An upstream secondary DNA motif within the *IL3* insulator CTCF binding site is required for enhancer-blocking insulator activity.

Sarion R. Bowers[1,2] and Peter N. Cockerill[1,3*]

[1]Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

[2]Leeds Institute of Molecular Research, University of Leeds, St James's University Hospital, Leeds LS9 7TF, UK

[3]Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, Birmingham B15 2TT, UK

[*]Address correspondence to : Peter N Cockerill, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, Birmingham B15 2TT, UK
Email: p.n.cockerill@bham.ac.uk

Short title

CTCF insulator activity requires binding to 2 specific DNA motifs

Abstract

CCCTC-binding Factor (CCTF) is the only known vertebrate protein that functions to organise the genome into distinct functional chromatin domains. A subset of CTCF sites also provide insulator activity and barrier activity to block inappropriate enhancer-promoter communication and spreading of histone modifications and non-coding transcription into loci where it is unwarranted. Paradoxically, other CTCF sites mediate enhancer and promoter communication, partly by supporting DNA loop extrusion within chromatin domains. Despite intensive study and abundant data, it remains poorly understood how CTCF directs these different functions. In this study we provide new data and mine published data that show that CTCF utilises zinc fingers 9 to 11 and an upstream DNA binding consensus sequence resembling CAGCTGTTCC to mediate high affinity binding and enhancer-blocking insulator activity. A single high affinity CTCF binding site from the *IL3* locus is able block *IL3* promoter activation by an upstream enhancer. Mutation of a CTGCAGCTTT sequence upstream of the CTCF core motif abolishes insulator activity. We propose that CTCF is able to confer insulator and barrier activity upon a specific subset of CTCF sites that contain the upstream DNA consensus binding motif, thereby allowing CTCF to function differently in different contexts.

Introduction

CTCF (CCCTC-binding Factor) is a highly conserved, 11-fingered zinc finger (ZF) protein [1-3] able to use different combinations of ZFs to bind to a highly divergent set of DNA sequences [1, 4]. CTCF sites are highly diverse in that their functions range from defining genome domain boundaries, blocking spreading of chromatin modifications, and insulating enhancers from promoters, while conversely some CTCF sites mediate enhancer-promoter communication. CTCF plays a major role in maintaining chromosomal 3D architecture whereby it is found at the borders of independently folded chromatin domains termed Topologically Associating Domains (TADs) that are typically ~1000 kilobase (kb) in size [5-7]. These interaction domains are bordered by a subset CTCF sites that act as roadblocks to directional DNA loop extrusion through the ring of the Cohesin complex at each boundary of the TAD [8-10]. DNA loops extrude through the Cohesin ring on the NH2-terminal or ZF3 side of the bound CTCF complex which faces into the TAD. This provides a mechanism whereby loop extrusion assists enhancer interaction with genes on one side of the CTCF boundary but not the other, and allows some CTCF sites to act as enhancer-blocking insulators in functional enhancer assays [11-13].

It is generally accepted that CTCF is the principal vertebrate insulator-binding protein [14, 15], and it is known to bind to most previously characterized enhancer-blocking insulators, including those in the human *IL3*/*CSF2* locus [11], the chicken beta-globin locus [16], the Igf2/H19 locus [17-20], the HLA-DQA1 locus [21], and the mouse alpha-globin locus [13]. In addition, some CTCF sites, such as ones found in the HOXA cluster and the *FOXJ3* locus, function as barrier elements segregating domains of active chromatin marked by histone H3K4me3 from the spreading of repressive domains of chromatin marked by H3K27me3 [22-24]. The +2.9 and +4.2 kb CTCF sites in the *IL3* insulator also block the spread of read-through non-coding transcription coming from both upstream and downstream [25]. Furthermore, CTCF is the only protein identified thus far that is directly associated with insulator function in vertebrates (other than the CTCF homolog BORIS which is expressed only in germ cells [26]). It is through CTCF that other insulator-associated proteins, such as the Cohesin complex, are recruited to CTCF sites [11, 27-29]. Despite the obvious importance of CTCF, it remains unclear how some CTCF sites can block enhancer function, while others promote enhancer function.

CTCF sites account for ~10% of all ubiquitous DNase I hypersensitive sites (DHSs) in the genome [30]. CTCF employs up to 11 ZFs to bind to its target sites, generating considerable complexity in the sequences that it can recognize. Global CTCF chromatin immunoprecipitation (ChIP) studies defined a core sequence resembling CCACCAG(G/A)GGG that is present in all CTCF sites at least 13,000 times in the human genome [4, 31, 32] and this core is now known to interact with CTCF ZFs 4 to 7 [33-35]. However, other studies revealed that in 13 to 15% of cases CTCF sites make additional contacts via ZFs 9 to 11 to a second upstream consensus sequence resembling CAGCTGTTCC, with ZF 8 bridging a flexible length DNA linker region in between [1, 11, 13, 34-44]. Individual deletions of ZFs 8 to 11 led to preferential loss of CTCF sites containing the upstream sequence [43]. The functions of the upstream CTCF motif and ZFs 9 to 11 remain poorly defined, but they are likely to support higher affinity binding than CTCF sites containing just the core consensus. Furthermore, an inspection of published studies of CTCF sites with known insulator, repressor or chromatin barrier activity reveals that they typically have this upstream motif (Fig 1) [1, 11, 14, 16, 21, 37, 38, 40, 45, 46]. A recent study of CTCF sites in the alpha globin locus found that the CTCF sites that function the most efficiently as enhancer-blocking insulators are also the CTCF sites that have the upstream motif, and this study observed that insulation is greater when the NH2-terminal or ZF3 side of the bound CTCF complex faces towards the enhancer [13]. In this orientation of the CTCF site, the enhancer is downstream of the CCACCAG(G/A)GGG sequence, and any extrusion loops are predicted to progress along the genome in the opposite direction and away from the promoter. A similar previous functional screen for the insulator activity of CTCF sites inserted into the *Sox2* locus

also showed a required for upstream CTCF motif, but found that 2 to 4 tandem copies of CTCF sites were needed for insulator activity, and that the orientation was less important [12].

