enhancers were genes encoding Wnt signaling proteins (*WNT3A* and *WNT10B*), which are critical regulators balancing self-renewal, proliferation, and differentiation in hematopoietic cells[205].

We also identified a class of over 100 latent enhancers, which gained H3K4me1 and/or H3K27ac upon Epo stimulation (**Figure 2.6E**). Remarkably, nearly half (45) of these latent enhancers were activated within 1 hour of Epo stimulation. **Figure 2.6D** illustrates the Epo-induced dynamics of the class of latent activated enhancers at the *BOP1* (Block of Proliferation 1) gene locus. The precise timing of the *BOP1* latent enhancer activation with Epo-stimulated erythropoiesis is consistent with the cellular transition to the inherently non-proliferating mode of terminal differentiation. Interestingly, as a protein involved in rRNA processing, Bop1 has also been implicated in ribosomopathies[206,207], such as Diamond-Blackfan Anemia.

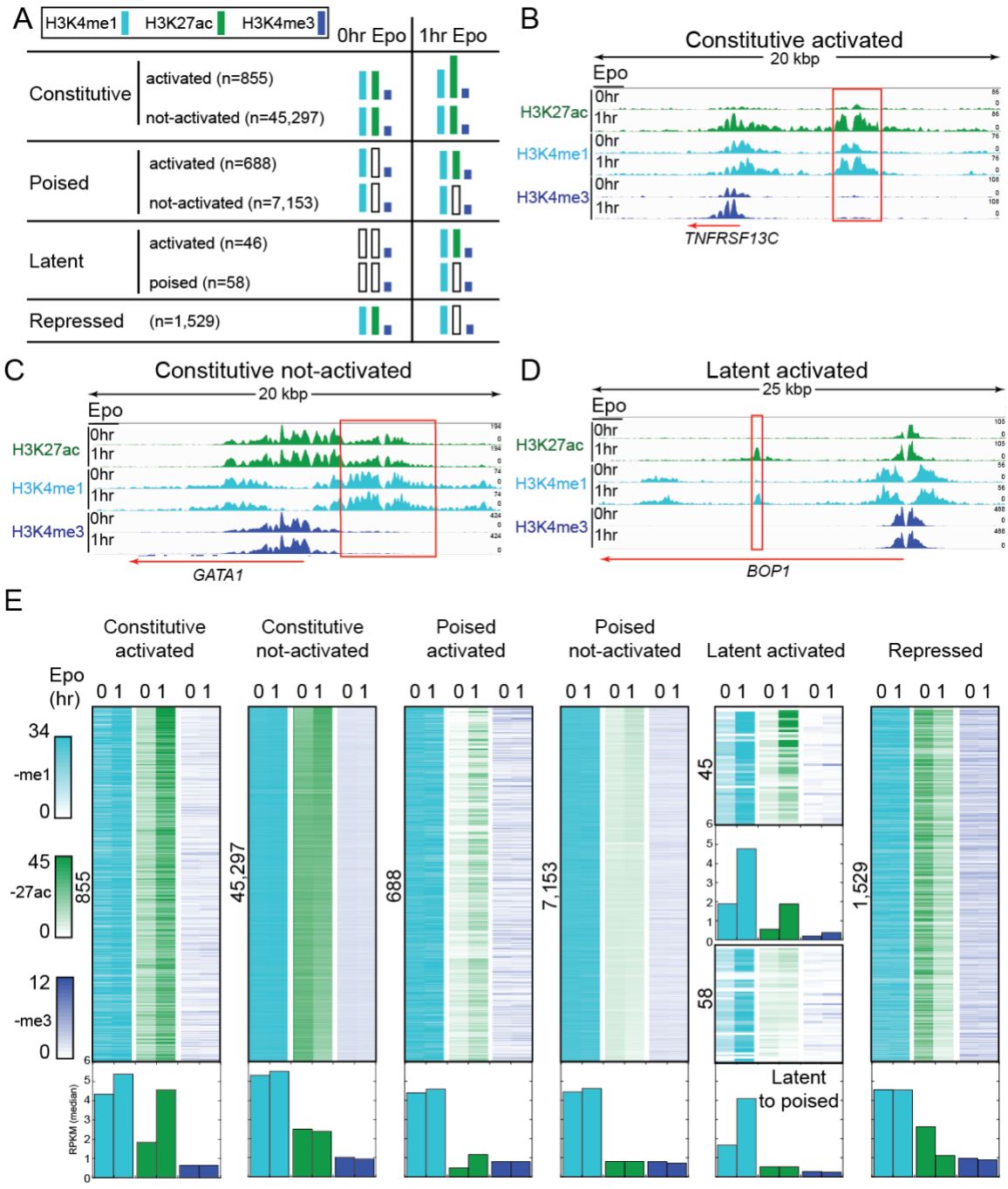

Figure 2.6. Epo-induced remodeling of the erythroid enhancer landscape. (**A**) Classification of erythroid enhancers based on their response to Epo. (**B-D**) Representative genome browser views showing Epo-responsive enhancers (B, D; constitutive activated and latent activated, respectively) and an Epo non-responsive enhancer (C, constitutive not activated). Red boxes denote genomic intervals identified as enhancers by hidden Markov modeling. (**E**) Density plots showing the dynamic behavior of histone marks for each enhancer interval induced by a 1 hour Epo treatment. The number of enhancer intervals shown is indicated to the left of each plot. The normalized RPKM median value for each column is represented as a bar graph below the density

plot. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Finally, the class of enhancers consistent with repression displayed both H3K4me1 and H3K27ac in untreated proerythroblasts, but lost or showed diminished H3K27ac within 1 hour after Epo stimulation (**Figure 2.6E**). This class of enhancers were enriched near numerous genes encoding TF regulators (P=10^{-7}), including Max. Max is the primary partner for Myc, conferring the ability of Myc/Max dimers to bind E-boxes. Together, Myc/Max dimers control the opposing cell fate decisions of cell cycle progression and apoptosis[208], suggesting that Epo-induced enhancer remodeling may contribute to cell fate decisions during erythropoiesis via Max. Consistent with the observation that Epo simulation leads to repression of an enhancer linked to the binding partner for Myc, a previous report showed that down-regulation of Myc is important during erythropoiesis[209].

In summary, we identified over 3,000 Epo-responsive enhancers, of which 1,588 and 1,529 were activated and repressed, respectively, within the first hour of terminal erythroid differentiation. This suggests that Epo rapidly remodels a subset of the epigenetic landscape of proerythroblasts undergoing erythropoiesis. Consistent with a previous report[182] showing that chromatin states remain largely unchanged during erythroid differentiation, the vast majority of the candidate enhancer regions identified in this study (n=52,508) were static during the first hour of Epo exposure. This indicates either that they are established at a prior precursor stage and not Epo-responsive, or instead respond to Epo at some point after 1 hour of treatment.

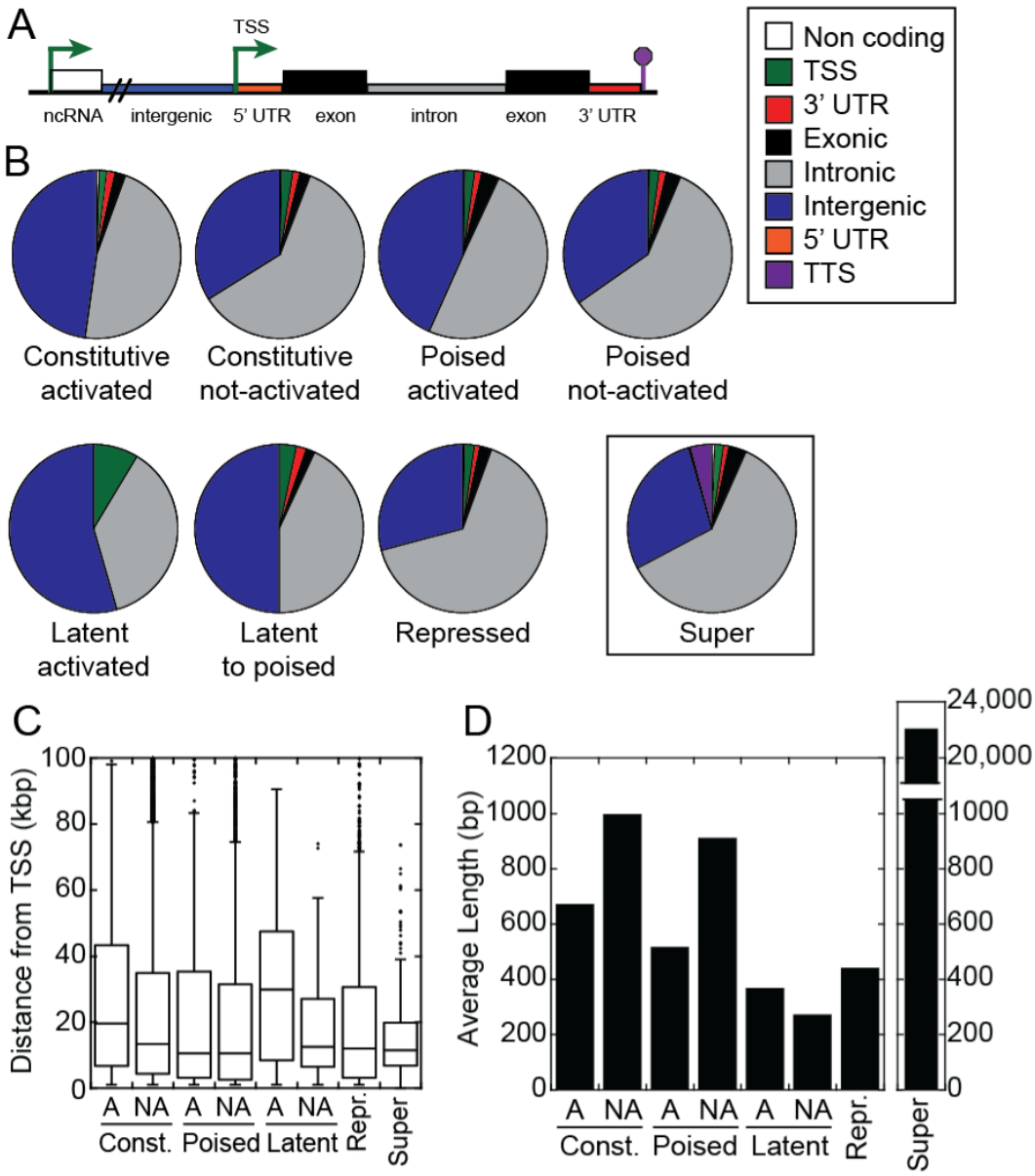

Figure 2.7. Additional characteristics of identified enhancer classes. (**A**) Cartoon depicting areas of the genome where enhancers were located. (**B**) Distribution of locations of enhancers by class. (**C**) Box-plot of distance from TSS for each class of enhancers. (**D**) Average length of each class of enhancers. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Validation and evolutionary conservation of candidate enhancers

A caveat to algorithms that predict enhancer regions based on functional genomic data is the possibility of false positives. Therefore, to provide corroborative evidence for the validity of candidate enhancers, we applied four complementary computational and experimental approaches: 1) DNA motif analysis, 2) correlation to promoter acetylation, 3) overlap analysis with orthogonal published data sets, and 4) TF binding via ChIP-exo analysis. Transcription factors regulate cell-type specific gene expression patterns by binding to their cognate motifs within enhancer regions and recruiting the transcription machinery to target genes[93]. In erythroid cells, Gata1, Klf1, and Tal1 are master TF regulators that play a major role in coordinating the precise timing of gene expression patterns during erythropoiesis[114]. Gata1, the most well-studied among the trio of master regulators, binds to the WGATAR consensus motif [210,211], which influences transcription positively or negatively, depending on the composition of the Ldb1-nucleated complex[182,212,213].

Therefore, we hypothesized that candidate enhancers identified by HMM that are physiologically relevant in erythroid cells would be enriched for motifs corresponding to the master regulator TFs of erythropoiesis. To test this hypothesis, we conducted de novo motif discovery analysis of all ~60,000 enhancer regions from **Figure 4.6**. Strikingly, the Gata motif was the most commonly enriched ($P=10^{-13}$ to 10^{-598}) across the enhancer classes (**Figure 2.8A**). Other erythroid related motifs were also found in the de novo search, including the Klf1, ETS, and Stat5 consensus motifs.

Since de novo motif analysis provides initial clues into overrepresented motifs in the genomic regions, we next conducted a more sensitive, directed motif search. Given the

connection between Epo signaling and Gata1 phosphorylation[214], we focused on identifying all occurrences of the Gata consensus motif with enhancers (**Figure 2.8B**). Indeed, for each class of enhancers the observed frequency of the Gata motif was significantly greater than an equivalent number of randomly sampled genomic regions (**Figure 2.8B**, empirical *P*-value $< 10^{-3}$). Altogether, our motif analysis provides corroborative evidence in validating candidate enhancers.

Previous studies demonstrated that enhancer-promoter interactions may be inferred by correlating H3K27ac patterns[30]. Therefore, we examined the relationship between Epo-stimulated changes in H3K27ac at enhancers with that of the nearest gene promoter. In general, genes that increased in promoter acetylation were linked to nearby enhancers that also acquired H3K27ac in response to Epo (e.g. activated enhancer subclasses denoted by "A", **Figure 2.8C**), with the repressed enhancer class least represented. Overall, the correlation between decreased promoter and enhancer acetylation was far weaker, although the repressed class of enhancers was most associated with a decrease in promoter acetylation. Not only do these results support the enhancer classification methodology (**Figure 2.6**), they provide further corroborative evidence for the candidate enhancers in this study.

Given the previous finding that TF binding is an accurate predictor of enhancer activity [41] and that Epo stimulation rapidly activates Tal1 in FVA cells[215], we performed Tal1 ChIP-exo analysis in FVA-derived proerythroblasts and identified 19,025 and 13,913 enriched peaks for 0 and 1 hour Epo treatment, respectively. Strikingly, nearly 60% (20,289 of 32,938) of all Tal1 peaks resided within candidate enhancer regions identified in this study, which are separated by class in **Figure 2.8D**. In other words, Tal1 alone occupied 33% of all candidate enhancers (18,370 of 55,626 enhancers with one or more Tal1 peaks), suggesting that these enhancer regions are valid.

Lastly, we compared our set of enhancer locations to published orthogonal data sets from mouse and human studies[43,182]. We found that nearly 24,000 enhancers identified in this study were also present in G1E-ER4 cells[182] (**Figure 2.8E**). In addition, 44% (3,995 of 9,059) of enhancers previously identified in primary human proerythroblasts[43] were shared with the current study (**Figure 2.8F**), indicating substantial evolutionary conservation of erythroid cis-regulatory modules. In summary, based on evidence from four complementary approaches presented above used to validate enhancers, we conclude that candidate enhancer regions identified in this study generally represent bona fide enhancers and thus are likely physiologically and functionally relevant to terminal erythroid differentiation.

Identification of super enhancers at erythroid genes involved in cell fate decisions

Super enhancers are an emerging class of long, clustered enhancers that are important for establishing cell fate and identity[53]. A recent study examined super enhancers in CD34+ derived proerythroblasts and murine G1E cell lines[43], but whether Epo stimulation impacts super enhancers remains unclear. Therefore, to investigate the extent to which Epo signaling influences super enhancer dynamics in proerythroblasts, we applied the algorithm previously described to identify super enhancers[53] to our H3K4me1 and H3K27ac ChIP-exo data sets. We identified 395 super enhancers dispersed throughout the mouse genome that were characterized by their length (21,192 bp on average (**Figure 2.7**), and exceptionally high H3K4me1 and H3K27ac levels (**Figure 2.9A-B**). We found that the super enhancers were generally non-responsive to Epo with respect to H3K27ac (**Figure 2.9B**). Consistent with the notion that super enhancers tend to be associated with genes controlling cell fate and identity, they were discovered nearby a number of genes known to play roles in erythroid cell physiology, such as *TAL1*, *BCL11A*, and *MIR144/451* (**Figure 2.9C-E**). *TAL1* and *BCL11A* encode TFs that are well known to regulate erythroid gene

expression patterns[94,215,216]. The *MIR144/451* polycistronic gene produces microRNAs that promote erythroid maturation and are strongly induced by Gata1 during erythropoiesis[217,218]. Interestingly beyond the representative genes shown in **Figure 2.9C-E**, Tal1 bound to 92% of all super enhancer regions (365 of 395 enhancers). In addition, Gata motifs were found in nearly all super enhancer regions (99.7%, 394 of 395 super enhancers). Indeed, the Gata motif overlapped with 95.6% of Tal1 peaks within super enhancers (345 of 361 Tal1 peaks for 0 hour Epo). Overlap analysis between super enhancer locations revealed that 57% (225 of 395, **Figure 2.9F**) of super enhancer locations from this study were shared with a previous report in mouse G1E-ER4 cells[43], thereby corroborating our super enhancer analysis. In contrast, super enhancer locations were evolutionarily poorly conserved when compared to super enhancer locations from CD34+ derived proerythroblasts (16%, 64 of 395, **Figure 2.9G**).

Epo modulated enhancers integrate erythroid signaling pathways

To investigate what types of genes are associated with Epo-modulated enhancers, we conducted gene ontology and pathway enrichment analysis (**Figure 2.10A**). We found that many enhancers modulated by Epo stimulation were linked to genes involved in cell signaling pathways, such as Jak-Stat (n=24), PI3K (n=45), and FoxO (n=25) signaling. While these pathways are known to regulate erythroid differentiation[219], it is surprising that enhancers linked to these pathways are remodeled in response to Epo. Nevertheless, this observation is consistent with the notion that Epo-driven epigenetic changes may form feed forward loops that serve to strengthen the kinetics and robustness of Epo-responsive pathways.

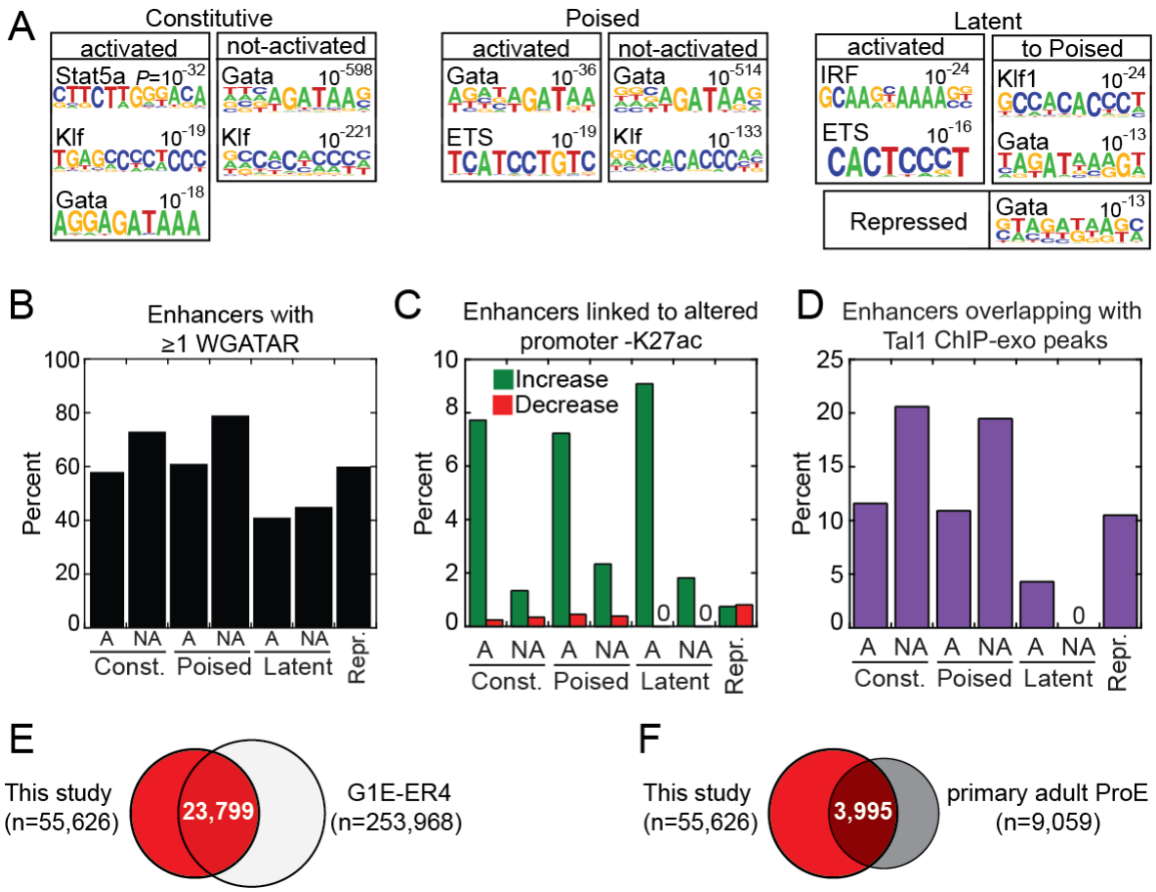

Figure 2.8. Validation of candidate enhancers. (**A**) Transcription factor (TF) binding motifs overrepresented in each enhancer class are shown as motif logos based on their respective position weight matrices (PWMs). De novo discovery of PWMs, P-values, and best Jaspar motif match were generated using HOMER. (**B**) The percent of each enhancer class containing at least one Gata1 consensus motif (WGATAR) with zero mismatches is shown as a bar graph. Activated and not-activated subclasses are denoted by "A" and "NA", respectively. (**C**) Shown is a bar graph of the percent of enhancers, separated by class, that are nearest to genes with greater than 2-fold increase (green) or decrease (red) in H3K27ac at promoters (+/-1kb). Activated and not-activated subclasses are denoted by "A" and "NA", respectively. (**D**) Percent of enhancer intervals that are occupied by ≥1 ChIP-exo Tal1 peaks, and broken out by enhancer class. Activated and not-activated subclasses are denoted by "A" and "NA", respectively. (**E**) Venn overlap between enhancer intervals in this study (red fill) and mouse G1E-ER4 chromatin states 2 and 3 (H3Kme1 and H3K4me1/H3K27me3, respectively) from Wu et al. (**F**) Venn overlap between enhancer intervals in this study (red fill) and enhancer regions in human primary adult proerythroblasts from Huang et al[43] [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

During terminal erythroid differentiation, the cytoskeletal structure undergoes a dramatic transformation, which involves actin and the glycosylation of a number of erythroid transmembrane proteins, such as Band3, Glycophorin A, and others[220]. Indeed, within the first hour of exposure to Epo, the epigenetic landscape is altered nearby genes involved in regulating the actin cytoskeleton (n=31), glycosylation (n=153), and transport (n=221). This suggests that cytoskeletal reorganization may be regulated in part at the epigenetic level.

Collectively, TFs orchestrate cell-type specific transcription programs by binding to their cognate sequence motifs within specific enhancers. TFs then recruit co-regulators and RNA polymerase II to promoters to initiate erythroid transcription programs. Strikingly, analysis of genes linked to Epo-modulated enhancers revealed an enrichment of numerous regulators of transcription (n=284), chromatin regulators (remodeling and modification, n=60), and phosphorylation regulators (kinase and phosphatase activity, n=97). In particular, Epo simulation altered enhancers linked to *BACH1*, *BCL11A*, *ELF1*, *KLF10*, *SOX6*, *TCF3*, and *TIF1*γ, each of which have previously been reported to involved in controlling erythroid expression programs[95,221-225]. Together, these observations suggest that Epo stimulation influences a network of erythroid relevant TFs at the epigenetic level, and perhaps serves to coordinately mobilize TF activity to orchestrate erythroid expression programs.

