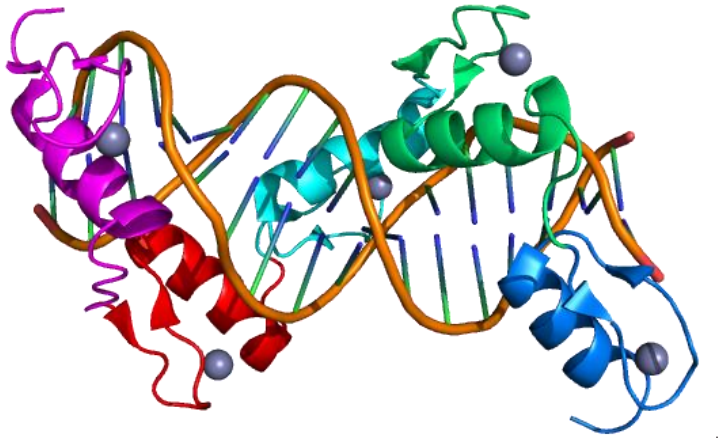
2019-2020

Master's Degree in Molecular Biology and Biomedicine

CTCF: regulator of erythroid transcription factors and its alterations in leukemia

Master's Thesis Project



Student: Vanessa Junco Ruisánchez

Institute: Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC)

Laboratory: Transcriptional control of cancer

Supervisor: Prof. M. Dolores Delgado Villar

ABBREVIATIONS

AML: Acute Myeloid Leukemia

APP: Amyloid Precursor Protein

Ara-C: 1-β-D-arabinofuranosylcytosine or cytosine arabinoside

BFU-E: Burst-Forming Unit Erythroid

BSA: Bovine Serum Albumin

CEN: Core Erythroid Network

CFU-E: Colony-Forming Unit Erythroid

ChIA-PET: Chromatin Interaction Analysis by Paired-End Tag Sequencing

ChIP: Chromatin Immunoprecipitation

ChIP-seq: Chromatin Immunoprecipitation followed by sequencing

CML: Chronic Myeloid Leukemia

CpG: Cytosine-phosphate-Guanine

CTCF: CCCTC-binding factor

CTS: CTCF binding (or target) site

DMEM: Dubelcco´s Modified Eagle Medium

DMSO: Dimethyl Sulfoxide

EDTA: Ethylenediaminetetraacetic acid

ENCODE: Encyclopedia of DNA Elements

Epo: Erythropoietin

EpoR: Erythropoietin Receptor

EV: Empty Vector

FBS: Fetal Bovine Serum

FLI1: Friend Leukemia Integration 1

Fw: Forward

GATA: GATA-binding factor

Hb: Hemoglobin

HEY1: Hairy/enhancer-of-split related with YRPW motif protein 1

HSC: Hematopoietic Stem Cell

KLF1: Krüppel-like factor 1

LDB1: LIM-Domain-Binding protein 1

LMO2: LIM Domain Only 2

MEP: Megakaryocyte Erythroid Progenitor

MOI: Multiplicity of Infection

MYB: Myeloblastosis oncogene

NFE2L2: Nuclear Factor, Erythroid like 2

PARP1: Poly(ADP-ribose) Polymerase 1

PARylation: Poly(ADP-ribosyl)ation

PBS: Phosphate-buffered saline

PEG: Polyethylene glycol

PEI: Polyethylenimine

PTM: Post-translational Modification

RBC: Red Blood Cell

RPMI: Culture medium Roswell Park Memorial Institute

RT: Room Temperature

RT-qPCR: Quantitative Real-Time Polymerase Chain Reaction

Rv: Reverse

SCF: Stem Cell Factor

SCL: Stem Cell Leukemia

s.d.: Standard Deviation

SDS: Sodium Dodecyl Sulfate

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

shCTCF: Short hairpin RNA against CTCF

shRNA: Short hairpin RNA

TAD: Topologically Associated Domain

TAL1: T-cell Acute Lymphocytic Leukemia 1

TCF3: Transcription Factor 3

TF: Transcription Factor

TSS: Transcription Start Site

UCSC: University of California, Santa Cruz

ZN: Zinc-Finger

ABSTRACT

CCCTC-binding factor (CTCF) is a zinc finger protein that binds to many DNA sequences and interacts with several partners to perform multiple functions, including transcriptional regulation, chromatin insulation, epigenetic modulation, and long-range interactions. Red blood cells are generated from hematopoietic stem cells during the erythropoiesis. Previous results of our group revealed that CTCF is involved in this process. The aim of this study was to better know the role of CTCF in the control of erythropoiesis and in the regulation of specific erythroid transcription factors. CTCF was downregulated using shRNA in K562 cells, derived from human chronic myeloid leukemia, and in primary $CD34^+$ progenitor cells purified from cord blood. Erythroid differentiation was induced by Ara-C or Imatinib in K562 cells and by erythropoietin in $CD34^+$ cells. Expression of erythroid markers and transcription factors was analyzed by Western blot and RT-qPCR. We observed that CTCF downregulation inhibits erythroid differentiation and regulates the expression of specific erythroid transcription factors. Using the ENCODE project, we identified CTCF binding sites to the regulatory regions of selected erythroid transcription factors and the chromatin interactions mediated by CTCF around them, which vary across cell lines. Moreover, we analyzed the histone modifications in the CTCF binding sites to erythroid genes. Furthermore, we analyzed clinical profiles from cBioPortal, and we observed that genetic alterations in *CTCF* gene appear to be involved in the development of acute myeloid leukemia, which supports CTCF role as tumor suppressor gene. In conclusion, our data suggest that CTCF plays an essential role in the erythroid differentiation through the formation of long-range chromatin interactions that modulate the expression of erythroid transcription factors.

INDEX

1. INTRODUCTION

1.1. CTCF STRUCTURE AND GENERAL FEATURES

The multifunctional CCCTC-binding factor (CTCF) is a highly conserved DNA-binding protein with 11-zinc-finger (ZF) domain. CTCF was firstly described for the presence of the repeating CCCTC sequence in its DNA binding site at the chicken c-Myc gene 5´-flanking sequence (Lobanenkov et al., 1990). Then, CTCF was characterized as a transcriptional repressor of chicken Myc gene (Klenova et al., 1993) and of human *MYC* gene (Filippova et al., 1996). Shortly after, CTCF was described as a transcriptional activator of the amyloid β-precursor protein (APP) (Vostrov & Quitschke, 1997). Furthermore, negative protein 1 (NeP1) was described as a binding protein of the silencer element of the chicken lysozyme that synergistically with the thyroid hormone receptor represses transcriptional activity (Baniahmad et al., 1990). Soon after, NeP1 was turned out to be CTCF in many organisms, including humans, which reveals the role of CTCF in the organization of chromatin structure (Burcin et al., 1997).

CTCF levels markedly impact cellular functions, such as differentiation, proliferation and apoptosis, at least in part, by regulating gene expression (Torrano et al., 2005). CTCF is encoded by a single-copy gene localized at chromosome locus 16q22.1 (Filippova et al., 1998). The mammalian *CTCF* gene has 5 introns and 10 exons. Exons E1 and E10 encode transcriptional repressor domains, exons E2 to E8 encode 11 ZFs and E9 encodes an AT-rich DNA-binding motif (Ohlsson et al., 2001). The *CTCF* gene is cell-cycle-regulated, with a maximum expression at S-G2 phase (Klenova et al., 1998). Additionally, CTCF associates with the nuclear matrix, with the centrosomes and the midbody at the end of mitosis, suggesting that CTCF plays a role in cell-cycle control (Rosa-Garrido et al., 2012; Zhang et al., 2004).

In adult organisms, CTCF protein is ubiquitously expressed and localized in the nuclei. The nucleolar localization of CTCF regulating ribosomal genes expression has also been reported (Torrano et al., 2006; Van De Nobelen et al., 2010). CTCF is present at 80,000 sites on mammalian chromosomes (Hashimoto et al., 2017) and 50,000 CTCF binding sites can be found in a single cell line (Wang et al., 2012). CTCF is remarkably evolutionary conserved among bilaterians, being 100% identical in the ZF domain (Filippova et al., 1996), and phylogenetic analysis suggest an early origin in the evolution of Metazoa (Arzate-Mejía et al., 2018; Heger et al., 2012). Expression levels and nuclear distribution patterns vary in a cell type-specific manner, indicating an important role in phenotypic diversity and gene expression patterns (Phillips & Corces, 2009). The differences in the affinity of CTCF binding sites are related to their conservation between species and tissues. Generally, low occupancy sites tend to exhibit higher expression and are cell-specific, while high occupancy sites are associated with repressive histone marks and appear to be conserved across cell types (Essien et al., 2009).

CTCF is a single polypeptide chain of 727 amino acid residues, with a molecular weight of 130 kDa, that folds into a secondary structure with three domains: a N-terminal region, a central domain containing 11 ZFs, each one consisting of an α-helix and two parallel β strands (**Figure 1.1**), and a C-terminal region (Zlatanova & Caiafa, 2009). The first 10 fingers are C2H2-type ZFs, which are typical units of ~30 residues containing a pair of cysteine residues separated by 12 amino acids from a pair of histidines. These four residues are coordinated through a Zn atom to form a compact structure with a DNA-recognition

α-helix (Ohlsson et al., 2001). In vertebrates, the eleventh C-terminal ZF is C2HC-type ZF and reassembles the C2H2 fingers in structure (Guo et al., 2018). The C-terminal and N-terminal CTCF domains exhibit repressor activities (Filippova et al., 1996). The ZF region is the major DNA binding region, although both N- and C-terminal regions of CTCF have the ability to weakly interact with the DNA (Guo et al., 2018).

The three domains contain sites for distinct post-translational modifications (PTMs), which can modulate CTCF activity: poly(ADP-ribosyl)ation (PARylation) in the N-terminal domain confers insulator properties to CTCF and is involved in the control of ribosomal gene expression (Yu et al., 2004; Zlatanova & Caiafa, 2009), the C-terminal domain contains several sites for phosphorylation, more specifically phosphorylation of Ser612 switches CTCF to an activator of the chicken Myc (Klenova et al., 2001), and SUMOylation at two sites in the polypeptide chain contributes to the repressive function of CTCF on the Myc P2 promoter (MacPherson et al., 2009).

1.2. CTCF BINDING TO DNA

CTCF is a multivalent factor that can recognize and bind to different DNA sequences by employing different combinations of ZFs (Filippova et al., 1996), which is known as the "CTCF code" (Wang et al., 2019). However, the finger recognition patterns are not predictable based on the DNA sequences (Guo et al., 2018). CTCF binding sequences of ∼50–60 bp are termed CTCF binding (or target) sites (CTSs). More than 60% of CTSs are located in intergenic regions and sites upstream of the transcription start sites (TSS) (Lee et al., 2012), and only 20% of CTSs display promoter-proximal localization (Kim et al., 2007).

Genome-wide data sets have enabled the identification of a 12-15 bp core consensus sequence (5´-NCA-NNA-G(G/A)N-GGC-(G/A)(C/G)(T/C)-3´) that is remarkably conserved in al cell types (Hashimoto et al., 2017; Phillips & Corces, 2009). ZFs 4-7 bind to 80% of CTCF target sites containing the 15 bp core motif and ZF8 on its own stabilizes CTCF to core sequence (Nakahashi et al., 2013). Analysis of CTCF consensus binding sequences identified the 8 most conserved nucleotides in the 2^{nd}, 3^{rd}, 6^{th}, 7^{th}, 8^{th}, 10^{th}, 11^{th}, and 12^{th} positions, which are recognized primarily by H-bonds between the bases of the top strand and residues of ZFs 4-7 (Hashimoto et al., 2017; Wang et al., 2019). Furthermore, CTCF tolerates base changes at specific positions within the degenerated CTSs sequences, permitting genome-wide CTCF binding to a diverse range of CTSs (Yin et al., 2017).

The six fingers ZFs 2-7 interact with the major groove of DNA making specific contacts with nucleotides by amino acid side chains at positions -1, +2, +3 and +6 (Hashimoto et al., 2017; Ohlsson et al., 2001). Each ZF recognizes 3-4bp in the DNA and the protein sequence runs in the opposite direction (from C to N) to the 15 bp core sequence (5´→3´) (Hashimoto et al., 2017) (**Figure 1.1**). There is controversy about the binding of the ZFs 9-11 to the target DNA (Hashimoto et al., 2017; Yin et al., 2017). DNA motifs flanking the core sequence modulate CTCF binding by enhancing (upstream motif) or reducing (downstream motif) its affinity to DNA. The conserved upstream motif is bound by ZFs 9-11, while the downstream motif is bound by ZFs 1-2 (Nakahashi et al., 2013).

Nuclear magnetic resonance (NMR) spectroscopy experiments revealed the mechanism of CTCF binding to the DNA. Before binding, individual ZFs are well structured with flexible linker regions. Then,

specific combinations of ZFs recognize DNA sequences and the other ZFs remain flexible, which may protrude out for further DNA binding. The flexibility of the inter-ZF linkers facilitates to CTCF adopt various conformations depending on the DNA sequences (Xu et al., 2018). A mutagenesis study has recently identified ineffective fingers and dominant effective fingers, which form distinctive patterns on different DNA targets (Guo et al., 2018).

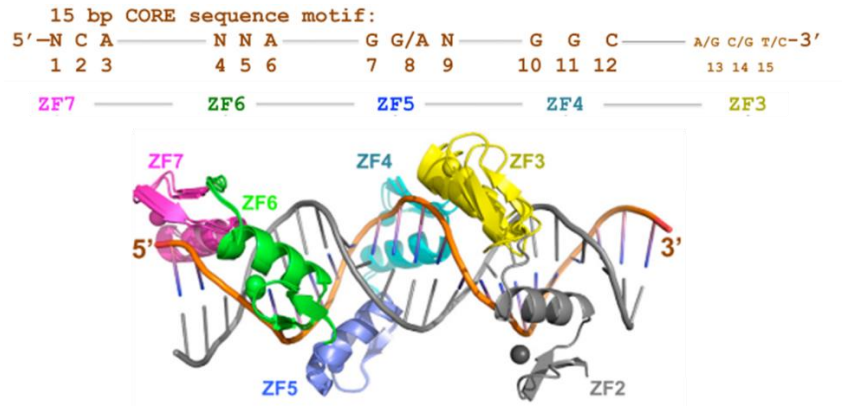


Figure 1.1. CTCF binding to DNA. ZFs 2-7 interact with the major groove of DNA. Each finger of ZFs 3–7 contacts 3 bases of the 15-bp consensus sequence. ZF2 continues to follow in the major groove, but the side chains within the DNA-interacting helix were too far away to make base-specific hydrogen bonds. Adapted from Hashimoto et al., 2017.

CTCF dimerization, this is the interaction of CTCF with another CTCF molecule via ZF and C-terminal domains, has an important role in the interaction between distant DNA regions (Ohlsson et al., 2010; Yusufzai et al., 2004). However, the dimerization through the N-terminal domain is not discarded (Yin et al., 2017). The surfaces available for dimerization vary depending on the underlying CTS (Ohlsson et al., 2010). Apart from dimerization, CTCF can interact with many partners, that will be detailed below.

1.3. CTCF INTERACTING PARTNERS

All 3 domains of CTCF may interact with other proteins (Chernukhin et al., 2007; Lee et al., 2017) or RNA (Saldaña-Meyer et al., 2019). More than 90 proteins were found to interact with CTCF with high confidence (Braccioli & de Wit, 2019), some of them bind to specific domains, but other interactions have not been mapped to specific domain (**Figure 1.2**). CTCF partners can be divided into several functional groups (Arzate-Mejía et al., 2018; Zlatanova & Caiafa, 2009): (i) architectural proteins: DNA topoisomerase II and cohesin, that has a great influence on CTCF function and will be detailed below; (ii) DNA-binding proteins: transcription factors and cofactors, including YB-1, YY1, Kaiso, RFX5, CIITA, OCT1, OCT4, TAF3, UBF; (iii) basal transcription: TFII-I and RNA Pol II; (iv) epigenetic modifiers and chromatin modelers: SIN3A, NURF, CDH8, BRD2, Suz12, Wdr5, DNMT1, H2A.Z, Taf1/SET, BRG1 and TET1; (v) miscellaneous: for example, RNA, PARP1, NPM, p68, LDB1 and CENP-E.

Cohesin complex has a central role in holding the two sister chromatids in close contact from DNA replication to their separation. The cohesin complex comprises four subunits, formed by a Smc1-Smc3 heterodimer that is complemented by Rad21/Scc1 and Scc3/SA1/SA2 subunits. The SA1 component directly interacts with CTCF (Steiner et al., 2016). There are two pools of cohesin at interphase (immobile and dynamic fractions) that form chromatin loops (Holwerda & de Laat, 2013). CTCF and cohesin have shared and independent functions at regulatory sequences in the genome. Interactions between cohesin and

CTCF mediate cell-type specific long-range chromatin contacts, modulate the enhancer-blocker activity of CTCF (Steiner et al., 2016) and can pause RNAPII activity, facilitating alternative splicing (Holwerda & de Laat, 2013). CTCF/cohesin binding sites consist of four modules defined as module 1-4 (Liu et al., 2019). ZF3, ZFs 4-7 and ZFs 9-11 recognize module 4, 3-2 and module 1 respectively, while ZF8 accommodates variable distances between modules 1 and 2 (Yin et al., 2017).

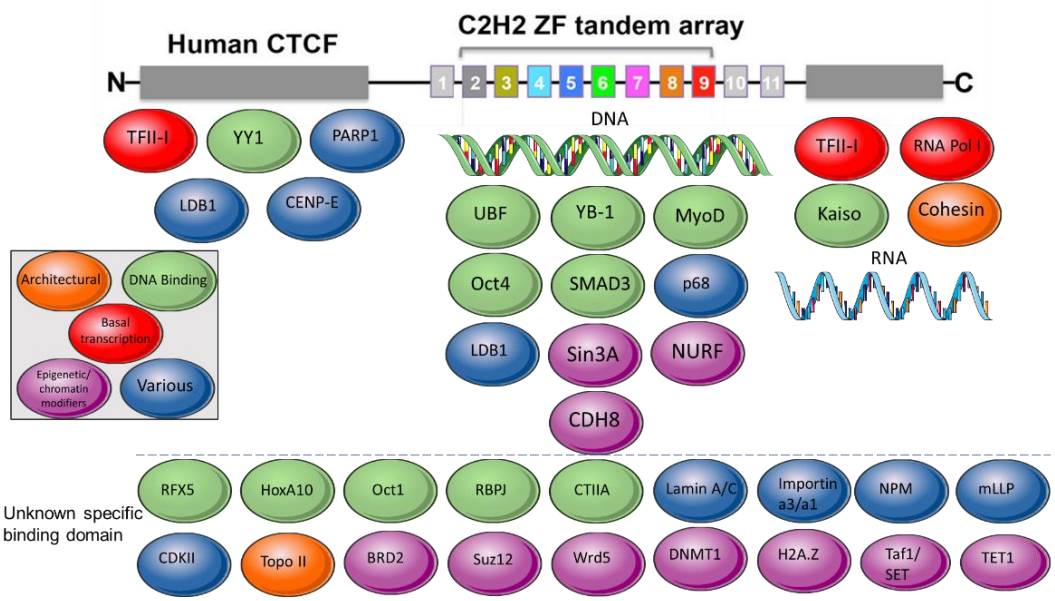


Figure 1.2 CTCF-interacting partners with their binding domains. Not only DNA, but also multiple proteins interact with the ZF domain. Other partners interact with the C-terminal domain, other proteins interact with the N-terminal domain and several additional proteins interact with CTCF in unknown specific domains.

1.4. CTCF FUNCTIONS

1.4.1. Transcription factor

CTCF is one of the most studied transcription factors (TFs). CTCF functions as a classical transcription factor that activates (Vostrov & Quitschke, 1997) or represses (Burcin et al., 1997; Filippova et al., 1996; Klenova et al., 1993; Lobanenkov et al., 1990; Lutz, 2000) the transcription of its target genes. The interaction between CTCF and APP promoter can facilitate the transcription of *APPβ* gene (Vostrov & Quitschke, 1997). Regarding with transcriptional repression, CTCF was discovered as a repressor of *MYC* gene (Filippova et al., 1996; Klenova et al., 1993; Lobanenkov et al., 1990), and also inhibits the transcription of *BAG-1* by interacting with its promoter (Sun et al., 2008). The transcription inhibition mediated by CTCF involves histone deacetylases and deacetylation of CTCF via binding the co-repressor SIN3 transcriptional regulator family member A (Lutz, 2000).

1.4.2. Insulator binding protein

Chromatin insulators are DNA elements that prevent inappropriate interactions between neighboring genes and mediate both intra- and inter-chromosomal interactions (Ong & Corces, 2014). Insulators have two main functions: (1) enhancer-blocking insulators prevent enhancer-promoter interactions if placed between these elements, thus repressing transcription; and (2) chromatin barrier insulators prevent spreading of heterochromatin, a repressive chromatin environment, into neighboring active domains (Bushey et al., 2009).

CTCF functions as the only known chromatin insulator in vertebrates that prevents interaction between a promoter and nearby enhancers or silencers, thereby preventing inappropriate gene activation or silencing and establishing independently regulated chromatin domains (Arzate-Mejía et al., 2018; Ong & Corces, 2014; Wallace & Felsenfeld, 2007). The role of CTCF as enhancer-blocking insulator was firstly described for the chicken HS4 element in the 5´ boundary of the β globin locus (Bell et al., 1999), at the *H19* imprinting control region (ICR) (Kanduri et al., 2000). In the case of the β globin locus, CTCF binding to the 5´HS4 insulator element also acts as a barrier to the spread of chromosomal silencing (West et al., 2004). The role of CTCF as chromatin barrier insulator is linked to its function as epigenetic remodeling, that will be further explained.

1.4.3. Epigenetic regulator

Epigenetics plays an important role in normal development and disease including cancer. The epigenetic modifications of histones are versatile marks that are involved in the regulation of gene expression. Methylation of histone H3 at lysine 4 (H3K4) is associated with transcriptionally active and/or poised genes in euchromatic regions of the genome. More specifically, H3K4me1 is a marker of enhancer elements and H3K4me3 is a marker of active promoter regions. Trimethylation of histone H3 at lysine 27 (H3K27me3) is a repressive epigenetic mark of enhancer elements (Chen & Lodish, 2014). Acetylation of histone H3 at lysine 9 (H3K9Ac) is associated with transcriptionally active genes, while its methylation is associated with inactive genes (Mahajan et al., 2009).

CTCF is involved in many aspects of epigenetic regulation, such as DNA methylation, high order chromatin structure and lineage-specific gene expression (Phillips & Corces, 2009). The role of CTCF as epigenetic regulator was firstly characterized by the ability of CTCF to compensate epigenetic silencing events in the human *RB* gene promoter (De La Rosa-Velázquez et al., 2007) and to protect the *p53* gene promoter against repressive histone marks (Soto-Reyes & Recillas-Targa, 2010). Our group described how CTCF regulates the local epigenetic state of ribosomal DNA repeats (Van De Nobelen et al., 2010) and its role in the epigenetic regulation of *BCL6* in B cells lymphoma (Batlle-López et al., 2015).

CTCF binding can prevent the spreading of cytosine-phosphate-guanine (CpG) methylation and protect nearby promoters from silencing by keeping them free of DNA methylation in a tissue-specific manner (De La Rosa-Velázquez et al., 2007; Filippova et al., 2005). Moreover, CTCF consensus binding sequences contain CpG susceptible to DNA methylation and CpG methylation prevents CTCF binding, which preferentially binds to unmethylated sequences (De La Rosa-Velázquez et al., 2007; Kanduri et al., 2000). CTCF binding is associated with chromatin boundaries by preventing the spread of repressive heterochromatin (chromatin barrier function) (Chen et al., 2012). Loss of CTCF binding is associated with the spread of heterochromatin and regional CpG methylation at CTCF binding sites (Cho et al., 2005). As an example, CTCF binding at the human *MYC* locus is required for protection from DNA methylation, which would lead to progressive loss of transcriptional activity (Gombert & Krumm, 2009).

A portion of CTCF binding sites is found enriched at transitions between active chromatin (high in H2K5Ac) and inactive chromatin domains (high in H3K27me3) (Cuddapah et al., 2009). CTCF binding events at promoters tend to be conserved across tissues, whereas CTCF binding to enhancers is more tissue restricted (Holwerda & de Laat, 2013). Additionally, CTCF plays a key role in establishing local

heterochromatin structure at repetitive elements. CTCF insulators flank repetitive elements in the genome (Cho et al., 2005).

1.4.4. Three-dimensional organization of the genome

Linear DNA is condensed ~100,000-fold by packaging with histones, forming nucleosomes into chromatin. At the interphase, de-condensed chromosomes occupy specific regions in the nucleus termed chromosome territories (Branco & Pombo, 2006). The positioning of the genes within the nucleus, in relation to other genes or nuclear structures, is important and is able to modulate gene function. The genome can further spatially segregate into distinct transcriptionally active (A-type) or inactive (B-type) regions termed compartments, harboring chromatin of similar states (Chen & Lei, 2019). High-throughput chromosome conformation capture (Hi-3C) methods allowed to show that the genome can be subdivided into approximately 2,000 topologically associated domain (TADs), the sub megabase region with strong contacts within each TAD, but limited contacts between different TADs. Higher resolution chromosome conformation capture carbon copy (5C) methods allowed to resolve sub-TADs within the TADs (Ghirlando & Felsenfeld, 2016). Inherent to TAD organization are chromatin loop that form between distant loci within the TAD or between sequences at TAD boundaries (Chen & Lei, 2019).

The three-dimensional organization of the genome (**Figure 1.3**) is the result of the interplay between the transcriptional state of genes and the activity of architectural proteins, such as cohesin, YY1 and CTCF (Arzate-Mejía et al., 2018). CTCF was described as a universal "master weaver" of genome in all multicellular organisms, which may be important for the organization and regulation of genomic functions in three dimensions, together with cohesin, including the formation of TADs, long-range DNA interactions, delimiting TAD borders and the formation loops between two CTCF sites in opposite or convergent orientation (Braccioli & de Wit, 2019; Phillips & Corces, 2009).

