bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Poly(ADP-ribosyl)ation dependent changes in CTCF-chromatin binding and gene expression in breast cells

Ioanna Pavlaki[1,2], France Docquier[3], Igor Chernukhin[4], Vladimir B. Teif[1*] and Elena Klenova[1*]

[1]University of Essex, School of Biological Sciences, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK

Present addresses:

[2] Ioanna Pavlaki: Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

[3] France Docquier: Department of Medical Genetics, Academic Laboratory of Medical Genetics, University of Cambridge, Box 238, Lv 6 Addenbrooke's Treatment Centre, Cambridge Biomedical Campus, Cambridge CB2 0QQ

[4] Igor Chernukhin: Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE

* To whom correspondence should be addressed. Tel: (+44) (0)1206-874868; Fax: +44(0)1206872592; Email: klenovae@essex.ac.uk. Correspondence may also be addressed to Vladimir Teif, Email: vteif@essex.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

CTCF is an evolutionary conserved and ubiquitously expressed architectural protein, which regulates a plethora of cellular functions using different molecular mechanisms. The main form of CTCF with an apparent molecular mass of 130 kDa (CTCF130) has been extensively studied, however the properties of CTCF180, highly modified by poly(ADP-ribosyl)ation (PARylation) are not well understood. In this study we performed ChIP-seq and RNA-seq analyses in breast cells 226LDM, proliferating (with CTCF130 as the most abundant CTCF form) and arrested in the cell cycle (with only CTCF180). A dramatic reorganization of CTCF binding was observed during this transition, whereby CTCF was evacuated from many sites ("lost" group), although some sites retained modified CTCF ("common") and others acquired CTCF180 ("gained"). The classic CTCF binding motifs were identified for the former two groups, whereas the latter had no CTCF binding motif. Changes in CTCF occupancies in lost/common (but not gained) sites were associated with increased chromatin densities and altered expression from the neighboring genes. We propose a model integrating CTCF130/180 transition with CTCF-DNA binding and gene expression patterns, and functional outcomes. This study issues an important cautionary note concerning design and interpretation of any experiments using cells and tissues where CTCF180 may be present.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

The CCCTC-binding factor (CTCF), an evolutionarily conserved and ubiquitous transcription factor, is a key regulator of chromatin architecture and multiple cellular functions including transcriptional activation, silencing, insulation, mediation of long range chromatin interactions and others (1-6). Great efforts are currently being made to integrate CTCF ChIP-sequencing (ChIP-seq) data with other types of high-throughput data such as RNA-sequencing (RNA-seq) and chromatin signatures to better understand gene regulatory mechanisms and the interplay between CTCF genetic and epigenetic regulation (7,8). It is particularly important because CTCF binds to numerous sites of unclear functions in the human genome, and some of these binding sites differ between different cell types (9). Post-translational modifications of chromatin proteins (histones, transcription factors and others) play a significant role in the regulation of epigenetic processes. Poly(ADP-ribosyl)ation (PARylation) is one of such modification performed by poly(ADP-ribose) (PAR) polymerases (PARPs) (10,11). Phylogenetically ancient PARylation is involved in the regulation of numerous cellular functions, such as DNA repair, replication, transcription, translation, telomere maintenance, and chromatin remodeling (12-15).

A growing body of evidence demonstrates the link between CTCF and PARylation; for example, the insulator and transcription factor functions of CTCF have been found to be regulated by PARylation (16,17). The link between PARylation and CTCF is important in DNA damage response (18). Direct interaction between CTCF and poly(ADP-ribose) polymerase 1 (PARP1) and their co-localization in the genome have been reported (19-21). Furthermore, PARP1 and CTCF have been found to regulate the transition between active and repressed chromatin at the lamina (22). A highly PARylated form of CTCF is represented by a protein with molecular mass 180 kDa (CTCF180), whereas the commonly observed CTCF130, is hypo- or non-PARylated and appears in many immortalized cell lines and in cancer tissues (19,23-25). Interestingly, only CTCF180 was detected in the normal breast tissues, whereas both CTCF130 and CTCF180 were present in breast tumours (25). Generally, CTCF130 is associated with cell proliferation whereas CTCF180 is characteristic for non-proliferating cells of different types. Among those are cells from healthy breast tissues with very low proliferative index (25), cells with induced cell cycle arrest, DNA damage (25), senescence (26) or apoptosis (24,25). Currently, all existing information regarding the binding characteristics of CTCF has been mined from the experimental data obtained for CTCF130, but not CTCF180. It is not known whether the sets of targets for CTCF130 and CTCF180 are the same, completely different or overlap, and how binding of different forms of CTCF may be associated with alteration in gene expression.