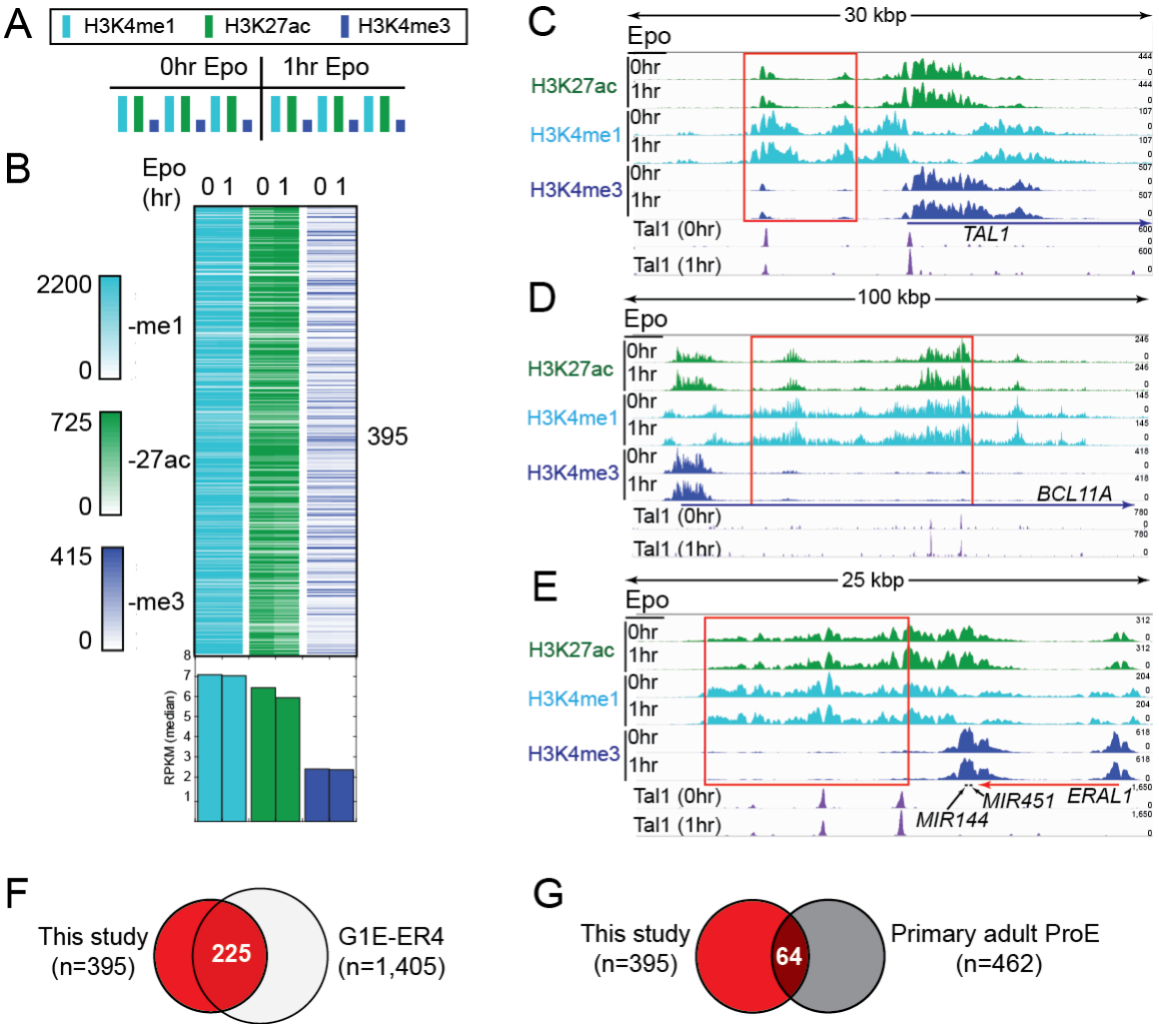

Figure 2.9. Identification of super-enhancers at erythroid genes involved in cell fate decisions. (**A**) Histone modification enrichment across super enhancers in erythroid cells before and after Epo stimulation. (**B**) Density plots showing the H3K4me1, H3K27ac, and H3K4me3 ChIP-exo signal (RPKM normalized) across super enhancer intervals before and after Epo treatment. The normalized RPKM median value for each column is represented as a bar graph below the density plots. (**C-E**) Genome browser views showing super enhancers before and after Epo stimulation at the *TAL1*, *BCL11A*, and *MIR144/MIR451* genes, respectively. Red boxes denote genomic intervals identified as super enhancers by the Whyte et al. algorithm employed by HOMER. (**F-G**) Venn overlap of super enhancer intervals identified in this study (red fill) compared to super enhancer regions from Huang et al[43] in mouse (panel F, G1E-ER4) and human (panel G, primary adult ProE) erythroid cells, respectively. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Discussion

How hormone signaling is manifested in chromatin dynamics remains an important, unanswered question. Here, we show that the hormone Epo reprograms a repertoire of enhancers in murine proerythroblasts that are linked to integrated networks of erythroid transcriptional regulators and signaling pathways. By showing where and how these histone marks are altered in response to stimuli reveals the plasticity of erythroid enhancers, and provides new insights into the molecular action of Epo.

Our experimental murine model uniquely allowed us to investigate how the enhancer landscape changes upon Epo stimulation. We identified ~60,000 enhancers in proerythroblasts, and their histone mark dynamics in response to Epo stimulation enabled stratification of enhancers into four major classes, in addition to the super enhancer class. Consistent with a previous study[182], we find that the enhancer landscape remains largely unchanged before and after Epo stimulation of proerythroblasts. However, nearly 3,000 enhancers were remodeled within the first hour of Epo treatment, suggesting that Epo initiates terminal erythroid differentiation in part by influencing a portion of the erythroid enhancer repertoire. Whether Epo continues to influence and reshape the epigenetic landscape throughout erythropoiesis remains unclear.

Two common themes emerged among the genes linked to Epo-responsive enhancers: 1) overrepresentation of pathways involved in coordinating the transition from a proliferative state to a non-proliferative differentiating state (**Figure 2.10**), and 2) enrichment of genes encoding TFs. The transition to a non-proliferative differentiating state is exemplified by Epo-responsive enhancers linked to *TNFRSF13C* and *BOP1* (**Figure 2.6B**, **D**) and is consistent with the physiological transitions that accompany the cell fate decision to undergo terminal erythroid

differentiation. In particular, once Epo triggers the cellular commitment to undergo erythropoiesis, the cell must halt proliferative signaling pathways and commence differentiation pathways, while at the same time blocking the cellular inclination toward apoptosis. Interestingly, Epo starvation and add back experiments resulted in strongly up regulated of *TNFRSF13C* expression in murine fetal proerythroblasts[178]. This finding is in agreement with our discovery of a constitutive enhancer upstream of *TNFRSF13C* that is further activated within the first hour of Epo stimulation.

Mechanistically, this study provides insight into how Epo stimulation alters the enhancer landscape in erythroid cells, which are presumably bound by TFs to orchestrate the dynamic gene expression programs during erythropoiesis. Indeed, the top Gene Ontology category for genes linked to the Epo-responsive class of constitutive activated enhancers is the positive regulation of transcription from Pol II promoters, which included the genes for TFs *LMO2*, *KLF5*, *KLF13*, and the *EP300* gene, which encodes an acetyltransferase. Furthermore, we found super enhancers near a number of important genes encoding TFs that regulate erythropoiesis, such as *BCL11A* and *TAL1*. It is interesting to note that a recent study reported a *TAL1* super enhancer specific to T-cell acute lymphoblastic leukemia cells, leading the authors to suggest that the *TAL1* super enhancer is oncogenic[226]. However, in the context of normal erythropoiesis, we propose that the *TAL1* super enhancer identified in the present study serves a physiological function controlling early *TAL1* expression kinetics in murine proerythroblasts. In regards to super enhancers, a recent report described a super enhancer at the α-globin locus and elegantly dissected the cis-regulatory requirements controlling α-globin expression [54]. This super enhancer was also found in the present study, further supporting the validity of the enhancers presented.

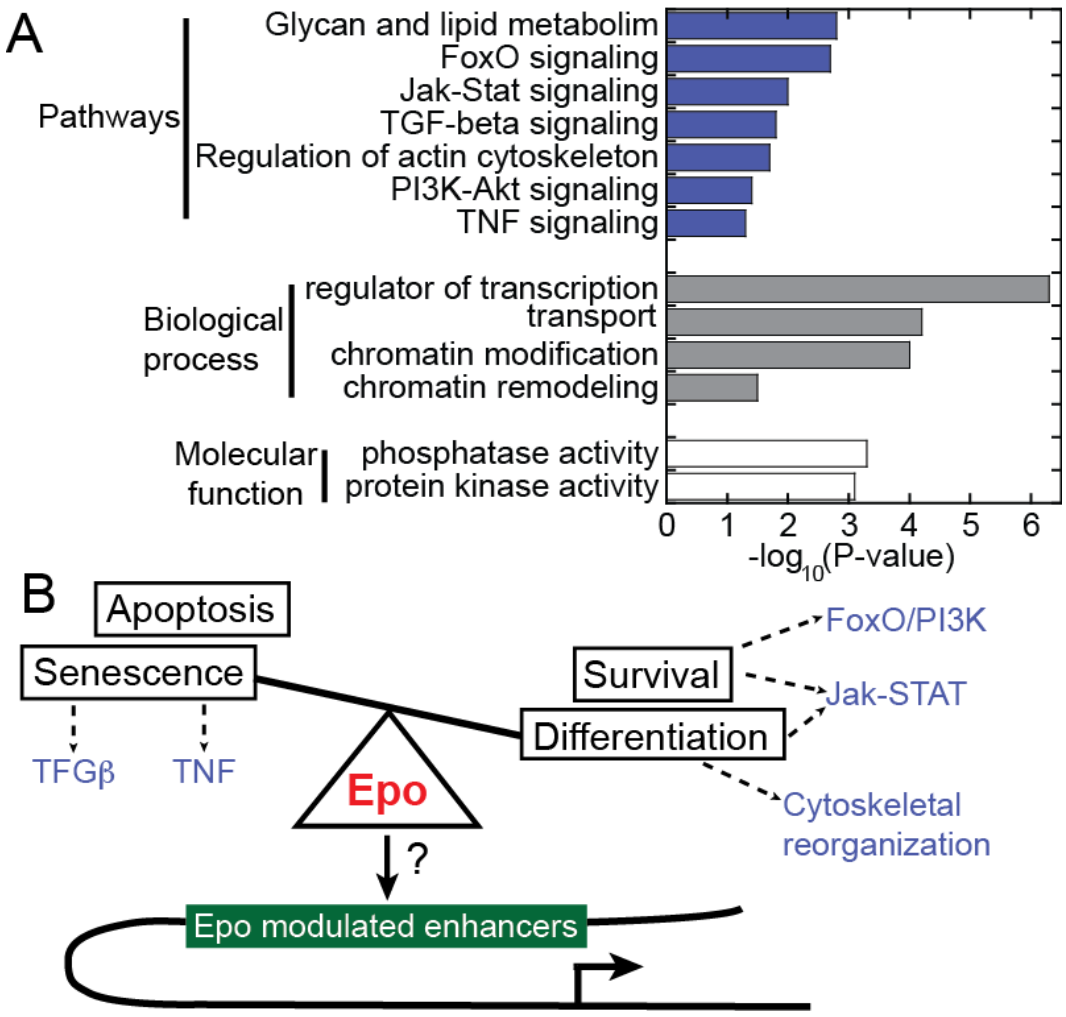

Figure 2.10. Epigenetic integration of signaling pathways by Epo modulated enhancers. (**A**) Gene ontology and pathway enrichment analysis for genes linked to Epo-modulated enhancers. Top enriched pathways, biological processes, and molecular functions are shown in blue, gray, and white fill bars, respectively. (**B**) Model illustrating the interplay of pathways enriched in the set of genes linked to Epo-modulated enhancers. [Figure from Perreault et al. (2017) Experimental Hematology[186] and used in accordance with Copyright, Elsevier]

Several previous studies have used epigenetic profiling to identify candidate enhancers in human and mouse erythroid cells[43,182,183]. While these studies provided detailed insights into erythroid enhancers during erythroid differentiation, it is important to note that the inherent limitations of these model systems precluded an assessment of Epo-dependent enhancer

dynamics. Furthermore, putative enhancers were identified without considering H3K4me1[43,183] or H3K27ac[182], which would miss the dynamic behavior of activated, repressed, poised, and latent enhancer states.

Specifically, our analysis of both H3K4me1 and H3K27ac histone modifications by ChIP-exo analysis in the FVA system circumvents these limitations and provides new insights into erythroid biology by uniquely tying dynamic enhancer states directly to Epo stimulation. In addition, our finding that Epo uncovers nearly 100 latent enhancers in erythroid cells further expands the repertoire of environmental stimuli that reveal cryptic enhancer locations[45,202].

The primary strength of using chromatin state mapping to identify enhancers is that it provides insights into the plasticity of the enhancer landscape in an unbiased and genome-wide manner, particularly in response to stimuli. However, as for any experimental model or approach, we acknowledge the limitations of the present study. First, enhancer predictions based on epigenetic profiling alone does not address functional significance. Corroborative evidence presented in **Figure 2.8** sought to mitigate this limitation, particularly with functional genomic assays, such as Tal1 ChIP-exo analysis (**Figure 2.8F**). Still, establishing a causal link between enhancer activation or repression with a corresponding gene activity is a challenging task, particularly on a genome-wide scale. Further confounding this pairwise enhancer-promoter association is the observation in previous studies that multiple enhancers can together fine-tune the expression of a single gene[47,48]. In agreement, the present study found that on average, a given gene was linked to ~5 discrete enhancer regions. However, whether multiple non-clustered enhancers (i.e. not super enhancers) synergistically or antagonistically regulate a given gene locus requires a gene-by-gene analysis. Lastly, our study, as well as previous erythroid enhancer reports[43,182,183], inferred enhancer-gene linkages using the nearest gene paradigm. A previous

report using human cancer cell lines suggested that as many as 40% of enhancers can skip over the nearest gene to loop to a more distant gene[83], but the extent to which this gene skipping occurs in a normal physiological context remains unclear. Although limited by resolution, long range chromatin interaction assays performed on a genome-wide scale would extend previous locus-specific looping studies in erythroid cells[227] and would complement the current study by furthering our understanding of the dynamics of enhancer repertoire usage during terminal erythroid differentiation.

Here, we used the FVA murine model system of erythropoiesis that uniquely allowed us to investigate how the hormone Epo shapes the enhancer landscape. This study provides novel insights into how hormone signaling pathways are able to control transcriptional programs during cellular differentiation by altering enhancer associated marks. Together, these findings define a cis-regulatory enhancer network for Epo signaling during erythropoiesis, and provide the framework for future studies involving the interplay of epigenetics and Epo signaling.

CHAPTER 3

EPO REGULATES YY1 DYNAMICS in a PRE-ESTABLISHED CHROMATIN ARCHITECTURE[2]

Introduction

Transcription control is a primary mechanism for regulating gene expression in eukaryotes. There are three major steps in the transcription cycle: 1) formation of the preinitiation complex (PIC) comprised of RNA Polymerase II (Pol II) and general transcription factors (GTFs); 2) pause release to productive elongation of Pol II; and 3) transcription termination[228]. Multiple mechanisms exist to regulate each step in this cycle, thereby providing precise control over the magnitude and kinetics of transcription and global gene expression programs. One such mechanism, promoter proximal pausing, is recognized as a general feature of transcription at many eukaryotic genes. However, there is interestingly a prominence of paused Pol II at signal-response genes, suggesting a process that primes these genes for quick transcription in response to stimuli[229-231].

Cis-regulatory DNA elements, called enhancers, play an essential role in regulating transcription[33]. These regulatory regions are identified using multiple assays, such as 1) regions of accessible chromatin[232-234]; 2) patterns of histone modifications[29,32]; and 3) genome-wide occupancy of transcription factors (TFs) and other coactivator proteins, such as p300[193,235,236]. Active enhancers are marked with specific histone modifications, namely histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac)[34]. Enhancers have

[2] Portions of this have been submitted for publication as Perreault et al, (2020).

also been classified as latent, poised, or repressive based on the changes in the patterns of these histone modifications[35,45,186]. Integration of these individual datasets can be used to generate a composite map of chromatin states that defines cell identity[30,42].

Genome-wide maps of enhancers have demonstrated that these cis-regulatory elements control transcription from long ranges in linear space. However, the 3D organization of the genome facilitates these interactions. Structural proteins, such as CCCTC-binding factor (CTCF) and Yin-Yang 1 (YY1), tether distal TF-bound enhancers to their target gene promoters. CTCF is an evolutionarily conserved zinc-finger TF that co-localizes with cohesin[61]. Together, these two factors establish and maintain chromatin loops[74,81,84]. Maps of chromatin contacts have been established genome-wide in many cell and tissue types using a variety of chromosome conformation capture (3C) based assays. Specifically, the use of HiC revealed that the genome is organized into active and inactive domains, which are delineated by CTCF and largely conserved between cell types[59,60]. These larger domains are further separated into topologically associated domains (TADs) and subTADs that contain higher interaction frequencies between regions of the genome, many of which are not limited to one-to-one[78,79,83-85]. Importantly, CTCF binding maintains enhancer-promoter (E-P) interactions and therefore cell-type specific gene expression[81,237]. Together, these findings support the function of CTCF as structural foci for chromatin organization whereby Pol II can selectively target cell-type-specific genes for transcription through interactions with looping factors and enhancers.

However, recent studies have suggested that YY1 may be a more precise indicator of E-P specific chromatin loops. YY1 is a ubiquitously expressed zinc-finger TF that plays an important role in cellular differentiation[87,88]. When the YY1 binding motif was deleted in a locus-specific manner using CRISPR-Cas9, Weintraub and colleagues found there was decreased YY1 binding

at the promoter, reduced contact frequency between the enhancer and promoter, and a decrease in mRNA levels[89]. These findings support the essential role of YY1 in controlling gene expression through facilitating E-P interactions. Although CTCF occupancy has been assessed in a variety of erythroid contexts (Dean, Higgs, Blobel, Gallagher), YY1 binding locations in erythroid cells have yet to be determined, resulting in a gap of important data for understanding E-P interactions during erythropoiesis.

The role of erythropoietin (Epo), the hormone that is required for terminal erythroid differentiation[101,102], in these regulatory interactions remains uncharacterized. Indeed, we have identified Epo-responsive enhancers in erythroid precursors[186], but have yet to investigate Epo's role in chromatin interactions. To address this gap in understanding, we leveraged an *ex vivo* murine cell system that undergoes synchronous erythroid maturation in response to Epo stimulation (**Figure 3.1**)[109,129,196]. Outside of erythropoiesis, very few studies have investigated the impact of hormone signaling on chromatin dynamics, highlighting the need for studies such as the one presented here.

Here, we show that there are acute transcriptional changes in proerythroblasts in response to Epo stimulation (**Figure 3.2**). There is a subset of genes that are significantly up- and down-regulated, which are associated with genes important in signal transduction, cell survival, and proliferation. YY1 binding has a dynamic response to Epo, as opposed to CTCF, which is comparatively stable during the same time period (**Figure 3.3**). Additionally, there is little overlap in the regions bound by both of these structural TFs. Using the novel chromosome conformation capture assay HiChIP, we identified chromatin contacts mediated by H3K27ac and YY1, revealing that although these loops have similar loop length and score, they delineate unique subsets of E-P interactions between enhancers and Epo-responsive genes (**Figure 3.8**).

Materials & methods

Isolation of proerythroblasts from FVA infected mice

Highly purified proerythroblasts were obtained from spleens of mice infected with the Friend virus as previously described[109,129], with the following modifications. All animal procedures were performed in compliance with and approval from the Vanderbilt Division of Animal Care (DAC) and Institutional Animal Care and Use Committee (IACUC). Female BALB/cJ mice (12 weeks old, Jackson Laboratories) were infected via intraperitoneal injection of ~10^4 spleen focus-forming units of Anemia-inducing strain of the Friend virus (FVA). At 13 to 15 days post-infection, the mice were sacrificed and spleens removed. The spleens were homogenized to a single cell suspension by passing the minced spleens through a sterile 100 micron nylon mesh filter into sterile solution of 0.2% bovine serum albumin (BSA) in 1x PBS. The filtrate was then repeatedly pipetted to ensure a single cell suspension. The homogenized spleen cells were size-separated by gravity sedimentation for 4 hours at 4°C in a continuous gradient of 1% to 2% deionized BSA. The sedimentation apparatus consisted of a 25cm diameter sedimentation chamber containing a 2.4L BSA gradient, two BSA gradient chambers containing 1.2L 1% and 2% deionized BSA in 1x PBS, and a cell loading chamber (ProScience Inc.) containing the 50ml cell suspension. After 4 hour sedimentation, cells were collected in 50ml fractions, with proerythroblasts typically enriched in fractions 5-20 of 24 total fractions. Typically about 10^9 proerythroblasts were obtained from the separation of 10^{10} nucleated spleen cells (6-7g spleen weight) across three 25cm sedimentation chambers.

Cell culture conditions

To study the effects of erythropoietin (Epo) on terminal erythroid differentiation, FVA-derived proerythroblasts were cultured at 10^6 cells/ml in Iscove-modified Dulbecco medium (IMDM, Life Technologies #12440043), 30% heat-inactivated fetal bovine serum (Gibco, 26140-079), 1% Penicillin-Streptomycin (Gibco #15140-122), 10% deionized BSA, and 100uM alpha-thioglycerol (MP Biomedicals #155723). Terminal erythroid differentiation of purified proerythroblasts was induced by the addition of 0.4 U/ml human recombinant Epo (10kU/ml Epogen by Amgen, NDC 55513-144-10) to media. At the desired times after the addition of Epo, cells were crosslinked by the addition of 1% formaldehyde for 10 minutes for ChIP analysis and 2% formaldehyde for 20 minutes for HiChIP analysis. Crosslinking was then quenched by the addition of 125mM glycine. Crosslinked cells were collected by centrifugation for 5 minutes at 1,000g at 4°C, washed once with 1x PBS, flash frozen in liquid nitrogen, and stored at -80°C until used. For RNA-seq, cells were removed from culture before crosslinking. Samples were spun for 5 minutes at 1,000g at 4°C and the supernatant was aspirated. Pellets were flash frozen in liquid nitrogen and stored at -80°C until used.

HiChIP

HiChIP was performed as described[170] with a few modifications. 50 million cell pellets were resuspended in 2.5ml ice cold Hi-C Lysis Buffer (10mM Tris HCl, 10mM NaCl, 0.2% NP-40, 1X protease inhibitors (Roche, 04693124001)) and split into 10 million cell amounts. Samples were incubated at 4°C for 30 minutes with rotation. Nuclei were pelleted by centrifugation at 2,500g for 5 minutes at 4°C and washed once with 500ul of ice cold Hi-C Lysis Buffer. After removing supernatant, nuclei were resuspended in 100ul of 0.5% SDS and

incubated at 62°C for 10 minutes. SDS was quenched by adding 285ul water and 50ul 10% Triton X-100. Samples were vortexed and incubated for 15 minutes at 37°C. After the addition of 50ul of 10X NEBBuffer 2 (NEB, B7002) and 1ul of MboI restriction enzyme (NEB, R0147), chromatin was digested at 37°C for 1 hour at 700rpm on Thermomixer. Following digestion, MboI enzyme was heat inactivated by incubating the nuclei at 62°C for 20 minutes. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 52ul of fill-in master mix, containing 15ul of 1mM biotin-dATP (Jena BioScience, NU-835-BIO14-L), 1.5ul of 10mM dCTP (NEB, N044_S), 1.5ul of 10mM dGTP (NEB, N044_S), 1.5ul of 10mM dTTP (NEB, N044_S), and 10ul of 5 U/ul DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210), was added and the tubes were incubated at 37°C for 1 hour at 700rpm on Thermomixer. Proximity ligation was performed by addition of 948ul of ligation master mix, containing 150ul of 10X NEB T4 DNA ligase buffer (NEB, B0202), 125ul of 10% Triton X-100, 15ul of 10 mg/mL BSA (NEB, B9000), 10ul of 400 U/mL T4 DNA ligase (NEB, M0202), and 648ul of water, and incubation at room temperature for 4 hours with rotation.