CTCF is involved in chromatin interactions between and within chromosomes across the genome. CTCF loops can contain active chromatin separated from inactive chromatin outside the loops and vice versa (Holwerda & de Laat, 2013). One of the main roles of CTCF in loop formation may be to facilitate the interaction between regulatory sequences and promoters during transcription initiation, as it has been proven for *MYC* gene (Hyle et al., 2019). Most CTCF-mediated interactions are local and involve loci less than 1 Mb apart (Ghirlando & Felsenfeld, 2016). Sequences adjacent to loop anchors are enriched in active histone modifications, RNA polymerase II (RNAPII), housekeeping genes and TSSs, suggesting that CTCF chromatin loops could represent topological structures within which transcription can take place (Arzate-Mejía et al., 2018). Nevertheless, chromatin looping requires, not only CTCF binding to cohesin, but also to non-coding RNAs (ncRNAs), which are implicated in curving the domain borders and in the formation of large regions of repressive or active chromatin (Yamamoto & Saitoh, 2019).

In mammals, most long-range interactions preferentially occur between convergently oriented CTCF binding sites, in a head-to-head configuration (Chen & Lei, 2019). Reversal of orientation of paired CTCF sites leads to disruption of looping and formation of a new loop (Ghirlando & Felsenfeld, 2016). Cohesin and CTCF form chromatin loops according to the currently favored loop extrusion model. The loop is first stabilized by a pair of cohesin molecules that acts as the extrusion complex and they enlarge the loop as they move away, either carrying along CTCF molecules with them until oriented binding sites

are reached or stopping when properly oriented bound CTCFs are encountered. This process stops once the cohesin complex encounters CTCF-bound sites in a convergent orientation (Arzate-Mejía et al., 2018). Cohesin's DNA release factor WAPL is essential in limiting loop length by releasing cohesin from the DNA, avoiding the formation of abnormally enlarged loops (Braccioli & de Wit, 2019).

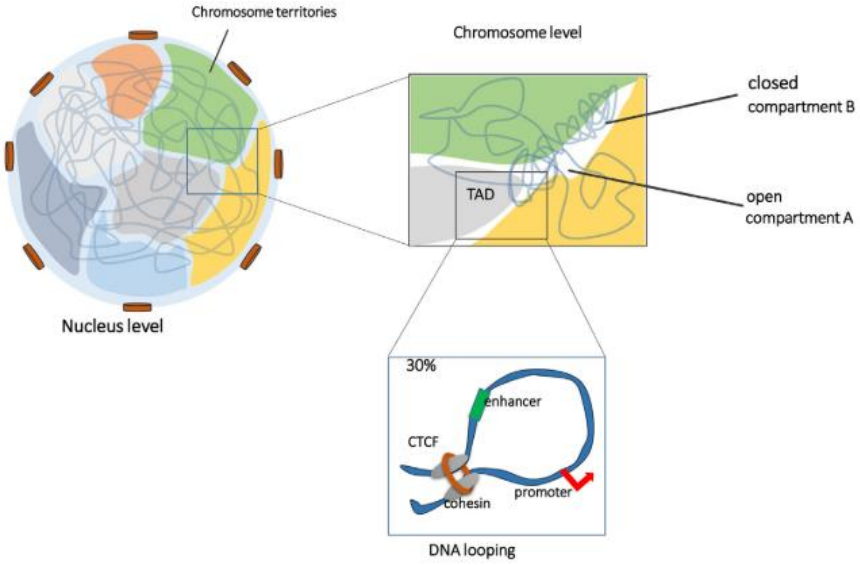


Figure 1.3. Three-dimensional genome organization. Chromosomes are organized into compartments followed by topologically associated domains (TADs) and loops. CTCF and cohesin are involved in chromatin loop formation and 30% of the loops bring together promoters and enhancers. Adapted from Wang et al., 2019.

Insulator proteins often occupy TAD boundaries and CTCF binding sites are enriched in 20 kb windows surrounding the boundaries of TADs (Holwerda & de Laat, 2013; Chen & Lei, 2019). Interruption CTCF or cohesin binding may damage TAD boundary, leading to different interactions and genes located in the affected domains can become deregulated (Gröschel et al., 2014; Liu et al., 2019). For example, an inversion on chromosome 3 in acute myeloid leukemia (AML) disrupts TAD boundaries, resulting in the formation of new hybrid TADs containing parts of each flanking genomic region. As a result, an enhancer that normally regulates the *GATA2* gene is repositioned within the TAD that contains the *EVI* oncogene (Gröschel et al., 2014).

1.5. CTCF IN DEVELOPMENT AND CANCER

CTCF is involved in many aspects of human development, including very early, postnatal and adult development. The effect on CTCF on cell physiology is highly dependent on the developmental stage and on the chromatin organization of its target genes (Arzate-Mejía et al., 2018). CTCF appears to play a role in the change of intra- and interdomain interactions that modulates gene expression during the development (Holwerda & de Laat, 2013). Additionally, CTCF is a regulator of mRNA alternative splicing, which also plays an important role in cellular differentiation and development (Braccioli & de Wit, 2019). CTCF is important for neuronal development (Arzate-Mejía et al., 2018) and it has been reported that frameshift mutations in the *CTCF* gene can cause intellectual disabilities, such as autosomal dominant mental retardation 21 (Gregor et al., 2013). Other disorders that affect human growth, such as Beckwith-Wiedemann and Silver-Russell syndromes, characterized by alterations in the *IGF2-H19* ICR, are closely linked to CTCF insulator function (Herold et al., 2012).

As mentioned above, CTCF levels markedly impact cellular functions, and the alteration of CTCF functions is related with many diseases, not only developmental disorders, but it is also dysregulated in cancer. *CTCF* gene is frequently deleted in several human malignancies (Filippova et al., 1998). Therefore, and given the essential role of CTCF, it has been proposed as a candidate tumor suppressor gene, whose expression is essential for cell viability and protection against cancer (Bailey et al., 2018).

In order to elucidate the cause that might disrupt CTCF activity in human tumors, both CTCF ZF mutations and CTCF binding sites mutations have been widely studied (Debaugny & Skok, 2020). On one hand, mutations of the ZFs of CTCF were detected in several cancers, such as endometrial cancer (ZFs 4-5 spacer region R377C mutation), prostate cancer (ZF3 H345R mutation), Wilms´ tumor (ZF3 R339W and ZF7 R448Q mutations) and breast cancer (ZF3 K344E mutation) (Filippova et al., 2002; Oh et al., 2017). Heterozygous deletion or mutations of CTCF, together with cohesin and other epigenetic modulators mutations, were observed in Down syndrome–related leukemia (Yoshida et al., 2013). On the other hand, mutations in the CTSs, mainly AT>GC and AT>GC substitutions, can disrupt the binding of CTCF and cohesin and are frequent in some types of cancer (Katainen et al., 2015), such as T-cell acute lymphoblastic leukemia (T-ALL) (Hnisz et al., 2016), liver cancer (Fujimoto et al., 2016) and gastrointestinal cancers (Guo et al., 2018).

Regarding with the epigenetics of cancer, abnormal CTCF binding due to aberrantly methylated CTCF binding sites has also been reported in various cancers, including testicular cancer, colorectal cancer, bladder cancer and ovarian cancer (Oh et al., 2017). Dysregulation of CTCF binding events in cancer is not always associated with changes in DNA methylation or mutations in CTCF binding sites, but can also be associated with altered chromatin interactions mediated by CTCF and cohesin. Specific alterations in CTCF occupancy, including both gained and lost sites, have been documented across different types of cancers, such as T-ALL, AML, breast cancer, colorectal cancer, lung cancer and prostate cancer (Fang et al., 2020).

The cBioPortal for Cancer Genomics platform (http://www.cbioportal.org) contains several studies of cancer genomics performed by The Cancer Genome Atlas (TCGA) and other groups. Due to the role of CTCF in cancer, mutations of *CTCF* and its partners have been analyzed, including amplifications, deletions, and copy number alterations. Interestingly, leukemia has not been profiled for CTCF mutations in the cBioPortal analysis performed by other authors (Debaugny & Skok, 2020; Voutsadakis, 2018).

1.6. ERYTHROID DIFFERENTIATION AND MODELS

Mammalian hematopoiesis produces approximately 10 distinct cell types, the most abundant of which belongs to the erythroid lineage (Barminko et al., 2015). An adult human produces approximately 2.5×10^{11} red blood cells (RBCs) and 1×10^{9} white blood cells per day in the bone marrow that are released into the peripheral blood, in order to replenish the cells that are taken out of the circulation by macrophages 120 days after (Valent et al., 2018). Maintenance of this rapid turnover requires efficient erythropoiesis through proliferation and differentiation of hematopoietic stem cells (HSCs) (Tumburu & Thein, 2017).

HSCs can reconstitute the entire blood system of an adult and sustain the lifelong production of all blood lineages (Pater & Dzierzak, 2015). In mammals, the sequential sites of hematopoiesis include the yolk sac, the aorta-gonad mesonephros region, the placenta, the fetal liver and finally the bone marrow

(Orkin & Zon, 2008). Positive selection for human HSCs relies on expression of CD34, c-Kit or stem cell factor (SCF) receptor, IL-6R, Thy-1 and CD45RA markers (Pater & Dzierzak, 2015), but CD34 is considered the most inclusive marker for HSCs (Huang et al., 2011). HSCs have the ability to self-renew through asymmetric cell divisions and its perturbation results in leukemia (Pater & Dzierzak, 2015). In fact, chronic myeloid leukemia (CML) arises from a multipotent HSC containing a reciprocal translocation between chromosomes 9 and 22 (Philadelphia chromosome), encoding the BCR-ABL chimeric protein with constitutively active tyrosine kinase (Yang et al., 2015).

RBCs are generated from HSCs through the process of differentiation known as erythropoiesis (**Figure 1.4**). The earliest restricted erythroid progenitors arise from megakaryocyte erythroid progenitors (MEPs) that form burst-forming unit erythroid (BFU-Es), which respond to erythropoietin (Epo), SCF, IGF1, corticosteroids, IL3 and IL6. BFU-Es give rise to colony-forming unit erythroid (CFU-Es) that are committed to terminal erythroid differentiation and depend on Epo for survival. The earliest morphologically identifiable erythroid precursor is the proerythroblast (ProE), that differentiates into basophilic erythroblast, polychromatophilic erythroblast, orthochromatic erythroblast, and reticulocyte (**Figure 1.4**). With each division, the erythroblast becomes progressively smaller, its nucleus condenses to be finally extruded, organelles are lost via autophagy and hemoglobin (Hb) accumulates in the cytoplasm (Barminko et al., 2015; Nandakumar et al., 2016). Erythroid terminal differentiation requires proliferation arrest and exit from the cell cycle, with a balance between proliferation and maturation (Yang et al., 2017).

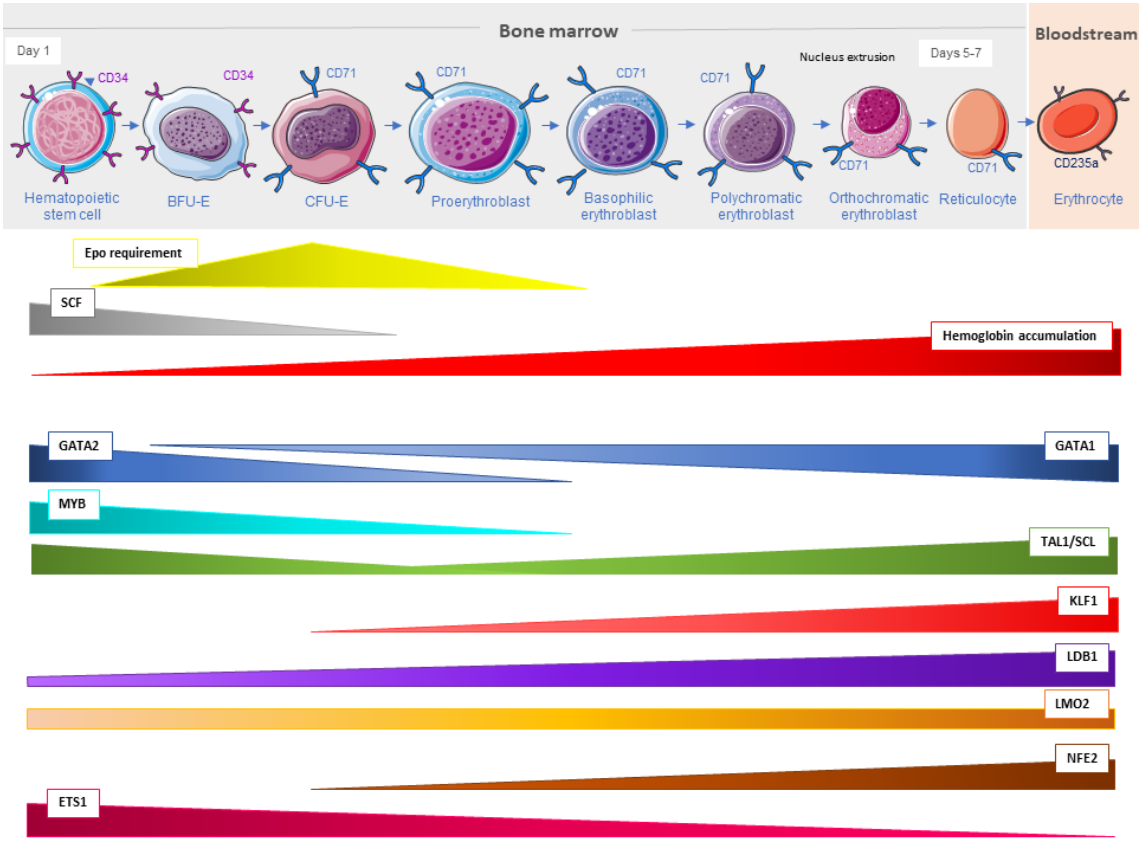


Figure 1.4. Erythroid cell differentiation. Differentiation from HSC to erythrocyte involves different progenitors characterized by different morphology, that express different surface markers, respond differently to Epo and SCF, among other growth factors, exhibit an increase in hemoglobin synthesis and are regulated by many transcription factors in a different way, including the TFs studied in this work.

Erythropoiesis in healthy individuals involves balanced high levels of transcription of α- and β-like globin genes and mutations in those genes or their regulatory regions lead to disease (Mahajan et al., 2009). A variety of β-like globin molecules are produced because the human β-globin locus is developmentally regulated (Sankaran & Orkin, 2013). Hemoglobin switching is the sequential change in globin gene expression during development (Barminko et al., 2015). Primitive erythroblasts (EryPs) mainly express the embryonic form of a β-like globin known as ε-globin (forming $\zeta_2\varepsilon_2$ and $\alpha_2\varepsilon_2$) (Demirci & Tisdale, 2018). Then, definitive erythroblasts (EryDs) mainly produce γ-globin, that in combination with adult α-globin chains into a stable tetramer forming fetal hemoglobin (HbF, $\alpha_2\gamma_2$). Reticulocytes that enter the circulation become fully mature RBCs expressing adult hemoglobin (HbA, $\alpha_2\beta_2$) through a transcriptional switch from γ- to β-globin (Sankaran & Orkin, 2013).

CTCF bound chromatin positions surrounding the α-globin TAD serve to restrict and guide local enhancer interactions to α-globin promoters (Hanssen et al., 2017). Importantly, the insulator CTCF is removed of the α-globin locus and rearranged on other specific site facilitating the transcription of α-globin gene during erythroid development (Mahajan et al., 2009). The transcriptionally active β-like globin genes are closely positioned with the locus control region (LCR) in erythroid cells and form enhancer-promoter loops mediated by five CTCF/cohesin binding sites around the β-globin locus. Recruitment of CTCF and Rad21 to CTCF sites near the β-globin locus and loop formation are dependent on GATA1, TAL1 and KLF1 (Kang et al., 2017).

The development and maturation of RBCs in the bone marrow is regulated by the supporting microenvironment, consisting of macrophages and stromal cells (Valent et al., 2018), and a complex genetic program involving a wide variety of molecules (Orkin & Zon, 2008; Pater & Dzierzak, 2015). The main regulator of RBC production is Epo, a glycoprotein hormone that activates intracellular signaling through binding to its receptor, EpoR. This stimulation upregulates the expression of globins, transferrin receptor and some membrane proteins that are characteristic of erythrocytes. Epo initiates expression of erythroid specific genes and protects against apoptosis (Perry & Soreq, 2002).

The binding of Epo to its receptor activates the cytoplasmic tyrosine kinase Janus Kinase 2 (JAK2), which triggers a downstream signaling cascade via the signal transducer and activator of transcription 5 (STAT5), phosphoinositide-3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways to support the proliferation and differentiation of erythropoietic progenitors (Hanssen et al., 2017; Moore & von Lindern, 2018). STAT5, together with TAL1, GATA1 and KLF1, binds enhancers to direct erythroid differentiation (Hanssen et al., 2017) and induces the expression of BCL-XL to maintain viability during terminal maturation (Moore & von Lindern, 2018). SCF, as well as Epo, activates PI3K/PKB signaling, which activates mammalian target of rapamycin (mTOR), a critical factor in protein synthesis (Moore & von Lindern, 2018). Epo also modulates the erythroid epigenome through the alteration of histone mark signatures across several thousand enhancer locations during erythropoiesis (Hanssen et al., 2017). The first phase of CFU-E erythroid differentiation is highly Epo-dependent, whereas later stages are no longer dependent on Epo and EpoRs are lost during terminal differentiation (Hattangadi et al., 2011).

The most extensively used model organisms in the study of erythropoiesis are mice and zebrafish. The most important cellular model is the primary culture of HSCs (i.e., $CD34^+$ cells) obtained from the

peripheral blood, BM, cord blood or fetal liver (Nandakumar et al., 2016). Erythroid differentiation can be induced in $CD34^+$ cells from cord blood using Epo (Mahajan et al., 2009). The human erythroleukemia cell lines are very important too, including K562 cells, although they show neither an adult-type hemoglobin expression pattern, nor terminal differentiation (Nandakumar et al., 2016).

K562, derived from CML, are multipotent cells that can be induced to differentiate into several hematopoietic pathways. More specifically, erythroid differentiation can be induced in K562 cells using the anti-metabolite cytosine arabinoside (Ara-C), one of the key drugs for the treatment of leukemia, that inhibits the initiation of replication and incorporates into DNA leading to the inhibition of DNA elongation (Delgado et al., 1995). High concentrations of Ara-C provoke apoptosis, but low concentrations induce erythroid differentiation (Yamada et al., 1998). Additionally, Imatinib is a BCR-ABL inhibitor that is currently being used to treat CML and in low concentration induces proliferation arrest by regulating RAS-ERK cascade, and erythroid differentiation of CML cells (Gómez-Casares et al., 2013; Yang et al., 2015). Erythroid differentiation can be confirmed with an increase in hemoglobin accumulation and is accompanied by an increased inhibition of cell growth (Woessmann et al., 2004).

CTCF is critical for hematopoiesis and the loss of CTCF results in hematopoiesis failure (Kim et al., 2017). Previous studies of our group in K562 cells showed that CTCF is differentially expressed and post-translationally modified depending on the differentiation pathway (Delgado et al., 1999). Additionally, CTCF overexpression has a stimulating effect on differentiation into the erythroid lineage, but not into the megakaryocytic lineage (Torrano et al., 2005). However, the mechanisms by which CTCF regulates the expression of several hematopoietic genes are still unknown. Previous unpublished results indicate that CTCF overexpression modulates the expression of certain genes encoding erythroid TFs, that will be the focus of this work.

1.7. ERYTHROID TRANSCRIPTION FACTORS

Erythropoiesis is tightly controlled by a combination of TFs that regulate erythroid-specific gene expression in a stage-specific and context-dependent manner, and their dysregulation is associated with chromosomal translocation and leukemias (Nandakumar et al., 2016; Perry & Soreq, 2002). The core erythroid network (CEN) (**Figure 1.5**) is necessary for proper RBC development, and it is comprised of DNA-binding TFs (GATA1 or its close relative GATA2, TAL1 and KLF1) and non-DNA-binding (LDB1 and LMO2) (Nandakumar et al., 2016).

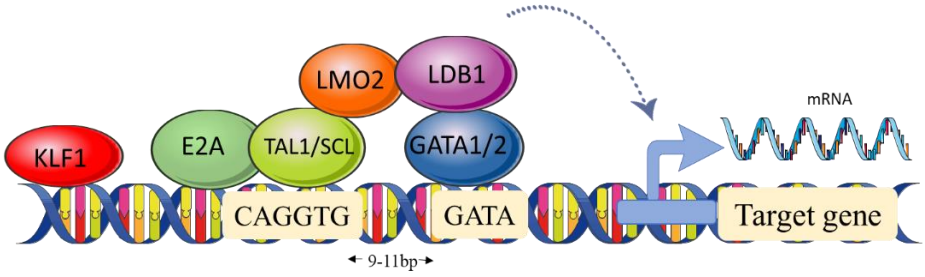


Figure 1.5. Erythroid transcription factors regulate gene expression. The core erythroid network (CEN) is a pentameric complex formed by DNA-binding factors (TAL1/SCL, KLF1 and GATA1/2) and non-DNA binding transcription factors (LMO2 and LDB1). The CEN, together with other TFs from LDB1 complexes, such as E2A, regulates the expression of erythroid genes.

- **GATA-binding factor 2 (GATA2)** is a TF that, together with LMO2, is involved in HSC production, survival, self-renewal or function (Doré & Crispino, 2011; Orkin & Zon, 2008). GATA family transcription factors bind the DNA consensus sequence (A/T)GATA(A/G) through 2 ZFs. Commitment of HSCs to the erythroid lineage is associated with the "GATA switch", which implicates GATA2 downregulation after the activation of *GATA1*, which is itself a GATA2 target gene (Love et al., 2014). As GATA1 protein levels increase, GATA1 replaces GATA2 at the β-globin locus and induces massive upregulation of hemoglobin gene expression (Doré & Crispino, 2011).

- **GATA-binding factor 1 (GATA1)** is a ZF TF that was identified to bind important DNA regulatory sequences in globin genes (Tsai et al., 1989). GATA1 is mainly expressed in MEPs and erythroblasts and is required for the maturation of several blood lineages, including erythrocytes, megakaryocytes, eosinophils, and mast cells (Doré & Crispino, 2011; Love et al., 2014). GATA1 is one of the master regulators of erythropoiesis, inducing the expression of EpoR, commitment to the erythroid lineage and transcription of adult β-globin gene (*HBB*) (Moore & von Lindern, 2018; Valent et al., 2018). There are GATA-binding motifs in promoters and enhancers of erythroid genes, TFs and co-factors, such as LMO2 (Osada et al., 1995), TAL1 and KLF1 (Merika & Orkin, 1995). The transcription factor hairy/enhancer-of-split related with YRPW motif protein 1 (HEY1) inhibits the expression of *GATA1*, maintaining an undifferentiated state of erythroid precursors (Elagib et al., 2004).

- **T-cell acute lymphocytic leukemia protein 1 (TAL1),** also known as stem cell leukemia (SCL), is a basic helix-loop-helix TF that is detected in early hematopoietic progenitors and in more mature megakaryocytes, erythroid and mast cells. TAL1/SCL expression increases during erythroid differentiation and represses differentiation of myeloid progenitors (Doré & Crispino, 2011; Perry & Soreq, 2002). Despite the requirement for TAL1/SCL in HSC generation, it is not specifically required for HSC long-term repopulating due to functional redundancy with LYL1 (Love et al., 2014). TAL1/SCL binds E-box (CAGGTG) DNA elements and participates in a DNA-bound complex containing the transcription factors E12/E47, GATA1, LDB1 and LMO2. TAL1/SCL depletion results in the inhibition of blood formation because it regulates a broad network of hematopoietic transcription factors (Doré & Crispino, 2011; Perry & Soreq, 2002).

- **Krüppel-like factor 1 (KLF1),** also known as Erythroid Krüppel-like factor (EKLF), is a ZF TF encoded by the *KLF1* gene and it is essential for the final steps of definitive erythropoiesis, promoting the maturation of erythroblasts to erythrocytes (Orkin & Zon, 2008). KLF1 regulates approximately 700 erythroid genes, primarily as a transcriptional activator, and promotes chromatin loop formation (Tumburu & Thein, 2017). In erythroid cells, *KLF1* gene is a direct target of GATA1 (Orkin & Zon, 2008). KLF1, as well as GATA1, is involved in bipotential decisions of the MEPs, inhibiting the megakaryocyte development and promoting erythroid lineage development (Siatecka & Bieker, 2011). More specifically, KLF1 represses the megakaryocytic factor Friend leukemia integration 1 (FLI1) (Doré & Crispino, 2011). KLF1 binds to the DNA consensus sequence 5´-NCNCNCCCN-3´, located in *HBB* promoter and recruits a chromatin remodeling complex for the transcription of the *HBB* gene (Orkin & Zon, 2008). In mature erythroid cells, KLF1 regulates heme synthesis, membrane stability and iron procurement pathway (Doré & Crispino, 2011).

- **LIM Domain Binding 1 (LDB1)** is a transcription co-factor that is essential for chromatin loop formation and long-range interaction of the β-globin LCR enhancer with β-globin genes, which is required for their activation (Doré & Crispino, 2011). LDB1 is necessary for both primitive and definitive erythropoiesis (Love et al., 2014). LDB1 is a critical downstream effector of the transcriptional activation network of GATA1 (Doré & Crispino, 2011). LDB1 interacts with CTCF, which is important for enhancer looping in erythroid cells. In erythroid cells, many enhancers of erythroid genes are bound by LDB1. LDB1 complex may specify the enhancers to be involved in looping, as LDB1-bound enhancers loop to genes that are occupied by CTCF (Lee et al., 2017).

- **LIM domain-only protein 2 (LMO2)** co-factor, encoded by *LMO2*, contains two cysteine-rich LIM domains and is essential for the development of hematopoietic system (Orkin & Zon, 2008). Aberrant expression of LMO2 results in the development of T-cell related disease, including T-ALL (Cantor & Orkin, 2002). LMO2 does not bind DNA by itself but acts as a bridge between DNA-binding transcription factors such as SCL and GATA1 (Perry & Soreq, 2002). In erythroid cells, LMO2 forms a complex with the transcriptional regulators TAL1, E2A, LDB1 and GATA1 (Bhattacharya et al., 2012; Brandt & Koury, 2009). Transcriptional regulation of *LMO2* in erythroid cells appears to involve multiple distal regulatory elements (DRE) (90, 75, 40 and 12) upstream of *LMO2* which overlap TFs (KLF1, GATA1, TAL1, LDB1) and CTCF/cohesin binding regions (Bhattacharya et al., 2012).

