Máster en Biología Molecular y Biomedicina

Universidad de Cantabria

2016-17

Effects of the CTCF transcriptional factor in the erythroid cell differentiation and regulation of erythroid genes

Alexandra Wiesinger

Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), University of Cantabria

Supervisor: Prof. M. Dolores Delgado

ABSTRACT

CCCTC-binding factor (CTCF) is a highly conserved zinc finger protein, which was first identified as a transcription factor regulating the c-*MYC* gene. Further studies revealed several new functions of CTCF including regulation of transcription, epigenetics as well as genome architecture. Due to its multivalent functions, it was reported by our group that CTCF also plays a pivotal role during differentiation of human myeloid leukemia cells. In this study, the effects of CTCF on human erythroid differentiation were investigated and CTCF target genes involved in erythroid differentiation were identified and analyzed. As a model system, K562, a pluripotent human leukemia cell line, was utilized. These cells can be induced chemically to differentiation by cytosine arabinoside (Ara-C) and Imatinib. Proliferation, differentiation and apoptosis analysis were performed upon induction. In order to investigate the role of CTCF in differentiation, CTCF was silenced using shRNA. K562 cells were infected with lentiviral particles containing shCTCF and the effect on erythroid differentiation was evaluated. Expression of erythroid markers was analyzed by western blot and RT-qPCR. Furthermore, the binding of CTCF to the regulatory regions of its putative target genes and changes upon differentiation were studied using ChIP (Chromatin Immunoprecipitation). It could be shown that CTCF knock-down inhibited erythroid differentiation of K562 cells. We aimed to extend these studies to primary hematopoietic precursors $CD34^+$ cells purified from cord blood. Erythroid differentiation was induced with erythropoietin. Our data indicate a role of CTCF in the erythroid differentiation of hematopoietic cells.

Abbreviations

Ara-C	cytosine arabinoside
BasoE	basophilic erythroblast
BFU-E	burst-forming unit erythroid cells
bp	base pair
BSA	Bovine Serum Albumin
cDNA	complementary DNA
CEN	core erythroid network
CFU-E	colony forming unit erythroid cells
ChIP	chromatin immunoprecipitation
Ct	cycle threshold
CTCF	CCCTC-binding factor
CTRL -	negative control
CTRL +	positive control
CTS	CTCF binding site
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetracetic acid
EP	erythroid progenitor
EPO	erythropoietin
ETS1	Erythroblastosis oncogene 1
EV	empty vector
FBS	fetal bovine serum
GPA	Glycophorin A
h	hour
HSC	Hematopoietic stem cells
ICR	imprinting control region
Inr	insulator element
IP	immunoprecipitation
KLF1	Erythroid Kruppel-like factor 1
LMO2	LIM domain-only protein 2
MEP	erythroid- and megakaryocyte progenitors
min	minutes
MOI	multiplicity of infection
mRNA	messenger RNA
MYB	Myeloblastosis oncogene
NuRD	nucleosome remodelling deacetylase
O/N	over night
OrthoE	orthochromatic erythroblast
PAGE	Polyacrilamide Gel Electrophoresis
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate Buffer Saline
PCR	Polyacrilamide Gel Electrophoresis

PolyE	polychromatophilic erythroblast
ProE	proerythroblasts
qPCR	quantitative PCR
RBS	red blood cells
Retic	reticulocytes
RNA	Ribonucleic acid
RNA Pol II	RNA Polymerase II
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute medium
RT	room temperature
RT-qPCR	reverse transcriptase qPCR
s.e.m	standard error of the mean
SCF	stem cell factor
SDS	Sodium Dodecyl Sulfate
SDV	standard deviation
s	seconds
shRNA	short hairpin RNA
TAD	topologically associating domains
TBS-T	Tris-buffer saline-Tween 20
TF	transcription factor
ZF	zinc finger

CONTENTS

1. INTRODUCTION

1.1 CTCF, a multivalent factor

The factor was firstly identified as a protein interacting with three repeats of the CCCTC motif within the chicken c-myc promoter and was therefore named CTCF (CCTC-binding factor). CTCF is highly conserved from Drosophila to humans. The CTCF gene encodes an 11 zinc-finger (ZF) transcription factor, which is ubiquitously expressed and located in the nucleus. There, CTCF tethers diverse and uncommonly long DNA sequences (~50 bp) using different combinations of zinc fingers. Due to its ability of recognizing multiple target sites, CTCF is called a multivalent transcriptional factor mediating many diverse functions in gene regulation (Filippova et al., 1996; Ohlsson et al., 2010)

1.1.1 CTCF structure and functional domains

The CTCF gene encodes a protein of 727 amino acids, which can be separated in three domains: The N-terminal domain, spanning from amino acid 1-265, the central domain containing the 11 zinc finger motifs (265-580) and the C-terminal domain from amino acid 580 to 727 (figure 1, upper). The human CTCF gene consists of 10 exons (NCBI;Gene ID: 10664), from which exon E2 to E9 contain the 11 zinc finger motifs (figure 1, lower). Furthermore, it could be revealed that 93% of the amino acids are identical comparing avian and mammalian CTCF. However, the 11 zinc fingers were demonstrated to be completely identical (Filippova et al., 1996). From ZF1 to ZF10, each zinc finger has ~30 residues harboring 2 cysteines, as well as 2 histidine, separated exactly by 12 amino acids. Together with a zinc ion, those residues are able to tightly interact with the major groove of the DNA.

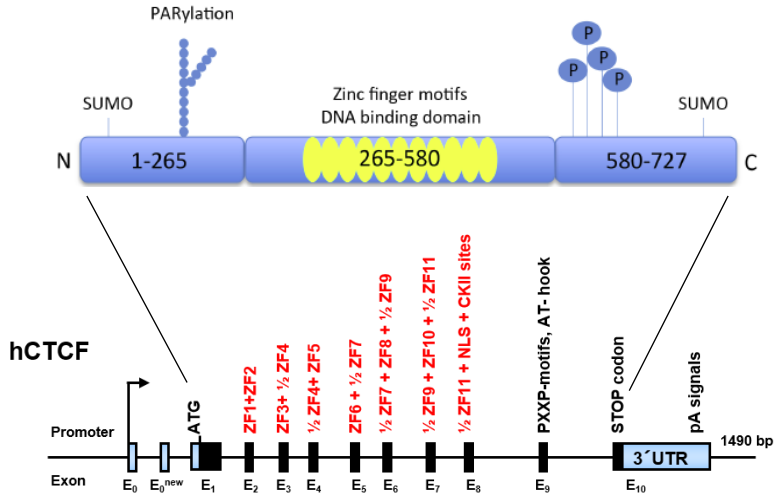

Figure 1. human CTCF gene and CTCF protein scheme. The upper represents CTCF protein. Zinc fingers are represented in yellow from amino acid 265 to 580. Posttranslational modifications are depicted including SUMOylation (SUMO), poly(ADP-ribosyl)ation (PARylation) and phosphorylation (P).The lower represents the human CTCF gene includiing the localization of each zinc finger.
(figure adapted from Ohlsson et al., 2010 and Klenova et al., 2002)

1.1.2 CTCF binding to DNA

Mentioned above, CTCF uses different combinations of zinc fingers to specifically bind DNA. Due to its multivalent nature recognizing remarkable varying sequences, no definite single consensus sequence for CTCF binding site (CTS) has been described. However, Kim et al. propose a 20 bp binding motif, which is present in over 75% of the experimentally identified CTCF-binding sites. A quite significant number of CTSs, however, do not correspond with this motif (Kim et al., 2007).

Since CTCF is able to tether a great number of highly variable sequences, the question of how this specific interaction is accomplished, remains. Systematic mutations of all eleven zinc fingers unveil that zinc finger 4 to 7 adhere CTCF to 80% of target sites incorporating the core motif. The zinc fingers surrounding this core seem to secure the DNA bond. Furthermore, the residence time varies corresponding to the combination of zinc fingers.(Nakahashi et al., 2013). Nakahashi et al., 2013, therefore, suggest that CTCF utilizes combinatorial clustering of its 11 ZFs to accomplish anchoring to diverse DNA sequences.

Within the mammalian genome, between 55,000 and 65,000 CTCF binding sites (CTS) were identified. These sites were further analyzed and grouped in cell type specific CTS, common CTS and ubiquitous CTS. In K562, for instance, only 6% are specific, 66% are classified as common and 28% are found in all cell lines analyzed. Additionally, the location of the CTSs was examined using the ENCODE annotation Genome Browser, which was published by the University of California, Santa Cruz (UCSC). They found that more than half of all binding sites were located within intergenic regions (53%), but only around 12 % were positioned in the proximal promoter. Interestingly, a considerable number of binding sites was situated intragenic, 30% in the introns and 5% in the exons (Chen et al., 2012). This ability of binding many different sites within the genome let assume that CTCF mediates many diverse functions in gene regulation.

1.1.3 Functions of CTCF

Initially, CTCF was identified as a regulator of transcription repressing the Myc gene (Filippova et al., 1996). Consecutive research revealed that CTCF plays a pivotal role in very diverse functions and is not limited to positive and negative regulation of transcription. These functions include insulator activities, epigenetic regulations and regulation of the genomic architecture. Recent studies, furthermore, found that CTCF is also involved in the regulation of noncoding transcription in mammalian genomes.

Transcriptional repressor and activator

Originally, CTCF was classified as a transcription factor due to its regulative role of the c-Myc gene and the chicken lysozyme gene. Later, the role of CTCF as a transcriptional activator of the APP gene was described. Furthermore, several genes involved in cell cycle progression, like the human Rb gene (De La Rosa-Velazquez et al., 2007), the mouse p19ARF, the human p14ARF (Filippova et al., 2002) and the human and mouse PIM-1 and Polo-like kinase were found to be positively regulated by CTCF (Ohlsson et al., 2010).

Insulator functions

CTCF has been revealed to provide transcription regulation properties, but it was also depicted having a binding site within the insulator sequence of the chicken β-globin gene suggesting that CTCF is involved in directional enhancer blocking. Chromatin insulators do not directly act on gene expression, however inhibit the interaction of enhancers/silencers with promotors. Together with cohesin, CTCF manages to form so called chromatin loops that allow close proximity between enhancer and promoter and therefore, activation of transcription. Contrarily, CTCF and cohesin can also form loops in order to exclude the enhancer from making contacts with the promoter and, as a result, repress transcription (Kim et al., 2015). All known vertebrate chromatin insulators involve CTCF binding (Bell et al., 1999). The CTCF mediated

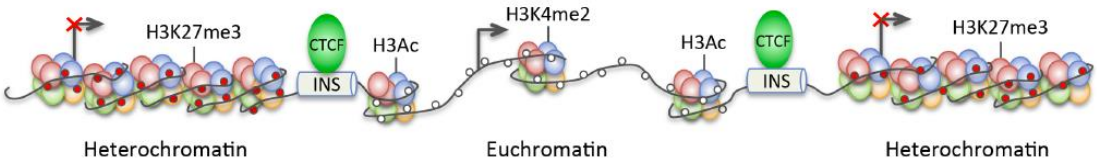

Figure 2. CTCF barrier function. Representative histone marks of either heterochromatin (H3K27me3) or euchromatin (H3Ac, H3K4me2) states are depicted. Red dots represent DNA methylation. CTCF binds to insulator (INS) sequences avoid spreading of heterochromatin. (adapted from doctoral thesis of Maria Cortiguera)

insulator function within the HS4 locus, however, can be functionally distinguished from the insulator's barrier action (figure 2). The barrier function is associated with maintenance of a high level of histone acetylation near the insulator and therefore avoids the spread of heterochromatin.

Additionally, CTCF was also found to interact with a differentially methylated region upstream of the H19 gene, where it also acts as an insulator resulting in the silencing of Igf2 expression from the maternal allele (Bell and Felsenfeld, 2000). It is also shown that CpG methylation abrogates CTCF binding and allows Igf2 expression from the paternal allele. In this case, methylation of CpGs inhibits CTCF binding, nevertheless, *in vitro* studies showed that CTCF binding is not exclusive for unmethylated areas. To sum up, CTCF has a dual insulator function that is on the one hand able to block regulatory elements, but on the other hand also capable of inhibiting DNA methylation spreading (Kim et al., 2015; Ong and Corces, 2014).

Interestingly, these insulator functions are not restricted to intrachromosomal sites. Interaction between two transcriptionally active genes on two different chromosomes (Igf2/H19 and Wsb1/Nf1) was mediated by CTCF meaning that its insulator function may be also a mechanism of bringing regulatory elements in close proximity with genes within the whole nucleus. It is assumed that homodimerization of CTCF accomplishes those long-distance interactions, however, in order to from homodimers, CTCF must be bound to specific DNA sequences (Kim et al., 2015; Ong and Corces, 2014).

Epigenetic regulation

Since CTCF seems to be methylation sensitive, as in the case of the Igf2/H19 locus, the protein is suggested to play a critical role in genomic imprinting. CTCF, furthermore, showed the ability to prevent methylation when bound to DNA (Engel et al., 2006) indicating that CTCF is reading and maintaining epigenetic marks at this locus.

CTCF regulates genes involved in cancer such as p53, p16 or Rb through epigenetic mechanisms. Our group has previously described the epigenetic regulation of ribosomal genes (van de Nobelen et al., 2010) and BCL6 oncogene (Batlle-Lopez et al., 2015). In these cases, CTCF is able to modulate the histone modifications status and local chromatin structure.

Regulation of genome architecture

Chromosomes are dynamic structures, especially during cell division resulting in wide distances between regulatory elements and their targets. In order to connect those elements while maintaining specificity, chromatin compartments are formed within the nucleus. Compartments are formed by self-association of chromatin termed TADs (topologically associating domains). TADs are DNA regions in the genome, within which physical interaction occurs quite frequently. CTCF has the ability to affect gene expression by bringing together elements, which are far apart from each other often forming loop structures. At TADs, a great number of CTCF binding sites was found together with Rad21 (a cohesin subunit) binding sites. CTCF, if located within a TAD, was shown to facilitate the interaction between regulatory elements. Binding sites for CTCF were not only identified within TADs but also at their borders suggesting that CTCF establishes TAD borders together with other proteins, like TFIIIC. TAD borders are thought to inhibit the interaction of sequences located in two adjacent TADs. The majority of CTSs, however, were found within TADs mediating so called short-range intra-TAD interactions (Lin et al., 2012). Those interactions allow appropriate binding of enhancers within a TAD to their promoter.

In summary, the transcriptional regulator CTCF is an important chromatin organizer and epigenetic regulator of genes involved in many cellular processes such as proliferation, differentiation and apoptosis, and in cancer.

Hematopoietic cell differentiation

1.2.1 Hematopoiesis

Hematopoiesis is the process in which all components of the blood are formed. All cell types of the blood originate from a common precursor, the self-renewing hematopoietic stem cells (HSCs). In adult mammals, HSCs reside in small number in the bone marrow and produce around 300 million of blood cells every minute in a healthy adult (Orkin and Zon, 2008). During embryogenesis, three waves of hematopoiesis occur involving many different anatomical sites (the yolk sac, the aorta-gonad-mesonephros region, the placenta and the fetal liver). At birth, HSCs migrate to the bone marrow, where they are maintained in specialized niches. There, multipotency is conserved through cell divisions while their progeny are directed towards lineage differentiation. Mature blood cells comprise myeloid and lymphoid lineage-derived cell types. Myeloid lineage involves erythrocytes, megakaryocytes, granulocytes (basophils, neutrophils and eosinophils), monocyte/macrophages and mast cells, while lymphoid cells are T cells, B cells and natural killer cells. Dendritic cells can derive from either the myeloid or the lymphoid lineage (not shown in figure) (figure 3) (reviewed in (Nandakumar et al., 2016). Additionally to the classical model of hemopoietic differentiation (figure 3, black arrows), recent studies suggest an alternative model (figure 3, red arrows), which is thought to be used only under special circumstances or in combination with the classical model. At the top of the hierarchy, both models depict the hemopoietic stem cell (HSC). The alternative model suggests that erythroid- and megakaryocyte progenitors (MEP) emerge either directly from HSCs or from their immediate downstream progenitors (Nandakumar et al., 2016).

