CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide.

Igor Chernukhin, Shaharum Shamsuddin, Sung Yun Kang, Rosita Bergström, Yoo-Wook Kwon, Wenqiang Yu, Joanne Whitehead, Rituparna Mukhopadhyay, France Docquier, Dawn Farrar, et al.

CTCF INTERACTS WITH AND RECRUITS THE LARGEST SUBUNIT OF RNA POLYMERASE II TO CTCF TARGET SITES GENOME-WIDE

Igor Chernukhin[#], Shaharum Shamsuddin[5#], Sung Yun Kang[#], Rosita Bergström[1], Yoo-Wook Kwon[2], WenQiang Yu[1], Joanne Whitehead[1], Rituparna Mukhopadhyay[1], France Docquier, Dawn Farrar, Ian Morrison, Marc Vigneron[3], Shwu-Yuan Wu[4], Cheng-Ming Chiang[4], Dmitri Loukinov [2] , Victor Lobanenkov [2], Rolf Ohlsson[1], Elena Klenova[*]

Running title: CTCF interacts with RNA Polymerase II genome-wide

Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK

[1] Department of Development & Genetics, Evolution Biology Centre, Uppsala University, Norbyvägen 18A, S-752 36 Uppsala, Sweden; [2] Molecular Pathology Section, LIP NIAD NIH, Building 7, Room 303, MSC-0760, Bethesda, MD 20892-0760, USA [3]; CR1 INSERM, UMR 7100 CNRS-ULP, E.S.B.S., Boulevard Sébastien Brant, BP 10413, 67412 ILLKIRCH, Cedex, France; [4] Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106-4935, USA; [5] Present address: School of Health Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia.

[#] These authors contributed equally to this work.

[*] Address correspondence to: Elena Klenova, Department of Biological Sciences, Central Campus, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ UK

e-mail: klenovae@essex.ac.uk, fax: +44(0)1206-872592, tel: +44(0)1206-874868

Abstract

CTCF is a transcription factor with highly versatile functions varying from gene activation and repression to regulation of insulators function and imprinting. Although many of these functions rely on CTCF-DNA interactions, it is an emerging realization that CTCF-dependant molecular processes involve CTCF interactions with other proteins. In this study we report association of a subpopulation of CTCF with the RNA Polymerase II (Pol II) protein complex. We identified the largest subunit of Pol II (LS Pol II) as a protein significantly co-localizing with CTCF in the nucleus and specifically interacting with CTCF *in vivo* and *in vitro*. The role of CTCF as a link between DNA and LS Pol II has been reinforced by the observation that LS Pol II association with CTCF –target sites *in vivo* depends on intact CTCF binding sequences. "Serial" chromatin immunoprecipitation analysis revealed that both CTCF and LS-Pol II were present at the β-globin insulator in proliferating HD3 cells, but not in differentiated globin synthesizing HD3 cells. Further, a single wild type CTCF-target site (N-Myc-CTCF), but not the mutant site deficient for CTCF binding, was sufficient to activate transcription from the promoterless reporter gene in stably transfected cells. Finally, a ChIP on ChIP hybridization assay using microarrays of a library of CTCF target sites revealed that many intergenic CTCF target sequences interacted with both CTCF and LS Pol II. We discuss possible implications of our observations with respect to plausible mechanisms of transcriptional regulation via CTCF-mediated direct link of LS Pol II to the DNA.

INTRODUCTION

CTCF, or CCCTC - binding factor, is an 11 Zn-finger transcription factor with highly versatile functions and a candidate tumour suppressor (30, 42). CTCF is localized to the nucleus and it is ubiquitous and highly conserved. CTCF regulates transcription in diverse modes, such as promoter activation and repression, silencing, constitutive and methylation dependent chromatin insulation; CTCF also organizes epigenetically controlled chromatin insulators that regulate imprinted genes in soma (30, 42). The characterized genes regulated by CTCF include *c-Myc* (16, 31), chicken lysozyme (7), *BRCA1* (8), *hTERT* (49), *IRAK2* (35), amyloid beta-protein precursor (*APP*) (63) and others genes (42). Among vertebrate insulators controlled by CTCF are the *β*-globin (3) and the *H19* imprinting control region (ICR) (42, 44) insulators. In our previous report the number of CTCF binding sites in the mouse genome was estimated as ~ 4000 (40), but the real number may be much higher (~ 30,000 in the human genome), as suggested in a more recent publication (62). Many of these sites are methylation sensitive and map to promoter, inter- and intragenic regions, and introns; some sites contain Alu-like repeated elements CTCF targets (40, 62).

Post-translational modifications of CTCF were found to be involved in the regulation of CTCF function(s). Thus specific phosphorylation of CTCF by the protein kinase CK2 (former casein kinase II) affects CTCF function in transcriptional regulation (15, 29). Poly(ADP-ribosyl)ation is another recently discovered modification of CTCF that is important for insulator function (27, 68) and nucleolar transcription (61). Post-translational modifications of CTCF have also been implicated in human myeloid cell differentiation (14).

CTCF association with other proteins is also important for regulation of CTCF-dependant molecular processes. Thus CTCF interactions with sin3 (37) and YB-1 (10, 28) are shown to modulate CTCF function as a transcriptional repressor. Cooperation of CTCF with nucleophosmin (69), Kaiso (13) and helicase protein CHD8 (22) have been linked to the control of insulator function of CTCF and epigenetic regulation.

In this report we describe interaction of CTCF with RNA Polymerase II (Pol II). The eukaryotic Pol II enzyme transcribes all protein-coding genes and also non-coding regulatory RNAs (e.g. snRNA, miRNA) (52). The Pol II enzyme is composed of 12 subunits (termed Rpb1–Rpb12) (67). The Rpb1, the largest subunit of Pol II (LS Pol II), is highly conserved among eukaryotic RNA polymerases. Its characteristic feature is the carboxyl-terminal domain (CTD), which contains multiple copies of a heptapeptide repeat Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The CTD can be modified by phosphorylation which results in the appearance of two forms of LS Pol II: hypophosphorylated (LS Pol IIa), migrating at 220 kDa and hyperphosphorylated (LS Pol IIo), migrating at 240 kDa. The LS Pol IIa has been associated with the initiation complex, whereas LS Pol IIo has been found in elongating complexes (12).

Accurate initiation of transcription by Pol II can be directed by the TATA box, INR and possibly other less characterized promoter elements. The mechanisms of TATA-mediated transcription initiation are very well understood. The TATA-binding protein (TBP) subunit of the TFIID complex is necessary for the recognition of the TATA-box and accurate initiation of transcription by Pol II (19, 58). Very little is however known at present about the mechanisms of transcription initiation mediated by other promoter elements, in particular proteins that recognize these elements and aid Pol II (19).

The views on how the transcriptional machinery is assembled and targeted to specific promoters do not harmonize. Thus, a "stepwise assembly" model proposes a coordinated step-by-step recruitment of the proteins in the transcription pre-initiation complex. The alternative "pre-assembly complex" model suggests the recruitment of a pre-assembled Pol II complex for transcription initiation (36). In both models, general transcription factors are required to form a stable initiation complex at promoters, and mediators and coactivators are necessary to communicate signals from transcriptional activators and repressors (39).

In this report we describe the association of a subpopulation of CTCF with the Pol II protein complex. A component of this complex, the LS Pol II protein, has been identified as a

protein interacting partner with CTCF. We demonstrate that CTCF is associated *in vivo* with LS Pol II at the selected known CTCF target sequences (CTSs). Furthermore, we show that a single CTCF-binding site is sufficient to activate transcription of the reporter gene in a stably transfected cell line, which is likely to occur through interaction between CTCF and LS Pol II. Analysis of CTCF and LS Pol II interactions genome-wide indicates that CTCF may recruit LS Pol II to a certain subpopulation of CTSs. These findings may provide a basis to link the transcriptional machinery directly to CTSs on the DNA with various potential functional implications.

MATERIALS AND METHODS

Cell lines, stable transfections, luciferase assay. Human HeLa (cervical carcinoma), MCF7 (breast carcinoma) and K562 (myeloid leukaemia) cells were maintained in RPMI 1640 medium supplemented with HEPES, GlutaMAX, sodium bicarbonate, 50μg/ml gentamicin, 10% foetal calf serum (all from Life Technologies). The chicken erythroblast cells HD3 were grown in DMEM medium, supplemented with 50μg/ml gentamicin, 8% foetal calf serum and 2% chicken serum. Cells were induced to differentiate according to the protocol previously described by Nicolas et al .(41). Briefly, 10^7 cells from a logarithmically growing culture were plated at $1x10^6$ cells /ml in the above medium containing 10mM HEPES, pH8 and 20 μM protein kinase C inhibitor H7, and incubated at 42^0C for 2 days. Staining with benzidine (5) was performed to confirm differentiation; cultures with > 80% of benzidine-positive cells were used in the experiments. The human choriocarcinoma JEG-3 cells were propagated as described previously (17). DNA transfection into JEG-3 cells was performed using the calcium phosphate method (51); 10^6 cells were transfected with 5 μg of plasmid DNA in 10cm plates. NIH 3T3 mouse fibroblasts were maintained in DMEM supplemented with 10% donor serum and 50μg/ml gentamicin. Growth arrest of NIH 3T3 was induced by serum starvation (0.05%) for 48 hours.

To generate p N-Myc or deficient for CTCF-binding p N-Sac-Myc mutant plasmids (pN-Myc-Luc wt and pN-MycLuc mut) for stable transfection in NIH 3T3 cells, the 90 bp N-Myc and N-

Myc-Sac mutant fragments were excised by Hind III from pBend-N-Myc and pBend-N-Myc-Sac, respectively (38). The fragments were then cloned into the Hind III digested dephosphorylated pGL2-basic promoter-less plasmid. For stable transfections, the FuGENE 6 reagent (Roche Applied Science) was used. Three μg of pMyc- N or mutant pMyc- N-Sac were mixed with 0.3 μg of pCIIN containing the neo-resistance selection marker and 10^6 cells were used for each transfection. For selection of transfectants cells were incubated with 500 μg/ml of G-418 for two weeks, followed by sub-cloning of single cells. The colonies found to be positive in a luciferase assay were pooled and grown as a mass culture.

Luciferase assay. For Luciferase Reporter Assay cells were grown in six-well plates for 24 hours; the luciferase activity was measured using the Luciferase Reporter Assay System kit (Promega) according to the manufacturer's instructions. The luciferase activites were normalized to transgene copy number, estimated by comparing the band intensities of the Cla I/ Xho I genomic fragments from pN-Myc-Luc wt and pN-MycLuc mut cells (see Southern blot procedures below). All assays were performed in triplicate.

Immunoprecipitations (IP). HeLa cells ($2x10^6$) were collected and washed twice with ice-cold PBS and lysed in 500μl of the high salt RIPA buffer, as previously described (10, 26). A panel of antibodies at a final concentration of 5 μg/ml was used for IP. They included: rabbit polyclonal antibodies anti- LS Pol II (N-20), Sp1 (PEP2), TBP (SI-1), all from Santa Cruz; Monoclonal Antibodies anti - LS Pol II (8WG16) raised against the 220 kDa form of Pol II (hypophosphorylated, Pol IIa) and anti-LS Pol II (H14) raised against the 240 kDa form (hyperphosphorylated, Pol IIo), both from Covance Research Products; the anti - histone H2A and - histone H3 antibodies were kindly provided by S. Muller and J. Dadoune. The anti-CTCF rabbit polyclonal antibodies were raised against the bacterially expressed CTCF N-terminal domain and are able to recognize CTCF in different species. We also used mouse monoclonal antibody Rb1

(Ab-6) and p53 (Ab1 through 6 from the sampler kit) from Oncogene Research Products. For neutralization of the antibody we followed the protocol developed at Santa Cruz. The 5 fold excess of the N-20 peptide (250 μg/ml) was added to 100 μl of the anti-LS Pol II antibody diluted to 50mg/ml in 1xPBS and incubated for 2 hours at room temperature. The blocked antibodies were then used at 5 μg/ml in co-IP. In some experiments, RNase-free DNase I (Roche Diagnostics) was added to the extracts at 300 U /ml for 50 min at 32°C prior to immunoprecipitation.

CTCF expression in the baculovirus system. The CTCF – producing recombinant SF9 cells were grown following the manufacturer's instructions (BacVector System Manual, Novagen). The CTCF protein (termed "baculoCTCF") was purified to 80-90% purity from infected SF9 cells using Ni-affinity chromatography with a linear gradient of imidazole for elution and subsequent gel filtration on an S-200 column.

Purification of the Pol II and TFIIH complexes. The Pol II and TFIIH complexes were prepared from HeLa cells as previously described (25, 65, 66).

Expression of the His-tagged N-terminal, Zn-finger and the C- terminal domains of CTCF in the bacterial system. Preparation of the vectors expressing the His-tagged N-terminal, Zn-finger, C-terminal domains of CTCF was described in detail in our previous report (10); their detailed maps are available on request. To generate proteins in a bacterial system, transformants carrying the plasmids expressing the His-tagged N-, C -terminal, Zn-finger and the C- terminal domains of CTCF, were grown in LB media supplemented with ampicillin (50 μg/ml) for 3 h at 37 °C. Protein expression was induced by the addition of 0.4 mM isopropyl 3-D-thiogalactoside with further incubation for 3 h at 37 °C. For purification of each of the desired protein, the bacterial cells were collected by centrifugation, washed twice with 0.1 volume of cold phosphate-buffered saline followed by lysis in 0.1 volume of the original culture in the cold freshly prepared lysis buffer (8 M urea, 0.1 M

NaH_2PO_4, 0.01 M Tris-HCl, pH 8.0). The lysates were then subjected to immobilized metal ion affinity chromatography for further purification. For this purpose the total bacterial lysates were supplemented with 20 mM imidazole, then loaded onto the nickel-charged His-Bind resin (R&D Systems, Europe Ltd.), washed with one bed volume of the washing buffer (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, pH 8.0, and 20 mM imidazole) and, finally, eluted with 10 ml of the elution buffer (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, pH 8.0, and 0.5 M imidazole).

Production of GST- LS Pol II (bactPol II) from E. coli. The construct for expression of LS Pol II in *E.coli* contained the cDNA of LS Poll II (hRPB1) cloned as a GST fusion using bacterial-IPTG inducible pGEX-2T vector (1). To produce the GST-LS Pol II protein in *E. coli* DH 5α , we used the standard protocols (54, 64) with some modifications to solubilize bactPolII, which was extracted from the inclusion bodies by incubation in HEPES/Guanidine buffer (50 mM HEPES pH 7.5, 6 M Guanidine HCl, 25 mM DTT), renaturing in ice-cold folding buffer (50 mM HEPES pH 7.5, 0.2M NaCl, 1 mM DTT, 1M NDSB201 [3(1-pyridinio)-1-propane sulfonate (Fluka Chem.), followed by dialysis against 100 volumes of the dialysis buffer (50 mM HEPES pH 7.5, 0.2 M NaCl, 1 mM DTT and 10% Glycerol) at +4^{o}C overnight.

Immobilization of the bacterially expressed proteins on the matrix. To generate matrices for interaction assays, the *in vitro* produced proteins were immobilized onto cystamine-coupled Sepharose 4B. Cystamine was first converted into aminoethylthiol after a reduction reaction with 50 mM dithiothreitol in TE buffer (50 mM Tris-HCl, 1 mM EDTA, pH 8.3) for 30 min at room temperature, then treated with 5 mM 2,2-dipyridyldisulfide for 2 hrs. The activated matrix was washed with the TE buffer. Each protein was reduced by incubation with 5 mM dithiothreitol for 1 hr at room temperature, desalted through a G50 column equilibrated with TE, and then incubated with the activated matrix overnight at +4 °C (protein to Sepharose v/w ratio was 5 mg/1 ml). The amounts of protein retained on the matrix were monitored by protein assay (Bio-Rad), according to

the manufacturer's instructions. The protein-Sepharose conjugates were finally washed with the TE buffer to remove non-incorporated materials and stored in the buffer containing 20% glycerol, 50 mM KH_2PO_4, pH 7.0, and 0.2% Na_3N.