After proximity ligation, nuclei were pelleted by centrifugation at 2500g for 5 minutes and resuspended in 880ul Nuclear Lysis Buffer (50mM Tris HCl, 10mM EDTA, 1% SDS, 1X protease inhibitors (Roche, 04693124001)). Samples were vortexed and nuclei were sonicated with a Bioruptor (Diagenode) for 10 minutes on the low setting to solubilize chromatin. Sonicated chromatin was clarified by centrifugation at 16,100g for 15 min at 4°C and supernatant from 10 million cell samples are pooled to a total of 50 million cells. Sample was diluted with 2X ChIP Dilution Buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris HCl, 167mM NaCl). 300ul Protein A beads (Thermo, 21348) were washed in 2ml ChIP Dilution Buffer and resuspended in 250ul ChIP Dilution Buffer. Beads were added to 50 million

cell sample and incubated at 4°C for 1 hour with rotation. Beads were then separated on a magnetic rack and supernatant was transferred to a new tube. 10ug of antibody for Pol II (Santa Cruz, sc-17798), H3K27ac (Abcam, ab4729), or YY1 (Abcam, ab109237) were added to the tube. Samples were incubated overnight at 4°C with rotation. The next day, 300ul Protein A beads were washed in 2ml ChIP Dilution Buffer and resuspended in 500ul ChIP Dilution Buffer. Beads were added to 50 million cell sample with antibody and incubated at 4°C for 2 hours with rotation. Beads were then separated on a magnetic rack and washed three times with 750ul Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris HCl, 150mM NaCl), three times with 750ul High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris HCl, 500mM NaCl), and three times with 750ul LiCl Wash Buffer (10mM Tris HCl, 250mM LiCl, 1% NP-40, 1% Na-Doc, 1mM EDTA).

Beads were then resuspended in 200ul of DNA Elution Buffer (50mM $NaHCO_3$, 1% SDS), which is made fresh, and incubated at room temperature for 10 minutes with rotation, followed by 37°C for 3 minutes at 700rpm. Samples were placed on a magnetic rack and supernatant transferred to a new tube. This was repeated once more. 10ul of Proteinase K (Roche, 03115828001) was added to each tube and samples were incubated at 55°C for 45 minutes at 700rpm, followed by 67°C for 1.5 hours at 700rpm. DNA was then purified using Zymo DNA Clean and Concentrator (Zymo, D4003) according to manufacturer's protocol and eluted in 10ul water. The amount of eluted DNA was quantified by Qubit dsDNA HS kit (Invitrogen, Q32854).

25ul of Streptavidin C-1 beads (Invitrogen, 65001) were washed with 1ml Tween Wash Buffer (5MM Tris HCl, 0.5mM EDTA, 1M NaCl, 0.05% Tween-20) and resuspended in 10ul of 2X Biotin Binding Buffer (10mM Tris HCl, 1mM EDTA, 2M NaCl). 10ul of bead mixture was

added to 50ng of purified DNA for each sample, incubating at room temperature for 15 minutes, agitating every 5 minutes. After capture, beads were separated with a magnet and the supernatant was discarded. Beads were then washed twice with 500ul of Tween Wash Buffer, incubating at 55°C for 2 minutes at 700rpm. Beads were washed with 100ul 1X TD Buffer (diluted from 2X TD Buffer (20mM Tris HCl, 10mM $MgCl_2$, 20% Dimethylformamide)). Beads were resuspended in 50ul of master mix, containing 25ul 2X TD Buffer, 2.5ul Tn5 Tagment DNA enzyme (Illumina, 15027865), and 22.5ul water. Samples were incubated at 55°C for 10 minutes at 700rpm. Beads were separated on a magnet and supernatant was discarded. Beads were washed with 750ul of 50mMEDTA at 50°C for 30 minutes, washed twice with 750ul of 50mMEDTA at 50°C for 3 minutes each, then washed twice with 750ul of Tween Wash Buffer at 55°C for 2 minutes each, and finally washed once with 750uL of 10mM Tris HCl pH 7.5. Beads were separated on a magnet and supernatant was discarded.

To generate the sequencing library, PCR amplification of the tagmented DNA was performed while the DNA is still bound to the beads. Beads were resuspended in a PCR master mix, consisting of 36ul water, 1.25 unique Nextera Ad2.X primer, 10ul Phusion HF 5X buffer (NEB, E0553), 1ul 10mM dNTPs, 1.25ul universal Nextera Ad1 primer, and 0.5ul Phusion DNA Polymerase (NEB, E0553). DNA was amplified with 8 cycles of PCR. After PCR, beads were separated on a magnet and the supernatant containing the PCR amplified library was transferred to a new tube, purified using the Zymo DNA Clean and Concentrator (Zymo D4003) kit according to manufacturer's protocol and eluted in 52ul water. Purified HiChIP libraries were size selected to 300-700 basepairs using a double size selection with AMPure XP beads (Beckman Coulter, A68831). HiChIP libraries were paired-end sequenced on an Illumina NextSeq500 with reads 75 nucleotides in length.

HiChiP data analysis

HiChIP library sequence reads were aligned to the mouse mm10 reference genome using HiC-Pro[238] with the following options in the configuration file:

BOWTIE2_OPTIONS = --very-sensitive --end-to-end –reorder

LIGATION_SITE = GATCGATC

GET_ALL_INTERACTION_CLASSES = 1

GET_PROCESS_SAM = 1

RM_SINGLETON = 1

RM_MULTI = 1

RM_DUP = 0

Hichipper[239] was applied to HiC-Pro output files to identify high confidence chromatin contacts using EACH, ALL peak finding settings. The quickAssoc and annotateLoops functions in the diffloop R package[240] were used to find differential loops and annotate epigenetic features, respectively. Enhancers were denoted as the intersection of H3K4me1 and H3K27ac peaks (previously published data, Key Resources Table) and promoters were identified using the getMouseTSS function. To visualize chromatin interactions identified using HiChIP, the –make-ucsc option was added when analyzing the data using hichipper[239].

Chromatin Immunoprecipitation with Lambda Exonuclease Digestion (ChIP-exo)

With the following modifications, ChIP-exo was performed as previously described[152,153] with chromatin extracted from 50 million cells, ProteinG MagSepharose resin (GE Healthcare), and 10ug of antibody directed against Pol II (Santa Cruz, sc-17798), YY1 (Abcam, ab109237),

or CTCF (Millipore, 07-729). First, formaldehyde crosslinked cells were lysed with buffer 1 (50mM HEPES–KOH pH 7.5, 140mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100), washed once with buffer 2 (10mM Tris HCL pH 8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA), and the nuclei lysed with buffer 3 (10mM Tris HCl pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na–Deoxycholate, 0.5% *N*-lauroylsarcosine). All cell lysis buffers were supplemented with fresh EDTA-free complete protease inhibitor cocktail (CPI, Roche #11836153001). Purified chromatin was sonicated with a Bioruptor (Diagenode) to obtain fragments with a size range between 100 and 500 base pairs. Triton X-100 was added to extract at 1% to neutralize sarcosine. Insoluble chromatin debris was removed by centrifugation, and sonication extracts stored at -80°C until used for ChIP analysis. Libraries were sequenced using an Illumina NextSeq500 sequencer as single-end reads 75 nucleotides in length on high output mode. To assess reproducibility of biological replicates, Pearson's correlation was calculated.

ChIP-exo data analysis

ChIP-exo library sequence reads were aligned to the mouse mm10 reference genome using BWA-MEM algorithm[241] using default parameters. The resulting bam files were first sorted using the Samtools Sort function[242], and then bam index files were generated using the Samtools Index function[242]. Since patterns described here were evident among individual biological replicates, and replicates were well correlated, we merged all tags from biological replicate data sets for final analyses.

ChIP-exo peaks were annotated and quantified using the Hypergeometric Optimization of Motif EnRichment (HOMER) suite[192]. Briefly, bam files were converted to tag directories using the makeTagDirectory function with the –genome, –checkGC, and –format options. The

findPeaks function was used to identify ChIP peaks using –o auto and –style gro-seq or factor for Pol II or CTCF/YY1 libraries, respectively. To quantify and normalize tags to RPKM, the analyzeRepeats function was used with the –rpkm, –count genes, –strand both, –condenseGenes, and –d options.

bigWig files for CTCF and YY1 libraries were generated using the deepTools bamCoverage function[188]. To create aligned heatmaps, first a matrix was generated using the computeMatrix function with the following options: reference-point –S, –a 2000, –b 2000, -–referencePoint center, –verbose, –missingDataAsZero, and –p max/2. Then, the heatmap was created using the plotHeatmap function with the following options: –verbose and –sortRegions descend.

Raw sequencing tags were smoothed (20 basepair bin, 100 basepair sliding window) and normalized to reads per kilobase per million (RPKM) using deepTOOLS[188] and visualized with Integrative Genomics Viewer (IGV)[189].

RNA-seq

RNA was isolated using the Qiagen RNAeasy kit (Qiagen, 74104) per manufacturer's instructions. Stranded polyA selected libraries were prepared using NEBNext PolyA mRNA isolation standard protocol, NEBNext rRNA Depletion standard protocol, and finally NEBNext Ultra II Directional DNA library preparation kit (Illumina, E75530S) per manufacturer's protocol. PCR amplified RNA-seq libraries were size selected using AMPure XP beads (Beckman Coulter, A68831). RNA-seq libraries were subjected to 75 basepair single end sequencing on Illumina NextSeq500 sequencer.

RNA-seq data analysis

RNA-seq library sequence reads were aligned to the mouse mm10 reference genome using TopHat[243] using default parameters. Cufflinks[244] was used to assemble transcripts and quantify expression of transcripts. Cuffmerge[244] merges all transcript assemblies to create a single merged transcriptome annotation for final analyses. CummeRbund visualizes RNA-seq data analyzed using cufflinks.

Data Availability

The accession number for the data reported in this paper is GEO: GSE142006.

Results

The FVA murine system faithfully recapitulates erythroid differentiation during erythropoiesis

We are able to study the molecular effect of Epo on gene expression and genome architecture using the unique anemia-inducing strain of the Friend virus (FVA) murine model system. In this system, we inject mice with FVA. The anemia-inducing Friend virus glycoprotein, gp55A, directly interacts with the naturally occurring truncated form of the Stk receptor tyrosine kinase (sf-Stk) receptor to induce proerythroblast proliferation in the absence of Epo. After approximately 14 days, we harvest spleens from the mice, enabling the large-scale procurement of highly purified murine proerythroblast. Once cultured ex-vivo with Epo, these cell populations synchronously respond to Epo to fully differentiate to mature erythrocytes (**Figure 3.1A**)[129].

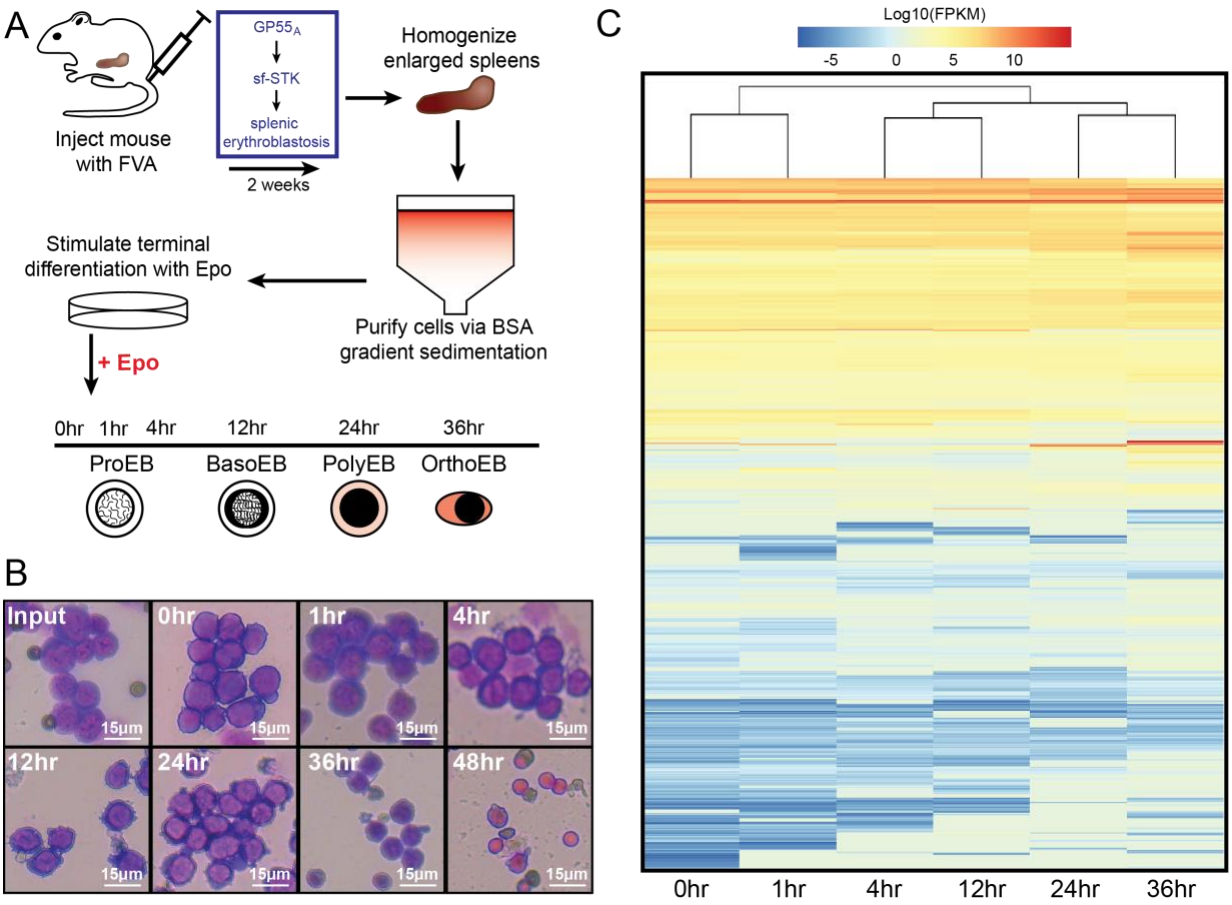

Figure 3.1. The FVA murine system faithfully recapitulates erythroid differentiation during erythropoiesis. (**A**) The workflow for generating and isolating highly purified Epo-responsive proerythroblasts from a mouse injected with the Friend Virus that induces Anemia (FVA). (**B**) Microscopy images highlighting morphological changes of proerythroblasts isolated using the FVA system during differentiation. (**C**) Heatmap of RNA-seq gene expression through erythroid differentiation.

Indeed, we see a dramatic shift in size and shape of maturing erythroid precursors during erythropoiesis. Before purification, there is a heterogeneous population of cells (**Figure 3.1B**, Input). After purification, there is a uniform population of proerythroblasts (ProEB), which are large, round cells, and this morphological stage persists until approximately 12 hours of Epo stimulation (**Figure 3.1B**). After 24 hours in the presence of Epo, cells are polychromatic erythroblasts (PolyEB), characterized by the accumulation of hemoglobin, which coincides with

the continued increase in globin gene expression. Finally, after about 48 hours of ex vivo culture in Epo, the precursors are terminally differentiated into reticulocytes, identified by high hemoglobin production and nucleus extrusion (**Figure 3.1B**, 48hr). Importantly, recent work has used flow cytometry and the unique cell surface markers expressed at each of these stages to isolate populations of cells[99,245]. The FVA model system has been shown to recapitulate normal erythropoiesis through a variety of validation experiments conducted by Koury and colleagues, as well as others, including globin expression kinetics, cell morphology, cell viability, and enucleation, which are all essential for mature erythrocytes[101,102,109,196]. These morphological changes are accompanied by changes in gene expression programs (**Figure 3.1C**).

Epo stimulation results in acute transcriptional changes in proerythroblasts

With the knowledge that erythropoiesis enacts different gene expression programs based on the cellular stage in the differentiation process on a large time scale, we wanted to investigate the immediate transcriptional response to Epo stimulation in proerythroblasts. The FVA murine system is ideal to study this question because naïve proerythroblasts are only exposed to low levels of Epo within the mouse before isolation. To assess the effects of hormone stimulation on acute transcriptional changes, we performed ChIP-exo on RNA polymerase II (Pol II) after one hour of Epo stimulation. In comparing ChIP-exo signal pre and post Epo stimulation, we see that Pol II occupancy is highly correlated (**Figure 3.2A**). This indicates that Pol II occupancy does not change dramatically after the short exposure to Epo.

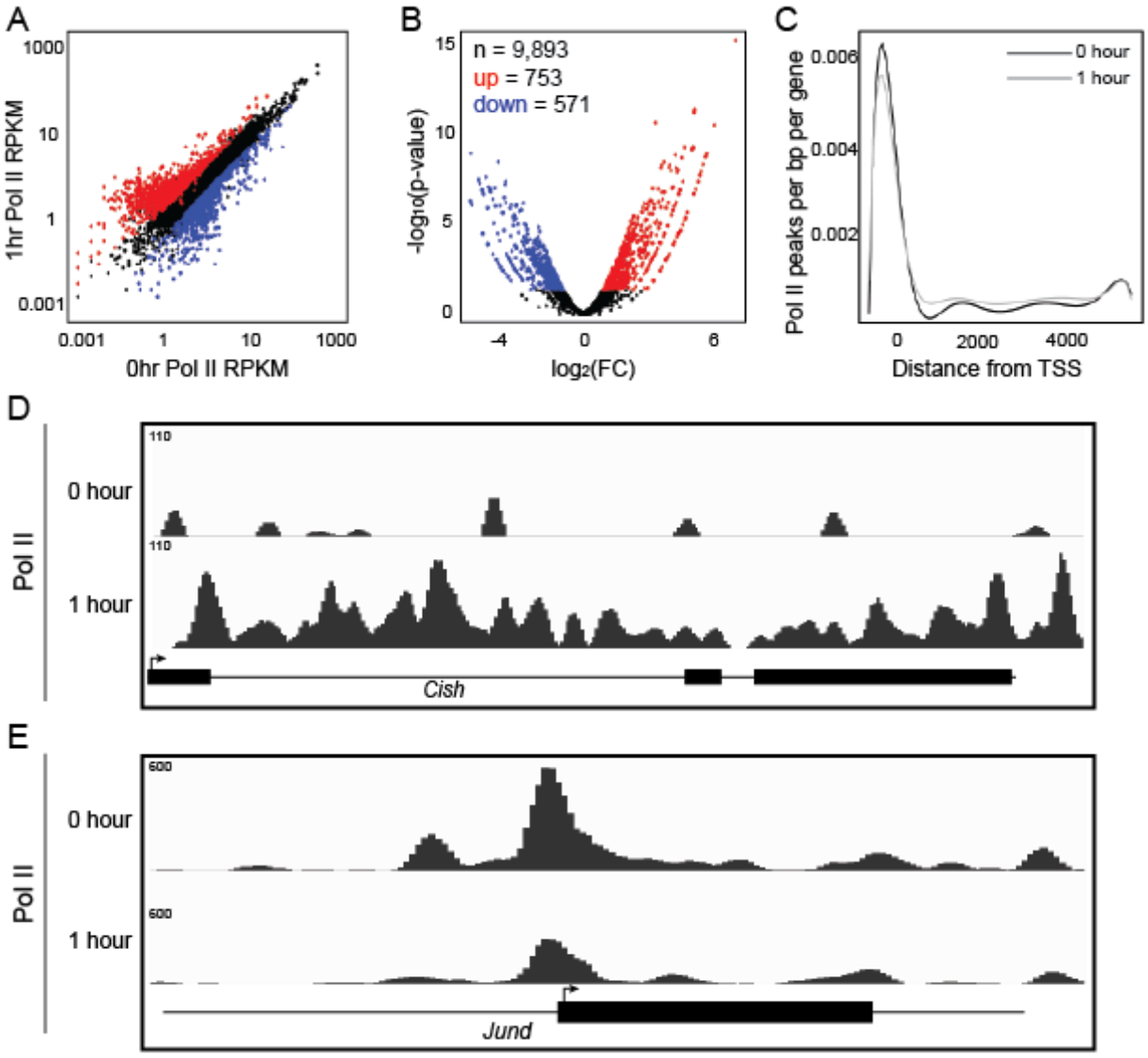

Figure 3.2. Epo stimulation results in acute transcriptional changes in proerythroblasts. (**A**) Scatterplot comparing Pol II RPKM before and after 1 hour Epo stimulation. (**B**) Volcano plot showing significant (p-value < 0.05) differential occupancy of increased (red) and decreased (blue) Pol II after 1 hour Epo stimulation. (**C**) Metagene plot comparing the position of Pol II peaks relative to transcription start site (TSS) (paired Wilcoxon ranked-sign test, p = 4.882 x 10-11). Genome browser view of ChIP-exo signal for Pol II at the upregulated *Cish* locus (**D**) and down regulated *Jund* locus (**E**).

However, when we investigate the fold change of Pol II signal between pre and post Epo stimulation, we find that there is significant differential occupancy of Pol II at a subset of genes (**Figure 3.2B**). This reveals a subset of Epo-responsive genes that are either upregulated (red) or downregulated (blue) after one hour of Epo stimulation.

Furthermore, when we investigate the localization of Pol II across the gene, we find that Pol II is more abundant at the transcription start site (TSS) of genes before Epo stimulation (**Figure 3.2C**). Pol II then transitions away from the TSS and towards the body of the gene, which supports an acute transcriptional response to Epo. We can see the dynamics of Pol II at both an upregulated (*Cish*) and downregulated (*Jund*) genes after one hour of Epo stimulation (**Figure 3.2D, E**). Cytokine Inducible SH2 Containing Protein (Cish) protein contains both a SH2 domain, which aids in signal transduction of receptor tyrosine kinase pathways, and a SOCS box domain, which has proposed function in suppressing cytokine signaling. *Cish* is a target of the JAK-STAT signaling pathway, which is induced through Epo binding to its receptor[246,247]. Literature details the integral function of Cish's components, SH2 and SOCS, in proper erythropoiesis[248-250]. *Jund* is a member of the Jun family and is part of a subunit of the AP1 TF complex, which is critical in gene expression regulation in response to cytokine, growth factor, stress, and infections in a variety of cellular contexts[251,252]. The importance of *Cish* and *Jund* in signaling mediated differentiation supports the biological significance of Epo-responsive genes as presented here.