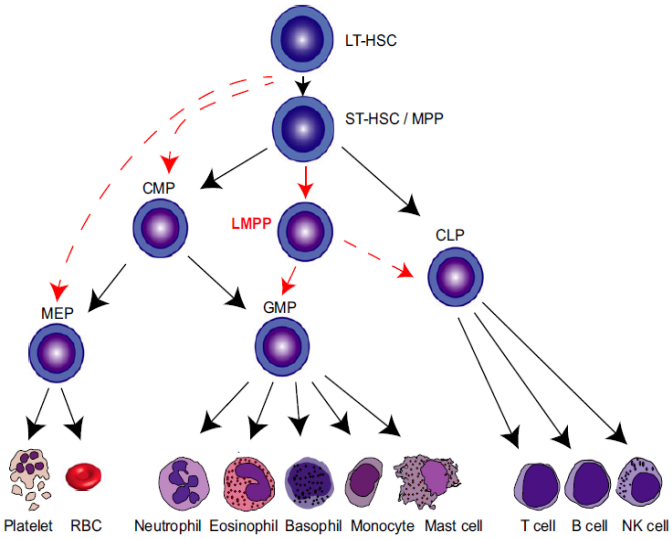

Figure 3. The hierarchy of hematopoietic cells: LT-HSC: Long Term-Hemtopoietic Stem Cell; ST-HSC:Short Term-Hematopoietic Stem Cell; MPP: Multipotent Progenitor; CMP: Common Myeloid Progenitor; CLP: Common Lymphoid Progenitor; MEP: Megakaryocyte-Erythroid Progenitor; GMP: Granulocyte-Macrophage Progenitor (reviewed in Nandakumar, Ulirsch and Sankaran, 2016)

1.2.2 Erythropoiesis

Erythropoiesis is a very active process generating around 2 x 10^{11} red blood cells (erythrocytes) per day, which supply all tissues within the body with oxygen. After birth, the HSCs remain in the bone marrow and give rise to their progenitors (figure 4). During erythropoiesis, MEPs differentiate to committed erythroid progenitors (EPs), which are the burst-forming unit erythroid cells (BFU-E) and the colony forming unit erythroid cells (CFU-E). Both cell types are able to form colonies, however, only BFU-Es can be found in the circulation under non-pathological conditions. Further differentiation to proerythroblasts (ProE) allow

first distinction due to morphology. They then mature to basophilic (BasoE), polychromatophilic (PolyE) and orthochromatic (OrthoE) erythroblasts decreasing constantly in size while increasing in hemoglobin production. Enucleation leads to the formation of reticulocytes (Retic), which are able to enter the circulation. After around 24 hours, they further mature to red blood cells (RBC), which are not longer capable of proliferation (Nandakumar et al., 2016).In order to be able to study erythroid differentiation, several erythroid markers have been established. MEPs and EPs express CD34 and CD38 but lack several lineage markers. Nevertheless, isolated MEPs are quite heterogenous and most probably contain committed erythroid progenitors (EPs). More differentiated cells can be discerned by the expression of surface markers, like glycophorin A (CD235a) or transferrin receptor (CD71) (Nandakumar et al., 2016).

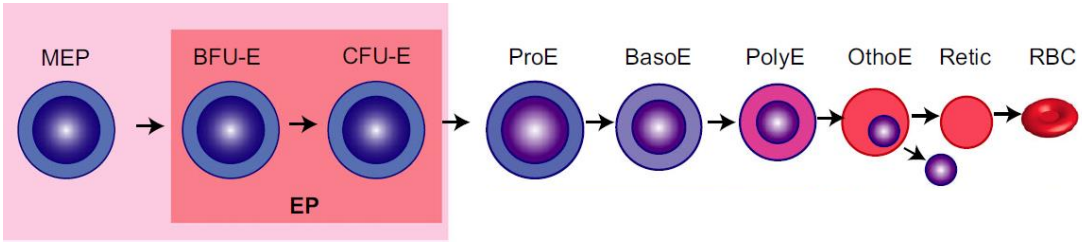

Figure 4. The differentiation steps from the megakaryocyte erythroid progenitor (MEP) to the mature red blood cell (RBC). MEP, Megakaryocyte erythroid progenitor; BFU-E, blast colony forming unit - erythroid; CFU-E, colony forming unit- erythroid; ProE, proerythroblast; BasoE, basophilic erythroblast; PolyE, polychromatic erythroblast; OrthoE, orthochromatic erythroblast; Retic,reticulocyte. (adapted from Nandakumar, Ulirsch and Sankaran, 2016)

1.2.3 Molecular regulation of erythropoiesis

Several growth factors and cytokines play pivotal roles during the erythroid differentiation process. One important cytokine is erythropoietin (EPO) acting primarily on erythroid precursors. *In vitro* studies showed that EPO is essential for CFU-Es and also BFU-Es may be dependent on EPO. The stem cell factor (SCF, KITLG) is another important factor, which binds to the KIT receptor and acts as a receptor tyrosine kinase. Together with EPO, insulin, insulin-like factor, activin and angiotensin II, SCF stimulates erythropoiesis. Erythropoiesis is moreover regulated on the transcriptional level by so called master transcription factors (TFs). It is believed that there is a core erythroid network (CEN) involving DNA binding factors, like GATA1, TAL1 and KLF1, and non-DNA binding factors, as LDB1 and LMO2 (figure 5A). Lacking one of the above mentioned transcription factors led to inhibition of erythropoiesis in mice. Additionally to the CEN TFs, other factors interacting with the CEN were found to exhibit pivotal roles. NuRD Complex was found to be associated with GATA1 and TAL1 resulting in repression of gene expression (figure 5B). Furthermore, other factors, like FOG1, NFE2, GF1B, ETO2, TAF10, MYB and ZBP-89, seem to be involved in the transcriptional regulation of erythropoiesis (Nandakumar et al., 2016).

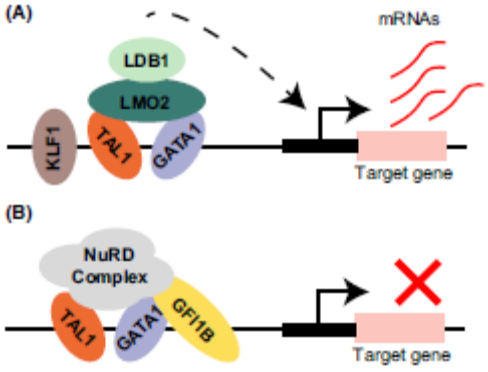

Figure 5. The core erythroid network (CEN) of transcription factors (TFs). The CEN loops from enhancers to promoters to activate target gene expression. (B) TFs in the CEN can interact with additional TFs. GATA1 and TAL1 interact with the nucleosome remodelling deacetylase (NuRD) complex, resulting in target gene repression. (adapted from Nandakumar, Ulirsch and Sankaran, 2016)

2. PROBLEM STATEMENT, HYPOTHESIS, AIMS

2.1 Previous results on CTCF and erythroid differentiation

As described above, CTCF has the ability to bind a huge variety of DNA sequences all over the genome and therefore, a pivotal role in regulation of transcription, insulation, epigenetics and genomic architecture are supposed. Our previous data suggest a possible role of CTCF in the regulation of erythroid cell differentiation. It could be shown that CTCF was differentially expressed and post-translationally modified, comparing several human myeloid cells, depending on the particular differentiation pathway. Furthermore, it could be shown that overexpression of CTCF promotes differentiation into the erythroid lineage. Previous results of Affymetrix microarrays comparing the transcriptome of parental K562 cells and cells overexpressing CTCF suggest that CTCF regulates erythroid specific genes. Those results indicate a feasible role of CTCF during myeloid differentiation.

2.2 Aims

Our published data demonstrate that CTCF induces the erythroid differentiation in pluripotent K562 cells but the underlying molecular events of these processes are unknown. To understand those events, we aim to:

a. Verify a pivotal role of CTCF during differentiation.
b. Silence CTCF expression to explore its effect on K562 cell differentiation and target gene expression using lentiviral particles.
c. Characterize genes regulated by CTCF, particularly genes encoding for erythroid transcription factors.
d. Study the binding of CTCF to the regulatory regions of its putative target genes upon differentiation.
e. Elucidate the role of CTCF in human pluripotent $CD34^+$ hematopoietic stem cells

3. MATERIALS AND METHODS

3.1 Cell Culture

Cell lines and maintenance

Cells are grown in either RPMI-1640 or DMEM basal media (Lonza) supplemented with 10 % (v/v) fetal bovine serum (FBS, Lonza), 150 µg/mL of gentamycin (Lab. Normon) and 2 µg/mL of ciprofloxacin. $CD34^+$ cells are cultured in Stemline® Hematopoietic Stem Cell Expansion Medium (Sigma). The stemline® medium is supplemented with 2mM L-glutamine 1% P/S, 1% anfotericina, 10µg/µL human insulin, 100ng/mL SCF, 100ng/mL TPO, 100ng/mL Flt3-Ligand, 20ng/mL IL-3, 100ng/mL holo transferring and 0.0001M 2-mercaptoethanol Cells are maintained exponentially growing in a humidified atmosphere at 37°C and 5% CO2 unless indicated.

Table 1 provides a list of the human cell lines used in this study.

Table 1: Human cell lines

Cell line	Background	Culture medium	Origin/Reference
HeLa	Human cervical	DMEM 10% FBS	Laboratory collection ATCC
HEK-293T	Human embryonic kidney expressing the T antigen	DMEM 10% FBS	Laboratory collection ATCC
K562	Human chronic myeloid leukemia	RPMI 10% FBS	Laboratory collection ATCC
$CD34^+$	Human primitive blood- and bone marrow-derived progenitor cells	Stemline® Hematopoietic Stem Cell Expansion Medium	Blood cord

For investigating in erythroid differentiation, K562 multipotential cell line and $CD34^+$ primary pluripotent cell line purified from cord blood are utilized. K562 is a human chronic myeloid leukemia cell line in blast crisis, which have the ability to differentiate into progenies of the erythroid, megakaryocytic and monocytic line. $CD34^+$ cells were purified from cord blood kindly donated from the Banco de Sangre y Tejidos de Cantabria. $CD34^+$ cells are induced with 3U/mL and 6U/mL EPO (500U/mL stock) together with 2mM L-glutamine 1% P/S, 1% anfotericina, 10µg/µL human insulin, 25ng/mL SCF. As per definition, 1U EPO elicits the same erythropoiesis-stimulating response in rodents as 5µM of cobaltous chloride (Published by Oxford University Press on behalf of ERA-EDTA, 2009)

Purification of $CD34^+$

$CD34^+$ were purified from cord blood using ficoll for separation of the blood components. Then the mononuclear cells were further purified using α-CD34 antibodies and µbead (Miltenyi Biotec) to obtain a more homogenous cell population. All steps were carried out according to the manufacturer's instructions.

Cell proliferation and viability assay

K562 cells are seeded at $2.5x10^5$ cells/mL and proliferation is measured by counting cell number. Cells are counted in a hemocytometer or the NucleoCounter® NC-100™ system (Chemometec). If working with drugs, the drug is added at the same time as the cells are seeded (t=0 hours). Viability of cells was determined using trypan blue, which is not taken up by viable cells with an intact cell membrane. Trypan blue (0.4%) is added 1:1 to a small amount of cell suspension (10µL) and applied to a neubauer chamber. Stained (blue) and unstained cells are counted and the percentage of viable cells is calculated.

Benzidine test

In order to study erythroid differentiation, 5×10^4 cells are subjected to benzidine test. This test is based on the catalytic reaction occurring between benzidine and hemoglobin in presence of H_2O_2. Cells are collected and centrifuged at 1500 rpm for 3 min at RT and supernatant is discarded by aspiration. Cells are resuspended in 20µL of growing medium and placed on ice. 20µl of the benzidine-H_2O_2 (30%) mixture (50:1 v/v) are added and incubated on ice in the dark for 5-10 min for the color reaction to occur. Resuspended cells are then applied to a neubauer chamber and the percentage of hemoglobin producing cells (blue) to non-hemoglobin producing cells was evaluated by microscopy. Several pictures are taken and cells are counted using ImageJ software.

Benzidine: dissolved in 0.5M acetic acid solution; stock concentration: 0.2%; stored at 4°C protected from light.

Drug treatments

Imatinib (0.5 or 1µM; as indicated) and 1µM Ara-C (cytosine arabinoside) were used in K562 cells to induce erythroid differentiation. Cytosine arabinoside is incorporated into DNA while synthesis and damages DNA due to its rapid conversion to cytosine arabinoside triphosphate. Imatinib is a specific inhibitor of tyrosine kinases by binding close to the ATP binding site and therefore, locking the enzyme in a non-active conformation.

Ara-C (Sigma-Aldrich®) dissolved in distilled water; stock concentration: 100mM; stored at -20°C.

Imatinib (LC Laboratories, Woburn, MA, USA) dissolved in DMSO; stock concentration: 1mM; stored at -20°C

3.2 Lentivirus production and infection

Transfection of HEK293T cells with PEI

The mixture of the three different plasmids was performed in a ratio of 1:3:4 (VSVG: psPAX2: transfer plasmid, respectively). The following transfer plasmids were used: pLKO (Addgene), pTRIPZ (Dharmacon™ GE Healthcare) (vector maps in appendix figure A2). HEK293T cells were seeded in 150 mm Ø plates at 70-80 % of confluence. Total amount of transfected DNA was 50 µg (6:19:25 µg). Mixture of DNA+PEI was added to the cells containing 15 mL of DMEM (without serum and antibiotics) and after

12 hours, mixture + DMEM were replaced by 15 mL of complete medium. 48 hours after transfection, lentiviral particles-containing supernatant was collected and stored at 4°C and 15 mL of new complete medium were added. 24 hours afterwards, the lentiviral particles-containing supernatant was collected again. Supernatants were clarified by centrifugation at 1500 rpm for 10 min and then filtered through a 45µm pore size sterile syringe filters.

Lentivirus concentration: PEG8000 was added to a final concentration of 15%. Mixture was homogenized by inversion and incubated at 4°C for at least 6 hours and a maximum of 7 days allowing PEG - lentivirus complex formation. After centrifugation at 1500xg for 30 min at 4°C in order to concentrate the lentiviral particles, the lentivirus-containing pellet was resuspended in serum-free media (150µL of serum free media for each 15 mL of collected supernatant), aliquoted and stored at -80°C.

PEI (Polysciences, Inc.) dissolved in distilled water; stock concentration 1mg/mL, stored at -80°C

PEG8000 (Sigma-Aldrich®): dissolved in PBS1X; stock concentration 40% (w/v); autoclaved; stored at RT.

Cell transduction

For transducing, the amounts of lentivirus used vary between a MOI (multiplicity of infection) from 3 to 5. The MOI indicates the ratio of the number of virus particles to the number of target cells. In all cases, polybrene was added in order to increase the efficiency of the transduction (5µg/mL for K562 cells). $2.5x10^5$ cells/mL were seeded in half of the normal volume in serum-free media and the corresponding amount of lentiviral particles and the polybrene were added. After 12 hours, complete media was added until the corresponding total volume and 48 hours after infection, the cells were centrifuged and the lentivirus-containing media was discarded. Cells were plated at $2.5x10^5$ cells/mL in complete media and 1µM/mL puromycine was added to select the infected cell population for at least 2 days. 0 hours correspond to two days of selection with 1µM/mL puromycin (in total t=0 correspond to 4 days after infection). If the titer was very low, $0.5x10^5$ cells /condition were used, centrifuged at 1500 rpm for 3 min and resuspended in 50µL RPMI (without FBS and antibiotics). Those cells are then added to the total amount of produced viruses and centrifuged for 60 min at RT (1500rmp) for higher transduction efficiency. The mixture is plated in a T6 well plate and incubated for 12 hours at 37°C. On the next day, the media was changed in order to remove dead cells. 48 hours after transfection, selection with 1µM/mL puromycin was carried out for 2 days.

Polybrene (Sigma-Aldrich®) dissolved in distilled water; stock concentration: 5mg/mL; stored at -20°C.

Lentivirus titer

HeLa cells were seeded at a concentration of 2×10^4 cells in a 6-well plate. When attached, media was replaced by 1.5 mL of free-serum media and 4 µg/mL of polybrene was added. Each well was transduced with a different amount of the concentrated lentivirus (10µL, 5µL, 1µL and 0.5µL). After 12 hours, complete media was added until 3 mL in total to every well and 48 hours after infection, lentivirus

containing media was replaced by complete media and the selective antibiotic (1 μM/mL of puromycin) to select the infected cells. Puromycin-containing media was replaced every 1 or 2 days, until control cells (not infected cells) die due to the puromycin and the single-infected cells form colonies of more than 6-10 cells. Then, the medium is removed and the plate washed 2x with PBS. Cells were stained with Crystal Violet and colony number was determined by counting under the microscope. The titer of the lentivirus was calculated as following:

$$\frac{x\ number\ of\ colonies}{\mu L\ of\ virus\ added}\ x\ 10^3 = U.F.C/mL$$

3.3 DNA and RNA analysis

DNA purification

For DNA purification, the QIAquick PCR Purification Kit was utilized and carried out according to the included protocol. DNA was eluted in a volume of 30-60μL Elution Buffer.

RNA extraction and purification

Total RNA extraction from cell cultures was performed using TRizol® Reagent (InvitrogenTM) following the manufacturer's instructions. Between 2-5x10^6 cells were used for RNA extraction and they were lysed in a total volume of 1mL of TRizol®. After purification, the RNA pellet was resuspended in pre-warmed (65°C) RNAse-free water in a final volume from 30μL and stored at -80°C. RNA concentration was determined by measuring A260nm using a microvolume spectrophotometer (ThermoScientificTMNanoDrop 2000).