The interaction between erythroid transcription factors and their DNA targets is summarized here below. SCL/TAL1 and its heterodimeric partner transcription factor E2A (also known as TCF3) recognizes a consensus E-box DNA sequence and GATA1 binds a GATA element located about 9-11bp away, which forms a paired E-box-GATA DNA motif. LMO2 and its partner LDB1 acts to bridge a SCL/TAL1-E2A and GATA1, forming the called LDB1 complexes. LDB1 complexes that contain GATA2, instead of GATA1 regulate a transcriptional program required for the generation and maintenance of HSCs. Substitution of GATA1 for GATA2 within the LDB1 complexes results in the induction of KLF1 and global erythroid gene activation (Cantor & Orkin, 2002; Love et al., 2014). LDB1 complexes and KLF1 have independent binding sites in the genome, but frequently colocalize in many erythroid lineage–specific promoters (Li et al., 2013). *KLF1* gene is a target of LDB1 complexes. GATA1 regulates the expression of *KLF1*, by binding to a paired E-box-GATA motif within the *KLF1* promoter. GATA1, TAL1 and KLF1 form an erythroid transcriptional network co-occupying >300 genes and are required for the expression of erythroid genes, including β-globin (*HBB*), and α-globin (*HBA*) (Love et al., 2014) (**Figure 1.5**).

Other chromatin-associated proteins are also involved in this process. For instance, KLF1 recruits BRG1, a critical component of the Swi/Snf ATP-dependent chromatin remodeling complex. Another erythroid factor, Nuclear factor erythroid 2 (NFE2) associates with the MLL2 complex, which is responsible for H3K4 histone methylation (Orkin & Zon, 2008). Globin gene transcription is increased only after the binding of NFE2, TFIIB and Pol II (Hattangadi et al., 2011). Although LDB1 complexes function mainly as transcriptional activators, repressive potential can be conferred by recruitment of additional factors, such as nucleosome remodeling and deacetylase (NuRD) or PU.1 inhibitor (Love et al., 2014).

Other TFs that play an important role in erythropoiesis are:

- **ETS1** belongs to the Erythroblast Transformation Specific (ETS) family of TFs and contains a winged

helix-turn-helix DNA binding domain that recognizes purine-rich DNA sequences containing a GGA(A/T) core. ETS1 stimulates the megakaryocytic differentiation pathway, while its downregulation is required for terminal erythroid maturation. ETS1 expression is detectable from HSCs to basophilic erythroblasts, and undetectable in successive stages. *GATA2* promoter is a direct target of ETS1, with two ETS binding sites (Garrett-Sinha, 2013; Lulli et al., 2006).

- **MYB** (myeloblastosis) is a TF encoded by *MYB* proto-oncogene that controls stem and progenitor cells of all hematopoietic lineages and lack of *MYB* is lethal (Stadhouders et al., 2012). MYB is abundantly expressed in immature hematopoietic cells of erythroid, myeloid, and lymphoid lineages, but needs to be downregulated to allow terminal differentiation (Perry & Soreq, 2002). MYB regulates several genes involved in cell-cycle and proliferation (*MYC*), erythroid transcription factors (KLF1 and LMO2), globin genes, erythroid cytoskeleton proteins and surface proteins and antigens associated with erythroid differentiation (Hattangadi et al., 2011; Tumburu & Thein, 2017). There are five LDB1 complex binding sites in the *MYB-HBS11* intergenic region, consistent with active transcription of both *MYB* and *HBS11* genes. The interaction between the -36 and -81 kb sites upstream *MYB* and the CTCF binding site in the intron 1 of *MYB* locus is required for *MYB* expression (Stadhouders et al., 2012).
- **MYC** is a TF encoded by *MYC* protooncogene and is required for the balance between self-renewal and differentiation of HSCs. Different lineages of hematopoietic cells differ in their requirement for MYC to regulate their proliferation and differentiation. In K562 cells, MYC impairs Ara-C induced erythroid differentiation, but not megakaryocytic differentiation. MYC impairs the upregulation of many erythroid-specific genes, as well as that of TFs that determine erythroid differentiation (Delgado & León, 2010).

2. HYPOTHESIS AND OBJECTIVES

As it has been previously explained, CTCF may play a key role in the regulation of erythroid differentiation. Among hematopoiesis, overexpression of CTCF promotes erythroid cell differentiation of human myeloid leukemia cells (Torrano et al., 2005). Additionally, microarray assays performed in our group have identified several erythroid genes that are regulated by CTCF, including transcription factors genes with an important role in erythroid differentiation (unpublished data). However, the underlying molecular events of these effects remain unknown. We hypothesized that CTCF could bind to and regulate the expression of specific erythroid transcription factor genes.

The general aim of this work is to gain further insight into CTCF function in the regulation of erythroid cell differentiation in human hematopoietic cells. For this purpose, we stablished the following aims:

1. Analyze the effects of CTCF downregulation on erythroid cell differentiation in K562 cell line and $CD34^+$ primary progenitor cells.
2. Analyze the effects of CTCF downregulation in the expression of previously identified erythroid transcription factors, at both mRNA and protein levels.
3. Analyze CTCF binding sites to the erythroid transcription factor genes, and the possible chromatin

interactions, as well as the histone marks, in different human cell lines using the ENCODE project.

4. Search for possible correlations between CTCF expression or mutations or CTCF binding site mutations and erythroid transcription factors expression or mutations in hematological cancers using cBioPortal.

3. MATERIALS AND METHODS

3.1. CELL CULTURE

3.1.1. Cell lines

Three human cell lines were used in this work: K562 (ATCC), HEK-293T (Laboratory collection) and HeLa (ATCC). K562 cell line is derived from human CML. K562 cells have the capability of differentiating along the erythroid lineage and they grow in suspension. 293T cell line is a highly transfectable derivative of human embryonic kidney 293 cells and they grow adhered to the plate. HeLa cell line is derived from human cervical cancer and it is also an adherent cell line. All cell lines were grown in basal media (RPMI (Roswell Park Memorial Institute) medium for suspension cell culture and DMEM (Dulbecco's Modified Eagle Medium) for adherent cell culture; Gibco™) supplemented with 10% fetal bovine serum (FBS, Gibco™), 150 µg/ml gentamycin and 2 µg/ml ciprofloxacin. All cell lines were maintained at 37 ºC in a humidified 5% CO_2 atmosphere.

3.1.2. Purification and culture of human CD34⁺ cells

In order to carry out the experiments with HSCs, primary human $CD34^+$ cells were isolated from human umbilical cord blood (Banco de Sangre y Tejidos de Cantabria). First, 15 ml of fresh blood were diluted with phosphate-buffer saline (PBS) supplemented with 2 mM ethylenediaminetetraacetic acid (EDTA) and 100 U/ml DNase in a final volume of 40 ml. Then, 20 ml of diluted blood were carefully added to centrifuge tubes containing 8.6 ml of Ficoll-Paque Plus™ (GE Healthcare), avoiding the disruption of Ficoll layer. Tubes were centrifuged at 400 g for 35 min at 20 ºC without brake. The upper layer containing plasma and platelets was removed, and the mononuclear cell layer was transferred to a clean tube and washed two times with PBS-2 mM EDTA. Mononuclear cells were counted and resuspended in freezing medium (90% FBS and 10% dimethyl sulfoxide (DMSO) as cryoprotectant) to freeze at -80 ºC until the purification of $CD34^+$ cells.

The purification of $CD34^+$ cells was performed with a magnetic beads separation kit (CD34 MicroBead Kit Ultrapure and MACS Columns, Miltenyi Biotec). Mononuclear cells were centrifuged, and the cellular pellet was resuspended in RPMI medium and filtered with a column. It was centrifuged 5 min at 1200 rpm and the pellet was resuspended in a final volume of 300 µl of buffer PBS-2mM EDTA-0.5% bovine serum albumin (BSA) per 10^8 cells. To carry out the magnetic labelling, blocking reagent (100 µl) and $CD34^+$ beads (100 µl) were added to the cell suspension and it was maintained 30 min at 4 ºC with rotation. The beads were washed with buffer PBS-2mM EDTA-0.5%BSA and centrifuged 10 min at 300 g. The pellet was resuspended in 500 µl of buffer and filtered with a MACS column in the presence of a magnetic field. The column was washed 3 times to remove unlabeled cells. Then, the column was separated from the magnet and placed on a collection tube. The $CD34^+$ cells were eluted with buffer PBS-2mM

EDTA-0.5%BSA by firmly pushing the plunger into the column. Finally, the cells were counted, centrifuged, and cultured in pre-stimulation medium.

3.1.3. Induction and assessment of erythroid cell differentiation

Exponentially growing K562 cells (2.5×10^5 cell/ml) were treated with 1 µM 1-β-D-arabinofuranosylcytosine (Ara-C, Sigma-Aldrich) or 0.5 µM Imatinib (LC Laboratories) up to 3 days to induce erythroid cell differentiation (Cañelles et al., 1997; García-Gaipo, 2019). Imatinib has emerged as the lead compound to treat the chronic phase of CML (O'Brien et al., 2003) targeting the BCR-ABL oncogene that causes CML in humans, so it is an inhibitor of its oncogenic tyrosine kinase (Jacquel et al., 2007; Roskoski, 2003). Besides triggering apoptosis in K562 cells, Imatinib also mediates their erythroid differentiation (Jacquel et al., 2007). Ara-C is also being widely used in the treatment of acute and chronic leukemias, since it is a cytosine analogue that inhibits DNA polymerases α, δ, and ε, and RNA polymerases resulting in the inhibition of DNA synthesis and repair (Reiter et al., 2001).

On the other hand, expanded $CD34^+$ cells were cultured in StemSpan™ Erythroid Expansion Supplement (100X), containing SCF, IL-3 and Epo up to 5 days to selectively promote the expansion and differentiation of erythroid progenitor cells from $CD34^+$ cells. Epo is the primary regulator of erythropoiesis and it has been widely used to induce erythroid differentiation of human $CD34^+$ cells in culture (García Gaipo, 2019; Kumkhaek et al., 2013; Mahajan et al., 2009).

Cells were collected at day 0 as negative control and the following 3 days after the induction of the erythroid differentiation. Erythroid differentiation was assessed by the expression of some specific erythroid markers, such as the production of hemoglobin, as well as by the benzidine test. Benzidine test is based on the catalytic reaction between a solution of benzidine (1,1'-biphenyl-4,4'-diamine) in acetic acid and hemoglobin in presence of hydrogen peroxide (H_2O_2). It has been widely used in clinics for the detection of blood in the urine because hemoglobin-containing cells turn into blue (Ali et al., 2019; Ingham, 1932; Viola et al., 2008). To perform this test, a mixture of benzidine:H_2O_2 (50:1 v/v) was prepared and 20 µl were added to 5×10^4cells in 20 µl of R10F. Catalytic reaction occurred on ice and, after 10 min of incubation, 5 images of each conditions were obtained with the microscope. A minimum of 200 cells were counted with the ImageJ software. The percentage of benzidine-positive cells was obtained from the quotient between the number of hemoglobin-producing cells (blue) and the number of non-hemoglobin-producing cells (white).

3.1.4. Cell proliferation assay

Exponentially growing cells were counted using the NucleoCounter® NC-100™ (ChemoMetec), a high precision automated cell counter that counts suspension cells based on the staining of DNA. First, a volume of cell suspension was mixed with an equal volume of lysis buffer (Reagent A100) to disrupt plasma membranes and an equal volume of stabilizing buffer (Reagent B) to stabilize the cell nuclei and avoid DNA degradation. Then, the diluted sample was introduced in the NucleoCassette™ containing immobilized propidium iodide that stains the nuclei and the system displayed total cell number. Considering the dilution of the sample, the cellular concentration was obtained to plot the proliferation curve.

3.2. LENTIVIRAL INFECTION

3.2.1. Plasmid DNA purification

Plasmid DNA for transfections was extracted from plasmid-containing bacteria and subsequently purified using the Plasmid Midi Kit (Qiagen). The plasmids used in this work are shown in **Table 3.1** and their elements are shown in **Figure 3.1**. Both plasmids are provided with the bacterial and eukaryotic promoter (human phosphoglycerate kinase (hPGK) in pLKO and doxycycline-inducible promoter in the case of pTRIPTZ), an origin of replication (pUC) for double stranded replication, f1 origin of replication to enable single stranded replication and packaging into phage particles, puromycin resistance gene for mammalian selection, antibiotic resistance for bacterial selection (ampicillin or hygromycin), 3′ self-inactivating long terminal repeat (SIN/3′ LTR) for increased lentivirus safety, 5′ long terminal repeat (5′ LTR), and RNA packaging signal (Psi).

Table 3.1. Plasmids used in this work and their origin.

PLASMID	CONSTRUCT	ORIGIN
pLKO control	Lentiviral empty vector.	Sigma-Mission®
pLKO shCTCF	Lentiviral shRNA for human *CTCF* gene.	Sigma-Mission®
pTRIPTZ control	TRIPTZ lentiviral empty vector.	Dharmacon™ GE Healthcare
pTRIPTZ shCTCF	TRIPTZ lentiviral shRNA for human *CTCF* gene, inducible by tetracyclines.	Dharmacon™ GE Healthcare
pCMV-VSV-G	*VSV-G* gene encoding envelope protein for producing lentiviral particles.	Addgene
psPAX2	*GAG* and *POL* genes encoding packaging lentiviral proteins.	Addgene

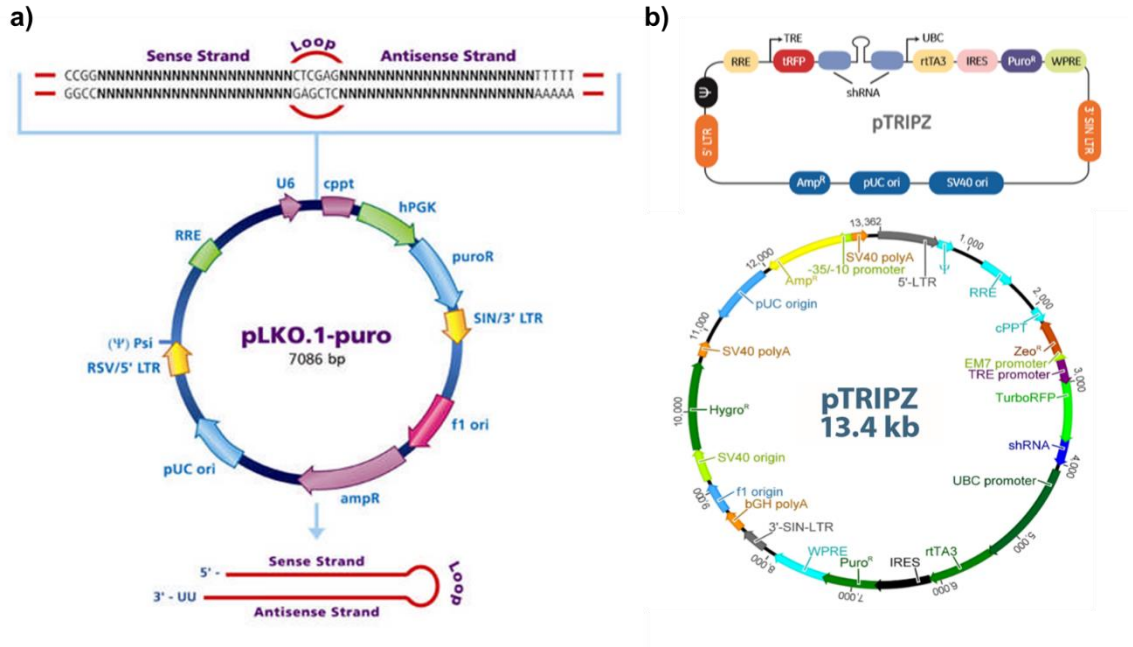


Figure 3.1. Maps of the vectors used in this work. Elements of a) lentiviral plasmid vector pLKO.1-puro and b) pTRIPZ inducible lentiviral shRNA vector. Adapted from Addgene.

Plasmid-containing bacteria were grown by streaking from a glycerol stock in LB-agar plate with ampicillin, which was incubated overnight at 37 ºC. A single colony was inoculated into 10 ml LB growth medium containing antibiotic (100 µg/ml ampicillin) and incubated for 6 h in an orbital shaking incubator at 37 ºC and 160 rpm. After incubation, the culture was added to 200 ml LB growth medium containing the same selection antibiotic and grown overnight in the same conditions. The following day, bacterial culture was centrifuged at 6000 rpm for 10 min at 4 ºC and plasmid DNA was purified following manufacturer´s instructions.

As a summary, for plasmid DNA extraction the cellular pellet was resuspended in a buffer with RNase to degrade RNA, then it was lysed with detergent at alkaline pH (SDS and NaOH) to break the plasma membrane and denaturalize proteins, chromosomal DNA and plasmid DNA. The pH was neutralized with potassium acetate, making the proteins and chromosomal DNA to precipitate, while the plasmid DNA remains in solution. The plasmid DNA was filtered and washed with an equilibrated column and then it was eluted with NaCl-Tris-HCl-isopropanol. It was centrifuged, the DNA pellet was washed with ethanol, and finally resuspended in MilliQ water. Plasmid DNA concentration was determined by measuring A_{260nm} using a microvolume spectrophotometer (Thermo Scientific™ NanoDrop 2000).

3.2.2. Lentivirus production

Lentiviral particles were produced to transduce different cell lines in order to downregulate CTCF expression. Lentivirus were produced in HEK293T transfected with the plasmid DNA using the Polyethylenimine (PEI) cationic lipids-based method. HEK293T were transfected with 3 distinct plasmids:

- **pCMV-VSV-G:** plasmid containing VSV-G gene encoding envelope protein.
- **psPAX2:** plasmid containing GAG and POL genes from human immunodeficiency virus (HIV) encoding packaging proteins.
- **The plasmid of interest** (pLKO EV, pLKO shCTCF, pTRIPTZ EV or pTRIPTZ shCTCF) containing the short hairpin RNA (shRNA) for gene knockdown and other elements mentioned before, necessary to produce lentiviral particles.

Cells were seeded in 150 mm plates at 70-80% of confluence and the complete medium D10F was changed to 13 ml of serum-free DMEM. The plasmids were mixed in the ratio 1:3:4 (pCMV-VSV-G:psPAX2:plasmid of interest), this means 6:19:25 µg for 50 µg of total DNA. 100 µg PEI (Polysciences, Inc.) and 1 ml of serum-free DMEM were added to the mixture and incubated 10 min at RT. The final mixture of DNA, PEI and DMEM was added to the cells and, after 10 hours, the medium was replaced by 15 ml of complete medium D10F. 48 hours after the transfection, the supernatant containing lentivirus was collected and stored at 4 ºC and 15 ml of complete medium were added. 72 hours after the transfection, the supernatant was collected and mixed with the previous one to be centrifuged 10 min at 1500 rpm. The supernatant was filtered through a 0.45 µm pore size sterile syringe filter (Merck Millipore) and the eluate was collected in a clean tube. 15% Polyethylene Glycol (PEG8000 40% in PBS) (Sigma-Aldrich) was added to make the lentiviral particles precipitate and they were stored at 4 ºC between 6 hours and one week. Finally, it was centrifuged for 30 min at 1500 rpm in order to concentrate the lentiviral particles, the pellet was resuspended in 150 µl serum-free medium and aliquots of 50 µl were stored at -80 ºC.

3.2.3. Lentivirus tittering

Lentiviral particles carrying antibiotic selection were tittered in HeLa cells grown in a concentration of 2×10^4 cell/well in a 6-well plate in 1.5 ml of D10F medium. The medium was replaced by 1.5 ml of serum-free DMEM and incompletely adhered cells were infected with different concentrations of lentivirus (0.1, 0.5, 1 and 5 µl), in the presence of 5 µg/ml Polybrene (Sigma-Aldrich). 12 hours after the infection, 1.5 ml of complete medium D10F was added to the plates. 48 hours after the infection, the complete medium was refreshed and the selective antibiotic (500 µg/ml puromycin, Sigma-Aldrich) was added to the plates to select the infected cells, except for the negative control. Puromycin selection was maintained until control cells died (approximately, 3 days) and in the infected cells there were countable colonies. Then, the medium was removed, and the wells were washed twice with PBS. Cell colonies were stained with Crystal Violet solution (1% crystal violet dye in methanol) for 15 min. The dye was washed, and the number of colonies was counted under the microscope. Assuming that each colony was infected by one lentiviral particle, the lentivirus tittering (in colony forming units (C.F.U.)/ml) was calculated as the quotient between the number of colonies and the volume of virus that was added to the counted well.

3.2.4. Cell transduction

K562 cells were seeded in a concentration of 2.5×10^5 cell/ml in 10 ml of serum-free RPMI medium. Cells were transduced with a multiplicity of infection (MOI) of 2 virus particle/cell and 5 µg/ml Polybrene was added to facilitate the infection. Cells were infected with pLKO EV and pLKO shCTCF lentiviral particles, produced as explained before. Cells were grown at 37 ºC o/n and, 12 hours after the infection, 10 ml of complete medium R10F were added to the cells. 48 hours after the infection, the medium was refreshed and 25 ml of D10F were added. Additionally, 1 µg/ml puromycin was added to select the infected cell population and the cells were grown 48 hours with the selective antibiotic at 37 ºC. At that point, erythroid differentiation was induced in transfected K562 cells with Ara-C or Imatinib in order to study the effect of CTCF downregulation in erythropoiesis.

$CD34^+$ cells were transduced with a MOI of 2 virus particle/cell. After 4 days growing, 5×10^5 cells were infected with 10^6 viral particles. For this purpose, plates were treated with 20 µg/µl retronectin for 2 h at room temperature (RT). Then, the plates were washed with an equal volume of PBS-2%BSA for 30 min at RT. When the plates were ready to be used, the inducible viral particles pTRIPRZ EV and pTRIPTZ shCTCF were added for 5 h at 37 ºC. Finally, the viruses were removed, the cells were added to the plates in a concentration of 1×10^6 cell/ml with medium without cytokines and they were incubated at 37 ºC. The following day, an equal volume of StemSpan™ SFEM II medium supplemented with StemSpan™ $CD34^+$ Expansion supplement 10X (StemCell Technologies) was added to the plates. 48 hours after the infection, the selection with puromycin 0.5 µg/ml was performed. 72 or 96 h after the infection, 1 µg/ml doxycycline was added in one half of cells infected with each virus. Doxycycline induces the silencing of CTCF in the cells transfected with pTRIPTZ shCTCF but has no effect on the cells transfected with pTRIPTZ EV. The medium was refreshed and the selection with 0.5 µg/ml puromycin was maintained. 48 hours after the addition of doxycycline, one half of each condition was separated in two new conditions: with and without Epo, which induces erythroid differentiation. The selection with puromycin and the induction of shCTCF expression with doxycycline was refreshed.

3.3. RNA EXTRACTION AND GENE EXPRESSION QUANTIFICATION

3.3.1. RNA extraction and purification

Total RNA was extracted and purified using TRIzolTM reagent (Invitrogen), which is a monophasic solution of phenol, guanidine isothiocyanate, and other components. It maintains the integrity of the RNA due to highly effective inhibition of RNase activity while disrupting cells. TRIzol™ Reagent is an improvement to the single-step RNA isolation method that was initially described with an acid guanidinium thiocyanate-phenol-chloroform mixture (Chomczynski & Sacchi, 1987). In the case of K562 cells, between 2-10 x 10^6 cells were collected by centrifugation and the pellet was stored at -80 ºC until RNA extraction. For lysis and homogenization, 1 ml of Trizol reagent was added per sample. After 5 min of incubation at RT to permit complete dissociation of the nucleoproteins complex, 0.2 ml of chloroform per 1 ml of Trizol Reagent was added, mixed, incubated 2-3 min at RT and centrifuged for 15 min at 12000 rpm and 4 ºC. The homogenate was separated into a colorless upper aqueous layer (containing RNA), an interphase, and a red lower organic layer (containing the DNA and proteins). The upper layer was collected in a new tube, mixed with 0.5 ml of isopropanol, incubated 10 min at RT and centrifuged for 20 min at 12000 rpm at 4 ºC to precipitate RNA. The supernatant was discarded, and the RNA pellet was washed with 75% ethanol to remove impurities and centrifuged for 5 min at 7500 rpm and 4 ºC. The supernatant was discarded, and the pellet was dried at RT for 10 min. Then, the precipitated RNA was resuspended in 25 µl of RNAse-free water for use in downstream applications. Finally, RNA concentration was determined by measuring A_{260nm} using a microvolume spectrophotometer. Isolated RNA was stored at -80 ºC and can be further used in RT-PCR.

3.3.2. Reverse transcription and quantitative polymerase chain reaction (RT-qPCR)

Gene expression was analyzed using quantitative real time polymerase chain reaction (RT-qPCR). For that purpose, complementary DNA (cDNA) was synthetized from RNA in a reverse transcription reaction, using the iScript™ cDNA Synthesis Kit (Bio-Rad). First, 1 µg of RNA was mixed with nuclease-free water in a total volume of 15 µl. Then, 4 µl of iScript Reaction Mix 5x were added, containing oligo(dT) and random hexamer primers. Finally, 1 µl of iScript Reverse Transcriptase was added. This enzyme is a modified Moloney murine leukemia virus reverse transcriptase that is preblended with RNase inhibitor. The complete reaction mix was incubated in a thermal cycler using the following protocol, according to manufacturer´s instructions: 5 min at 25 ºC (priming), 20 min at 46 ºC (reverse transcription) and 1 min at 95 ºC (RT inactivation). The cDNA was more stable than RNA and was stored at -20 ºC until used.