The interaction (pull down) assay. Fifty µl of the Sepharose suspension carrying each of different of CTCF protein domains produced in E.coli, or BSA as control, were mixed with 1 ml of K562 cell lysate in 0.25 M RIPA buffer and incubated for at least 6 hrs on a rotating platform. Each suspension was then washed six times with 0.25 M RIPA buffer, boiled in SDS sample-loading solution for 5 min, run on 10% SDS-PAGE, and analyzed by Western blot assay. The presence of LS Pol II was detected with the anti-LS Pol II (N-20) antibody.

Western blot analysis. Protein were resolved, blotted and visualized as described in Chernukhin et al (10). The primary anti-CTCF antibody was used at 1:300 dilution, the anti-LS Pol II antibodies at 1:100 (1µg/ml final concentration) and the anti-His tag monoclonal antibodies (Sigma) at 1:500. The secondary anti-rabbit-peroxidase (Abcam) or anti-mouse-peroxidase conjugated antibodies (Abcam) were used at 1:10,000 dilutions. Before re-probing, a membrane was stripped in a buffer containing 100 mM mercaptoethanol, 2% SDS 62.5 mM Tris HCL, pH 6.7, incubated twice at 55^{o}C 15 min, then rinsed three times for 15 min at room temperature in PBS (pH 7.5) supplemented with 0.1% Tween-20. Quantification of the bands was performed by using the Image J software (http://rsb.info.nih.gov/ij/).

Mass-spectrometry. The protein bands were subjected to in gel digestion and peptide mass fingerprint analysis using MALDI-TOF (Bruker Daltonics Reflex 4) was performed as described in Chernukhin et al (10). The obtained spectra were interpreted using Bruker Daltonics flexAnalysis 2.0 software and the sequence retrieval was done with Mascot Peptide Mass Fingerprint (http://www.matrixscience.com) using SWISS-Prot database.

Chromatin immunopurification (ChIP) assay and "Serial" ChIP assays. Harvested cells were cross-linked with formaldehyde according to Kuo and Allis (34) and the DNA concentration was adjusted to 100μg/ml. DNA-protein complexes were immunopurified using anti-CTCF or anti-Pol II antibodies (N-20, Santa Cruz) and protein A 4 Fast Flow Sepharose beads (Sigma).

The "serial" ChIP assay is a modification of a standard ChIP assay designed to assess the DNA occupancy by two protein molecules simultaneously. In this assay, the formaldehyde cross-linked DNA-protein complexes are first passed through the matrix linked with the antibody against one of the partner proteins; after elution, the retained complexes are subjected to the subsequent IP with the second partner antibody. The resulting complexes therefore contain DNA fragments associated with the two partner proteins. The matrices containing the conjugated anti-CTCF, anti-LS Pol II antibody (N-20) or pre-immune serum were prepared, as previously described (9). The DNA purified from ChIP assays was measured and 1-10 μl of the DNA was used for PCR amplification. The primers and conditions for PCR are described in Table 1. Full protocols for ChIP and "serial" ChIP assays used in these experiments are available on request.

ChIP on ChIP analysis. ChIP samples were prepared from $5x10^6$ NIH 3T3 cells as described in the precious section. A ChIP on chip hybridization assay was performed using microarrays of a library of CTCF target sites derived from a ChIP of mouse fetal liver (40). Briefly, the ChIP samples were amplified with SR1/SR2 primers (40) and PCR-labeled with Cy3/Cy5 dyes (Amersham) using the nested primers. Labeled targets were purified with Qiaquick PCR purification kit (Qiagen) and eluted in hybridization solution (GlassHyb™ Hybridization solution, Clontech). The Cy3 and Cy5-labeled targets were denatured and incubated at 45°C for 1 hour in the presence of 100 μg of Cot-1 DNA (Clontech). The samples were pooled and hybridized to borohydride-pretreated slides. Following a washing procedure according to GlassHyb™

Hybridization Solution kit User Manual (Clontech), the slides were scanned using a Scanarray 4000 and analysed with ScanarrayExpress 3.0 (Packard Biosciences).

Indirect immunofluorescence and analysis of colocalization. For indirect immunofluorescent staining, the original protocol of Harlow and Lane (20) was used with an additional modification (53). Cells were incubated at +4^{o}C overnight with anti-CTCF rabbit-polyclonal antibody (Abcam) (dilution 1:5) and the anti-Pol II CTD monoclonal antibody 7G5 (4) (dilution 1:50), followed by the subsequent incubation with the secondary antibodies: swine anti-rabbit FITC (Dako) and rabbit anti-mouse TRITC (Dako), both diluted 1:50. The cells were visualized using Confocal Laser Scanning microscopy (Bio-Rad). Images were obtained using a Bio-Rad Radiance 2000 confocal unit on an Olympus IX70 microscope. The fluorescein labels were illuminated by the 488nm laser line and detected via a 500-560nm bandpass filter, while the rhodamine probes were excited by a 543nm laser with a 570nm longpass filter; transmission images were also recorded. Colocalization of the two probes was analyzed by in-house software, using the methods of Costes et al (11) to estimate the background intensity. Bleed-through of rhodamine signals into the fluorescein images was measured to be 17.5% using single-labelled samples, and was corrected for in the colocalization analyis.

Southern Blot Procedures. DNA was extracted from NIH3T3 stable transfectant cells using the DNeasy tissue kit (Qiagen) and digested with Cla I/Xho I restriction enzymes followed by electrophoresis in a 0.8% agarose gel. DNA was then transferred to Hybond-N+ membrane (Amersham Biosciences). Blots were probed with a ^{32}P-labeled luciferase cDNA probe (Cla I-Xho I fragment) synthesized using a random priming labeling kit (Roche Applied Science). Membranes were hybridized at 68°C for 4–6 h in a buffer containing 0.5 M sodium phosphate, pH 6.8, 1 mM EDTA, 7% SDS, and 0.2 mg/ml herring sperm DNA. Following hybridization, the membranes were washed twice for 10 min in a 5% SDS, 0.04 M sodium phosphate, pH 6.8, 1 mM EDTA

solution and then four times for 10 min in the same solution containing 1% SDS and exposed to the X-film (Kodak) for 24 hrs. Quantification of the bands was performed using the ImageQuaNT 5.0 software program. Similar blotting and hybridization procedures were used to analyze PCR products obtained in a ChIP experiments from HD3 cells. The primers and conditions for PCR are described in Table 1.

RESULTS

CTCF is a component of the Pol II protein complex. Potential CTCF-Pol II interactions were first hinted at when one of the CTCF binding sites in human and mouse *MYC* promoters were found to map precisely within the region of Pol II pausing and release (33, 56) (CTCF site A, Figure 4A). In our ensuing experiments to isolate proteins interacting with CTCF, affinity chromatography on a matrix with immobilized purified recombinant CTCF was employed. Routinely, a doublet of two proteins of about 200kDa and 240kDa, reminiscent of two differentially phosphorylated forms of the LS Pol II (LS Pol IIa and LS Pol IIo) was retained by CTCF from nuclear extracts of different cellular origins (10). Based on these observations, we hypothesized that CTCF could be a part of the Pol II complex and interact with the large proteins from this complex, such as LS Pol II. This supposition was further examined by biochemical analyzes.

First we tested whether CTCF may be a part of the Pol II protein complex. In these experiments, the Pol II complex was purified from a cell line conditionally expressing the FLAG-tagged RPB9 subunit of human Pol II (65) and the TFIIH complex was obtained from a cell line conditionally expressing the p62 subunit of human TFIIH (25). These complexes had been previously purified and characterized (66); the same preparations were used in this study. When purified the Pol II and TFIIH complexes were resolved by SDS-PAGE and then subjected to a Western blot analysis with the anti-CTCF antibody, the band specific for CTCF appeared in the Pol II complex, but not in the TFIIH complex (Figure 1A) thus confirming that this association is

specific for the Pol II complex. The amount of CTCF associated with the Pol II protein complex in cell extracts is relatively small. This indicates that only a proportion of CTCF in the nucleus may exist in a complex with Pol II or/and this interaction *in vivo* is not strong, with CTCF being lost from the complex in a process of lengthy purification. The former explanation is consistent with the partial overlap of CTCF and Pol II staining in the K562 and HeLa cells' nuclei (Figure 2), however the latter can not be ruled out.

CTCF interacts with the largest subunit (LS) of Pol II in vivo. Next we carried out a series of co-IP assays with a panel of antibodies against proteins known to be associated with the Pol II protein complex and also proteins known to form functional interactions important for transcriptional regulation. Figure 1B shows that while the anti-LS Pol II and anti-YB-1 antibodies were able to co-IP CTCF from cell extracts, CTCF was absent from the complexes precipitated with the anti-TBP, Sp1, Rb1, histone H2A and histone H3 antibodies. No CTCF was observed with any of six anti-p53 antibodies (Figure 1B and data not shown). Our earlier study revealed no association between CTCF and other nuclear factors, such as p21, the ubiquitous nuclear receptor UR, thyroid receptor TRα, hTAF $_{II}$130, and MYC (10). The CTCF- LS Pol II association was abolished when the peptide N-20, originally employed to raise the anti-LS Pol II antibodies, was pre-incubated with the anti-RNA-Pol II antibody (Figure 1B).

CTCF was also co-immunoprecipitated by the anti - LS Pol II (8WG16) that recognizes predominantly the hypophosphorylated LS Pol II (LS Pol IIa) and the anti-LS Pol II (H14) that is specific only to the hyperphosphorylated LS Pol II (LS Pol IIo) (6) (Figure 1C). Interestingly, CTCF was 1.6 fold more efficiently co-immunoprecipitated with the anti –LS Pol IIa, than with the anti-LS Pol IIo antibody. Treatment with DNAse has not significantly changed the interaction between CTCF and LS Pol II (96%), which rules out possible contamination by chromatin fragments (Figure 1C). On the other hand, interaction between CTCF and YB-1 decreased to a higher degree (63%).

The specificity of this association was further corroborated by our observation that the anti-CTCF antibody co-immunoprecipitated LS Pol II from HeLa cell extracts, while there was no LS Pol II band in co-immunoprecipitates from the pre-immune serum (Figure 1D). Notably, in cell lysates the hypophosphorylated form of Pol II (Pol IIa) was precipitated more efficiently, thus reciprocating previously made observations (Figure 1C). This may reflect the nature of the interaction between CTCF and LS Pol II *in vivo* when post-translational modifications of CTCF and LS Pol II or presence of other proteins in the complex may be important in the establishment of the specific protein association. Our preliminary results indeed show that phosphorylation of CTCF results in decreased binding to LS Pol II (I. Chernukhin and S. Shamsuddin, unpublished data). The results similar to described above showing interaction between CTCF and LS Pol II were obtained when the lysates from other cell lines (K562, NIH 3T3) were used in co-IP experimentations (data not shown).

The two proteins sized ~ 220 kDa and ~ 240 kDa were also observed after a preparative immunoprecipitation with the anti-CTCF antibody (data not shown). These bands were excised, subjected to the "in-gel" digestion and peptides were analyzed by MALDI-TOF-MS. The database interrogation revealed the presence of peptides matching the DNA-directed RNA polymerase II largest subunit, RPB1 (Swiss-Prot protein data base accession number P24928). These data complement the co-IP results showing that the LS Pol II is a proteins interacting with CTCF.

CTCF-LS Pol II interaction is mediated via the C-terminal domain of CTCF and is direct. To define which portion of the CTCF protein is involved in the interaction with LS Pol II, the three His-tagged domains of CTCF (N-, Zn - and C; a map is shown in Figure 1G) were generated in a bacterial system, coupled to the matrix and employed in the interaction assay. Figure 1E shows that the K562 cell-derived LS Pol II interacts with the CTCF C-terminal domain immobilized on the matrix. As in Figure 1C, two bands of the LS Pol II, hypophosphorylated LS Pol IIa and hyperphosphorylated LS Pol IIo, were observed. However, in this case, both bands were retained

efficiently by the C-terminal domain of CTCF, which may be due to the absence of the post-translational modifications in the CTCF-C. No LS Pol II was seen with the CTCF-N, CTCF-Zn or BSA. Equal loading of the proteins was verified by the subsequent probing of the membrane with the anti-His tag antibodies (Figure 1F).

The directness of the association between CTCF and LS Pol II was further confirmed in the interaction assay between the *in vitro* generated CTCF and LS Pol II. In this experiment, the full length recombinant CTCF produced in the baculovirus system (baculoCTCF) was immobilized on the matrix, whereas the LS Pol II protein produced in E.coli (bactLS Pol II) was present in solution. As shown in Figure 1H, in this experiment, bactLS Pol II was retained after incubation of the lysate containing bactLS Pol II with the matrix conjugated with baculoCTCF. On the other hand, no bactLS Pol II was observed in the control experiment when the matrix contained immobilized BSA.

CTCF and LS Pol II are significantly colocalized in the nucleus. Interaction between CTCF and LS Pol II was confirmed by imaging techniques such as immunofluorescent staining using the anti-CTCF polyclonal antibody and anti- LS Pol II monoclonal antibody. For immunofluorescent staining, HeLa and K562 cell lines were chosen because of their difference in CTCF distribution, uniform in HeLa and patchy in K562. The staining revealed that CTCF and LS Pol II proteins are significantly co-localized in the nucleus (typical images are shown in Figure 2). This was confirmed by further analysis of the merged images shown in the two-dimensional histograms of the fluorescence (Figure 2, right). Signals were corrected for background and bleed-through (11), and the Pearson coefficient found; the average for three sets of images of HeLa cells was 0.83 and for three sets of images of K562 cells was 0.85, which shows good correlation between the LS Pol II and CTCF staining patterns. However, these results also indicate that there are pools of CTCF and LS Pol II which are not co-localized and therefore may not be involved in the interaction.

Analysis of the in vivo distribution of CTCF and LS Pol II at the chicken β-globin insulator in proliferating and differentiated HD3 cells. To explore the possibility that the interaction of DNA-bound CTCF and LS Pol II occurs *in vivo* at the β-globin insulator (site FII) (Figure 3B) and may be important for regulation of gene activity, we chose the erythroblast cell line, HD3. In proliferating HD3 cells, the globin genes are inactive and in differentiated HD3 cells, ρ-and β^A –globin genes are transcriptionally active (41), prompting a hypothesis that LS Pol II might be held by CTCF at the insulator in non-globin synthesized cells. To assess the simultaneous presence of CTCF and LS Pol II at CTCF-binding sites, we developed the standard ChIP assay into a modified version, which we termed the "serial" ChIP assay. In this assay, two subsequent IP reactions of formaldehyde cross-linked DNA-protein complexes were performed to specifically precipitate the DNA-CTCF-LS Pol II complexes. The input samples were first passed through the matrix with the immobilized anti-CTCF antibodies, or with the pre-immune serum as a control. The advantage of using the anti-CTCF antibody covalently bound to the matrix was that only retained protein complexes, but not the unwanted free antibodies, were anticipated after the first IP. After a subsequent IP with the anti-LS Pol II antibody, the composition of the resulting complexes was expected to be DNA-CTCF-LS Pol II. The fragment sizes of the sonicated DNA are on average 300-400bp, hence it is very unlikely that CTCF would be associated with the same DNA fragment as LS Pol II, unless they form a protein-protein complex, because the β-globin FII site is located more than 10 kb upstream from the first transcription start site (50) (Figure 3B).