Figure 1 Legend. Conservation among insulator and boundary CTCF sites.
Comparison of sequence homology between previously characterized CTCF binding sites that have been defined as insulators, repressors or chromatin boundaries [1, 11, 13, 14, 16, 21-24, 28, 37-40, 45-47]. The three lines of sequence in the global CTCF consensus above the table represent the predominant alternative bases allowed at each position with the preferred bases shown above in uppercase [41-43]. Regions protected in published DNase I or DMS footprinting assays are highlighted in bold. Underlined bases represent bases where methylation or carboxyethylation blocks binding of CTCF in interference assays. Above the table are the predicted binding sites of Zn finger domains 3-7 and 9-11 (indicated by brackets and numbers) [33]. Zn Finger 8 is not shown because this domain sits away from the DNA above the variable length spacer region. Zn fingers 3-7 bind the core CTCF consensus sequence common to all CTCF sites (primary motif) and Zn fingers 9-11 bind the secondary upstream consensus sequence which is present in ~13% of CTCF sites. Underneath the table are shown a web logo [48] derived from the sequences in the table, and the published global consensus sequence for CTCF sites that contain the upstream binding site [42, 43].

CTCF controls many aspects of chromatin structure, chromosomal architecture and the regulation of transcription. These functions can include regulating the intrusion of histone modifications or non-coding transcription into gene loci, and long-range chromatin looping. The complex nature of CTCF sequences and their functions makes it very difficult to dissect the individual specific functions of CTCF when sites are embedded within chromosomal DNA. Even in circular plasmids it is difficult to evaluate enhancer-blocking insulators unless they are inserted both upstream and downstream of the gene loci under study. To circumvent these difficulties, we previously developed a novel enhancer-blocking insulator assay based of the transfection of linearised luciferase reporter gene plasmids containing insulator elements in between the *IL3* promoter an upstream *CSF2* enhancer [11]. Using this model we previously reported that insertion of 3 tandem copies of the *IL3* +2.9 kb CTCF site completely blocked *CSF2* enhancer activity in linear plasmids, and blocked 80% of activity in circular plasmids [11]. This model system was able to assay for the enhancer-blocking activity of CTCF in isolation from other activities relating to chromatin modifications or chromatin architecture.

In this new study we employed one of the enhancer-blocking CTCF sites from the *IL3* insulator [11] as a model to investigate the functions of the secondary upstream CTCF binding motif. We first showed that *in vitro* CTCF binding is greatly reduced in the absence of this upstream DNA motif. We next employed our previously described enhancer-blocking insulator assay based on linearised DNA fragments containing the *CSF2* enhancer and the *IL3* promoter [11]. We demonstrated that a single CTCF site from the *IL3* insulator was sufficient to confer robust enhancer-blocking insulator activity and that the secondary upstream consensus sequence was essential for this activity. This observation now further builds upon our understanding of mechanisms that determine why only a subset of CTCF sites function as insulators.

Results

The *IL3* insulator encompasses 2 nearly identical CTCF binding sites that include both the core consensus and an upstream CTGCAGTTTC motif (Fig 2A). Each of these motifs is highly conserved across mammalian species for both sites (S3 Table). Our previous study of this insulator employed *in vivo* footprinting using dimethyl sulphate to confirm that CTCF bound to both the upstream and core motifs, while the flanking sequences were hypersensitive to DNaseI (summarised in Fig 2A) [11].

Figure 2 Legend. EMSAs of CTCF binding to the *IL3* insulator CTCF site.
(A) Locations of the core and upstream CTCF motifs within CTCF sites found within the human IL-3 gene insulator [11]. The boxes represent regions protected from DNase I in footprinting assays. G bases protected from modification by DMS in *in vivo* footprinting assays are indicated by open circles. Regions hypersensitive to DNase I (arrows) or modification by DMS (filled circles) are also indicated.
(B) Sequences of CTCF EMSA probes used in this study. Mutated residues are shown in bold lowercase and the two consensus sequences are underlined.
(C-E) EMSAs using intact or mutated *IL3* +2.9 kb CTCF sites. EMSAs included 4 µg of Jurkat nuclear extract with either the +2.9, Δ1°, Δ2°(1), Δ2°(2), Δ1°/Δ2°(2), HS4, or ΔHS4 probes (C). Specificity of CTCF binding was confirmed by incubating 4 µg of nuclear extract with 2 µl of anti-CTCF or control IgG antibodies, and by comparison to the chicken beta globin HS4 CTCF site (C). The vertical black lines indicated parts of the gel where 2 irrelevant lanes were deleted. Binding data was quantified by densitometry (D).
(E - G) Competition assays of CTCF binding. EMSAs included 4 µg of Jurkat nuclear extract in the presence of increasing amounts unlabelled competitor as indicated above. (H) Densitometric quantification of the EMSAs in panels E and F depicting the efficiency of the mutated +2.9 CTCF sites to compete with wild-type +2.9 site. Values are expressed as the fold decrease in the presence of competitor. The unmanipulated EMSA images are shown in S2 Figure.

The upstream CTCF motif promotes high affinity binding to DNA
We employed Electrophoretic Mobility Shift Assays (EMSAs) using Jurkat human T cell nuclear extracts to investigate the relative contributions of the primary core (1°) and secondary upstream (2°) motifs to *in vitro* CTCF binding. Because CTCF is a ZnF protein, EMSAs were performed in the presence of Zn^{2+}, which we previously found to be essential for efficient *in vitro* binding of CTCF using nuclear extracts [11]. We synthesised various mutated derivatives of the +2.9 kb *IL3* CTCF site for use in EMSAs and included intact and mutated versions of the chicken beta-globin HS4 CTCF site as controls (Fig 2B). The identity of the CTCF complex was confirmed using CTCF antibodies and control IgG (Fig 2C). EMSAs demonstrated that *in vitro* CTCF binding was highly dependent on the presence of the conserved core motif, and binding was abolished if the core consensus was mutated in either the *IL3* or the HS4 CTCF site (Δ1° and ΔHS4) (Fig 2C). High affinity CTCF binding was dependent upon maintaining the upstream motif intact, whereby binding was reduced 3-fold when it was mutated (Δ2°) (Fig. 2C). These binding data were quantified by densitometry (Fig 2D). The relative strength of binding of the intact and mutated CTCF sites in EMSAs was also evaluated by including 3, 10 or 30 ng of unlabelled oligonucleotides as competitors (Figs 2E - G). The mutation of the core sequence eliminated all competing binding activity while mutation of the upstream sequence resulted in a modest decrease in competition.