Quantitative PCR was performed to amplify the previously obtained cDNA using specific primers for gene of interest to quantify its expression, in terms of mRNA level. Primer sequences used in RT-qPCR assays are shown in **Table 3.2**. The mix for qPCR reaction was prepared mixing 11 µl of SyBR® Select Master Mix (Applied Biosystems), 8.7 µl of distilled water, 1.3 µl of specific primers (0.3 µM; stock concentration 100 mM in distilled water) and 1 µl of cDNA, for a final volume of 22 µl used for two duplicate wells. 10 µl of the mix were loaded in each well of a 96-well PCR plate. Reaction mix for each primer without cDNA was used as negative control to detect possible amplification from contaminating DNA or primer dimers. The qPCR reaction was performed in the CFX Connect™ Real-Time PCR Detection System (Bio-Rad). The qPCR protocol for cDNA amplification was the following: 40 cycles of

10 min at 95 ºC (5 min at 95 ºC; 30 s at 55-60 ºC), 1 min at 95 ºC and 1 min at 55 ºC. The protocol for real time melting curve was 80 cycles of 10 s starting at 55 ºC and decreasing by half a degree each cycle.

The CFX Connect Real-Time PCR Detection System was connected to a computer provided with the CFX Manager™ software to analyze the data. Threshold cycles (Ct) were determined by default at the beginning of DNA amplification in the exponential phase. The Ct value is associated with the amount of amplified product in the reaction: the lower the Ct value, the more PCR product that is present because it takes fewer PCR cycles for that product to be detected over the background signal.

The delta-delta Ct ($2^{-\Delta\Delta Ct}$) method (Rao et al., 2013) was used in order to calculate the relative gene expression of our samples, normalized to mRNA expression of the ribosomal protein S14. First, the average of the Ct for the replicates was calculated for all the genes. Next, ΔCt was calculated for all the samples as the difference of the average Ct of the gene of interest and the average Ct of the housekeeping gene (S14) from the same sample: ΔCt = Gene Ct – housekeeping gene Ct. After that, the average ΔCt value for the control group was calculated. Then, the ΔΔCt was calculated for all the samples as the difference between the ΔCt of all samples and the average ΔCt of the control: ΔΔCt = sample ΔCt – average control ΔCt. After that, the $2^{-(\Delta\Delta Ct)}$ was calculated, which represents the comparative fold gene expression values for all individual samples. Finally, the average $2^{-(\Delta\Delta Ct)}$ and the standard deviation were calculated.

Table 3.2. Primers used for RT-qPCR analysis.

Amplified gene	Primer sequences (5´→ 3´)	Amplicon size
CTCF	Fw: TTA CAC GTG TCC ACG GCG TTC	365 bp
	Rv: GCT TGT ATG TGT CCC TGC TGG CA	
ETS1	Fw: TCC AGA CAG ACA CCT TGC AG	153 bp
	Rv: TGA GGC GAT CAC AAC TAT CG	
MYB	Fw: AGC AAG GTG CAT GAT CGT C	157 bp
	Rv: GGG GGT GGA AGT TAA AGA AGG	
GATA2	Fw: CAA GAT GAA TGG GCA GAA CC	113 bp
	Rv: GCC ATA AGG TGG TGG TTG TC	
HEY1	Fw: GAC CGT GGA TCA CCT GAA A	123 bp
	Rv: ATT CCC GAA ATC CCA AAC TC	
LMO2	Fw: CTG AGC TGC GAC CTC TGT G	164 bp
	Rv: CGC ATT GTC ATC TCA TAG GC	
KLF1	Fw: CAG GTG TGA TAG CCG AGA CC	241 bp
	Rv: CCG TGT GTT TCC GGT AGT G	
NFE2L2	Fw: CGG TAT GCA ACA GGA CAT TG	246 bp
	Rv: AGA GGA TGC TGC TGA AGG AA	
MYC	Fw: TCG GAT TCT CTG CTC TCC TC	157 bp
	Rv: CCT GCC TCT TTT CCA CAG AA	
TCF3	Fw: CAT CTG CAT CCT CCT TCT CC	
	Rv: CCT CGT CCA GGT GGT CTT C	

Fw: forward primer; Rv: reverse primer.

3.4. PROTEIN EXTRACTION AND QUANTIFICATION

3.4.1. Cell lysis and protein extraction

Protein levels were analyzed by Western blot. Between 2 to 10 x 10^6 K562 cells were collected by centrifugation and the pellet was stored at -80 ºC until protein extraction. For cell lysis, 50 ml of RIPA lysis buffer were prepared, containing: 1.5 ml NaCl 5M, 2.5 ml Tris pH 6.8, 0.25 g Sodium deoxycholate (NaDOC), 500 µl NP-40 (Igepal®) and 500 µl SDS. Protease inhibitor and phosphatase inhibitor were added to the RIPA lysis buffer immediately before using. Cell pellets were lysate in 200 µl of lysis buffer per 5 x 10^6 cells and samples were incubated for 15 min on ice. Then, samples were sonicated at 4 ºC using the Bioruptor® Plus sonication device (Diagenode) with high power mode for 10 cycles, each sonication cycle consisting of 30 s ON, 30 s OFF. After sonication, cell lysates were incubated on ice and centrifuged for 20 min at 14000 rpm and 4 ºC. The pellets were discarded and the supernatants, which contained the proteins, were placed on a new tube, and stored at -80 ºC until protein analysis by Western blot.

3.4.2. Western blot

We performed Western blot to detect a specific protein from a sample and this technique consists basically in three steps: gel electrophoresis, blotting, and hybridization. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is used to separate proteins according to their molecular weights. We performed an initial Western blot with the same volume of all samples in order to equal the total protein concentration of our samples. The volume of each sample was mixed with distilled water and 5X Laemli buffer for a final concentration 1X in a final volume between 20-25 µl. Then, the samples were heated for 5 min at 95 ºC and loaded in an SDS-PAGE gel, which acrylamide percentage depends on the molecular weight of the analyzed protein. 8% acrylamide gels were used to analyze the biggest proteins, 15% acrylamide gels for the smallest proteins and percentages in between for intermediate proteins. As marker of known molecular weight, Precision PlusProtein™ Dual Colors Standars (Bio-Rad) was used.

Vertical electrophoresis was performed at constant voltage of 175 V for 1 hour and a half, using electrophoresis running buffer TGS (25 mM Tris, 192 mM glycine and 0.1% SDS). When protein migrated until the end of the gel, electrophoresis had finished and proteins were transferred to a nitrocellulose membrane AmershamProtan Supported 0.45 NC (GE Healthcare Life Sciences), with 0.45 µm pore size, in a Mini-Trans Blot cell (Bio-Rad) using transfer buffer TG-MeOH (25 mM Tris, 192 mM glycine, and 10% methanol). The transference was performed at constant amperage of 400 mA for 30 min to 1 hour, depending on the molecular weight of the protein of interest because the greater the time, the bigger the proteins that are transferred. In our conditions, we assumed 1 min per kDa, for proteins smaller than 60 kDa. The transference was performed on ice to avoid the overheating.

After the transference, the gel was stained with Coomassie Brilliant Blue solution (0.025% Coomassie Brilliant Blue R-250, 40% methanol and 10% acetic acid) for 10 min at RT with shaking and distained with water to check proteins load. To block unspecific binding sites, the membrane containing the proteins was incubated with TTBS (20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.05% Tween 20 in distilled-water) containing 10% non-fat dry milk for 1 h at RT. Then, the membrane was incubated with the primary antibody against the protein of interest (**Table 3.3**) overnight at 4 ºC or 4 h at RT in agitation.

Table 3.3. Primary antibodies used for Western blot.

Primary antibody	Molecular weight (kDa)	Specie	Dilution	% BSA (in TTBS)	Origin
α-CTCF	130	Mouse	1:1000	1	Santa Cruz
α-MYB	80	Rabbit	1:1000	5	Cell Signaling
α-MYC	67	Rabbit	1:3000	5	Cell Signaling
α-β-Tubulin	55	Mouse	1:1000	1	Santa Cruz
α-GATA1	46	Rabbit	1:1000	1	Santa Cruz
α-β-Actin	43	Mouse	1:3000	1	Santa Cruz
α-HEY1	33	Mouse	1:1000	1	Santa Cruz
α-LMO2	24	Mouse	1:1000	1	Santa Cruz
α-γ-Hemoglobin	14	Mouse	1:1000	1	Santa Cruz

After the incubation with primary antibody, the membrane was washed 3 times for 10 min with TTBS solution at RT with shaking. Then, the membrane was incubated with the corresponding fluorescent secondary antibody against the specie of primary antibody (**Table 3.4**) for 45 min at RT with shaking and in darkness. The membrane was washed three times for 10 min each with TTBS and the immunocomplexes were detected with the Odyssey® Infrared Imaging System (LI-COR, Biosciences). Finally, the membranes were scanned with the Odyssey® software and the amount of protein was quantified with ImageJ and normalized to a constitutively expressed protein, such as actin or tubulin.

Table 3.4. Secondary antibodies used for Western blot.

Secondary antibody	Type	Origin	Dilution	% BSA (in TTBS)
Anti-Mouse IRDye®680	Donkey polyclonal	LI-COR (926-68072)	1:10000	1
Anti-Mouse IRDye®800	Donkey polyclonal	LI-COR (926-32212)	1:10000	1
Anti-Rabbit IRDye®680	Donkey polyclonal	LI-COR (926-68073)	1:10000	1
Anti-Rabbit IRDye®800	Donkey polyclonal	LI-COR (926-32213)	1:10000	1

3.5. ENCODE ANALYSIS

The Encyclopedia of DNA Elements (ENCODE) is a public research consortium funded by the National Human Genome Research Institute (NHGRI) and it is aimed at identifying all functional elements in the human and mouse genomes. All ENCODE production data is submitted to the Data Coordinating Center at the University of California, Santa Cruz (UCSC). The human assembly of February 2009 (GRCh37/hg19) of the UCSC Genome Browser (https://www.genome.ucsc.edu/) was used in this work.

On one hand, chromatin immunoprecipitation sequencing (ChIP-seq) data were used to identify CTCF binding sites to the regulatory regions of the TFs involved in erythroid differentiation. This analysis was performed, not only in K562, but also in other human cell lines, to see if the interactions were cell type-specific or if, however, they were conserved among different cell lines. ENCODE analysis for the subunit of cohesin complex Rad21 was also carried out to confirm the predicted CTSs.

On the other hand, Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) data were used to analyze the genome-wide long-range chromatin interactions bound by CTCF, involving the CTCF binding site to the erythroid genes and other CTCF binding site (CTS). This analysis was also performed in K562 and other human cell lines available to make the comparison.

Finally, Histone Modifications by ChIP-seq from ENCODE in the CTCF binding site to the genes of interest were analyzed. Three active histone marks (H3K4me1, H3K4me3 and H3K9Ac) and one repressive histone mark (H3K27me3) were analyzed in all human cell lines available.

3.6. CBIOPORTAL ANALYSIS

The cBioPortal for Cancer Genomics (http://cbioportal.org) provides genomics data from different large-scale projects, including The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). Three studies of AML were used in this work: AML (OHSU, Nature 2018), AML (TCGA, PanCancer Atlas) and Pediatric AML (TARGET, 2018). Ten genes were queried for mutations and copy number alterations: *CTCF, ETS1, MYB, GATA2, HEY1, LMO2, KLF1, NFE2L2, TCF3* and *MYC*. The portal shows different results:

- **OncoPrint** was used to analyze the genomic alterations in queried genes across the samples, including missense mutations, nonsense mutations, truncating mutations, amplifications, and deletions.
- The **Mutual Exclusivity tab** was used to analyze if two mutations tend to be mutually exclusive or, instead, two genetic alterations of different genes occurred simultaneously in the same sample.
- The **Mutations tab** was used to analyze the position of each mutation in the queried genes and their frequency, as well as the amino acid change in the protein sequence. Cancer hotspots were further analyzed in the website https://www.cancerhotspots.org/ and OncoKB (https://www.oncokb.org/) was used to find more information about specific alterations. Additionally, those positions were localized in the UCSC Genome Browser to search for possible correlations between mutations in erythroid transcription factor genes and CTCF binding or chromatin interactions in AML.
- The **Overall survival** graph was used to compare the overall survival of patients without mutations and the survival or patients with mutations in the genes of interest.
- The **CN (copy number) segment** data was downloaded and visualized in the Integrative Genomics Viewer (IGV) in order to analyze the copy-number status of all queried genes. IGV provides a quantitative result of the cases with amplifications and deletions in the genes of interest.
- The **RNA seq** data were analyzed to determine if the genetic mutations in the queried genes encompass changes in the mRNA expression compared with the average of the group.

3.7. STATISTICAL ANALYSIS

Data were represented as the mean of two or three independent experiments or one single experiment. Error bars represent the positive and negative standard deviation. The significance of differences was determined by the unpaired Student´s *t*-test. The differences were considered statistically significant when $p<0.05$ (*$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$).

4. RESULTS

4.1. EFFECTS OF CTCF DOWNREGULATION IN THE ERYTHROID CELL DIFFERENTIATION

K562 cells can be induced to differentiate into several hematopoietic pathways in response to the treatment with different drugs (Torrano et al., 2005). For example, K562 can be differentiated into the erythroid lineage by treatment with Ara-C (Delgado et al., 1995; Torrano et al., 2005) or Imatinib (Gómez-Casares et al., 2013). To know more about the role of CTCF in erythropoiesis, we aimed to down-regulate the expression of CTCF in K562 cells and analyze erythroid differentiation upon induction by treatment. First, we produced lentiviral particles containing a specific shRNA for *CTCF* gene (shCTCF) and for the corresponding empty vector pLKO (EV). K562 cells were infected with the lentiviral particles, selected with puromycin for two days and induced for erythroid differentiation with 1 µM Ara-C and 0.5 µM Imatinib (**Figure 4.1a**). CTCF downregulation was assessed by RT-qPCR (**Figure 4.1b)** and Western blot (**Figure 4.1c)**. CTCF levels decreased in K562 cells infected with pLKO shCTCF, compared with those infected with the empty vector, and the downregulation was maintained for 3 days. Although there was a slightly reduction in CTCF levels upon differentiation with the empty vector, the reduction of CTCF was strongly higher with shCTCF.

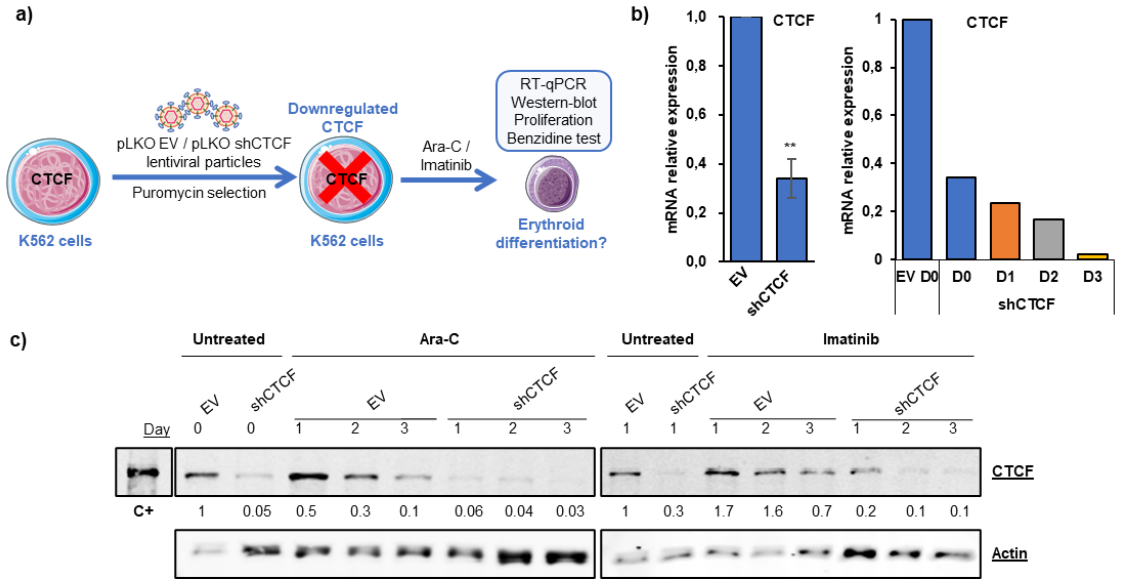


Figure 4.1. Assessment of CTCF downregulation in K562 cells upon erythroid differentiation. a) Schematic representation of the experiment, including the infection with lentiviral particles pLKO empty vector (EV) and pLKO shCTCF, the selection with puromycin for 48 hours, the induction of erythroid differentiation with 1 µM Ara-C or 0.5 µM Imatinib and its subsequent analysis. b) Assessment of CTCF downregulation in K562 cells at day 0 (left) and its maintenance for 3 days (right) with mRNA expression of CTCF analyzed by RT-qPCR. Expression was normalized against RPS14 levels. Bars indicate mean of three experiments (Day 0) ± s.d. or one representative experiment (Days 1, 2 and 3). Significance difference (**, p<0.01) from the untreated cells. c) Assessment of CTCF downregulation in K562 cells and its maintenance for 3 days with protein expression levels of CTCF analyzed by Western blot. Actin levels were used as loading control. Densitometry values are shown at the bottom, normalized to the control. C+, 293T cells transfected with pcDNA-CTCF as positive control.

Cell proliferation was analyzed by cell counting until 3 days after induction of differentiation and untreated cells infected with the empty vector were considered the control (**Figure 4.2a**). On one hand, the cells that growth faster were K562 cells infected with the EV and untreated. Both Ara-C and Imatinib

treatments inhibited cell proliferation in K562 cells almost completely, compared to the control. Imatinib induced cell death the last day of treatment, as previously observed. On the other hand, CTCF downregulation provoked a slight reduction in cell proliferation, highlighting that CTCF is required for cell survival and proliferation.

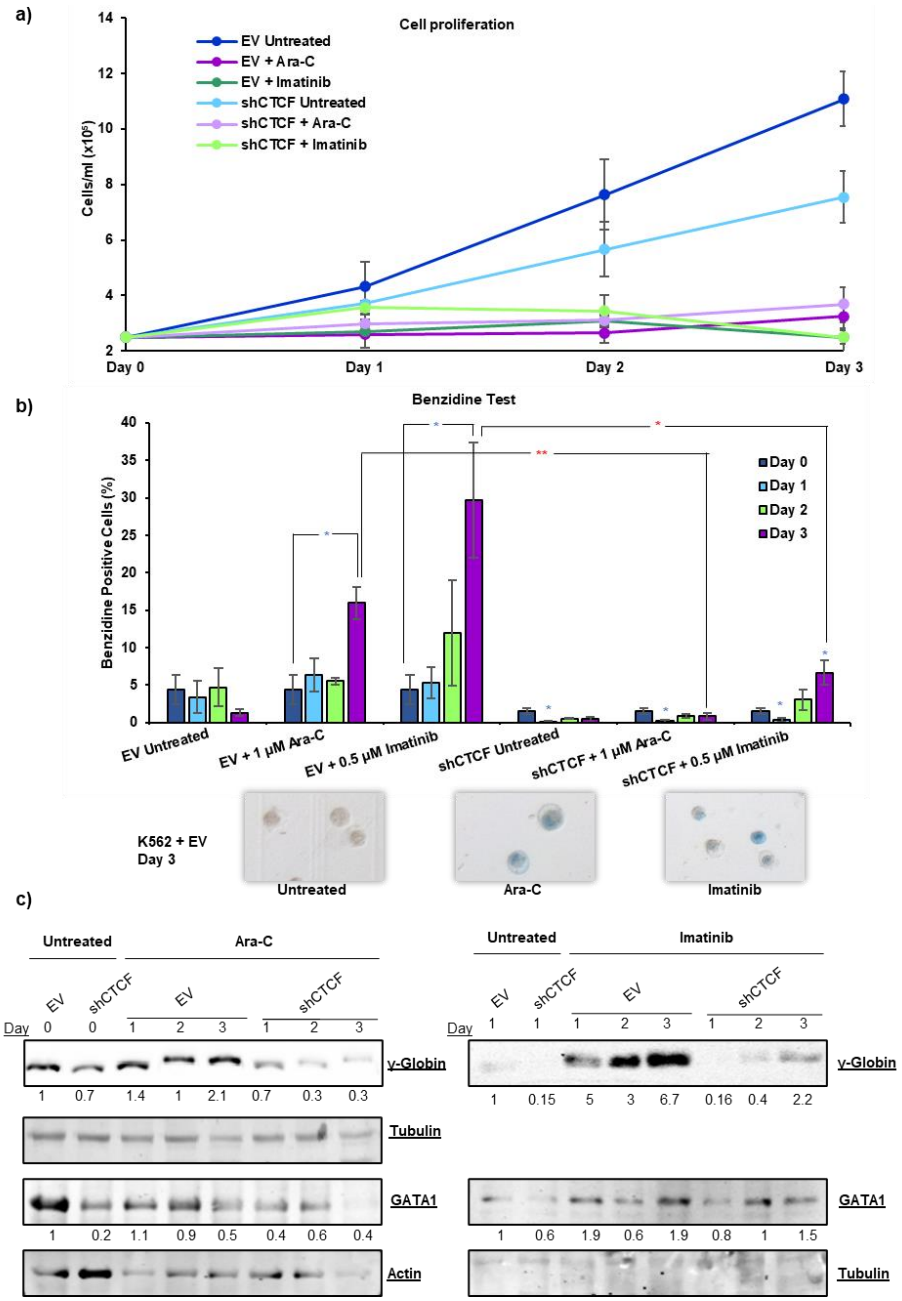


Figure 4.2. Effects of CTCF downregulation on erythroid differentiation of K562 cells. a) Cell proliferation curves for K562. 2.5 x 10^5 cells were seeded, infected with pLKO EV or shCTCF and treated with 1 µM Ara-C or 0.5 µM Imatinib for 3 days. Bars indicate mean ± s.d. of 3 independent experiments. b) Benzidine test of K562 cells infected with pLKO EV or shCTCF and treated as above. Under the bar charts, pictures taken during benzidine test for K562 cells infected with pLKO EV at day 3 of treatment are shown. Bars indicate mean ± s.d. of 2 or 3 independent experiments. Significance difference (*, $p<0.05$; **, $p<0.01$) from the untreated cells (blue asterisks) or from the pLKO EV treated cells (red asterisks). c) Protein expression of γ-globin and GATA1 analyzed by Western blot after infection of K562 cells with pLKO EV or shCTCF and treatment with Ara-C and Imatinib for 3 days. Actin or tubulin levels were used as loading control. Densitometry values are shown at the bottom, normalized to the control.

We assessed the erythroid differentiation using the benzidine test that analyzes hemoglobin production, exclusive of erythroid cells (**Figure 4.2b**). We confirmed the erythroid differentiation induced

by Ara-C and Imatinib in K562 cells up to 3 days with an increase in the percentage of benzidine positive cells compared to the untreated cells that was statistically significant for the day 3 of treatment. The effect of Imatinib was higher than the effect of Ara-C and the increase in the percentage of benzidine positive cells was gradual. Regarding with the changes in cell morphology and size, we observed that the cells treated with Ara-C, in addition to becoming blue, were bigger than the untreated cells and the cells treated with Imatinib were smaller.

Furthermore, there was a reduction of benzidine positive cells in the cells infected with shCTCF compared to the ones infected with the EV at day 0, before the treatment with Ara-C or Imatinib. When we compared the K562 cells infected with shCTCF with the ones infected with the EV at day 3 of treatment with Ara-C or Imatinib, we observed a significant reduction in the percentage of benzidine positive cells upon CTCF downregulation.

Additionally, erythroid differentiation was assessed by detecting the hemoglobin protein, which is exclusive of erythroid cells, by Western blot (**Figure 4.2c**). The expression of the γ-globin increased upon Ara-C and Imatinib treatment within 3 days compared with untreated cells, while γ-globin levels remained low upon CTCF downregulation. We analyzed the expression of another key erythroid marker, GATA1, which also decreased upon CTCF downregulation.

Taken together, these results show that CTCF downregulation strongly reduces the percentage of hemoglobin producing cells and the protein levels of erythroid markers, indicating that CTCF downregulation could inhibit erythroid cell differentiation induced by Ara-C and Imatinib and spontaneous erythroid differentiation of untreated cells. Therefore, CTCF seems to be necessary for erythroid differentiation of K562 cells.

In order to confirm the role of CTCF in erythropoiesis, similar experiments were performed with $CD34^+$ progenitor cells obtained from cord blood of newborns, as a more physiological model. First, we isolated the mononuclear cells by density gradient centrifugation on Ficoll and then we purified $CD34^+$ cells using magnetic beads. Expanded $CD34^+$ cells were treated with Epo to induce erythroid differentiation, which was confirmed with the benzidine test and the expression of γ-globin and GATA1 proteins by Western blot (data from Lorena García PhD Thesis). In order to study the effects of CTCF downregulation, $CD34^+$ cells were infected with inducible lentiviral particles containing pTRIPTZ EV and pTRIPTZ shCTCF and selected for puromycin resistance for 2 days. Then, treatment with doxycycline was carried out for 2 days to induce the knock down of CTCF and erythroid differentiation was induced with Epo treatment for 5 days (**Figure 4.3**). In $CD34^+$ cells infected with the EV, independently of doxycycline induction, we observed erythroid differentiation, as we expected. In $CD34^+$ infected with shCTCF, we observed erythroid differentiation without doxycycline induction, but erythroid differentiation was inhibited with doxycycline induction, that is, upon CTCF downregulation (these are preliminary data obtained by Lorena Garcia that I couldn't pursue and extend given the extraordinary sanitary situation).