As shown in Figure 3A, in proliferating HD3 cells binding of CTCF to the DNA at the β-globin insulator can be detected after ChIP with anti-CTCF and anti LS Pol II antibodies. Importantly, the globin insulator could be specifically amplified in the samples subjected to the "serial" ChIP. The interpretation of this result is that LS Pol II can interact with the β-globin insulator either directly or indirectly via CTCF in these cells. The level of β-globin sequences precipitated with the anti- LS Pol II antibody is lower than with the anti-CTCF (Figure 3A); this can be explained by relative inefficiency of formaldehyde in protein-protein cross linking in HD3 cells.

In differentiated HD3 cells, no signal was detected in samples precipitated with the anti-CTCF, anti-LS Pol II antibodies and also after the "serial" ChIP suggesting that CTCF and LS Pol II were no longer associated with the β-globin insulator after induction of HD3 cells. In both, non-differentiated and differentiated HD3 cell, no amplification was seen when the pre-immune serum (PS) was used for precipitation and when primers from a region from exon 8 of the chicken CTCF gene lacking CTCF binding sites were employed for PCR. The same ChIP samples were subjected to amplification with primers designed to overlap the TATA-box within the promoter region of the chicken β–actin gene, which served as a control for LS Pol II loading. The PCR products could only be seen in the samples precipitated with the anti-LS Pol II antibodies prepared from both proliferating and differentiated HD3 cells.

From these data we conclude that CTCF and LS Pol II epitopes colocalize to the β-globin insulator, despite the absence of any known transcriptional unit at this domain.

Association of the LS Pol II to the H19 ICR requires functional CTCF target sites. The *H19* imprinting control region (ICR) is devoid of any promoter except for the *H19* promoter separated from the *H19* ICR by 2 kb (24) (Figure 3D). To ascertain that the CTCF-LS Pol II signal depended on the CTCF target sites within the *H19* ICR, we used previously described vectors (68) carrying a 1,2 kb region of the *H19* ICR. This region contains CTCF target sites 3 and 4 in a wild type or with mutations in the CTCF binding sites, in a pGEM vector devoid of any known eukaryotic regulatory cis elements. We have previously characterized these mutations in the *H19* ICR (44) and shown that they abolished CTCF binding *in vivo and in vitro* (43, 44). The vectors with the wild type and mutant *H19* ICR were mixed in equimolar amounts and transfected in JEG-3 cells followed by ChIP analysis using CTCF or LS Pol II antibodies. The ChIP material was subsequently amplified and analyzed by using a diagnostic Eco RV restriction site which is present only in the mutated *H19* ICR allele (Figure 3D). These experiments revealed that using CTCF target site 3 as a diagnostic marker the sequences pulled down by the CTCF and LS Pol II

antibodies contained exclusively the wild type *H19* ICR sequences while both the wild type and mutant *H19* ICR sequences were present in the input DNA extracted from the cell lysates used for ChIP (Figure 3C). Therefore we conclude that the association of the LS Pol II to the *H19* ICR requires functional CTCF target sites.

A single CTCF-binding site is sufficient to activate a reporter gene. One of a possible function of the interaction between CTCF and LS Pol II could be activation of transcription followed by the recruitment of LS Pol II by CTCF at the CTCF target site. To test this hypothesis and explore the functional dimension of this interaction we prepared two vectors containing a CTCF binding site N-Myc and its mutated variant incapable of CTCF binding, fused to the promoterless luciferase reporter gene (pN-Myc-Luc wt and pN-MycLuc mut, Figures 4 A and B). The N-Myc site is located 2.1 kb downstream from the P1 promoter of the human c-myc gene (38) and was chosen randomly. The NIH 3T3 cell lines containing stably integrated constructs N-Myc –Luc wt and N-Myc-Luc mut were then generated.

When assessed for luciferase activity, normalized to the integrated vector copy number, the cells containing the wild type N- site fused with the luciferase reporter showed significantly higher levels of luciferase activity compared with the cells containing the mutant element deficient for CTCF binding (Figure 4B and C). This implies that CTCF bound to DNA may recruit LS Pol II and factors associated with LS Pol II, which could be sufficient to initiate transcription in the absence of the promoter elements.

To confirm that both, CTCF and LS Pol II are present at the wild type N-Myc site, we used transgenic pN-Myc-Luc wt NI 3T3 and and pN-MycLuc mut 3T3 cells to perform a series of “single” and “serial” ChIP asays with the anti-CTCF and anti- LS Pol II antibodies. In these experiments, in addition to the anti- LS Pol II antibodies that detects both forms of Pol II (N-20), we also tested the anti - LS Pol II (8WG16) that recognizes predominantly the hypophosphorylated LS Pol II (LS Pol IIa) and the anti-LS Pol II (H14) that is specific only to the

hyperphosphorylated LS Pol II (Pol IIo) (6). This analysis demonstrated that only the wild type N- site sequences could be detected after ChIP with all these antibodies individually and together, whereas no DNA was detected in ChIP samples from cells containing the mutant N-site (Figure 4D). The same ChIP samples were amplified with primers designed to overlap the TATA-box within the promoter region of the mouse GAPDH gene used as a control for LS Pol II loading. Intriguingly, both forms of LS Pol II, hyperphosphorylated and hypophosphorylated, were present at the N-site, whereas only the hypophosphorylated form of LS Pol II was detected in the control (GAPDH promoter). This is likely to reflect the fact that the non-elongating Pol II is associated with the GAPDH promoter, some distance away from the elongating RNA Pol II complex. On the other hand, presence of both forms of Pol II at the N-Myc site may indicate that non-elongating and elongating Pol II are confined to closely located promoter regions. No amplification was seen when the pre-immune serum (PS) was used for precipitation and when primers from a region from exon 1 of the mouse GAPDH gene lacking CTCF binding sites were employed for PCR. Additional control experiments with the anti-Pol II (N-20) as the first antibody in the "serial" ChIP assay confirmed the simultaneous presence of CTCF and LS Pol II at the-N-Myc site, whereas no signal was detected when the pre-immune serum was used as the second antibody in the "serial" ChIP with both anti-CTCF and anti-LS Pol II used as the first antibodies (Figure 4D).

From these experiments we conclude that a single CTCF-binding site is sufficient to activate a reporter gene in the transgenic context and the site occupancy of the N-Myc site by CTCF and LS Pol II depends on functional CTCF sequences. Presence of both CTCF and elongating Pol II at the wild type N-Myc site indicates that CTCF may be responsible for recruiting Pol II to the site which then can lead to transcription of the reporter gene.

DNA-bound CTCF and the largest subunit of Pol II simultaneously interact genome-wide to a subset of CTCF binding sites. To gain insight into a more genome-wide perspective of this

association, we utilized a ChIP -on -chip hybridization assay using microarrays of a library of CTCF target sites (CTSs) derived from a ChIP of mouse fetal liver. This library has been characterized both in terms of patterns of CTCF occupancy and DNA methylation status in mouse fetal liver, as well as its ability to prevent enhancer-promoter communications (40). Although it represents only a proportion of total CTSs, the library gives a genome-scale impression of occupancy of binding sites. For our experimentations, proliferating and resting mouse NIH 3T3 cells were used to prepare DNA samples from a standard ChIP or a "serial" ChIP assays for hybridization to the CTSs microarrays. Following amplification and labelling with Cy3/Cy5, the ChIP samples were hybridized to the target microarray.

To determine the specificity of the assay, we first compared the CTCF-LS Pol II and preimmune serum-LS Pol II "serial" ChIP samples in resting and growing cells (Figure 5A, panels a,b). The hybridization signals of 266 different CTCF target sites were quantified and represented on a scatter plot diagram. The signal intensities in the compared samples in both cases were low indicating the non-specific or background nature of the signals.

Having established the background levels for hybridizations, we examined the CTCF-LS Pol II "serial" ChIP samples from both resting and growing NIH 3T3 cells. The analysis of hybridizations revealed highly specific signals, with at least 8-fold enrichment over the preimmune serum-LS Pol II "serial" ChIP samples (Figure 5A, panels c and d). This is a conservative estimate as multiplex PCR of the original "serial" ChIP samples revealed minimally a 10-fold enrichment (data not shown). Of note, in this and all subsequent analyzes, all sequences harbouring repeat elements were excluded to avoid ambiguity.

Next we compared the "serial" CTCF-LS Pol II ChIP samples with CTCF occupancy as determined by the single CTCF ChIP samples. As shown in Figure 5A (panels c and d), in both resting and growing NIH 3T3 cells, only a subpopulation of CTCF target sites, approximately 10 %, was pulled down with the LS Pol II antibody.

To characterize these sequences further, we compared the hybridization signals between "serial" ChIP samples derived from resting and growing NIH 3T3 cells. Summary of the results of the "serial" ChIP assay in growing and resting cells is given in Figure 5A (panels e and f) and Tables 2 and 3. The scatter plot analyses reveal that while a majority of the sequences interact with both LS Pol II and CTCF in resting cells and growing cells, a subset of the sequences were present in the "serial" ChIP material from primarily resting cells. The finding that a subpopulation of CTCF target sites is occupied by CTCF and LS-Pol II only in resting cells may be linked to the nature of CTCF as an inhibitor of cell growth and proliferation (47, 48, 60). It is also in agreement with the fact that CTCF interacts with low-affinity sites just downstream of each of the three *MYC* promoters in only resting B cells (Lobanenkov et al, unpublished data). The hypothesis that CTCF may sequester LS Pol II at such sites and thus support the establishment and maintenance of transcriptional repression or pausing states is further examined in the "Discussion" section.

The simultaneous binding of CTCF and LS Pol II to the eleven identified targets in growing or resting NIH 3T3 cells was further confirmed by ChIP and "serial" ChIP. In these experiments, DNA material retained by IP with the anti-CTCF, anti-LS Pol II ("single ChIP"), or subsequent imunoprecipitation with the immobilized anti-CTCF antibody and then with the anti-LS Pol II antibody ("serial" ChIP), was amplified with the primers specific for each target. In almost all cases the DNA sequences were precipitated individually by the anti-CTCF and anti- LS Pol II antibodies, and also by both antibodies in "serial" ChIP assays (see below). The intensities of the signals in these assays varied, which may reflect differences of the individual targets in the affinities to CTCF – LS Pol II. These experiments confirm the simultaneous presence of CTCF and LS Pol II on the identified microarray sites in growing and resting cells (Figure 5B). In agreement with the microarray hybridization data, sequence 267 was not present in the DNA sample immunoprecipitated with the anti-LS Pol II antibody, or with both antibodies from the resting cells. Similarly, sequence 6 was not present in the DNA sample immunoprecipitated with

these antibodies from growing cells. Thus, although both of these sites are occupied by CTCF in growing and resting cells, interaction with the LS Pol II at these sites may depend on the functional state of the cells.

The same ChIP samples were subjected to amplification with primers designed to overlap the TATA-box within the promoter region of the mouse GAPDH gene, which served as a control for LS Pol II loading. The PCR products could only be seen in the samples precipitated with the anti-LS Pol II antibodies prepared from both resting and growing cells. No amplification was seen when the pre-immune serum (PS) was used for precipitation and when primers from a region from exon 1 of the mouse GAPDH gene lacking CTCF binding sites were employed for PCR.

Fifteen of the 26 different sequences that interact with both CTCF and LS Pol II could not be identified in the mouse genome database, which currently contains almost exclusively euchromatic sequence (57). Taken together with absence of known ESTs, this observation points at a heterochromatic origin of these sequences. The striking conclusion, that non-transcribed sequences nonetheless interact with LS Pol II, can be extended to several intergenic sequences. Figure 5C identifies four such clones that contain a CTCF target site that is pulled down with the LS Pol II antibody. In all of these instances, there are one or several ESTs separated from the CTCF target site by 1,5 to 15 kb. This observation prompts the proposal that CTCF recruits the Pol II to a subset of the CTCF target sites and that these complexes remain intact until the signal for the release of Pol II is received. We speculate that post-translational modifications of CTCF (29, 68) may lead to the release of the Pol II, with ensuing activation of transcription from neighbouring cryptic promoters.

DISCUSSION

In this report we have investigated the interaction between CTCF and the Pol II and discovered that the largest subunit of Pol II (LS Pol II) can be physically associated with CTCF. Both proteins are ubiquitous and essential for cell viability ((32) and V. Lobanenkov et al., unpublished

data). While the LS Pol II is an important subunit of the Pol II complex, which is an essential component of the transcriptional machinery, CTCF is a multivalent, versatile factor that activates or represses gene transcription in various modes, including chromatin insulation. Given the biological importance of the two proteins one can envisage that their interaction may have important functional implications.

In this study we firstly documented that CTCF is a component of the Pol II protein complex and this association is specific, since CTCF is not part of the TFIIH complex. Next, in a series of IP experiments the LS Pol II was identified as the protein interacting with CTCF. The specific nature of this interaction is evident because (i) CTCF can be precipitated from cell lysates with at least three different anti-LS Pol II antibodies; (ii) conversely, the anti-CTCF antibody can co-IP both isoforms of LS Pol II; (iii) the CTCF-LS Pol II co-IP reaction can be blocked by peptide N-20, originally employed to raise the anti-LS Pol II antibodies N-20; (iv) neither CTCF nor LS Pol II were retained when pre-immune serum was used in co-IP; (v) CTCF was not detected in the co-IP reactions with a large panel of the antibodies against various proteins, nuclear and cytoplasmic, (vi) CTCF and LS Pol II could be immunoprecipitated from cell extracts treated with DNase and (vii) peptides matching LS Pol II were detected in the high molecular weight bands obtained after co-IP with the anti-CTCF antibody. Notably, CTCF – LS Pol II complexes were detected in various cell types thus pointing to the "universal" functions for the association of these two ubiquitous proteins.

Further *in vitro* binding analyses revealed that the interaction between CTCF and LS Pol II is direct because CTCF and LS Pol II produced and purified from the baculoviral and bacterial systems are still able to interact *in vitro*. Since this interaction occurs via its C-terminal portion, CTCF can be subjected to regulatory influences while bound to DNA. This may be achieved for example by reversible post-translational modifications of CTCF. We previously reported the presence of several functional phosphorylation sites for protein kinase CK2 within the C-terminal domain (29) and our preliminary results show that phosphorylation of the C-terminal domain *in*

vitro with protein kinase CK2 decreases binding of the LS Pol II (I. Chernukhin and S. Shamsuddin, unpublished observations). It is therefore conceivable that phosphorylation of CTCF may be important for the regulation of CTCF –LS Pol II interaction *in vivo.* Similar mechanisms may be involved in the regulation of CTCF and Kaiso interaction, which also occurs via the C-terminal domain of CTCF (13). On the other hand, these mechanisms may differ from those relying on CTCF interactions with YB-1 (10, 28), Sin3A (37) and the helicase protein CHD8 (22), which occur through the DNA –binding Zinc finger domain of CTCF.