Epo dynamically regulates YY1 occupancy genome-wide

Next, we examined the global ChIP-exo patterns for CTCF and YY1. The ChIP-exo signal for each structural TF before and after Epo treatment (0 and 1 hour, respectively) was aligned peaks after Epo stimulation and displayed as a heatmap (**Figure 3.3A, C**). CTCF binding is somewhat affected by Epo treatment, as shown by the quantification in the composite plot below the heatmap (**Figure 3.3B**). Although CTCF binding varies from cell type to cell type[61,253,254], the largely invariant binding in a population of cells that are the same type is

supported by recent literature detailing the invariant nature of domain boundaries delineated by CTCF to decrease variability in cell-to-cell gene expression[81].

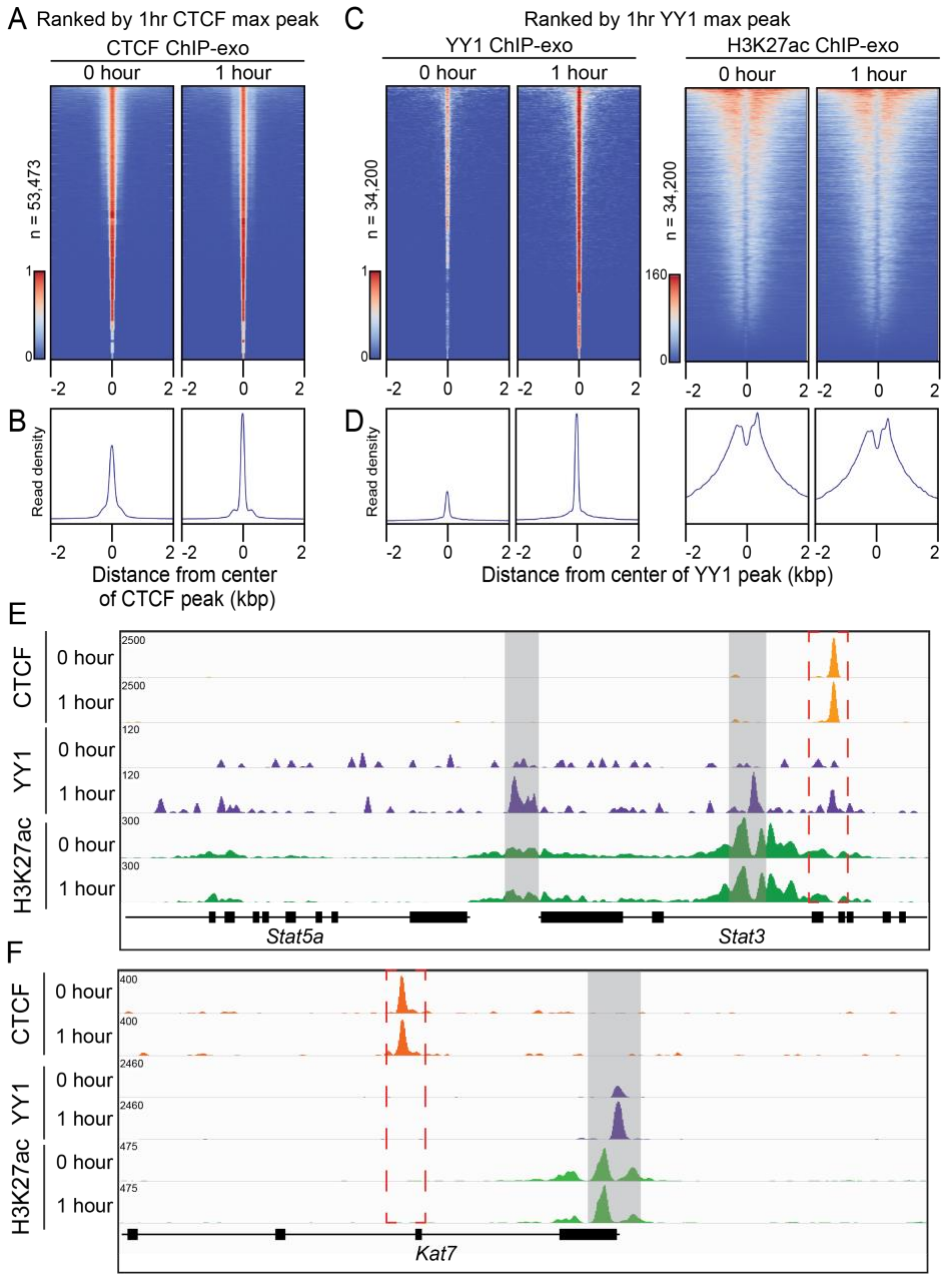

Figure 3.3. Epo dynamically regulates YY1 occupancy genome-wide. (**A**) Heatmap of CTCF peaks pre and post Epo stimulation, ranked by 1 hour CTCF max peak. (**C**) Heatmap of YY1 and H3K27ac peaks pre and post Epo stimulation, ranked by 1 hour YY1 max peak. (**B**) and (**D**) Composite plots below each heatmap quantifying the normalized tag density. (**E-F**) Representative genome browser view of CTCF, YY1, and H3K27ac occupancy in response to Epo stimulation, highlighted in light gray bars and red dashed box.

Alternatively, YY1 is highly dynamic (**Figure 3.3C, D**). This suggests that YY1 occupancy is influenced by hormone stimulation and could be more intricately involved in the remodeling of domains during cellular differentiation. Additionally, when YY1 is compared to H3K27ac, a histone modification canonically associated with activated regions of the genome involved in gene regulation, YY1 is much more dynamic within the short time period of Epo treatment (**Figure 3.3C, D**).

With the new understanding that YY1 has a dynamic response to Epo, we wanted to understand its localization in the genome and compare this to CTCF. In Epo-naïve proerythroblasts, the majority of YY1 is in intergenic regions (48%, **Figure 3.4A**). However, this shifts after Epo stimulation, where the majority of YY1 is in intronic regions (42%, **Figure 3.4A**). There is also a marked increase in YY1 at TSS of genes, from 5% before Epo to 17% after stimulation (**Figure 3.4A**). This suggests YY1 may have an important role in the transcriptional, and therefore gene expression, changes that occur during erythropoiesis in response to Epo.

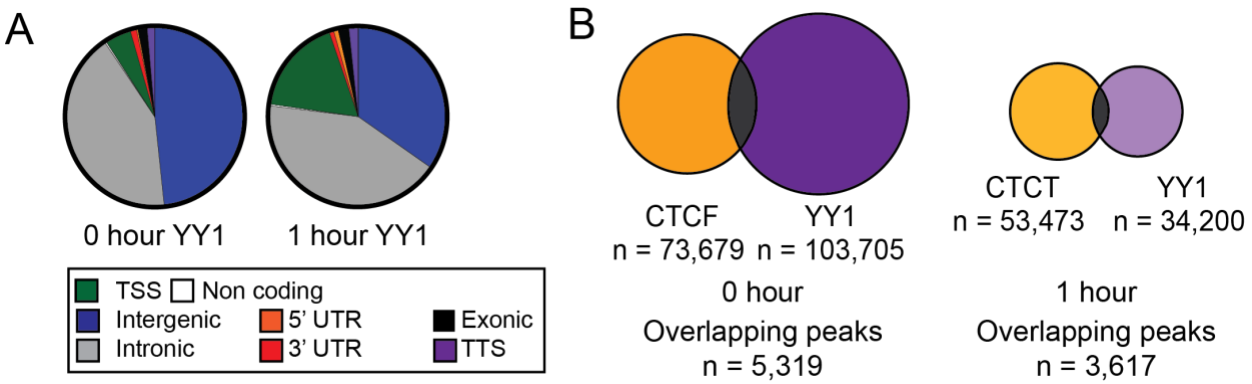

Figure 3.4. Characterization of CTCF and YY1 occupancy genome-wide. (**A**) YY1 binding locations in the genome. (**B**) Comparison of CTCF and YY1 peak overlap before and after 1 hour Epo stimulation.

We then wanted to understand if CTCF and YY1 were co-localized. We found that the regions of the genome bound by CTCF, termed domain boundaries, are not greatly affected by Epo stimulation in the short term (**Figure 3.3A, C**). However, the regions of the genome bound by YY1 are dynamic (**Figure 3.3C, D**). Out of the CTCF and YY1 bound regions pre stimulation, only 5,319 regions were bound by both TFs (Figure 3.4B, 7% and 5%, respectively. This small overlap in peaks was not greatly altered by Epo stimulation, where 6% of CTCF and 10% of YY1 peaks were shared between the two TFs (**Figure 3.4B**). This suggests that the chromatin domains established by CTCF and YY1 are unique and these structural proteins may have unique functions in delineating chromatin architecture. In **Figure 3.3E** and **3F**, the localization and dynamics of CTCF compared to YY1 are illustrated at the *Stat5a* and *Stat3* and *Kat7* loci, respectively. Interestingly, non-dynamic H3K27ac is found in regions of both increased YY1 and stable CTCF. The JAK-STAT signaling pathway is essential for erythropoiesis and is initiated by the hormone Epo, which is necessary and sufficient for terminal erythroid differentiation[100,219,255]. Lysine Acetyltransferase 7 (Kat7), is a subunit of the HBO1 complex, which has histone H4-specific acetyltransferase activity[256]. Due to its acetyltransferase functions, the HBO1 complex is involved in transcriptional activation through chromatin reorganization[257]. This suggests that YY1 dynamism is found at loci that are involved in signal transduction and chromatin modification.

Epo regulates transcription in a pre-established chromatin conformation

Recent literature has implicated YY1 in chromatin domains that specifically contact enhancers and their target genes[88,89]. Additionally, chromatin interactions mediated by H3K27ac have been studied in a variety of cellular contexts to understand active regions of the

genome[170,171,258,259]. We therefore wanted to investigate domains defined by YY1 and H3K27ac using HiChIP, a novel chromosome conformation capture assay developed by Mumbach and colleagues[170]. To observe intrachromosomal interactions mediated by H3K27ac (**Figure 3.5**) and YY1 (**Figure 3.6**) at a variety of resolutions, chromatin maps were generated and visualized using Juicer[260]. We further refined the chromatin interactions using hichipper, a program designed by Lareau and colleagues specifically to analyze HiChIP data, which has a unique data structure that cannot be leveraged by simply using programs developed for HiC or other higher order 3C assays[239]. Because of this definition, HiChIP anchors have a wide range of lengths. However, the average anchor lengths for H3K27ac and YY1 HiChIP libraries pre and post stimulation is approximately 4 kb (**Figure 3.7A**). The average loop lengths for H3K27ac and YY1 HiChIP libraries pre and post stimulation is approximately 317 kb (**Figure 3.7B**). Since chromatin interactions greater than 2 Mb are rare, we limit our analysis to the interactions up to this distance.

Using diffloop, we identified 151,032 H3K27ac and 138,210 YY1 chromatin interactions that were shared between pre and post Epo treatment conditions[239]. Loop scores are defined as the number of paired end tags (PETs) that support the interaction. The majority of these loops had a loop score less than 5 (**Figure 3.7C, D**). This suggests that the chromatin landscape in early erythroid precursors is delineated by weak interactions, potentially due to the upcoming shift in chromatin domains that initiates transcriptional changes.

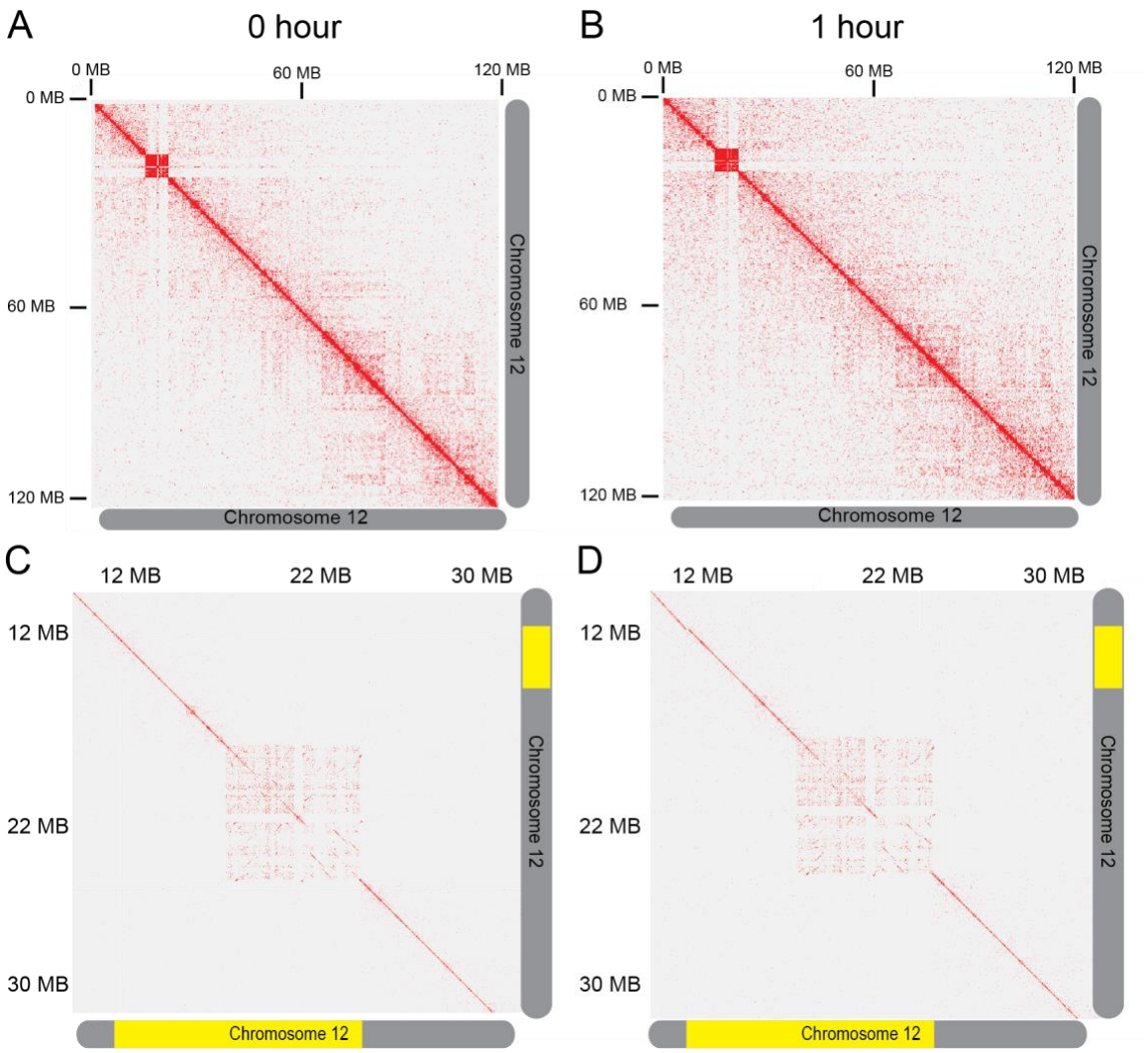

Figure 3.5. Chromatin contact maps for H3K27ac HiChIP. (**A**) Chromatin contacts at 0 hour Epo on chromosome 12 at 250KB resolution. (**B**) Chromatin contacts at 1 hour Epo on chromosome 12 at 250KB resolution. (**C**) Chromatin contacts at 0 hour Epo on chromosome 12 at 25KB resolution. (**D**) Chromatin contacts at 1 hour Epo on chromosome 12 at 25KB resolution.

To investigate the biological relevance of the identified loops to erythroid precursors, we conducted de novo motif discovery analysis of all HiChIP anchors. Strikingly, we found an enrichment of consensus motifs for important erythroid and structural TFs (**Figure 3.8A**). The *SMAD* family act as signal transducers for TGFβ receptors, which are critically important for regulating cell development and growth (P-value = 10^{-197} and 10^{-131} for H3K27 and YY1,

respectively)[261]. Interestingly, *Smad* was found to be critical in primitive erythropoiesis by maintaining Gata1 expression[262]. The proteins in the *STAT* family, as described previously, are critical in signal transduction and transcriptional activation (P-value = 10^{-185})[100,255]. The *KLF* family regulate important cellular functions, such as proliferation and apoptosis (P-value = 10^{-44}). Specifically, *Klf1*, which is a master regulator of erythropoiesis, is a transcriptional activator and is uniquely expressed in erythrocytes and megakaryocytes[115,119,263]. The *GATA* family members are involved in a variety of cellular processes, mainly differentiation and development (P-value = 10^{-195})[264]. *Gata1* is considered a master regulator of erythropoiesis as it is required for primitive and definitive erythropoiesis[115,264,265]. The *ETS* family is involved in a wide variety of cellular functions, including the regulation of cellular differentiation, cell cycle control, cell migration, cell proliferation, apoptosis, and angiogenesis (P-value = 10^{-44}). Studies have shown Ets and Ets-related TFs to be important in lineage commitment and proliferation[266,267]. The enrichment of these TFs suggests that the identified chromatin interactions have biological significance to early erythroid development.

The 151,032 H3K27ac and 138,210 YY1 chromatin interactions identified are classified into enhancer-promoter (E-P), enhancer-enhancer (E-E) or none classes based on epigenetic annotation. 23,423 H3K27ac and 21,777 YY1 loops were classified as E-P, which is approximately 16% for each library. Recent literature has shown that transcriptional response to external stimulation occurs within a pre-established chromatin landscape[82,90-92]. Therefore, we investigated the invariant loops in proerythroblasts to test if erythroid precursors also respond to Epo stimulation within a defined chromatin context. We define an invariant loop as a chromatin interaction whose fold change between loop scores in between -2 and 2. Using this definition, we analyzed 16,698 and 1,144 H3K27ac and YY1 E-P loops, respectively.

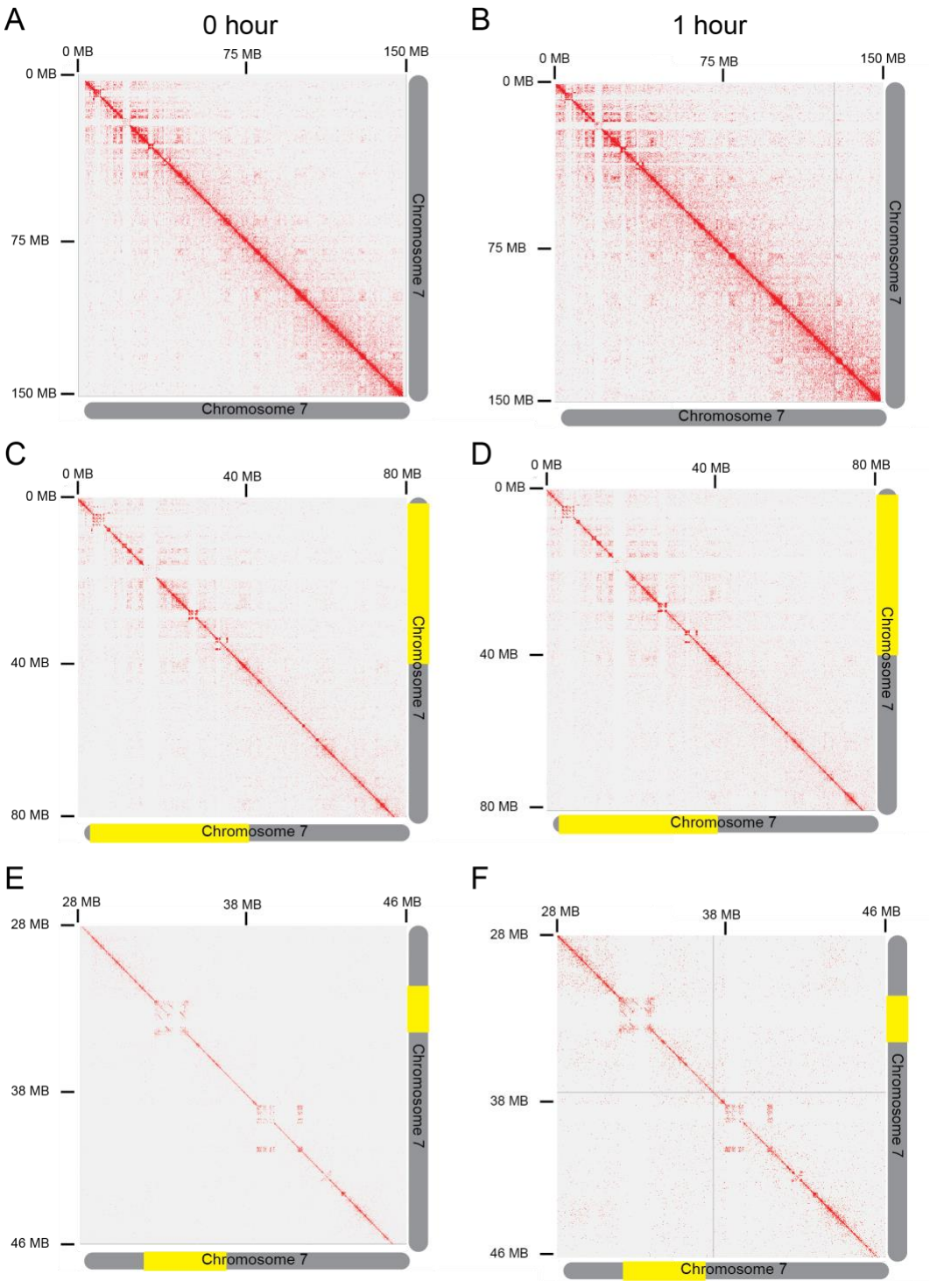

Figure 3.6. Chromatin contact maps for YY1 HiChIP. (**A**) Chromatin contacts at 0 hour Epo on chromosome 7 at 250KB resolution. (**B**) Chromatin contacts at 1 hour Epo on chromosome 7 at 250KB resolution. (**C**) Chromatin contacts at 0 hour Epo on chromosome 7 at 100KB resolution. (**D**) Chromatin contacts at 1 hour Epo on chromosome 7 at 100KB resolution. (**E**) Chromatin contacts at 0 hour Epo on chromosome 7 at 25KB resolution. (**F**) Chromatin contacts at 1 hour Epo on chromosome 7 at 25KB resolution.

Since we found that there is an acute transcriptional response to Epo (**Figure 3.2**) and changes in YY1 occupancy (**Figure 3.3**), we wanted to understand how these dynamics could occur within invariant chromatin architecture. To do this, we studied the location of HiChIP anchors related to specific transcriptomic and epigenetic regions, as described in the diagram in **Figure 3.8B**. First, as an additional form of validation, we compared UCSC annotated mm10 TSS and HiChIP anchors of E-P loops. We would expect 50% of the anchors to be found in TSS regions if each promoter was connected to one enhancer in the E-P loop. Indeed, half of the anchors of chromatin interactions mediated by H3K27ac were found in annotated TSSs (50%, **Figure 3.8C**). However, 73% of anchors of chromatin interactions mediated by YY1 were found in annotated TSSs, indicating that more anchors are found at promoters compared to enhancer regions (**Figure 3.8C**). This could suggest that a single enhancer regulates the transcription of multiple target genes. This also supports the need to investigate chromatin interactions mediated by specific proteins, as the regions these proteins encapsulate could have unique regulatory features, such as connectivity. We also conducted a null test, where we overlapped HiChIP anchors with randomly generated regions, which resulted in a low percent overlap (**Figure 3.8C**, light bars).

Knowing that the E-P interactions have anchors in promoters, we wanted to connect these loops to changes in transcription as described in **Figure 3.2**. Using the subset of Epo-responsive genes (n = 1,462), we found that interactions mediated by H3K27ac were in higher concordance than interactions mediated by YY1 (**Figure 3.8D, Figure 3.9A**). 50% of H3K27ac anchors were in promoters of Epo-responsive genes, where only 6% of YY1 anchors were in promoters of Epo-responsive genes. We studied this further by breaking all Epo-responsive genes into upregulated (red, n = 752) and downregulated (blue, n = 709) genes. As expected, fewer

downregulated gene promoters were in H3K27ac anchors. The presence of H3K27ac in downregulated promoters is consistent with the finding that this activating histone mark lags behind changes in transcription. **Figure 3.4E** shows a representative example of this overlap at the *Fadh1* gene, which is intricately involved in mitochondrial activity and metabolism[268].