Reverse Transcription (RT) and quantitative polymerase chain reaction (qPCR)

Reverse transcription

Reverse transcription (RT) was performed of total extracted RNA. 1μg was used for reverse transcription using the iScript cDNA Synthesis Kit (Bio-Rad) according to manufacturer's instructions in a total volume of 20μl and diluted 1:2 in nuclease free water. The obtained cDNA was stored at -20°C until used.

The following protocol was set for the reaction: 25°C 5min; 42°C 30 min; 85°C 5min; 6°C ∞

Quantitative polymerase chain reaction (qPCR)

Primers for the specific genes of interest were designed using the online Primer 3 software tool (http:\\frodo.wi.mit.edu/primer3/). Primer sequences and amplicon sizes used in RT-qPCR assays can be found in the appendix, table A1.

The SYBR® Select Master Mix (Applied BiosystemsTM) supplied in a 2x concentration premix was used in a CFX ConnectTM Real-Time PCR Detection System (Bio-Rad). The qPCR reaction was prepared as follows: 2x SYBR® Select Master Mix were mixed with primers final concentration 0.2μM, 3μL of DNA/cDNA and distilled water until 35μL of final volume. Then, the reaction mix was added to a 96-well

white PCR plate in duplicates (15µL of the reaction mix in each well). Reaction mix without DNA/cDNA was used as negative control.

The protocol used for amplification was the following: 95°C 3min; (95°C10s; 56°C 30s; 72°C 10s) (40 x cycles); 72°C 30s.

The protocol for real time melting curve was the following: 55°C 10 s; 60°C 10 s; x °C +0.5°C 10 s; 95°C 10s

qPCRs were analyzed with the CFX ManagerTM software. Threshold cycles (Ct) were determined by default at the beginning of DNA amplification in the exponential phase. The mRNA expression of genes of interest was normalized to mRNA expression of housekeeping gene (S14) using the comparative Delta Ct (ΔCt) method: $\Delta Ct = 2^{Ct\ normalizing\ gene - Ct\ gene\ of\ interest}$

3.4 Protein analysis

Westernblot

Cell pellet lysis was performed using RIPA lysis buffer. Suspension cells were harvested and centrifuged at 1500 rpm for 3 min The cell pellet was washed once with 1X PBS, centrifuged at 1500 rpm for 3 min and stored at - 80°C or directly lysed by adding 100µL/10 million cells of lysis buffer (pipetting up and down). Homogenized samples were kept on ice for 30 min. Then, protein samples were sonicated using the Bioruptor® Plus sonication device (Diagenode) set at high powersetting for 10 cycles (30s ON, 30s OFF). After sonication, cell lysates were clarified by centrifuging 20 min at 4°C at 15000 rpm. The supernatant containing the proteins was collected and stored at -80°C until used. Quantification of protein concentration was carried out using the Qubit®Protein Assay Kit in a Qubit® 2.0 Fluorometer.

50µg per sample were mixed with 5x loading buffer (Laemli) and boiled at 95°C for 5 min for denaturation. Samples were loaded and separated according to their molecular weight in a polyacrylamide-SDS (SDS-PAGE) gel (8%-15%). Electrophoresis was carried out in a Mini Protean III cuvette powered by a basic PowerPAC supply (Bio-Rad) at constant voltage of 175 V for ~1hour, using electrophoresis running buffer. "Precision PlusProteinTMDual Color Standards" (Bio-Rad) were used as markers. Proteins were transferred from acrylamide gel to a nitrocellulose membrane (AmershamProtran Supported 0.45 NC, GEHealthcare Life Sciences) in a Mini-Trans Blot cell (Bio-Rad) using transfer buffer at constant amperage of 400 mA for 30 min-1 hour depending on the molecular weight of the proteins of interest. Then, the polyacrylamide-SDS gels were stained with Coomassie Brilliant Blue solution for 10 min at RT and distained with water to check proteins load and integrity. The membrane was blocked with 10% (w/v) milkpowder in TBS (10mL) for 1-1.5h in agitation at RT to avoid unspecific antibody binding and then washed 3x. Every washing step was carried out with TBS-T at RT for 10 min. Primary antibodies were incubated overnight at 4°C or 3-4 hours at RT in agitation. After washing the membrane 3x, the secondary antibody was incubated at RT for 45 min in the dark. The membrane was washed 3x and recorded with an

Odyssey® Infrared Imaging Scanner (Li-Cor® Biosciences). Immunoblot quantification and densitometry analysis was carried out using the ImageJ software.

Data for primary and secondary antibodies are available in the appendix, table A2 and A3.

RIPA Lysis Buffer: 150mM NaCl, 50mM Tris pH 6.8, Sodium Deoxychloride 0.5% (w/v), NP-40 1% (w/v), 0.1 % SDS (v/v) supplemented with protease inhibitor cocktail Set I (1:100; Calbiochem) and phosphatase inhibitor cocktail I (1:100; Sigma-Aldrich®) immediately before using; stock concentration: 1%; stored at 4°C.

5X-SDS-PAGE loading buffer: 100 mM Tris-HCl pH6.8, 5% β-mercaptoethanol (v/v), 5% SDS (w/v), 0.1%bromophenol blue (w/v), 50% glycerol (v/v); stock concentration: 5X; stored at -20°C.

CoomassieBrillant Blue solution: 0.025 % Coomassie Brilliant Blue R-250 (w/v), 40 % methanol (v/v) and 10 % (v/v) glacial acetic acid; stock concentration: 1X; stored at RT.

Running buffer: 25 mM Trizma pH8.3, 192 mM glycine and 0.1% SDS (w/v); stock concentration: 1X; stored at RT.

TBS: 20 mM Tris-HCl pH7.5, 150 mM NaCl; stock concentration: 1X; stored at RT.

TBS-T: 0.05 % Tween 20 (v/v) diluted in TBS; stock concentration: 1X; stored at RT protected from light.

Transfer buffer: 25 mMTris pH8.3, 192 mM glycine and 10% methanol (v/v); stock concentration: 1X; stored at RT.

3.5 Chromatin Immunoprecipitation Analysis (ChIP)

K562 cells were treated with Ara-C and Imatinib. $20x10^6$ cells/condition were collected, washed with 1x PBS and fixed with 10mL 1% formaldehyde diluted in PBS for 10 min at RT rotating. Fixation was stopped with Glycine (final concentration of 125mM) for 5 min at RT rotating. After centrifugation for 5 min at RT 15000 rpm, the pellet was washed 2x with 1x PBS and frozen at -80°C or lysed with between 800µL and 1.2 mL of Lysis Buffer (+protease inhibitor 1:100) (1mL for $20x10^6$ cells) (appendix table A5). From now on, all steps are performed on ice. The samples are resuspended with a syringe (5x), incubated for 10 min on ice and then split in aliquots of 200-250µL for sonication using the Bioruptor® Plus sonication device (Diagenode) set at high powersetting for 10 cycles (30s ON, 30s OFF). After centrifugation for 20 min at 4°C (14000rpm), the samples are stored at -80°C.

Input: 30µL of each sample are diluted in 170µL Dilution Buffer (appendix table A5) supplemented with 6µL RNAse and 12µL NaCl (5M) for decrosslinking. After incubation of the samples at 65°C shaking (1500rpm) O/N, the DNA is purified and run on a 1% gel to check if the sonication worked well. The majority of the bands should be visible between 200 and 500 bp. Only if this is the case, the samples can be used for next procedure.

Chromatin immunoprecipitation (ChIP) was performed with a mixture of 3 CTCF antibodies (2µL of each per sample) (appendix, table A2). The ChIP was performed at 4°C O/N rotating.

On the next day, 30µL/sample of Dynabeads® Protein G (Invitrogen) were washed with 1mL Dilution Buffer and blocked with 14µL/sample Salmon Sperm DNA for 1 hour at 4°C rotating. Then 44µL of the beads are added to each sample and incubated for 1-2 hours at RT or 6 hours at 4°C rotating. Then, washing

steps are performed with 1mL of the following buffers for 5 min each, rotating. With the help of a magnet, the beads were hold back during the washing steps: Low Salt Wash Buffer, High Salt Wash Buffer, LiCl Wash Buffer and 2x with TE Buffer (appendix table A5). The samples are eluted wit 200µL Elution Buffer for 30-40 min at 65°C shaking (1500rpm). Then 6µL RNAse and 12µL NaCl (5M) are added and incubated at 65°C shaking (1500rpm) O/N. On the next day, 1µL Proteinase K, 2µL 0.5M EDTA and 4µL Tris pH 6.5 1M are added and incubated at 45°C shaking (1500rpm) for 3 hours. DNA is purified and qPCR is performed using primers found in the appendix, table A4

3.6 Data analysis

Results were presented as the mean of two or three independent experiments with error bars representing the standard deviation or the standard error of the mean.

4. RESULTS

4.1 Cytosine arabinosid (Ara-C) and Imatinib induce erythroid differentiation in K562

Previous results have suggested that K562 cells, a multipotent chronic myeloid leukemia cell line in blast crisis, can differentiate along the erythroid pathway upon induction with 1µM Ara-C and 1µM Imatinib (figure 6A). Since the differentiation process is usually associated with cell cycle arrest, proliferation and viability of K562 cells upon induction, was examined. In figure 6B, the proliferation curve of K562 cells is depicted showing reduced proliferation upon treatment with Ara-C and Imatinib. Furthermore, trypan blue assay was carried out in order to check cell viability upon treatment. As shown in figure 6C, treatment with 1µM Ara-C lead to a slight reduction in viability. However, treatment with 1µM Imatinib seemed to induce cell death. Therefore, in order to check apoptosis, PARP (Poly (ADP-ribose) polymerase) was analyzed by westernblot. PARP can be cleaved to PAR, which further leads to programmed cell death. In figure 6D it can be seen, that after treatment with Ara-C, cells start undergoing apoptosis after 4 days of treatment, however, when treated with Imatinib, high levels of cleaved PAR (*) are visible already after 2 days of treatment. After 5 days of treatment with Imatinib, no viable cells could be collected and therefore, no proteins extracted.

The presence of hemoglobin within the cell is an exclusive characteristic of erythroid cells. In order to examine hemoglobin production, we decided to perform benzidine test, with which we are able to identify hemoglobin producing cells due to a chemical color reaction (blue). In figure 7A, the percentage of benzidine positive cells is shown. Treatment with both, Ara-C and Imatinib, increase the number of hemoglobin producing cells compared to control cells, indicating erythroid differentiation. Furthermore, pictures showing benzidine positive and negative cells are shown. Blue cells indicate hemoglobin production.

Fetal hemoglobin is built up of two α (alpha) subunits and two γ (gamma) subunits, whereas adult hemoglobin involves two α and two β (beta) subunits. Erythroid differentiation was further assessed by γ-globin expression. In figure 7B, expression levels of γ-globin were examined by westernblot. Treatment with Ara-C lead to a clear increase of γ-globin within 5 days compared to control untreated cells. Treatment with Imatinib also showed an increase in γ-globin levels within the first two days, however due to a reduced amount of proteins extracted, no clear increase can be detected after 3 days.

Additionally, our protein of interest, CTCF, was analyzed upon differentiation by westernblot. In figure 7C, it is visible that CTCF levels remain quite stable upon differentiation. However, MYC, an important driver of cell cycle, is shown to decrease upon differentiation associated with reduced cell proliferation.

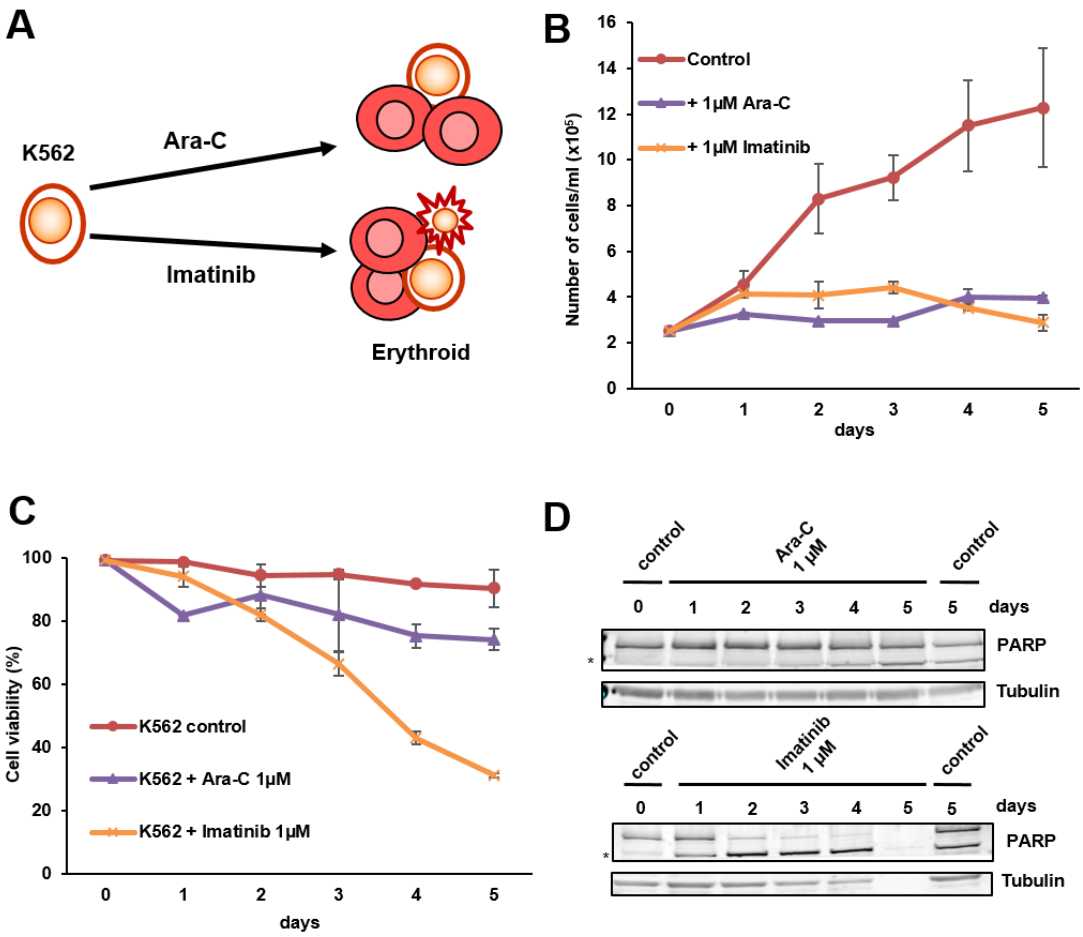

Figure 6. Proliferation and viability of K562 upon induction of differentiation with Ara-C and Imatinib.
A) Schematic representation of experimental workflow B) Proliferation curve of K562 treated with 1µM Ara-C and 1µM Imatinib. $2.5x10^5$ cells/mL were seeded. Error bars mean the STD of two independent experiments. C) Cell viability (trypan blue assay) of K562 treated with 1µM Ara-C and 1µM Imatinib. $2.5x10^5$ cells/mL were seeded. Error bars mean the STD of two independent experiments. D) Analysis of PARP (poly(ADP-ribose) polymerase) as marker for apoptosis by westernblot after treatment with Ara-C and Imatinib. * indicate cleavage of PARP. Tubulin was used as loading control.