The interaction between CTCF and LS Pol II was reinforced by the finding that CTCF and the LS Pol II significantly co-localize in the nucleus, which indicates that subpopulations of these two proteins may be involved in the execution of the same biological processes. We hypothesized that if this was the case then CTCF and LS Pol II could be found *in vivo* in association with the same functional element of DNA (insulator or promoter) via a CTCF – binding site. To assess the simultaneous presence of two proteins at the same DNA sequence we developed a "serial" ChIP assay. In this assay, the DNA-protein complexes are first passed through the matrix linked with the antibody against one of the partner proteins. This modification has advantages over subsequent IP in solution as only protein complexes, retained on the matrix, but not the unwanted free antibodies, will be involved in the subsequent IP with the second partner antibody. The analysis of the *in vivo* occupancies of the CTCF –binding sites at the β-globin insulator by CTCF and LS Pol II, using this approach, revealed that in non-differentiated HD3 cells with no globin expression, CTCF and LS Pol II are associated with the β-globin insulator. Since FII is positioned more than 10kb upstream of the transcription start site and the likelihood of precipitation of the same fragment with the two individual antibodies is very small, the "serial" ChIP assay data also point out that LS Pol II is associated with CTCF bound to the FII site.

Although these results suggested that association of CTCF with LS Pol II depends on functional CTCF target sites, possible indirect or non-specific effects still could not be ruled out. Co-transfection of plasmids containing the wild type and CTCF target site-mutated *H19* ICR

followed by ChIP assays showed that both the anti-CTCF and anti-Pol II antibodies retained the wild type but not the mutated allele (Figure 3C). Furthermore, in the chromatin context, only the wild type CTCF-binding site, N-Myc, but not its mutated variant deficient for CTCF binding, could be precipitated by the anti-CTCF and anti-LS Pol II antibodies, individually or in a "serial" ChIP format (Figures 4D and 4E). These results further confirm that LS Pol II is associated with CTCF *via* the CTSs.

A genome-wide screen of the CTSs microarray subsequently revealed that the majority of the sequences that were pulled down with both the CTCF and LS Pol II antibodies are not transcribed. We interpret this information to mean that a CTCF- LS Pol II complex is recruited to these CTCF target sites in a transcription-independent manner. In some instances, where the sequences could be identified, those CTCF target sites map in the relative vicinity of ESTs, i.e. at a distance from 1, 5 to 15 kb (Figure 5C). Given that the sonicated fragments were on average less than 1 kb, it is less likely that the low-abundant ESTs signified an interaction between CTCF and LS Pol II in a transcription-dependent manner. Similarly, it remains intriguing how LS Pol II can be associated with the murine β-globin locus positioned far from promoters in a transcription-independent manner (23). We speculate that such association could be mediated by CTCF. The formation of the LCR-promoter loops known to exist within the β-globin locus (55) may be responsible for bringing LS Pol-II and CTCF in direct physical proximity.

A common theme of this report is that the abundant and ubiquitous CTCF may provide a novel pathway to recruit transcription complexes to particular targets. In this scenario, non-coding transcripts, known to originate throughout the genome, might be transcribed in an enhancer-independent manner. We thus hypothesize that the release of the LS Pol II from a DNA-bound CTCF complex might transcriptionally activate nearby cryptic promoters or, alternatively, CTCF site itself could act as a promoter in a certain genomic context. As a result, the genome may be represented by low-abundant non-coding transcripts in a manner that is dictated by CTCF target site occupancy. Indeed, we demonstrate here that only the wild type version of the CTCF –

binding site (N-Myc) fused with the luciferase reporter gene, was able to activate the reporter gene efficiently. As *in vivo* binding of CTCF and LS Pol II to the wild-type N-Myc site was verified by ChIP and "serial" ChIP assays, it is possible that binding of the CTCF-LS Pol II complex to this site was sufficient to activate transcription from this "artificial" promoter. Another and not mutually exclusive possibility is that the DNA-bound CTCF-LS Pol II complex might be mistaken for a promoter by nearby enhancers. Such promoter decoys have been proposed to provide one of several essential mechanisms by which chromatin insulators block enhancer-promoter communications (18).

The identification of intronic / exonic sequences that simultaneously interact with CTCF and LS Pol II hint at another possibility: the tracking Pol II encounters the DNA-bound CTCF stalling the transcriptional elongation process. Such pause elements have previously been described downstream of each of the *MYC* promoters (33, 56). Intriguingly, these pause elements map to or are identical with CTCF target sites, which are occupied in growth arrested cells (45). It is conceivable therefore that dissociation of the CTCF-DNA complex during G0/G1 transition might subsequently release LS Pol II to complete the transcriptional elongation process. The commonality of such a scenario is indicated by our demonstrations here, that CTCF is dissociated from a significant subpopulation of CTCF target sites in growing cells, but not in resting NIH 3T3 cells.

Although the CTSs- microarray has been a valuable tool in this investigation, its limitations should be acknowledged. Firstly, this microarray represented only a relatively small subpopulation of CTCF binding sites, which is approximately 5%–7% of all potential CTCF target sites (40). This number however may be lower (~ 0.75% -1%), if the criteria described by Vetchinova et al are used for assessment (62). Secondly, some of the previously identified and characterized CTSs (e.g. *MYC*, *H19* ICR) could not be assessed in the screening as they were not present on the microarray (40). Instead, a representative panel of these CTSs were investigated separately in this report. Thirdly, origin of cells used for microarray experimentations (NIH3T3

fibroblasts) should be taken into consideration as CTSs in NIH 3T3 and foetal liver (source for microarray CTS) may have different occupancy patterns and only partially overlap.

Our observations indicate that only a relatively small subpopulation (approximately 10%) of CTSs, as determined in the microarray screening, can simultaneously interact with CTCF and LS Pol II. However, given the abundance of these two proteins in the nucleus, it is likely that the number of specific CTCF- LS Pol II complexes associated with CTSs may be sufficient to meet the functional requirements of cells. On the other hand, as the immunostaining reveals (Figure 2), the actual number of CTCF- LS Pol II complexes in the nucleus may be greater as some of these complexes may not be linked to the CTSs. Of note, association of CTCF with other interacting proteins (e.g. helicase protein CHD8 and YB-1) has also been shown to be partial ((22), S. Shamsuddin and F.Docquier, unpublished observations). We hypothesize that different sub-populations of CTCF may generate different specific complexes with various protein partners as a means of creating molecular and functional diversity.

Based on the experimental data presented in this study we suggest several possible functions of CTCF interaction with LS Pol II. Firstly, CTCF may engage RNA Polymerase II to potentially generate "storage" of proteins necessary for transcription in promoter-proximal positions, thereby modulating transcription in response to a stimulus. Data in Figure 3A demonstrating the presence of CTCF and LS Pol II at the CTCF-binding site at the insulator in non-differentiated chicken HD3 cells, and absence of CTCF and LS Pol II at this site in globin-producing differentiated HD3 cells support this case. Secondly, CTCF may "piggybacks" LS Pol II to a certain set of DNA targets to establish an appropriate configuration for pausing of transcriptional elongation once a CTCF target site has been recognized. Alternatively, CTCF may play a role of a functional equivalent of TBP allowing accurate initiation of transcription at some promoters. In our model system, just the presence of a single CTCF binding site was sufficient to activate transcription from the adjacent luciferase gene (Figure 4B). Finally, we have here also discussed the potential roles of CTCF-LS Pol II complexes on the expression of intergenic, non-

coding transcripts and chromatin insulation. Irrespective of these considerations, the "marriage" between a versatile chromatin insulator protein, CTCF, and a Pol II enzyme complex constitutes a novel angle on the genome-wide regulation of gene transcription.

ACKNOWLEDGEMENTS

We thank S. Muller and J. Dadoune for anti - histone H2A and histone H3 antibodies. We are also grateful to M. Metodiev, A. Harrison, A. Akoulitchev and A. Ramadass for helpful discussions and P. O'Toole for assistance with confocal microscopy. We gratefully acknowledge the assistance of A. Isaksson and the Wallenberg microarray platform at the Rudbeck laboratory. This research was supported by the Association for International Cancer Research (I.C., E.K.), the Breast Cancer Campaign (F.D., E.K.), the Medical Research Council (D.F., E.K.), the Research Promotion Fund from University of Essex (E.K.), the National Institutes of Health grants GM59643 and CA81017 (C. -M.C.), Scholarship from the Malaysian Government (S.S.), the Swedish Science Research Council (R.O.), the Juvenile Diabetes Research Foundation International (R.O.), the Swedish Cancer Research Foundation (R.O.), the Swedish Pediatric Cancer Foundation (R.O.), the Wallenberg and Lundberg Foundations (R.O.), and Stiftelsen Wenner-Grenska Samfundet (R.O.).

FIGURE LEGENDS

Figure 1. CTCF is associated with LS Pol II *in vivo* and *in vitro*.

(A) **CTCF is a part of the Pol II protein complex.** The Pol II and TFIIH complexes were purified, resolved by 10% SDS-PAGE, transferred onto a membrane and then probed with the anti-CTCF antibody. The band, specific for CTCF (indicated) can be seen in nuclear extract (NE) and in the Pol II complex, but not in the TFIIH complex. Position of the molecular marker is indicated on the left.

(B) **Analysis of the *in vivo* interactions between CTCF and LS Pol II by co-immunoprecipitation with the anti- LS Pol II antibody.** The co-IP reactions were performed with a series of antibodies shown on top of the image; lysates from $5x10^5$ HeLa cells were used in each reaction. The arrow signals the position of CTCF co-immunoprecipiated with the anti-LS Pol II antibody (N-20) and anti-YB-1 antibody. Pre-incubation with peptide N-20 can block co-IP with the anti-LS Pol II antibody. On the other hand, CTCF does not co-IP with TBP, Sp1, Rb1, histone H2A, histone H3 or Ab-1 p53.

(C) **Analysis of the interactions between CTCF and LS Pol II pre-treated with DNAse I and interactions between CTCF, LS Pol IIa and LS Pol IIo.** Antibodies used for co-IP are shown on top of the image. The anti- LS Pol II antibodies were as follows: the anti-LS Pol II (N-20) antibody that detects both forms of Pol II, the anti - LS Pol II (8WG16) recognizes the hypophosphorylated LS Pol II (LS Pol IIa) and the anti-LS Pol II (H14) recognizes the hyperphosphorylated LS Pol II (Pol IIo). Lysates from $5x10^5$ HeLa cells were used for each reaction. Samples were electrophoretically separated, blotted and probed with the anti-CTCF antibody. The arrow signals the position of CTCF. The developed films were scanned and images quantified. Levels of CTCF precipitated by the anti-Pol IIa, anti-Pol IIo antibodies and anti-Pol II treated with DNAse I prior to co-IP are presented as a percentage from the co-IP reactions with the anti-Pol II (N-20) antibody (designated as 100%). Amount of CTCF precipitated by the anti-YB-1 antibody after treatment with DNAse I is presented as a percentage from the co-IP reaction with the anti-YB- 1 antibody (designated as 100%). Numbers below the lanes represent these results.

(D) **Analysis of the *in vivo* interactions between CTCF and LS Pol II by immunoprecipitation with the anti-CTCF antibodies.** Western blot assay with anti- LS Pol II antibody (N-20) was performed after co-IP from HeLa lysates with pre-immune serum (PS) or anti-CTCF antibody (CTCF); $5x10^5$ HeLa cells were used for each reaction. The immunocomplexes were resolved by SDS-PAGE, blotted, and the membrane was probed with

the anti-LS Pol II antibody N-20. Arrows on the right indicate the positions of the hypophosphorylated LS Pol II (IIa) sized 220 kDa and the hyperphosphorylated LS Pol II (IIo) sized 240 kDa. An ~85kDa protein depicted by the asterisk is most likely partially reduced IgG (21).

(E) **The C-terminal domain of CTCF interacts with LS Pol II *in vitro*.** The three domains of CTCF (CTCF-N, CTCF-Zn and CTCF-C) expressed in *E.coli* and BSA (control) were coupled to the matrix and incubated with the whole lysate from K562 cells, washed with 0.25M RIPA buffer, and the retained proteins were analyzed by Western blot assay with the anti-LS-Pol II antibody. Arrows indicate the positions of two forms of the LS Pol II. K562 – 20µl of K562 cell lysate used in the assay. Position of the molecular marker is indicated on the left.

(F) **Analysis of the proteins used in the interaction assay.** The membrane utilized in the experiment described in Figure 1E was stripped and subsequently probed with the anti-His-tag antibodies. Position of the molecular markers are indicated on the right.

(G) **The three domain structure of CTCF.** The three domains of CTCF are depicted as follows: N – N terminal (patterned box), ZF – 11 Zn finger (grey box) and C- C-terminal (open box) domains. The His-tag is shown as an open circle. Amino acids are numbered according to Filippova et al (16).

(H) **The full-length CTCF and LS Pol II interact directly *in vitro*.** The complete peptides of CTCF (baculoCTCF) and LS Pol II (bact LS Pol II) were generated *in vitro* using baculoviral and bacterial systems, respectively. BaculoCTCF and BSA were coupled to the matrix and incubated with the lysate containing bactLS Pol II, washed with 0.25M RIPA buffer, and the retained proteins were analyzed by Western blot assay with anti-LS-Pol II antibody. The position of the bactLS Pol II in shown.

Figure 2. Confocal analysis of CTCF and LS Pol II in HeLa and K562 cells. HeLa and K562 cell lines were prepared and immunostained as described under “Materials and Methods”. The endogenous CTCF and LS Pol II proteins are extensively co-localized in the nucleus in both cell

lines (HeLa, upper panel and K562, lower panel) as shown by the merge of the CTCF (FITC, green) and Pol II (TRITC, red) staining and colocalization analysis using the methods of Costes et al (11). The typical two-dimensional histograms of the fluorescence for a K562 cell (indicated by a white arrow) and a HeLa cell are shown on the right.

Figure 3. CTCF and LS Pol II interact *in vivo*, at the β-globin insulator and the *H19* ICR.

(A) **CTCF and LS Pol II are associated at the β-globin insulator in proliferating HD3 cells** as shown by ChIP and "serial" ChIP assays. Nuclear extracts were prepared from $5x10^6$ of proliferating and differentiated HD3 cells; the standard ChIP assay was performed to assess the *in vivo* occupancies at the DNA target sites and "serial" ChIP assay was performed to assess simultaneous presence of CTCF and LS Pol II at the β-globin insulator. PCR products were resolved by a 1% agarose gel and Southern blot assay was performed with the ^{32}P-labeled β-globin insulator FII probe. PCR and hybridization with the CTCF exon 8 and chicken β-actin probes were used as background control and as LS Pol II loading control, respectively (see Table 1 for details of the hybridization probes).

The antibodies used in ChIP and "serial" ChIP assays are indicated above the corresponding lanes as follows: PS/CTCF- "serial" ChIP with pre-immune serum (PS) followed by the anti-CTCF antibody; CTCF - ChIP with the anti-CTCF antibody; Pol II- ChIP with the anti-LS Pol II antibody; CTCF/Pol II - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol II antibody; PS/Pol II – "serial" ChIP with the pre-immune serum followed by the anti-LS Pol II antibody. Input – DNA from HD3 cell lysates.

(B) **Cartoon illustration of the chicken β-globin domain** (adapted from Prioleau et al. and Bell et al.) (3, 46). The 1.2 kb insulator core element is shown as an open box; the detailed structure is represented in the enlarged image. CTCF binds to the 42 bp F II region within the insulator (grey box). The four β-globin genes are shown as black boxes. The hypersensitive site HS4 is indicated with a vertical arrow. Primers used for amplification of the FII are shown by horizontal arrows (sequences of the primers are given in Table 1).

(C) **LS Pol II association with the H19 ICR requires functional CTCF target sites.** pGEM vectors containing the wild type (wt) and mutant (mut) 1.2 kb *H19* ICR were transfected into JEG-3 cells, individually or mixed together as indicated. The image shows DNA amplified from ChIP material pulled down by CTCF and LS Pol II antibodies, or control material, not subjected to ChIP, digested with EcoRV. The antibodies used in the assay are indicated above the corresponding lanes. The PCR products were resolved by a 1% agarose gel. M – DNA Marker, 100bp DNA ladder.

