Charles University
Faculty of Science

Study programme: Special Chemical and Biological Programmes
Branch of study: Molecular Biology and Biochemistry of Organisms

Linda Pokorná

Multifunkční protein CTCF a jeho role v regulaci genové exprese
Multifunctional protein CTCF and its role in regulation of gene expression

Bachelor´s thesis

Supervisor: Ing. Tomáš Vacík, Ph.D.

Praha, 2017

Chcem poďakovať svojmu školiteľovi Ing. Tomášovi Vacíkovi, Ph.D. za jeho obrovskú pomoc a cenné rady, ktoré mi poskytoval pri písaní bakalárskej práce. Ďalej chcem poďakovať svojej rodine a kamarátkam za veľkú podporu.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 12.05.2017

Podpis

Abstrakt

CTCF je všudypřítomně exprimovaný jaderný protein, který se váže na DNA skrze svojí centrální doménu zinkových prstů. Tisíce CTCF vazebných míst bylo identifikováno v genomu na promotorech genů, v mezigénových oblastech a v nekódujících sekvencích. CTCF může fungovat jako pozitivní tak i jako negativní regulátor genové exprese a také se podílí na vytváření a udržovaní chromozómových interakcí na velkou vzdálenost. U různých vývojově důležitých genů se ukázalo, že jsou regulované CTCF proteinem a poruchy funkce CTCF jsou asociovány s vývojovými vadami nebo nemocemi. CTCF je různě posttranslačně modifikován například fosforylací nebo SUMOylací, což ovlivňuje jeho funkci v regulaci genové exprese.

Klíčová slova: CTCF, trojdimenzionální genom, kohesin, regulace genové exprese, HOX geny

Abstract

CTCF is a ubiquitously expressed nuclear protein that binds to DNA through its central zinc finger domain. Thousands of CTCF binding sites have been identified throughout the human genome at gene promoters, in intergenic regions or in non-coding sequences. CTCF can function either as a positive or as a negative regulator of gene expression and is also involved in creating and maintaining long-range chromosomal interactions. Various developmentally important genes have been shown to be regulated by CTCF and its malfunction is frequently associated with developmental defects or diseases. CTCF undergoes various posttranslational modifications such as phosphorylation or SUMOylation which also affect its function in the regulation of gene expression.

Keywords: CTCF, three dimensional genome, cohesin, regulation of gene expression, insulation, HOX genes

List of shortcuts

ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CK2	Caseine kinase 2
CTCF	CCCTC binding factor
CTCFL	CTCF-like protein
ELK1	ETS domain-containing protein
FIRE	Frequently interacting region
H3K20me	Methylation of lysine20 of histone H3
H3K27ac	Acetylation of lysine27 of histone H3
H3K27me	Methylation of lysine27 of histone H3
H3K36me	Methylation of lysine36 of histone H3
H3K4me	Methylation of lysine4 of histone H3
H3K9ac	Acetylation of lysine9 of histone H3
H3K9me	Methylation of lysine9 of histone H3
HOX	Homeotic box
IDH	Isocitrate dehydrogenase
IPO13	Importin-13
LAD	Lamina associating domain
mRNA	Messenger RNA
MAX	Myc-associated factor X
P300	Protein 300
PARylation	Poly(ADP-ribosyl)ation
POZ domain	Pox virus and zinc finger domain
PUMA	P53 upregulated modulator of apoptosis
rDNA	Ribosomal DNA

RNAPII	RNA polymerase II
rRNA	Ribosomal RNA
RXRA	Retinoid X receptor alpha
SENP	SUMO-specific isopeptidase
SNP	Single nucleotide polymorphism
SOX2	Sex determining region Y-box 2
Su(Hw)	Suppressor of hairy wing
SUMO	Small ubiquitin-like modifier
TAD	Topologically associating domain
TOP2B	DNA topoisomerase 2-beta
YY1	Ying Yang 1
ZNF143	Zinc finger protein 143
Zw5	Zeste-white 5

Table of Contents

1. Introduction

The first reports about CTCF are almost 30 years old. CTCF was originally characterized in chicken embryonic blood cells as a DNA binding nuclear protein. These first scientific experiments showed that CTCF has a negative effect on the expression of the gene encoding MYC protein. To act as a negative regulator of the *C-MYC* promoter, CTCF binds to this promoter at the CCCTC repeats. Although CTCF was first isolated from embryonic blood cells, it was also detected in various other cell types (Lobanenkov et al., 1990).

Following the discovery of CTCF, a huge number of scientific studies have focused on this protein and have shown its potential in regulation of gene expression. It has been shown that CTCF is able to regulate not only the *C-MYC* promoter, but also a lot of other genes. CTCF can act as a positive transcriptional regulator by forming a heterocomplex with RNAPII, which then binds to the DNA regions to be transcribed (Chernukhin et al., 2007). The DNA regions that are regulated by CTCF have been identified throughout the whole human genome (Kim et al., 2007). CTCF has been also found to modulate the three dimensional architecture of chromatin. CTCF, together with other architectural proteins, forms chromatin loops and maintains these higher order domain structures around the nuclear center or around the nuclear lamina. These higher order chromatin structures are important for long-distance inter- or intra-chromosomal interactions and for insulator activity (Reviewed in Bouwman and de Laat, 2015).

Nuclear multifunctional CTCF has an impact on an extremely large amount of genetic elements throughout the whole genome. This work will summarize the various very important functions of CTCF.

2. CTCF protein structure

The single CTCF polypeptide chain consists of the N-terminal, central and C-terminal parts (Fig.1). The central part contains 11 zinc finger motifs, 10 of which belong to the Cys2His2 class, whereas the eleventh zinc finger belongs to the Cys2HisCys class (Klenova et al., 1993). CTCF zinc fingers 4 to 7 are essential for binding of CTCF to DNA and for recognizing the CTCF binding sites. The CTCF-DNA complex is stabilized by CTCF zinc fingers 4 to 7 much more efficiently than by CTCF

zinc fingers 1 to 3 and 8 to 11 (Nakahashi et al., 2013; Renda et al., 2007). The DNA strands are asymmetrically bent when CTCF is bound to them (Liu and Heermann, 2015).

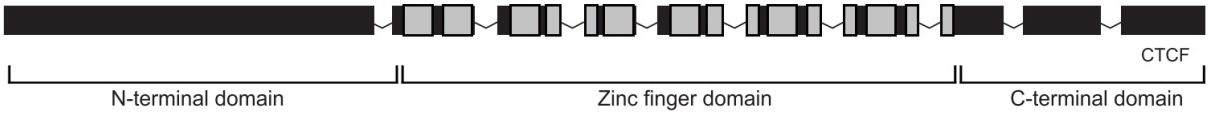

Figure 1. The human CTCF protein structure: The N- and C-terminal domains do not have any secondary structures or structural motifs (Martinez and Miranda, 2010), whereas the central domain consists of 11 zinc finger motifs (Adapted from Hore et al., 2008).

In various organisms various combinations of CTCF zinc finger motifs are used for ideal binding to DNA. For example, for optimal binding to the *P2 C-MYC* promoter region the mammalian CTCF protein needs its zinc finger motifs 3 to 11, whereas the chicken CTCF protein needs its zinc finger motifs 2 to 7 (Filippova et al., 1996).

Almost one half of the identified CTCF binding sites lies in the intergenic regions. Most of the CTCF binding sites are conserved among various cell types and organisms and a single gene locus can contain multiple CTCF binding sites. The consensus CTCF binding site consists of an approximately 20 bp long GC rich core motif and some 3' and 5' sequences around it (Kim et al., 2007). The sequences outside of the core motif have the ability to weaken or to stimulate the binding of CTCF to DNA. The sequence stimulating the affinity of CTCF to DNA lies upstream from the core motif, whereas the sequence weakening the affinity lies downstream (Nakahashi et al., 2013). The large amount of the CTCF binding sites was created mainly by the tandem gene duplication (Liao and Chang, 2014) or by nucleotide point mutations and little of them were created through the insertion of transposable elements (Schmidt et al., 2012).