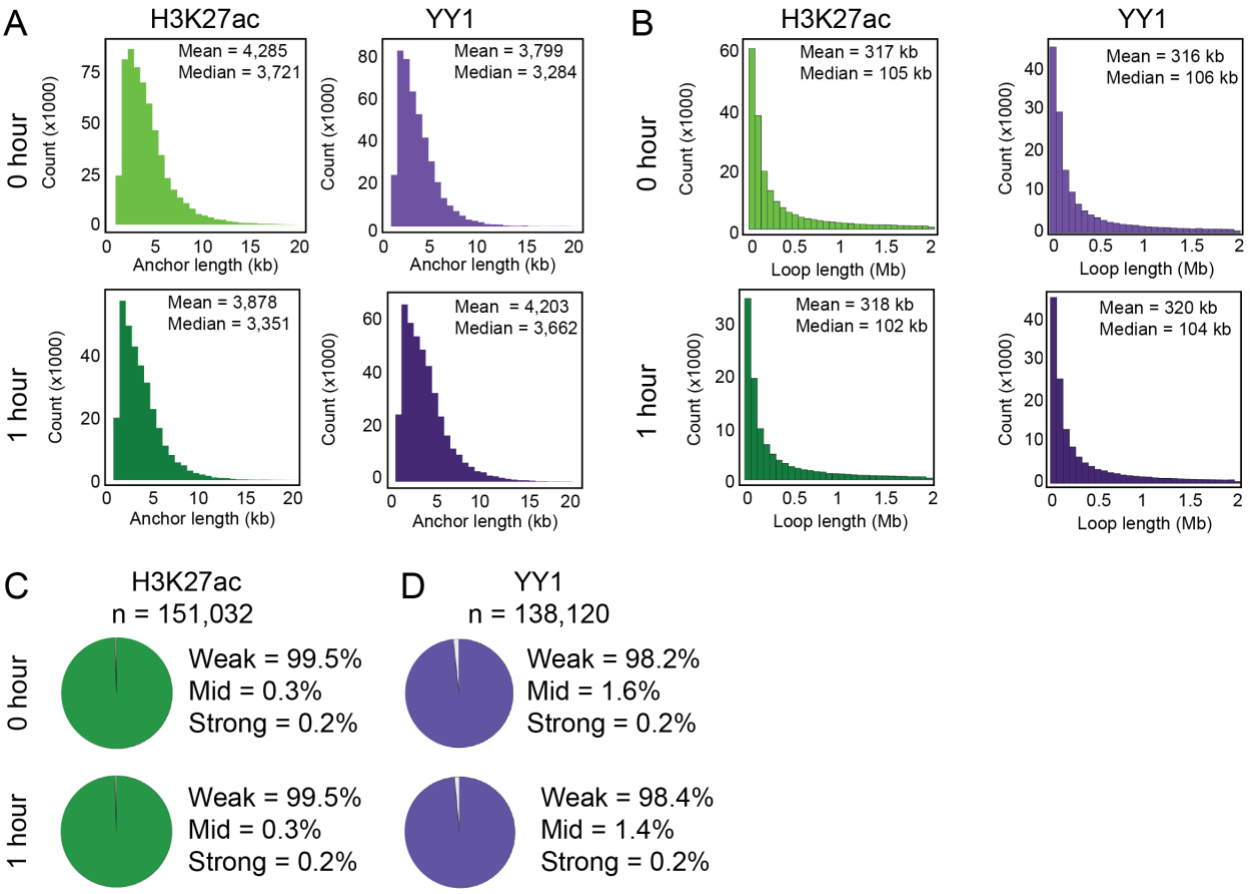

Figure 3.7. Characterization of chromatin loops mediated by H3K27ac and YY1. (**A**) Histogram showing the size distribution of anchors for HiChIP loops in H3K27ac and YY1 libraries pre and post Epo stimulation. (**B**) Histogram showing the size distribution of HiChIP loops in H3K27ac and YY1 libraries pre and post Epo stimulation. (**C**) Fraction of weak (score < 5), mid (score between 5 and 10), and strong (score > 10) H3K27ac chromatin interactions. (**D**) Fraction of weak (score < 5), mid (score between 5 and 10), and strong (score > 10) YY1 chromatin interactions.

As described in **Figure 3.3**, H3K27ac is relatively invariant in response to Epo as compared to YY1 peaks, which are highly variant. Therefore, we wanted to understand how the dynamics of H3K27ac and YY1 ChIP peaks related to invariant chromatin loops. In invariant

H3K27ac loops, anchors were enriched at loci with differential YY1 peaks (**Figure 3.8F**). Additionally, invariant H3K27ac loops were associated with invariant H3K27ac ChIP peaks (**Figure 3.8F**). **Figure 3.8G** shows a representative example of this overlap at the *Cdkn1b* gene, which controls cell cycle progression. In fact, *Cdkn1b* and its associated long noncoding RNA *Lockd* have been studied in erythroid cells, revealing the *Lockd* gene positively regulates *Cdkn1b* transcription[269]. Interestingly, a similar pattern was seen in invariant YY1 chromatin loops, whose anchors were also enriched at loci with differential YY1 peaks (**Figure 3.9B**).

With these individual findings, we wanted to understand how transcription, TF binding, and chromatin loops all interact in gene regulation. To do this, we looked at anchors in promoters of Epo-responsive genes with differential YY1 peaks. In concordance with the findings in **Figure 3.8D**, H3K27ac HiChIP anchors were more enriched in promoters of Epo-responsive genes with differential YY1 peaks than YY1 HiChIP anchors (**Figure 3.8H**), as opposed to interactions mediated by YY1 (**Figure 3.9C**). **Figure 3.8I** shows a representative example of this overlap at the *Supt4a* gene, which encodes the SPT4 protein, a component of the DSIF elongation complex, implicating this loci in transcriptional regulation[270,271]

Discussion

We set out to understand the gene regulatory mechanisms governing transcriptional, epigenetic, and chromatin dynamic response to Epo stimulation during early erythroid differentiation. First, we present the diverse gene expression programs induced by Epo stimulation during erythropoiesis in the FVA murine model system. We further investigate the acute transcriptional dynamics that occur within one hour of Epo treatment, a unique insight that allows for study of the direct impact of Epo on Pol II.

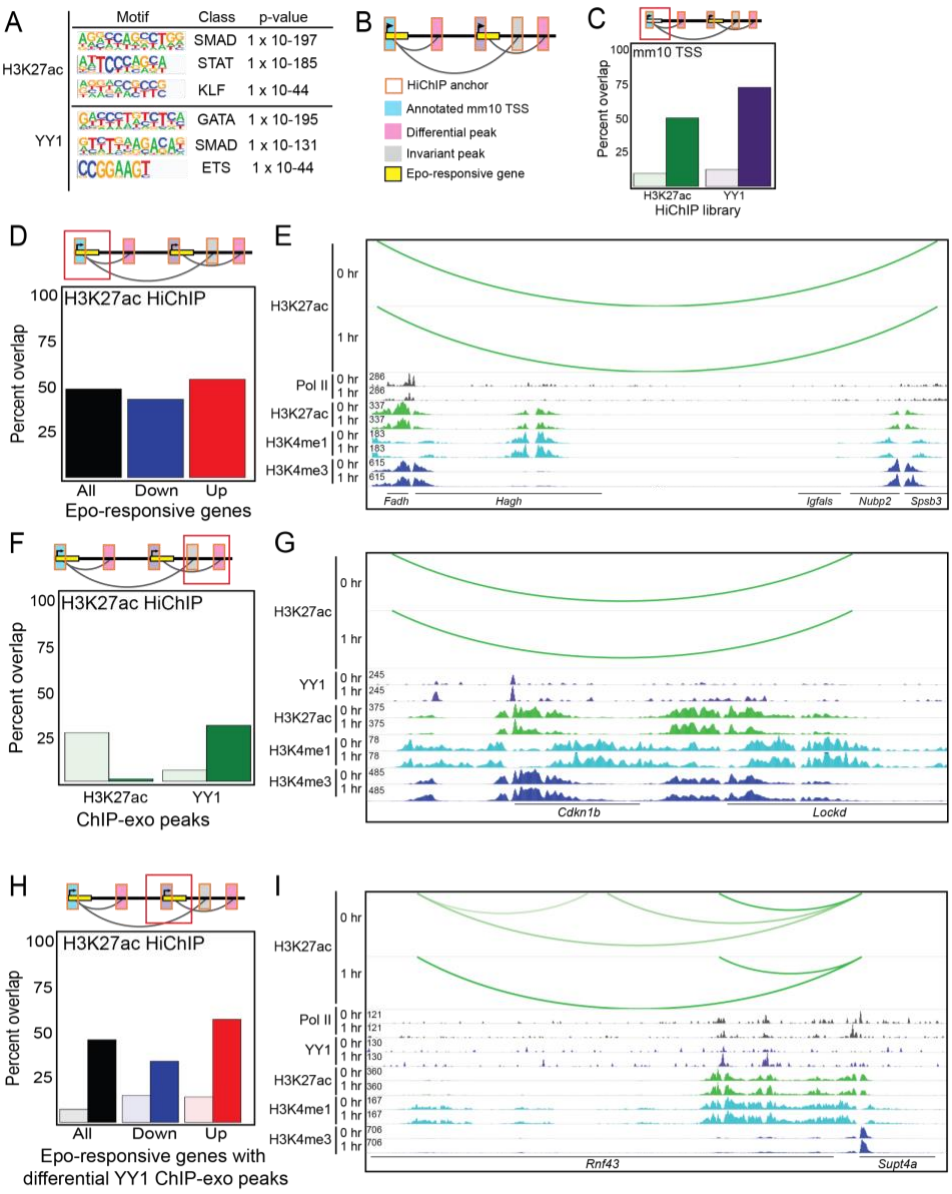

Figure 3.8. Epo regulates transcription in a pre-established chromatin conformation mediated by H3K27ac. (**A**) TF binding motifs overrepresented in HiChIP loop anchors. (**B**) A schematic of invariant chromatin landscape. (**C**) Proportion of interactions with annotated mm10 TSS within their anchor regions. Dark bars represent annotated mm10 TSS and light bars represent randomly generated sequences. (**D**) Proportion of interactions with promoters of Epo-responsive genes within H3K27ac HiChIP anchor regions. (**E**) Representative genome browser view of overlap described in (D). (**F**) Proportion of interactions with differential H3K27ac or YY1 ChIP-exo peaks within anchor regions of H3K27ac HiChIP. Dark bars represent differential ChIP-exo peaks TSS and light bars represent invariant ChIP-exo peaks. (**G**) Representative genome browser view of overlap described in (F). (**H**) Proportion of interactions with differential YY1 ChIP-exo peaks at promoters of Epo-responsive genes within anchor

regions of H3K27ac HiChIP. Dark bars represent Epo-responsive genes and light bars represent non-responsive genes. (**I**) Representative genome browser view of overlap described in (H).

We find that there is a subset of genes that are significantly up- and down-regulated by Epo, as defined by differential Pol II occupancy, within this short time frame, despite the similarity in overall gene expression by RNA-seq. Next, we observe the non-dynamic occupancy of CTCF juxtaposed by the highly dynamic occupancy of YY1 in response to Epo. Finally, we use the novel chromosome conformation capture assay HiChIP to identify chromatin interactions mediated by H3K27ac and YY1, finding a subset of (E-P) interactions that are not altered in response to Epo. Our data supports a model in which enhancers regulate transcription of target genes within established chromatin domains.

It is known that there is a dramatic shift in gene expression during erythropoiesis. Although several studies have investigated the changes between primitive and definitive erythropoiesis[176,272] and the regions of open chromatin during erythropoiesis that indicate an environment conducive to transcription[273], the gene expression profiles for the distinct cellular stages of definitive erythropoiesis had yet to be identified. Leveraging the FVA murine model system, we were able to isolate erythroid precursors at the distinct erythroid cell stages and generate a time course of gene expression profiles with RNA-seq (**Figure 3.1**). Studying the gene expression dynamics during erythropoiesis could provide insights to understanding the deregulation of these normal programs in disease.

Despite the similarity in gene expression between naïve proerythroblasts and those that have been exposed to Epo for one hour, there are significant changes in Pol II occupancy and dynamics (**Figure 3.2**). This was somewhat unexpected, as we hypothesized that genes with differential gene expression would also have differential Pol II occupancy. We surmise this is due to the early time points investigated in the present study, as well as their proximity to one

another, such that the transcriptional machinery has a quicker response time to stimuli than the processes of transcription and translation together to generate a fully processed, stable transcripts. An area of further study would be to conduct Pol II ChIP-exo at the remaining time points, creating a parallel set of data to the RNA-seq presented here. This would enable the comparison of transcription and overall gene expression, providing a method to understand gene regulatory mechanisms related to transcriptional response during erythropoiesis.

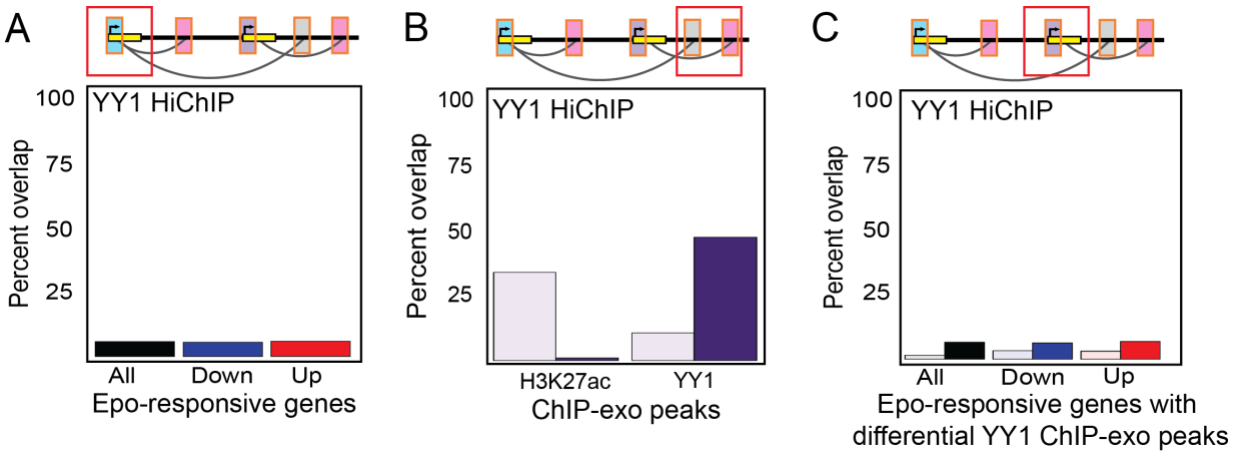

Figure 3.9. Epo regulates transcription in a pre-established chromatin conformation mediated by YY1. (**A**) Proportion of interactions with promoters of Epo-responsive genes within YY1 HiChIP anchor regions. (**B**) Proportion of interactions with differential H3K27ac or YY1 ChIP-exo peaks within anchor regions of YY1 HiChIP. Dark bars represent differential ChIP-exo peaks TSS and light bars represent invariant ChIP-exo peaks. (**C**) Proportion of interactions with differential YY1 ChIP-exo peaks at promoters of Epo-responsive genes within anchor regions of YY1 HiChIP. Dark bars represent Epo-responsive genes and light bars represent non-responsive genes.

CTCF has been extensively studied in a variety of cell types, as well as during several differentiation model systems. Indeed, it has been credited as the structural protein that delineates TADs, which shift during development. We find in the present study that CTCF is not very dynamic (**Figure 3.3**). This seems unexpected, but is supported by the literature because the time points investigated in the current study are both in the proerythroblast stage of erythropoiesis. Additionally, there was a decrease in CTCF peaks after Epo stimulation,

supporting the idea of selective pruning of CTCF binding sites during development[88]. An interesting avenue of further study would be to identify CTCF binding locations throughout the FVA time course, establishing sequential CTCF maps during erythroid maturation. This would provide unique insights to the fields of erythroid biology, chromatin organization, and response to stimuli since the majority of erythroid-related studies cannot isolate the distinct cell stages with the ease and purity of the FVA model system.

Although YY1 has been studied since the 1990's, it has become the subject of recent studies focusing on chromatin architecture due to its implications in cell-type specific (E-P) interactions[88,89]. The data in the present study is supported by the literature, as we found that YY1 responds dynamically to Epo stimulation within an hour in concordance with acute transcriptional responses (**Figure 3.3**). Additionally, there was little overlap in binding locations of YY1 and CTCF, which supports the accepted model that YY1-mediated chromatin loops that bring enhancers and promoters into contact are contained with larger CTCF-mediated chromatin interactions[88,89]. In the present study, the majority of YY1-mediated loops had weak scores (PET < 5, **Figure 3.7**) despite the abundant occupancy throughout the genome (**Figure 3.3**). This intriguingly suggests that the abundance of YY1 does not necessarily indicate the strength of the loop it mediates. It is possible that YY1 binding locations are establishing chromatin loops that will gain strength, and therefore delineate cell-type specific contacts more decisively as maturation continues. It would be interesting to establish YY1-mediated contacts through erythropoiesis to investigate the changes in strength, as well as the role of Epo in this process.

HiC and other chromosome conformation capture (3C) assays have been employed to gain insights to chromatin architecture on a genome-wide and loci-specific scale[59,84,162,169]. HiChIP allows for the investigation of genome-wide chromatin interactions that are mediated by

a protein of interest, which facilitates the study of particular chromatin interactions that participate in gene regulation. For example, the H3K27ac and YY1 HiChIP libraries discussed here provide novel insights in conjunction with one another (**Figure 3.8, Figure 3.9**). Previous studies have shown that H3K27ac HiChIP generates contacts between promoters and regulatory regions, such as enhancers[171]. This finding is supported in the present study and suggests that chromatin loops mediated by H3K27ac present a link between active enhancers and the activation of their target genes. Additionally, recent literature described the occupancy of YY1 at promoters and enhancers, highlighting YY1's role in regulation through E-P loops[89]. Investigating the chromatin interactions mediated by H3K27ac and YY1 in the same model system facilitates the comparison of gene regulatory mechanisms within these individual data sets. Although outside the scope of the present study, it would be interesting to understand how the combination of factors, i.e. loops mediated by both H3K27ac and YY1, impact gene regulation.

Together, the findings presented here examine the transcriptional and epigenetic responses to hormone stimulation in erythroid precursors. These are analyzed within the context of an invariant chromatin architecture, indicating that chromatin domains may not be as dynamic as these other features. Additionally, these results suggest a pre-established chromatin organization that facilitates transcriptional and epigenetic responses to external stimuli, as has been discussed in other studies[82,90-92]. Future research should aim to integrate various types of genomic data, as described here, to comprehensively understand gene regulatory mechanisms from multiple perspectives.

CHAPTER 4

ChIP-SEQ and ChIP-EXO PROFILING OF POL II, H2A.Z, and H3K4me3 in HUMAN K562 CELLS[3]

Introduction

Control of eukaryotic transcription patterns involves the interplay of Pol II and chromatin. In the paused position, Pol II is juxtaposed with the first nucleosome downstream of the TSS[274,275]. The +1 nucleosome is specifically enriched with the histone variant H2A.Z and tri-methylation of the fourth N-terminal lysine on the histone H3 tail (H3K4me3). It has been known for several decades that Pol II must overcome nucleosomal obstacles during transcription[276]. However, questions remain regarding the molecular mechanisms underlying how chromatin regulates Pol II activity, and vice versa.

Since functional genomic approaches often require tens of millions of cells per assay, immortalized mammalian cell lines are frequently used in these studies. Due to its facile growth characteristics and its designation as an ENCODE tier 1 cell line, K562 cells are one of the most commonly used mammalian cell lines. The K562 cell line was originally established from a female patient with chronic myeloid leukemia[277]. K562 cells are considered erythroleukemic, displaying characteristics of undifferentiated granulocytes and erythrocytes[278]. In the presence of specific chemical inducers, K562 cells will differentiate along the erythroid lineage and upregulate globin expression[279-281].

[3] Portions of this have been published as Mchaourab and Perreault et al, (2018). *Scientific Data.*

Chromatin immunoprecipitation coupled to high throughput sequencing (ChIP-seq) is a powerful tool to study mechanisms of gene regulation by selectively enriching for DNA fragments that interact with a given protein in living cells. A more recently developed technology, called ChIP-exo, improves upon ChIP-seq by providing near base pair mapping resolution for protein-DNA interactions. The key innovation of the ChIP-exo methodology is the incorporation of lambda exonuclease digestion in the library preparation workflow to effectively footprint the left and right 5' DNA borders of the protein-DNA crosslink site. Thus, rather than sequencing from the distal sonication borders as in ChIP-seq, ChIP-exo enriched DNA fragments are sequenced from the left and right 5' DNA borders of the protein-DNA crosslink site. The precision of the resulting data can be leveraged to provide unique and ultra-high resolution insights into the functional organization of the genome. Given its high base pair resolution, ChIP-exo is uniquely capable of spatially resolving divergent, initiating, paused, and elongating RNA polymerase II on a genome-wide scale.

Here, we extend the value of previous Pol II ChIP-exo data by generating 12 new ChIP-seq and ChIP-exo data sets for Pol II, H2A.Z, and H3K4me3 in K562 cells. ChIP-exo mapping of Pol II, a histone variant, and a histone modification should enable other investigators to use these data sets for their own research to further understand the detailed interplay of Pol II and chromatin. Further, paired libraries generated side-by-side should enable direct comparisons between the quality of ChIP-seq and ChIP-exo mapping genome-wide. To facilitate interpretation of these data, we provide detailed information on experimental design (**Figure 4.1**), sequence quality control analyses (**Figure 4.2**), and biological validation (**Figure 4.4**).

Materials & methods

Tissue culture

Human chronic myelogenous leukemia cells (K562, ATCC) were maintained at 37$_{o}$C in 5% CO_2 between 0.1-1 million cells/ml in DMEM (Dulbecco's Modified Eagle Media) containing 10% bovine calf serum and 1% Penicillin/Streptomycin.

ChIP-seq and ChIP-exo library preparation

ChIP-exo was performed as previously described[153,187] with chromatin extracted from 50 million cells, ProteinG MagSepharose resin (GE Healthcare), and 5 ug of antibody directed against RNA polymerase II, H2A.Z, or H3K4me3, (Santa Cruz sc899, EMD Millipore 07-594, or Abcam ab8580, respectively). For each biological replicate, ChIP-seq and ChIP-exo libraries were prepared using the same starting sonicated nuclear extract. Importantly, this controls for more direct comparisons ChIP-seq and ChIP-exo for each antibody used. Libraries were sequenced using an Illumina NextSeq500 sequencer as single-end reads 50 or 75 nucleotides in length.

Sequence read alignment and quality control

The base call quality for each sequenced read was assessed using the FastQC program (bioinformatics.babraham.ac.uk/projects/fastqc/) (**Figure 4.2A**). Sequence reads (fastq files) were aligned to the human hg19 reference genome build using BWA-MEM algorithm with default parameters[241]. The resulting bam files were first sorted using the Samtools Sort function, and then bam index files were generated using the Samtools Index function[242]. The purpose of bam index files is to enable viewing of raw sequencing data in a genome browser. Next, genome-

wide read coverage and enrichment were assesses using deepTOOLS fingerprint plots[188] (Fig 4.2B).