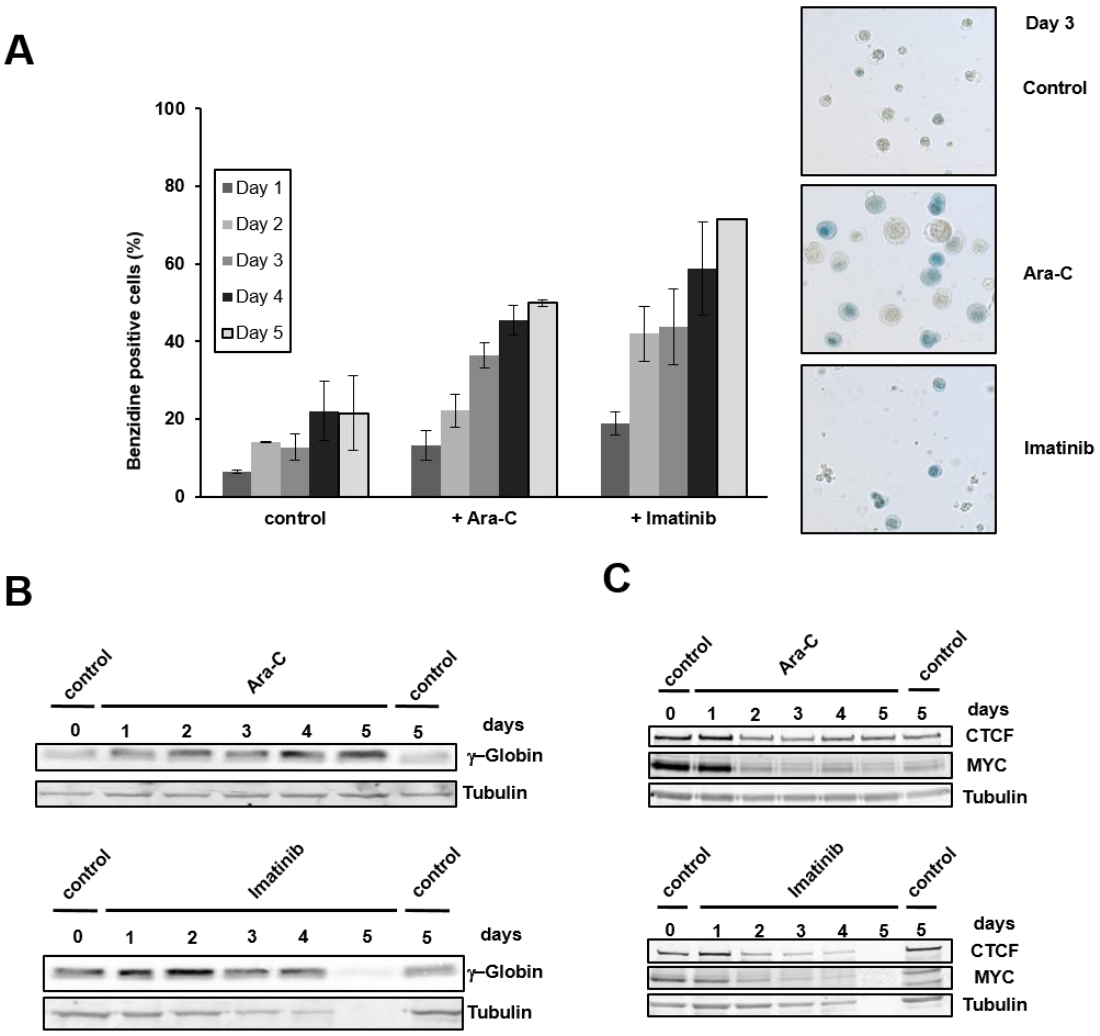

Figure 7. Assessment of differentiation of K562 treated with Ara-C and Imatinib A) Benzidine test of K562 cells. Treatment with 1μM Ara-C and 1μM Imatinib lead to an increase in benzidine positive cells compared to control cells. Error bars mean the STD of two independent experiments. B) Erythroid differentiation was evaluated by westernblot analyzing expression of γ–globin, an erthroid marker. Tubulin was used as a loading control. C) CTCF and MYC expression was analyzed upon treatment with Ara-C and Imatinib. Tubulin was used as a loading control.

4.2 Silencing of CTCF inhibits erythroid differentiation of K562 cells

In order to study the role of CTCF during the erythroid differentiation process, lentiviral particles containing specific shRNA against CTCF mRNA were used to knockdown CTCF in K562 cells. K562 cells were transfected with lentiviral particles containing the empty vector pLKO (EV) and the vector containing shCTCF (shCTCF). After infection, the cells were selected for puromycin. In figure 8A, the validation of CTCF knockdown by westernblot and RT-qPCR is shown. In order to induce differentiation, K562 cells infected with lentiviral particles were treated with 1µM Ara-C or 0.5µM Imatinib. Proliferation and viability were assessed for 3 days. In figure 8B, the proliferation curve is depicted showing that both treatments lead to a reduction of proliferation, whereas untreated (control) cells did not show a reduction of proliferation.

Trypan blue assay for cell viability evaluation is shown in figure 8C. K562 cells infected with the pLKO empty vector (EV) show a similar viability pattern as uninfected cells (comparing figure 6C and 8C). Treatment with Imatinib reduced cell viability to ~40% even after reduction of the concentration when infected with the EV. Ara-C did not affect cell viability in cells infected with the pLKO empty vector. Untreated cells infected with shCTCF did not show reduced viability. Treatment with Ara-C reduced viability to ~80% and treatment with Imatinib reduced viability to ~70%. Interestingly, knockdown of CTCF seems to increase viability of K562 cells upon treatment with Imatinib.

In order to verify those results, PARP expression was analyzed by westernblot. K562 cells with CTCF knockdown do not show a significant difference in PARP cleavage compared to cell infected with the empty vector upon Ara-C treatment. However, when comparing the treatment with Imatinib, a reduced cleavage pattern of PARP is visible when CTCF is knocked down confirming the results of increased viability upon silencing of CTCF and treatment with Imatinib.

In order to analyze erythroid differentiation upon CTCF knockdown, benzidine test was performed to identify hemoglobin producing cells. In figure 9A, the percentage of benzidine positive cells is visualized. K562 cells infected with the empty vector (EV) show an increased number of benzidine positive cells upon induction of differentiation with Ara-C and Imatinib over 3 days. However, K562 cells infected with lentiviral particles containing shCTCF do not show an increase in benzidine positive cells indicating a pivotal role of CTCF during differentiation. Furthermore, pictures taken during benzidine test evaluation are revealed (figure 9A).

CTCF knockdown, as well as erythroid gene expression, were evaluated by westernblot. In figure 9B, K562 cells infected with lentiviral particles, containing the empty vector (EV) or specific shRNA against CTCF mRNA and Ara-C treatment, are shown. The westernblot shows a clear decrease of CTCF protein levels. Furthermore, an increase in γ-globin when infected with the empty vector (EV) could be revealed. However, upon CTCF silencing, no increase in γ-globin levels can be seen. In order to verify those results, additional erythroid genes were analyzed, GATA1 and LMO2. GATA1 is an indicator for megakaryocytic and/or erythroid commitment and LMO2 is another important transcriptional regulator found in erythropoiesis. Gene expression of both genes was shown to increase upon treatment with Ara-C, but there was no increase

visible when CTCF was silenced, which is consistent with γ-globin expression indicating a pivotal role of CTCF

In figure 9C, K562 cells infected with lentiviral particles, containing the empty vector (EV) or specific shRNA against CTCF mRNA and Imatinib treatment, are shown. Decreased CTCF levels can be shown comparing 24 hours and 48 hours. After 72 hours of Imatinib treatment, less protein could be extracted due to reduced cell viability. Moreover, an increase in γ-globin after 48 hours when infected with the empty vector (EV) can be seen. γ-globin levels, however, do not increase if CTCF is silenced. Looking at the expression levels of GATA1 and LMO2, similar results are shown compared to the treatment with Ara-C).

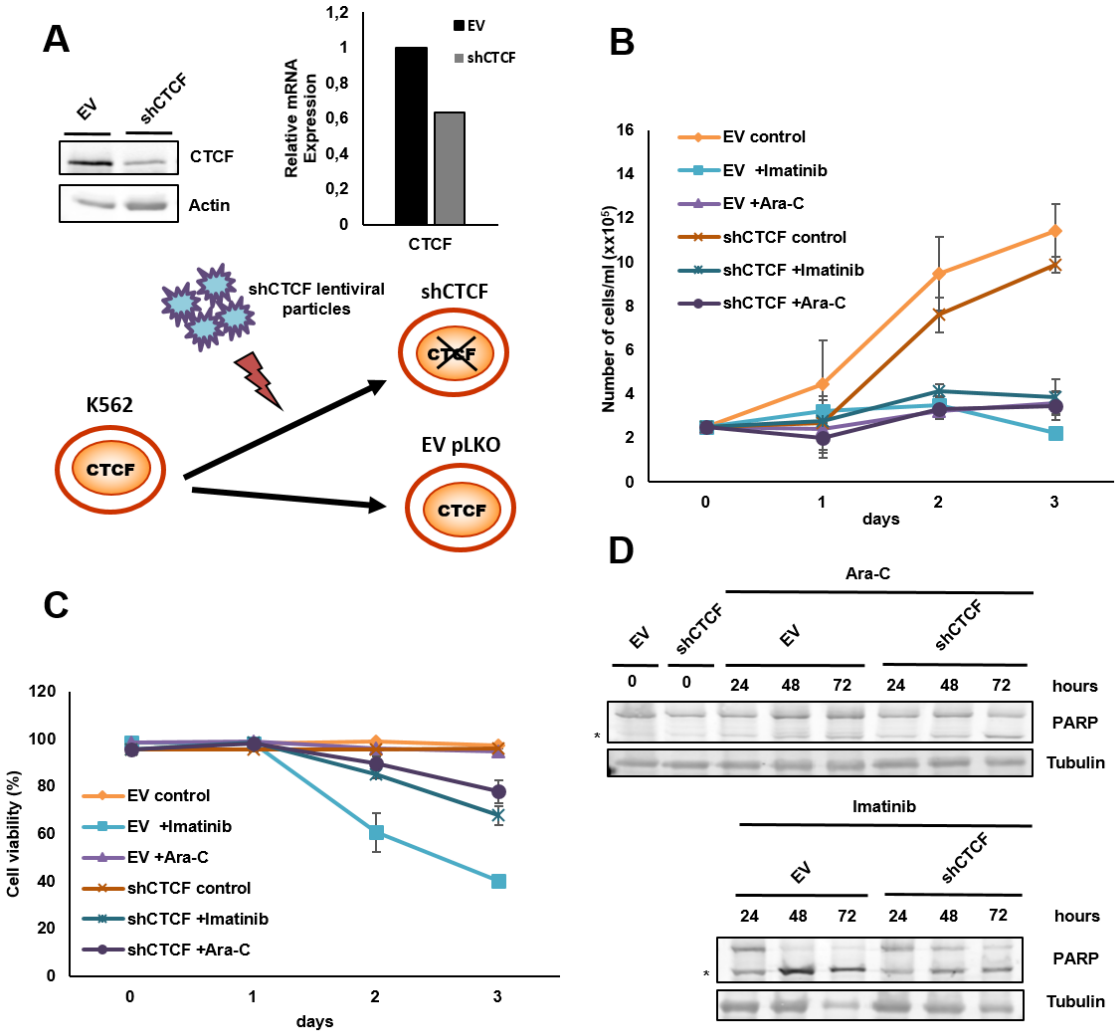

Figure 8. Silencing of CTCF in K562 A) Schematic representation of experimental workflow and validation of CTCF silencing with shCTCF lentiviral particles by westernblot and RT-qPCR B) Proliferation curve of K562 cells infected with lentiviral particles containing specific shRNA against CTCF gene (shCTCF) or the pLKO empty vector (EV). $2.5x10^5$ cells/mL were seeded after 48 hours of puromycin selection. 24 hours correspond to three days of puromycin selection + 1µM Ara-C or 0.5µM Imatinib for 24 hours. Error bars mean the STD of two independent experiments. C) Cell viability (trypan blue assay) of K562 cells infected with lentiviral particles containing specific shRNA against CTCF gene (shCTCF) or the pLKO empty vector (EV). $2.5x10^5$ cells/mL were seeded after 48 hours of puromycin selection. 24h correspond to three days of puromycin selection + treatment with 1µM Ara-C or 0.5µM Imatinib for 24 hours Error bars mean the STD. of two independent experiments. D) Analysis of PARP (poly(ADP-ribose) polymeraseas) as marker for apoptosis by westernblot after silencing of CTCF and treatment with 1µM Ara-C or 0.5µM Imatinib for the hours indicated. Tubulin was used as loading control.

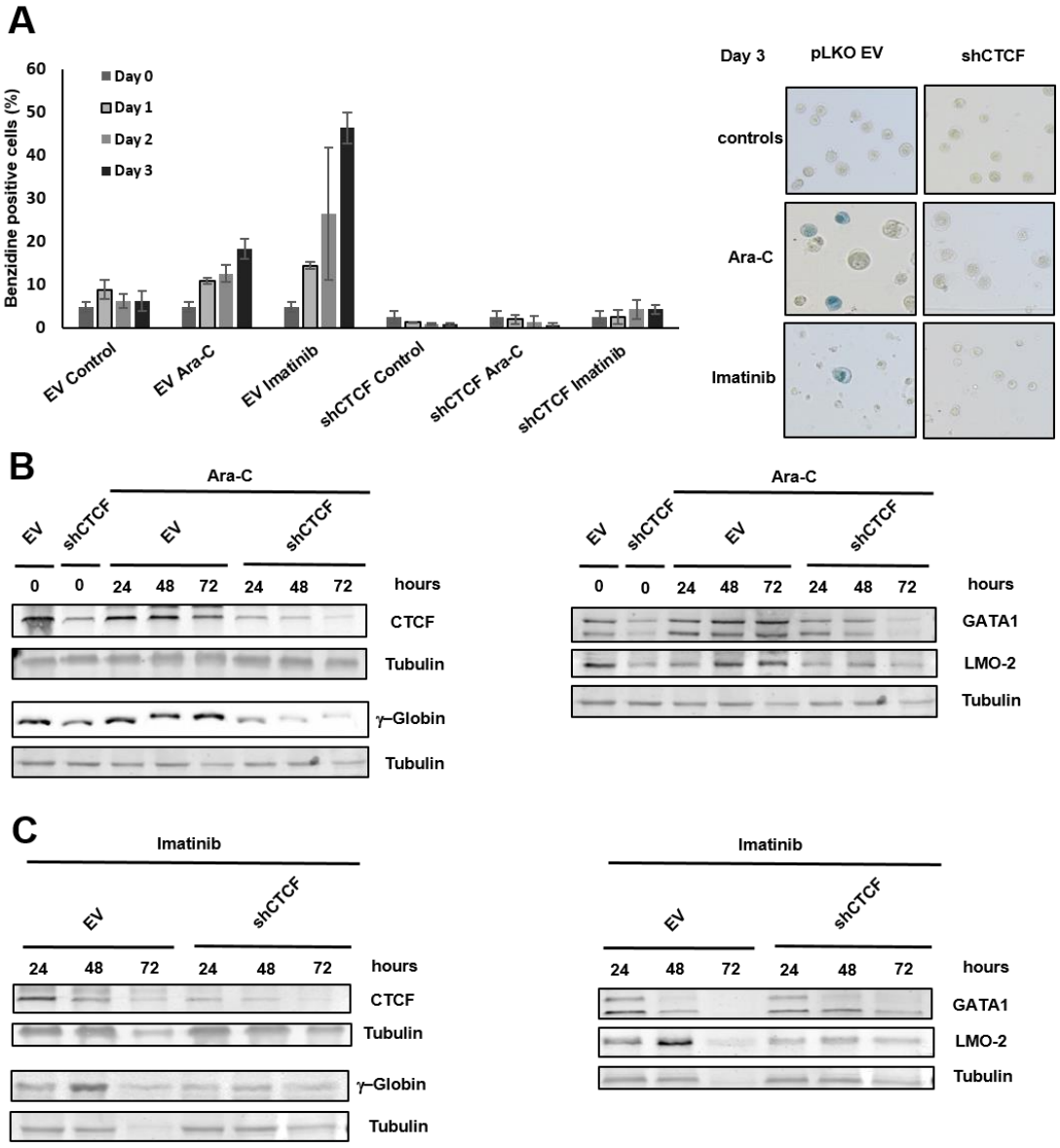

Figure 9. CTCF silencing inhibits differentiation and reduces percentage of benzidine positive cells in K562 cell line. A) Benzidine test of K562 cells infected with lentiviral particles containing specific shRNA against CTCF gene (shCTCF) or the pLKO empty vector (EV). $2.5x10^5$ cells/mL were seeded after 48hours of puromycin selection. 24h correspond to three days of puromycin selection + treatment with 1µM Ara-C or 0.5µM Imatinib for 24 hours. Error bars mean the STD of two independent experiments. Treatment with 1µM Ara-C and 0.5µM Imatinib lead to an increase in benzidine positive cells in cell infected with the empty vector, but not in cells with CTCF silenced. Error bars mean the STD. of two independent experiments. B) Erythroid differentiation was evaluated by westernblot analyzing expression of γ–globin, GATA1 and LMO2 upon treatment with Ara-C. Tubulin was used as a loading control. C) Erythroid differentiation was evaluated by westernblot analyzing expression of γ–globin, GATA1 and LMO2 upon treatment with Imatinib. Tubulin was used as a loading control.

4.3 Preparation of lentiviral particles using the pTRIPZ inducible vector

In order to extend our results using the pLKO constitutive system, a tet-inducible system was chosen to study the effects of CTCF silencing on erythroid differentiation. The pTRIPZ inducible lentiviral shRNA vector is designed as a Tet-On system producing shRNA expression in the presence of doxycycline, a derivate of tetracycline (figure 10A). In the presence of doxycycline, the transactivator is able to bind the TRE and as a result, the shRNA as well as a reporter (TurboRFP) are expressed allowing visual tracking of transduction and shRNA expression. The puromycin resistance marker approve the selection of a stable cell line.

HEK293T cells were transfected with pTRIPZ containing a specific shRNA against CTCF mRNA or pTRIPZ empty vector (EV) together with PAX2 and VSGS in order to produce lentiviral particles (figure 10B). Particles were collected and concentrated.

For titration, HeLa cells were seeded and infected with different amounts of virus. After selection with puromycin and cristal violet staining, a titer of 344.8 $x10^5$ virus/mL for the empty vector and 234.4 $x10^5$ virus/mL for the shCTCF containing vector were obtained (figure 10C).