A single nucleotide polymorphism (SNP) in a CTCF binding motif results in a different affinity of CTCF binding. For example, when the 14th guanine of the core sequence is changed to a cytosine, the CTCF binding affinity to DNA is much lower (Tang et al., 2015). Binding of CTCF to its target sites located between exons of the same gene could lead to activation of an alternative promoter and to transcription of an alternative mRNA isoform. This phenomenon has been described for a number of genes such as the *TRB*, *PCDHG*, or *WNT5A* gene (Kim et al., 2007; Vaidya et al.,

2016). It has been shown that the usage of the alternative promoter of the *WNT5A* gene is inhibited in osteogenic sarcoma by methylation of the CTCF binding sites lying in the first intron around the 1β exon. When these CTCF binding site are methylated, CTCF cannot bind here and cannot activate the alternative promoter. Therefore, the osteogenic sarcoma tumor cells produce the longer *WNT5A* transcript controlled by the first promoter rather than the shorter *WNT5A* transcript whose promoter is not active in these cells. On the other hand, in healthy bone cells the level of the shorter *WNT5A* mRNA is higher than that of the longer *WNT5A* mRNA because the CTCF sites are not methylated and the alternative *WNT5A* promoter is active. The experiments also revealed that the demethylating agents can reactivate the usage of the alternative promoter (Vaidya et al., 2016).

CTCF is well known as a nuclear protein. The transport of a newly synthesized CTCF polypeptide from the cytoplasm into the nucleus is mediated by the IPO13 (Importin-13) transporter. Interestingly, the transport of CTCF into the nucleus is positively regulated by CTCF itself, since CTCF can activate the *IPO13* promoter through a CTCF binding site. In the breast tumor cells the *IPO13* gene is downregulated and CTCF is therefore not transported into the nucleus and accumulates in the cytoplasm (Wang et al., 2013).

3. Posttranslational modifications of CTCF

Proteins are frequently modified by various posttranslational modifications which affect the cellular functions of the modified proteins. CTCF has been shown to undergo four different posttranslational modifications: phosphorylation (Klenova et al., 2001), poly(ADP-ribosyl)ation (Yu et al., 2004), SUMOylation (MacPherson et al., 2009) and ubiquitination (Qi et al., 2012).

3.1 Phosphorylation

CTCF is phosphorylated by CK2 (Caseine kinase) at the C-terminus containing the CK2 consensus sequence (S/T)XX(E/D) (Reviewed in Allende and Allende, 1995; Klenova et al., 2001). The phosphorylated protein sequence by CK2 in the chicken CTCF is S^{604}KKD$S^{600}$$S^{610}DS^{612}$E (Klenova et al., 2001). Although most of the phosphorylated amino acid residues at the CTCF C-terminus are serines, phosphorylation of tyrosine has been also detected in the capacitated male sperms

(Tang and Chen, 2006) or in the *Xenopus laevis* CTCF (Les J. Burke et al., 2001).

Phosphorylation of CTCF stimulates the transcriptional activator properties of CTCF. The *C-MYC* gene is one of the genes whose transcription is stimulated by CTCF depending on the CTCF phosphorylation status. The authors also suggest that an abnormal phosphorylation status of CTCF that is caused by misregulated CK2 may lead to upregulation of the *C-MYC* gene and to the cancer development (El-Kady and Klenova, 2005).

3.2 Poly(ADP-ribosyl)ation

The CTCF N-terminus undergoes Poly(ADP-ribosyl)ation, which is also called PARylation. One of the functions of the CTCF PARylation is to target CTCF to the nucleolus where CTCF represses transcription of rRNA (Guerrero and Maggert, 2011; Yu et al., 2004). Another function of the CTCF PARylation is to target CTCF to the sites of damaged DNA. The PARylated CTCF binds to the damaged sites through its zinc finger domain, mainly through its zinc finger motifs number 4, 5 and 6. Then, the DNA repairing complex is recruited to these CTCF bound DNA damaged sites (Han et al., 2017; Venkatraman and Klenova, 2015). PARylation of CTCF also stimulates the enhancer-blocking properties of CTCF (see 5.2.2). It has been shown that although an unPARylated CTCF is still bound to boundary elements, the enhancer blocking activity of this unPARylated CTCF is not as strong as that of PARylated CTCF and the nearby genes are transcriptional active (Farrar et al., 2010).

3.3 SUMOylation

CTCF contains two sites for SUMOylation, one of them lies at the N-terminus and the other one lies at the C-terminus. In mouse, lysines Lys^{74} at the N-terminal domain of the mouse CTCF and Lys^{698} at the C-terminal domain of the mouse CTCF are SUMOylated by SUMO (Small ubiquitin-like modifier) 1, SUMO 2 or SUMO 3 proteins. These SUMO proteins form an isopeptide bond between their C-terminal diglycine residues and the CTCF lysine residues. SUMOylation of CTCF is known to stimulate the repressive effect of CTCF on the *C-MYC* gene (MacPherson et al., 2009).

The CTCF SUMOylation level is affected by various stress factors such as UV irradiation or hypoxic conditions. Cells exposed to such stress situations show lower levels of SUMOylated CTCF due to the overexpression of SENP (SUMO-specific

isopeptidase) 1 which hydrolyzes the isopeptide bond between SUMO and CTCF. De-SUMOylation of CTCF results in transcriptional downregulation of various genes such as the *PAX6* gene (Wang et al., 2012).

3.4 Ubiquitination

Ubiquitination of the CTCF central zinc finger domain is catalyzed by the HUWE1 E3 ubiquitin-protein ligase and is responsible for proteasomal degradation of CTCF. The extremely low levels of CTCF in the B-cell lymphomas is caused by elevated levels of HUWE1 and the subsequent proteosomal degradation of CTCF. The reduced CTCF levels in the B-cell lymphomas result in activation of the *C-MYC* gene which consequently contributes to the growth of the B-cell lymphoma (Qi et al., 2012).

4. *CTCF* gene

The human *CTCF* gene is located on chromosome 16, q22.1 (Fig.2). The whole locus is approximately 76 kb long and it gives rise to two *CTCF* mRNA isoforms. The first isoform contains 10 exons and 9 introns, whereas the second isoform contains 12 exons and 11 introns. In different organisms, the *CTCF* gene is localized on different chromosomes and has different exon structures (Internet source 1). For example, the approximately 46.4 kb long mouse *Ctcf* gene is localized on chromosome 8 and its exon structure is the same as that of the human *CTCF* gene (Internet source 2), whereas the approximately 30 kb long chicken *CTCF* gene is localized on chromosome 11 and has 11 exons and 10 introns (Internet source 3).