The secondary upstream CTCF motif is essential for enhancer-blocking *IL3* insulator function
We next dissected the relative contributions that the upstream and core consensus sequences made towards the enhancer-blocking activity of the *IL3* +2.9 kb insulator CTCF site. We performed transfection assays in Jurkat T cells with plasmids containing a 717 bp fragment of the inducible human *CSF2* enhancer (GME) upstream of the -559 to +50 region of the human *IL3* promoter in front of the luciferase gene in pXPG (Fig 3). The activity of the inducible *IL3* promoter was measured in linearised plasmids 40 hours after transfection into Jurkat T cells, following stimulation for 8 h with 20 ng/ml PMA and 1 mM calcium ionophore A23187 to induce T cell receptor signalling. Enhancer-blocking insulation activity was measured after placing a single CTCF site in the reverse orientation in between the enhancer and promoter. In this configuration the CTCF ZF 3 to 7 binding sequence CCACTAGGGGGAGACA is facing towards the enhancer, which is predicted to maximise the insulation

activity of CTCF. Insertion of just a single copy of the *IL3* CTCF site was sufficient to almost entirely block *CSF2* enhancer activity (Fig 3). However, mutation of either the upstream motif or the core CTCF consensus sequence abolished enhancer-blocking insulator activity (Fig 3).

Figure 3 Legend. Enhancer blocking assay of a CTCF site.
Linearised luciferase reporter gene plasmids containing the human IL-3 gene promoter (pIL3) were assayed in Jurkat T cells. As depicted in the cartoon at the right, derivatives of pIL3H contained either just the *CSF2* enhancer (pIL3-GME) or the enhancer plus a CTCF site inserted between the promoter and the enhancer (pIL3-C-GME). CTCF site plasmids contained either the intact *IL3* +2.9 kb CTCF site, or this CTCF with either the primary or the secondary consensus motif mutated, as indicated below. The CTCF sites are inserted into the plasmids in the opposite orientation to that shown which means that the NH2 terminus of bound CTCF faces towards the enhancer. Luciferase activities were measured 40 h after transfection plus an additional 8 h after stimulating transfected Jurkat cells with PMA and calcium ionophore. Luciferase activities were expressed relative to pIL3-GME having a value of 1, and represent the mean of 6 transfections, using two independently derived DNA clones. Error bars indicate Standard Deviation. Data was normalised by comparison with the cotransfected Renilla luciferase control plasmid pRL-TK.

The upstream CTCF motif is a common feature of enhancer insulators and chromatin barriers

The role of the upstream consensus sequence at insulators remains poorly understood. To find further evidence for its significance we performed a brief survey of other well-defined CTCF-dependent insulators and found that they typically included a near ideal match to the 10 bp upstream binding site (highlighted in yellow in Fig 1). Figure 1 also summarises previously published *in vitro* DNA binding data and *in vivo* footprinting data confirming contacts between CTCF and the upstream motif where ZFs 9 to 11 are predicted to bind. Regions protected by CTCF from either DNase I digestion and/or methylation by DMS *in vivo* are highlighted in bold, whereas regions where CTCF binding is blocked by *in vitro* modification of DNA by either methylation or carboxyethylation are underlined. Studies of nuclease accessibility at these sites invariably find that the regions occupied by ZF 4 to 7 and Z F9 to 11 are separated by an accessible variable length linker approximately 8 bp in length where ZF8 is thought to sit away from an accessible region of linker DNA [11, 13, 33, 44].

Figure 1 also lists CTCF sites within previously defined chromatin boundary elements from the *HOXA9* and *FOXJ3* loci that similarly include the upstream CTCF motif [22-24]. To find further evidence for the presence of the upstream CTCF motif at histone H3 K27me3 chromatin boundaries we next performed a limited scan of the publicly available genome data from a global study of histone methylation and CTCF binding, provided by the authors online via a link to the UCSC genome browser (https://dir.nhlbi.nih.gov/papers/lmi/epigenomes/hgtcell.aspx) [24, 47]. This allowed us to identify several additional CTCF sites with conserved upstream CTCF motifs at H3K27me3 chromatin boundaries (Fig 1 and S1 Figure). We examined the conservation of a sample of the above sites and found that like the *IL3* insulator (Fig 1 and S3 Table), the upstream CTCF site is highly conserved in at least the UPK2, HOXA9, FOXJ3 and ABCG8 CTCF sites (S4 to S6 Tables).

Taken together, the above data collectively provide strong evidence that the upstream CTCF consensus present in ~13-15% of all CTCF sites plays a major role in both enhancer insulator and chromatin barrier function. A sequence logo derived from the insulator sites in Fig 1 is essentially identical to that derived for the upstream consensus, suggesting a strong functional association.

Discussion

The aim of this study was to further raise the consciousness of the scientific community about the potential role of the upstream CTCF consensus in insulator and barrier function. Despite numerous

studies over the last 25 years revealing binding of CTCF to an upstream CTCF motif at known insulators, the role and mechanism of action of this secondary binding site is still not fully understood or appreciated. Our findings are in agreement with another recent study of CTCF sites in the mouse alpha globin locus which similarly found a correlation between the presence of the upstream CTCF motif and the ability of CTCF sites to function as enhancer-blocking insulators when inserted into the genome [13]. They also agree with the findings of a genomic screen for insulators inserted into the mouse *Sox2* locus whereby CTCF sites that functioned as insulators, or were found at TAD boundaries, were also those CTCF sites that included the upstream CAGCTGTTCC-like motif [12]. However, one major difference to our study was that the *Sox2* study found that single CTCF sites had no enhancer-blocking insulator activity and four tandem CTCF sites were required to block 38% of gene expression [12]. In our study, a single CTCF site functioned as an insulator to block 74% of enhancer activity in the context of non-integrated linear DNA, although we also previously observed that 100% insulation was only achieved when 3 tandem copies were used [11]. The above study of CTCF sites in the alpha globin locus also found that insulation strength increased with insertion of multiple CTCF sites [13].

The mechanism of insulation by CTCF remains an enigma

The mechanism of action of insulation mediated via the upstream CTCF motif is far from clear. In our study we found that the upstream motif was needed for high affinity binding, but the work of others indicated that high affinity binding alone is not sufficient for insulator activity [12, 13]. This raises the possibility that CTCF has 2 independent functional DNA binding domains that contribute different activities. This is supported by the fact that CTCF uses ZFs 4 - 7 and ZFs 9 – 11 as two independent clusters of DNA binding domains whereby the ZFs bind side by side to adjacent 3 bp ZF recognition sequences to the two separate motifs which are interrupted by a variable length (~7 – 10 bp) DNA spacer bridged by the more flexible non-bound ZF8. The data suggest that interactions with the core CTCF motif via ZFs 4 to 7 are essential for all CTCF binding whereas binding via ZFs 9 -11 contributes additional functionality at a subset of sites. It is evident that the 2 DNA binding domains bind somewhat independently because the DNA linker is of flexible length and is highly accessible to nucleases when CTCF is bound.