(D) **Cartoon illustration of the *IGF2 –H19* locus** (adapted from Bell and Felsenfeld (2)). Positions of *IGF2* (open box) and *H19* (black box) genes are shown. The 2.4 kb *H19* ICR element (grey box) is located –2 kb to –4.4 kb relative to the transcription start site of *H19* (59). The *IGF2* and *H19* ICR are separated by more than 80 kb of intervening sequences. Transcription start sites of *IGF2* and *H19* are presented by bent arrows. The 1.2 kb *H19* ICR fragment cloned into pGEM vector is shown as a black bar. Primers used for *H19* ICR amplification are denoted by straight arrows (sequences of the primers are given in Table 1). The sequence recognized by Eco RV is specific for the mutated CTCF target site 3 (indicated by an asterisk) (44).

Figure 4. CTCF and LS Pol II are associated with wild type N-Myc, which alone can activate transcription from the Luciferase reporter gene.

(A) **Cartoon illustration of the 5' non-coding region of the human *c-Myc* gene promoter** (adapted from Filippova et al (16). Grey boxes depict the CTCF-binding sites A, B and N (38).

(B) **The wild type N-Myc sequence activates the Luciferase reporter gene.** The NIH 3T3 cells stably transfected with pN-Myc-Luc wt and pN-MycLuc mut were harvested, assayed for luciferase activity as described in Materials and Methods. The luciferase activity normalized to the plasmid copy number is shown in relative luciferase units (RLU). Each bar represents an average of three experiments performed in triplicate. Error bars indicate standard deviations. The diagrams on top show the structure of the two plasmids, pN-Myc-Luc wt and pN-MycLuc mut. Grey boxes depict N-Myc sites, Luc – luciferase reporter gene.

(C) **Southern blot analysis of the DNA extracted from NIH 3T3 cells** (pN-Myc-Luc wt and pN-MycLuc mut). Genomic DNA was extracted, digested with ClaI /Xho I, blotted and hybridized as described under Materials and Methods.

(D) **CTCF and LS Pol II are associated with the wild type N-Myc site in stably transfected NIH 3T3 cells**. The standard ChIP and "serial" ChIP assays were performed to assess the *in vivo* occupancies by CTCF and Pol II at the N-Myc target sites. The antibodies used in ChIP and "serial" ChIP are indicated above the corresponding lanes as follows: CTCF - ChIP with the anti-CTCF antibody; Pol II- ChIP with the anti-LS Pol II antibody; Pol IIa – ChIP with the anti- LS Pol IIa antibody (hypophosphorylated form); Pol IIo – ChIP with the anti LS Pol IIo antibody (hyperphosphorylated form); CTCF/Pol II - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol II antibody; CTCF/Pol IIa - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol IIa antibody; CTCF/Pol IIo - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol IIo antibody; PS – ChIP with pre-immune serum (PS). Input – DNA from NIH 3T3 cell lysates. DNA prepared from these samples was amplified using corresponding pairs of primers as described under Materials and Methods and in Table 1. The PCR products were resolved in a 1% agarose gel. M – DNA Marker, 100bp DNA ladder.

(E) **CTCF and LS Pol II association with the wild type N-Myc site in stably transfected NIH 3T3 cells is specific.** The "serial" ChIP assays were performed to further assess the specificity of the *in vivo* occupancies by CTCF and Pol II at the N-Myc target sites. The antibodies used in ChIP and "serial" ChIP are indicated above the corresponding lanes as follows: CTCF/Pol II - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol II antibody; Pol II/CTCF - "serial" ChIP with the anti-LS Pol II antibody followed by the anti-CTCF antibody; CTCF/PS - "serial" ChIP with the anti-CTCF antibody followed by pre-immune serum; Pol II/PS - "serial" ChIP with the anti-LS Pol II antibody followed by pre-immune serum. Input – DNA from NIH 3T3 cell lysates. DNA prepared from these samples was amplified using corresponding pairs of primers as described

under Materials and Methods and in Table 1. The PCR products were resolved in a 1% agarose gel. M – DNA Marker, 100bp DNA ladder.

Figure 5. Genome-wide interaction between CTCF and LS Pol II.

(A) **Chip on chip hybridization analysis revealing the simultaneous presence of CTCF and LS–Pol II epitopes genome-wide.** DNA samples from the standard ChIP or "serial" ChIP assays from proliferating and resting mouse NIH 3T3 cells were prepared and hybridized to CTCF –target site microarrays. Hybridization signals are expressed in relative fluorescence units (RFU); the results of analyzes are presented in scatter plots as follows:

a) Comparison of hybridization data between "serial" ChIP samples CTCF-Pol II versus pre-immune serum-Pol II in resting cells.

b) Comparison of hybridization data between "serial" ChIP samples CTCF-Pol II versus pre-immune serum-Pol II in growing cells.

c) Comparison of the CTCF ChIP with the CTCF-Pol II "serial" ChIP signals in resting cells.

d) Comparison of the CTCF ChIP with the CTCF-Pol II "serial" ChIP signals in growing cells.

e) Comparison between "serial" ChIP CTCF-Pol II samples in resting and growing cells.

f) Comparison between single CTCF ChIP/CTCF ChIP signals in resting and growing cells.

(B) Analysis of the eleven sequences in growing or resting NIH 3T3 identified by screening of the CTCF-target site microarrays demonstrating the simultaneous presence of CTCF and LS–Pol II. Proliferating and resting mouse NIH 3T3 cells were used to perform the standard ChIP or "serial" ChIP assays. The antibodies used in ChIP and "serial" ChIP assays are indicated above the corresponding lanes as follows: CTCF/Pol II - "serial" ChIP with the anti-CTCF antibody followed by the anti-LS Pol II antibody; CTCF - ChIP with the anti-CTCF antibody; Pol II- ChIP with the anti-LS Pol II antibody (Pol II); PS – ChIP with the pre-immune serum. Input – DNA from NIH 3T3 cell lysates. DNA prepared from these samples was amplified using corresponding pairs of primers as described in Materials and Methods and Table 1, and resolved by a 1% agarose gel. G –

growing cells; R – resting cells. M – DNA Marker, 100bp DNA ladder. GAPDH(p) – promoter region of *GAPDH*, GAPDH (e) – exon 1 region of *GAPDH*.

(C) **A gene map depicting the location of transcriptional units of identified genes (black arrows) or ESTs (green arrows).** The numbers below each row indicate the distance between the CTCF target site and the closest known transcriptional unit. Additional sequences are described in Table 3.

Table 1. PCR primers used in ChIP analysis and generation of probes for hybridization.

Table 2. Summary of number of intragenic (Intra), intergenic (Inter) and unidentified (Unident) sequences interacting with both CTCF and the LS Pol II in growing and resting cells.

Table 3. Summary of the identified CTCF target sites in growing and resting cells.

REFERENCES

1. **Acker, J., M. de Graaff, I. Cheynel, V. Khazak, C. Kedinger, and M. Vigneron.** 1997. Interactions between the human RNA polymerase II subunits. J Biol Chem **272:**16815-21.
2. **Bell, A. C., and G. Felsenfeld.** 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature **405:**482-5.
3. **Bell, A. C., A. G. West, and G. Felsenfeld.** 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell **98:**387-96.
4. **Besse, S., M. Vigneron, E. Pichard, and F. Puvion-Dutilleul.** 1995. Synthesis and maturation of viral transcripts in herpes simplex virus type 1 infected HeLa cells: the role of interchromatin granules. Gene Expr **4:**143-61.
5. **Beug, H., G. Doederlein, C. Freudenstein, and T. Graf.** 1982. Erythroblast cell lines transformed by a temperature-sensitive mutant of avian erythroblastosis virus: a model system to study erythroid differentiation in vitro. J Cell Physiol Suppl **1:**195-207.

6. **Bregman, D. B., L. Du, S. van der Zee, and S. L. Warren.** 1995. Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. J Cell Biol **129:**287-98.

7. **Burcin, M., R. Arnold, M. Lutz, B. Kaiser, D. Runge, F. Lottspeich, G. N. Filippova, V. V. Lobanenkov, and R. Renkawitz.** 1997. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol Cell Biol **17:**1281-8.

8. **Butcher, D. T., D. N. Mancini-DiNardo, T. K. Archer, and D. I. Rodenhiser.** 2004. DNA binding sites for putative methylation boundaries in the unmethylated region of the BRCA1 promoter. Int J Cancer **111:**669-78.

9. **Chernukhin, I. V., and E. M. Klenova.** 2000. A method of immobilization on the solid support of complex and simple enzymes retaining their activity. Anal Biochem **280:**178-81.

10. **Chernukhin, I. V., S. Shamsuddin, A. F. Robinson, A. F. Carne, A. Paul, A. I. El-Kady, V. V. Lobanenkov, and E. M. Klenova.** 2000. Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. J Biol Chem **275:**29915-21.

11. **Costes, S. V., D. Daelemans, E. H. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett.** 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. Biophys J **86:**3993-4003.

12. **Dahmus, M. E.** 1996. Phosphorylation of mammalian RNA polymerase II. Methods Enzymol **273:**185-93.

13. **Defossez, P. A., K. F. Kelly, G. J. Filion, R. Perez-Torrado, F. Magdinier, H. Menoni, C. L. Nordgaard, J. M. Daniel, and E. Gilson.** 2005. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. J Biol Chem **280:**43017-23.

14. **Delgado, M. D., I. V. Chernukhin, A. Bigas, E. M. Klenova, and J. Leon.** 1999. Differential expression and phosphorylation of CTCF, a c-myc transcriptional regulator, during differentiation of human myeloid cells. FEBS Lett **444:**5-10.

15. **El-Kady, A., and E. Klenova.** 2005. Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. FEBS Lett **579:**1424-34.

16. **Filippova, G. N., S. Fagerlie, E. M. Klenova, C. Myers, Y. Dehner, G. Goodwin, P. E. Neiman, S. J. Collins, and V. V. Lobanenkov.** 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol **16:**2802-13.

17. **Franklin, G. C., M. Donovan, G. I. Adam, L. Holmgren, S. Pfeifer-Ohlsson, and R. Ohlsson.** 1991. Expression of the human PDGF-B gene is regulated by both positively and negatively acting cell type-specific regulatory elements located in the first intron. Embo J **10:**1365-73.

18. **Geyer, P. K.** 1997. The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev **7:**242-8.

19. **Gross, P., and T. Oelgeschlager.** 2006. Core promoter-selective RNA polymerase II transcription. Biochem Soc Symp**:**225-36.

20. **Harlow, E., and D. Lane.** 1999. Using antibodies : a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

21. **Hay, F. C., O. M. R. Westwood, P. N. Nelson, and L. Hudson.** 2002. Practical immunology, 4th ed. Blackwell Science, Malden, MA.

22. **Ishihara, K., M. Oshimura, and M. Nakao.** 2006. CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. Mol Cell **23:**733-42.

23. **Johnson, K. D., J. A. Grass, C. Park, H. Im, K. Choi, and E. H. Bresnick.** 2003. Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. Mol Cell Biol **23:**6484-93.

24. **Kanduri, C., V. Pant, D. Loukinov, E. Pugacheva, C. F. Qi, A. Wolffe, R. Ohlsson, and V. V. Lobanenkov.** 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol **10:**853-6.

25. **Kershnar, E., S. Y. Wu, and C. M. Chiang.** 1998. Immunoaffinity purification and functional characterization of human transcription factor IIH and RNA polymerase II from clonal cell lines that conditionally express epitope-tagged subunits of the multiprotein complexes. J Biol Chem **273:**34444-53.

26. **Klenova, E., I. Chernukhin, T. Inoue, S. Shamsuddin, and J. Norton.** 2002. Immunoprecipitation techniques for the analysis of transcription factor complexes. Methods **26:**254-9.

27. **Klenova, E., and R. Ohlsson.** 2005. Poly(ADP-ribosyl)ation and epigenetics. Is CTCF PARt of the plot? Cell Cycle **4:**96-101.

28. **Klenova, E., A. C. Scott, J. Roberts, S. Shamsuddin, E. A. Lovejoy, S. Bergmann, V. J. Bubb, H. D. Royer, and J. P. Quinn.** 2004. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. J Neurosci **24:**5966-73.

29. **Klenova, E. M., I. V. Chernukhin, A. El-Kady, R. E. Lee, E. M. Pugacheva, D. I. Loukinov, G. H. Goodwin, D. Delgado, G. N. Filippova, J. Leon, H. C. Morse, 3rd, P. E. Neiman, and V. V. Lobanenkov.** 2001. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol Cell Biol **21:**2221-34.

30. **Klenova, E. M., H. C. Morse, 3rd, R. Ohlsson, and V. V. Lobanenkov.** 2002. The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. Semin Cancer Biol **12:**399-414.

31. **Klenova, E. M., R. H. Nicolas, H. F. Paterson, A. F. Carne, C. M. Heath, G. H. Goodwin, P. E. Neiman, and V. V. Lobanenkov.** 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. Mol Cell Biol **13:**7612-24.

32. **Kobor, M. S., and J. Greenblatt.** 2002. Regulation of transcription elongation by phosphorylation. Biochim Biophys Acta **1577:**261-275.
33. **Krumm, A., T. Meulia, M. Brunvand, and M. Groudine.** 1992. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes Dev **6:**2201-13.
34. **Kuo, M. H., and C. D. Allis.** 1999. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods **19:**425-33.
35. **Kuzmin, I., L. Geil, L. Gibson, T. Cavinato, D. Loukinov, V. Lobanenkov, and M. I. Lerman.** 2005. Transcriptional regulator CTCF controls human interleukin 1 receptor-associated kinase 2 promoter. J Mol Biol **346:**411-22.
36. **Lemon, B., and R. Tjian.** 2000. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev **14:**2551-69.
37. **Lutz, M., L. J. Burke, G. Barreto, F. Goeman, H. Greb, R. Arnold, H. Schultheiss, A. Brehm, T. Kouzarides, V. Lobanenkov, and R. Renkawitz.** 2000. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res **28:**1707-13.
38. **Lutz, M., L. J. Burke, P. LeFevre, F. A. Myers, A. W. Thorne, C. Crane-Robinson, C. Bonifer, G. N. Filippova, V. Lobanenkov, and R. Renkawitz.** 2003. Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. Embo J **22:**1579-87.
39. **Malik, S., and R. G. Roeder.** 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci **30:**256-63.
40. **Mukhopadhyay, R., W. Yu, J. Whitehead, J. Xu, M. Lezcano, S. Pack, C. Kanduri, M. Kanduri, V. Ginjala, A. Vostrov, W. Quitschke, I. Chernukhin, E. Klenova, V. Lobanenkov, and R. Ohlsson.** 2004. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome Res **14:**1594-602.

41. **Nicolas, R. H., G. Partington, G. N. Major, B. Smith, A. F. Carne, N. Huskisson, and G. Goodwin.** 1991. Induction of differentiation of avian erythroblastosis virus-transformed erythroblasts by the protein kinase inhibitor H7: analysis of the transcription factor EF1. Cell Growth Differ **2:**129-35.

42. **Ohlsson, R., R. Renkawitz, and V. Lobanenkov.** 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet **17:**520-7.

43. **Pant, V., S. Kurukuti, E. Pugacheva, S. Shamsuddin, P. Mariano, R. Renkawitz, E. Klenova, V. Lobanenkov, and R. Ohlsson.** 2004. Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. Mol Cell Biol **24:**3497-504.

44. **Pant, V., P. Mariano, C. Kanduri, A. Mattsson, V. Lobanenkov, R. Heuchel, and R. Ohlsson.** 2003. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev **17:**586-90.