Due to low titers, an infection favored by centrifugation was chosen (figure 10D). Therefore, $0.5x10^5$ cells/condition resuspended in 50µL RPMI (without FBS and antibiotics) and the total amount of viruses were centrifuged. After 2 days of infection, cells were selected with 1µM/mL puromycine for 2 days. Then, induction with doxycycline was carried out and expression of reporter was checked by microscopy. In figure 11A, pictures with and without induction are shown. On the left, bright field pictures are presented showing the total cell amount. On the right, cells expressing the TurboRFP reporter (red) upon induction with doxycycline are depicted. It can be seen that upon induction with doxycycline, the reporter gene is expressed giving information about transfection efficiency and shRNA expression. For the empty vector, a higher transfection efficiency was obtained than for the shCTCF containing vector.

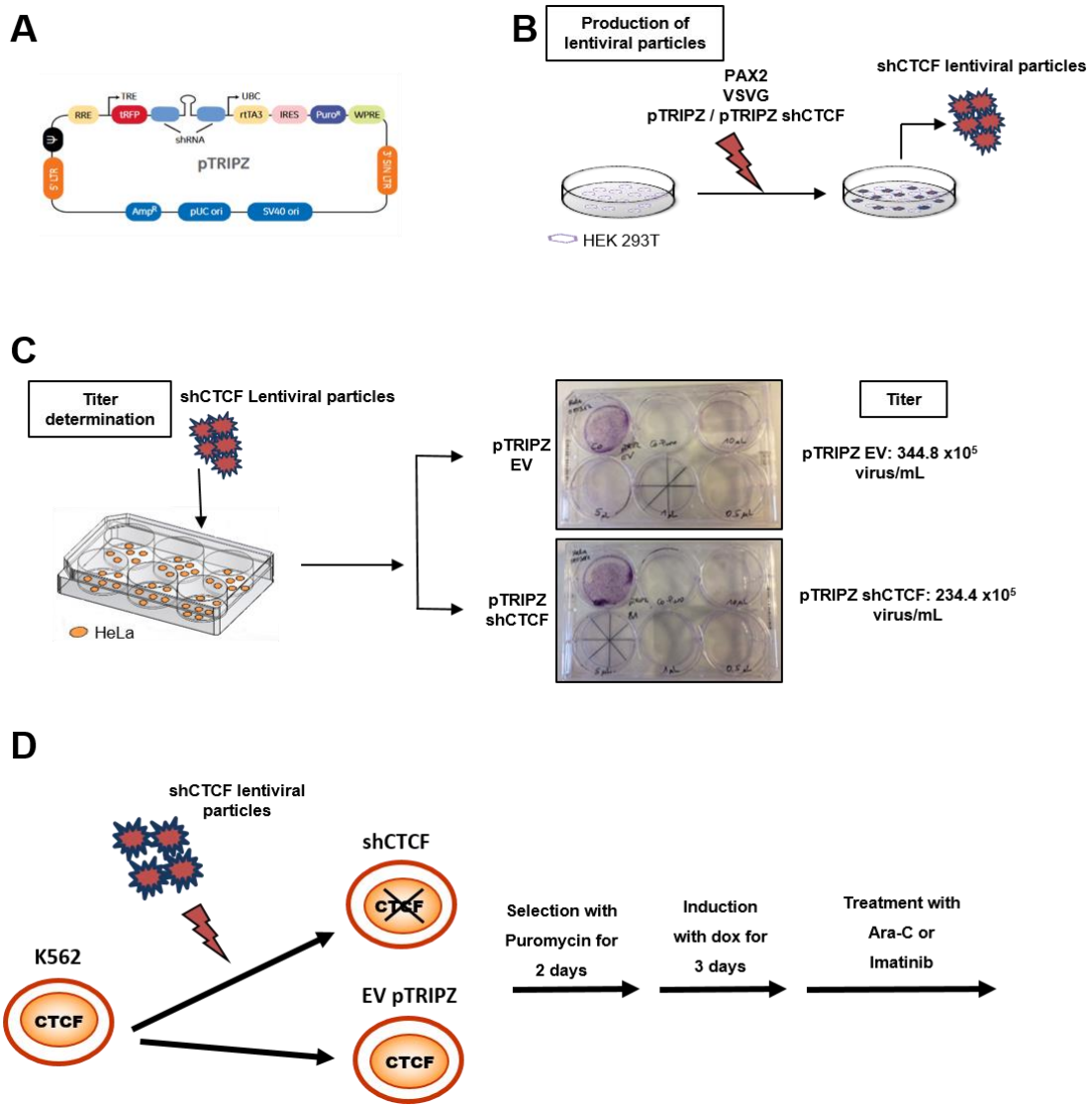

Figure 10. Preparation of inducible shCTCF lentiviral particles A) Vector map of pTRIPZ inducible lentiviral shRNA vector containing puromycin resistance, tetracycline-inducible promoter and TurboRFP reporter for visual tracking of transduction and shRNA expression (13.4 kb). B) Schematic representation of preparation of lentiviral particles. HEK293T cells were seeded and transfected using PEI. The supernatants were collected twice. C) Titer determination in HeLa selecting with puromycin. HeLa were infected with different volumes of viruses (0.5µL, 1µL, 5µL and 10µL). Virus titer of 344.8 x 10^5 virus/mL (EV) and 234.4 x10^5 virus/mL were obtained. D) Schematic representation of infection of K562 with lentiviral particles containing specific shRNA against CTCF gene (shCTCF) or the pTRIPZ empty vector (EV). 0.5x10^5 cells/condition resuspended in 50µL RPMI (without FBS and antibiotics) and the total amount of viruses were centrifuged for 60 minutes at RT. After 2 days of infection, cells were selected with puromycine for 2 days and then induced with doxycycline for 3 more days. Infected cells were treated with 1µM Ara-C and 0.5µM Imatinib.

4.4 Knockdown of CTCF using pTRIPZ inducible system inhibits erythroid differentiation of K562 cells

Silencing of CTCF was assessed by westernblot, shown in figure 11B. Two different shRNA constructs specific for CTCF mRNA were tested, shCTCF 31 and shCTCF 81. As it can be seen in figure 11B, knockdown of CTCF was accomplished with shCTCF81, but not with shCTCF31. The lower figure shows the relative quantification of the westernblot indicating a reduction of CTCF when infected with lentiviral particles containing the shCTCF81 construct. Therefore, only this construct was used for further investigation.

In figure 11C, proliferation curves of K562 cells infected with lentiviral particles containing shRNA specific for CTCF (shCTCF) or the pTRIPZ empty vector (EV) are shown. In figure 11C, upper, K562 cells were infected, selected with puromycin for two days and treated with Ara-C or Imatinib for 3 days (no induction). In figure 11C, lower, K562 cells were infected, selected with puromycin for two days and induced with doxycycline for 3 days. Then, the cells were treated with Ara-C and Imatinib for 3 days. K562 cells show a similar growth curve as shown in figure 6B, showing that treatment with Ara-C and Imatinib leads to a reduction of proliferation. In order to receive better insight, cell viability was assessed. In figure 12A, the percentage of viable K562 cells infected with lentiviral particles containing shRNA specific for CTCF (shCTCF) or the pTRIPZ empty vector (EV) are shown. On the left, K562 cells were not induced with doxycycline and therefore, CTCF should not be silenced. It can be seen, that the results are consistent with figure 6C and that the treatment with Imatinib leads to cell death. However, if the K562 cells are induced with doxycycline and CTCF is silenced, less cells are found to be dead when treated with Imatinib and therefore, the results are consistent with figure 8C.

In order to evaluate erythroid differentiation, benzidine test was performed. In figure 12B, the percentage of benzidine positive cells is shown. Figure 12B, left, shows K562 cells infected with lentiviral particles containing pTRIPZ empty vector (EV) without (-) and with (+) induction with doxycycline and treated with Ara-C or Imatinib. It can be seen that there is not a significant difference between induced and non-induced and that the results are consistent with figure 7A. Figure 12B, right, depicts K562 cells infected with lentiviral particles containing shRNA specific for CTCF (shCTCF) without (-) and with (+) induction with doxycycline and treated with Ara-C or Imatinib. It can be seen that silencing of CTCF leads to a reduced amount of benzidine positive cells and therefore, inhibition of erythroid differentiation. Those results are consistent with the results obtained by silencing of CTCF with the constitutive pLKO vector.

Altogether our results show that silencing of CTCF inhibits erythroid differentiation in K562 cells. This effect was demonstrated using two different inducers of differentiation (Ara-C and Imatinib) and two different knockdown systems (constitutive and inducible shCTCF vectors).

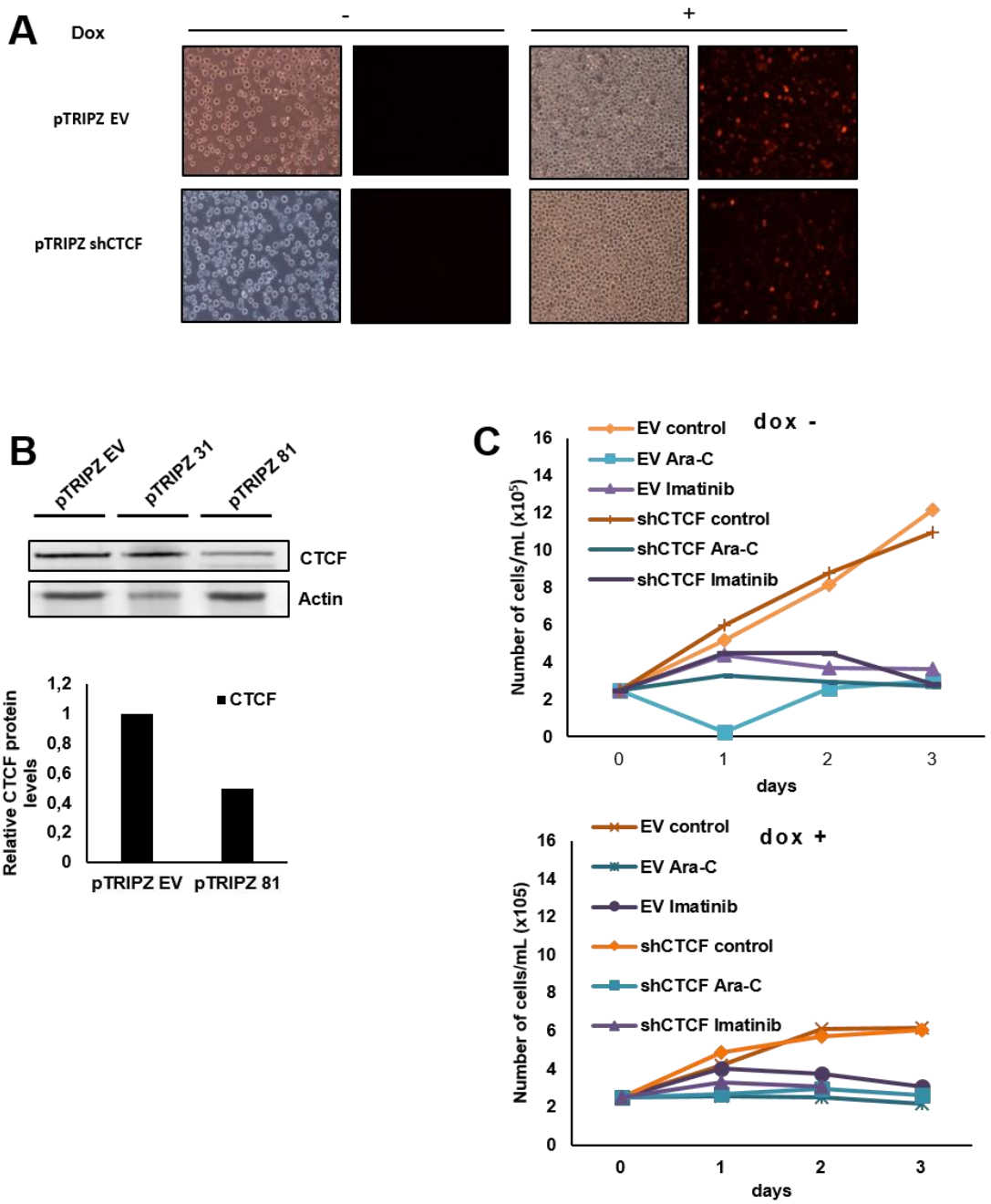

Figure 11. Silencing of CTCF with pTRIPZ-shCTCF lentiviral particles in K562 cells A) Transduction efficiency and reporter expression were checked by microscopy after 3 days with (+) and without (-) doxycycline induction. Bright field pictures show total cell number (left). Red cells indicate expression of TurboRFP reporter in both vectors and shRNA in pTRIPZ shCTCF. B) Silencing of CTCF was evaluated by westernblot. Silencing was achieved with shCTCF 81 but not with shCTCF 31. Relative CTCF protein level of westernblot was quantified using ImageJ. C) Proliferation curves of K562 cells infected with pTRIPZ empty vector (EV) or shRNA specific for CTCF gene (shCTCF) and treatment with 1µM Ara-C or 0.5 µM Imatinib. Upper shows proliferation curve without doxycycline, lower shows proliferation curve upon induction with doxycycline.

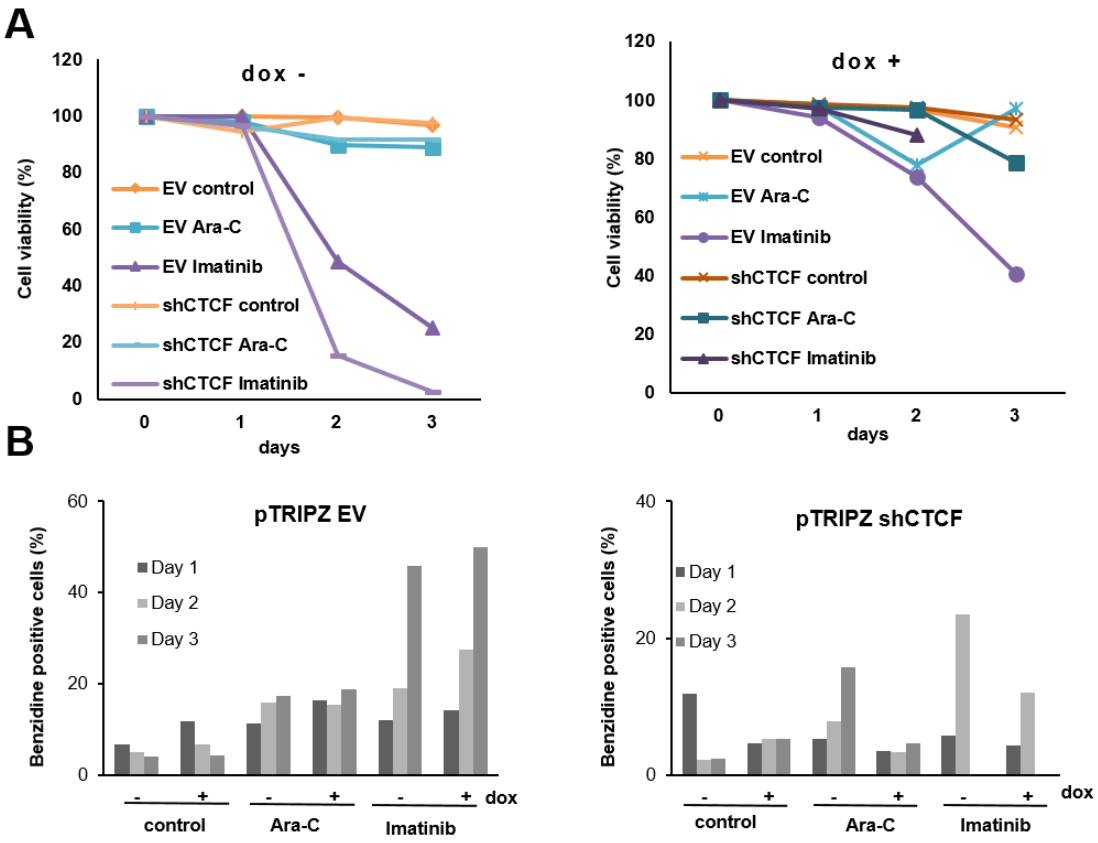

Figure 12. Silencing of CTCF using inducible pTRIPZ reduces benzidine positive cells A) Cell viability (trypan blue assay) of K562 cells infected with lentiviral particles containing specific shRNA against CTCF gene (shCTCF) or the pTRIPZ empty vector (EV). $2.5x10^5$ cells/mL were seeded after 48 hours of puromycin selection. 24 hours correspond to three days of puromycin selection + treatment with 1µM Ara-C or 0.5µM Imatinib for 24 hours. B) Benzidine test of K562 cells infected with lentiviral particles containing specific shRNA against CTCF mRNA (shCTCF) or the pTRIPZ empty vector (EV). $2.5x10^5$ cells/mL were seeded after 48 hours of puromycin selection. 24 hours correspond to three days of puromycin selection + treatment with 1µM Ara-C or 0.5µM Imatinib for 24 hours. Treatment with Ara-C and Imatinib lead to an increase in benzidine positive cells in K562 infected with the empty vector (EV) pTRIPZ independent of doxycycline induction (left). Induction of shCTCF in K562 lead to a decrease in the number of benzidine positive cells.