Biological validation

To estimate variance across biological replicates, the Pearson correlation coefficient for pairwise gene Reads Per Kilobase of genome per Million reads (RPKM) was computed (**Figure 4.2C**) using the Hypergeometric Optimization of Motif EnRichment (HOMER) suite[192]. Briefly, bam files were converted to tag directories using the makeTagDirectory function with the –genome, –checkGC, and –format options. To quantify and normalize tags within gene body regions to RPKM, the analyzeRepeats function was used with the –rpkm and –d options.

Chromatin Analysis and Exploration (ChAsE) visualization suite[190] was used to display the distribution of Pol II, H2A.Z, and H3K4me3 relative to the TSS (**Figure 4.4, Figure 4.5**). Raw sequencing tags were binned, smoothed, and RPKM computed using the deepTOOLS genomeCoverage tool (20bp bin, 100bp sliding window)[188]. Smoothed RPKM signal was visualized with Integrative Genomics Viewer (IGV) (**Figure 4.4C**)[189].

Data records

ChIP-seq and ChIP-exo bigwig data files were deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE108323. GEO linked ChIP-seq and ChIP-exo bam data files were deposited in the Sequence Read Archives (SRA) under accession number SRP116017.

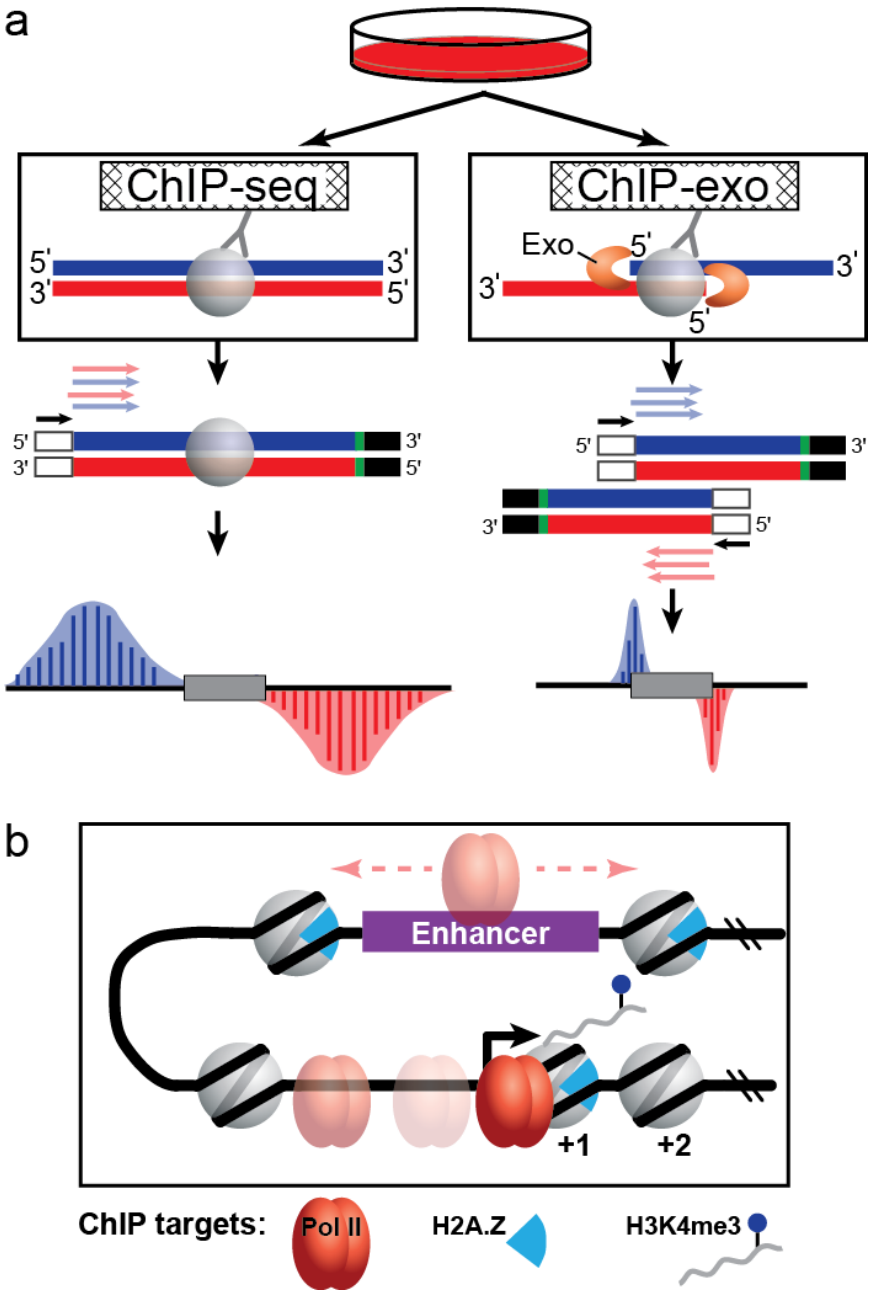

Figure 4.1. Experimental design and overview of ChIP targets. (**A**) K562 cells were cultured using standard conditions and harvested for ChIP-seq and ChIP-exo. ChIP-seq reports on the sonication borders of ChIP-enriched DNA fragments, wherein the location of the protein-DNA crosslink is deduced. In contrast, ChIP-exo sequences the exonuclease left and right borders that flank protein-DNA interactions. (**B**) Illustration of biological context of ChIP targets: Pol II, H2A.Z, and H3K4me3. [Figure from Mchaourab and Perreault et al. (2018) Scientific Data[282] and used in accordance with Copyright, SpringerNature]

Results & discussion

Overview of experimental design

In this study, functional genomic experiments using K562 cells were designed with two primary goals in mind. First, to facilitate direct comparisons for each biological replicate, ChIP-seq and ChIP-exo were performed on pooled fractions of sonicated nuclear extracts. Second, the ChIP targets (Pol II, H2A.Z, and H3K4me3) were selected so that the spatial relationships between Pol II and nucleosome positions may be examined on a genome-scale at high precision (**Figure 4.1A B**). H2A.Z and H3K4me3 are associated with both proximal promoters and distal enhancers. Indeed, recent reports have underscored the interplay of these proteins in Pol II recruitment, enhancer RNA transcription, and enhancer-promoter interactions[283-285]. Taken together, reanalysis of this collection of data should enable new biological insights into chromatin dynamics during transcriptional activation. Below, we briefly describe the rationale and considerations for sequencing data analysis with respect to general read quality, genome alignment, ChIP enrichment, replicate correlation, and biological validation.

Raw sequence quality control analyses

To assess the quality of the raw sequencing data sets, base call scores were analyzed using the FastQC program and displayed as a box plot distribution at each base position (**Figure 4.2A**). The average base quality score for all 12 ChIP data sets in the present study fell within the high confidence range (base quality score of 30-40, green region).

Raw sequence reads were aligned to the hg19 build of the human genome. On average, 46 million total aligned reads were generated for each ChIP-seq and ChIP-exo data set, ranging from 20-95 million reads. Because of the ambiguity of reads that align to multiple locations

throughout the genome, we only retain uniquely aligned reads for subsequent analyses. On average, 42 million uniquely aligned reads were obtained per data set, representing unique alignment rates between 84-93%.

Two critical questions for assessing ChIP sequencing data quality are: 1) how much of the genome is represented by a given experiment? and 2) to what extent did the ChIP assay enrich for specific regions of the genome? Typically, high genome coverage and strong ChIP enrichment are desirable in ChIP experiments. To determine genome coverage and ChIP enrichment simultaneously, we used the deepTOOLS suite to perform a fingerprint analysis (**Figure 4.2B**). In the case of Pol II ChIP-exo (**Figure 4.2B**), the fingerprint plot trace intersects the x-axis at 15, indicating 85% genome coverage. In fingerprint plots, a rightward deflection of the trace indicates the extent of ChIP enrichment. Given a point along the trace that is the point of intersection from the axes, the corresponding values on the x- and y-axes denote the percent of genome and the percent of all uniquely aligned reads, respectively. Together, these values reflect ChIP enrichment.

For example, the Pol II ChIP-exo fingerprint trace reveals that 20% of the genome (x-axis, 100-80) is enriched with 60% of all uniquely aligned reads (y-axis, 100-40), suggesting strong enrichment Pol II ChIP-exo data. Fingerprint plots for other replicates showed similar patterns of genome coverage and ChIP enrichment. Theoretically, complete genome coverage with no enrichment would be result in a trace with a slope equal to one that intersects the origin (eg: whole genome sequencing wherein 50% of the genome is contains 50% of all aligned reads).

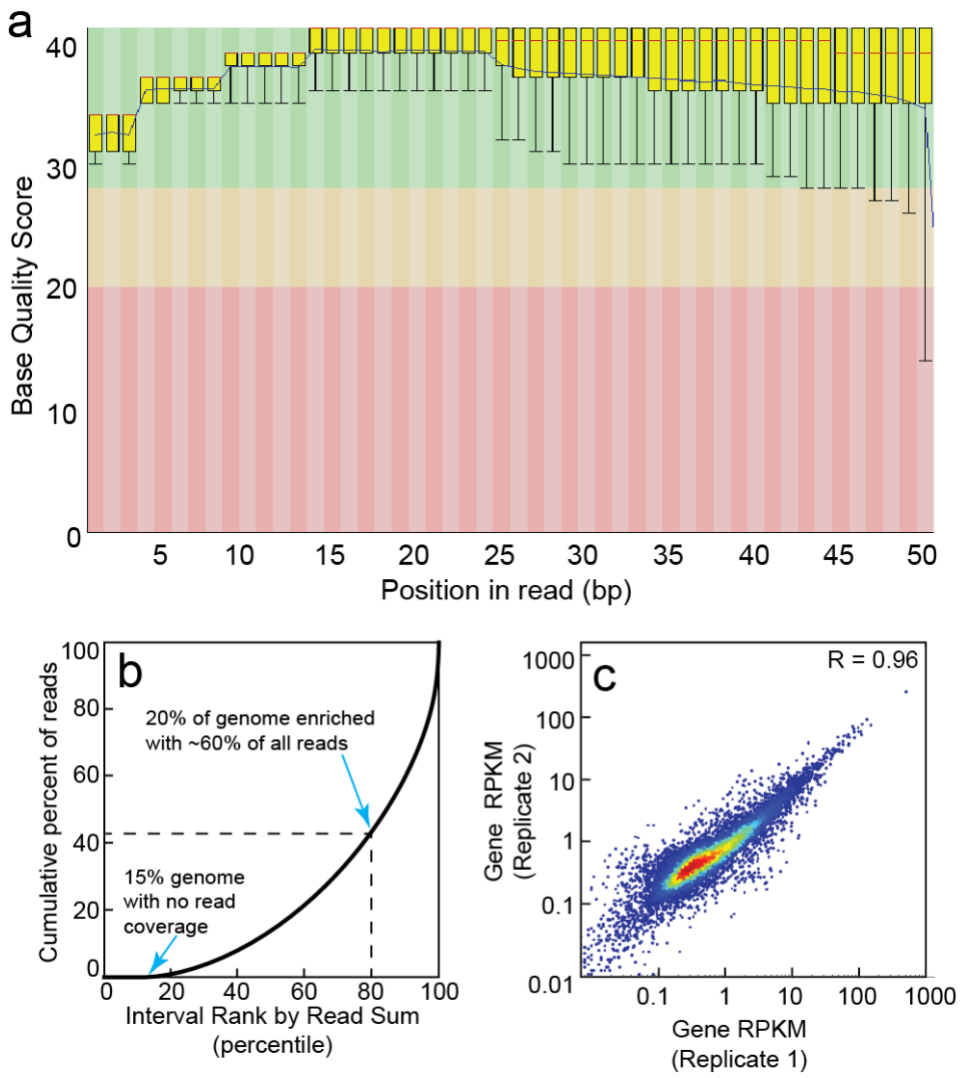

Figure 4.2. Quality control, enrichment analysis, and reproducibility for ChIP-seq and ChIP-exo data. (**A**) Box-plot distribution of base quality scores are shown for Pol II ChIP-exo replicate 1. A score greater than 30 (green region) indicates a high confidence base call. (**B**) ChIP-enrichment analysis plot that displays the cumulative percent of total reads found in a given percent of the mappable human genome. No ChIP enrichment would result in a diagonal trace. (**C**) Scatter plot correlation analysis for Pol II ChIP-exo biological replicates as measured by the Spearman correlation coefficient R-values (upper right corner). [Figure from Mchaourab and Perreault et al. (2018) Scientific Data[282] and used in accordance with Copyright, SpringerNature]

Biological validation

After verifying the quality of the raw sequencing data, we next sought to provide evidence of biological validity for the data. First, we determined the extent to which biological replicates were reproducible using correlation scatter plots (**Figure 4.2C**). For each gene, the RPKM was computed using the HOMER suite. Pearson correlation coefficients (R values) were computed for pairwise correlation plots of gene RPKMs across biological replicates. For example, biological replicates for Pol II ChIP-exo analysis displayed an R-value of 0.96, indicating high reproducibility (**Figure 4.2C**). Correlation analysis of other data resulted in positive R-values between 0.56 and 0.99. Similarity across ChIP-seq and ChIP-exo for each factor were assessed by correlation analysis between merged ChIP-exo and ChIP-seq data sets, which displayed R-values between 0.86 and 0.99 (**Figure 4.3**).

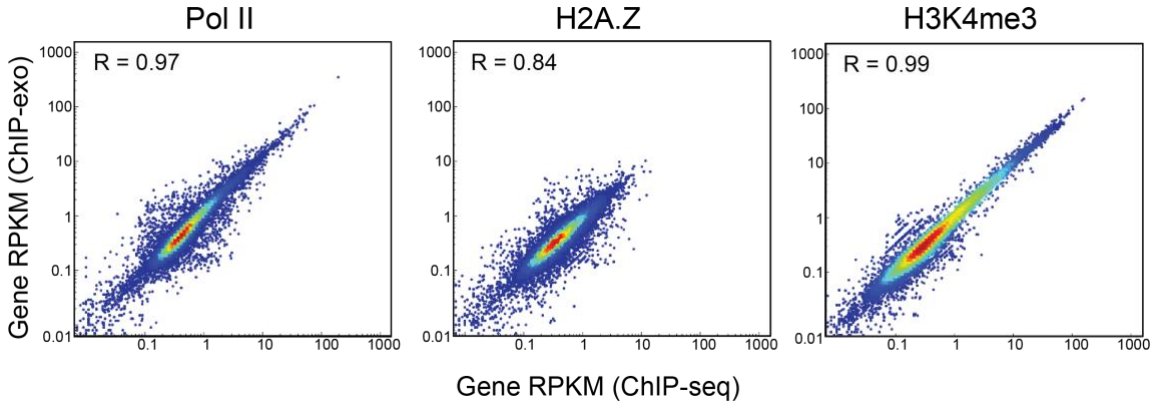

Figure 4.3. Comparison of ChIP-exo and ChIP-seq datasets. Scatter plot correlation analysis for ChIP-exo and ChIP-seq libraries as measured by the Spearman correlation coefficient R-values (upper left corner). [Figure from Mchaourab and Perreault et al. (2018) Scientific Data[282] and used in accordance with Copyright, SpringerNature]

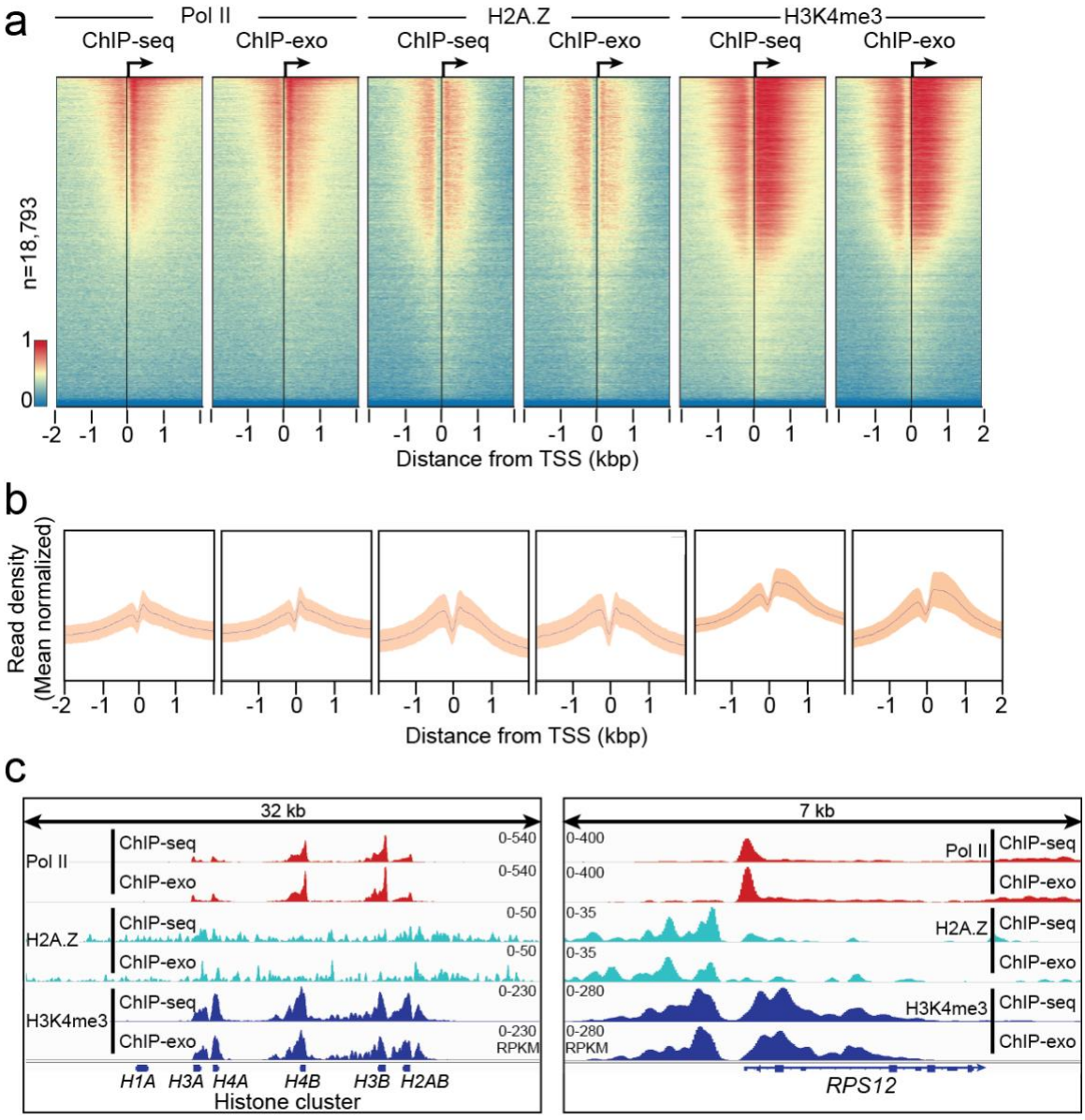

Figure 4.4. Genomic distribution of Pol II, H2A.Z, and H3K4me3 by average row tag density. (**A**) Row-linked heatmaps show RPKM normalized number of reads across a 4kb genomic interval in 20bp bins relative to the TSS. Heatmaps were generated from merged biological replicate pairs for Pol II, H2A.Z, and H3K4me3. Regions are sorted in descending order based on average row tag density for Pol II ChIP-seq. Each row represents a gene, with 18,793 genes displayed. Red and blue reflect high and low read densities, respectively. (**B**) Composite plots below each heatmap quantify the normalized tag density. The central trace denotes the average tag density for each 20bp bin and the orange fill reflects the standard deviation. (**C**) Genome browser view of ChIP-seq and ChIP-exo signal for Pol II, H2A.Z, and H3K4me3 in K562 cells shown at a histone cluster locus and the RPS12 gene. Tag distributions were smoothed and RPKM normalized using deepTOOLS. Traces were generated from merged biological replicate pairs for Pol II, H2A.Z, and H3K4me3. [Figure from Mchaourab and Perreault et al. (2018) Scientific Data[282] and used in accordance with Copyright, SpringerNature]

Given that certain transcription factors operate at a consistent distance from TSSs, analyzing global patterns of ChIP signal relative to TSSs is a useful method to assess biological validation. It is well established that once Pol II initiates transcription of genes in metazoans, Pol II moves into a stable paused state 30-50 bp downstream of the TSS[6]. Likewise, H2A.Z and H3K4me3 are consistently incorporated primarily into the +1 nucleosome of actively transcribed genes[286]. Thus, to examine global patterns of ChIP enrichment, the Chromatin Analysis and Exploration (ChAsE) heatmap tool was used to align ChIP signal merged from both biological replicates to TSSs (**Figure 4.4A**, sorted by max peak; and **Figure 4.5**, sorted by max peak position). Quantification of signal density relative to TSSs is displayed as a composite plot below each heatmap (**Figure 4.4B**). As expected, Pol II ChIP signal was sharply enriched just downstream of the TSS at the pause site for both ChIP-seq and ChIP-exo data. H2A.Z and H3K4me3 signals were broadly enriched up- and downstream of the TSS, consistent with the -1 and +1 nucleosome positions. To examine individual examples of global patterns, RPKM normalized tracks for ChIP signal were displayed using the Integrative Genome Viewer (IGV). The distribution of ChIP signal at a histone cluster and the *RPS12* gene recapitulated the global patterns of Pol II, H2A.Z, and H3K4me4 (**Figure 4.4C**). Taken together, we conclude that the data presented in this Data Descriptor represent high quality next generation sequencing data that are biologically valid, and should be useful to future studies that seek to understand the interplay of Pol II and chromatin in high resolution on a global scale.

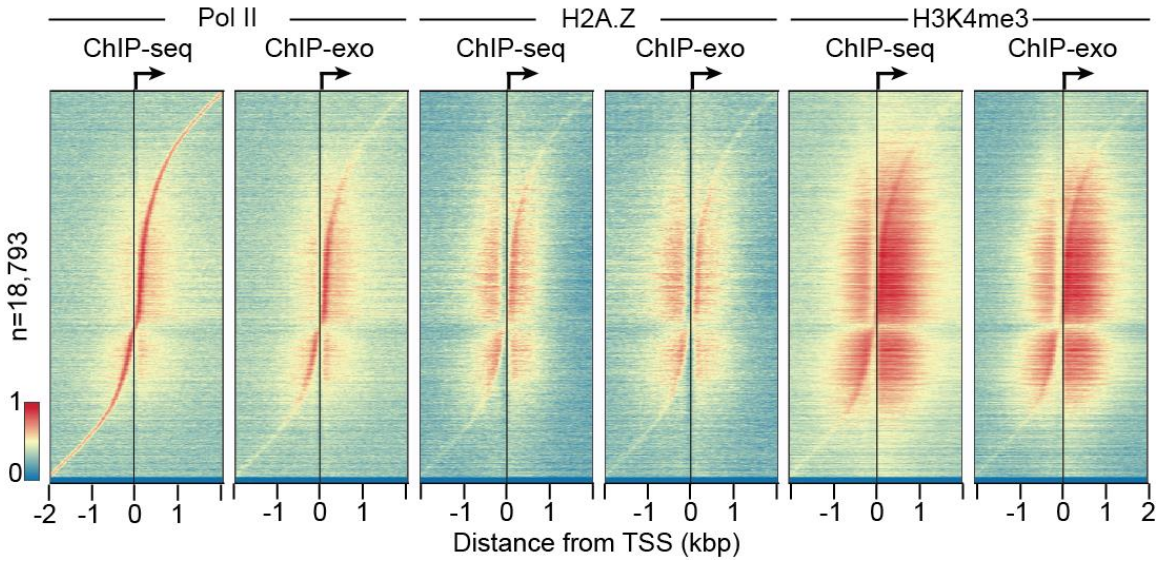

Figure 4.5. Genomic distribution of Pol II, H2A.Z, and H3K4me3 by max peak position. ChAsE heatmap display for Pol II, H2A.Z, and H3K4me3 ChIP signal from merged biological replicates. Rows are linked and sorted by Pol II ChIP-seq max peak position. [Figure from Mchaourab and Perreault et al. (2018) Scientific Data[282] and used in accordance with Copyright, SpringerNature]

CHAPTER 5

SUMMARY

Purpose of studies

The objective of these studies was to understand the epigenetic regulation of gene expression during early erythroid differentiation compared to the erythroleukemic state. This lead to the central hypothesis that erythropoietin (Epo) initiates epigenetic modifications genome-wide that alter transcriptional programs during erythropoiesis. This hypothesis was addressed using the following specific aims:

Aim 1. Establish enhancer locations genome-wide in FVA cells.