45. **Perez-Juste, G., S. Garcia-Silva, and A. Aranda.** 2000. An element in the region responsible for premature termination of transcription mediates repression of c-myc gene expression by thyroid hormone in neuroblastoma cells. J Biol Chem **275:**1307-14.

46. **Prioleau, M. N., P. Nony, M. Simpson, and G. Felsenfeld.** 1999. An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. Embo J **18:**4035-48.

47. **Qi, C. F., A. Martensson, M. Mattioli, R. Dalla-Favera, V. V. Lobanenkov, and H. C. Morse, 3rd.** 2003. CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. Proc Natl Acad Sci U S A **100:**633-8.

48. **Rasko, J. E., E. M. Klenova, J. Leon, G. N. Filippova, D. I. Loukinov, S. Vatolin, A. F. Robinson, Y. J. Hu, J. Ulmer, M. D. Ward, E. M. Pugacheva, P. E. Neiman, H. C. Morse,**

3rd, S. J. Collins, and V. V. Lobanenkov. 2001. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. Cancer Res **61:**6002-7.

49. **Renaud, S., D. Loukinov, F. T. Bosman, V. Lobanenkov, and J. Benhattar.** 2005. CTCF binds the proximal exonic region of hTERT and inhibits its transcription. Nucleic Acids Res **33:**6850-60.
50. **Saitoh, N., A. C. Bell, F. Recillas-Targa, A. G. West, M. Simpson, M. Pikaart, and G. Felsenfeld.** 2000. Structural and functional conservation at the boundaries of the chicken beta-globin domain. Embo J **19:**2315-22.
51. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning : a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
52. **Saunders, A., L. J. Core, and J. T. Lis.** 2006. Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol **7:**557-67.
53. **Shi, S. R., M. E. Key, and K. L. Kalra.** 1991. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. J Histochem Cytochem **39:**741-8.
54. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67:**31-40.
55. **Splinter, E., H. Heath, J. Kooren, R. J. Palstra, P. Klous, F. Grosveld, N. Galjart, and W. de Laat.** 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev **20:**2349-54.
56. **Strobl, L. J., and D. Eick.** 1992. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. Embo J **11:**3307-14.
57. **The Mouse Genome Sequencing Consortium.** 2006. http://www.ensembl.org/Mus_musculus/.
58. **Thomas, M. C., and C. M. Chiang.** 2006. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol **41:**105-78.

59. **Thorvaldsen, J. L., K. L. Duran, and M. S. Bartolomei.** 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev **12:**3693-702.

60. **Torrano, V., I. Chernukhin, F. Docquier, V. D'Arcy, J. Leon, E. Klenova, and M. D. Delgado.** 2005. CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. J Biol Chem **280:**28152-61.

61. **Torrano, V., J. Navascues, F. Docquier, R. Zhang, L. J. Burke, I. Chernukhin, D. Farrar, J. Leon, M. T. Berciano, R. Renkawitz, E. Klenova, M. Lafarga, and M. D. Delgado.** 2006. Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. J Cell Sci **119:**1746-59.

62. **Vetchinova, A. S., S. B. Akopov, I. P. Chernov, L. G. Nikolaev, and E. D. Sverdlov.** 2006. Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19. Anal Biochem **354:**85-93.

63. **Vostrov, A. A., and W. W. Quitschke.** 1997. 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 **272:**33353-9.

64. **Vuillard, L., T. Rabilloud, and M. E. Goldberg.** 1998. Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. Eur J Biochem **256:**128-35.

65. **Wu, S. Y., and C. M. Chiang.** 1998. Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. J Biol Chem **273:**12492-8.

66. **Wu, S. Y., E. Kershnar, and C. M. Chiang.** 1998. TAFII-independent activation mediated by human TBP in the presence of the positive cofactor PC4. Embo J **17:**4478-90.

67. **Young, R. A.** 1991. RNA polymerase II. Annu Rev Biochem **60:**689-715.

68. **Yu, W., V. Ginjala, V. Pant, I. Chernukhin, J. Whitehead, F. Docquier, D. Farrar, G. Tavoosidana, R. Mukhopadhyay, C. Kanduri, M. Oshimura, A. P. Feinberg, V. Lobanenkov, E. Klenova, and R. Ohlsson.** 2004. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. Nat Genet **36:**1105-10.

69. **Yusufzai, T. M., H. Tagami, Y. Nakatani, and G. Felsenfeld.** 2004. CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. Mol Cell **13:**291-8.

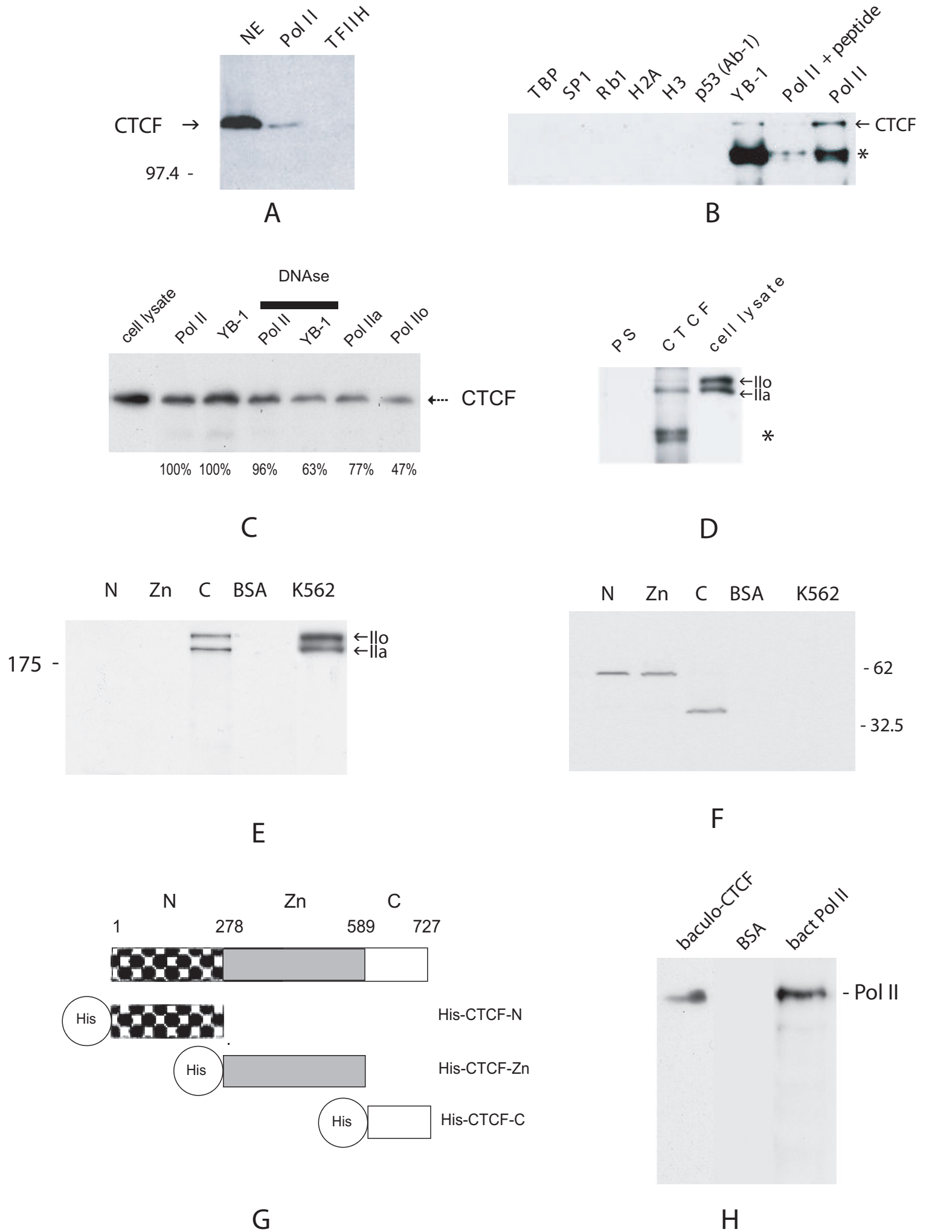
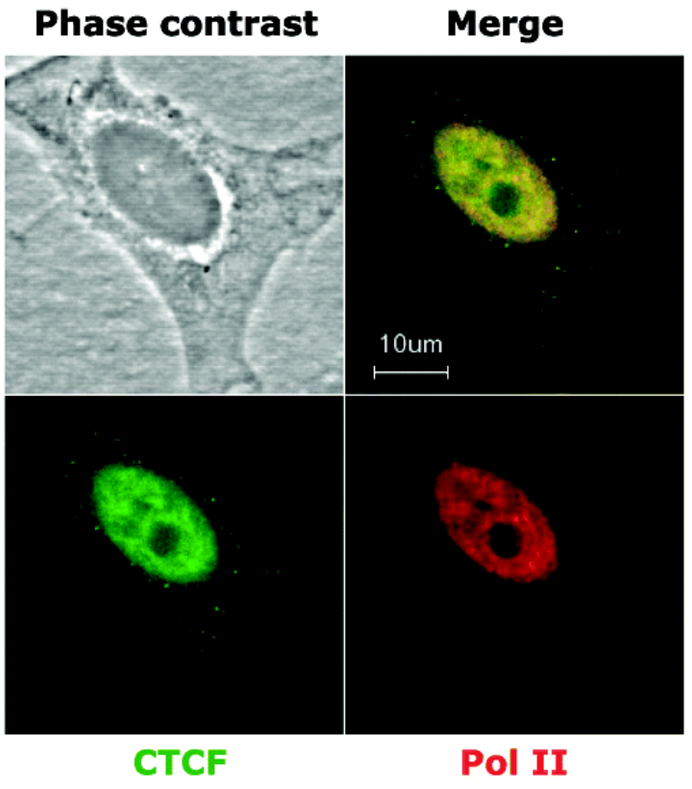
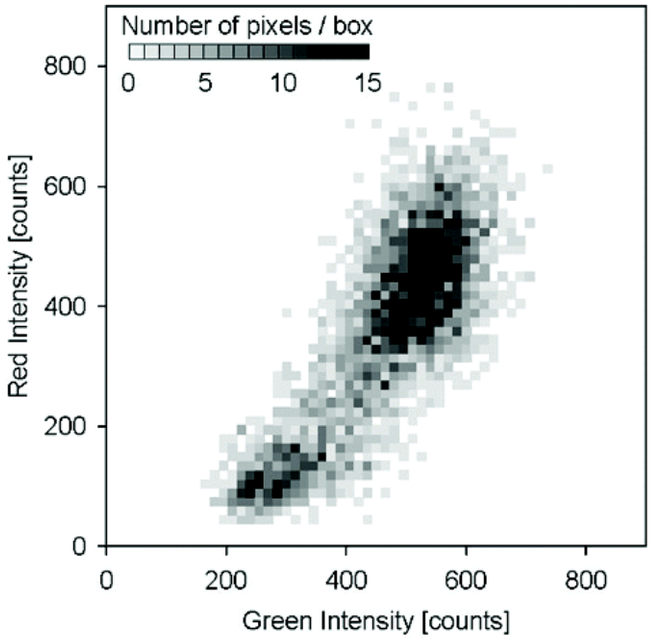

NE
Pol II
TFIIH
CTCF
97.4 -
A
TBP
SP1
Rb1
H2A
H3
p53 (Ab-1)
YB-1
Pol II + peptide
Pol II
CTCF
*
B
DNAse
cell lysate
Pol II
YB-1
Pol II
YB-1
Pol IIa
Pol IIo
CTCF
100% 100% 96% 63% 77% 47%
C
P S
C T C F
cell lysate
IIo
IIa
*
D
N
Zn
C
BSA
K562
175 -
IIo
IIa
E
N
Zn
C
BSA
K562
- 62
- 32.5
F
N
Zn
C
1
278
589
727
His
His-CTCF-N
His
His-CTCF-Zn
His
His-CTCF-C
G
baculo-CTCF
BSA
bact Pol II
- Pol II
H