4.5 Erythropoietin (EPO) induces erythroid differentiation in CD34$^+$ cells

CD34$^+$ cells are pluripotent hematopoietic stem cells with the ability to self-renewal and giving rise to myeloid and lymphoid progenitor cells. Compared to K562 cells, which is a human immortalized myelogenous leukemia line, CD34$^+$ are primary cells and therefore, closer to the physiological state. In order to verify that CTCF plays a pivotal role during erythropoiesis, CD34$^+$ cells were chosen as a model.

CD34$^+$ cells were obtained from cord blood, as shown in figure 13A. Ficoll was used to separate mononuclear cells from the rest of the blood components. From this fraction, CD34$^+$ were purified and induced to differentiation with EPO. As described in the introduction, erythropoietin is an important cytokine primarily acting on erythroid precursors. Two different concentrations of EPO were utilized: 3U/mL and 6U/mL and erythroid differentiation was analyzed using the benzidine test and westernblot.

In figure 13B, the percentage of benzidine positive cells after 5, 7 and 10 days of induction with EPO are shown. It can be seen that the amount of benzidine positive cells increase after treatment with both concentrations of EPO. In figure 13C, erythroid differentiation was analyzed by westernblot showing an increase of γ-globin after 10 days of treatment with EPO. Furthermore, an increase in GATA1 expression is visible indicating erythroid differentiation. Further experiments silencing CTCF in CD34$^+$ cells to analyze the effects of CTCF in this physiological system are in progress.

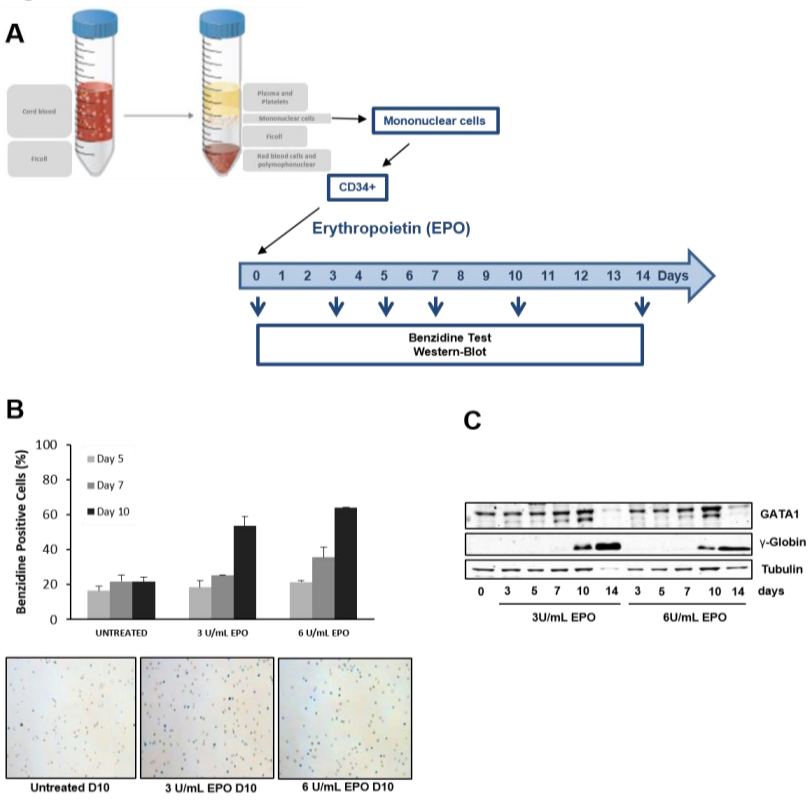

Figure 13. Erythroid differentiation of CD34$^+$ cells induced by EPO. A) Schematic representation of purification of mononuclear cells using ficoll and induction of differentiation with Erythropoietin (EPO). B) Benzidine test of CD34$^+$ cells induced with 3U/mL and 6U/mL EPO lead to an increase of benzidine positive cells compared to control cells. Error bars mean the STD of two independent experiments. C) Erythroid differentiation was evaluated by westernblot analyzing expression of γ–globin and GATA1. Tubulin was used as a loading control.

4.6 Identification of erythroid genes regulated by CTCF

Since it was shown that CTCF has an important function during erythroid differentiation, the next focus lies on the identification of erythroid genes, which may be regulated by CTCF. From previous microarray data we know that a lot of genes are up or down regulated by CTCF. Those data was filtered for genes involved in erythropoiesis (appendix, figure A3) and an ENCODE analysis was carried out. ENCODE Genome Browser contains a broad collection of vertebrate and model organism assemblies and annotations developed by University of California Santa Cruz (UCSC) and the other members of the International Human Genome Project consortium (https://genome.ucsc.edu/). In figure 14 and appendix figure A1, all genes analyzed with ENCODE are shown. In this case, CTCF binding sites were evaluated (shown as blue peaks) and the DNA sequence from the binding site was obtained for primer design. In figure 14, genes analyzed by chromatin immunoprecipitation (ChIP) are shown.

A CTCF binding site was found located upstream of the ETS1 (Erythroblastosis oncogene 1) gene. ETS1 was described to inhibit erythroid differentiation in human hematopoietic progenitor cells (Lulli et al., 2006). Microarray analysis have revealed that CTCF downregulates ETS1. In order to get a better insight, CTCF binding upstream to ETS1 was analyzed before and after induction of differentiation using ChIP.

MYB (Myeloblastosis oncogene) was also found to be downregulated by CTCF. MYB was described to inhibit induced erythroid differentiation in K562 cells (O'Brien, 1995). According to ENCODE Genome Browser, Intron 1 of the MYB gene harbors a CTCF binding site. Primer were designed specifically to this region and ChIP analysis was performed.

LIM domain-only protein 2 (LMO2) is an important transcriptional regulator in erythropoiesis and, as described in the introduction of this report, belongs to the master transcription factors of the core erythroid network. It was found to bind together with GATA1 and SCL to the beta-globin locus control region, recruit RNA Pol II and therefore, ß-globin transcription (Song et al., 2007). Microarray analysis revealed that CTCF upregulates LMO2 expression (appendix, figure A3). A CTCF binding site was localized downstream of the LMO2 gene indicating possible regulation of LMO2 by CTCF. To further investigate in CTCF binding downstream of LMO2 upon differentiation, ChIP analysis was carried out.

Another gene analyzed by ChIP in this report is KLF1 (Erythroid Kruppel-like factor 1). KLF1 was also identified as one of the master transcriptions factors of the core erythroid network. Interestingly, recent studies found that KLF1 plays an important role in controlling apoptotic gene expression to drive the terminal stages of erythroid differentiation, which are important for enucleation (Tallack et al., 2012). According to microarray analysis, CTCF upregulates KLF1, indicating an important role of CTCF during the process of erythropoiesis. ENCODE studies have shown that CTCF binds KLF1 at Exon 2. The green area within the depicted binding site of CTCF indicates that several groups have discovered the same binding site in this area.

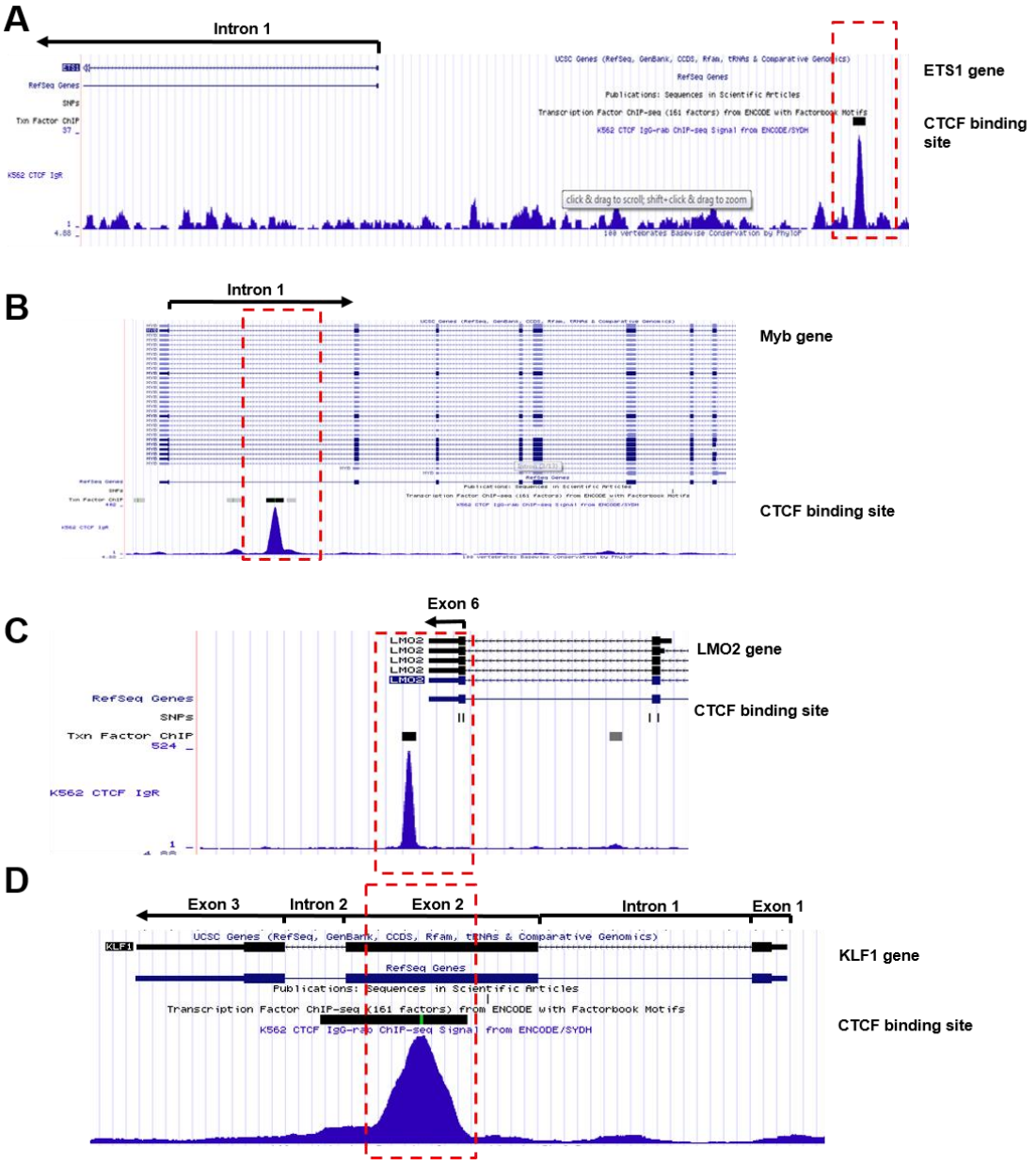

Figure 14. Encode ChIP-seq data (K562 cells) of erythroid genes analyzed by ChIP. A) ChIP-seq profiles of CTCF for ETS1 gene in chromosome 11. CTCF intergenic binding site is shown (red rectangular). B) ChIP-seq profiles of CTCF for Myb gene in chromosome 6. CTCF Intron 1 binding site is shown (red rectangular). C) ChIP-seq profiles of CTCF for LMO2 gene in chromosome 11. CTCF Exon 6 binding site is shown (red rectangular). D) ChIP-seq profiles of CTCF for KLF1 gene in chromosome 19. CTCF Exon 2 binding site is shown (red rectangular).

4.7 Analysis of CTCF binding to erythroid genes upon differentiation

Taking the results from the ENCODE study into account, primers were designed for the CTCF binding sites and ChIP analysis was performed. For immunoprecipitation a mixture of CTCF antibodies was utilized (appendix, table A2).

In figure 15A, positive and negative controls of ChIP analysis in untreated K562 cells are shown. As a positive control, the H42.1 rDNA site, which was identified by our group as a significant CTCF binding site (Van de Nobelen et al., 2010) was used. As a negative control, H4 rDNA site was chosen, since CTCF was reported to do not bind to this site (Van de Nobelen et al., 2010). High chromatin enrichment were detected in the H42.1site, whereas only rare binding was found in the H4 site, indicating that the ChIP works well. Furthermore, the specificity of the pull down with beads was verified, since there is no signal without adding CTCF antibodies (beads).

In figure 15B, the results of ChIP analysis in K562 cells are shown. CTCF binding sites were identified using the ENCODE Genome browser (figure 14) for which primers were designed (appendix, table A4). K562 cells were fixed after 24 and 72 hours and an immunoprecipitation using CTCF antibodies was performed. In figure 15B, it can be seen that all analyzed sites are bound by CTCF. Those results confirm binding sites of CTCF in erythroid genes provided by ENCODE.

In the next step, CTCF binding to erythroid genes upon differentiation was examined. Therefore, K562 cells were treated with 1µM Ara-C for 72 hours and with 0.5µM Imatinib for 48 hours. In order to determine cell differentiation, cell pellets were collected and RNA extracted. In figure 15C and D, results from RT-qPCR are shown. The relative mRNA expression is shown from the genes MYC, ε-globin and glycophorin A (GPA). MYC is a transcription factor driving the cell cycle and should be downregulated upon differentiation. ε-globin is an embryonic globin, which is initially expressed during embryogenesis and then replaced by fetal and adult globins. Glycophorin A is a protein localized in the membrane of erythrocytes. Both, ε-globin and GPA serve as markers indicating differentiation, whereas MYC indicates undifferentiated status. Downregulation of MYC and upregulation of ε-globin and GPA indicate that the K562 cells treated with Ara-C and Imatinib undergo differentiation compared to control untreated cells.

For ChIP analysis, K562 cells treated with Ara-C and with Imatinib. The IP was performed with three different CTCF antibodies (appendix, table A2) and the relative enrichment was quantified by qPCR. Shown is the relative fold enrichment normalized to K562 without treatment. In figure 16A, a CTCF binding site upstream of the ETS1 gene was analyzed. Both treatments lead to an increase in CTCF binding upon differentiation. In figure 16B, the CTCF binding site in Exon 2 of KLF1 gene was inspected. Increase in CTCF binding to this site upon differentiation was also found. When investigating in the Intron 1 binding site within the MYB gene, no significant change in binding upon differentiation could be detected (figure 16C). Furthermore, preliminary results do not allow clear insight into binding behavior of CTCF to the CTS downstream of LMO2 gene (figure 16D). All in all, more replicates are required for clear understanding of CTCF binding behavior upon erythroid differentiation.