CTCF sites that function as insulators or TAD boundaries normally function in a specific orientation whereby the core motif and ZF3 face into the loop [8]. This may be related to the fact that CTCF recruits cohesin, and DNA loop extrusion through the cohesin ring proceeds by drawing in DNA from the ZF3 end of the CTCF/cohesin complex [8, 13], which is downstream and not upstream of the CTCF core motif [10]. As TAD DNA loop extrusion can extend for hundreds of kilobases, most CTCF sites probably have the ability to pass through the cohesin ring. Taken together, these observations support a model whereby high affinity binding of ZFs 9 – 11 to the upstream motif provides a double lock to the CTCF complex, stabilising its binding, and rendering domain boundaries and loop ends more stable. Enhancer-blocking insulation also works most effectively when ZF3 is facing towards the enhancer [13]. However, it seems unlikely that enhancer insulation can be regulated solely on the basis of effects on loop extrusion and enhancer looping to promoters because many insulator assays have the enhancer less than 1 kb from the promoter. Nevertheless, cohesin almost certainly contributes to insulator activity, and co-association with the cohesin protein Rad21 provided a better predictor of insulator activity that just strength of CTCF binding [13].

There is also evidence that CTCF binds less rigidly to CTCF sites lacking the upstream motif [49]. Some CTCF sites bind in a highly tissue-specific manner that relies on cooperation with other transcription factors binding to adjacent sequences. A study of these dynamic CTCF sites in blood cells found that only 6% of these sites included the upstream motif whereas most insulator and barrier CTCF sites have it, whereas 13 – 15% of CTCF sites globally include the upstream motif [49].

While we still have a poor understanding of the molecular mechanisms of enhancer blocking, there may also be a role for CTCF ZF8 which interacts with the phosphate backbone in the minor groove of the linker DNA separating the core and upstream motifs [33]. The other ZFs interact

with the DNA bases in the major groove. It is likely that any structure formed by ZF8 on DNA will itself be much more rigid, and fixed like a bridge above the linker, when both motifs are tightly bound. Such a rigid structure would be expected to behave differently than one where binding via ZFs 8 – 11 is dynamic.

The upstream CTCF binding motif is not conserved in insects

The role of the upstream CTCF binding motif at insulators described here may be restricted to vertebrates. A study of CTCF ChIP peaks in Drosphila S2 cells found no enrichment for the closely linked CTGCAG-like motifs identified in CTCF sites in human cells, but just the same core consensus motif as is found in vertebrates [50]. This is most likely due to the fact that ZFs 9 – 11 are not as highly conserved between human and Drosophila CTCF as are ZFs 4 – 7 [51]. This includes 2 amino acids in ZF9 and 1 amino acid in ZF11 that are predicted to contact DNA in human CTCF [50]. Although Drosophila do express both CTCF and cohesin, they may not interact or function in the same way as they do in vertebrates. This is partly because Drosophila utilise multiple insulator proteins and not just CTCF [50, 52]. Drosophila insulator CTCF sites also showed a partial dependency upon additional unidentified flanking sequences located within 90 bp of the core motif [50].

CTCF is a multifunctional protein

Taken together, our data and recent published studies all predict that CTCF can use two distinct modes of binding to contribute very different and seemingly contradictory functions. Binding via just ZFs 4 – 7 may support more dynamic binding, and generate a structure able to pass through cohesin rings. This mode of binding is more likely to fit with the vast majority of CTCF sites that probably assist enhancer-promoter communication. Conversely, binding that also includes tight binding to ZFs 9-11 may generate a more rigid structure that is either unable to pass though cohesin rings or is tightly bound to cohesin itself. This mode of binding appears to support insulator or barrier function. Future studies will need to determine the structures that CTCF adopts when bound to DNA together with its partners to gain a better understanding of the mechanisms of insulation.

Materials and Methods

Electrophoretic Mobility Shift Assays (EMSAs)

Double stranded oligonucleotides containing the wild type sequence, mutations or truncations of the +2.9 kb CTCF binding site from the IL-3/GM-CSF insulator [11] or chicken β-globin gene [14] were labeled by end-filling with α-[32P]-dCTP plus unlabelled dATP, dGTP and dTTP. The oligonucleotide sequences were as follows:

+2.9:
TCGATATCCTGTGGCATGTCTCCCCCTAGTGGTCCTTCCAGAAACTGCAGCCGCCGCCCCTGCTCCTCCCTCGA
Δ2°(1):
TCGATATCCTGTGGCATGTCTCCCCCTAGTGGTCCTTCCAGAAAATTGTTATGCCGCCCCTGCTCCTCCCTCGA
Δ2°(2):
TCGATATCCTGTGGCATGTCTCCCCCTAGTGGTCCTTCCCTTTGATTGTTATGCCGCCCCTGCTCCTCCCTCGA
Δ1°:
TCGATATCCTGTGGCATGTCATTCCAGAGGAATCCTTCCAGAAACTGCAGCCGCCGCCCCTGCTCCTCCCTCGA
HS4:
TCGATCCCCCAAAGCCCCCAGGGATGTATTTACGTCCCTCCCCCGCTAGGGGGCAGCAGCGAGCCGCCCCTCG
A
ΔHS4:
TCGATCCCCCAAAGCCCCCAGGGATGTATTTACGTCCCTCCCTTCCTCTGGAATAGCAGCGAGCCGCCCCTCGA
Δ1°/Δ2°:
TCGATATCCTGTGGCATGTCATTCCAGAGGAATCCTTCCCTTTGATTGTTATGCCGCCCCTGCTCCTCCCTCGA

Nuclear extract from unstimulated Jurkat cells were prepared as previously described [53]. 4 µg of nuclear extract was incubated with 4 µg of poly(dI-dC) and 100 ng of an unrelated 22 bp

oligonucleotide duplex competitor (containing the GM450 Runx-1 element from the GM-CSF enhancer [54] for 10 minutes before addition of 0.2 ng of labeled oligonucleotide in a 17 µl reaction mixture containing 10% glycerol, 20 mM HEPES, 30 mM KCl, 30 mM NaCl, 0.1 mM ZnSO4, 0.1 mM MgCl2, 1% dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, 5 µg/ml aprotinin, 5 µg/ml leupeptin for 15 minutes at room temperature (~ 22 °C). 4% polyacrylamide gels containing 25 mM Tris borate, 0.5 mM EDTA, were pre-electrophoresed at 10 V/cm for 1 h and run at 10 V/cm for 1 h 20 min at room temperature. Gels were fixed in 0.1% cetyl trimethylammonium bromide, 50 mM sodium acetate before drying and visualization on a phosphorimager screen. Supershift assays were carried out by addition of 2 µl of CTCF antibody (07-729 Millipore) or IgG antibody (12-370 Millipore) to 4 µg of Jurkat nuclear extract and incubation for 10 min at room temperature before subsequent addition of poly(dI-dC) and oligonucleotide. Competition assays were carried out by addition of the specified amount of unlabelled oligonucleotide competitor in 1 µl to 4 µg of nuclear extract and incubated for 15 mins at room temperature before addition of 0.2 ng of +2.9 labeled oligonucleotide.