Aim 2. Identify enhancer-promoter interactions in FVA cells.

Aim 3. Investigate the epigenetic landscape of the erythroleukemic K562 cell line.

Outcome of studies

Through these studies, we sought to examine the epigenetic landscape of erythroid precursors and assess the impact of hormone stimulation on this landscape. To do this, we identified enhancer locations in proerythroblasts isolated from FVA-injected mice. To identify enhancer locations, we conducted ChIP-exo for three histone modifications, H3K4me1, H4K4me3, and H3K27ac, before and after one hour of Epo stimulation. Using the ChromHMM program, we were able to predict enhancer regions genome wide at these two time points leveraging patterns of histone modifications that form distinct chromatin states. We classified enhancers based on presence and abundance of the histone modifications studied. Interestingly, a subset of Epo-responsive enhancers was revealed. Epo-responsive enhancers were marked by a

change in H4K4me1, H3K27ac, or both. Enhancer regions were validated using transcription factor (TF) consensus binding motifs, Tal1 binding, and comparison to published data sets.

This study evaluated the epigenome in erythroid cells, but also left certain aspects of enhancer biology unstudied. Specifically, the question of enhancer-promoter assignment was left to be determined. The previous study employed the nearest gene paradigm to assign enhancers to target genes, which uses linear proximity of the regulatory regions and gene promoters. However, a large body of work suggests that this can be incorrect up to 40% of the time and enhancers frequently skip over the nearest gene. This instigated the reevaluation of the enhancer-promoter designations we had previously published. To do this, we leveraged the novel chromosome conformation capture assay HiChIP to identify chromatin contacts mediated by a protein of interest. This work has revealed that despite significant changes in gene expression, RNA polymerase II occupancy, and enhancer marks of activation, the chromatin architecture is invariant in response to Epo stimulation.

Together, these studies have addressed the larger question of the cellular response to external stimulation on a transcriptomic and epigenetic level. Recent work has investigated the response to stimuli in a variety of cellular contexts. For example, Bing Ren's group characterized the dynamics of enhancer-promoter interactions after TNF-alpha signaling in human fibroblasts[82]. This paper preceded several more, including studies that examined the chromatin organization in response to glucocorticoid treatment[90], serum- compared to 2i-cultured embryonic stem cells[92], and heat shock[91], revealing that the chromatin architecture remains stable. These findings are especially interesting when taken in consideration with the extensive enhancer activation and changes in gene expression that accompany stimuli response.

To complement this work differentiating proerythroblasts, we wanted to investigate key transcriptomic and epigenetic features in the widely used K562 cell line, which is an immortalized line produced from a patient with chronic myelogenous leukemia and is a Tier 1 ENCODE cell line. Pol II ChIP-exo data generated by the Venters lab has been reanalyzed by several groups, providing extended insights into transcription initiation and PIC formation and highlighting the multipurpose use of high-throughput sequencing data. To further study the transcriptional landscape in K562 cells and assess the importance of resolution in functional genomic studies, we complemented the original Pol II ChIP-exo libraries with 12 new ChIP-seq and ChIP-exo libraries. ChIP-exo mapping of Pol II, the histone variant H2A.Z, and the H3K4me3 histone modification will enable studies that aim to understand the detailed interplay of Pol II and chromatin. Furthermore, paired libraries generated side-by-side facilitate direct comparisons between the quality of ChIP-seq and ChIP-exo mapping genome-wide.

Future directions

These findings are consistent with research in the fields of gene regulation, epigenetics, and chromatin structure. While they contribute to important understanding of gene expression, transcription, enhancer biology, and chromatin looping in erythroid cells, the studies presented here also generate new research questions that should be addressed.

First, the FVA model system employed here allows for the direct study of Epo's effect on proerythroblasts. Thus, we purposely selected early time points to explicitly study the mechanism of Epo on transcription, enhancers, and chromatin conformation and to avoid confounding factors, such as downstream signaling. However, this does not preclude the need to systematically characterize the epigenome of developing erythroid precursors. Using the FVA

system, future work can isolate five unique cell types (proerythroblasts, basophilic erythroblasts, polychromatophilic erythroblasts, orthochromatic erythroblasts, and reticulocytes) and identify the enhancer repertoire in each cell type. Although the presumed changes in the enhancer landscape cannot be directly attributed to Epo through terminal differentiation, there are interesting biologically insights to be gained from this work.

This work would be nicely complemented by time course TF binding data, specifically the erythroid master regulators Gata1, Klf1, and Tal1, as well as important signaling proteins such as Stat5. TF binding data would provide a source of functional validation for enhancer regions, as well as demonstrate the dynamics of key transcriptional regulators during erythropoiesis. Since these TFs are intimately involved in chromatin looping, it would also benefit the field to examine the chromatin architecture at these same time points during terminal differentiation. This could be approached in two ways- description of the overall chromatin structure using ChIP-based assays of structural proteins, such as CTCF, cohesin, Mediator, and YY1, and delineation of the subset of enhancer-promoter interactions using 3C-based assays. Specifically, this potential future work would follow recent findings in the field that established cell-type specific chromatin architecture during neuronal development, classifying interactions as shared or unique between stages of differentiation[88].

Another methodology that has recently been employed in the epigenetics field is the modification or deletion of regulatory regions and relevant factors. This has been done using a variety of techniques, but the most common now is Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), which leverages an immune system component of bacteria to edit DNA sequences. Recent work has deleted CTCF[63,64,80,81,287,288] and YY1[89] binding sites at regions of the genome, revealing distinct modes of chromatin looping regulation for these

structural TFs. Additionally, future work could dissect enhancer region functionality by removing it and assessing transcription of the proposed target gene. Although this type of genetic manipulation is quite complicated in the FVA model system due to limitations in the ex vivo culture requirements, studies conducting CRISPR deletions would provide powerful data regarding the functionality of enhancer regions, especially those that are identified and deleted throughout the differentiation time course. Alternatively, this type of work could be done in another erythroid model system, such as CD34+ cells[289], the MEL cell line[290], or the G1E cell line[118].

This type of study design could also be leveraged to assess enhancer connectivity. This idea has been a result of extensive research that has shown one enhancer can regulate multiple genes and conversely one gene can be regulated by multiple enhancers. If enhancer-promoter interactions are identified and there is a context in which there are multiple interactions, CRISPR can be used to assess the connectivity of the genes and associated putative regulatory regions by deleting the region itself, the promoter of the gene, or knocking down an important TF or structural protein.

Conclusions

Altogether, these studies indicate that Epo reprograms the epigenome of differentiating erythroid cells through dynamically regulating histone modifications, TF binding, and transcription. There is a specific subset of enhancers that respond to Epo stimulation and regulate the transcription of key cell stage-specific genes. This finding was expanded upon by investigating chromatin looping through structural TF occupancy and chromatin interactions between enhancers and promoters. Importantly, Epo stimulation does not alter chromatin contacts after one hour of

treatment, despite dynamic epigenetic and transcriptional profiles. The body of work presented here integrates a variety of high-throughput sequencing data to gain understanding in the field of gene regulation, specifically addressing the role of external stimuli to epigenetic regulation of gene expression during differentiation.

REFERENCES

1 Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-131, doi:10.1101/sqb.1953.018.01.020 (1953).

2 Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-967, doi:10.1038/171964b0 (1953).

3 Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563, doi:10.1038/227561a0 (1970).

4 Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev* **10**, 2657-2683, doi:10.1101/gad.10.21.2657 (1996).

5 Venters, B. J. & Pugh, B. F. How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* **44**, 117-141, doi:10.1080/10409230902858785 (2009).

6 Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731, doi:10.1038/nrg3293 (2012).

7 Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982, doi:10.1101/gad.325142.119 (2019).

8 Nechaev, S. & Adelman, K. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* **1809**, 34-45, doi:10.1016/j.bbagrm.2010.11.001 (2011).

9 Nechaev, S. & Adelman, K. Promoter-proximal Pol II: when stalling speeds things up. *Cell Cycle* **7**, 1539-1544, doi:10.4161/cc.7.11.6006 (2008).

10 Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, doi:10.1038/s41580-018-0010-5 (2018).

11 Bascom, G. & Schlick, T. Linking Chromatin Fibers to Gene Folding by Hierarchical Looping. *Biophys J* **112**, 434-445, doi:10.1016/j.bpj.2017.01.003 (2017).

12 Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39-55, doi:10.1016/j.cell.2013.09.011 (2013).

13 Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272-286, doi:10.1038/nrg3682 (2014).

14 Fischle, W. In nucleo enzymatic assays for the identification and characterization of histone modifying activities. *Methods* **36**, 362-367, doi:10.1016/j.ymeth.2005.03.008 (2005).

15 Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116-1122, doi:10.1038/nature04219 (2005).

16 Smith, C. L. & Peterson, C. L. ATP-dependent chromatin remodeling. *Curr Top Dev Biol* **65**, 115-148, doi:10.1016/S0070-2153(04)65004-6 (2005).

17 Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422, doi:10.1038/nrm.2017.26 (2017).

18 Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat Biotechnol* **28**, 1057-1068, doi:10.1038/nbt.1685 (2010).

19 Jin, J. *et al.* A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J Biol Chem* **280**, 41207-41212, doi:10.1074/jbc.M509128200 (2005).

20 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).
21 Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707-719, doi:10.1016/j.cell.2007.01.015 (2007).
22 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080, doi:10.1126/science.1063127 (2001).
23 Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**, 163-176, doi:10.1016/j.gde.2005.01.005 (2005).
24 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).
25 Giaimo, B. D., Ferrante, F., Herchenrother, A., Hake, S. B. & Borggrefe, T. The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* **12**, 37, doi:10.1186/s13072-019-0274-9 (2019).
26 Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327-339, doi:10.1016/j.cell.2011.01.024 (2011).
27 Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-837, doi:10.1016/j.molcel.2013.01.038 (2013).
28 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:10.1016/j.cell.2016.09.018 (2016).
29 Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817-825, doi:10.1038/nbt.1662 (2010).
30 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).
31 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
32 Hon, G. C., Hawkins, R. D. & Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* **18**, R195-201, doi:10.1093/hmg/ddp409 (2009).
33 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).
34 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318, doi:10.1038/ng1966 (2007).
35 Creyghton, M. P. *et al.* 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**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
36 Peters, A. H. *et al.* Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet* **30**, 77-80, doi:10.1038/ng789 (2002).
37 Consortium, E. P. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816, doi:10.1038/nature05874 (2007).
38 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
39 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048, doi:10.1038/nbt1010-1045 (2010).

40 Li, X. Y. *et al.* Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* **6**, e27, doi:10.1371/journal.pbio.0060027 (2008).
41 Dogan, N. *et al.* Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* **8**, 16, doi:10.1186/s13072-015-0009-5 (2015).
42 Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642-654, doi:10.1016/j.cell.2012.12.033 (2013).
43 Huang, J. *et al.* Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell* **36**, 9-23, doi:10.1016/j.devcel.2015.12.014 (2016).
44 Xu, J. *et al.* Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell* **23**, 796-811, doi:10.1016/j.devcel.2012.09.003 (2012).
45 Ostuni, R. *et al.* Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157-171, doi:10.1016/j.cell.2012.12.018 (2013).
46 Barolo, S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135-141, doi:10.1002/bies.201100121 (2012).
47 Hong, J. W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314, doi:10.1126/science.1160631 (2008).
48 Guerrero, L., Marco-Ferreres, R., Serrano, A. L., Arredondo, J. J. & Cervera, M. Secondary enhancers synergise with primary enhancers to guarantee fine-tuned muscle gene expression. *Dev Biol* **337**, 16-28, doi:10.1016/j.ydbio.2009.10.006 (2010).
49 Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493, doi:10.1038/nature09158 (2010).
50 Cannavo, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* **26**, 38-51, doi:10.1016/j.cub.2015.11.034 (2016).
51 Hobert, O. Gene regulation: enhancers stepping out of the shadow. *Curr Biol* **20**, R697-699, doi:10.1016/j.cub.2010.07.035 (2010).
52 Wunderlich, Z. *et al.* Kruppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep* **14**, 3030, doi:10.1016/j.celrep.2016.03.032 (2016).
53 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
54 Hay, D. *et al.* Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* **48**, 895-903, doi:10.1038/ng.3605 (2016).
55 Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun* **9**, 943, doi:10.1038/s41467-018-03279-9 (2018).
56 Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**, 271-277, doi:10.1038/nbt.2137 (2012).
57 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
58 Gisselbrecht, S. S. *et al.* Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. *Nat Methods* **10**, 774-780, doi:10.1038/nmeth.2558 (2013).

59 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).
60 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
61 Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194-1211, doi:10.1016/j.cell.2009.06.001 (2009).
62 Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456-6465, doi:10.1073/pnas.1518552112 (2015).
63 de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676-684, doi:10.1016/j.molcel.2015.09.023 (2015).
64 Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111**, 996-1001, doi:10.1073/pnas.1317788111 (2014).
65 Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232-235, doi:10.1038/nature23263 (2017).
66 Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433, doi:10.1016/j.cell.2008.01.011 (2008).
67 Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* **105**, 8309-8314, doi:10.1073/pnas.0801273105 (2008).
68 Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801, doi:10.1038/nature06634 (2008).
69 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).
70 Nichols, M. H. & Corces, V. G. A CTCF Code for 3D Genome Architecture. *Cell* **162**, 703-705, doi:10.1016/j.cell.2015.07.053 (2015).
71 Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60-63, doi:10.1126/science.281.5373.60 (1998).
72 Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13**, 2465-2477, doi:10.1101/gad.13.19.2465 (1999).
73 Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345, doi:10.1126/science.aau0320 (2018).
74 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324, doi:10.1016/j.cell.2017.09.026 (2017).
75 Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800, doi:10.1038/s41576-018-0060-8 (2018).
76 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).
77 Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* **20**, 578-588, doi:10.1101/gr.100479.109 (2010).
78 Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387, doi:10.1016/j.cell.2014.09.030 (2014).
79 Ji, X. *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**, 262-275, doi:10.1016/j.stem.2015.11.007 (2016).

80 Hanssen, L. L. P. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol*, doi:10.1038/ncb3573 (2017).
81 Ren, G. *et al.* CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell* **67**, 1049-1058 e1046, doi:10.1016/j.molcel.2017.08.026 (2017).
82 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).
83 Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).
84 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).
85 Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64, doi:10.1038/nature08497 (2009).
86 Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**, 582-597, doi:10.1101/gr.185272.114 (2015).
87 Kleiman, E., Jia, H., Loguercio, S., Su, A. I. & Feeney, A. J. YY1 plays an essential role at all stages of B-cell differentiation. *Proc Natl Acad Sci U S A* **113**, E3911-3920, doi:10.1073/pnas.1606297113 (2016).
88 Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res* **27**, 1139-1152, doi:10.1101/gr.215160.116 (2017).
89 Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e1528, doi:10.1016/j.cell.2017.11.008 (2017).
90 D'Ippolito, A. M. *et al.* Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. *Cell Syst* **7**, 146-160 e147, doi:10.1016/j.cels.2018.06.007 (2018).
91 Ray, J. *et al.* Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 19431-19439, doi:10.1073/pnas.1901244116 (2019).
92 Atlasi, Y. *et al.* Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat Cell Biol* **21**, 568-578, doi:10.1038/s41556-019-0310-9 (2019).
93 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251, doi:10.1016/j.cell.2013.02.014 (2013).
94 Sankaran, V. G. *et al.* Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**, 1093-1097, doi:10.1038/nature08243 (2009).
95 Bauer, D. E. & Orkin, S. H. Hemoglobin switching's surprise: the versatile transcription factor BCL11A is a master repressor of fetal hemoglobin. *Curr Opin Genet Dev* **33**, 62-70, doi:10.1016/j.gde.2015.08.001 (2015).
96 Liu, J. & Krantz, I. D. Cohesin and human disease. *Annu Rev Genomics Hum Genet* **9**, 303-320, doi:10.1146/annurev.genom.9.081307.164211 (2008).
97 Miyake, N. *et al.* MLL2 and KDM6A mutations in patients with Kabuki syndrome. *Am J Med Genet A* **161A**, 2234-2243, doi:10.1002/ajmg.a.36072 (2013).

98 Nandakumar, S. K., Ulirsch, J. C. & Sankaran, V. G. Advances in understanding erythropoiesis: evolving perspectives. *Br J Haematol* **173**, 206-218, doi:10.1111/bjh.13938 (2016).
99 Chen, K. *et al.* Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. *Proc Natl Acad Sci U S A* **106**, 17413-17418, doi:10.1073/pnas.0909296106 (2009).
100 Watowich, S. S. The erythropoietin receptor: molecular structure and hematopoietic signaling pathways. *J Investig Med* **59**, 1067-1072, doi:10.2310/JIM.0b013e31820fb28c (2011).
101 Koury, M. J. & Bondurant, M. C. Maintenance by erythropoietin of viability and maturation of murine erythroid precursor cells. *Journal of cellular physiology* **137**, 65-74, doi:10.1002/jcp.1041370108 (1988).
102 Koury, M. J. & Bondurant, M. C. Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells. *Science* **248**, 378-381 (1990).
103 Koury, M. J. & Bondurant, M. C. The molecular mechanism of erythropoietin action. *Eur J Biochem* **210**, 649-663 (1992).
104 Friend, C. Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia. *J Exp Med* **105**, 307-318 (1957).
105 Friend, C. Leukemia of adult mice caused by a transmissible agent. *Ann N Y Acad Sci* **68**, 522-532, doi:10.1111/j.1749-6632.1957.tb56105.x (1957).
106 Mirand, E. A. Murine viral-induced polycyhemia. *Ann N Y Acad Sci* **149**, 486-496, doi:10.1111/j.1749-6632.1968.tb15187.x (1968).
107 Hankins, W. D. & Troxler, D. Polycythemia- and anemia-inducing erythroleukemia viruses exhibit differential erythroid transforming effects in vitro. *Cell* **22**, 693-699, doi:10.1016/0092-8674(80)90545-0 (1980).
108 Koury, M. J., Bondurant, M. C., Duncan, D. T., Krantz, S. B. & Hankins, W. D. Specific differentiation events induced by erythropoietin in cells infected in vitro with the anemia strain of Friend virus. *Proc Natl Acad Sci U S A* **79**, 635-639 (1982).
109 Koury, M. J., Sawyer, S. T. & Bondurant, M. C. Splenic erythroblasts in anemia-inducing Friend disease: a source of cells for studies of erythropoietin-mediated differentiation. *J Cell Physiol* **121**, 526-532, doi:10.1002/jcp.1041210311 (1984).
110 Noguchi, C. T., Asavaritikrai, P., Teng, R. & Jia, Y. Role of erythropoietin in the brain. *Crit Rev Oncol Hematol* **64**, 159-171, doi:10.1016/j.critrevonc.2007.03.001 (2007).
111 Timmer, S. A., De Boer, K., Knaapen, P., Gotte, M. J. & Van Rossum, A. C. The potential role of erythropoietin in chronic heart failure: from the correction of anemia to improved perfusion and reduced apoptosis? *J Card Fail* **15**, 353-361, doi:10.1016/j.cardfail.2008.10.024 (2009).
112 Wang, L. *et al.* PPARalpha and Sirt1 mediate erythropoietin action in increasing metabolic activity and browning of white adipocytes to protect against obesity and metabolic disorders. *Diabetes* **62**, 4122-4131, doi:10.2337/db13-0518 (2013).
113 Broudy, V. C., Lin, N., Brice, M., Nakamoto, B. & Papayannopoulou, T. Erythropoietin receptor characteristics on primary human erythroid cells. *Blood* **77**, 2583-2590 (1991).
114 Love, P. E., Warzecha, C. & Li, L. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet* **30**, 1-9, doi:10.1016/j.tig.2013.10.001 (2014).

115 Cantor, A. B. & Orkin, S. H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-3376, doi:10.1038/sj.onc.1205326 (2002).
116 Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* **8**, 1184-1197, doi:10.1101/gad.8.10.1184 (1994).
117 Weiss, M. J. & Orkin, S. H. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A* **92**, 9623-9627, doi:10.1073/pnas.92.21.9623 (1995).
118 Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* **17**, 1642-1651, doi:10.1128/mcb.17.3.1642 (1997).
119 Miller, I. J. & Bieker, J. J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Molecular and cellular biology* **13**, 2776-2786, doi:10.1128/mcb.13.5.2776 (1993).
120 Levings, P. P. & Bungert, J. The human beta-globin locus control region. *Eur J Biochem* **269**, 1589-1599, doi:10.1046/j.1432-1327.2002.02797.x (2002).
121 Perreault, A. A. & Venters, B. J. Integrative view on how erythropoietin signaling controls transcription patterns in erythroid cells. *Current opinion in hematology* **25**, 189-195, doi:10.1097/MOH.0000000000000415 (2018).
122 Hardison, R. *et al.* Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205**, 73-94, doi:10.1016/s0378-1119(97)00474-5 (1997).
123 Grosveld, F. *et al.* The dominant control region of the human beta-globin domain. *Ann N Y Acad Sci* **612**, 152-159 (1990).
124 Sankaran, V. G. & Orkin, S. H. The switch from fetal to adult hemoglobin. *Cold Spring Harb Perspect Med* **3**, a011643, doi:10.1101/cshperspect.a011643 (2013).
125 Noordermeer, D. & de Laat, W. Joining the loops: beta-globin gene regulation. *IUBMB Life* **60**, 824-833, doi:10.1002/iub.129 (2008).
126 Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**, 975-985 (1987).
127 Talbot, D. *et al.* A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* **338**, 352-355, doi:10.1038/338352a0 (1989).
128 Reik, A. *et al.* The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18**, 5992-6000, doi:10.1128/mcb.18.10.5992 (1998).
129 Sawyer, S. T., Koury, M. J. & Bondurant, M. C. Large-scale procurement of erythropoietin-responsive erythroid cells: assay for biological activity of erythropoietin. *Methods in enzymology* **147**, 340-352 (1987).
130 Sanger, F. & Thompson, E. O. The amino-acid sequence in the glycyl chain of insulin. *Biochem J* **52**, iii (1952).
131 Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H. & Fischer, D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in

phage fl DNA. *Proc Natl Acad Sci U S A* **70**, 1209-1213, doi:10.1073/pnas.70.4.1209 (1973).
132 Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H. & Fischer, D. Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. *J Mol Biol* **90**, 315-333, doi:10.1016/0022-2836(74)90376-3 (1974).
133 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-448, doi:10.1016/0022-2836(75)90213-2 (1975).
134 Sanger, F. *et al.* Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-695, doi:10.1038/265687a0 (1977).
135 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467, doi:10.1073/pnas.74.12.5463 (1977).
136 Sanger, F. Determination of nucleotide sequences in DNA. *Science* **214**, 1205-1210, doi:10.1126/science.7302589 (1981).
137 Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* **162**, 729-773, doi:10.1016/0022-2836(82)90546-0 (1982).
138 Yamey, G. Scientists unveil first draft of human genome. *BMJ* **321**, 7, doi:10.1136/bmj.321.7252.7 (2000).
139 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).
140 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).
141 Green, E. D., Watson, J. D. & Collins, F. S. Human Genome Project: Twenty-five years of big biology. *Nature* **526**, 29-31, doi:10.1038/526029a (2015).
142 Gilmour, D. S. & Lis, J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci U S A* **81**, 4275-4279, doi:10.1073/pnas.81.14.4275 (1984).
143 Gilmour, D. S. & Lis, J. T. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Mol Cell Biol* **5**, 2009-2018, doi:10.1128/mcb.5.8.2009 (1985).
144 Blat, Y. & Kleckner, N. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* **98**, 249-259, doi:10.1016/s0092-8674(00)81019-3 (1999).
145 Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309, doi:10.1126/science.290.5500.2306 (2000).
146 Lee, T. I. *et al.* Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804, doi:10.1126/science.1075090 (2002).
147 Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. & Farnham, P. J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* **16**, 235-244, doi:10.1101/gad.943102 (2002).
148 Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240-248, doi:10.1016/j.ymeth.2009.03.001 (2009).
149 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).