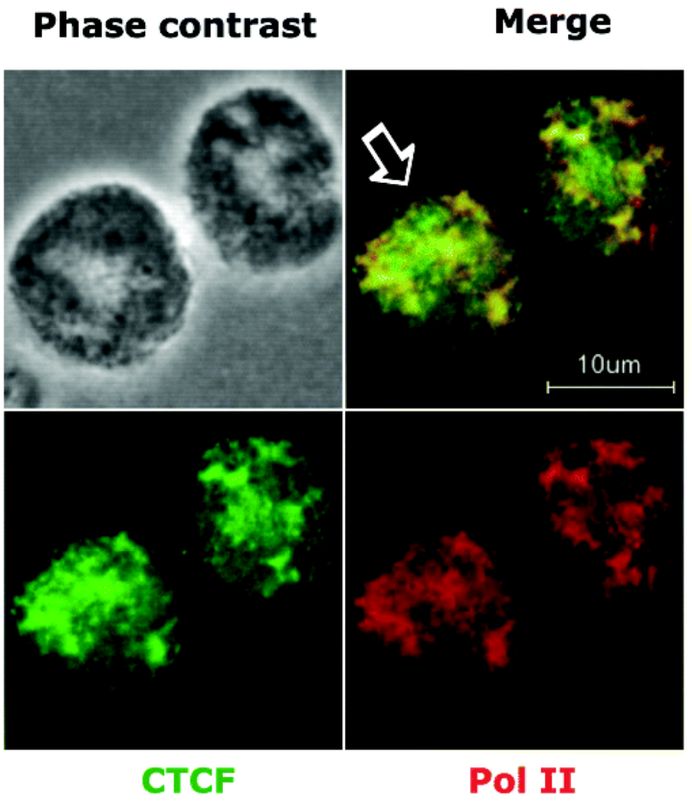
Figure 1

HeLa





K562



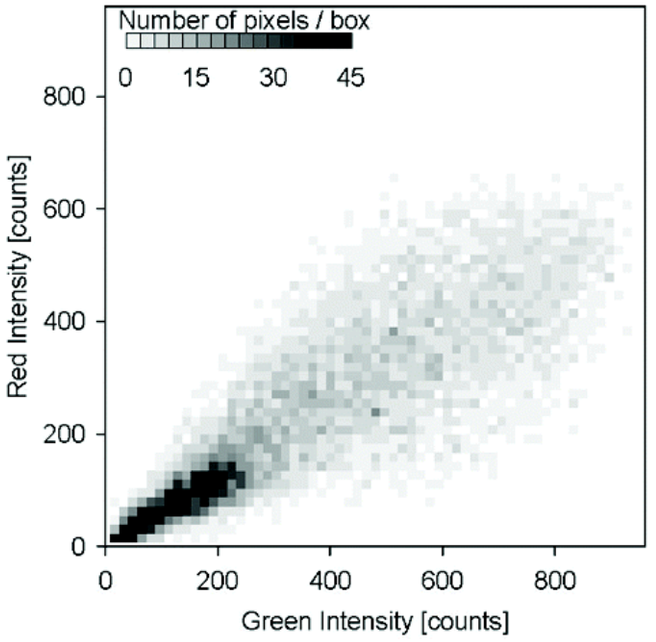


Figure 2

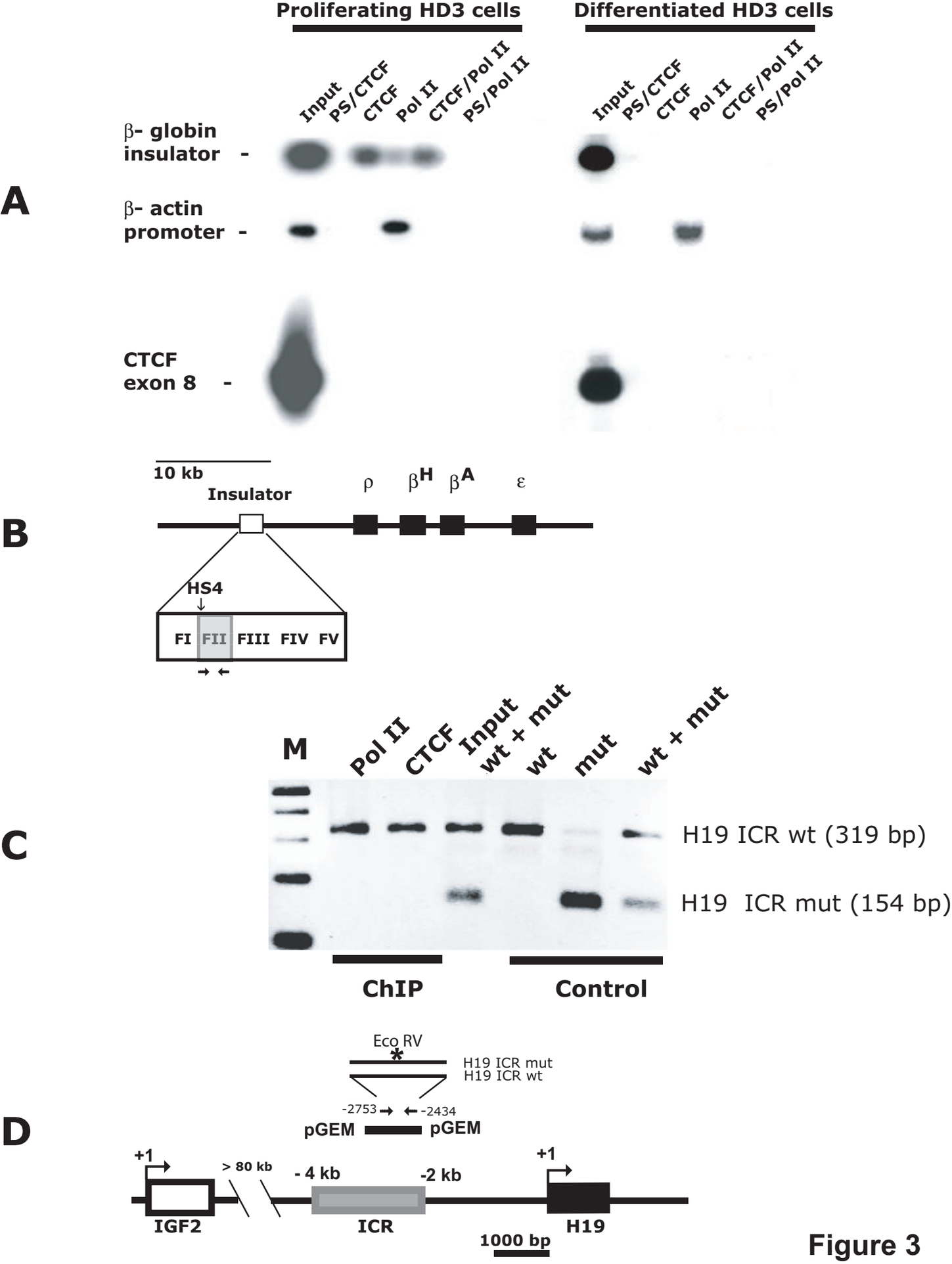
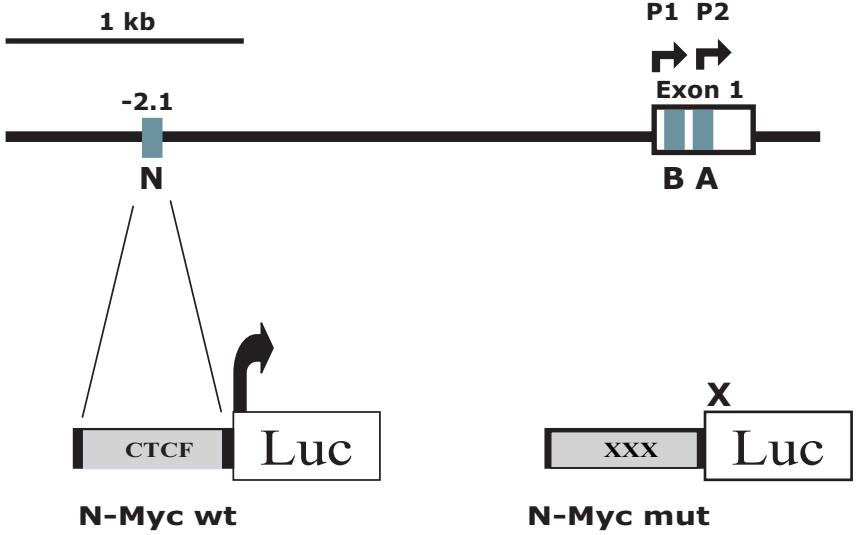
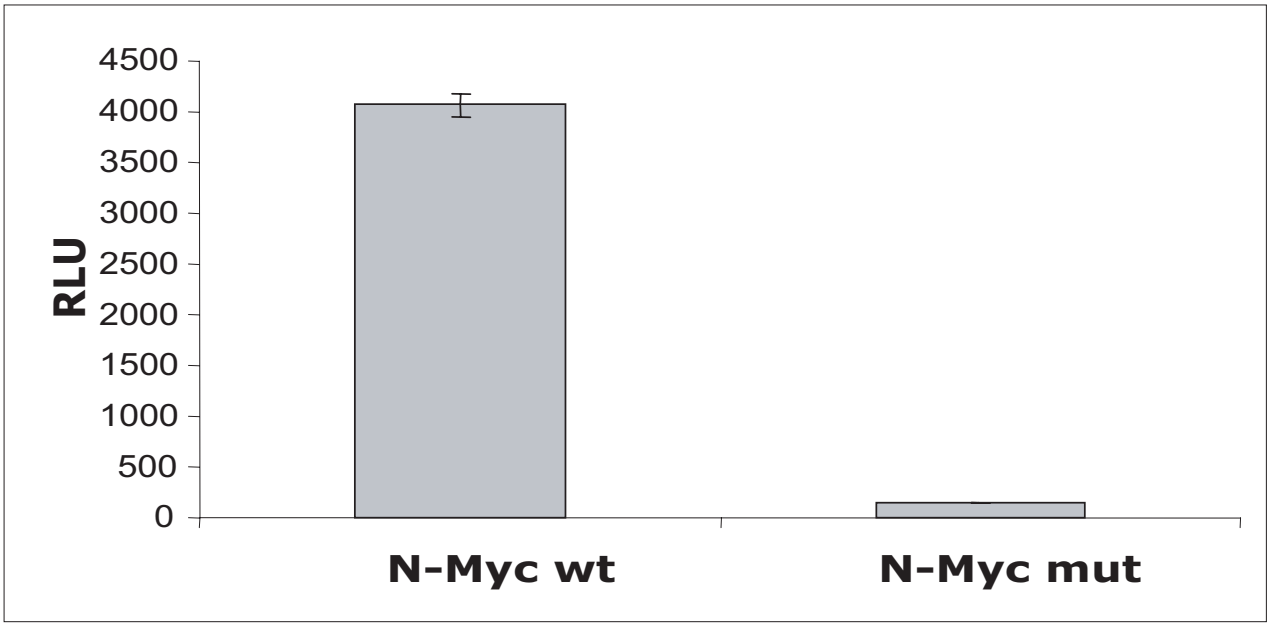

A
Proliferating HD3 cells
Differentiated HD3 cells
Input
PS/CTCF
CTCF
Pol II
CTCF/Pol II
PS/Pol II
β- globin insulator -
β- actin promoter -
CTCF exon 8 -
B
10 kb
Insulator
ρ
βH
βA
ε
HS4
FI
FII
FIII
FIV
FV
C
M
Pol II
CTCF
Input
wt + mut
wt
mut
wt + mut
H19 ICR wt (319 bp)
H19 ICR mut (154 bp)
ChIP
Control
D
Eco RV
H19 ICR mut
H19 ICR wt
-2753
-2434
pGEM
pGEM
+1
> 80 kb
- 4 kb
-2 kb
+1
IGF2
ICR
H19
1000 bp

Figure 3

A



B



C

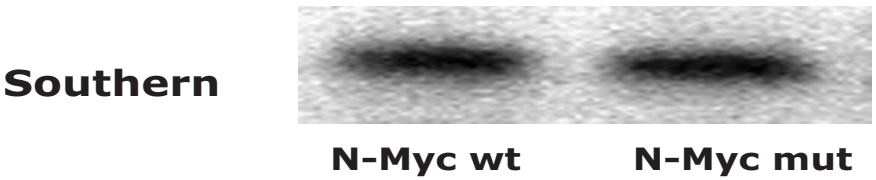
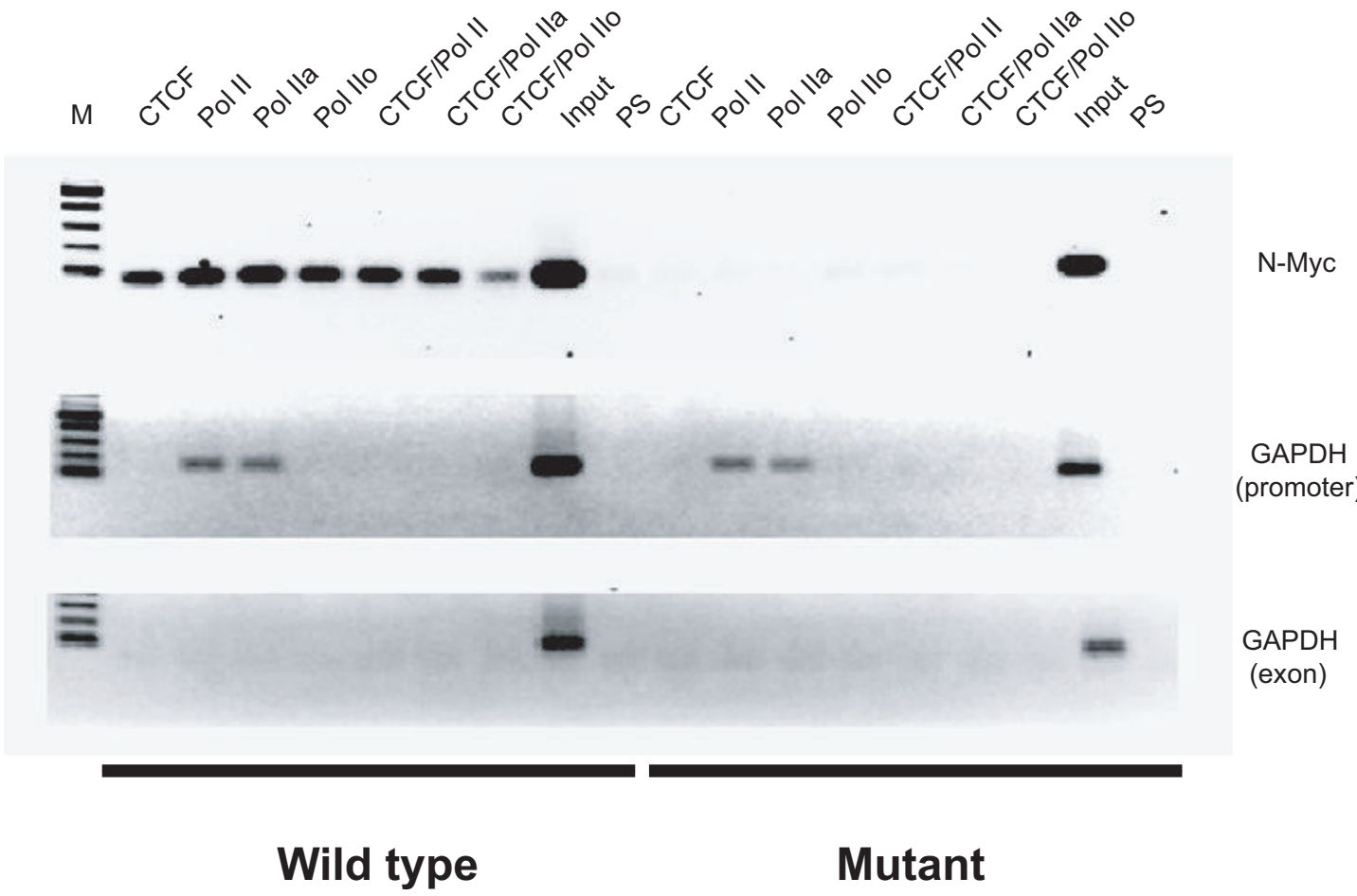