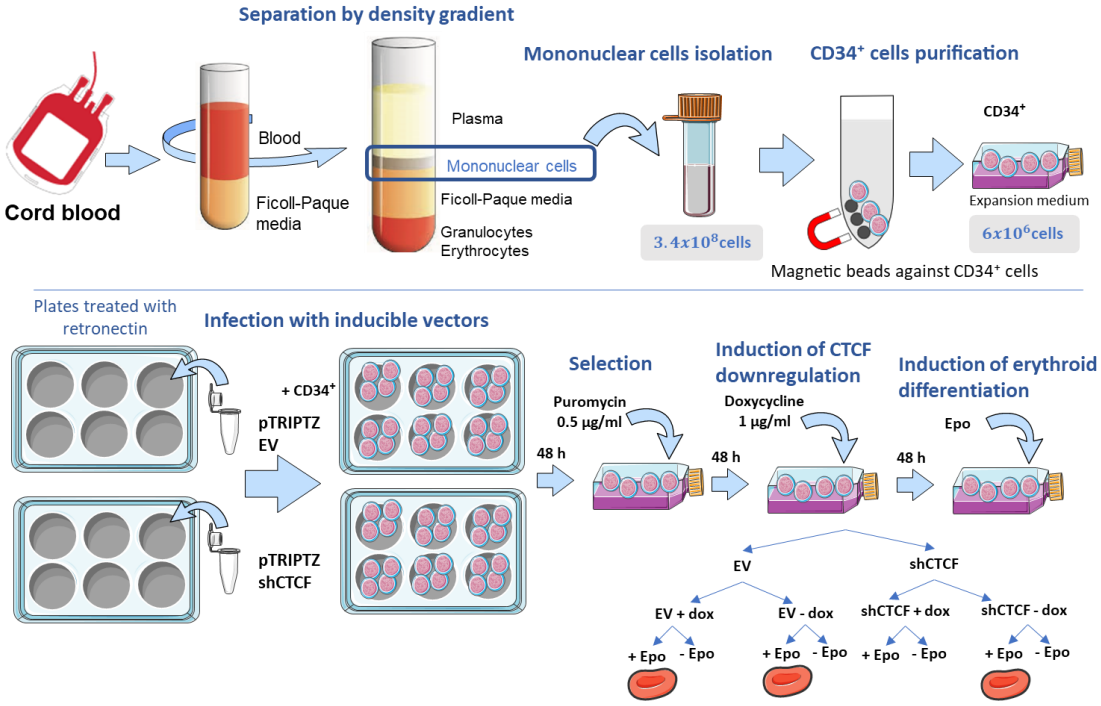


Figure 4.3. Effects of CTCF downregulation on erythroid differentiation of CD34⁺ cells. Schematic representation of the experiment, including the infection with lentiviral particles pTRIPTZ empty vector (EV) and pTRIPTZ shCTCF, the selection with puromycin for 48 hours, the induction of CTCF downregulation with doxycycline (dox) for 48 hours and the erythroid differentiation with Erythropoietin (Epo).

Taken together, our results reveal that CTCF downregulation, with constitutive or inducible systems, inhibits erythroid differentiation in K562 and primary CD34⁺ cells, highlighting the role of CTCF in the regulation of erythropoiesis.

4.2. REGULATION OF ERYTHROID TRANSCRIPTION FACTORS MEDIATED BY CTCF

4.2.1. Erythroid genes mRNA expression

Once we have demonstrated that CTCF has an essential role in erythroid differentiation, we aimed to analyze how CTCF modulates the expression of certain erythroid genes. Previous work of our group revealed, with microarray analysis, that CTCF overexpression in K562 cells modulates the expression of genes encoding, among others, several transcription factors that determine erythroid lineage differentiation.

In this work, we selected some important transcription factors implicated in erythroid differentiation: 4 TFs that inhibit erythroid differentiation (ETS1, MYB, GATA2 and HEY1) and 4 that induce erythroid differentiation (LMO2, KLF1, NFE2L2 and TCF3). We first analyzed their mRNA expression upon induction of erythroid cell differentiation with Ara-C or Imatinib for the 3 days of treatment, and the results that were more consistent with previous results of our group were represented in **Figure 4.4a**. As preliminary results, we observed a reduction in the expression of *ETS1*, *MYB*, *GATA2* and *HEY1* upon differentiation and an increase in the expression of *LMO2* and *KLF1*. In the case of *NFE2L2*, its expression did not change upon differentiation, and *TCF3* expression was slightly increased with Ara-C treatment and did not change with Imatinib. In general, these results are in concordance with the role of those genes during differentiation although more experiments are needed to confirm them.

We next analyzed the effect of CTCF downregulation in the mRNA expression of erythroid

transcription factor genes, prior to Ara-C or Imatinib treatment (day 0 of treatment) (**Figure 4.4b**). Because of the role of CTCF in erythropoiesis, we expected that the expression of the genes that ihibit erythroid differentiation decreased upon CTCF downregulation and the expression of those that induce erythroid differentiation increased. As preliminary results, we observed a reduction in the expression of *MYB*, *HEY1*, *LMO2* and *KLF1* upon CTCF downregulation and an increase in the expression of *NFE2L2*, while the expression of *ETS1*, *GATA2* and *TCF3* have hardly changed.

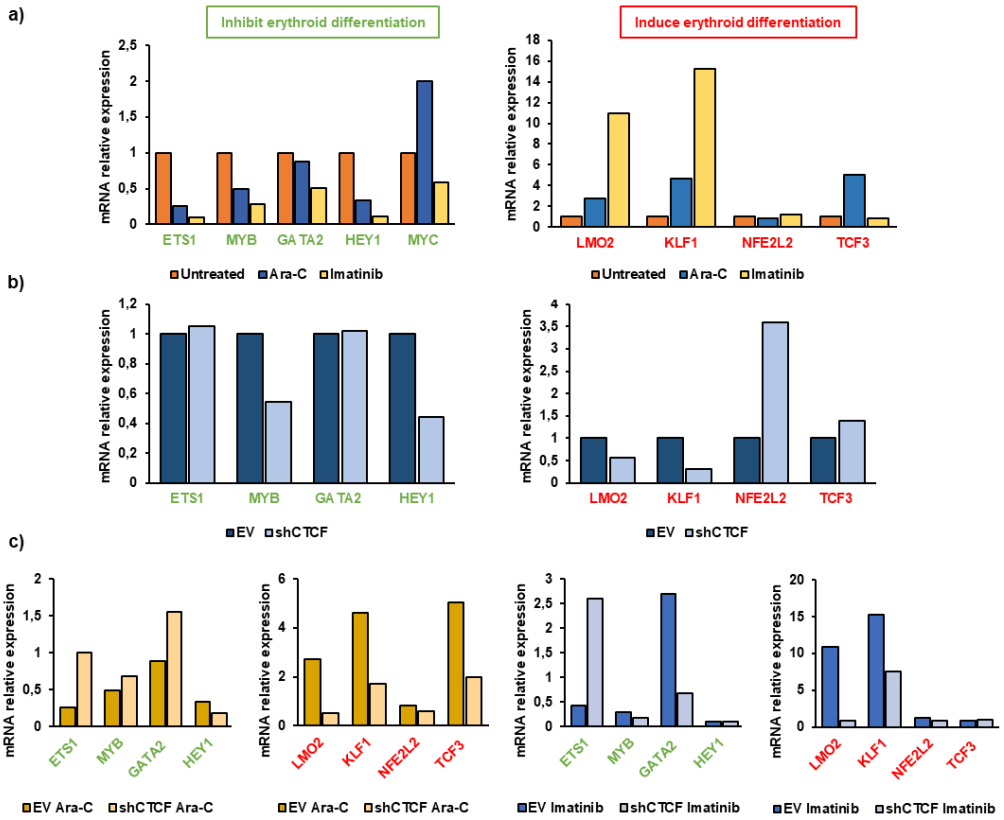


Figure 4.4. Erythroid genes mRNA expression. mRNA expression levels of *ETS1, MYB, GATA2, HEY1, LMO2, KLF1, NFE2L2* and *TCF3* analyzed by RT-qPCR in K562 cells upon: a) erythroid differentiation induced by treatment with 1 µM Ara-C for 72 hours or with 0.5 µM Imatinib for 24 (left) or 48 (right) hours, b) infection of K562 cells with pLKO EV or pLKO shCTCF without Ara-C or Imatinib treatment, and c) infection of K562 cells with pLKO EV or shCTCF and treated with 1 µM Ara-C for 72 hours or with 0.5 µM Imatinib for 48 hours. Expression was normalized against RPS14 levels. Bars indicate mean of two experiments or one representative experiment. mRNA expression values were relativized to the values obtained for untreated cells infected with pLKO EV that are represented in b).

In order to better stablish the role of CTCF during erythroid cell differentiation, we analyzed the effect of CTCF downregulation in the mRNA expression of erythroid transcription factor genes in K562 cells treated with 1 µM Ara-C for 72 hours or with 0.5 µM Imatinib for 48 hours (**Figure 4.4c**). As preliminary results, we observed an increase in the expression of the genes that inhibit erythroid differentiation (*ETS1*, *MYB* and *GATA2*) in cells with shCTCF treated with Ara-C, while a decrease in *HEY1* expression was observed. A reduction in the expression of the genes that induce erythroid differentiation (*LMO2*, *KLF1*, *NFE2L2* and *TCF3*) was observed. In K562 cells treated with Imatinib, we observed an increase in the expression of *ETS1* upon CTCF downregulation and a decrease in the expression of *GATA2*, *LMO2* and *KLF1*. The expression of *MYB*, *HEY1*, *NFE2L2* and *TCF3* has hardly changed upon Imatinib treatment in the cells with downregulated CTCF. Nevertheless, those are preliminary results and those experiments should be repeated a few times to perform statistical analysis.

4.2.2. Erythroid transcription factors protein expression

In order to confirm and extend the results obtained with RT-qPCR, we next aimed to analyze by Western blot the effects of CTCF downregulation on the protein expression levels of some erythroid transcription factors, such as LMO2, MYB and HEY1 during erythroid differentiation induced by Ara-C (**Figure 4.5a**) and Imatinib (**Figure 4.5b**). When we compared the K562 cells infected with shCTCF with the ones infected with the EV, prior to Ara-C or Imatinib treatment, we observed a decrease in LMO2 and MYB protein levels. LMO2 levels increased during erythroid differentiation induced by Ara-C and Imatinib, as we expected due to its role in the induction of erythropoiesis. When we downregulated CTCF, we observed an increase in the LMO2 protein levels upon erythroid differentiation induced by Ara-C and Imatinib during the 3 days of the treatment.

MYB protein levels were reduced during erythroid cell differentiation induced by Ara-C and Imatinib for 3 days, as we expected due to its role in the inhibition of erythropoiesis. When we downregulated CTCF, we observed a reduction in MYB protein levels upon erythroid differentiation induced by Ara-C and Imatinib during the 3 days of the treatment.

The expression of HEY1 has been normalized to K562 cells infected with pLKO EV treated with Ara-C at day 1 of treatment. We observed a slightly decrease in the expression of HEY1 protein levels in the K562 cells infected with the EV during erythroid differentiation induced by Ara-C and an increase in the ones infected with shCTCF.

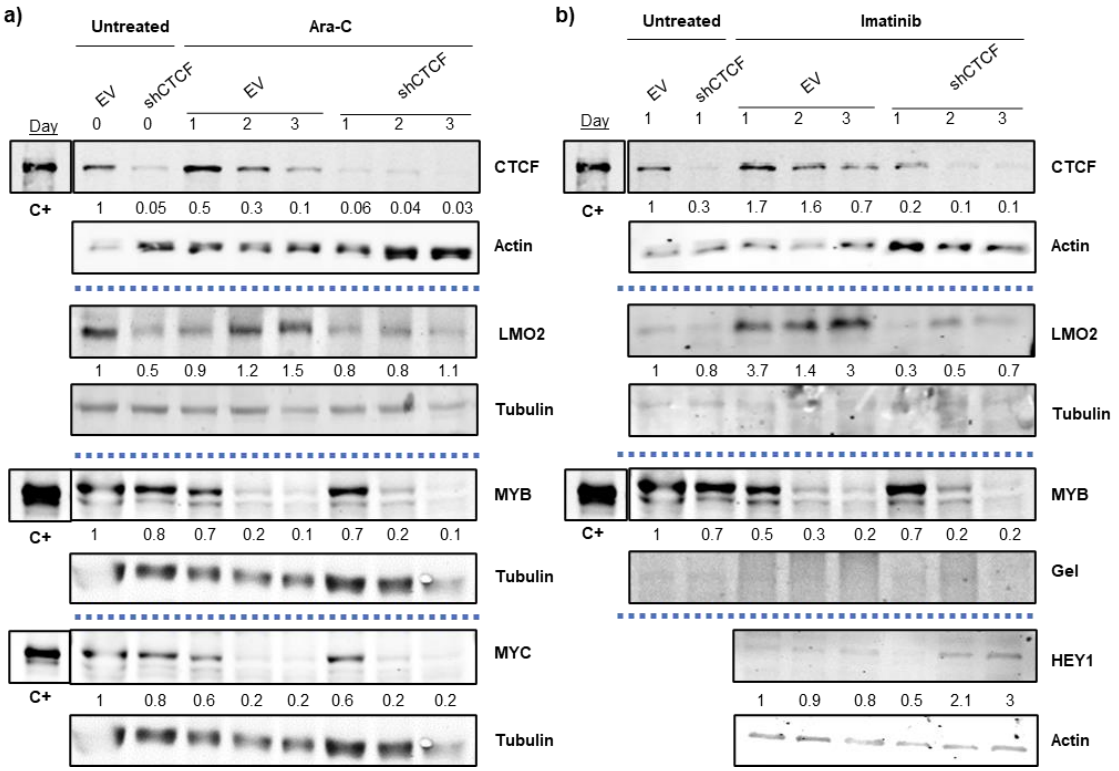


Figure 4.5. Effect of CTCF downregulation on erythroid transcription factors protein expression. Protein expression of CTCF, LMO2, MYB, HEY1 and MYC analyzed by Western blot after infection of K562 cells with pLKO EV or shCTCF and treated with a) 1 µM Ara-C or b) 0.5 µM Imatinib for 3 days. Actin or tubulin levels or coomasie-blue stained gel were used as loading control. Densitometry values are shown at the bottom, normalized to the control. C+, 293T cells transfected with pCEFL-CTCF as positive control of CTCF and Jurkat cells as positive control of MYB and MYC.

We also analyzed the expression of MYC protein because it plays an important role during cell proliferation and differentiation and it has been widely studied in our group. Previous results of our group confirmed that MYC expression decreases during erythroid differentiation (Delgado et al., 1995). MYC levels decreased upon CTCF downregulation and this reduction was maintained during erythroid differentiation induced by Ara-C.

4.3. ENCODE ANALYSIS IN K562 AND OTHER HUMAN CELL LINES

We have demonstrated that CTCF has an essential role in erythroid differentiation and we hypothesized that CTCF could be regulating specific erythroid genes. Therefore, we analyzed *in silico* the CTCF binding to the regulatory regions of selected genes encoding erythroid transcription factors using ENCODE Project data available on the UCSC Genome Browser.

4.3.1. CTCF binding sites to erythroid genes

We analyzed the ChIP-seq data published in the ENCODE project to identify possible CTCF binding sites (CTSs) to the selected erythroid genes: *ETS1*, *MYB*, *GATA2*, *HEY1*, *LMO2*, *KLF1*, *NFE2L2* and *TCF3* in five human cell lines (**Figure 4.6a, Supplementary figures S1-S7 a**) (Due to space limitations, most figures are shown as Supplementary information).

- **GATA2** is a TF highly expressed in HSCs that is repressed by GATA1 during erythropoiesis. ENCODE ChIP-seq analysis revealed a CTCF binding site 5 kb upstream *GATA2* gene in K562 cells (**Figure 4.6a**). This CTS is also a binding site of the Rad21 subunit of cohesin complex, which strongly supports CTCF binding in this position. This CTS is conserved in lymphoblastoid (GM78), lung adenocarcinoma (A549), lung fibroblast (IMR90) and neuroblastoma (SK-N-SH) cell lines.

- **ETS1** is a TF that inhibits erythroid differentiation, and, during erythropoiesis, its expression must be downregulated. ENCODE ChIP-seq analysis revealed a CTCF/Rad21 binding site 43 kb upstream *ETS1* gene (between *ETS1* and *FLI1* genes) in K562 cells that is conserved in different cell lines (**Supplementary Fig. S1a**).

- **MYB** is a TF that is highly expressed in HSCs and must be downregulated during erythroid cell differentiation. ENCODE ChIP-seq analysis revealed a CTCF/Rad21 binding site in the intron 1 of *MYB* gene (2.5 kb downstream from the TSS) in K562 cells (**Supplementary Fig. S2a**). This CTS is conserved in different cell lines. However, in some of them, such as A549, IMR90 and SK-N-SH, a novel CTS 15.5 kb upstream *MYB* gene show a higher peak than the CTS in the intron 1.

- **HEY1** is a TF that inhibits the expression of *GATA1*, maintaining an undifferentiated state of erythroid precursors. ENCODE ChIP-seq analysis revealed a CTCF/Rad21 binding site in the exon 5 of *HEY1* gene (2.1 kb downstream from the TSS) in K562 cells that is conserved in different cell lines (**Supplementary Fig. S3a**).

- **LMO2** is a non-DNA binding component of the CEN that promotes the expression of erythroid genes. ENCODE ChIP-seq analysis revealed a CTCF/Rad21 binding site 34 kb downstream *LMO2* gene in K562 cells that is conserved in different cell lines (**Supplementary Fig. S4a**).

- **KLF1** is a TF of the CEN that is essential for terminal erythroid cell differentiation. ENCODE ChIP-

seq analysis revealed a CTCF/Rad21 binding site in the exon 2 of *KLF1* gene in K562 cells, which is conserved in different cell lines (**Supplementary Fig. S5a**).

- **NRF2**, encoded by *NFE2L2*, is an erythroid TF involved in the regulation of β-globin gene transcription. There is a CTCF/Rad21 binding site in the intron 1 of *NFE2L2* gene in K562 cells that is conserved in different cell lines (**Supplementary Fig. S6a**). Some cell lines, including GM, A549, IMR90 and SK-N-SH, show other CTS in the exon 1 of *NFE2L2* gene, higher than the CTS in the intron 1 of *NFE2L2*.

- **TCF3**, also called E2A, is a TF that regulates HSC differentiation and interacts with the CEN to promote erythroid gene transcription. There is a CTS 23 kb upstream *TCF3* gene in K562 cells, which is also a Rad21 binding site, and it is conserved in different cell lines (**Supplementary Fig. S7a**).

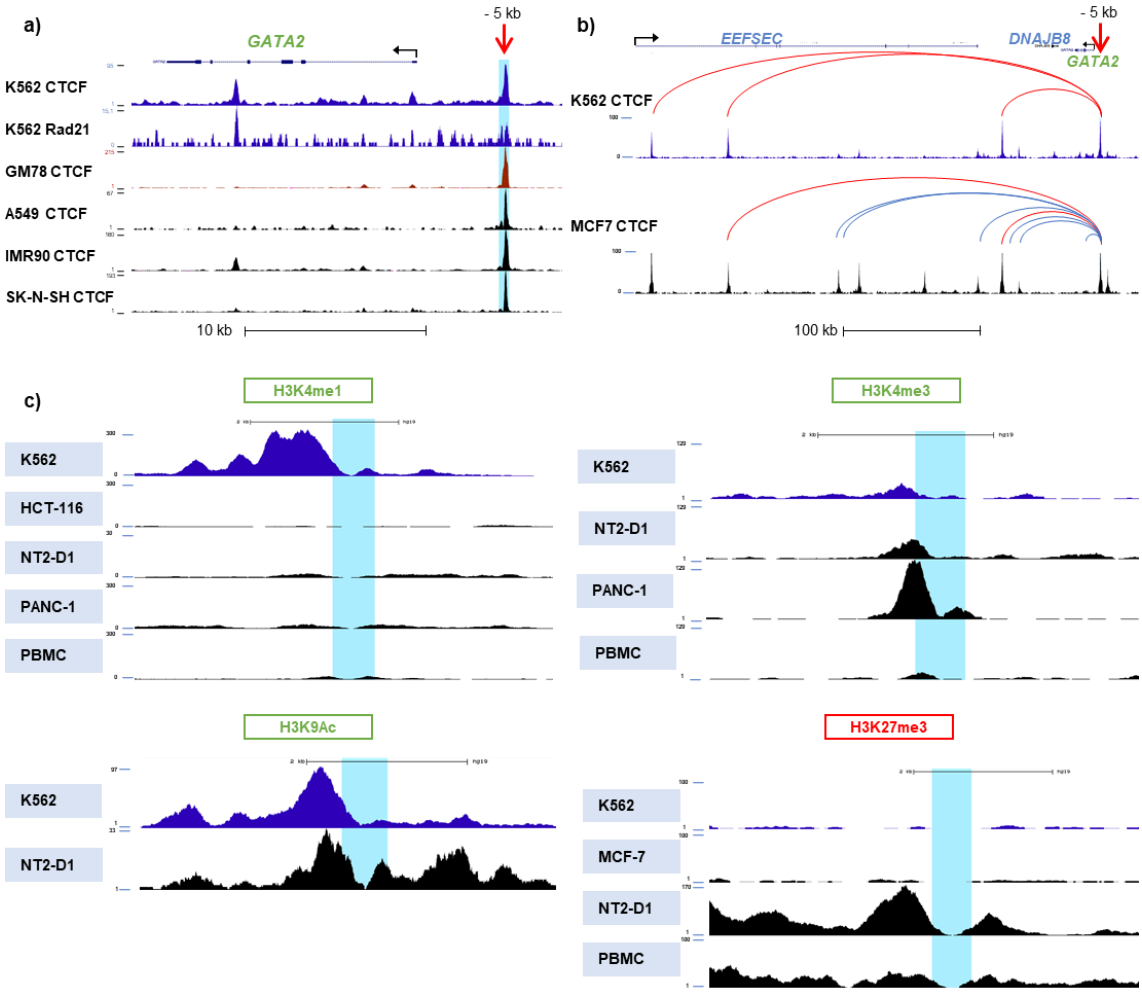


Figure 4.6. Analysis of CTCF binding to *GATA2* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *GATA2* gene in chromosome 3 from 5 human cell lines. CTCF binding site 5 kb upstream *GATA2* gene is shown with a red arrow. b) ChIA-PET interactions from the *GATA2* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *GATA2* CTS (highlighted in blue) in different cell lines.

Once identified the CTCF binding sites to the selected erythroid transcription factor genes, the next aim should be to perform chromatin immunoprecipitation experiments to confirm the CTCF occupancy at these sites in K562 cells during erythroid differentiation as well as in other cell lines. Part of these experiments in K562 have been already performed in our lab.

4.3.2. Chromatin interactions involving erythroid genes

We analyzed the ChIA-PET data published in the ENCODE project to identify possible interactions involving the previously identified CTSs around erythroid genes (**Figure 4.6b, Supplementary figures S1-S7 b).**

- ENCODE ChIA-PET data for ***GATA2*** gene revealed three interactions including the CTS 5 kb upstream *GATA2*, all of them located downstream (75, 280 and 335 kb) *GATA2* in K562 cells (**Figure 4.6b, red arcs)**. The smallest loop includes *DNAJB8* gene and the biggest one includes also *EEFSEC* gene. The longest loop is not conserved in breast adenocarcinoma (MCF-7) cell line, which present six new interactions around *GATA2* (blue arcs).

- ENCODE ChIA-PET data for ***ETS1*** revealed two interactions between the CTS upstream *ETS1* in K562 cells and two CTSs located downstream (440 and 300 kb) *ETS1* gene (**Supplementary Fig. S1b**). Because the loops contain *ETS1*, CTCF may function as an insulator between *ETS1* and *FLI1* genes. These two interactions are conserved in MCF-7 cells, that present two new interactions.

- ENCODE ChIA-PET data for ***MYB*** revealed one interaction involving the CTS in the intron 1 of *MYB* and one CTS 40 kb upstream *MYB* in K562 cells (**Supplementary Fig. S2b)**. This loop is conserved in MCF-7 cells, and in this cell line there was a new interaction with a CTS downstream *MYB*.

- ENCODE ChIA-PET data for ***HEY1*** revealed one loop involving the CTS in the exon 5 of *HEY1* and other CTS 250 kb upstream *HEY1* in K562 cells (**Supplementary Fig. S3b)**. The loop also contains the *MRPS28* gene (mitochondrial ribosomal protein S28). The interaction is conserved in MCF-7 cell line, and there is a new loop involving the CTS in the exon 5 of *HEY1*.

- ENCODE ChIA-PET data for ***LMO2*** revealed four loops that involve CTS 34 kb downstream *LMO2* in K562 cells, all of them are located upstream *LMO2* (**Supplementary Fig. S4b)**. Three interactions involving the CTCF binding site downstream *LMO2* are conserved in MCF-7 cell line, but the smallest loop is not conserved and there is a new interaction, which is longer.

- ENCODE ChIA-PET data for ***KLF1*** revealed one loop formed between the CTS in the exon 2 of *KLF1* and a CTS located 34 kb upstream *KLF1* in K562 cells (**Supplementary Fig. S5b)**. The loop also contains the *GCDH* gene (glutaryl-CoA dehydrogenase) and *SYCE2* gene (Synaptonemal complex central element protein 2). The interaction involving the CTS in the exon 2 of *KLF1* gene is not conserved in MCF-7 cell line, which have a new longer interaction.

- ENCODE ChIA-PET data for ***NFE2L2*** revealed two interactions involving the CTS in the intron 1 of *NFE2L2* gene in K562 cells. One loop is formed with a CTS in the exon 1 of *NFE2L2* gene and other loop is formed with a CTS 105 kb downstream *NFE2L2* gene (**Supplementary Fig. S6b)**. The biggest loop includes *HNRNPA3* gene (pre-mRNA splicing). Any interaction was observed involving the CTS in the intron 1 of *NFE2L2* gene in MCF-7 cells. However, two loops are formed around the *NFE2L2* gene in MCF-7 cells, involving the CTS in the exon 1 of *NFE2L2*.

- ENCODE ChIA-PET data for ***TCF3*** revealed seven interactions involving the CTS upstream *TCF3* in K562 cells. Two loops are formed with two CTSs located downstream (100 and 43 kb) *TCF3* gene and

five loops with CTSs located upstream (1, 116, 141, 142 and 163 kb) *TCF3* gene. In the loops that are formed with CTSs located downstream *TCF3*, CTCF could act as an insulator between *TCF3* and *ONECUT3*, *ATP8B3* and *REXO1* genes. Four of the interactions are conserved in MCF-7, that have two new loops (**Supplementary Fig. S7b)**.

In summary, different long-range interactions involving CTCF binding sites have been identified in the erythroid transcription factor genes. It should be interesting to analyze possible changes in these interactions during erythroid differentiation although these experiments are out of the scope of this work.

4.3.3. Histone marks in CTSs of erythroid genes

We analyzed histone modifications levels in the CTSs of erythroid genes obtained by ChIP-seq from the ENCODE project (**Figure 4.6c, Supplementary figures S1-S7 c**). Three active (H3K4me1, H3K4me3 and H3K9Ac) and one repressive (H3K27me3) histone marks have been reported in K562 cells.