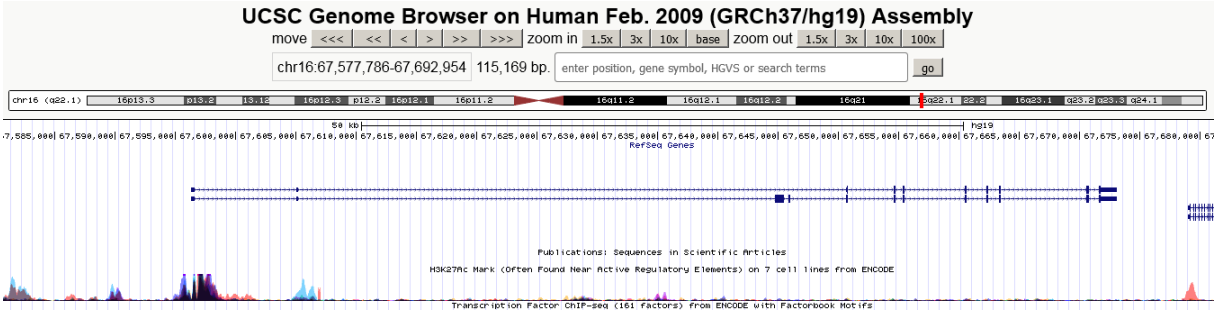

Figure 2. The *CTCF* locus on human chromosome 16. Two human *CTCF* mRNA isoforms differing in their exon structures are produced from the same *CTCF* gene locus. Blue rectangles represent exons of RefSeq mRNAs. High levels of the H3K27Ac

histone modification are associated with active promoter regions and mark the *CTCF* promoter region (Adapted from Internet source 1).

The two human *CTCF* mRNA isoforms are likely to encode two CTCF protein isoforms which differ in their length. The full length canonical CTCF protein isoform shown in Fig. 1 is formed by 727 amino acids, whereas the shorter CTCF protein isoform contains only 399 amino acids. Since the shorter CTCF isoform (Fig. 3) lacks 328 N-terminal amino acids, this isoform is likely to lack the first three zinc finger motifs and therefore to be functionally distinct from the canonical full length CTCF isoform (Internet source 4).

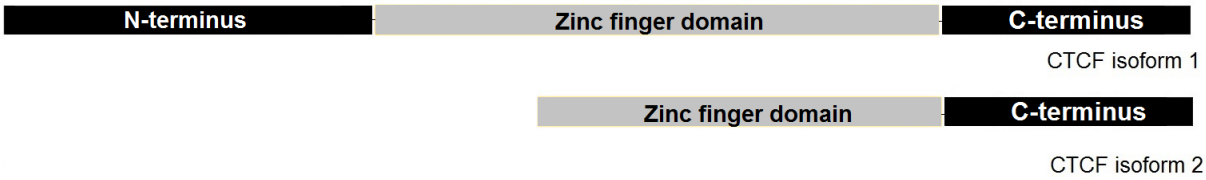

Figure 3. CTCF protein isoforms. The human canonical full-length CTCF protein containing 727 amino acids and all 11 zinc finger motifs is shown as CTCF isoform 1. CTCF isoform 2 is shown as a shorter protein without the first 329 amino acids (Internet source 4).

Various genome-wide binding studies have been performed using a combination of ChIP (Chromatin immunoprecipitation) and the next generation sequencing approach (ChIP-seq) to analyze binding of various transcription factors to their target sequences (Reviewed in Furey, 2012). Some of the ChIP-seq data tracks in the UCSC Genome Browser suggest that the human *CTCF* gene is regulated by various transcription factors (Fig.4). The *CTCF* promoter could have binding sites for proteins such as MYC, YY1 or interestingly, for CTCF itself (Internet source 1). YY1 has been found to be an activating factor of *CTCF* gene (Klenova et al., 1998).

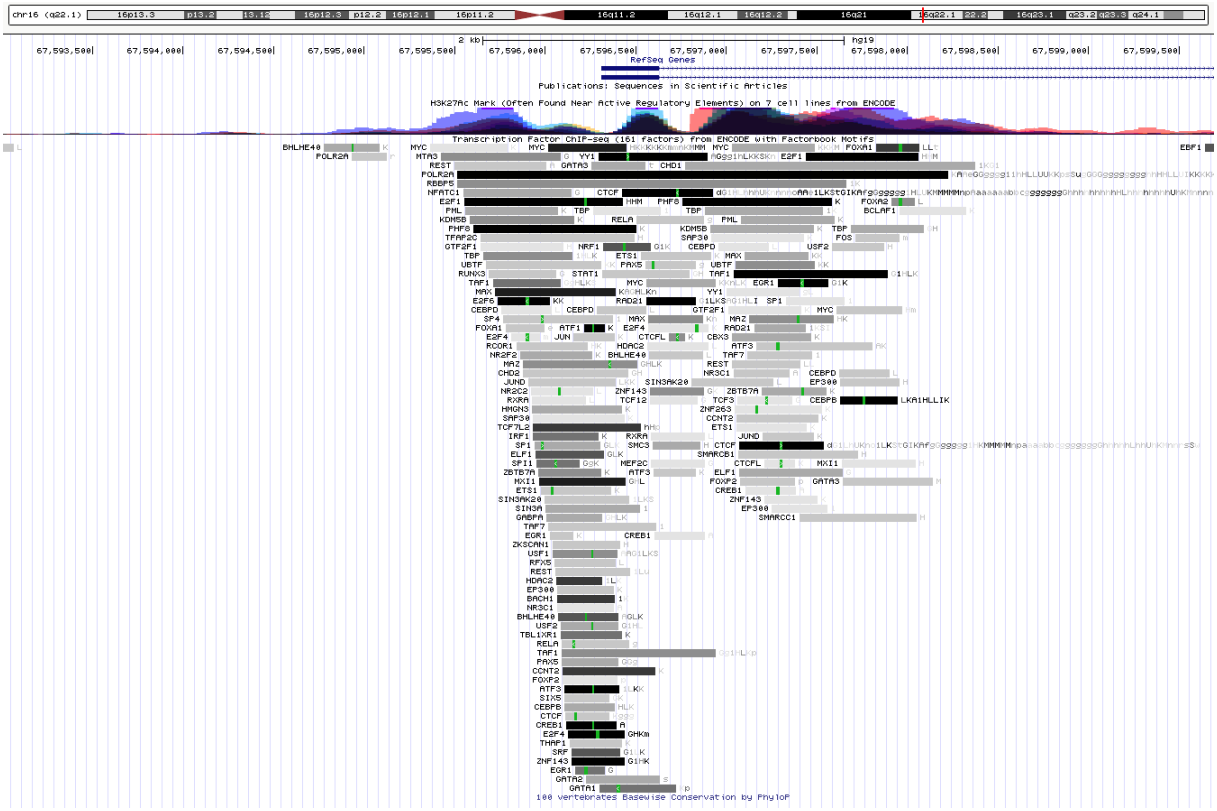

Figure 4. The *CTCF* promoter region on human chromosome 16. Blue rectangles represent the first *CTCF* exon of the two *CTCF* mRNA isoforms (the UCSC genome browser RefSeq track). The gray and black bars represent binding of various transcription factors as analyzed by ChIP-seq (various UCSC genome browser ChIP-seq tracks). The darker the bar the higher the amount of the protein bound to the region. The active human *CTCF* promoter is marked by the H3K27ac histone modification mark (UCSC genome browser H3K27ac ChIP-seq track) (Adapted from Internet source 1).

5. CTCF and three dimensional chromatin

It is now a well-known and generally accepted fact that chromatin is not distributed randomly within the nucleus, but rather that it is arranged in higher order structures. It has been shown that CTCF and its various architectural protein partners are involved in creating and maintaining these higher order chromatin structures (Fig. 5). These higher order chromatin structures play an important role in regulation of gene expression (Reviewed in Bouwman and de Laat, 2015).