Relative binding and competition efficiency were determined by densitometric analysis using Quantity One (Bio-Rad) software.

Enhancer-blocking assays

The insulator function of individual CTCF sites was tested using linearised firefly luciferase reporter gene plasmids as previously described [11]. Briefly, gene fragments cloned into pXPG [55] were cotransfected with the Renilla luciferase control plasmid pRL-TK into Jurkat cells which were then cultured for 40 h before stimulation for 8 h with 20 ng/ml Phorbol 12-myristate 13-acetate (PMA) and 1 mM calcium ionophore A23187. Cell extracts were then harvested for dual luciferase activity. The plasmid pIL3 (also known as pIL3H [56]) contains the -559 to +50 region of the human IL-3 gene promoter in the Hind III site of pXPG [55], and pIL3-GME contains the 717 bp Bgl II fragment of the human GM-CSF gene enhancer cloned upstream in the Bgl II site of pIL3. The pIL3-C-GME series of insulator plasmids were made by inserting a single copy of the indicated DNA segments containing the intact or disrupted IL-3 +2.9 kb CTCF site into the Xho I site of pIL3-GME using Xho I adapters at each end. The CTCF sites are inserted into the plasmids in the opposite orientation to that shown in Fig 1A and Fig 2 which means that the NH2 terminus of bound CTCF and ZF3 face towards the enhancer while ZF11 faces towards the promoter. Firefly luciferase plasmids were linearised with Aat II prior to transfection. Relative luciferase activity was calculated as the ratio of firefly and Renilla luciferase activity.

Acknowledgements

This work was supported by Biotechnology and Biological Sciences Research Council grant BB/E023002/1 awarded to P.N.C.

References

1. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, et al. 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. 1996;16(6):2802-13. Epub 1996/06/01. PubMed PMID: 8649389.

2. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, et al. 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. Mol Cell Biol. 1993;13(12):7612-24. Epub 1993/12/01. PubMed PMID: 8246978; PubMed Central PMCID: PMC364833.

3. Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, et al. 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. 1990;5(12):1743-53. Epub 1990/12/01. PubMed PMID: 2284094.
4. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007;128(6):1231-45. PubMed PMID: 17382889.
5. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148(3):458-72. Epub 20120119. doi: 10.1016/j.cell.2012.01.010. PubMed PMID: 22265598.
6. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381-5. Epub 20120411. doi: 10.1038/nature11049. PubMed PMID: 22495304; PubMed Central PMCID: PMCPMC3555144.
7. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376-80. Epub 20120411. doi: 10.1038/nature11082. PubMed PMID: 22495300; PubMed Central PMCID: PMCPMC3356448.
8. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-80. Epub 20141211. doi: 10.1016/j.cell.2014.11.021. PubMed PMID: 25497547; PubMed Central PMCID: PMCPMC5635824.
9. de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, et al. CTCF Binding Polarity Determines Chromatin Looping. Mol Cell. 2015;60(4):676-84. Epub 20151029. doi: 10.1016/j.molcel.2015.09.023. PubMed PMID: 26527277.
10. Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters JM. DNA loop extrusion by human cohesin. Science. 2019;366(6471):1338-45. Epub 20191121. doi: 10.1126/science.aaz3418. PubMed PMID: 31753851.
11. Bowers SR, Mirabella F, Calero-Nieto FJ, Valeaux S, Hadjur S, Baxter EW, et al. A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes. Mol Cell Biol. 2009;29(7):1682-93. Epub 2009/01/23. doi: 10.1128/MCB.01411-08. PubMed PMID: 19158269; PubMed Central PMCID: PMC2655614.
12. Huang H, Zhu Q, Jussila A, Han Y, Bintu B, Kern C, et al. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. Nat Genet. 2021;53(7):1064-74. Epub 20210517. doi: 10.1038/s41588-021-00863-6. PubMed PMID: 34002095; PubMed Central PMCID: PMCPMC8853952.
13. Tsang FH, Stolper RJ, Hanifi M, Cornell LJ, Francis HS, Davies B, et al. The characteristics of CTCF binding sequences contribute to enhancer blocking activity. Nucleic Acids Res. 2024;52(17):10180-93. doi: 10.1093/nar/gkae666. PubMed PMID: 39106157; PubMed Central PMCID: PMCPMC11417384.
14. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999;98(3):387-96. PubMed PMID: 10458613.
15. Liu M, Maurano MT, Wang H, Qi H, Song CZ, Navas PA, et al. Genomic discovery of potent chromatin insulators for human gene therapy. Nat Biotechnol. 2015;33(2):198-203. Epub 20150112. doi: 10.1038/nbt.3062. PubMed PMID: 25580597.
16. Chung JH, Bell AC, Felsenfeld G. Characterization of the chicken beta-globin insulator. Proc Natl Acad Sci U S A. 1997;94(2):575-80. Epub 1997/01/21. PubMed PMID: 9012826.
17. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000;405(6785):482-5. Epub 2000/06/06. doi: 10.1038/35013100. PubMed PMID: 10839546.

18. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 2000;405(6785):486-9. PubMed PMID: 10839547.
19. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, et al. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol. 2000;10(14):853-6. Epub 2000/07/19. doi: S0960-9822(00)00597-2 [pii]. PubMed PMID: 10899010.
20. Szabo P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol. 2000;10(10):607-10. Epub 2000/06/06. doi: S0960-9822(00)00489-9 [pii]. PubMed PMID: 10837224.
21. Majumder P, Gomez JA, Boss JM. The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. J Biol Chem. 2006;281(27):18435-43. Epub 2006/05/06. doi: M601298200 [pii]

10.1074/jbc.M601298200. PubMed PMID: 16675454.
22. Luo H, Wang F, Zha J, Li H, Yan B, Du Q, et al. CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. Blood. 2018;132(8):837-48. Epub 20180514. doi: 10.1182/blood-2017-11-814319. PubMed PMID: 29760161; PubMed Central PMCID: PMCPMC6107877.
23. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015;347(6225):1017-21. doi: 10.1126/science.1262088. PubMed PMID: 25722416; PubMed Central PMCID: PMCPMC4428148.
24. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 2009;19(1):24-32. Epub 2008/12/06. doi: gr.082800.108 [pii]