In this study we utilised the immortalized human luminal breast cell line, 226LDM, in which the CTCF130 to CTCF180 transition could be induced following growth arrest (25), with the aim to analyse the genomic targets for CTCF130 and CTCF180, together with the corresponding

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcriptomes, in two functional states of 226LDM cells. The first state consists of control, proliferating cells predominantly containing CTCF130, while the second state represents cells in which proliferation blockade is chemically induced leading to the presence of CTCF180 only. The 226LDM cell model therefore provides us with the unique opportunity to study both CTCF forms, whereby overcoming the problem with the absence of a specific antibody against CTCF180.

MATERIAL AND METHODS

Cell Culture

226LDM cells, derived from human luminal breast cells, were propagated and treated as previously described (25). In brief, cells were seeded in flasks and grown in DMEM/F-12 (PAA) supplemented with 5 µg / ml insulin, 1 µg / ml hydrocortisone, 20 ng / ml epidermal growth factor, 20 ng / ml cholera toxin (all from Sigma), 10 % fetal bovine serum (FBS) (Biosera), and 50 µg / ml gentamicin (Life Technologies-Invitrogen) at 37°C and 5 % CO_2. To induce cell cycle arrest cells were treated with 100 mM hydroxyurea for 24 h followed by 1 h of complete medium and a further 24 h with 500 ng / ml nocodazole (SIGMA), and cells in suspension were then harvested for further analyses. Untreated adherent proliferating 226LDM cells were used as control.

Immunoblotting

The endogenous protein levels of CTCF were observed by SDS-PAGE/ western blot analysis (27,28) in whole cell lysates of 226LDM cells from the control and treated populations using a polyclonal anti-CTCF antibody (Millipore, 07-729, lot # JBC1903613). Anti-tubulin specific antibody (SIGMA, T5168) was used as a loading control. Chemiluminescence detection was performed with the Fusion FX7 gel documentation system (PeqLab) and the UptiLight (Interchim) reagents according to the manufacturer's instructions.

Protein Immunoprecipitation (IP)

This method was used to detect and immunoprecipitate CTCF out of a solution containing thousands of proteins present in 226LDM cells, using an antibody that specifically recognizes this protein (29). 226LDM cells cultured in a T75 flask were trypsinized, washed twice with PBS and then lysed by vortexing in BF2 (25 mM Tris/Hepes - pH 8.0, 2 mM EDTA, 0.5% Tween20, 0.5 M NaCl, 1:100 Halt protease inhibitor cocktail). The lysate was incubated on ice for 15 min and then equal volume of BF1 (25 mM Tris/Hepes - pH 8.0, 2 mM EDTA, 0.5% Tween20 and 1:100 Halt protease inhibitor cocktail) was added. For Immunoprecipitation, the cell lysate was pre-cleared by incubating 500 µl of the lysate in 50 µl of pre-blocked Protein A/Sepharose beads for 30 minutes at 4°C on a rotor shaker. The sample was then centrifuged at 200 x g for 1 minute at RT and the pre-cleared supernatant was transferred into a fresh centrifuge tube. 50 µl of the sepharose beads were added to the pre-cleared lysate along with the anti-CTCF antibody (Millipore, 07-729, lot # JBC1903613) and the samples were incubated overnight at 4°C on a rotating wheel. On the following day, the immune-complexes were