Figure 4ABC

D



E

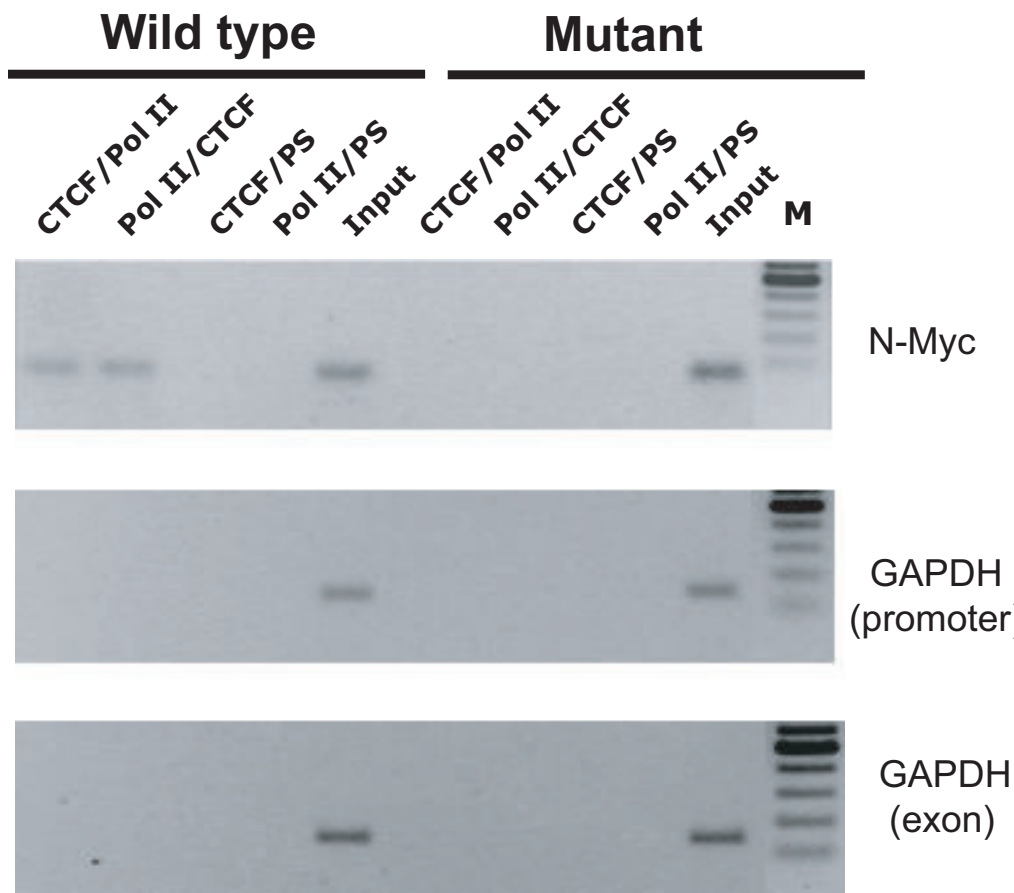


Figure 4 DE

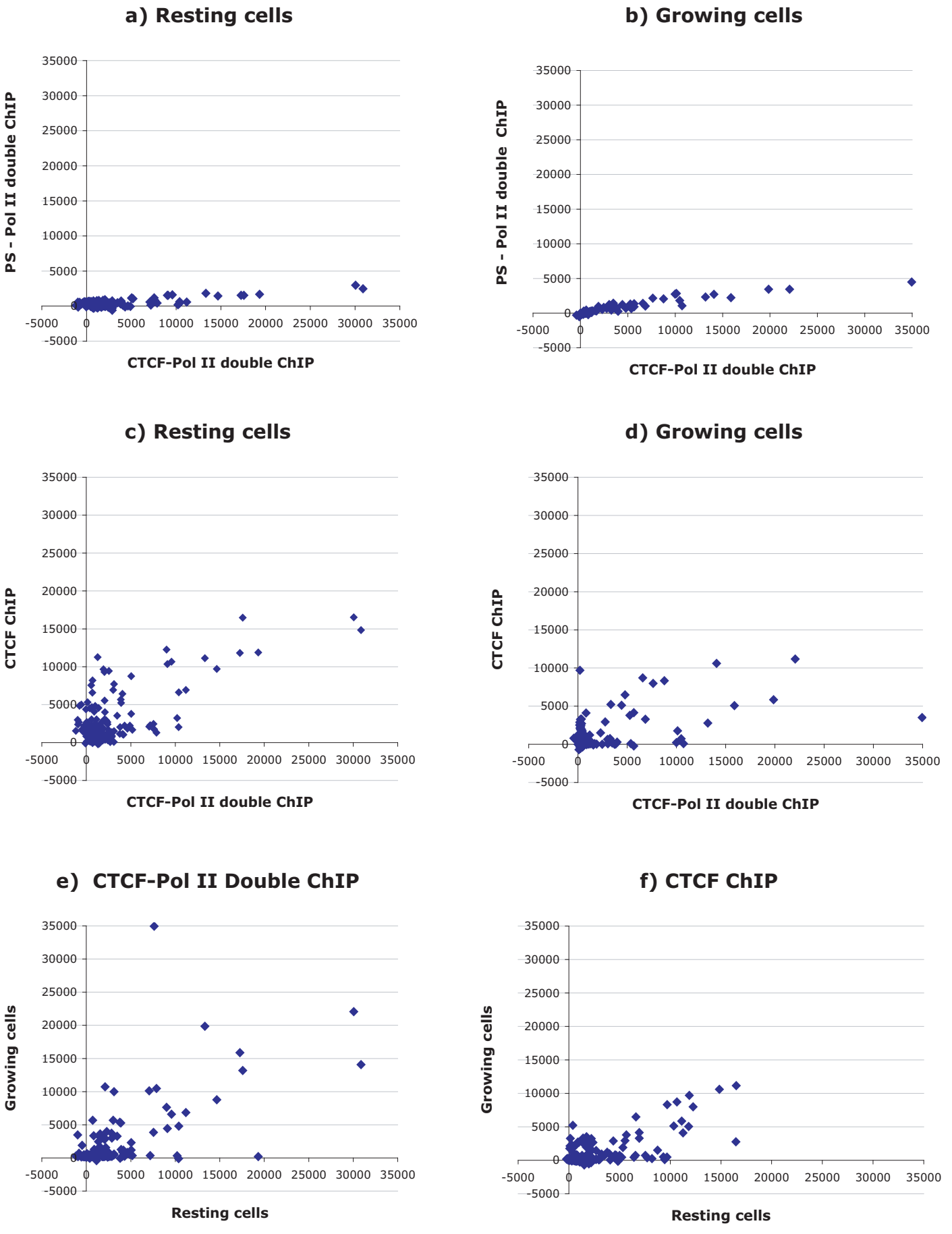

a) Resting cells
PS - Pol II double ChIP
CTCF-Pol II double ChIP
b) Growing cells
PS - Pol II double ChIP
CTCF-Pol II double ChIP
c) Resting cells
CTCF ChIP
CTCF-Pol II double ChIP
d) Growing cells
CTCF ChIP
CTCF-Pol II double ChIP
e) CTCF-Pol II Double ChIP
Growing cells
Resting cells
f) CTCF ChIP
Growing cells
Resting cells

Figure 5A

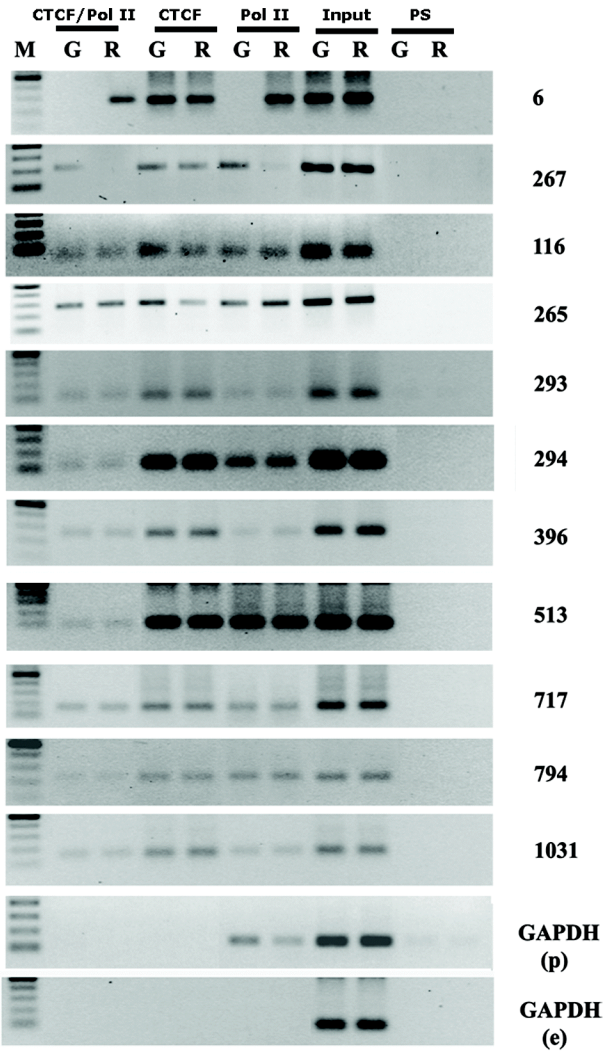

CTCF/Pol II
CTCF
Pol II
Input
PS
M G R G R G R G R G R
6
267
116
265
293
294
396
513
717
794
1031
GAPDH (p)
GAPDH (e)

FIgure 5B

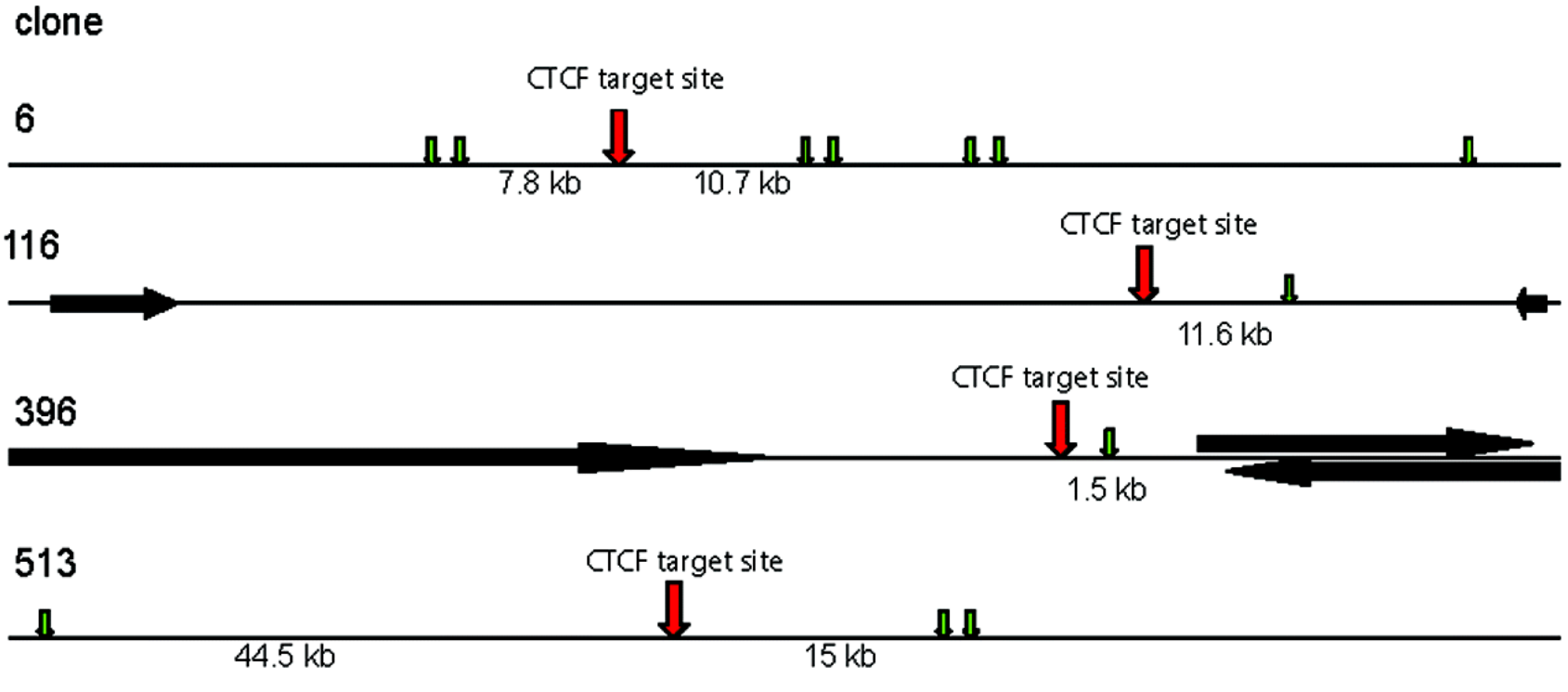

clone
6
CTCF target site
7.8 kb
10.7 kb
116
CTCF target site
11.6 kb
396
CTCF target site
1.5 kb
513
CTCF target site
44.5 kb
15 kb

Figure 5C

Table 1

Primers	GeneBank accession number	Forward primer (5' – 3' direction)	Reverse primer (5' – 3' direction)	PCR Fragment Size (bp)	PCR conditions (for all – initial denaturation: 94^0C 5 min; final elongation – 72^0C 10 min)
Microarray primers/ Clone numbers					
6	AY457178	tgcaggagaagcaatattat	tagaagaggtagaagaggaag	278	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.
116	AY457216	aggttttcaggctagacaga	tcttctggtcttctctgaatg	96	Cycle(x30): 95^0C, 30 sec; 53^0C 30 sec; 72^0C 20 sec.
265	AY457268	tctcagtggagaaaaactttgc	gtcgacaaatcagaagctga	270	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.
267	AY457269	gacctattgagaatgctcaca	atgattgttacctctctttg	235	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.
293	AY457285	ccactaaacctccttctcca	aagaaggctgtaggtggctgt	127	Cycle(x30): 95^0C, 30 sec; 63^0C 30 sec; 72^0C 15 sec.
294	AY457286	ccagcaagccctttaggaaca	tactcctacatcctgaaagtg	124	Cycle(x30): 95^0C, 30 sec; 59^0C 30 sec; 72^0C 20 sec.
396	AY457330	ctaatccttattgtacagga	catggaaattctactttgaa	264	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.
513	AY457372	cccttgctccatcttttgg	gtctgcagaagcacttgaag	94	Cycle(x30): 95^0C, 30 sec; 53^0C 30 sec; 72^0C 20 sec.
717	AY457431	gactaatattgaaaaatgtagc	ataggcattctggccttctgag	190	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.
794	AY457460	aagtattgaattttaggatt	ttgagaaccactgctctaac	234	Cycle(x30): 95^0C, 30 sec; 48^0C 30 sec; 72^0C 20 sec.
1031	AY457551	tcctctcggggttttctcca	aggaaaagacaaaaataaccc	138	Cycle(x30): 95^0C, 30 sec; 50^0C 30 sec; 72^0C 20 sec.

Primers for ChIP assays in HD3 cells					
β-globin insulator	E02199	gagctcacggggacagcccc	gatcccgtgccaccttcccc	174	Cycle (30x): 94°C 1 min, 60°C 1 min, 72°C 4 min
β-actin promoter	E02199	cgctccgaaagtttcctttt	agaaaagaaacgagccgtca	238	Cycle(x30): 94°C 1 min, 60°C 1 min, 72°C 4 min
CTCF exon 8	Z22605	cgagttttatgatacagaagtggaag	agttatttacaagcttgaccattacag	318	Cycle(x30): 94°C 1 min, 55°C 1 min, 72°C 4 min
Primers for Hybridization (nested)					
β-globin insulator	E02199	cccaaagcccccagggatgtaat	cccgggctgtccccgcacgct	130	Cycle (30x): 94°C 1 min, 60°C 1 min, 72°C 4 min
β-actin promoter	E02199	ccttttatggcgaggcggcggcgg	gggcgaaggcaacgcagcgactcc	84	Cycle(x30): 94°C 1 min, 60°C 1 min, 72°C 4 min
CTCF exon 8	Z22605	cgagttttatgatacagaagtggaag	agttatttacaagcttgaccattacag	301	Cycle(x30): 94°C 1 min, 55°C 1 min, 72°C 4 min

Primers for ChIP assays in transgenic NIH 3T3 cells					
N-site -Myc	X00364	acctgacccccgccctcgttga	ctctactggcagcagagatcat	58	Cycle(x30): 95^0C, 30 sec; 65^0C 30 sec; 72^0C 20 sec
GAPDH (exon 1)	X55448	atcatggcagagcaggtggc	gatgcacccatgatgataa	122	Cycle(x30): 95^0C, 30 sec; 58^0C 30 sec; 72^0C 20 sec.
GAPDH (promoter)	X55448	tcctgcaatgatagactag	ctgccaaacacgttcacaga	158	Cycle(x30): 95^0C 30 sec; 55^0C 30 sec; 72^0C 20 sec.
Primers for ChIP assays of wt and mut H19 ICR	AF049091	tccctttggtcactgaacc	aatccctatttgggtgaccc	319	Cycle(x30): 94^0 30 sec; 50^0 45 sec; 72^0 50sec

In both growing and resting cells			In only growing cells			In only resting cells		
Intra	Inter	Unident	Intra	Inter	Unident	Intra	Inter	Unident
5	4	10	1	0	0	0	1	5

Table 2

Clone ID	Gene	Location	Neighbouring gene (upstream)	Neighbouring genes (downstream)
Intronic/exonic CTCF target sites				
265	GDP-mannose dehydratase related	Intron 7	FOXc1	MAP/CAM kinase
294	Putative prostate cancer suppressor	Intron 1	Similar to cyclophillin A pseudogene	60S ribosomal protein pseudogene
717	Ahi1 isoform 1	Exon 20		myeloblastis oncogene
794		Intron 3	HnRNP-related	
1031	Krüppel-related Zn finger	Intron 1	GAPDH pseudogene	Zinc finger protein 140
* 267	Ring finger protein 144	Intron 3	NADH ubiquinone oxidoreductase related	Mo cofactor biosynthesis related
Intergenic CTCF target sites				
** 6			60S acidic ribosomal protein	Cbp/p300 interacting transactivator
116			basic helix-loop-helix	ADP-ribosylation related
293			adenomatosis polyposis coli binding; Dnmt3b	
396			membrane associated guanylate kinase superfamily	
513			cadherin-like	

* present only in growing cells
** present only in resting cells

Table 3