150 Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502, doi:10.1126/science.1141319 (2007).
151 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).
152 Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-1419, doi:10.1016/j.cell.2011.11.013 (2011).
153 Perreault, A. A. & Venters, B. J. The ChIP-exo Method: Identifying Protein-DNA Interactions with Near Base Pair Precision. *J Vis Exp*, doi:10.3791/55016 (2016).
154 He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotechnol* **33**, 395-401, doi:10.1038/nbt.3121 (2015).
155 Rossi, M. J., Lai, W. K. M. & Pugh, B. F. Simplified ChIP-exo assays. *Nat Commun* **9**, 2842, doi:10.1038/s41467-018-05265-7 (2018).
156 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).
157 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
158 Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344-1349, doi:10.1126/science.1158441 (2008).
159 Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523-536, doi:10.1016/j.cell.2008.03.029 (2008).
160 Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848, doi:10.1126/science.1162228 (2008).
161 Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953, doi:10.1126/science.1229386 (2013).
162 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
163 Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-1354, doi:10.1038/ng1896 (2006).
164 Wurtele, H. & Chartrand, P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res* **14**, 477-495, doi:10.1007/s10577-006-1075-0 (2006).
165 Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38**, 1341-1347, doi:10.1038/ng1891 (2006).
166 Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* **16**, 1299-1309, doi:10.1101/gr.5571506 (2006).
167 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661-678, doi:10.1038/nrg.2016.112 (2016).

168 Fullwood, M. J., Han, Y., Wei, C. L., Ruan, X. & Ruan, Y. Chromatin interaction analysis using paired-end tag sequencing. *Curr Protoc Mol Biol* **Chapter 21**, Unit 21 15 21-25, doi:10.1002/0471142727.mb2115s89 (2010).
169 Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* **107**, 30-39, doi:10.1002/jcb.22116 (2009).
170 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919-922, doi:10.1038/nmeth.3999 (2016).
171 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612, doi:10.1038/ng.3963 (2017).
172 Wojchowski, D. M., Sathyanarayana, P. & Dev, A. Erythropoietin receptor response circuits. *Current opinion in hematology* **17**, 169-176, doi:10.1097/MOH.0b013e328338008b (2010).
173 Koulnis, M., Porpiglia, E., Hidalgo, D. & Socolovsky, M. Erythropoiesis: from molecular pathways to system properties. *Adv Exp Med Biol* **844**, 37-58, doi:10.1007/978-1-4939-2095-2_3 (2014).
174 Richmond, T. D., Chohan, M. & Barber, D. L. Turning cells red: signal transduction mediated by erythropoietin. *Trends Cell Biol* **15**, 146-155, doi:10.1016/j.tcb.2005.01.007 (2005).
175 An, X. *et al.* Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* **123**, 3466-3477, doi:10.1182/blood-2014-01-548305 (2014).
176 Kingsley, P. D. *et al.* Ontogeny of erythroid gene expression. *Blood* **121**, e5-e13, doi:10.1182/blood-2012-04-422394 (2013).
177 Pishesha, N. *et al.* Transcriptional divergence and conservation of human and mouse erythropoiesis. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4103-4108, doi:10.1073/pnas.1401598111 (2014).
178 Singh, S. *et al.* Defining an EPOR- regulated transcriptome for primary progenitors, including Tnfr-sf13c as a novel mediator of EPO- dependent erythroblast formation. *PloS one* **7**, e38530, doi:10.1371/journal.pone.0038530 (2012).
179 Tallack, M. R. *et al.* Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Res* **22**, 2385-2398, doi:10.1101/gr.135707.111 (2012).
180 Cheng, Y. *et al.* Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19**, 2172-2184, doi:10.1101/gr.098921.109 (2009).
181 Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-223, doi:10.1038/nrg.2016.4 (2016).
182 Wu, W. *et al.* Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Res* **21**, 1659-1671, doi:10.1101/gr.125088.111 (2011).
183 Su, M. Y. *et al.* Identification of biologically relevant enhancers in human erythroid cells. *J Biol Chem* **288**, 8433-8444, doi:10.1074/jbc.M112.413260 (2013).
184 Finkelstein, L. D., Ney, P. A., Liu, Q. P., Paulson, R. F. & Correll, P. H. Sf-Stk kinase activity and the Grb2 binding site are required for Epo-independent growth of primary erythroblasts infected with Friend virus. *Oncogene* **21**, 3562-3570, doi:10.1038/sj.onc.1205442 (2002).

185 Zhang, J. *et al.* Role of erythropoietin receptor signaling in Friend virus-induced erythroblastosis and polycythemia. *Blood* **107**, 73-78, doi:10.1182/blood-2005-05-1784 (2006).
186 Perreault, A. A., Benton, M. L., Koury, M. J., Brandt, S. J. & Venters, B. J. Epo reprograms the epigenome of erythroid cells. *Exp Hematol*, doi:10.1016/j.exphem.2017.03.004 (2017).
187 Pugh, B. F. & Venters, B. J. Genomic Organization of Human Transcription Initiation Complexes. *PloS one* **11**, e0149339, doi:10.1371/journal.pone.0149339 (2016).
188 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365 (2014).
189 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
190 Younesy, H. *et al.* An interactive analysis and exploration tool for epigenomic data. *Computer Graphics Forum* **32: 91-100**, doi:10.1111/cgf.12096 (2013).
191 Saldanha, A. J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248, doi:10.1093/bioinformatics/bth349 (2004).
192 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
193 Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858, doi:10.1038/nature07730 (2009).
194 Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**, D142-147, doi:10.1093/nar/gkt997 (2014).
195 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11-11 12 34, doi:10.1002/0471250953.bi1112s47 (2014).
196 Bondurant, M. C., Lind, R. N., Koury, M. J. & Ferguson, M. E. Control of globin gene transcription by erythropoietin in erythroblasts from friend virus-infected mice. *Molecular and cellular biology* **5**, 675-683 (1985).
197 Rhodes, M. M., Kopsombut, P., Bondurant, M. C., Price, J. O. & Koury, M. J. Bcl-x(L) prevents apoptosis of late-stage erythroblasts but does not mediate the antiapoptotic effect of erythropoietin. *Blood* **106**, 1857-1863, doi:10.1182/blood-2004-11-4344 (2005).
198 Koury, M. J. Abnormal erythropoiesis and the pathophysiology of chronic anemia. *Blood reviews* **28**, 49-66, doi:10.1016/j.blre.2014.01.002 (2014).
199 Li, J. P., D'Andrea, A. D., Lodish, H. F. & Baltimore, D. Activation of cell growth by binding of Friend spleen focus-forming virus gp55 glycoprotein to the erythropoietin receptor. *Nature* **343**, 762-764, doi:10.1038/343762a0 (1990).
200 Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077-3086, doi:10.1182/blood-2002-04-1104 (2002).
201 Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364, doi:10.1038/nature13992 (2014).
202 Vahedi, G. *et al.* STATs shape the active enhancer landscape of T cell populations. *Cell* **151**, 981-993, doi:10.1016/j.cell.2012.09.044 (2012).

203 Brown, J. D. *et al.* NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol Cell* **56**, 219-231, doi:10.1016/j.molcel.2014.08.024 (2014).
204 Rauch, M., Tussiwand, R., Bosco, N. & Rolink, A. G. Crucial role for BAFF-BAFF-R signaling in the survival and maintenance of mature B cells. *PloS one* **4**, e5456, doi:10.1371/journal.pone.0005456 (2009).
205 Lento, W., Congdon, K., Voermans, C., Kritzik, M. & Reya, T. Wnt signaling in normal and malignant hematopoiesis. *Cold Spring Harb Perspect Biol* **5**, doi:10.1101/cshperspect.a008011 (2013).
206 Chiabrando, D. & Tolosano, E. Diamond Blackfan Anemia at the Crossroad between Ribosome Biogenesis and Heme Metabolism. *Adv Hematol* **2010**, 790632, doi:10.1155/2010/790632 (2010).
207 Pestov, D. G., Strezoska, Z. & Lau, L. F. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol* **21**, 4246-4255, doi:10.1128/MCB.21.13.4246-4255.2001 (2001).
208 Amati, B., Littlewood, T. D., Evan, G. I. & Land, H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO J* **12**, 5083-5087 (1993).
209 Jayapal, S. R. *et al.* Down-regulation of Myc is essential for terminal erythroid maturation. *J Biol Chem* **285**, 40252-40265, doi:10.1074/jbc.M110.181073 (2010).
210 Orkin, S. H. GATA-binding transcription factors in hematopoietic cells. *Blood* **80**, 575-581 (1992).
211 Evans, T., Reitman, M. & Felsenfeld, G. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci U S A* **85**, 5976-5980 (1988).
212 Wadman, I. A. *et al.* The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J* **16**, 3145-3157, doi:10.1093/emboj/16.11.3145 (1997).
213 Soler, E. *et al.* The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24**, 277-289, doi:10.1101/gad.551810 (2010).
214 Zhao, W., Kitidis, C., Fleming, M. D., Lodish, H. F. & Ghaffari, S. Erythropoietin stimulates phosphorylation and activation of GATA-1 via the PI3-kinase/AKT signaling pathway. *Blood* **107**, 907-915, doi:10.1182/blood-2005-06-2516 (2006).
215 Prasad, K. S., Jordan, J. E., Koury, M. J., Bondurant, M. C. & Brandt, S. J. Erythropoietin stimulates transcription of the TAL1/SCL gene and phosphorylation of its protein products. *J Biol Chem* **270**, 11603-11611 (1995).
216 Wu, W. *et al.* Dynamic shifts in occupancy by TAL1 are guided by GATA factors and drive large-scale reprogramming of gene expression during hematopoiesis. *Genome Res* **24**, 1945-1962, doi:10.1101/gr.164830.113 (2014).
217 Dore, L. C. *et al.* A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci U S A* **105**, 3333-3338, doi:10.1073/pnas.0712312105 (2008).
218 Yu, D. *et al.* miR-451 protects against erythroid oxidant stress by repressing 14-3-3zeta. *Genes Dev* **24**, 1620-1633, doi:10.1101/gad.1942110 (2010).
219 Kuhrt, D. & Wojchowski, D. M. Emerging EPO and EPO receptor regulators and signal transducers. *Blood* **125**, 3536-3541, doi:10.1182/blood-2014-11-575357 (2015).

220 Lux, S. E. t. Anatomy of the red cell membrane skeleton: unanswered questions. *Blood* **127**, 187-199, doi:10.1182/blood-2014-12-512772 (2016).
221 Perkins, A. *et al.* Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood* **127**, 1856-1862, doi:10.1182/blood-2016-01-694331 (2016).
222 Bai, X. *et al.* TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell* **142**, 133-143, doi:10.1016/j.cell.2010.05.028 (2010).
223 Tanimura, N. *et al.* Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO Rep* **17**, 249-265, doi:10.15252/embr.201541465 (2016).
224 Xu, J. *et al.* Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev* **24**, 783-798, doi:10.1101/gad.1897310 (2010).
225 Calero-Nieto, F. J. *et al.* Transcriptional regulation of Elf-1: locus-wide analysis reveals four distinct promoters, a tissue-specific enhancer, control by PU.1 and the importance of Elf-1 downregulation for erythroid maturation. *Nucleic Acids Res* **38**, 6363-6374, doi:10.1093/nar/gkq490 (2010).
226 Mansour, M. R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-1377, doi:10.1126/science.1259037 (2014).
227 Lee, K., Hsiung, C. C., Huang, P., Raj, A. & Blobel, G. A. Dynamic enhancer-gene body contacts during transcription elongation. *Genes Dev* **29**, 1992-1997, doi:10.1101/gad.255265.114 (2015).
228 Liu, X., Kraus, W. L. & Bai, X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem Sci* **40**, 516-525, doi:10.1016/j.tibs.2015.07.003 (2015).
229 Adelman, K. *et al.* Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18207-18212, doi:10.1073/pnas.0910177106 (2009).
230 Gaertner, B. *et al.* Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Rep* **2**, 1670-1683, doi:10.1016/j.celrep.2012.11.024 (2012).
231 Danko, C. G. *et al.* Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* **50**, 212-222, doi:10.1016/j.molcel.2013.02.015 (2013).
232 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-29, doi:10.1002/0471142727.mb2129s109 (2015).
233 Sheffield, N. C. & Furey, T. S. Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. *Genes (Basel)* **3**, 651-670, doi:10.3390/genes3040651 (2012).
234 Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**, 1757-1767, doi:10.1101/gr.121541.111 (2011).

235 Holmqvist, P. H. & Mannervik, M. Genomic occupancy of the transcriptional co-activators p300 and CBP. *Transcription* **4**, 18-23, doi:10.4161/trns.22601 (2013).
236 Witte, S., Bradley, A., Enright, A. J. & Muljo, S. A. High-density P300 enhancers control cell state transitions. *BMC Genomics* **16**, 903, doi:10.1186/s12864-015-1905-6 (2015).
237 Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627, doi:10.1016/j.cell.2015.11.024 (2015).
238 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
239 Lareau, C. A. & Aryee, M. J. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods* **15**, 155-156, doi:10.1038/nmeth.4583 (2018).
240 Lareau, C. A. & Aryee, M. J. diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics* **34**, 672-674, doi:10.1093/bioinformatics/btx623 (2018).
241 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
242 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
243 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
244 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).
245 An, X. & Chen, L. Flow Cytometry (FCM) Analysis and Fluorescence-Activated Cell Sorting (FACS) of Erythroid Cells. *Methods Mol Biol* **1698**, 153-174, doi:10.1007/978-1-4939-7428-3_9 (2018).
246 Matsumoto, A. *et al.* CIS, a cytokine inducible SH2 protein, is a target of the JAK-STAT5 pathway and modulates STAT5 activation. *Blood* **89**, 3148-3154 (1997).
247 Rascle, A. & Lees, E. Chromatin acetylation and remodeling at the Cis promoter during STAT5-induced transcription. *Nucleic Acids Res* **31**, 6882-6890, doi:10.1093/nar/gkg907 (2003).
248 Marine, J. C. *et al.* SOCS3 is essential in the regulation of fetal liver erythropoiesis. *Cell* **98**, 617-627, doi:10.1016/s0092-8674(00)80049-5 (1999).
249 Jegalian, A. G. & Wu, H. Differential roles of SOCS family members in EpoR signal transduction. *J Interferon Cytokine Res* **22**, 853-860, doi:10.1089/107999002760274863 (2002).
250 Funakoshi-Tago, M. *et al.* Phosphorylated CIS suppresses the Epo or JAK2 V617F mutant-triggered cell proliferation through binding to EpoR. *Cell Signal* **31**, 41-57, doi:10.1016/j.cellsig.2016.12.008 (2017).
251 Karin, M., Liu, Z. & Zandi, E. AP-1 function and regulation. *Curr Opin Cell Biol* **9**, 240-246, doi:10.1016/s0955-0674(97)80068-3 (1997).
252 Hernandez, J. M., Floyd, D. H., Weilbaecher, K. N., Green, P. L. & Boris-Lawrie, K. Multiple facets of junD gene expression are atypical among AP-1 family members. *Oncogene* **27**, 4757-4767, doi:10.1038/onc.2008.120 (2008).
253 Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234-246, doi:10.1038/nrg3663 (2014).

254 Arzate-Mejia, R. G., Recillas-Targa, F. & Corces, V. G. Developing in 3D: the role of CTCF in cell differentiation. *Development* **145**, doi:10.1242/dev.137729 (2018).
255 Kisseleva, T., Bhattacharya, S., Braunstein, J. & Schindler, C. W. Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene* **285**, 1-24 (2002).
256 Voss, A. K. & Thomas, T. Histone Lysine and Genomic Targets of Histone Acetyltransferases in Mammals. *Bioessays* **40**, e1800078, doi:10.1002/bies.201800078 (2018).
257 Kueh, A. J., Dixon, M. P., Voss, A. K. & Thomas, T. HBO1 is required for H3K14 acetylation and normal transcriptional activity during embryonic development. *Molecular and cellular biology* **31**, 845-860, doi:10.1128/MCB.00159-10 (2011).
258 Jeng, M. Y. *et al.* Enhancer Connectome Nominates Target Genes of Inherited Risk Variants from Inflammatory Skin Disorders. *J Invest Dermatol* **139**, 605-614, doi:10.1016/j.jid.2018.09.011 (2019).
259 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-1528, doi:10.1038/ng.3935 (2017).
260 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).
261 Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**, 577-584, doi:10.1038/nature02006 (2003).
262 Schmerer, M. & Evans, T. Primitive erythropoiesis is regulated by Smad-dependent signaling in postgastrulation mesoderm. *Blood* **102**, 3196-3205, doi:10.1182/blood-2003-04-1094 (2003).
263 Kang, Y., Kim, Y. W., Yun, J., Shin, J. & Kim, A. KLF1 stabilizes GATA-1 and TAL1 occupancy in the human beta-globin locus. *Biochim Biophys Acta* **1849**, 282-289, doi:10.1016/j.bbagrm.2014.12.010 (2015).
264 Lentjes, M. H. *et al.* The emerging role of GATA transcription factors in development and disease. *Expert Rev Mol Med* **18**, e3, doi:10.1017/erm.2016.2 (2016).
265 Weiss, M. J. & Orkin, S. H. GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol* **23**, 99-107 (1995).
266 Pimkin, M. *et al.* Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res* **24**, 1932-1944, doi:10.1101/gr.164178.113 (2014).
267 Schuetze, S., Stenberg, P. E. & Kabat, D. The Ets-related transcription factor PU.1 immortalizes erythroblasts. *Molecular and cellular biology* **13**, 5670-5678, doi:10.1128/mcb.13.9.5670 (1993).
268 Weiss, A. K. H. *et al.* Structural basis for the bi-functionality of human oxaloacetate decarboxylase FAHD1. *Biochem J* **475**, 3561-3576, doi:10.1042/BCJ20180750 (2018).
269 Paralkar, V. R. *et al.* Unlinking an lncRNA from Its Associated cis Element. *Mol Cell* **62**, 104-110, doi:10.1016/j.molcel.2016.02.029 (2016).
270 Schneider, D. A. *et al.* RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12707-12712, doi:10.1073/pnas.0605686103 (2006).

271 Crickard, J. B., Lee, J., Lee, T. H. & Reese, J. C. The elongation factor Spt4/5 regulates RNA polymerase II transcription through the nucleosome. *Nucleic Acids Res* **45**, 6362-6374, doi:10.1093/nar/gkx220 (2017).
272 Merryweather-Clarke, A. T. *et al.* Distinct gene expression program dynamics during erythropoiesis from human induced pluripotent stem cells compared with adult and cord blood progenitors. *BMC Genomics* **17**, 817, doi:10.1186/s12864-016-3134-z (2016).
273 Schulz, V. P. *et al.* A Unique Epigenomic Landscape Defines Human Erythropoiesis. *Cell Rep* **28**, 2996-3009 e2997, doi:10.1016/j.celrep.2019.08.020 (2019).
274 Gilchrist, D. A. *et al.* Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540-551, doi:10.1016/j.cell.2010.10.004 (2010).
275 Mavrich, T. N. *et al.* Nucleosome organization in the Drosophila genome. *Nature* **453**, 358-362, doi:10.1038/nature06929 (2008).
276 Armstrong, J. A. Negotiating the nucleosome: factors that allow RNA polymerase II to elongate through chromatin. *Biochem Cell Biol* **85**, 426-434, doi:10.1139/O07-054 (2007).
277 Lozzio, C. B. & Lozzio, B. B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**, 321-334 (1975).
278 Andersson, L. C., Nilsson, K. & Gahmberg, C. G. K562--a human erythroleukemic cell line. *Int J Cancer* **23**, 143-147 (1979).
279 Baliga, B. S., Mankad, M., Shah, A. K. & Mankad, V. N. Mechanism of differentiation of human erythroleukaemic cell line K562 by hemin. *Cell Prolif* **26**, 519-529 (1993).
280 Sutherland, J. A., Turner, A. R., Mannoni, P., McGann, L. E. & Turc, J. M. Differentiation of K562 leukemia cells along erythroid, macrophage, and megakaryocyte lineages. *J Biol Response Mod* **5**, 250-262 (1986).
281 Bianchi, N. *et al.* Induction of erythroid differentiation of human K562 cells by cisplatin analogs. *Biochem Pharmacol* **60**, 31-40 (2000).
282 McHaourab, Z. F., Perreault, A. A. & Venters, B. J. ChIP-seq and ChIP-exo profiling of Pol II, H2A.Z, and H3K4me3 in human K562 cells. *Sci Data* **5**, 180030, doi:10.1038/sdata.2018.30 (2018).
283 Brunelle, M. *et al.* The histone variant H2A.Z is an important regulator of enhancer activity. *Nucleic Acids Res* **43**, 9742-9756, doi:10.1093/nar/gkv825 (2015).
284 Cauchy, P., Koch, F. & Andrau, J. C. Two possible modes of pioneering associated with combinations of H2A.Z and p300/CBP at nucleosome-occupied enhancers. *Transcription* **8**, 179-184, doi:10.1080/21541264.2017.1291395 (2017).
285 Segala, G., Bennesch, M. A., Pandey, D. P., Hulo, N. & Picard, D. Monoubiquitination of Histone H2B Blocks Eviction of Histone Variant H2A.Z from Inducible Enhancers. *Mol Cell* **64**, 334-346, doi:10.1016/j.molcel.2016.08.034 (2016).
286 Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* **16**, 178-189, doi:10.1038/nrm3941 (2015).
287 Lee, J., Krivega, I., Dale, R. K. & Dean, A. The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. *Cell Rep* **19**, 2490-2502, doi:10.1016/j.celrep.2017.05.072 (2017).

288 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).
289 Hoban, M. D. *et al.* CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells. *Mol Ther* **24**, 1561-1569, doi:10.1038/mt.2016.148 (2016).
290 Bauer, D. E., Canver, M. C. & Orkin, S. H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J Vis Exp*, e52118, doi:10.3791/52118 (2015).