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

recovered by centrifugation at low speed and the supernatant was removed. The pellet was washed three times with immunoprecipitation buffer (BF1+BF2) and each time the beads were collected with centrifugation at low speed. The sepharose was then lysed in SDS-lysis buffer and analysed by SDS-PAGE and western blot analysis as described in previous section.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed using the ZymoSpin kit (Zymo Research USA) following the manufacturer's instructions. In brief, 5 x 10^6 of 226LDM cells from the control and the treated populations were cross-linked with formaldehyde. The cross-linking was quenched with glycine and the cells were washed twice with PBS with the addition of a protease inhibitor cocktail before pelleting at 1000 g for 1 min at 4°C. The pellet was lysed in Chromatin Shearing Buffer and sonicated using Bioruptor Plus (Diagenode) on high power to obtain fragments of 250-300 bp. ChIP reaction mixes containing sheared chromatin, Chromatin Dilution Buffer, anti-CTCF antibody (Millipore, 07-729, lot # JBC1903613 or no-antibody for negative control) and protease inhibitor cocktail were incubated rotating overnight at 4°C. The next day, ZymoMag Protein A beads were added to the mix and incubated for 1 h at 4°C. The complexes were washed with Washing Buffers I, II and III and then the beads were re-suspended in DNA Elution Buffer. Following de-crosslinking with Proteinase K at 65°C, the ChIP DNA was purified using the ZymoSpin IC columns. The samples were stored at -80°C. The concentration of DNA in the ChIP samples was measured using the NanoDrop 3300 fluorospectrophotometer (Thermo Scientific) along with the Quant-iT™ PicoGreen ds DNA assay kit according to the manufacturer's instructions.

RNA extraction

Total RNA from 226LDM cells (three biological replicates from the control and three from the treated population) was extracted using the TRIsure reagent (Bioline) according to the manufacturer's guidelines. Briefly, cells grown in a T75 flask were washed twice with PBS, then scraped off and pelleted at 300 g for 5 min. Following incubation with TRIsure for 5 min at RT, chloroform was added and the sample was incubated for 15 min at RT. After centrifugation at 9,500 g for 15 min at 4°C, the top aqueous layer was carefully extracted and the genetic material was precipitated with isopropanol for 20 min on ice. After centrifugation (9.500 g / 15 min / 4°C) the pellet was washed twice in 75 % ethanol before air-drying the obtained RNA pellet. The RNA was solubilized in sterile water (40-50 µl) and heated for 10 min at 55°C. The pellet was stored at -80°C. The RNA quality was tested using the Agilent Bioanalyzer system; the electropherographs are shown in Supplemental Figure S1.

Library Preparation and Sequencing

The library preparation and sequencing using the Illumina platforms were performed at the University College London (UCL) Genomics Centre.

ChIP-seq analysis

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Uniquely mappable reads were aligned to the human hg19 genome with the help of Bowtie (30) allowing up to 1 mismatch. Peak calling was done with MACS (31) using standard parameters, and taking into account the corresponding Input. Replicate experiments were analysed separately, and then obtained peaks were merged. The aggregate profiles were calculated using NucTools (32) and visualised using OriginPro (Origin Lab). Sequence motif analysis was performed using HOMER (33).

RNA-seq analysis

Reads were aligned using Novoalign 3.2 to the reference genome (hg19) and the raw counts were normalized to RPKM values using the Bam2rpkm tool from Galaxy. Differential expression was determined using DeSeq. Genes whose expression change was less than 1.5-fold were included in the "unchanged" gene expression category. Gene Ontology (GO) analysis was performed using DAVID (34), Revigo (https://www.ncbi.nlm.nih.gov/pubmed/21789182), Cytoscape (https://www.ncbi.nlm.nih.gov/pubmed/14597658) and Panther (https://www.ncbi.nlm.nih.gov/pubmed/27899595). The list of genes that were associated with CTCF binding sites in their promoter region was divided into upregulated/downregulated/no-change based on the RNA-seq data. The lists of upregulated/downregulated/no change genes associated with CTCF binding sites were intersected with the list of housekeeping genes from (35) in order to determine the enrichment of housekeeping genes in each category.

ChIP-seq and RNA-seq data have been deposited to the GEO archive (accession number GSE102237).