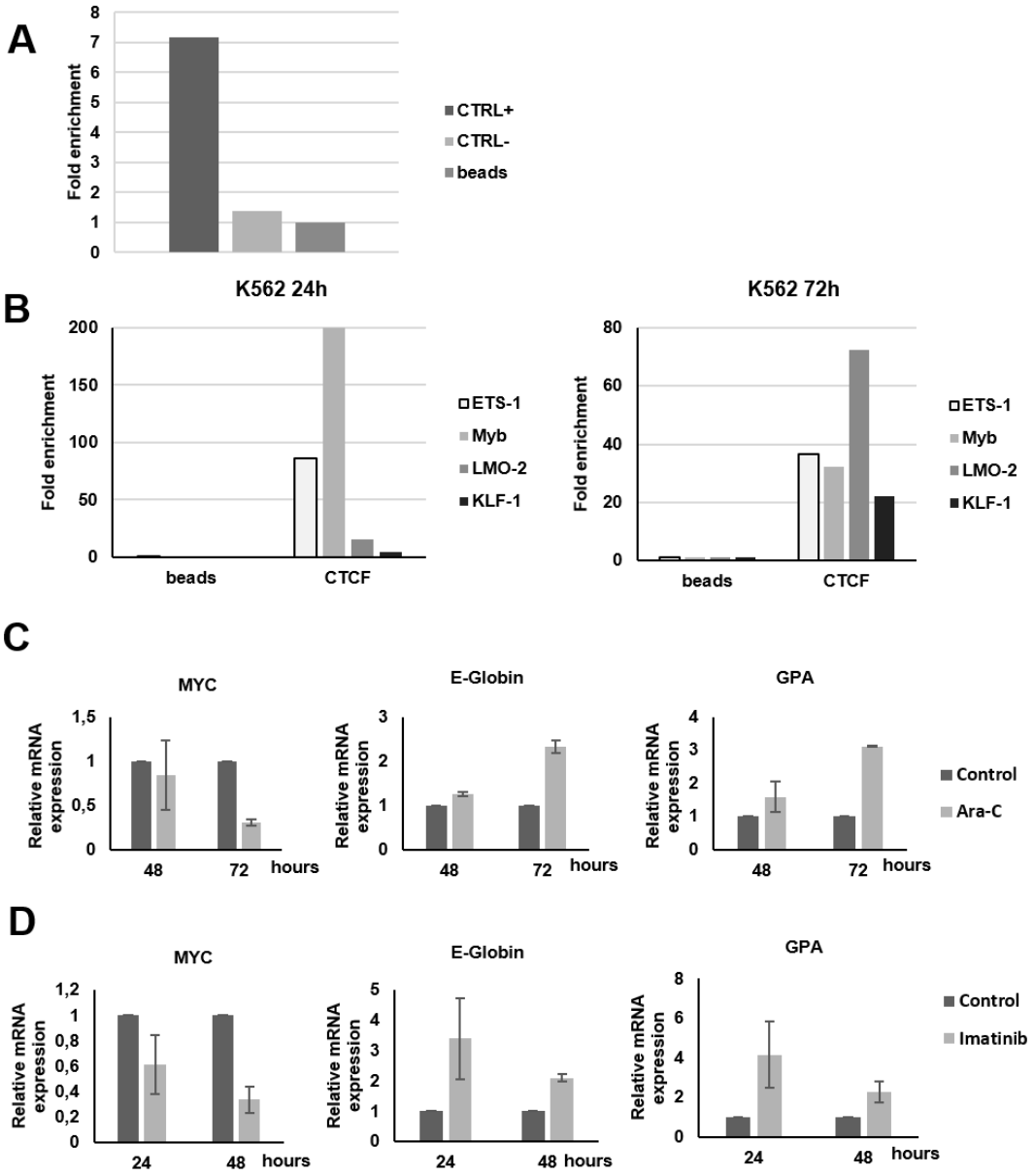

Figure 15. CTCF *in vivo* binding to erythroid genes. A) ChIP *in vivo* binding to well characterized CTCF target site (H42.1 rDNA (CTRL+) and non-target site H4 (CTRL-) in K562 cells. A mixture of three different CTCF-antibodies was used for ChIP. Specificity of beads was verified using no antibodies during IP (beads). Fold enrichment was quantified by qPCR and determined by calculating the ratio of the amount of the target sequence in the immunoprecipitation over the amount of the target sequence in the input DNA. B) Identification of CTCF binding sites within erythroid genes. K562 cells were fixed after 24 hours and 72 hours and analyzed for CTCF binding sites within erythroid genes. C) RT-qPCR of erythroid marker genes after treatment with 1µM Ara-C. Error bars mean the s.e.m. of two independent experiments. D) RT-qPCR of erythroid marker genes after treatment with 0.5µM Imatinib. Error bars mean the s.e.m. of two independent experiments.

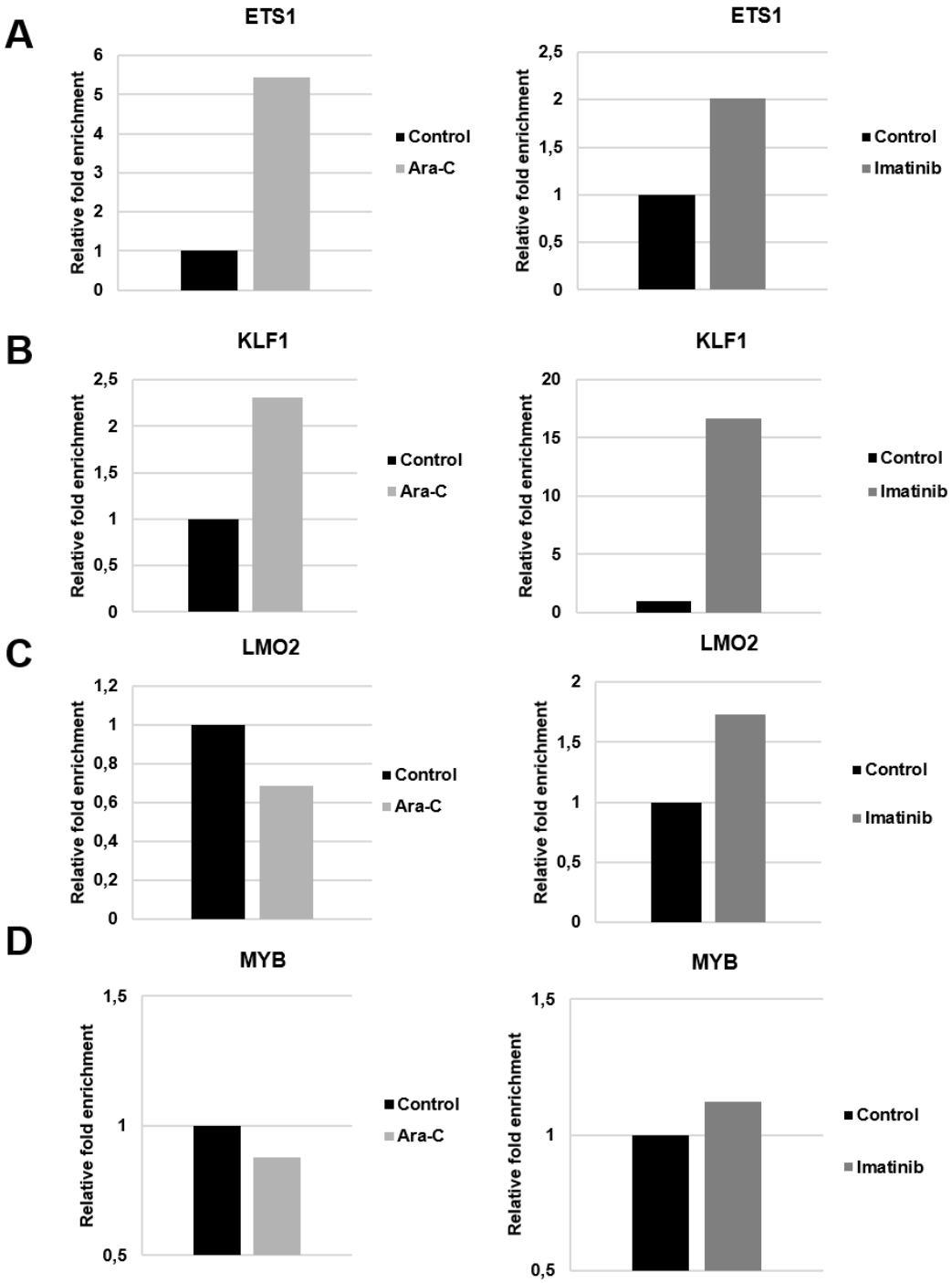

Figure 16. CTCF *in vivo* binding to erythroid genes upon differentiation A) ChIP analysis with three anti-CTCF antibodies showing the CTCF binding to erythroid genes upon treatment with 1µM Ara-C and 0.5µM Imatinib. Relative enrichment was quantified by qPCR. The fold enrichment of a particular target sequence was determined by calculating the ratio of the amount of the target sequence in the immunoprecipitation over the amount of the target sequence in the input DNA. Each value was normalized with respect to the control without antibodies. Shown is the relative fold enrichment normalized to K562 without treatment.

5. DISCUSSION

Role of CTCF in erythroid differentiation

The multipotent K562 cells derive from a chronic myeloid leukemia in blast crisis and retain the ability to differentiate towards the myeloid lineage. This common myeloid progenitor gives rise to the megakaryocyte-erythroid progenitor or the granulocyte-monocyte progenitor, which finally will differentiate into erythrocytes and megakaryocytes or granulocytes and monocyte/macrophage respectively. Our group's previous results have described that K562 cells can be differentiated along the erythroid pathway using Ara-C (Torrano et al., 2005).or Imatinib (Gomez Casares, Oncogene 2013) In this study, it could be confirmed that treatment with Ara-C and Imatinib induces erythroid differentiation in K562 cells. In our hands, due to decreased cell viability upon treatment with 1µM Imatinib, the concentration was reduced to 0.5 µM Imatinib for further experiments.

Additionally, expression levels of CTCF and MYC were analyzed. CTCF seems to be expressed constantly during the maturation process. Downregulation of MYC additionally coincide with reduced proliferation associated with cell differentiation, which was also revealed in previous studies from our lab (Delgado and Leon, 2010). All in all, the results are consistent with previous results indicating that Ara-C and Imatinib induce differentiation of K562.

In order to unravel the role of CTCF during erythroid differentiation, CTCF was silenced using lentiviral particles containing a vector, which constitutively expresses a shRNA specifically to CTCF gene. K562 were infected and treated with Ara-C and Imatinib in order to check erythroid differentiation. It could be demonstrated that silencing of CTCF inhibits erythroid differentiation in K562 cells by analysis of hemoglobin containing cells (benzidine positive) and erythroid genes expression. These data are in agreement with previous results from our group where CTCF overexpression increase the erythroid differentiation induced by Ara-C (Torrano et al., 2005). Interestingly, less cells undergo apoptosis upon Imatinib treatment if CTCF is silenced compared to CTCF expressing cells. In order to get better insight into the `possible effect of CTCF in apoptosis, further experiments are required. Nevertheless, it could be shown that CTCF plays a pivotal role during erythroid differentiation.

In order to verify and extend those results, a different system for CTCF knockdown was established using a doxycycline-inducible promoter for shCTCF expression. The construct was delivered by lentiviral particles and induced 3 days before treatment with Ara-C or Imatinib. Preliminary results of CTCF knockdown with the inducible system revealed similar findings compared to the constitutive expressing system and therefore, confirm that CTCF silencing inhibits erythroid differentiation, demonstrating a role of CTCF during erythroid differentiation. $CD34^+$ primary pluripotent cell line purified from cord blood serve as a second model for studying the importance of CTCF in erythroid differentiation. $CD34^+$ cell have the ability to give rise to all cell types of the blood and therefore, are utilized to reveal the role of CTCF during early stages of erythroid differentiation. Previous studies have shown that treatment with erythropoietin (EPO) induces erythroid differentiation (Jin et al., 2014). In this study, it could be confirmed

that EPO induces commitment to the erythroid pathway of $CD34^+$ cells. Experiments to explore the CTCF effect on these primary cells are in progress in our lab.

CTCF interacting with different erythroid genes

Previous results from our group have identified erythroid genes, which are up or downregulated by CTCF using microarray (appendix, figure A3). These genes were scanned for CTCF binding sites using the ENCODE annotation Genome Browser, which was published by the University of California, Santa Cruz (UCSC). The existence of CTCF binding sites in all of the identified erythroid genes could be unraveled according to ENCODE and also by ChIP during this study. In order to get better insight into the binding events occurring during the process of differentiation, chromatin immunoprecipitation (ChIP) assays were carried out before and after induction of differentiation.

Preliminary result indicate that CTCF binding events change upon induction of differentiation in erythroid genes. Microarray studies from our group have revealed that ETS1 and MYB expression is downregulated by CTCF (appendix, figure A3). In order to get a more detailed view of the occurring molecular events, one CTS per gene was analyzed upon induction of differentiation. The CTS upstream of the ETS1 gene showed increased binding of CTCF upon differentiation with both drug treatments. We hypothesize that binding of CTCF upstream of the ETS1 gene represses its expression, which may be required for the differentiation process. The occupancy of the Intron 1 CTS within MYB gene does not significantly change within the differentiation process and therefore, seems not to have a major role on MYB expression. According to previous findings, MYB is downregulated by CTCF, however we propose that the CTS within Intron 1 does not play a major role in this case. Nevertheless, those results must be verified by additional replicates.

Genes upregulated by CTCF were found to be KLF1 and LMO2 (appendix, figure A3). Both genes were furthermore identified to be part of the CEN, as described in the introduction (1.2.3.). CTCF binding to Exon 2 within KLF1 gene was found to change upon differentiation. Both treatments resulted in increased binding of CTCF to KLF1/Exon 2 indicating a possible role of CTCF on regulation of KLF1 expression. Furthermore, a binding site of CTCF downstream of LMO2 gene was scrutinized. The results do not indicate a clear change of binding upon induction of differentiation and therefore, we propose that this binding site does not have a major effect on LMO2 expression.

To sum up, it should be point out to the diverse functions of CTCF and that nearby binding to a gene does not necessary indicate its regulation. Furthermore, this also means that CTCF bound far apart from erythroid genes may have a regulatory influence by forming loops. In this study, binding sites upstream, in coding and non-coding regions and downstream of a genes were analyzed. We found that binding sites upstream and within the coding sequence may have an influence of the nearby genes, however binding sites within the non-coding area or downstream of a gene seem to have no influence on the nearby gene. In order to confirm those results, more binding sites of CTCF need to be analyzed.

6. CONCLUSION

- Treatment with Ara-C and Imatinib reduce proliferation and induce differentiation along the erythroid pathway in K562 cells.
- CTCF constitutive downregulation inhibits erythroid differentiation induced by Ara-C or Imatinib
- Inducible downregulation of CTCF confirms inhibition of erythroid differentiation.
- $CD34^+$ primary cell differentiate along the erythroid pathway upon treatment with EPO.
- CTCF *in vivo* binding to erythroid genes (ETS1, MYB, LMO-2 and KLF-1) is demonstrated
- Differential CTCF occupancy of erythroid genes is found upon erythroid differentiation

7. FUTURE WORK

- Silencing of CTCF at different time points during differentiation to obtain better insight into molecular events in K562 using the inducible system pTRIPZ.
- Silencing of CTCF using the inducible system pTRIPZ to assess its effect on $CD34^+$ erythroid differentiation.
- ChIP analysis upon downregulation of CTCF to identify crucial binding sites.
- ChIP analysis of all erythroid genes analyzed by ENCODE Genome Browser.

8. REFERENCES

Batlle-Lopez, A., Cortiguera, M.G., Rosa-Garrido, M., Blanco, R., del Cerro, E., Torrano, V., Wagner, S.D., and 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, 246-256.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

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

Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS One 7, e41374.

De La Rosa-Velazquez, I.A., Rincon-Arano, H., Benitez-Bribiesca, L., and Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. Cancer Res 67, 2577-2585.

Delgado, M.D., and Leon, J. (2010). Myc roles in hematopoiesis and leukemia. Genes Cancer 1, 605-616.

Engel, N., Thorvaldsen, J.L., and Bartolomei, M.S. (2006). CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. Hum Mol Genet 15, 2945-2954.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and 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. Mol Cell Biol 16, 2802-2813.

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., et al. (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer Res 62, 48-52.

Jin, H., Kim, H.S., Kim, S., and Kim, H.O. (2014). Erythropoietic potential of CD34+ hematopoietic stem cells from human cord blood and G-CSF-mobilized peripheral blood. Biomed Res Int 2014, 435215.

Kim, S., Yu, N.K., and Kaang, B.K. (2015). CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med 47, e166.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128, 1231-1245.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol 13, 1196-1204.

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

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep 3, 1678-1689.

Nandakumar, S.K., Ulirsch, J.C., and Sankaran, V.G. (2016). Advances in understanding erythropoiesis: evolving perspectives. Br J Haematol 173, 206-218.

O'Brien, D.R.a.T.G. (1995). Constitutive c-myb expression in K562 cells inhibits induced erythroid differentiation but not tetradecanoyl phorbol acetate-induced megakaryocytic differentiation. Molecular and Cellular Biology 15, 772–779.

Ohlsson, R., Bartkuhn, M., and Renkawitz, R. (2010). CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. Chromosoma 119, 351-360.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15, 234-246.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132, 631-644.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28, 810-822.

Tallack, M.R., Magor, G.W., Dartigues, B., Sun, L., Huang, S., Fittock, J.M., Fry, S.V., Glazov, E.A., Bailey, T.L., and Perkins, A.C. (2012). Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. Genome Res 22, 2385-2398.

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., Leon, J., Klenova, E., and Delgado, M.D. (2005). CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. J Biol Chem 280, 28152-28161.

van de Nobelen, S., Rosa-Garrido, M., Leers, J., Heath, H., Soochit, W., Joosen, L., Jonkers, I., Demmers, J., van der Reijden, M., Torrano, V., et al. (2010). CTCF regulates the local epigenetic state of ribosomal DNA repeats. Epigenetics Chromatin 3, 19.