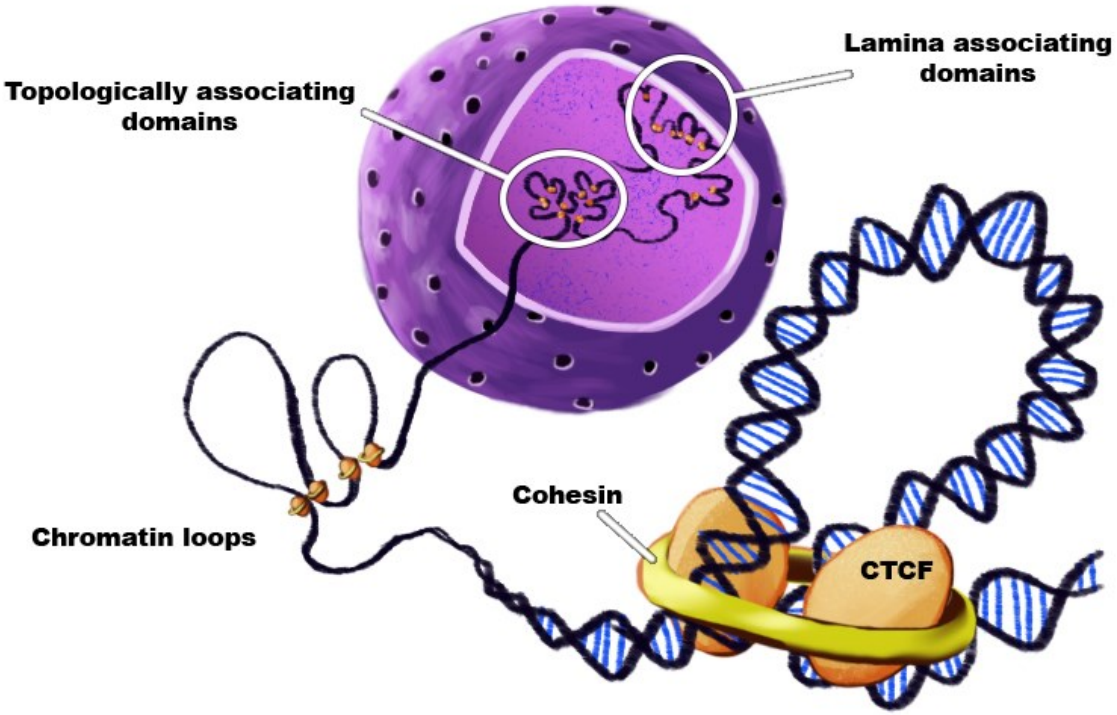

Figure 5. CTCF and higher chromatin structures. Chromatin loops frequently contain a large number of genes (Tang et al., 2015) and form higher genome structures and domains. These domains can be divided into two groups: The first group of domains named topologically associating domains inhabits mainly the center of the nucleus (Dixon et al., 2012), whereas lamina associating domains actively contact the nuclear bilayer (Guelen et al., 2008).

5.1 Architectural protein partners of CTCF

5.1.1 Cohesin

Cohesin and CTCF frequently form a complex that is involved in creating the higher order chromatin structures. Most of the cohesin binding sites overlap with those of CTCF mainly in intergenic regions (Wendt et al., 2008). To form a three-dimensional chromatin loop, the circular protein structure of cohesin encloses the CTCF polypeptide bound to DNA (Fig.5) (Tang et al., 2015).

5.1.2 Zinc finger protein 143

It has been shown that ZNF143 (Zinc finger protein 143) helps to maintain the boundaries of the three-dimensional chromatin structures alongside with the CTCF-cohesin complex. The ZNF143 binding sites on DNA have been found to partially

overlap with the CTCF or cohesin binding sites. When the three-dimensional chromatin structure is forming, ZNF143 can be targeted to its binding site by CTCF and cohesin or by other proteins like MAX (Ye et al., 2016). A SNP in the ZNF143 binding site lowers the binding affinity of ZNF143 to this site (Bailey et al., 2015).

5.2 CTCF-mediated chromatin loops

The CTCF-mediated chromatin loops stabilize the contact between the regions of the same chromosome - these are called intrachromosomal interactions and are relatively strong, and also to stabilize the contact between the regions on different chromosomes - these are called interchromosomal interactions, which are not as strong as intrachromorosomal interactions (Handoko et al., 2011). The CTCF mediated formation of the chromatin loops has been studied in great detail on the *β-globin* locus in the red erythrocytes. It has been found that the *β-globin* locus is surrounded by CTCF binding sites, through which CTCF creates and establishes the long-range chromosomal interactions (Splinter et al., 2006). Each CTCF binding site of the *β-globin* locus has even its own CTCF binding affinity (Farrell et al., 2002). The cells in which CTCF was knocked down using the siRNA approach showed an abberant 3D chromatin structure with disrupted long-range chromatin loops and the same phenotype was seen for the chromatin loops with mutated CTCF binding sites. Mutation of CTCF binding sites also resulted in a changed histone modification status of the region (Splinter et al., 2006).

Generally, the chromatin loops formed by CTCF can be divided into two groups according to their histone modification status: One group of predominantly shorter than 200 kb chromatin loops contains the histone modification marks that are associated with an open chromatin state and with transcriptionally active regions (e.g. H3K4me1, H3K4me2 and H3K36me3) (Handoko et al., 2011, see 5.2.1). The other group of predominantly longer than 200 kb chromatin loops contains the histone modification marks that are associated with a closed chromatin state and with transcriptionally repressed regions (e.g. H3K9me, H3K20me and H3K27me) (Handoko et al., 2011, also see 5.2.2).

5.2.1 Transcriptionally active chromatin loops

The transcriptional activity of the promoters lying inside the first type of chromatin loops formed by CTCF is also dependent on their long-range contacts with regulatory enhancer elements (Fig. 6). Certain cell types have their own specific long-range interactions, which results in transcriptional activity of the genes specifically expressed only in these cell types (Heidari et al., 2014). The interaction between enhancer and its target promoter is dependent on the presence of active histone modifications such as H3K9ac and H3K27ac (Sanyal et al., 2012). The stability of certain enhancer-promoter interactions is frequently strictly dependent on the presence of cohesin and ZNF143 (Bailey et al., 2015; Ing-Simmons et al., 2015).

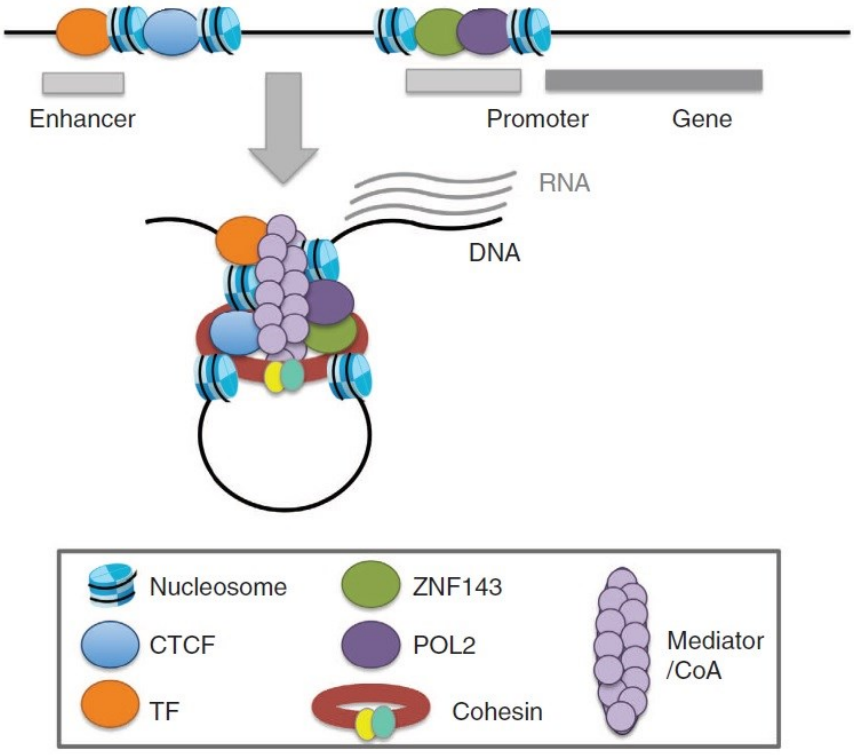

Figure 6. A transcriptionally active chromatin loop formed by CTCF. CTCF and its architectural protein partners (e.g. cohesin and ZNF143) bind to DNA and drag the enhancer element to its target gene promoter (Adapted from Bailey et al., 2015).