RESULTS

Analysis of CTCF binding and gene expression profiles in proliferating (control) and arrested (treated) 226LDM cells

The 226LDM cell line was chosen as a model to investigate binding patterns of CTCF130 and CTCF180 in the genome and correlate them with transcription activity. The main challenge was the absence of an antibody specifically recognising CTCF180; i.e. all existing anti-CTCF antibodies detect either just CTCF130 or both forms. The utilisation of 226LDM cells presents a unique opportunity to discriminate between these forms, because proliferating 226LDM cells predominantly contain CTCF130, whereas cells treated with hydroxyurea (HU) and nocodazole (NO), blocking the cell cycle in the S and G2-M phase, respectively, display only CTCF180 (25). The latter cells display clear morphological changes becoming rounded and suspended in the medium. Importantly, due to batch-to-batch variations, screening procedures are required to select the appropriate antibodies (25). Such tests were conducted in the current investigation and the antibodies which could recognize both CTCF130 and CTCF180 (Millipore, 07-729, lot # JBC1903613) were selected from the panel of several anti-CTC anti-CTCF antibodies (data not shown). Using these antibodies we confirmed the transition from CTCF130 to CTCF180 following 226LDM cells treatment ((Supplemental Figure S2)

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and also their ability to immunoprecipitate both CTCF forms (Supplemental Figure S3). These antibodies were then used for ChIP-sequencing (ChIP-seq) analysis of CTCF binding in control (proliferating) and arrested (treated) cells. RNA from these cells was also purified and RNA sequencing (RNA-seq) performed.

The analysis of total transcriptomes of control and treated cells revealed that 11,180 coding mRNAs were significantly differentially expressed in treated cells (2,276 up-regulated and 8,904 down-regulated). The ranked genes were then annotated using the Gene Ontology analysis; these data are shown in Supplemental Figures S4 and S5. The changes identified in the transcriptomes were consistent with the two biological states of the cells (proliferating *vs* arrested). Thus, genes involved in cell cycle arrest, development, differentiation and energy reserve metabolic processes were among up-regulated in treated cells, whereas genes associated with metabolic and cell signaling pathways, ion transport and cell adhesion were down-regulated. Interestingly, various RNA metabolic processes were affected in the latter group of genes. These datasets were later used to link changes in gene expression with CTCF130/CTCF180 binding (see below).

The analysis of CTCF ChIP-seq revealed that the number of CTCF target sites was considerably higher in control cells (n=9986) compared to treated cells (n=2271). Three groups of CTCF sites with different binding patterns in control and treated cells were observed, which were termed “common”, “lost” and “gained”. In the common group, CTCF was bound to the same sites in both cell states (as identified by ChIP-seq peaks), however, the CTCF occupancies varied. In the lost group CTCF binding was observed in control cells and not in treated cells, and in the gained group CTCF binding was only observed in treated cells (Figure 1A). The majority of the sites (9729) were lost after treatment; 2014 sites were gained and 257 sites were common (Figure 1B).

The analysis of expression of genes containing CTCF in their promoter regions, within +/- 10,000 bp from the Transcription Start Sites (TSS), showed that, collectively for all three groups, most of these genes were down-regulated (1169 or 49.6%); 443 (18.8%) were up-regulated and 744 (31.6%) unchanged (Figure 1C, left panel). The calculations of the relative number of genes showed similar patterns in individual groups (common, lost and gained) whereby the majority of the genes (~50-55%) were down-regulated and ~10-20% were up-regulated (Figure 1C, middle panel). The absolute numbers of genes associated with changes in their expression were considerably higher in the lost group (Figure 1C, right panel). This is not surprising since the majority of the CTCF binding sites belonged to this group (Figure 1B). Expression did not change in a considerable percentage of genes in these groups (~30-35%). A large proportion of genes associated with the presence of CTCF belongs to housekeeping genes (35), with 39%, 42% and 20% up-regulated, and 14%, 15% and 30% down-regulated in the lost, gained and common groups, respectively. No changes were found in 47% of lost and gained, and 50% of common groups (Figure 1D).

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene ontology analysis of transcriptional changes revealed genes highly up- or down-regulated in three different groups (Figure 1E). In the common group, highly up-regulated genes were associated with differentiation and down-regulated genes – with cell migration and apoptosis. In the lost, the largest group, highly up-regulated genes were enriched in categories associated with anti-apoptotic processes, whereas most of the highly down-regulated genes were associated with cell cycle and cell migration processes. In the gained group, both highly up-regulated and down-regulated genes were enriched in categories regulating RNA Pol II transcription.