9. APPENDIX

Primers used in this study

Table A1: Primers for RT-qPCR

Amplified gene (human)	Primer Sequence (5'→3')	Amplicon Size
S14	Fw: TATCACCGCCCTACACATCA Rv: GGGGTGACATCCTCAATCC	135 bp
MYC	Fw: TCGGATTCTCTGCTCTCCTC Rv: CCTGCCTCTTTTCCACAGAA	157 bp
MYB	Fw: AGCAAGGTGCATGATCGTC Rv: GGGGGTGGAAGTTAAAGAAGG	157 bp
E-Globin	Fw: GCAAGAAGGTGCTGACTTCC Rv: TGCCAAAGTGAGTAGCCAGA	168 bp
GPA	Fw: GAGAAAGGGTACAACTTGCC Rv: CATTGATCACTTGTCTCTGG	220 bp

Primary antibodies used in this study:

Table A2: Primary antibodies

Antibody	Immunogen	Type	Origin	Use and Dilution
CTCF	Human CTCF aa. 184-290	Mouse monoclonal	BD Transduction Laboratories™	WB (1:1000) and ChIP
CTCF	synthetic peptide corresponding to amino acids 659-675 of human CTCF	Rabbit polyclonal	EMD Millipore/Merck	ChIP
CTCF	Synthetic peptide, corresponding to a region within amino acids 650-700 of Human CTCF	Rabbit polyclonal	ABCAM	ChIP
MYC	raised against amino-terminal residues of c-Myc	Rabbit polyclonal	Cell Signaling Technology® 9402s	WB (1:3000)
GATA1 C20	C-terminus of GATA1 of human origin	Goat polyclonal	Santa Cruz Biotechnology sc-1233	WB (1:1000)
γ-Globin	raised against human Hemoglobin	Mouse monoclonal	Santa Cruz Biotechnology sc-21756	WB (1:1000)
LMO2	Synthetic peptide corresponding to a region within N terminal residues 1-100 of human LMO2	Rabbit polyclonal	ABCAM ab72841	WB (1:2000)
PARP1 (H-250)	raised against aa764-1014 of C-terminal residues (human origin)	Rabbit polyclonal	Santa Cruz Biotechnology sc-7150	WB (1:1000)
B-Actin (C4)	raised against gizzard Actin of chicken origin	Mouse monoclonal	Santa Cruz Biotechnology sc-47778	WB (1:1000)
Tubulin	-	Rabbit polyclonal	Kindly donated from the laboratories of Nick Cowan, NY, USA	WB (1:3000)

Secondary antibodies used in this study:

Table A3: Secondary antibodies

Antibody	Immunogen	Type	Origin	Use and Dilution
Anti-GoatIRDye®800	goat heavy and light IgG	Donkey polyclonal	LI-COR (926-32214)	WB (1:10000)
Anti-Mouse IRDye®800	mouse heavy and light IgG	Donkey polyclonal	LI-COR (926-32212)	WB (1:10000)
Anti-Mouse IRDye®680	mouse heavy and light IgG	Donkey polyclonal	LI-COR (926-68072)	WB (1:10000)
Anti-RabbitIRDye®800	rabbit heavy and light IgG	Donkey polyclonal	LI-COR (926-32213)	WB (1:10000)
Anti-RabbitIRDye®680	rabbit heavy and light IgG	Donkey polyclonal	LI-COR (926-68073)	WB (1:10000)

The following primers were designed for ChIP:

Table A4: Primers for ChIP

Amplified gene (human)	Location	Primer Sequence (5'→3')	Amplicon Size
ETS1	upstream	Fw: GAGGTCCTTCCTCCTGGAAC Rv: ATGCAGCTATTGGGTTTTGC	184 bp
MYB	Intron 1	Fw: TCCAAGCAAGCCCTTATTGT Rv: ACAACCCAGGAACAAGCAAC	198 bp
LMO2	downstream	Fw: TTAAGGTGATGGCCAGAAGG Rv: TTTTCCAAGACGGGTGTCTC	162 bp
KLF1	Exon 2	Fw: GGTGGGAGCTCTTGGTGTAG Rv: CCCCTCCTTCCTGAGTTGTT	191 bp
H42.1	rDNA repeats	Fw: 5´GCTTCTCGACTCACGGTTTC Rv: CCGAGAGCACGATCTCAAA	124 bp
H4	rDNA repeats	Fw: CGACGACCCATTCGAACGTCT Rv: CTCTCCGGAATCGAACCCTGA	103 bp
GATA2	upstream	Fw: TGCTTTGTCACTGCTGTTCC Rv: AAATTCAGTGGGATGCGTTC	198 bp
GATA2	Intron 4	Fw: CCAAAGCAGGGAACGATTTA Rv: TTTCCTATCCGGACATCTGC	226 bp
HES1	Exon 4	Fw: ATCAATGCCATGACCTACCC Rv: AGCCTCCAAACACCTTAGCC	208 bp
HEY1	Exon 5	Fw: GCCACTGAGGAGAGCAGAG Rv: GACCGTCTTCGGACATCAC	234 bp
NFE2L2	Upstream	Fw: CCTAGGGGAGATGTGGACAG Rv: AGAGGTTCTCTTGGGGTTCC	195 bp
NFE2L2	Intron 1	Fw: GCAACAGATCAACAGCTCCA Rv: CACCCGGGGCTTCTAGTT	185 bp
TCF3	Intron 15	Fw: CTCACCCTCAAGTCCTGTCC Rv: ATTTGGCCATGAGAAAGGTG	238 bp

TCF3	Intron 6	Fw: ACAGGGACAGATCCAACCAC Rv: AGGCTGGATTGGGATTGAG	155 bp

Table A5: Buffer used for ChIP

Lysis Buffer	0.7% SDS 2mM EDTA 20mM Tris-HCl pH 8 150mM NaCl
Dilution Buffer	0.01% SDS 1% Triton X-100 1.2mM EDTA 20mM Tris-HCL pH 8 150mM NaCl
Low Salt wash Buffer	0.1% SDS 1% Triton X-100 2mM EDTA 20mM Tris HCl pH 8 150mM NaCl
High Salt wash Buffer	0.1% SDS 1% Triton X-100 2mM EDTA 20mM Tris HCl pH 8 500mM NaCl
LiCl wash Buffer	0.25 M LiCl 1% NP 40 (IGEPAL) 1% sodium deoxycholate 1mM EDTA 10mM Tris-HCl pH8
TE Buffer	10mM Tis-HCl pH 7.5 1mM EDTA
Elution Buffer	1% SDS 50mM Tris pH 7.5 10mM EDTA

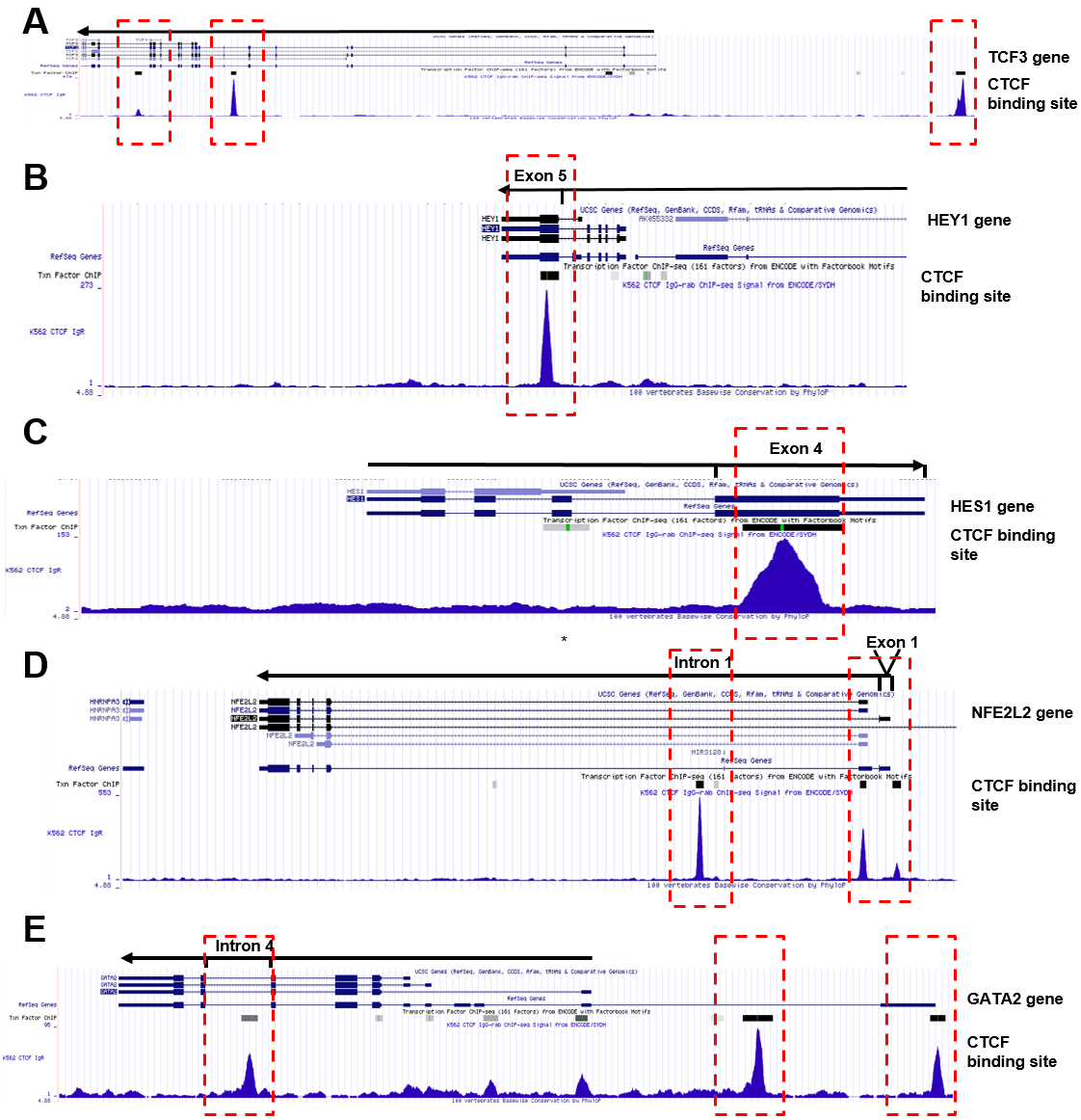

Figure A1. Encode ChIP-seq data (K562 cells) of erythroid genes. A) ChIP-seq profiles of CTCF for TCF3 gene in chromosome 19. CTCF Intron 15, Intron 6 and intergenic binding sites are shown (red rectangulars). B) ChIP-seq profiles of CTCF for HEY1 gene in chromosome 8. CTCF Exon 5 binding site is shown (red rectangular). C) ChIP-seq profiles of CTCF for HES1 gene in chromosome 3. CTCF Exon 4 binding site is shown (red rectangular). D) ChIP-seq profiles of CTCF for NFE2L2 gene in chromosome 2. CTCF Exon 1 and Intron 1 binding sites are shown (red rectangulars). E) ChIP-seq profiles of CTCF for GATA2 gene in chromosome 3. CTCF Intron 4 and intergenic binding sites are shown (red rectangulars).

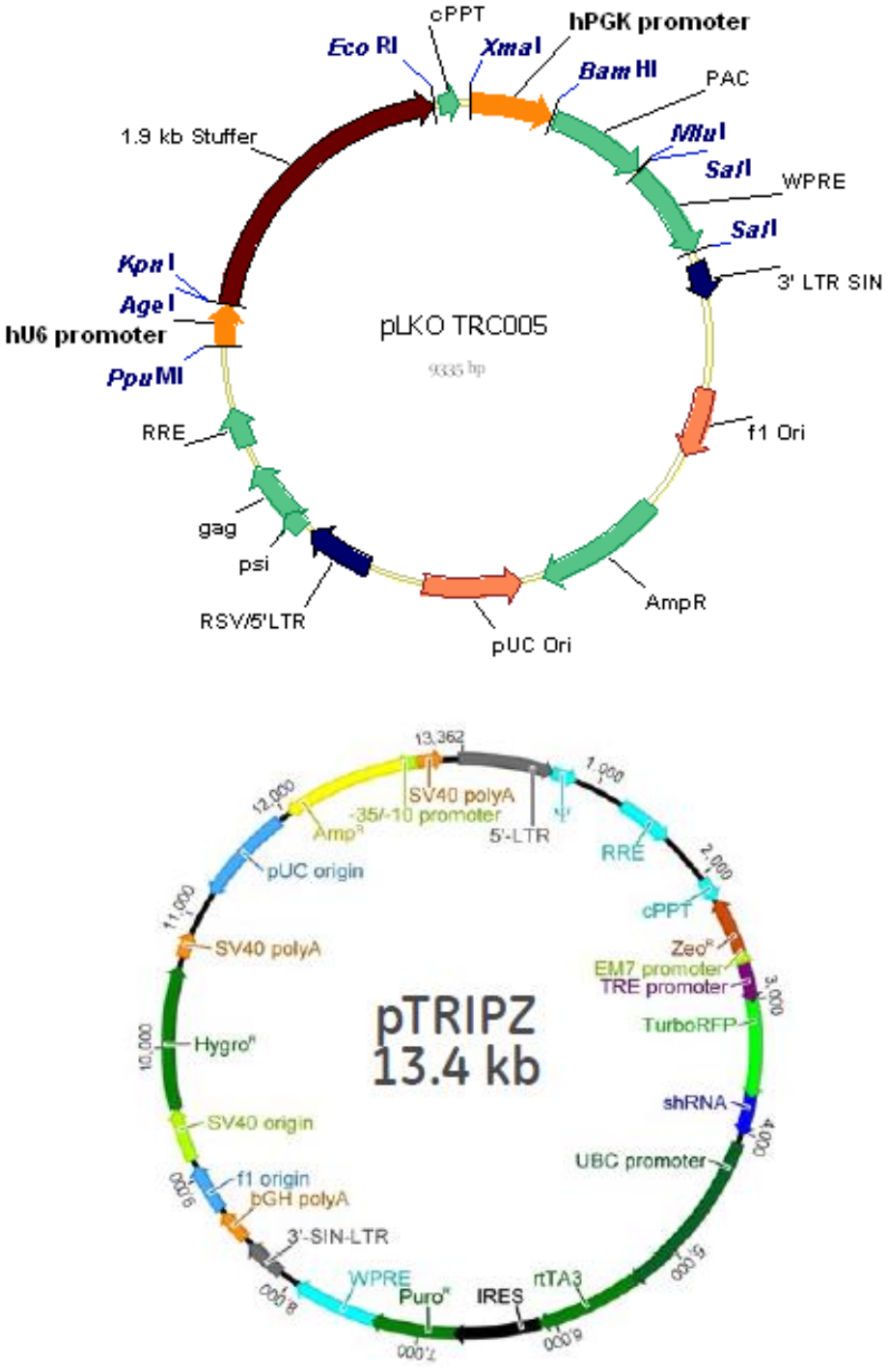

Figure A2: Vector maps of constitutive plasmid pLKO (upper) and inducible plasmid pTRIPZ (lower)

Erythroid-related genes upregulated by CTCF expression

Gene	Description	Function
GYPA	Glycophorin A	Erythroid-specific membrane protein
GYPE	Glycophorin E	Erythroid-specific membrane protein
ANK1	Ankirin 1, erythrocytic	Erythroid-specific membrane protein
SPTA1	Spectrin alpha, erythrocytic 1	Erythrocyte cytoskeletal protein
HBA2	Hemoglobin, alpha 2, adult chain 2	Hemoglobin component
HBA1	Hemoglobin, alpha 1, adult chain 1	Hemoglobin component
HBB	Hemoglobin, beta adult chain	Hemoglobin component
HBD	Hemoglobin, delta	Hemoglobin component
HBZ	Hemoglobin,zeta	Hemoglobin component
HBE1	Hemoglobin, epsilon1	Hemoglobin component
HEMGN	Hemogen	Hematopoietic nuclear protein (regulated by GATA1)
ID1	Inhibitor of DNA binding 1	Transcription factor. Erythroid function?
ID3	Inhibitor of DNA binding 3	Transcription factor 3 (its overexpression in K562 induces erythroid differentiation)
LMO2	LIM domain2 (rhombotin-like 1)	Transcription factor (TAL-1/SCL partner)
TCF3	Transcription factor (E2A)	Transcription factor (TAL-1/SCL partner. Coordinates erythopoietic differentiation)
NF-E2-like 2	Nuclear factor erythroid like 2	Transcription factor (binds to beta-globin promoter at NF-E2 site)
EKLF	Erythroid Kruppel-like factor 1	Transcription factor (erythrocytic; activates beta-globin gene)

Erythroid-related genes downregulated by CTCF expression

Gene	Description	Function
CD44	Cluster of differentiation 44	Transmembrane glycoprotein (leukocytes and erythrocytes)
CD164	Cluster of differentiation 164	Transmembrane glycoprotein (potent negative regulator of erythroid differentiation)
MYB	Myeloblastosis oncogene	Transcription factor (inhibitor of erythroid differentiation)
HES 1	Hairy and enhancer of split 1	Transcription factor (delay erythroid differentiation by inhibition of GATA1)
HEY1	Hairy related transcription factor 1	Transcription factor (inhibitor of erythroid differentiation, c-Jun pathway)
ID2	Inhibitor of DNA binding 2	Transcription factor (inhibited during erythroid differentiation)
GATA 2	GATA binding protein 2	Transcription factor (inhibits erythroid differ.; Its downregulation is required for erythroid differentiation)
ETS1	Erythroblastosis oncogene 1	Transcription factor (blocks erythroid differentiation in human hematopoietic progenitor cells)
STAT5B	Signal transducer and activator of transcription B	Transcription factor (antiapoptotic effect over erythroid cells)
VIM	Vimentin	Intermediate filament (downregulated in erythroid differentiated K562 cells)

Figure A3: Erythroid genes regulated by CTCF