A large number of genes lying within the chromatin loops formed by CTCF belong to the family of inducible genes, while the outside of the chromatin loops is inhabited by housekeeping genes. The transcriptionally active chromatin loops in various

immortalized cells contain many genes involved in proliferation or metabolism (Oti et al., 2016). CTCF and cohesin engage in creating the frequently interacting regions (FIREs). These regions are conserved in mammals and exhibit extremely high interacting frequency. Although most of the FIREs lie inside the transcriptionally active chromatin loops, some of them are also located at the loop boundaries. FIREs contain active histone modification marks and transcriptionally active genes. The repressive histone modifications are not usually present in FIREs. The regions flanking FIREs contain a lot of tissue-specific genes suggesting that FIREs participate in the regulation of tissue-specific gene expression (Schmitt et al., 2016).

5.2.2 CTCF enhancer-blocking activity

The chromatin loops formed and maintained by CTCF are also responsible for transcriptional repression of the genes lying inside them, which is caused by the CTCF enhancer-blocking (also called insulator) activity. CTCF binds to the insulator DNA sequence and forms a spherical barrier which prevents the enhancer from contacting and activating the gene promoter. The insulator DNA sequence can contain multiple CTCF binding sites (Bell et al., 1999). In the nucleolus the targeting of CTCF to the insulator DNA sequence is mediated by nucleophosmin which forms a heterodimer with CTCF (Yusufzai et al., 2004). The absence of CTCF enhancer blocking function in the prostate cancer cells results in transcription of fused chimeric *SLC45A3-ELK4* mRNAs, which supports the cell growth and the cell division. The creation of the *SLC45A3-ELK4* chimeric RNA occurs by the failure of CTCF to target the insulator DNA site flanking the *SLC45A3* gene and to target the insulator DNA sites between the *SLC45A3* and the *ELK4* genes. It has been observed that the agents such as androgene receptors interfere with the CTCF targeting to the insulator DNA sequence (Qin et al., 2016; Zhang et al., 2012).

The CTCF enhancer-blocking activity can be inhibited by the protein KAISO whose binding sites are near the CTCF insulator binding sites. To prevent CTCF from acting as an enhancer blocker, KAISO creates a complex with CTCF through its N-terminal POZ domain and the CTCF C-terminal domain (Defossez et al., 2005).

5.2.3 Polarity of CTCF bound to chromatin loops

The arrangement of chromatin loops is dependent on the polarity of CTCF bound to the DNA strands (Fig. 7). The most favored CTCF binding polarity is convergent. Although CTCF bound to DNA in the same orientation (tandem right or tandem left CTCF orientation) can also contribute to the establishment of chromatin loops, the stability of such chromatin loops is not as strong as that of the loop created with convergent oriented CTCF. Also, the visual structure of these two types of loops is different. The convergent CTCF oriented chromatin loops resemble a hairpin, whereas the tandem CTCF oriented chromatin loops resemble a coiled structure (de Wit et al., 2015; Tang et al., 2015).

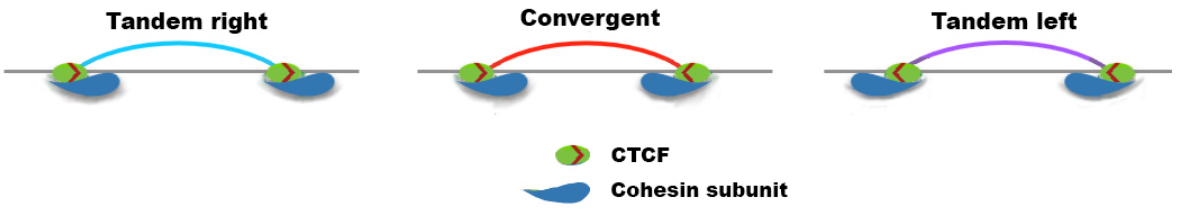

Figure 7. CTCF chromatin loop binding polarity. An illustration shows different CTCF and cohesin binding polarity on DNA - the tandem right polarity, the convergent polarity and the tandem left polarity (Adapted from Tang et al., 2015).

The CRISPR/Cas9-mediated inversion or elimination of the mouse convergent CTCF sites lead to relaxation of the chromatin loops and to misregulation of some genes (e.g. *Sox2* or *Fbn2*). Relaxation of one loop also negatively affects the stability of the neighboring loops (de Wit et al., 2015).

5.3 Topologically associating domains

A group of the chromatin loops formed by CTCF is called a topologically associating domain (TAD). TADs are conserved not only within various cell types but also within various animal species. Their average length is hundreds of kilobases (kb), although some can measure several megabases (Mb) (Dixon et al., 2012). Each TAD consists of several chromatin structures called sub-topologically associating domains (Phillips-Cremins et al., 2013). Long-range chromatin interactions between enhancers and promoters and subsequent gene activations occur primarily within the same TAD. Similarly to the chromatin loop boundaries (see 5.2.1), the TAD boundaries are enriched for housekeeping genes (Dixon et al., 2012). The TAD boundaries have been

shown to be bound by various transcription factors (e.g. ELK1, p300 or RXRA), when the corresponding TAD is transcriptionally active. A SNP in the CTCF binding sites contributes to the TAD boundary destabilization (Mourad and Cuvier, 2016).

It has been shown that cohesin specific protease cleavage affects the long-range interactions within a TAD in a different manner than knocking down CTCF by siRNA (Fig. 8.) (Zuin et al., 2014).

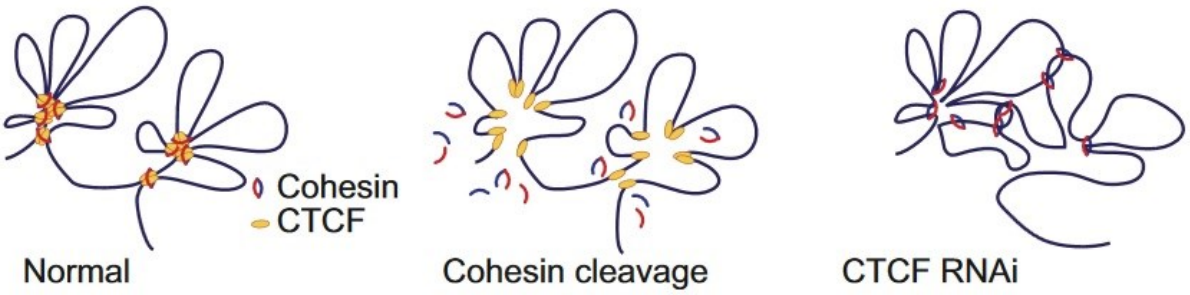

Figure 8. TADs, CTCF and cohesin. The reduced levels of cohesin lead to disruption of whole TADs, whereas the reduced level of CTCF results in partial disruption of TADs and to aberrant new interactions. It has been observed that downregulation of cohesin negatively affects the expression of the *HOX* genes (Zuin et al., 2014).

Another CTCF protein partner involved in building the TADs architecture is TOP2B (DNA topoisomerase 2-beta). TOP2B has been found to recognize and bind the CTCF or the cohesin DNA binding sequences. TOP2B can also interact indirectly with DNA through the CTCF-cohesin complex. In this study the authors hypothesize that TOP2B is involved in the supercoiling of the TAD DNA regions (Uuskula-Reimand et al., 2016).