Next, we assessed the relative enrichment of gene ontology terms of the genes with promoters containing CTCF (Supplementary Figure S6). In the common group, up-regulated genes were enriched in developmental processes and down-regulated genes were enriched in response to metal ions. In the lost group, up-regulated genes were enriched in ion binding and homeostasis processes, whereas most of the down-regulated genes were associated with signal transduction and adhesion processes. In the gained group, up-regulated genes were enriched in those involved in the regulation of macromolecular complexes and membrane transporter activities, whereas down-regulated genes were enriched in genes involved in nucleotide binding and biosynthetic processes. Taken together, these experiments demonstrate characteristic CTCF binding and the corresponding gene expression patterns for cells containing predominantly CTCF130 (control) and CTCF180 only (treated).

Relationship between CTCF occupancy and gene expression in control and treated cells

Next, we investigated the relationship between changes in CTCF binding and gene expression. By stratifying all CTCF-containing promoters (within +/-10,000 bp from TSS) according to the level of expression from the corresponding gene, we observed that it was more likely to find CTCF at a promoter of a higher expressed gene in both control and treated cells (Figures 2A and 2B). Furthermore, due to the loss of CTCF from promoters of many low-expressed genes upon treatment, this effect is more pronounced in treated cells (Figure 2B). Interestingly, when we stratified genes by their expression fold change upon treatment (Figure 2C), it appeared that there was a clear preference for retained CTCF at common sites at promoters to be associated with genes which did not change or changed their expression minimally. Genes considerably up- or down-regulated upon treatment have lost CTCF at their promoters (see the leftmost and rightmost parts of Figure 2C).

We also correlated changes in gene expression with the changes in CTCF occupancies for the three groups of CTCF sites within a more narrow window +/-1000 bp from TSS as typically done in this type of studies (36). In agreement with observations above (Figure 2C), for the group of genes contained common CTCF sites at their promoters the changes in gene expression were relatively small (Figure 3A), while promoters which lost or gained CTCF were associated with a much broader range of gene expression levels (Figure 3, panels B and C, respectively). No correlation was observed between CTCF occupancy and gene expression in the common and lost groups, although small positive correlation (r=0.15) was seen in the gained group.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Common and lost, but not gained, CTCF sites contain classical CTCF binding motifs

CTCF employs a combination of its eleven Zinc fingers to bind to diverse DNA sequences with a central ~20 bp core DNA motif critical for CTCF binding (37). Most of CTCF sites contain the classical consensus sequence, however CTCF sites with different consensus motifs and also sites which do not match any consensus motifs have been also identified previously (37-40). Since the composition of the motif may be linked to CTCF functions, we compared the CTCF binding sites identified in the three groups. We calculated the nucleotide frequencies as a function of distance from the summit of CTCF ChIP-seq peak for the subsets of the sited from the common, lost and gained groups. As shown in Figure 4, CTCF sites in the common and lost but not the gained groups contain classical CTCF recognition motif, enriched with the guanine and cytosine residues at the summit, although this pattern is more pronounced for the common sites. Interestingly, the nucleotide distribution in the 3' and 5' flanking regions of the motifs significantly differs between these groups, demonstrating higher GC content in the common group. This is in line with our previous observation that common but not lost/gained sites were enriched inside CpG islands for the system of mouse embryonic stem differentiation (41)

We have also assessed the strength of CTCF binding for different classes of CTCF sites calculated by the heights of the ChIP-seq peaks (Figure 4D). The strongest binding was observed in the control cells in the common group, which on average almost did not change after treatment. The initial CTCF signal in the lost group was smaller than in the common sites before treatment, and it significantly decreased after treatment. The lowest signal was in the gained group which most likely reflected the nature of CTCF180 binding (very weak or DNA-independent).