10.1101/gr.082800.108. PubMed PMID: 19056695; PubMed Central PMCID: PMC2612964.
25. Mirabella F, Baxter EW, Boissinot M, James SR, Cockerill PN. The human IL-3/granulocyte-macrophage colony-stimulating factor locus is epigenetically silent in immature thymocytes and is progressively activated during T cell development. J Immunol. 2010;184(6):3043-54. Epub 2010/02/12. doi: jimmunol.0901364 [pii]

10.4049/jimmunol.0901364. PubMed PMID: 20147630.
26. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, et al. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A. 2002;99(10):6806-11. Epub 2002/05/16. doi: 10.1073/pnas.092123699

99/10/6806 [pii]. PubMed PMID: 12011441; PubMed Central PMCID: PMC124484.
27. Stedman W, Kang H, Lin S, Kissil JL, Bartolomei MS, Lieberman PM. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. EMBO J. 2008;27(4):654-66. PubMed PMID: 18219272.
28. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell. 2008;132(3):422-33. PubMed PMID: 18237772.
29. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008;451(7180):796-801. PubMed PMID: 18235444.

30. Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, et al. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS genetics. 2007;3(8):e136. Epub 2007/08/22. doi: 10.1371/journal.pgen.0030136. PubMed PMID: 17708682; PubMed Central PMCID: PMC1950163.
31. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 2008;36(16):5221-31. Epub 2008/08/08. doi: gkn488 [pii]

10.1093/nar/gkn488. PubMed PMID: 18684996; PubMed Central PMCID: PMC2532738.
32. Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A. 2007;104(17):7145-50. PubMed PMID: 17442748.
33. Yang J, Horton JR, Liu B, Corces VG, Blumenthal RM, Zhang X, et al. Structures of CTCF-DNA complexes including all 11 zinc fingers. Nucleic Acids Res. 2023;51(16):8447-62. doi: 10.1093/nar/gkad594. PubMed PMID: 37439339; PubMed Central PMCID: PMCPMC10484683.
34. Yin M, Wang J, Wang M, Li X, Zhang M, Wu Q, et al. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res. 2017;27(11):1365-77. Epub 20171027. doi: 10.1038/cr.2017.131. PubMed PMID: 29076501; PubMed Central PMCID: PMCPMC5674162.
35. Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol Cell. 2017;66(5):711-20 e3. Epub 20170518. doi: 10.1016/j.molcel.2017.05.004. PubMed PMID: 28529057; PubMed Central PMCID: PMCPMC5542067.
36. Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, et al. 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. J Biol Chem. 2007;282(46):33336-45. PubMed PMID: 17827499.
37. Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, Moore JM, et al. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet. 2001;28(4):335-43. Epub 2001/08/02. doi: 10.1038/ng570

ng570 [pii]. PubMed PMID: 11479593.
38. Quitschke WW, Taheny MJ, Fochtmann LJ, Vostrov AA. Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. Nucleic Acids Res. 2000;28(17):3370-8. Epub 2000/08/23. PubMed PMID: 10954607; PubMed Central PMCID: PMC110710.
39. Vostrov AA, Quitschke WW. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. J Biol Chem. 1997;272(52):33353-9. Epub 1998/01/31. PubMed PMID: 9407128.
40. Burcin M, Arnold R, Lutz M, Kaiser B, Runge D, Lottspeich F, et al. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol Cell Biol. 1997;17(3):1281-8. PubMed PMID: 9032255.
41. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell. 2012;148(1-2):335-48. Epub 20120112. doi: 10.1016/j.cell.2011.11.058. PubMed PMID: 22244452; PubMed Central PMCID: PMCPMC3368268.
42. Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 2013;3(5):1678-89. Epub 20130523. doi: 10.1016/j.celrep.2013.04.024. PubMed PMID: 23707059; PubMed Central PMCID: PMCPMC3770538.

43. Soochit W, Sleutels F, Stik G, Bartkuhn M, Basu S, Hernandez SC, et al. CTCF chromatin residence time controls three-dimensional genome organization, gene expression and DNA methylation in pluripotent cells. Nat Cell Biol. 2021;23(8):881-93. Epub 20210729. doi: 10.1038/s41556-021-00722-w. PubMed PMID: 34326481.
44. Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, et al. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res. 2011:Advance publication doi 10.1101/gr.112656.110. Epub 2010/11/26. doi: 10.1101/gr.112656.110. PubMed PMID: 21106903.
45. Awad TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, et al. Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. J Biol Chem. 1999;274(38):27092-8. Epub 1999/09/10. PubMed PMID: 10480923.
46. Kohne AC, Baniahmad A, Renkawitz R. NeP1. A ubiquitous transcription factor synergizes with v-ERBA in transcriptional silencing. J Mol Biol. 1993;232(3):747-55. Epub 1993/08/05. doi: S0022-2836(83)71428-2 [pii]

10.1006/jmbi.1993.1428. PubMed PMID: 8102652.
47. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. Epub 2007/05/22. doi: S0092-8674(07)00600-9 [pii]

10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.
48. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90. doi: 10.1101/gr.849004. PubMed PMID: 15173120; PubMed Central PMCID: PMCPMC419797.
49. Qi Q, Cheng L, Tang X, He Y, Li Y, Yee T, et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. Blood. 2021;137(10):1327-39. doi: 10.1182/blood.2020005780. PubMed PMID: 33512425; PubMed Central PMCID: PMCPMC7955410.
50. Tonelli A, Cousin P, Jankowski A, Wang B, Dorier J, Barraud J, et al. Systematic screening of enhancer-blocking insulators in Drosophila identifies their DNA sequence determinants. Dev Cell. 2025;60(4):630-45 e9. Epub 20241111. doi: 10.1016/j.devcel.2024.10.017. PubMed PMID: 39532105.
51. Moon H, Filippova G, Loukinov D, Pugacheva E, Chen Q, Smith ST, et al. CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep. 2005;6(2):165-70. doi: 10.1038/sj.embor.7400334. PubMed PMID: 15678159; PubMed Central PMCID: PMCPMC1299244.
52. Dorsett D. The Many Roles of Cohesin in Drosophila Gene Transcription. Trends Genet. 2019;35(7):542-51. Epub 20190523. doi: 10.1016/j.tig.2019.04.002. PubMed PMID: 31130395; PubMed Central PMCID: PMCPMC6571051.
53. Cockerill PN, Shannon MF, Bert AG, Ryan GR, Vadas MA. The granulocyte-macrophage colony-stimulating factor/interleukin 3 locus is regulated by an inducible cyclosporin A-sensitive enhancer. Proc Natl Acad Sci U S A. 1993;90(6):2466-70. PubMed PMID: 8460159.
54. Bowers SR, Calero-Nieto FJ, Valeaux S, Fernandez-Fuentes N, Cockerill PN. Runx1 binds as a dimeric complex to overlapping Runx1 sites within a palindromic element in the human GM-CSF enhancer. Nucleic Acids Res. 2010. Epub 2010/05/21. doi: gkq356 [pii]

10.1093/nar/gkq356. PubMed PMID: 20483917.
55. Bert AG, Burrows J, Osborne CS, Cockerill PN. Generation of an improved luciferase reporter gene plasmid that employs a novel mechanism for high-copy replication. Plasmid. 2000;44(2):173-82. PubMed PMID: 10964627.