To better understand the TADs structure and its role in the regulation of gene expression, various advanced softwares have been developed to model the TAD structure based on the available experimental data (e.g. TADtree) (Weinreb and Raphael, 2016).

5.3.1 Disruption of TADs in disease and development

The CRISPR-Cas9 mediated modifications (duplication, deletion or inversion) of various TAD boundaries lead to misregulation of the genes lying within these TADs and to digit abnormalities such as syndactyly, brachydactyly or polydactyly (Fig. 9). The misregulated genes associated with these developmental abnormalities are *PAX3,*

IHH and *WNT6* and their upregulation is caused by their interaction with an enhancer element in the neighboring TAD. In a healthy cell the *PAX3, IHH* and *WNT6* genes do not interact with this enhancer element in the neighboring TAD and thus they remain silent. The CRISPR-Cas9 mediated deletion of the TAD boundary between the TAD containing an enhancer and the TAD containing the *PAX6* gene resulted in brachydactyly. The CRISPR-Cas9 mediated inversion between the TAD containing an enhancer and the TAD containing the *WNT6* gene resulted in syndactyly. The CRISPR-Cas9 mediated duplication between the TAD containing an enhancer and the TAD containing *IHH* resulted in polydactyly (Lupianez et al., 2015).

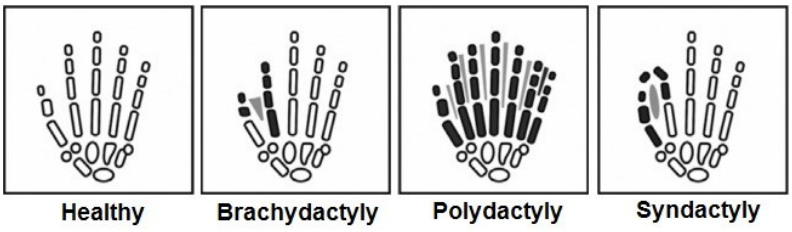

Figure 9. Disrupted TADs and limb development. Unmodified TAD boundary does not cause any misregulation of the *PAX6, WNT6* or *IHH* gene, whereas modifications of the TAD boundaries by CRISPR-Cas9 lead to the shown abnormal phenotypes (Adapted from Lupianez et al., 2015).

TADs can also be disrupted by methylation of CTCF binding sites lying in the TAD boundary. In gliomas the mutated isocitrate dehydrogenase (IDH) enzyme aberrantly methylates CTCF binding sites TAD boundaries which results in disruption of TADs and in upregulation of the *PDGFRA* gene. The resulting overexpression of the *PDGFRA* gene significantly contributes to the glioma development. The researchers has also observed that the correct TAD structure has been restored by inhibiting the IDH methylation activity (Flavahan et al., 2016).

5.4 Lamina associating domains

Lamina associating domains (LADs) are chromatin loops formed by CTCF at the inner layer of the nuclear membrane. LADs interact with the fibrous lamin polypeptides - predominantly with lamin B1 but also with lamin A/C (Amendola and van Steensel, 2015; Guelen et al., 2008). The size of LADs can reach up to a couple of megabases and approximately one third of DNA is estimated to be packed in LADs.

Although LADs are not transcriptionally active, some genes lying within LADs have been reported to be marked by active histone modification marks and to be transcribed. On the other hand, genes lying outside LADs are bound by various transcription activators and are expressed at high levels. These regions with high gene expression levels can block the expansion of mostly transcriptional silenced LADs to the other chromosomal regions (Guelen et al., 2008).

LADs interact with the inner layer of the nuclear membrane with various frequencies. Some LAD regions rarely interact with the nuclear periphery, whereas some LAD regions exhibit a lot of interactions. The high amount of interactions between the genome and the nuclear lamina are common for larger chromosomes and haploid cells (Kind et al., 2015).

6. CTCF and *HOX* genes

The *HOX* genes encode HOX transcription factors that bind to DNA through their homeobox domain and regulate the expression of their target genes. The activity of the *HOX* genes is very important for a number of processes such as the setting up the anterior-posterior early body axis or the dorso-ventral axis of developing limbs. The 39 mammalian *HOX* genes are located in four clusters: *HOXA*, *HOXB*, *HOXC* and *HOXD*. The *HOX* genes are transcriptionally co-regulated and their misregulation leads to severe body malformations, developmental defects and diseases (Reviewed in Pearson et al., 2005). The *HOX* gene clusters contain multiple consensus CTCF binding sites and they have been shown to be regulated by CTCF. Interestingly, the HOX proteins can form complex with CTCF and by doing so they can participate in regulating their own expression (Jerkovic et al., 2017).

6.1 *HOXA* gene cluster

The mammalian *HOXA* cluster contains 11 genes (Reviewed in Pearson et al., 2005). It has been shown that the CTCF binding site between the *HOXA5* and *HOXA6* genes is mainly responsible for silencing the *HOXA9* to *HOXA13* genes in the human fetal lung fibroblasts. On the other hand, in the human foreskin fibroblasts the *HOXA9* to *HOXA13* genes are repressed (Wang et al., 2015). Further, CTCF is also involved in regulating the expression of the *HoxA* genes during differentiation of the mouse motor neurons. In non-differentiated motor neurons the whole *HoxA* cluster is silenced

by CTCF and various repressive histone marks, whereas in differentiated motor neurons the *HoXA1-A6* genes are transcriptionally active. The *HoxA7* to *HoxA13* genes are flanked by various repressive histone modifications (e.g. H3K27me3) and remain transcriptionally repressed. When the CTCF binding sites between the individual *HoxA* genes were disrupted, the epigenetic profile of the *HoxA7* to *HoxA13* genes changed from repressed histone marks (e.g. H3K27me3) to active histone marks (e.g. H3K4me3), which results in their de-repression and transcription (Narendra et al., 2015). By negatively regulating the *HOX* genes, which can also function as tumor suppressors (Pilato et al., 2013), CTCF contributes to cancer proliferation. In the breast tumor cells, the *HOXA10* gene is constantly kept silenced by CTCF bound to the *HOXA10* promoter and by the presence of the H3K27me3 repressive histone modification. Knocking down CTCF with a siRNA results in the *HOXA10* promoter not being occupied by CTCF and therefore in transcriptional activation of this gene. The transcriptional activity of the *HOXA10* gene is accompanied by the presence of the active histone marks such as H3K4me3 on its promoter (Mustafa et al., 2015).

6.2 *HOXD* gene cluster

The mammalian *HOXD* gene cluster contains 9 genes (Reviewed in Pearson et al., 2005). The mouse *HoxD* genes are also flanked by a large number of CTCF binding sites. Transcriptional regulation of the *HoxD* genes by CTCF have been extensively studied for the process of limb development in the mouse. A tissue specific elimination of CTCF in the conditional CTCF knockout mice (Prx1Cre:Ctcf$^{flox/flox}$) results in either positive (*HoxD8* and *HoxD9*) or negative (*HoxD10* and *HoxD13*) regulation of the *HoxD* genes lying in the mouse *HoxD* gene cluster. Downregulation of the *HoxD10* gene has been detected predominantly in the anterior part of the mutant mouse limb, whereas downregulation of the *HoxD13* gene has been detected predominantly in the posterior part of the mutant limb. The absence of CTCF in these mutant mice results in severe developmental limb defects (Fig.10). Most of the bones of the mutant forelimbs do not develop at all, whereas some of the bones of the mutant hind limbs develop to some extent. The CTCF mutant malformations were mainly attributed to the result of a large number of cell apoptotic deaths associated with the misregulation of the apoptotic factor PUMA (Soshnikova et al., 2010).