- The CTS in ***GATA2*** is not enriched in any histone modification in K562 cells (**Figure 4.6c**), but pancreatic adenocarcinoma (PANC-1) cell line is enriched in active mark H3K4me3.
- The CTS in ***ETS1*** is not enriched in any histone mark in K562 cells (**Supplementary Fig. S1c**). However, this site is enriched in H3K4me1 in PANC-1 cells and peripheral blood mononuclear cells (PBMC). Lung malignant pluripotent embryonal carcinoma (NT2-D1) cells have small peaks of H3K9Ac and H3K27me3.
- The CTS in ***MYB*** is enriched in active histone modifications (H3K4me3 and H3K9Ac) in K562 cell line (**Supplementary Fig. S2c**). PBMC are also enriched in H3K4me3 and, additionally, they are enriched in H3K4me1. However, NT2-D1 cells are enriched in repressive mark H3K27me3.
- The CTS in ***HEY1*** is enriched in histone modifications related with active genes (H3K4me1 and H3K9Ac) in K562 cells (**Supplementary Fig. S3c**). PBMC are also enriched in H3K4me1, but they are enriched in repressive H3K27me3 too. PANC-1 are enriched in active mark H3K4me3.
- The CTS in ***LMO2*** is not enriched in any histone mark in any cell line (**Supplementary Fig. S4c**).
- The CTS in ***KLF1*** is enriched in the three active histone marks in K562 cells (**Supplementary Fig. S5c**). PBMC are also enriched in H3K4me1 and H3K9Ac, and PANC-1 are enriched in H3K4me3. NT2-D1 are enriched in active marks, but they are enriched in repressive H3K27me3 too.
- The CTS in ***NFE2L2*** is enriched in active histone modifications (H3K4me1 and H3K9Ac) in K562 cells (**Supplementary Fig. S6c**). PBMC and PANC-1 are also enriched in active mark H3K4me1.
- The CTS in ***TCF3*** is enriched in active histone marks (H3K4me1, H3K4me3 and H3K9Ac) in K562 cells (**Supplementary Fig. S7c**). PBMC cells are also enriched in H3K4me1 and PANC-1 cells are enriched in H3K4me3. NT2-D1 cells are also enriched in active marks (H3K4me3 and H3K9Ac), however they are enriched in repressive mark H3K27me3 too.

Once we have made *in silico* predictions of histone marks in the CTCF binding sites to erythroid genes, we would have performed ChIP experiments to analyze changes of histone modifications in those CTSs during erythroid differentiation and upon CTCF downregulation.

4.4. ANALYSIS OF GENETIC ALTERATIONS IN CTCF OR ERYTHROID GENES IN LEUKEMIA PATIENTS

In order to search for possible correlations between CTCF expression or mutations or CTCF binding site mutations and erythroid transcription factors expression or mutations, a combined study of hematological cancers was performed with cBioPortal. AML was chosen among all leukemia types because in this work we were interested in the differentiation of myeloid lineage. A query of genomic data was made for the analysis of ten genes (*CTCF*, *ETS1*, *MYB*, *GATA2*, *HEY1*, *LMO2*, *KLF1*, *NFE2L2*, *TCF3* and *MYC*) in a combined study including 3 studies of AML (**Figure 4.7a**). A total of 1897 samples from 1661 patients were analyzed and profiled for copy number alterations and mutations, among others.

CTCF mutations and deletions mostly result in loss-of-function and predominantly occur in worse prognosis subtypes of cancer, as it has been previously described for endometrial carcinoma (Marshal et al., 2017). We found that the overall survival of AML patients with alterations in CTCF was much lower than the survival of the unaltered group (11 vs. 55 median months overall survival) (**Figure 4.7b**).

Overall, queried genes were altered in 5% of the patients and *GATA2* was the most frequently mutated gene (**Figure 4.7c**). The most frequent alterations in those genes included genetic amplifications, deletions, missense mutations (very common in *MYC* and *GATA2*), and some truncating mutations were also found. We focused on *CTCF*, that was altered in 0.9% of the samples. Truncating mutations in *CTCF* are considered as driver mutations because they give rise to an incomplete protein and could play a role in AML development. Additionally, patients with this type of mutations deceased. On the other hand, deletions of *CTCF* and missense mutations that change one amino acid of CTCF protein were found in patients that survived and in patients that deceased.

Since DNA copy number variations (CNVs) have been strongly associated with cancer, we analyzed the CNVs of the queried genes, and we observed that *CTCF* was amplified in 2% of the samples and deleted in 3% of them. *CTCF* alterations were related with a lower *CTCF* mRNA expression analyzed by RNA-seq in the great majority of the cases (**Figure 4.7d)**.

We next aimed to examine *CTCF* alteration frequency depending on cancer type and we observed that *CTCF* gene was altered in 0.74% of 1088 cases of AML, with half of the alterations being mutations and the other half deep deletions (**Figure 4.7e**). Additionally, a truncating mutation of *CTCF* was observed in 16.67% of 6 cases of generically called myeloproliferative neoplasms (not shown). Further analysis on AML subtypes showed that *CTCF* mutation frequency was higher in AML not otherwise specified (NOS), which includes those AML subtypes not associated with recurrent genetic abnormalities. *CTCF* gene was also mutated in AML with myelodysplasia-related changes, which has a poor prognosis, and in AML with mutated *NPM1*, with favorable outcomes. All *CTCF* deletions and one mutation were detected in AML (subtype not specified). Furthermore, *CTCF* was altered neither in AML with recurrent genetic abnormalities, nor in AML with chromosomal translocations or inversions.

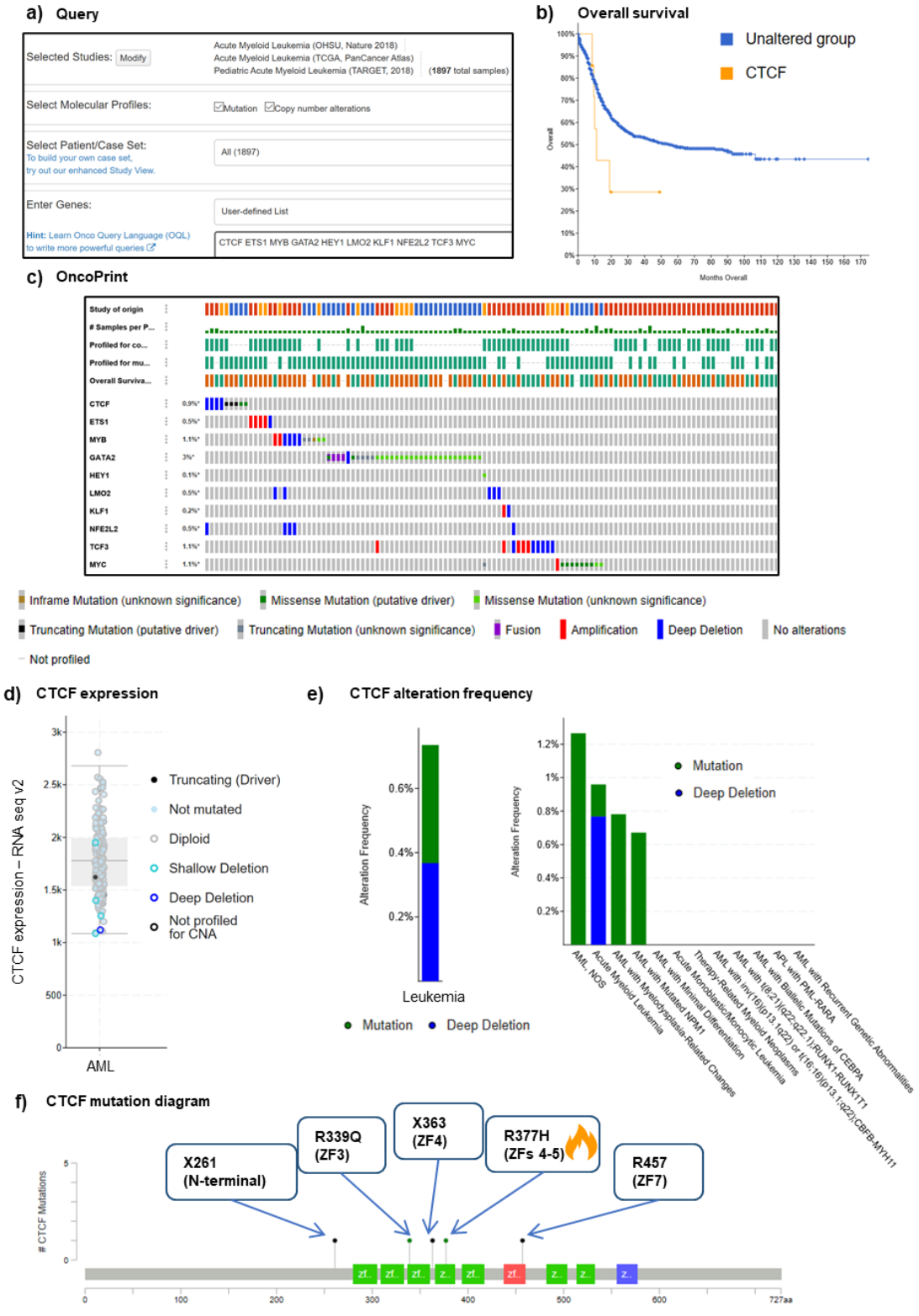


Figure 4.7. Analysis of mutations in hematological cancers with cBioPortal. a) Query of genomic data in cBioPortal for the analysis of ten genes: *CTCF, ETS1, MYB, GATA2, HEY1, LMO2, KLF1, NFE2L2, TCF3* and *MYC* in a combined study of AML. b) Survival analysis of AML cancer samples with (yellow) or without (blue) CTCF genetic alterations. c) The OncoPrint tab shows a visual summary of the genomic alterations in all queried genes across the study. Each row represents a gene and each column represents a sample. d) mRNA expression levels of *CTCF* by RNA-seq in relation to its copy-number status in each sample from AML (TCGA PanCancer Atlas) study. 1661 AML cases are shown. e) CTCF alteration frequency for leukemia (left) and detailed AML subtypes (right) including mutations (green) and deep deletions (blue). f) Position and frequency of CTCF mutations. Flames represent the most recurrent mutations (cancer hotspots).

Because biological processes in cancer are often deregulated through different genes, we aimed to analyze the mutual exclusivity events or, by contrast, co-occurrence in queried genes. Most of the samples showed mutations in a single gene, except for some cases such as co-occurrence of mutations in *MYB* and *NFE2L2*. Although it was not significant, alterations in *CTCF* appeared to be exclusive, and only one sample showed deletion of both *CTCF* and *NFE2L2* genes (**Table 4.1**).

Table 4.1. Mutual exclusivity tab for CTCF alterations.

A	B	Neither	A Not B	B Not A	Both	q-Value	Tendency
CTCF	*NFE2L2*	898	7	2	1	0.295	Co-occurrence
CTCF	*GATA2*	866	8	34	0	0.999	Mutual exclusivity
CTCF	*MYC*	890	8	10	0	0.999	Mutual exclusivity
CTCF	*MYB*	891	8	9	0	0.999	Mutual exclusivity
CTCF	*TCF3*	895	8	5	0	0.999	Mutual exclusivity
CTCF	*ETS1*	896	8	4	0	0.999	Mutual exclusivity
CTCF	*KLF1*	899	8	1	0	0.999	Mutual exclusivity
CTCF	*HEY1*	899	8	1	0	0.999	Mutual exclusivity
CTCF	*LMO2*	899	8	1	0	0.999	Mutual exclusivity

We next examined the position of the mutations and we found 5 mutations in *CTCF*, one of them in the N-terminal domain, three in the zinc fingers and one in a spacer region (**Figure 4.7f**). The mutations in the positions 261 in the N-terminal domain, and 363 in the ZF4, are splice-site mutations that can disrupt RNA splicing resulting in an abnormal CTCF protein. The mutation R339Q in the ZF3 corresponds to a putative driver missense mutation. The mutation R377H is localized in the spacer region between ZFs 4 and 5, also corresponds to a missense mutation and it is a cancer hotspot that has been reported as a recurrent mutation in some cancers. The mutation of the arginine 457 is located in the ZF7 that interacts with 5´ sequence of DNA and it is a truncating mutation. This last mutation was termed as "likely oncogenic" and it has been reported in endometrial cancer (Zighelboim et al., 2014).

Finally, we observed a mutation in the hairy orange motif of *HEY1* (R123Q) that is included in the CTCF binding site to the exon 5 of this gene that has been previously found with the UCSC Genome Browser (**Supplementary Fig. S3d**). We analyzed mutations in other genes, such as *MYB*, *GATA2* or *MYC*, but those regions did not appear to be relevant for CTCF binding, they were neither the CTCF binding sites nor the ends of the chromatin loops.

5. DISCUSSION

Red blood cells are generated through the differentiation and proliferation of hematopoietic stem cells during erythropoiesis. This process is controlled by a combination of transcription factors that regulate erythroid gene expression, and their dysregulation is associated with blood cell disorders. Human leukemia cell lines, such as K562, are commonly used in the study of erythroid differentiation. Multipotent K562 cells, derived from CML, can be induced to differentiate into several hematopoietic pathways (Torrano et al., 2005). More specifically, erythroid differentiation can be induced in K562 cells using cytosine arabinoside (Ara-C) or Imatinib (Delgado et al., 1995; Gómez-Casares et al., 2013). Primary $CD34^{+}$

progenitor cells obtained from cord blood of newborns are used as a more physiological model and can be induced to differentiate into erythroid lineage with erythropoietin (Epo) treatment (Nandakumar et al., 2016). In this work, erythroid differentiation was confirmed with an increase in hemoglobin accumulation, accompanied by a reduction of cell growth and an increase in the expression of erythroid markers (γ-globin and GATA1). The results were consistent with those obtained previously by our group upon the treatment of K562 cells with Ara-C.

CTCF is a zinc finger protein that interacts with many partners to perform multiple functions, including transcriptional activation or repression, chromatin insulation, chromatin looping, long-range interactions, epigenetic regulation. CTCF plays an essential role in hematopoiesis and has been shown to be differentially expressed and phosphorylated depending on the differentiation pathway of human myeloid leukemia cells (Delgado et al., 1999). Our group also described that CTCF overexpression promotes erythroid differentiation of K562 cells (Torrano et al., 2005). CTCF appears to be necessary for erythroid cell differentiation and globin genes expression (Kang et al., 2017; Stadhouders et al., 2012), but molecular mechanisms of this process are still unknown.

To gain further insight into the regulation of erythropoiesis by CTCF, we aimed to downregulate CTCF expression in K562 cells. For that purpose, K562 cells were infected with lentiviral particles containing a constitutive vector with specific shRNA against CTCF and treated with Ara-C or Imatinib. We observed that cell proliferation was reduced upon CTCF downregulation, which is consistent with the requirement of CTCF for cell viability (Bailey et al., 2018). A reduction in hemoglobin producing cells and erythroid markers expression (γ-globin and GATA1) was also observed upon CTCF downregulation and did not increase after the treatment with Ara-C and Imatinib. Hou *et al*. (2010) also reported that CTCF reduction markedly reduces globin gene expression, however it did not affect the protein levels of the important globin gene activator GATA1 (Hou et al., 2010). Additionally, downregulation of CTCF with a doxycycline inducible system inhibits erythroid cell differentiation induced by Epo in primary $CD34^+$ cells (preliminary data obtained by Lorena Garcia). Taken together, these results confirm that the reduction of CTCF inhibits spontaneous and induced erythroid cell differentiation and support that CTCF is required for erythropoiesis, which is consistent with previous results from our group (Torrano et al., 2005).

Erythropoiesis is regulated by a core erythroid network of transcription factors (CEN) and interacting proteins that control erythroid genes expression. Previous results of our group indicated that CTCF overexpression modulates the expression of certain genes encoding erythroid TFs (unpublished data), that were analyzed in this work. We analyzed the mRNA expression of four transcription factors that inhibit erythroid differentiation (ETS1, MYB, GATA2 and HEY1) and four that induce erythroid differentiation (LMO2, KLF1, NFE2L2 and TCF3). We observed that the expression of those genes after the treatment with Ara-C and Imatinib is consistent with their role during erythroid cell differentiation, except for *NFE2L2* and *TCF3*. Similarly, it has been reported that *NFE2L2* was absent in very early stages of induced erythropoiesis and appeared at late stages (Mahajan et al., 2009).

The reduction of CTCF alters the expression of erythroid TFs in K562 cells, which is consistent with CTCF requirement for erythropoiesis. As preliminary results, we observed a reduction in the expression of *MYB, HEY1, LMO2* and *KLF1* upon CTCF downregulation and an increase in the expression

of *NFE2L2*, while the expression of *ETS1*, *GATA2* and *TCF3* hardly changed. The reduction in the expression of *MYB* upon CTCF downregulation is consistent with the fact that *MYB* expression is regulated by CTCF interaction with LDB1 complexes around *MYB* gene (Stadhouders et al., 2012). Additionally, CTCF binding to *LMO2* regulatory regions has been reported (Bhattacharya et al., 2012).

Furthermore, we analyzed the changes in TFs expression after Ara-C or Imatinib treatment of K562 cells with downregulated CTCF and we observed that the results vary slightly depending on the treatment. On one hand, we observed an increase in the expression of the genes that inhibit erythroid differentiation (except *HEY1*) and a reduction in the expression of the genes that induce erythroid differentiation in cells with shCTCF treated with Ara-C, as we expected. On the other hand, we observed an increase in the expression of *ETS1* and a decrease in the expression of *GATA2*, *LMO2* and *KLF1*, while the expression of *MYB*, *HEY1*, *NFE2L2* and *TCF3* has hardly changed in K562 cells with shCTCF treated with Imatinib. Nevertheless, those are preliminary results and those experiments will be repeated. We confirmed by protein expression analysis that LMO2 levels increased during erythroid differentiation and were reduced upon CTCF downregulation. We also confirmed that MYB protein expression levels decreased during erythroid differentiation and were reduced upon CTCF downregulation, as we found by mRNA expression. We observed that HEY1 decreased during erythroid differentiation and increased with downregulated CTCF, which has not hardly changed in terms of mRNA expression.

CTCF and cohesin co-occupy regulatory regions in the genome and mediate long-range chromatin interactions in a cell-type specific manner (Steiner et al., 2016). We used ChIP-seq data from the ENCODE project to identify possible CTCF binding sites (CTSs) to the regulatory regions of the erythroid genes in different hematopoietic and non-hematopoietic human cell lines. All CTSs were supported by cohesin complex subunit Rad21 binding. ChIA-PET data was used to find chromatin interactions around them. Active (H3K4me1, H3K4me3 and H3K9Ac) and repressive (H3K27me3) histone modifications were also analyzed.

The expression of ETS1, MYB, GATA2 and HEY1 must be downregulated to allow erythroid cell differentiation (Perry & Soreq, 2002). We identified a CTS 43 kb upstream *ETS1* gene, which could interact with two CTSs downstream *ETS1* gene. CTCF could function as an insulator between *ETS1* gene and another member of the ETS family, Friend leukemia integration 1 (*FLI1*). FLI1 is involved in erythroid/megakaryocytic differentiation and its expression is indicative of megakaryocyte cell, while blocking the erythroid differentiation (Doré & Crispino, 2011). We identified a CTS in the first intron of *MYB* gene, that has been previously reported (Stadhouders et al., 2012), and interacts with another CTS upstream *MYB*. We identified a CTS 5 kb upstream *GATA2* gene that participates in different interactions with three CTSs located downstream *GATA2* in K562 cells. The biggest loop includes *GATA2*, *DNAJB8* and *EEFSEC*, which could bring together regulatory elements of these genes. We identified a CTS in the exon 5 of *HEY1* that interacts with another CTS upstream *HEY1* gene in K562 cells and this loop includes *MRPS28* gene.

By contrast, LMO2, KLF1, NFE2L2 and TCF3 are required for erythroid differentiation (Love et al., 2014). We identified a CTS 34 kb downstream *LMO2* that interacts with four CTSs upstream *LMO2*. There are different distal regulatory regions that modulate *LMO2* expression and several CTSs upstream

LMO2 have been reported (Bhattacharya et al., 2012). CTCF may function as an insulator between *LMO2* and the neighbouring cell cycle associated protein 1 (*CAPRIN1*). We identified a CTS in the exon 2 of *KLF1* that interacts with another CTS upstream *KLF1*, forming a loop that includes *GCDH* and *SYCE2* genes. KLF1 is necessary for the formation of chromatin loops around the β-globin gene mediated by CTCF (Kang et al., 2017), suggesting that CTCF binding to *KLF1* gene could increase KLF1 expression and, as a result, β-globin gene expression. We identified a CTS in the intron 1 of *NFE2L2* that interacts with two CTSs, one in the exon 1 of *NFE2L2* and another downstream *NFE2L2*. Both interactions appear to be exclusive of K562 cells. We identified a CTS 23 kb upstream *TCF3* that interacts with seven CTSs and is mainly enriched in active histone marks. CTCF could act as an insulator between *TCF3* and *ONECUT3*, *ATP8B3* and *REXO1* genes depending on the cell type.

In summary, we identified shared and non-shared chromatin interactions around erythroid genes in K562 cells and breast adenocarcinoma MCF-7 cell line, which is consistent with the cell-type specificity of certain chromatin interactions mediated by CTCF and cohesin (Steiner et al., 2016). We also found that the patterns of histone modifications in our CTSs were variable between different cell lines. As further experiments, ChIP experiments should be performed to analyze active (H3Ac and H3K4me2) and repressive (H3K9me3 and H3K27me3) histone marks in the CTCF binding sites to erythroid transcription factor genes, as it has been performed in our group for CTCF binding site to *BCL6* (Batlle-López et al., 2015). We would expect that CTCF downregulation reduced CTCF occupancy in all CTSs. Due to the important role of CTCF in erythropoiesis, we would also expect an increase in repressive histone marks and a reduction of active histone marks in the genes that induce erythroid differentiation when we decreased CTCF levels. We would expect the opposite effect in the genes that inhibit erythroid differentiation.

Finally, we analyzed genetic alterations in *CTCF* and erythroid transcription factor genes in AML patients with cBioPortal, that have not already been reported (Debaugny & Skok, 2020; Voutsadakis, 2018). AML starts in the bone marrow form immature cells that develop into abnormal white blood cells. AML is a genetically and molecularly heterogeneous disease and, because the subtype of AML can be important to determine a patient´s prognosis, the World Health Organization classified AML in 2016 into several groups (American Cancer Society, 2018). We observed that *CTCF* mutations and deletions in AML patients decrease *CTCF* mRNA expression. It has been reported that CTCF hemizygous loss in human tumors may be related with different clinical outcomes (Kemp et al., 2015). We observed a reduction in the survival of AML patients with CTCF alterations, which is in concordance with the association between the reduced levels of CTCF and a poor survival of cancer patients reported in other studies (Kemp et al., 2015). CTCF loss of function observed in AML studies, mainly due to truncating mutations, supports its role as oncosuppressor gene during cancer development (Bailey et al., 2018; Filippova et al., 1998). We observed that most of the deletions encompass large portions on the long arm (q) of chromosome 16, not only *CTCF* deletion, as it was observed in endometrial cancer (Walker et al., 2015), and the deletion of these regions could include other tumor suppressor genes. Additionally, *CTCF* mutations rarely co-occur with mutations in the studied erythroid genes in the same patients.

Mutations were observed in the N-terminal domain and ZF domain of CTCF, including the mutation in the position R339 that has been reported in Wilms tumor (Filippova et al., 2002); the mutation

R377H that is a cancer hotspot reported as a recurrent mutation in endometrial (Filippova et al., 2002; Oh et al., 2017), uterus, bowel and skin cancers (Chang et al., 2018); and the mutation in the position R457 that has been reported in endometrial cancer (Zighelboim et al., 2014). The mutation R339Q in the ZF3 of CTCF eliminates the arginine at position 6 of the alpha helix that is contacting guanine, therefore may be critical for DNA recognition. The mutation R377H is localized in the spacer region between ZFs 4 and 5 and does not directly affect the contact with DNA, but it could alter the formation of the ZFs. The mutation in the arginine 457, that does not directly interact with the DNA, is a nonsense mutation, which leads to a premature stop codon and may lead or not to a loss of binding. Cancer-associated CTCF stop codon mutations have been analyzed in many types of cancers, not including AML (Debaugny & Skok, 2020). The mutation in the N-terminal domain could affect the interaction of CTCF with distinct binding partners, and mutations in this domain have been reported in other cancers. In conclusion, the reduction in CTCF levels and function associated with mutations could modulate the expression of many genes with important roles in AML tumorigenesis, in part due to the alterations in chromatin looping mediated by CTCF, as it has been described in other cancers (Debaugny & Skok, 2020).

6. CONCLUSIONS

1. CTCF downregulation with constitutive or inducible systems based on siRNA inhibits erythroid differentiation in K562 cells and in primary $CD34^+$ cells, indicating that CTCF is required for the erythropoiesis.
2. The expression of erythroid transcription factors changed in K562 cells during erythroid differentiation according to their role during erythropoiesis, at both mRNA and protein levels. CTCF downregulation altered their expression after Ara-C and Imatinib treatment.
3. CTCF binding to erythroid transcription factor genes participates in the formation of long-range chromatin interactions that differ between cell lines, suggesting a cell-type specific expression of erythroid transcription factors mediated by CTCF.
4. CTCF expression is reduced in some AML cases and lower CTCF levels have been associated with lower AML survival. Therefore, CTCF could be implicated in the development of AML.

7. FUTURE WORK

In order to mechanistically unravel CTCF function as a transcriptional regulator of erythroid transcription factors, a summary of our future research could be:

1. Further analyze mRNA and protein expression to verify preliminary results about the effects of CTCF downregulation on the expression of erythroid transcription factors during erythroid differentiation.
2. Perform ChIP experiments to analyze CTCF occupancy and histone modifications at CTCF binding sites to the selected erythroid transcription factor genes in K562 cells during erythroid differentiation, as well as in other cell lines.
3. Perform ChIP experiments for histones upon CTCF downregulation to analyze the role of CTCF on local chromatin structure in the CTCF binding sites to erythroid genes.