56. Hawwari A, Burrows J, Vadas MA, Cockerill PN. The human IL-3 locus is regulated cooperatively by two NFAT-dependent enhancers that have distinct tissue-specific activities. J Immunol. 2002;169(4):1876-86. PubMed PMID: 12165512.

Supporting information

S1 Figure Locations of CTCF sites at histone H3K27me27 chromatin boundaries.

S2 Figure Unmanipulated original images used in figure 2

S3 Table Conservation of the IL3 insulator CTCF sites in mammalian species

S4 Table Conservation of the UPK2 intronic CTCF site in mammalian species

S5 Table Conservation of the HOXA9 chromatin boundary CTCF site in mammalian species

S6 Table Conservation of chromatin boundary CTCF sites in mammalian species

Figure 1

	2° upstream motif	1° core motif	References
Zinc fingers	11 10 9	7 6 5 4 3	[33]
Global CTCF	**CTGCAGTTCC**nnn--nn**TGGCCACCAGGGGGCGCC**		[41-43]
	G TTACGTT	C G AT AGT	
	C		
IL-3 +2.9 kb DHS	GCGGC**GGCTGCAG**TTTCTGG--AA**GGACCACTAGGGGGAGACA**		[11]
IL-3 +4.2 kb DHS	GAT**GCTGCTGCAG**TTTCTGG--AT**TGACCACTAGGGGGAGGTG**		[11]
Ch lysozyme F1	AAAA**GACATGTAAATACCAT-A**GCTATCCAGTAGAGGTCTCAA		[40, 46]
Ch HBB HS4	C**CCAGGGATGTAATTACGT**C-CC**TCCCCCGCTAGGGGGCAGCA**		[14, 16]
Human APP	CGAGC**GGGCGCAGTTCCC**CG-**GCGGCGCCGCTAGGGGTCTC**TC		[38, 39]
DM1 site 2	GAAAG**CTTTGCACTT**TGCG---**AACCAACGATAGGTGGGGGTG**		[37]
MHC-II XL9	CCAGG**CCCTGAAGAGCCCT-----GTTCCAGAGAATATGCTTG**		[21]
Human Myc	TAACTCGCTGTAGTAATTCC--AGCGAGAGGCAGAGGGAGCGA		[1, 40]
Mouse Myc	TGACTCGCTGTAGTAATTCC-ACGCGAGAGACAGAGGGAGTGA		[1]
THRE 144	CTCCTGAGTGCATTTCCCTC--ATGATCCAAAAGAGGGAATAA		[45]
Mouse HBA -94	CAGGAGGCTGCATTACCACA--GATGGACAGTAGAGGGAGACA		[13]
Mouse HBA -38	CCTGGTAGTGCTATGCACCC--TGAGGCCACCAGAGGGTAGCA		[13]
Mouse HBA +66	TGTGTGAATGCAGGTGCCTA-CAGTGGACAGAAGAGGGCGCCA		[13]
Human TCRa BEAD-A	CCCAGGCCTGCACTGCCGC-----CTGCCGGCAGGGGTCCAGT		[14]
Human UPK2 gene	GCAGCTTCTGCAGCCCAAGC----CTGCCACCTGGTGGTCATA		[1, 28]
HOXA9 boundary	GCTGAGGCTGCAGTACCAAAC-GGCGGCCAGCAGATGGCAGTG		[22, 23]
FoxJ3 boundary	AATTAGGCTGTAGTGAGTGC-CTGTGGCCACTAGGGGGCGGGG		[24]
SLC2A1 boundary	TGAGAACTTGCAGTACTGTA--GACAGACACCAGAGAGAGCAA		[47]
GLTP boundary	TGAGCCCCTGCAGCTCCCAC---GGGCCCGGTAGGTGGCGCTG		[47]
STKE17A boundary	CCGAAAGCTGCACCTTCTCCCCCTGAGCTGCAGAGGGCGCGCG		[47]
ABCG8 boundary	CAGGGTGCTGCAGTGGCACA-GACCAGCCACAGGATGGCAGTA		[47]
Weblogo of the above sites			[48]
Global consensus for upstream site			[42, 43]