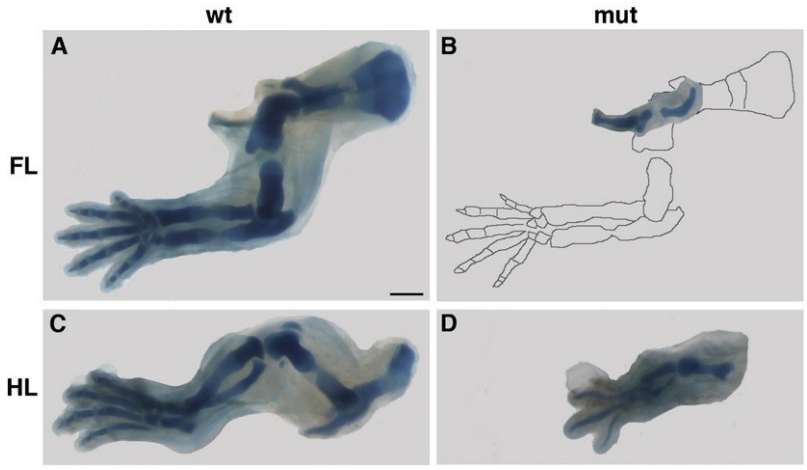

Figure 10. CTCF is essential for the mouse limb development. When CTCF is not present, the forelimbs (FL) and the hind limbs (HL) suffer from severe bone malformations compared to the healthy wild type mouse limbs. Bones are stained blue with β-galactosidase staining. The missing bone elements in the mouse mutants are a consequence of the high level of apoptosis in the mouse limb cells (Adapted from Soshnikova et al., 2010).

7. CTCF-like protein

The CTCF-like protein (CTCFL) is a CTCF paralog (Fig.11), which plays an important role in the spermatid and testes development by binding to and activating the promoter of the gene encoding cerebroside sulfotransferase - a protein important for the testes development (Suzuki et al., 2010).

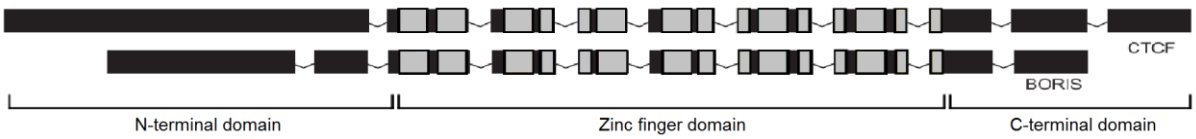

Figure 11. The protein structures of CTCF and CTCFL paralogues. Their central zinc-finger domain is conserved, whereas N- and C-termini show different composition and very low protein sequence identity (Adapted from Hore et al., 2008).

High levels of CTCFL have been observed in the K562 carcinoma cells. In these cells CTCFL and CTCF form a stable heterodimer which binds to the shared CTCFL-CTCF DNA binding sites throughout the genome. Both CTCFL and the CTCFL-CTCF heterodimer activate genes whose protein products stimulate tumor proliferation. Interestingly, no cohesin presence has been detected at the CTCFL-bound regions, which suggests that CTCFL does not participate in the formation of the higher order chromatin structures (Pugacheva et al., 2015).

8. CTCF in invertebrates

CTCF is likely to have evolved long before the protostomes and the deuterostomes since it is present in both groups. As for the protostomes, CTCF has been identified in arthropods and in some nematodes. No signs of the CTCF presence have been observed in such organisms as unicellulars, fungi and plants. (Heger et al., 2013; Heger et al., 2012; Heger et al., 2009)

8.1 Nematodes and Drosophila melanogaster

The bioinformatic analyses of the *Caenorhabditis* genome have revealed that this nematode species lacks CTCF, whereas the *Trichinella* or *Xiphinema* nematode species harbor *CTCF* genes. The *CTCF* genes of the *Trichinella* or *Xiphinema* species consist only of four exons and three introns. In these nematode species the central part of the CTCF protein structure is made up from 10 zinc finger motifs as opposed to 11 zinc fingers in vertebrates. The nematode CTCF protein shows approximately 50% similarity to that of *Drosophila* or mammals (Heger et al., 2009).

In *Drosophila* the central domain of CTCF contains 11 zinc finger motifs similarly to vertebrates (Moon et al., 2005). The *Drosophila* CTCF has been also shown to affect the development of the fruit fly larvae through a machinery involving the biosynthesis of ecdysone from cholesterol through the action of the Halloween proteins (dib, sad, nobo and spok). The Halloween genes are transcriptionally activated by CTCF and the *CTCF* mutant fruit flies had prolonged formations of the pupae as a consequence of the lower Halloween gene transcript levels. The reduced concentrations of the Halloween genes also delayed the expression of the *E74B* gene whose expression is strictly dependent on the expression of the Halloween genes (Fresan et al., 2015).

In *Drosophila melanogaster*, proteins which display a high similarity to the CTCF

zinc finger domain (e.g. Su(Hw) and Zw5) also exhibit an enhancer blocking activity similarly to CTCF (Heger et al., 2013).

9. Conclusions

This work summarizes various functions of the multifunctional CTCF protein. CTCF is a nuclear protein whose central zinc-finger domain is essential for targeting the protein to its binding sites which spread throughout the genome. By binding to its target sites CTCF regulates the expression of various genes. CTCF can transcriptionally activate or repress its target genes, creates long-range interactions and also exhibits an enhancer blocking activity. The formation and maintaining of the three-dimensional genome structure by CTCF is also supported by other architectural proteins such as cohesin and ZNF143. Various posttranslational modifications have been shown to modulate the CTCF functions - some modifications stimulate the activator activity of CTCF whereas other stimulate its repressive properties. CTCF is present not only in vertebrates but also in evolutionarily lower organisms such as arthropods or nematodes. The CTCF paralogue - CTCFL also participates in gene regulation either alone or together with CTCF in a heterodimer complex.

CTCF has been intensively studied by the researchers all over the world because of the extremely large amount of genetic elements regulated by this protein. It will be interesting to see what yet unknown protein partners co-operate with CTCF and whether the two *CTCF* mRNA isoforms encode two different CTCF protein isoforms with different functions. Since the current knowledge of the regulation of the expression of CTCF itself is very limited, it will be also very interesting to analyze the *CTCF* promoter and its regulation more in detail.

10. References

Reviews are marked with an asterisk (*)

* Allende, J.E., and Allende, C.C. (1995). Protein kinases. 4. Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *9*, 313-323.

Amendola, M., and van Steensel, B. (2015). Nuclear lamins are not required for lamina-associated domain organization in mouse embryonic stem cells. EMBO reports *16*, 610-617.

Bailey, S.D., Zhang, X., Desai, K., Aid, M., Corradin, O., Cowper-Sal Lari, R., Akhtar-Zaidi, B., Scacheri, P.C., Haibe-Kains, B., and Lupien, M. (2015). ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nature communications *2*, 6186.

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.

* Bouwman, B.A., and de Laat, W. (2015). Getting the genome in shape: the formation of loops, domains and compartments. Genome Biol *16*, 154.

de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol Cell *60*, 676-684.

Defossez, P.A., Kelly, K.F., Filion, G.J., Perez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C.L., Daniel, J.M., and Gilson, E. (2005). The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. The Journal of biological chemistry *280*, 43017-43023.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

El-Kady, A., and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. FEBS letters *579*, 1424-1434.

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A., and Klenova, E. (2010). Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. Molecular and cellular biology *30*, 1199-1216.

Farrell, C.M., West, A.G., and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol *22*, 3820-3831.

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. Molecular and cellular biology *16*, 2802-2813.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110-114.

Fresan, U., Cuartero, S., O'Connor, M.B., and Espinas, M.L. (2015). The insulator protein CTCF regulates Drosophila steroidogenesis. Biol Open *4*, 852-857.

* Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nature reviews Genetics *13*, 840-852.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Guerrero, P.A., and Maggert, K.A. (2011). The CCCTC-binding factor (CTCF) of Drosophila contributes to the regulation of the ribosomal DNA and nucleolar stability. PloS one *6*, e16401.

Han, D., Chen, Q., Shi, J., Zhang, F., and Yu, X. (2017). CTCF participates in DNA damage response via poly(ADP-ribosyl)ation. Scientific reports *7*, 43530.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F.*, et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics *43*, 630-638.

Heger, P., George, R., and Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. Evolution *67*, 2945-2956.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc Natl Acad Sci U S A *109*, 17507-17512.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC molecular biology *10*, 84.

Heidari, N., Phanstiel, D.H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q., and Snyder, M.P. (2014). Genome-wide map of regulatory interactions in the human genome. Genome Res *24*, 1905-1917.

Hore, T.A., Deakin, J.E., and Marshall Graves, J.A. (2008). The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. PLoS Genet *4*, e1000169.

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergstrom, R., Kwon, Y.W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D.*, et al.* (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. Molecular and cellular biology *27*, 1631-1648.

Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merkenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res *25*, 504-513.

Jerkovic, I., Ibrahim, D.M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., Gonzalez Navarrete, I., Robinson, P.N., Hecht, J., and Mundlos, S. (2017). Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF. PLoS Genet *13*, e1006567.

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.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M.*, et al.* (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134-147.

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

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J.*, et al.* (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Molecular and cellular biology *21*, 2221-2234.

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

Liao, B.Y., and Chang, A. (2014). Accumulation of CTCF-binding sites drives expression divergence between tandemly duplicated genes in humans. BMC Genomics *15 Suppl 1*, S8.

Liu, L., and Heermann, D.W. (2015). The interaction of DNA with multi-Cys2His2 zinc finger proteins. J Phys Condens Matter *27*, 064107.

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

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M., and Sadowski, P.D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. Molecular and cellular biology *29*, 714-725.

Martinez, S.R., and Miranda, J.L. (2010). CTCF terminal segments are unstructured. Protein Sci *19*, 1110-1116.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R.*, et al.* (2005). CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO reports *6*, 165-170.

Mourad, R., and Cuvier, O. (2016). Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. PLoS Comput Biol *12*, e1004908.

Mustafa, M., Lee, J.Y., and Kim, M.H. (2015). CTCF negatively regulates HOXA10 expression in breast cancer cells. Biochem Biophys Res Commun *467*, 828-834.

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.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Oti, M., Falck, J., Huynen, M.A., and Zhou, H. (2016). CTCF-mediated chromatin loops enclose inducible gene regulatory domains. BMC Genomics *17*, 252.

* Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nature reviews Genetics *6*, 893-904.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Pilato, B., Pinto, R., De Summa, S., Lambo, R., Paradiso, A., and Tommasi, S. (2013). HOX gene methylation status analysis in patients with hereditary breast cancer. J Hum Genet *58*, 51-53.

Pugacheva, E.M., Rivero-Hinojosa, S., Espinoza, C.A., Mendez-Catala, C.F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z.*, et al.* (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. Genome biology *16*, 161.

Qi, C.F., Kim, Y.S., Xiang, S., Abdullaev, Z., Torrey, T.A., Janz, S., Kovalchuk, A.L., Sun, J., Chen, D., Cho, W.C.*, et al.* (2012). Characterization of ARF-BP1/HUWE1 interactions with CTCF, MYC, ARF and p53 in MYC-driven B cell neoplasms. Int J Mol Sci *13*, 6204-6219.

Qin, F., Song, Y., Zhang, Y., Facemire, L., Frierson, H., and Li, H. (2016). Role of CTCF in Regulating SLC45A3-ELK4 Chimeric RNA. PloS one *11*, e0150382.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. The Journal of biological chemistry *282*, 33336-33345.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335-348.

Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L.*, et al.* (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep *17*, 2042-2059.

Soshnikova, N., Montavon, T., Leleu, M., Galjart, N., and Duboule, D. (2010). Functional analysis of CTCF during mammalian limb development. Dev Cell *19*, 819-830.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

Suzuki, T., Kosaka-Suzuki, N., Pack, S., Shin, D.M., Yoon, J., Abdullaev, Z., Pugacheva, E., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2010). Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. Molecular and cellular biology *30*, 2473-2484.

Tang, J.B., and Chen, Y.H. (2006). Identification of a tyrosine-phosphorylated CCCTC-binding nuclear factor in capacitated mouse spermatozoa. Proteomics *6*, 4800-4807.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B.*, et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611-1627.

Uuskula-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M.V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E.J.*, et al.* (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. Genome biology *17*, 182.

Vaidya, H., Rumph, C., and Katula, K.S. (2016). Inactivation of the WNT5A Alternative Promoter B Is Associated with DNA Methylation and Histone Modification in Osteosarcoma Cell Lines U2OS and SaOS-2. PloS one *11*, e0151392.

Venkatraman, B., and Klenova, E. (2015). Role of CTCF poly(ADP-Ribosyl)ation in the regulation of apoptosis in breast cancer cells. Indian journal of medical and paediatric oncology : official journal of Indian Society of Medical & Paediatric Oncology *36*, 49-54.

Wang, J., Wang, Y., and Lu, L. (2012). De-SUMOylation of CCCTC binding factor (CTCF) in hypoxic stress-induced human corneal epithelial cells. The Journal of biological chemistry *287*, 12469-12479.

Wang, R., Shen, J., Huang, P., and Zhu, X. (2013). CCCTC-binding factor controls its own nuclear transport via regulating the expression of importin 13. Mol Cells *35*, 388-395.

Wang, X., Xu, M., Zhao, G., Liu, G., Hao, D., Lv, X., and Liu, D. (2015). Exploring CTCF and cohesin related chromatin architecture at HOXA gene cluster in primary human fibroblasts. Sci China Life Sci *58*, 860-866.

Weinreb, C., and Raphael, B.J. (2016). Identification of hierarchical chromatin domains. Bioinformatics *32*, 1601-1609.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T.*, et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

Ye, B.Y., Shen, W.L., Wang, D., Li, P., Zhang, Z., Shi, M.L., Zhang, Y., Zhang, F.X., and Zhao, Z.H. (2016). [ZNF143 is involved in CTCF-mediated chromatin interactions by cooperation with cohesin and other partners]. Molekuliarnaia biologiia *50*, 496-503.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C.*, et al.* (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. Nature genetics *36*, 1105-1110.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell *13*, 291-298.

Zhang, Y., Gong, M., Yuan, H., Park, H.G., Frierson, H.F., and Li, H. (2012). Chimeric transcript generated by cis-splicing of adjacent genes regulates prostate cancer cell proliferation. Cancer Discov *2*, 598-607.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F.*, et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A *111*, 996-1001.

Internet sources:

1) UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly. Retrieved from: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr16%3A67596310-

67673088&hgsid=591216325_Glf54wWqiD6fEJUn0meGTKRK7ixp (accessed 8.5. 2017)

2) UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly. Retrieved from: https://genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr8%3A108160438-108206822&hgsid=591216325_Glf54wWqiD6fEJUn0meGTKRK7ixp (accessed 8.5. 2017)

3) UCSC Genome Browser on Chicken Nov. 2011 (ICGSC Gallus_gallus-4.0/galGal4) Assembly. Retrieved from: https://genome.ucsc.edu/cgi-bin/hgTracks?db=galGal4&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr11%3A1097769-1128266&hgsid=591216325_Glf54wWqiD6fEJUn0meGTKRK7ixp (accessed 8.5. 2017)

4) UniProtKB - P49711 (CTCF_HUMAN). Retrieved from: http://www.uniprot.org/uniprot/P49711 (accessed 8.5. 2017)