4. Extend cBioPortal analysis to search for possible alterations in CTCF or erythroid genes in other hematological malignancies, such as T-ALL, myeloproliferative neoplasms, and lymphomas. We could access to more data from other portals, including ICGC Data Portal, COSMIC or Tumorscape.

8. REFERENCES

Ali, H., Iftikhar, F., Shafi, S., Siddiqui, H., Khan, I. A., Choudhary, M. I., & Musharraf, S. G. (2019). Thiourea derivatives induce fetal hemoglobin production in-vitro: A new class of potential therapeutic agents for β-thalassemia. *European Journal of Pharmacology*, *855*, 285–293.

American Cancer Society. (2018). Acute Myeloid Leukemia (AML) Subtypes and Prognostic Factors. Retrieved from https://www.cancer.org/cancer/acute-myeloid-leukemia/detection-diagnosis-staging/how-classified.html

Arzate-Mejía, R. G., Recillas-Targa, F., & Corces, V. G. (2018). Developing in 3D: the role of CTCF in cell differentiation. *Development (Cambridge, England)*, *145*(6).

Bailey, C. G., Metierre, C., Feng, Y., Baidya, K., Filippova, G. N., Loukinov, D. I., … Rasko, J. E. J. (2018). CTCF expression is essential for somatic cell viability and protection against cancer. *International Journal of Molecular Sciences*, *19*(12), 1–20.

Baniahmad, A., Steiner, C., Kohne, A. C., & Renkawitz, R. (1990). Modular Structure of a Chicken Lysozyme Sirwet : lnvohrement of an Unusual Thyroid Hormone Fbceptor Binding Site. *Cell*, *61*, 505–514.

Barminko, J., Reinholt, B., & Baron, M. H. (2015). Development and differentiation of the erythroid lineage in mammals. *Developmental and Comparative Immunology*.

Batlle-López, A., Cortiguera, M. G., Rosa-Garrido, M., Blanco, R., Del Cerro, E., Torrano, V., … Delgado, M. D. (2015). Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus. *Oncogene*, *34*(2), 246–256.

Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, *98*(3), 387–396.

Bhattacharya, A., Chen, C. Y., Ho, S., & Mitchell, J. A. (2012). Upstream Distal Regulatory Elements Contact the Lmo2 Promoter in Mouse Erythroid Cells. *PLoS ONE*, *7*(12).

Braccioli, L., & de Wit, E. (2019). CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays in Biochemistry*, *63*(1), 157–165.

Branco, M. R., & Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biology*, *4*(5), 780–788.

Brandt, S. J., & Koury, M. J. (2009). Editorials and perspectives regulation of lmo2 mrna and protein expression in erythroid differentiation. *Haematologica*, *94*(4), 447–448.

Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., … Renkawitz, R. (1997). Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Molecular and Cellular Biology*, *17*(3), 1281–1288.

Bushey, A. M., Dorman, E. R., & Corces, V. G. (2009). *Chromatin insulators: regulatory mechanisms and epigenetic inheritance*. *32*(404), 1–9.

Cañelles, M., Delgado, M. D., Hyland, K. M., Lerga, A., Richard, C., Dang, C. V., & León, J. (1997). Max and inhibitory c-Myc mutants induce erythroid differentiation and resistance to apoptosis in human myeloid leukemia cells. *Oncogene*, *14*(11), 1315–1327.

Cantor, A. B., & Orkin, S. H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene*, *21*(21), 3368–3376.

Chang, M. T., Bhattarai, T. S., Schram, A. M., Bielski, C. M., Donoghue, M. T. A., Jonsson, P., … Taylor, B. S. (2018). Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discovery*, *8*(2), 174–183.

Chen, C., & Lodish, H. F. (2014). Global analysis of induced transcription factors and cofactors identifies Tfdp2 as an essential coregulator during terminal erythropoiesis. *Experimental Hematology*, *42*(6), 464-476.e5.

Chen, D., & Lei, E. P. (2019). Function and regulation of chromatin insulators in dynamic genome organization. *Current Opinion in Cell Biology*, *58*, 61–68.

Chen, H., Tian, Y., Shu, W., Bo, X., & Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE*, *7*(7).

Chernukhin, I., Shamsuddin, S., Kang, S. Y., Bergstrom, R., Kwon, Y.-W., Yu, W., … Klenova, E. (2007). CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. *Molecular and Cellular Biology*, *27*(5), 1631–1648.

Cho, D. H., Thienes, C. P., Mahoney, S. E., Analau, E., Filippova, G. N., & Tapscott, S. J. (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Molecular Cell*, *20*(3), 483–489.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, *162*(1), 156–159.

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., & Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research*, *19*(1), 24–32.

De La Rosa-Velázquez, I. A., Rincón-Arano, H., Benítez-Bribiesca, L., & Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Research*, *67*(6), 2577–2585.

Debaugny, R. E., & Skok, J. A. (2020). CTCF and CTCFL in cancer. *Current Opinion in Genetics and Development*, *61*, 44–52.

Delgado, M. D., Chernukhin, I. V., Bigas, A., Klenova, E. M., & León, J. (1999). Differential expression and phosphorylation of CTCF, a c-myc transcriptional regulator, during differentiation of human myeloid cells. *FEBS Letters*, *444*(1), 5–10.

Delgado, M. D., & León, J. (2010). Myc roles in hematopoiesis and leukemia. *Genes and Cancer*, *1*(6), 605–616.

Delgado, M. D., Lerga, A., Cañelles, M., Gómez-Casares, M. T., & León, J. (1995). Differential regulation of Max and role of c-Myc during erythroid and myelomonocytic differentiation of K562 cells. *Oncogene*, *10*(8), 1659–1665.

Demirci, S., & Tisdale, J. F. (2018). *Definitive Erythropoiesis from Pluripotent Stem Cells: Recent Advances and Perspectives BT - Cell Biology and Translational Medicine, Volume 3: Stem Cells, Bio-materials and Tissue Engineering* (K. Turksen, Ed.). Cham: Springer International Publishing.

Doré, L. C., & Crispino, J. D. (2011). Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*, *118*(2), 231–239.

Elagib, K. E., Xiao, M., Hussaini, I. M., Delehanty, L. L., Palmer, L. A., Racke, F. K., … Goldfarb, A. N. (2004). Jun blockade of erythropoiesis: role for repression of GATA-1 by HERP2. *Molecular and Cellular Biology*, *24*(17), 7779–7794.

Essien, K., Vigneau, S., Apreleva, S., Singh, L. N., Bartolomei, M. S., & Hannenhalli, S. (2009). CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biology*, *10*(11).

Fang, C., Wang, Z., Han, C., Safgren, S. L., Helmin, K. A., Adelman, E. R., … Zang, C. (2020). Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. *BioRxiv*, 2020.01.17.910687.

Filippova, G., Lindblom, A., Meincke, L. J., Klenova, E. M., Neiman, P. E., Collins, S. J., … Lobanenkov, V. V. (1998). *A Widely Expressed Transcription Factor With Multiple DNA Sequence Specificity,*. *36*, 26–36.

Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., … Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Developmental Cell*, *8*(1), 31–42.

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., … Lobanenkov, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology*, *16*(6), 2802–2813.

Filippova, G. N., Lindblom, A., Meincke, L. J., Klenova, E. M., Neiman, P. E., Collins, S. J., … Lobanenkov, V. V. (1998). A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes, Chromosomes and Cancer*, *22*(1), 26–36.

Filippova, G. N., Ulmer, J. E., Moore, J. M., Ward, M. D., Hu, Y. J., Neiman, P. E., … Dorion-Bonnet, F. (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Research*, *62*(1), 48–52.

Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., … Nakagawa, H. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nature Genetics*, *48*(5), 500–509.

García-Gaipo, L. (Universidad de C. (2019). *Epigenetic mechanisms in erythroid and lymphoid differentiation: CTCF transcriptional regulator and epigenetic drugs*. Universidad de Cantabria.

Garrett-Sinha, L. A. (2013). Review of Ets1 structure, function, and roles in immunity. *Cellular and Molecular Life Sciences*, *70*(18), 3375–3390.

Ghirlando, R., & Felsenfeld, G. (2016). *CTCF : making the right connections*. 881–891.

Gombert, W. M., & Krumm, A. (2009). Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. *PLoS ONE*, *4*(7), 1–8.

Gómez-Casares, M. T., García-Alegria, E., López-Jorge, C. E., Ferrándiz, N., Blanco, R., Alvarez, S., … León, J. (2013). MYC antagonizes the differentiation induced by imatinib in chronic myeloid leukemia cells through downregulation of p27(KIP1.). *Oncogene*, *32*(17), 2239–2246.

Gregor, A., Oti, M., Kouwenhoven, E. N., Hoyer, J., Sticht, H., Ekici, A. B., … Zweier, C. (2013). De novo mutations in the genome organizer CTCF cause intellectual disability. *American Journal of Human Genetics*, *93*(1), 124–131.

Gröschel, S., Sanders, M. A., Hoogenboezem, R., De Wit, E., Bouwman, B. A. M., Erpelinck, C., … Delwel, R. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell*, *157*(2), 369–381.

Guo, J., Li, N., Han, J., Pei, F., Wang, T., Lu, D., & Jiang, J. (2018). DNA recognition patterns of the multi-zinc-finger protein CTCF: a mutagenesis study. *Acta Pharmaceutica Sinica B*, *8*(6), 900–908.

Guo, Y. A., Chang, M. M., Huang, W., Ooi, W. F., Xing, M., Tan, P., & Skanderup, A. J. (2018). Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nature Communications*, *9*(1).

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., … Higgs, D. R. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, *19*(8), 952–961.

Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G., & Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Molecular Cell*, *66*(5), 711-720.e3.

Hattangadi, S. M., Wong, P., Zhang, L., Flygare, J., & Lodish, H. F. (2011). From stem cell to red cell:

Regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*, *118*(24), 6258–6268.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., & Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17507–17512.

Herold, M., Bartkuhn, M., & Renkawitz, R. (2012). CTCF: Insights into insulator function during development. *Development*, *139*(6), 1045–1057.

Hnisz, D., Weintrau, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., … Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, *351*(6280), 1454–1458.

Holwerda, S. J., & de Laat, W. (2013). CTCF: The protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1620).

Hou, C., Dale, R., & Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3651–3656.

Huang, X., Gschweng, E., Van Handel, B., Cheng, D., Mikkola, H. K. A., & Witte, O. N. (2011). Regulated expression of microRNAs-126/126* inhibits erythropoiesis from human embryonic stem cells. *Blood*, *117*(7), 2157–2165.

Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., … Li, C. (2019). Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Research*, *47*(13), 6699–6713.

Ingham, J. (1932). An improved and simplified benzidine test for blood in urine and other clinical material. *The Biochemical Journal*, *26*(4), 1124–1126.

Jacquel, A., Colosetti, P., Grosso, S., Belhacene, N., Puissant, A., Marchetti, S., … Auberger, P. (2007). Apoptosis and erythroid differentiation triggered by Bcr-Abl inhibitors in CML cell lines are fully distinguishable processes that exhibit different sensitivity to caspase inhibition. *Oncogene*, *26*(17), 2445–2458.

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., … Lobanenkov, V. V. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Current Biology*, *10*(14), 853–856.

Kang, Y., Kim, Y. W., Kang, J., Yun, W. J., & Kim, A. R. (2017). Erythroid specific activator GATA-1-dependent interactions between CTCF sites around the β-globin locus. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1860*(4), 416–426.

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., … Aaltonen, L. A. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nature Genetics*, *47*(7), 818–821.

Kemp, C. J., Moore, J. M., Moser, R., Bernard, B., Teater, M., Smith, L. E., … Filippova, G. N. (2015). *CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer*. *7*(4), 1020–1029.

Kim, T. G., Kim, S., Jung, S., Kim, M., Yang, B., Lee, M. G., & Kim, H. P. (2017). CCCTC-binding factor is essential to the maintenance and quiescence of hematopoietic stem cells in mice. *Experimental and Molecular Medicine*, *49*(8), 1–14.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D. D., … Ren, B. (2007). Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*, *128*(6), 1231–1245.

Klenova, E. ., Chernukhin, I. V, El-kady, A., Lee, R. E., Pugacheva, E. M., Loukinov, D. I., … Filippova, G. N. (2001). Functional Phosphorylation Sites in the C-Terminal Region of the Multivalent Multifunctional Transcriptional Factor CTCF. *Molecular and Cellular Biology*, *21 (6)*, 2221–2234.

Klenova, E. M., Fagerlie, S., Filippova, G. N., Kretzner, L., Goodwin, G. H., Loring, G., … Lobanenkov, V. V. (1998). Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *Journal of Biological Chemistry*, *273*(41), 26571–26579.

Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., … Lobanenkov, V. V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Molecular and Cellular Biology*, *13*(12), 7612–7624.

Kumkhaek, C., Aerbajinai, W., Liu, W., Zhu, J., Uchida, N., Kurlander, R., … Rodgers, G. P. (2013). MASL1 induces erythroid differentiation in human erythropoietin-dependent CD34+ cells through the Raf/MEK/ERK pathway. *Blood*, *121*(16), 3216–3227.

Lee, B. K., Bhinge, A. A., Battenhouse, A., McDaniell, R. M., Liu, Z., Song, L., … Iyer, V. R. (2012). Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Research*, *22*(1), 9–24.

Lee, J., Krivega, I., Dale, R. K., & Dean, A. (2017). The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. *Cell Reports*, *19*(12), 2490–2502.

Li, L., Freudenberg, J., Cui, K., Dale, R., Song, S.-H., Dean, A., … Love, P. E. (2013). Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood*, *121*(22), 4575–4585.

Liu, F., Wu, D., & Wang, X. (2019). Roles of CTCF in conformation and functions of chromosome. *Seminars in Cell and Developmental Biology*, *90*, 168–173.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene*, *5*(12), 1743–1753.

Love, P. E., Warzecha, C., & Li, L. Q. (2014). Ldb1 complexes: The new master regulators of erythroid gene transcription. *Trends in Genetics*, *30*(1), 1–9.

Lulli, V., Romania, P., Morsilli, O., Gabbianelli, M., Pagliuca, A., Mazzeo, S., … Marziali, G. (2006). Overexpression of Ets-1 in human hematopoietic progenitor cells blocks erythroid and promotes megakaryocytic differentiation. *Cell Death and Differentiation*, *13*(7), 1064–1074.

Lutz, M. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Research*, *28*(8), 1707–1713.

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M., & Sadowski, P. D. (2009). The CTCF Insulator Protein Is Posttranslationally Modified by SUMO. *Molecular and Cellular Biology*, *29*(3), 714–725.

Mahajan, M. C., Karmakar, S., Newburger, P. E., Krause, D. S., & Weissman, S. M. (2009). Dynamics of alpha-globin locus chromatin structure and gene expression during erythroid differentiation of human CD34(+) cells in culture. *Experimental Hematology*, *37*(10), 1143-1156.e3.

Marshal, A. D., Bailey, C. G., Champ, K., Vellozzi, M., O'Young, P., Metierre, C., … Rasko, J. E. J. (2017). CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic. *Oncogene*, *36*(29), 4100–4110.

Merika, M., & Orkin, S. H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. *Molecular and Cellular Biology*, *15*(5), 2437–2447.

Moore, K. S., & von Lindern, M. (2018). RNA binding proteins and regulation of mRNA translation in erythropoiesis. *Frontiers in Physiology*, *9*, 1–17.

Nakahashi, H., Kwon, K. R. K., Resch, W., Vian, L., Dose, M., Stavreva, D., … Casellas, R. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Reports*, *3*(5), 1678–1689.

Nandakumar, S. K., Ulirsch, J. C., & Sankaran, V. G. (2016). Advances in understanding erythropoiesis: Evolving perspectives. *British Journal of Haematology*, *173*(2), 206–218.

O'Brien, S. G., Guilhot, F., Larson, R. A., Gathmann, I., Baccarani, M., Cervantes, F., … Druker, B. J. (2003). Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *New England Journal of Medicine*, *348*(11), 994–1004.

Oh, S., Oh, C., & Yoo, K. H. (2017). Functional roles of CTCF in breast cancer. *BMB Reports*, *50*(9), 445–453.

Ohlsson, R., Lobanenkov, V., & Klenova, E. (2010). Does CTCF mediate between nuclear organization and gene expression? *BioEssays*, *32*(1), 37–50.

Ohlsson, R., Renkawitz, R., & Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics*, *17*(9), 520–527.

Ong, C. T., & Corces, V. G. (2014). CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*, *15*(4), 234–246.

Orkin, S. H., & Zon, L. I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*, *132*(4), 631–644.

Osada, H., Grutz, G., Axelson, H., Forster, A., & Rabbitts, T. H. (1995). Association of erythroid transcription factors: Complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(21), 9585–9589.

Pater, E. de, & Dzierzak, E. (2015). Stem Cells and Haemopoiesis. *Postgraduate Haematology*, pp. 1–10.

Perry, C., & Soreq, H. (2002). Transcriptional regulation of erythropoiesis fine tuning of combinatorial multi-domain elements. *European Journal of Biochemistry*, *269*(15), 3607–3618.

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, *137*(7), 1194–1211.

Rao, X., Huang, X., Zhou, Z., & Lin, X. (2013). An improvement of the 2ˆ(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, *3*(3), 71–85.

Reiter, A., Hochhaus, A., Berger, U., Kuhn, C., & Hehlmann, R. (2001). AraC-based pharmacotherapy of chronic myeloid leukaemia. *Expert Opinion on Pharmacotherapy*, *2*(7), 1129–1135.

Rosa-Garrido, M., Ceballos, L., Alonso-Lecue, P., Abraira, C., Delgado, M. D., & Gandarillas, A. (2012). A cell cycle role for the epigenetic factor CTCF-L/BORIS. *PLoS ONE*, *7*(6).

Roskoski, R. (2003). STI-571: An anticancer protein-tyrosine kinase inhibitor. *Biochemical and Biophysical Research Communications*, *309*(4), 709–717.

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E. P., … Reinberg, D. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Molecular Cell*, *76*(3), 412-422.e5.

Sankaran, V. G., & Orkin, S. H. (2013). The switch from fetal to adult hemoglobin. *Cold Spring Harbor Perspectives in Medicine*, *3*(1), a011643–a011643.

Siatecka, M., & Bieker, J. J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*, *118*(8), 2044–2054.

Soto-Reyes, E., & Recillas-Targa, F. (2010). Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene*, *29*(15), 2217–2227.

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R. J., Bryne, J. C., Van Den Heuvel, A., … Soler, E. (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO Journal*, *31*(4), 986–999.

Steiner, L. A., Schulz, V., Makismova, Y., Lezon-Geyda, K., & Gallagher, P. G. (2016). CTCF and cohesinSA-1 mark active promoters and boundaries of repressive chromatin domains in primary human erythroid cells. *PLoS ONE*, *11*(5), 1–15.

Sun, L., Huang, L., Nguyen, P., Bisht, K. S., Bar-Sela, G., Ho, A. S., … Gius, D. (2008). DNA methyltransferase 1 and 3B activate BAG-1 expression via recruitment of CTCFL/BORIS and modulation of promoter histone methylation. *Cancer Research*, *68*(8), 2726–2735.

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., León, J., Klenova, E., & Delgado, M. D. (2005).

CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *Journal of Biological Chemistry*, *280*(30), 28152–28161.

Torrano, V., Navascués, J., Docquier, F., Zhang, R., Burke, L. J., Chernukhin, I., … Delgado, M. D. (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *Journal of Cell Science*, *119*(9), 1746–1759.

Tsai, S.-F., Martin, D. I. K., Zon, L. I., D'Andrea, A. D., Wong, G. G., & Orkin, S. H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, *339*(6224), 446–451.

Tumburu, L., & Thein, S. L. (2017). Genetic control of erythropoiesis. *Current Opinion in Hematology*, *24*(3).

Valent, P., Büsche, G., Theurl, I., Uras, I. Z., Germing, U., Stauder, R., … Hermine, O. (2018). Normal and pathological erythropoiesis in adults: From gene regulation to targeted treatment concepts. *Haematologica*, *103*(10), 1593–1603.

Van De Nobelen, S., Rosa-Garrido, M., Leers, J., Heath, H., Soochit, W., Joosen, L., … Sleutels, F. (2010). CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics and Chromatin*, *3*(1), 1–21.

Viola, G., Vedaldi, D., Dall'Acqua, F., Fortunato, E., Basso, G., Bianchi, N., … Gambari, R. (2008). Induction of γ-globin mRNA, erythroid differentiation and apoptosis in UVA-irradiated human erythroid cells in the presence of furocumarin derivatives. *Biochemical Pharmacology*, *75*(4), 810–825.

Vostrov, A. A., & Quitschke, W. W. (1997). The zinc finger protein CTCF binds to the APBβ domain of the amyloid β-protein precursor promoter: Evidence for a role in transcriptional activation. *Journal of Biological Chemistry*, *272*(52), 33353–33359.

Voutsadakis, I. A. (2018). Molecular lesions of insulator CTCF and its paralogue CTCFL (BORIS) in cancer: An analysis from published genomic studies. *High-Throughput*, *7*(4).

Walker, C. J., Miranda, M. A., O'Hern, M. J., McElroy, J. P., Coombes, K. R., Bundschuh, R., … Goodfellow, P. J. (2015). Patterns of CTCF and ZFHX3 mutation and associated outcomes in endometrial cancer. *Journal of the National Cancer Institute*, *107*(11), 1–8.

Wallace, J. A., & Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Current Opinion in Genetics and Development*, *17*(5), 400–407.

Wang, D. C., Wang, W., Zhang, L., & Wang, X. (2019). A tour of 3D genome with a focus on CTCF. *Seminars in Cell and Developmental Biology*, *90*, 4–11.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., … Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, *22*(9), 1680–1688.

West, A. G., Huang, S., Gaszner, M., Litt, M. D., & Felsenfeld, G. (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Molecular Cell*, *16*(3), 453–463.

Woessmann, W., Zwanzger, D., & Borkhardt, A. (2004). ERK signaling pathway is differentially involved in erythroid differentiation of K562 cells depending on time and the inducing agent. *Cell Biology International*, *28*(5), 403–410.

Xu, D., Ma, R., Zhang, J., Liu, Z., Wu, B., Peng, J., … Ruan, K. (2018). Dynamic Nature of CTCF Tandem 11 Zinc Fingers in Multivalent Recognition of DNA As Revealed by NMR Spectroscopy. *The Journal of Physical Chemistry Letters*, *9*(14), 4020–4028.

Yamada, H., Horiguchi-Yamada, J., Nagai, M., Takahara, S., Sekikawa, T., Kawano, T., … Iwase, S. (1998). Biological effects of a relatively low concentration of 1-beta-D-arabinofuranosylcytosine in K562 cells: alterations of the cell cycle, erythroid-differentiation, and apoptosis. *Molecular and Cellular Biochemistry*, *187*(1–2), 211–220.

Yamamoto, T., & Saitoh, N. (2019). Non-coding RNAs and chromatin domains. *Current Opinion in Cell Biology*, *58*, 26–33.

Yang, C. T., Ma, R., Axton, R. A., Jackson, M., Taylor, A. H., Fidanza, A., … Forrester, L. M. (2017). Activation of KLF1 Enhances the Differentiation and Maturation of Red Blood Cells from Human Pluripotent Stem Cells. *Stem Cells*, *35*(4), 886–897.

Yang, Y., Liu, X., Xiao, F., Xue, S., Xu, Q., Yin, Y., … Wang, L. (2015). Spred2 modulates the erythroid differentiation induced by imatinib in chronic myeloid leukemia cells. *PLoS ONE*, *10*(2), 1–12.

Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., & Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Research*, *27*(11), 1365–1377.

Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., … Ogawa, S. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nature Genetics*, *45*(11), 1293–1301.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., … Ohlsson, R. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nature Genetics*, *36*(10), 1105–1110.

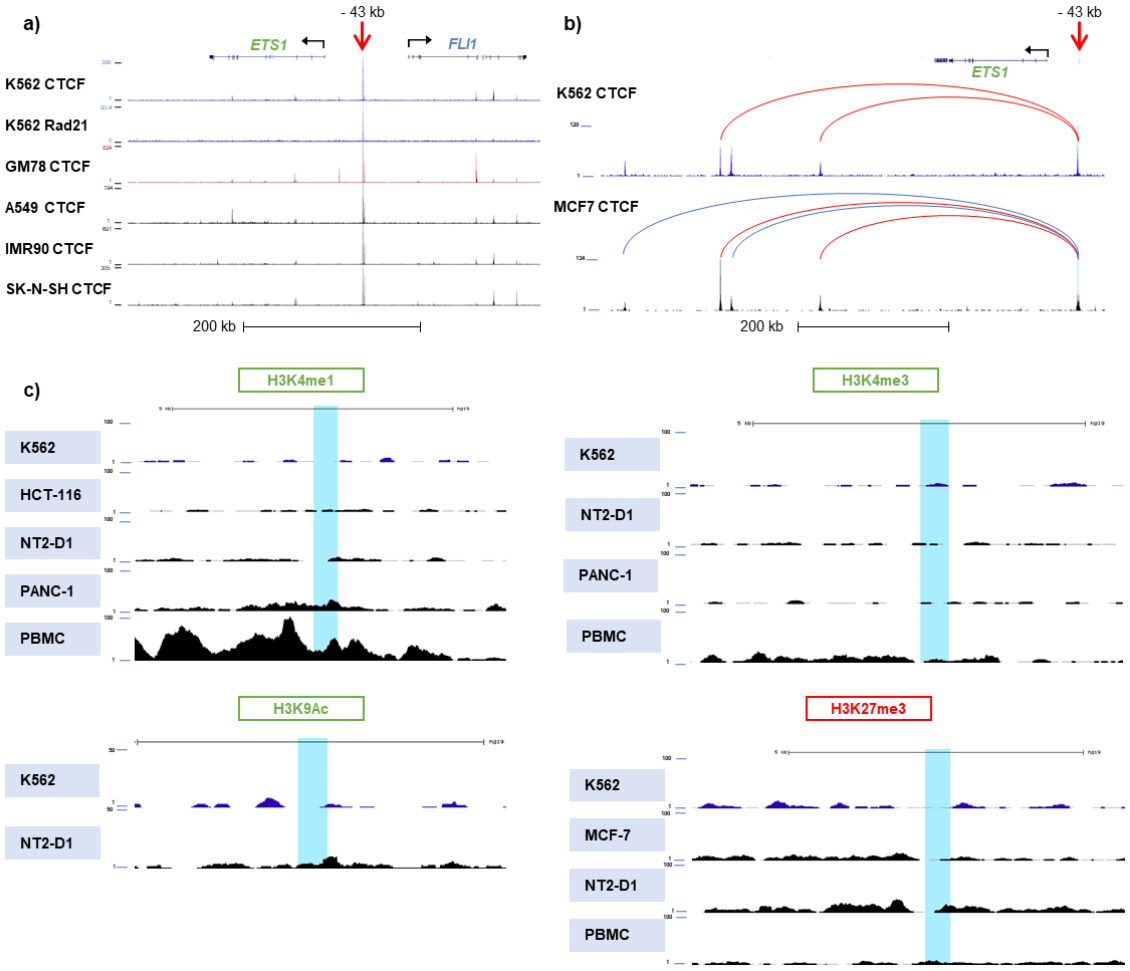
Yusufzai, T. M., Tagami, H., Nakatani, Y., & Felsenfeld, G. (2004). CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. *Molecular Cell*, *13*(2), 291–298.