Figure 2

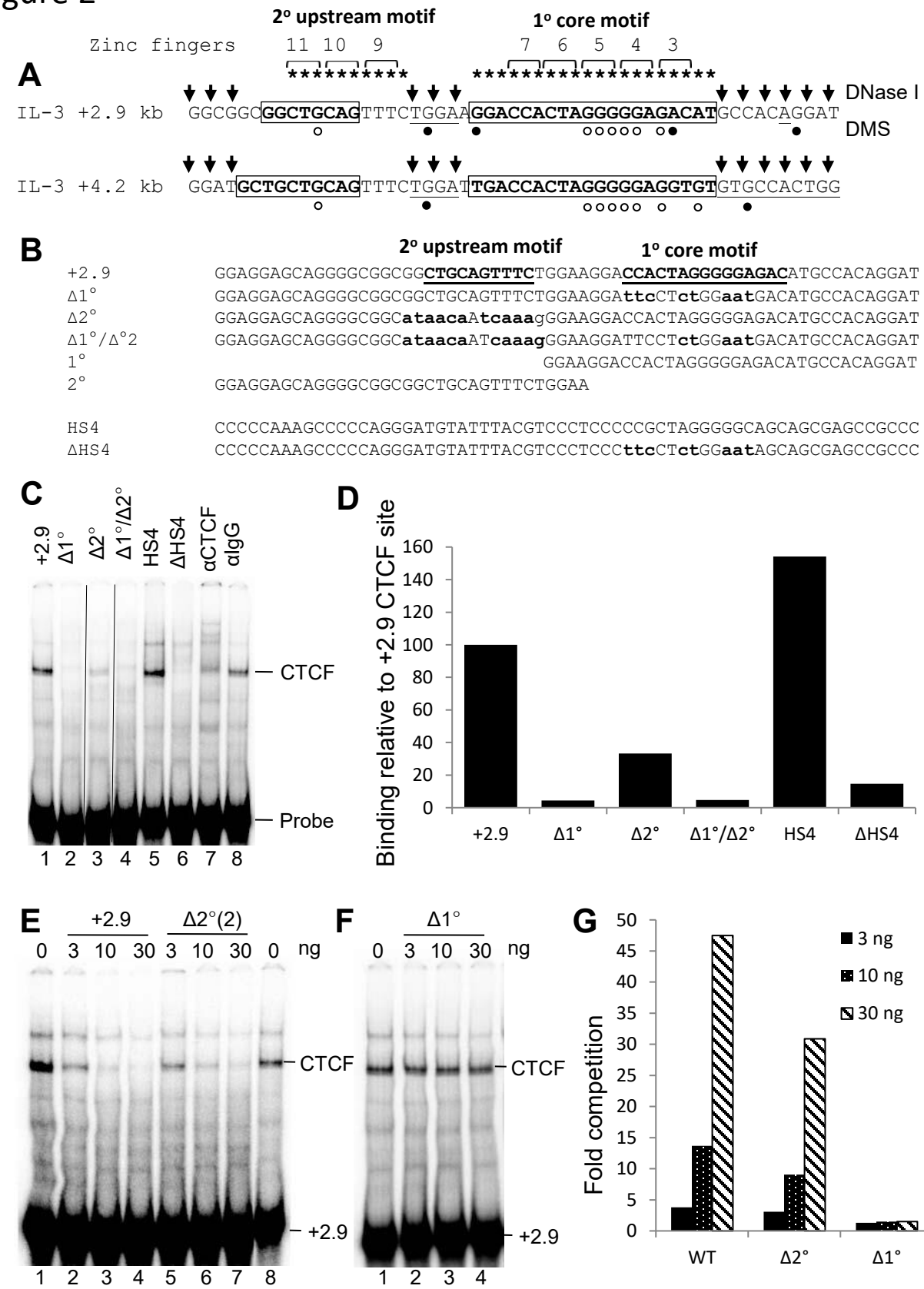
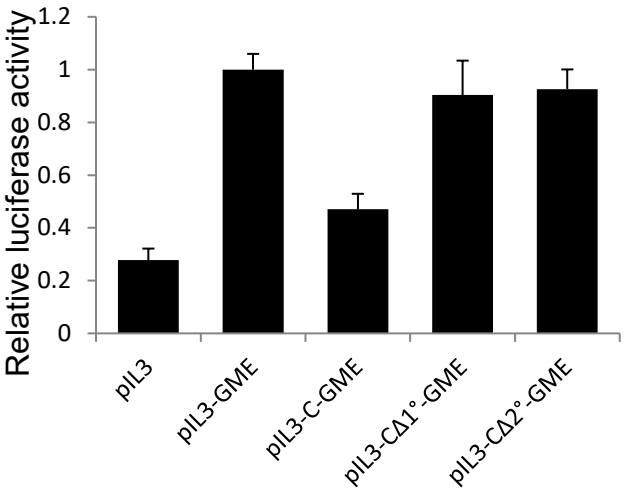
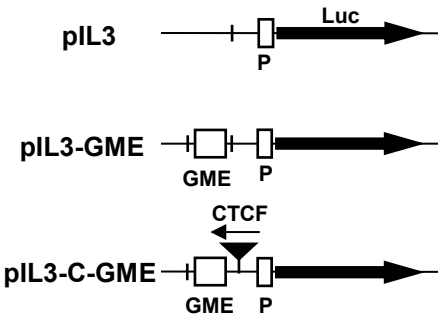

A
2° upstream motif
1° core motif
Zinc fingers 11 10 9 7 6 5 4 3
DNase I
DMS
IL-3 +2.9 kb GGCGGCGGCTGCAGTTTCTGGAAGGACCACTAGGGGGAGACATGCCACAGGAT
IL-3 +4.2 kb GGATGCTGCTGCAGTTTCTGGATTGACCACTAGGGGGAGGTGTGTGCCACTGG
B
2° upstream motif
1° core motif
+2.9 GGAGGAGCAGGGGCGGCGGCTGCAGTTTCTGGAAGGACCACTAGGGGGAGACATGCCACAGGAT
Δ1° GGAGGAGCAGGGGCGGCGGCTGCAGTTTCTGGAAGGATttcCTctGGaatGACATGCCACAGGAT
Δ2° GGAGGAGCAGGGGCGGCataacaAtcaaagGGAAGGACCACTAGGGGGAGACATGCCACAGGAT
Δ1°/Δ°2 GGAGGAGCAGGGGCGGCataacaATcaaagGGAAGGATTCCTctGGaatGACATGCCACAGGAT
1° GGAAGGACCACTAGGGGGAGACATGCCACAGGAT
2° GGAGGAGCAGGGGCGGCGGCTGCAGTTTCTGGAA
HS4 CCCCCAAAGCCCCCAGGGATGTATTTACGTCCCTCCCCCGCTAGGGGGCAGCAGCGAGCCGCCC
ΔHS4 CCCCCAAAGCCCCCAGGGATGTATTTACGTCCCTCCCttcCTctGGaatAGCAGCGAGCCGCCC
C
+2.9
Δ1°
Δ2°
Δ1°/Δ2°
HS4
ΔHS4
αCTCF
αIgG
CTCF
Probe
1 2 3 4 5 6 7 8
D
Binding relative to +2.9 CTCF site
160
140
120
100
80
60
40
20
0
+2.9
Δ1°
Δ2°
Δ1°/Δ2°
HS4
ΔHS4
E
+2.9
Δ2°(2)
0 3 10 30 3 10 30 0 ng
CTCF
+2.9
1 2 3 4 5 6 7 8
F
Δ1°
0 3 10 30 ng
CTCF
+2.9
1 2 3 4
G
Fold competition
50
45
40
35
30
25
20
15
10
5
0
3 ng
10 ng
30 ng
WT
Δ2°
Δ1°

Figure 3





	2° motif	1° motif
C	GGGGCGGCGG**CTGCAGTTTC**TGGAAGGA**CCACTAGGGGGAGA**CATGCCACAGGA	
CΔ1°	GGGGCGGCGGCTGCAGTTTCTGGAAGGA**ttc**CT**ct**GG**aat**GACATGCCACAGGA	
CΔ2°	GGGGCGGC**ataaca**A**tcaaa**gGGAAGGACCACTAGGGGGAGACATGCCACAGGA	