Zhang, R., Burke, L. J., Rasko, J. E. J., Lobanenkov, V., & Renkawitz, R. (2004). Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Experimental Cell Research*, *294*(1), 86–93.

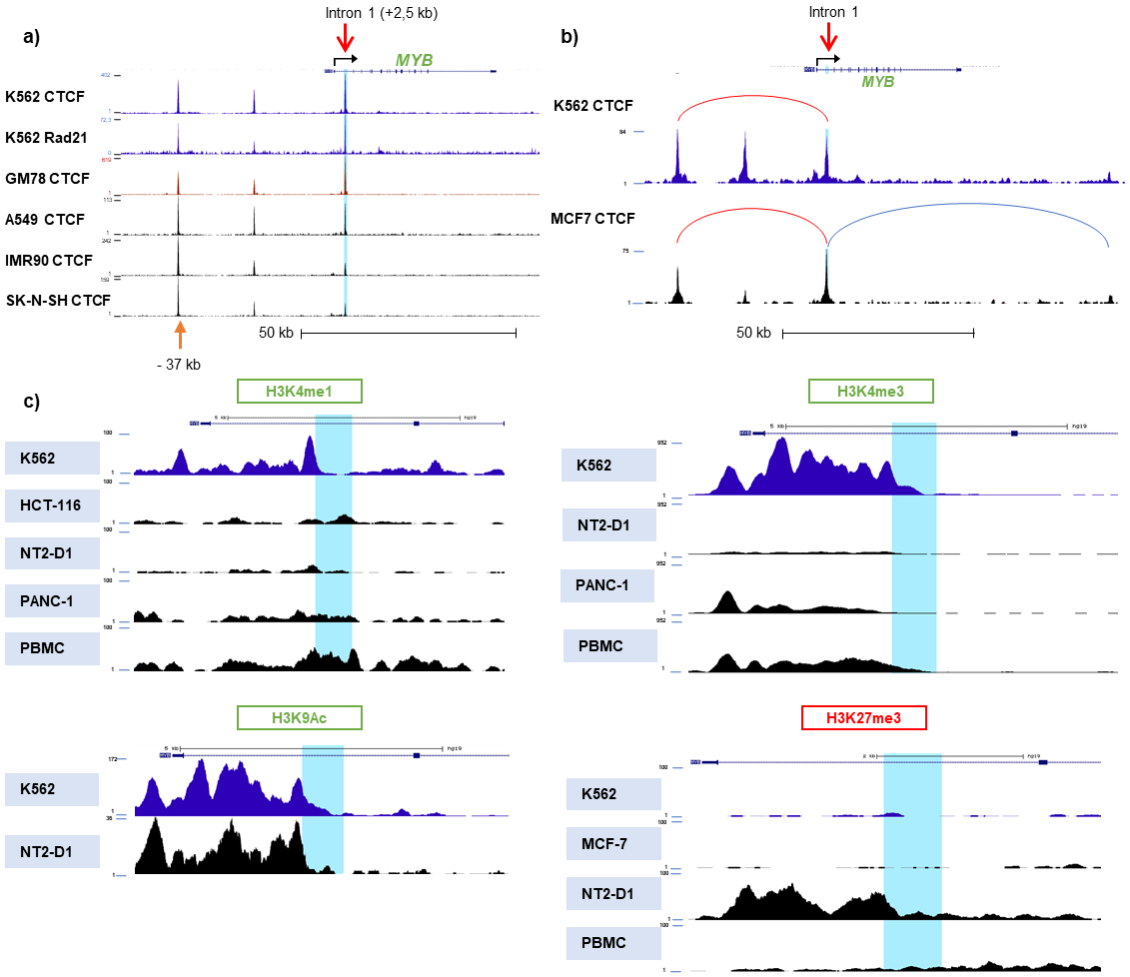
Zighelboim, I., Mutch, D. G., Knapp, A., Ding, L., Xie, M., Cohn, D. E., & Goodfellow, P. J. (2014). High frequency strand slippage mutations in CTCF in MSI-positive endometrial cancers. *Human Mutation*, *35*(1), 63–65.

Zlatanova, J., & Caiafa, P. (2009). CTCF and its protein partners: Divide and rule? *Journal of Cell Science*, *122*(9), 1275–1284.

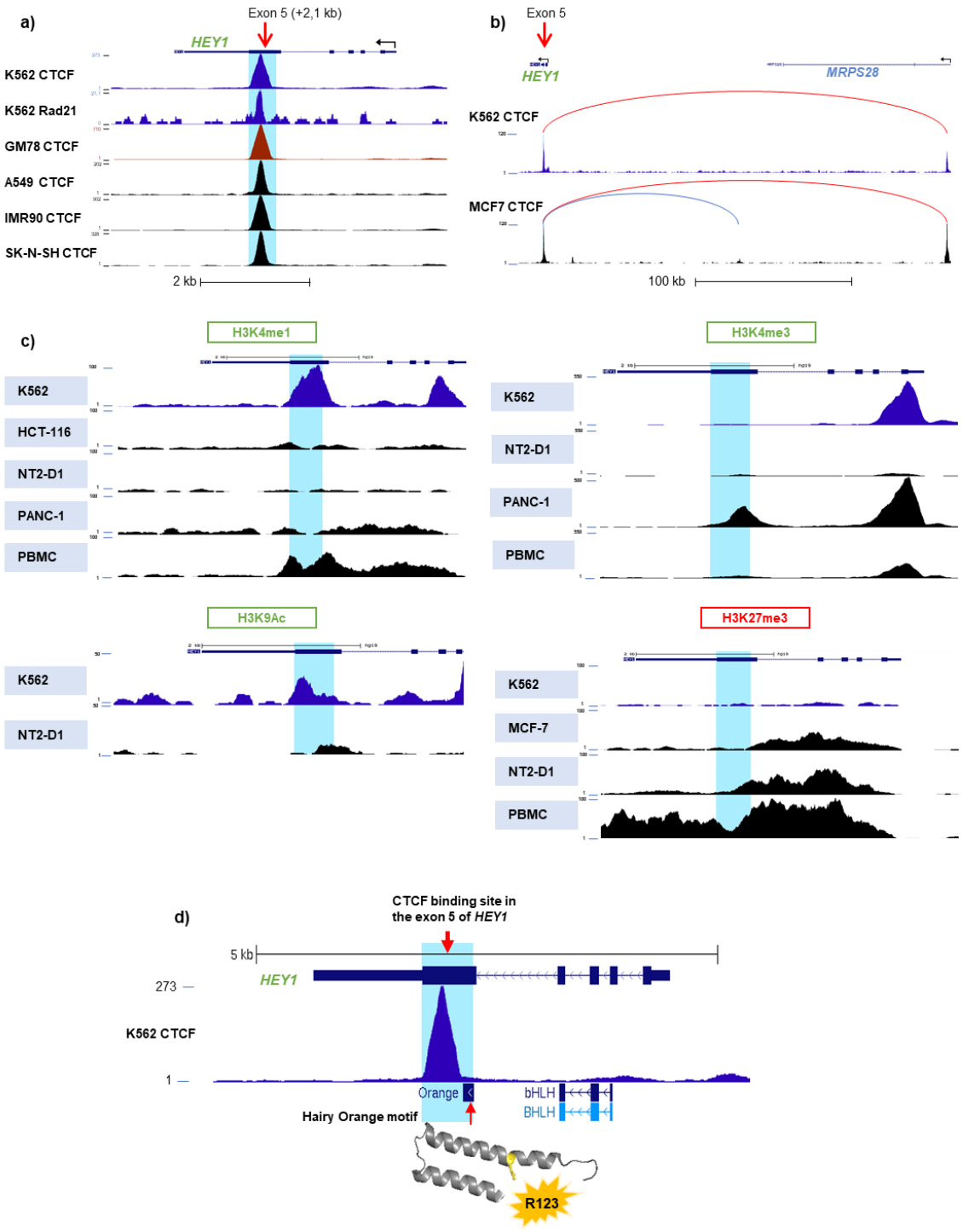
9. APPENDIX



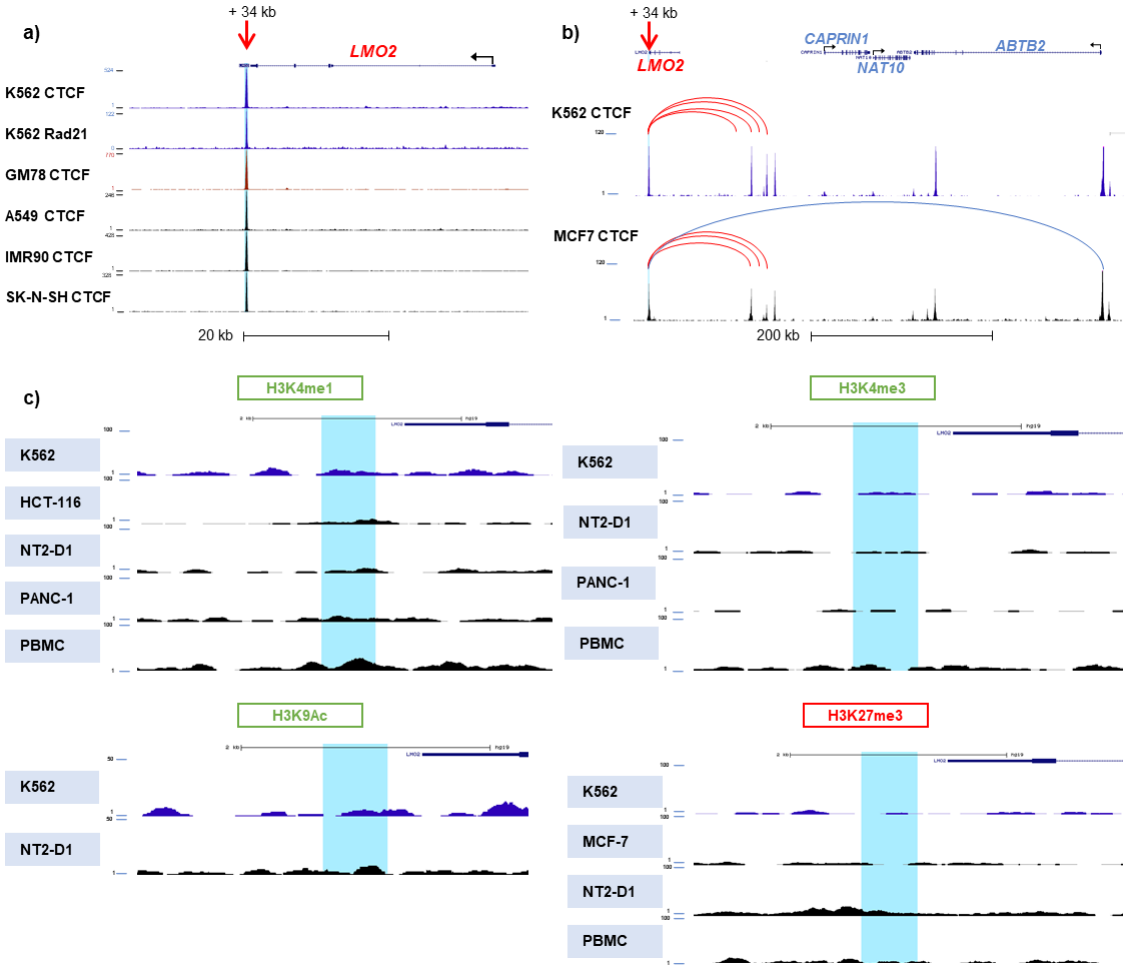
Supplementary Figure S1. Analysis of CTCF binding to *ETS1* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *ETS1* gene in chromosome 11 from 5 human cell lines. CTCF binding site 43 kb upstream *ETS1* gene is shown with a red arrow. b) ChIA-PET interactions from the *ETS1* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *ETS1* CTS (highlighted in blue) in different cell lines.



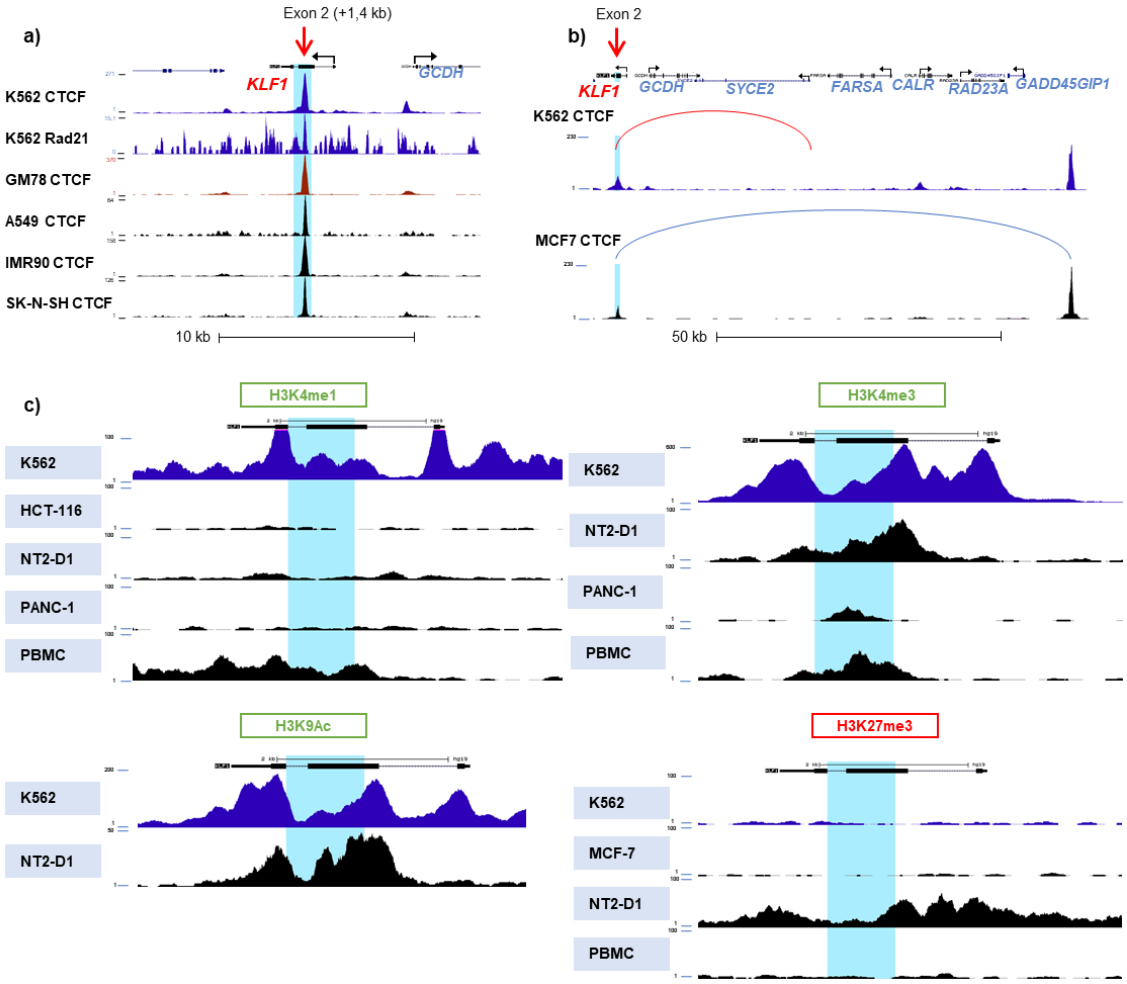
Supplementary Figure S2. Analysis of CTCF binding to *MYB* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *MYB* gene in chromosome 6 from 5 human cell lines. CTCF binding site in the intron 1 of *MYB* gene is shown with a red arrow and CTCF binding site 37 kb upstream *MYB* gene is shown with an orange arrow. b) ChIA-PET interactions from the *MYB* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *MYB* CTS (highlighted in blue) in different cell lines.



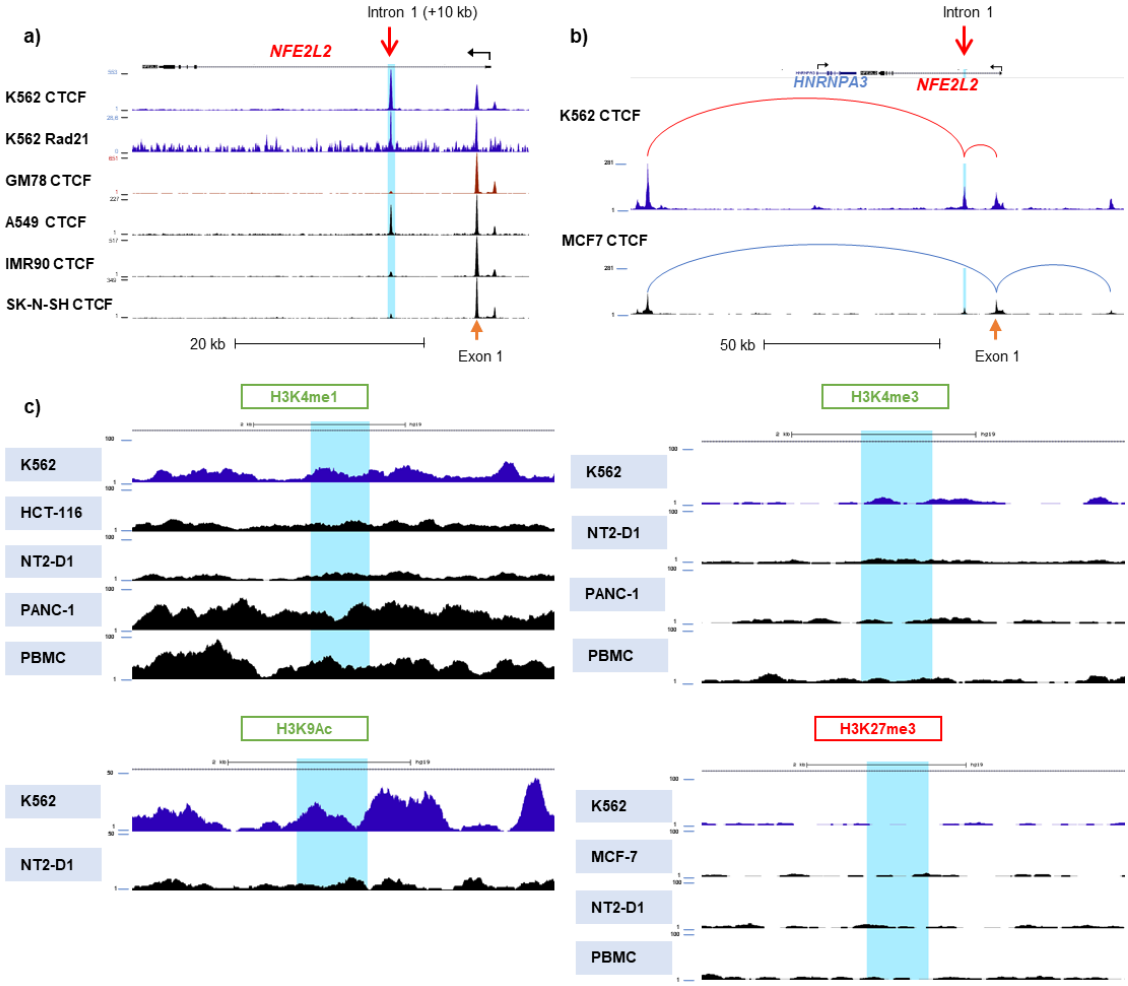
Supplementary Figure S3. Analysis of CTCF binding to *HEY1* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *HEY1* gene in chromosome 8 from 5 human cell lines. CTCF binding site in the exon 5 of *HEY1* gene is shown with a red arrow. b) ChIA-PET interactions from the *HEY1* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *HEY1* CTS (highlighted in blue) in different cell lines. d) ENCODE analysis showing CTCF binding site in the exon 5 of *HEY1* in K562 cells, involving Hairy orange motif of *HEY1*, which is mutated in the arginine 123 in AML studies from cBioPortal.



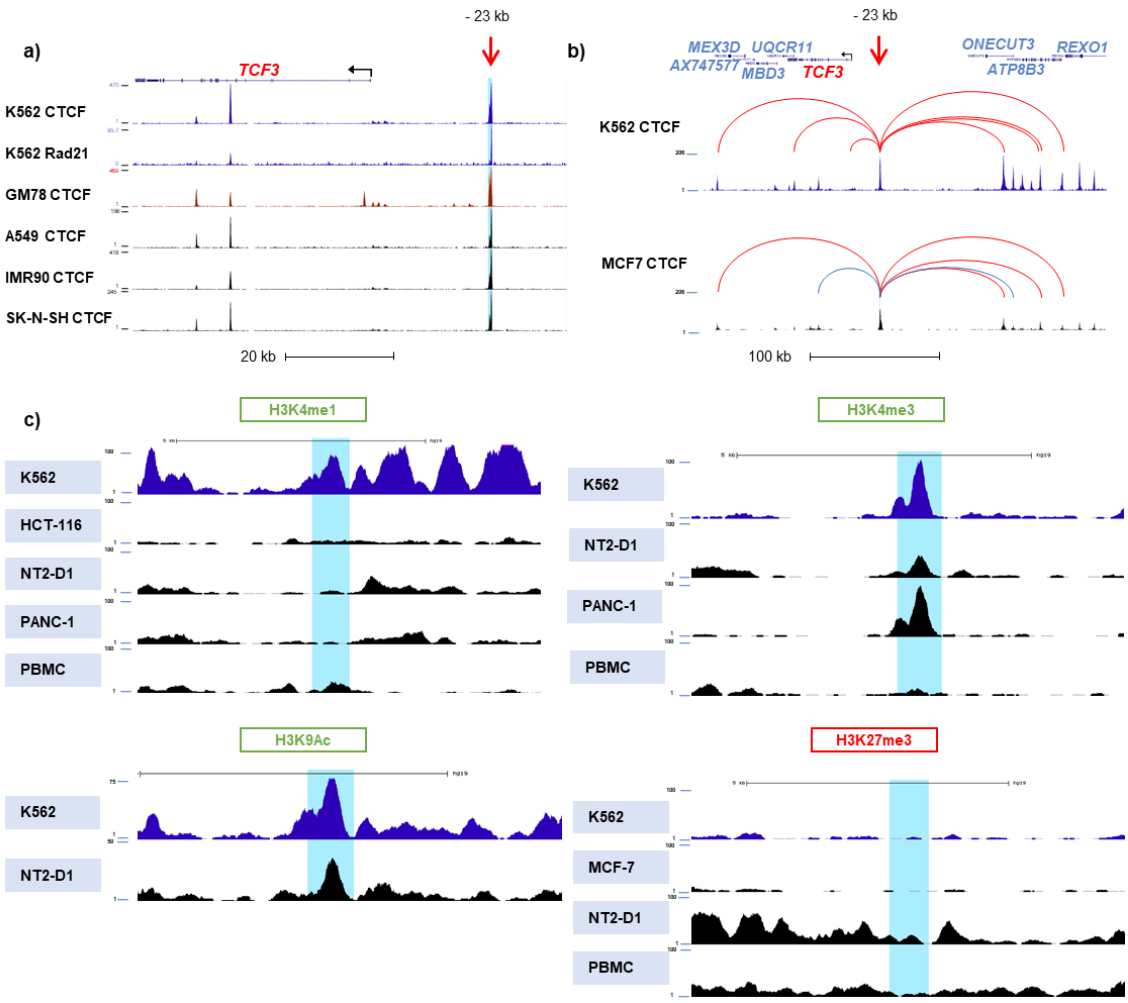
Supplementary Figure S4. Analysis of CTCF binding to *LMO2* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *LMO2* gene in chromosome 11 from 5 human cell lines. CTCF binding site 34 kb downstream *LMO2* gene is shown with a red arrow. b) ChIA-PET interactions from the *LMO2* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *LMO2* CTS (highlighted in blue) in different cell lines.



Supplementary Figure S5. Analysis of CTCF binding to *KLF1* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *KLF1* gene in chromosome 19 from 5 human cell lines. CTCF binding site in the exon 2 of *KLF1* gene is shown with a red arrow. b) ChIA-PET interactions from the *KLF1* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *KLF1* CTS (highlighted in blue) in different cell lines.



Supplementary Figure S6. Analysis of CTCF binding to *NFE2L2* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *NFE2L2* gene in chromosome 2 from 5 human cell lines. CTCF binding site in the intron 1 of *NFE2L2* gene is shown with a red arrow and CTCF binding site in the exon 1 of *NFE2L2* is shown with an orange arrow. b) ChIA-PET interactions from the *NFE2L2* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *NFE2L2* CTS (highlighted in blue) in different cell lines.



Supplementary Figure S7. Analysis of CTCF binding to *TCF3* gene with UCSC Genome Browser. a) ENCODE analysis showing ChIP-seq profile of CTCF and Rad21 for *TCF3* gene in chromosome 19 from 5 human cell lines. CTCF binding site 23 kb upstream *TCF3* gene is shown with a red arrow. b) ChIA-PET interactions from the *TCF3* CTS that are conserved are shown as red arcs and interactions that are not conserved are shown as blue arcs. c) Histone modifications by ChIP-seq from ENCODE in the *TCF3* CTS (highlighted in blue) in different cell lines.