From these analyses we conclude that common and lost sites are characterised by the presence of the CTCF consensus motif. The strongest CTCF binding is observed for common sites, whereas it is weaker for the lost sites. The gained sites have no classical CTCF motif and this may be a reason why their CTCF binding is the lowest.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Loss of CTCF binding upon cell treatment is associated with chromatin redistributions

The importance of CTCF in the regulation of the chromatin structure has been widely recognized. For example, CTCF binding often demarcates distinct chromatin states and protects DNA from methylation (2,5,42). Using our model system, we investigated whether the presence of CTCF130 and CTCT180 at the sites in three different groups was associated with the changes in the chromatin density. In our first analysis, the accessibility of chromatin to shearing in the input samples was considered, as in several previous studies (43,44). Figure 5 shows that CTCF130 bound regions in the common control group are associated with less dense chromatin than at the same regions in treated cells. This also correlated with the reduced strength of CTCF180 binding at these sites after treatment (Figure 4D). In the lost group, the density of chromatin increased after treatment and release of CTCF180 (Figure 5B), whereas in the gained group the density of chromatin at CTCF180 binding sites did not change following CTCF recruitment (Figure 5C). Taken together, these findings indicate that, in comparison with CTCF180 in treated cells, CTCF130 in control cells is associated with less dense chromatin and stronger binding to its sites. These data are in line with our previous reports on the competition of CTCF with nucleosomes (45-47). Note that the chromatin peaks in Figure 5 correspond to profiles averaged over hundreds of CTCF sites, and thus, they are expected to be symmetric with respect to CTCF. Of note, the nucleosome oscillations around CTCF observed in the previous studies (45-47) are not revealed here, since our samples were not digested with MNase, but sonicated.

Chromatin redistributions reported in Figure 5 can be associated with different covalent modifications of histones. It was beyond the scope of this work to test all histone modifications systematically, therefore, as a test case, we have profiled only the H3K9me3 modification which is known to be associated with higher nucleosome density (46). Figure 6 (panels A and B) shows the results of the H3K9me3 ChIP-seq analysis demonstrating significant H3K9me3 redistributions around common and lost CTCF sites. Interestingly, no such rearrangements around gained CTCF sites were observed, which is in line with our previous findings that the mechanisms of CTCF binding at gained sites is different from classical CTCF/nucleosome competition and CTCF-DNA recognition.

Since PARylation can change physical CTCF interactions with chromatin proteins, we have also looked at the chromatin profiles near individual CTCF sites (not at the CTCF peak summits, but in the neighboring regions). Examples of specific gene promoter regions where CTCF-associated chromatin rearrangements take place following treatment, together with changes in gene expression patterns, are given in Supplementary Figure S7. We noted that CTCF binding in control cells was associated with sharp chromatin peaks in the physical proximity to CTCF, which disappears upon cell treatment. In ddition, in some cases the chromatin peaks and CTCF happen at different ends of the gene (e.g. a chromatin peak at the transcription start site (TSS), and CTCF at the transcription end site (TES) in the case of E2F4. The latter suggests possible TSS-TES bridging by CTCF in control cells, which disappears after treatment. The effect of CTCF-dependent chromatin reorganization was observed for all three groups of sites (common, lost and gained), and was not correlated with changes of gene

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression (gene expression could go either up or down following treatment). The fact that CTCF-dependent chromatin peaks were next to CTCF but did not coincide with it, provides an argument that the chromatin peak is not formed by CTCF itself. The depletion of chromatin peaks near CTCF sites at the PARP3 and TP53 promoters was also confirmed by ChIP experiments with the primers residing inside the corresponding chromatin peaks (panels D and E in Supplementary Figure 7). We also analyzed profiles of H3K9me3 chromatin marks in the same promoters. As shown in Supplemental Figure S8, the strength of H3K9me3 signal increases around the regions near CTCF sites which lost chromatin peaks. This suggests that perhaps the lost sharp chromatin peaks represent specific chromatin complexes other than nucleosomes. It did not escape our attention that the situation depicted in Supplementary Figure S7 may represent a general effect, but exploring the nature of such chromatin peaks would require additional extensive experiments which are beyond the scope of the current work. One should be also cautious to not over-interpret such chromatin peaks due to possible artifacts (43,48-50).

DISCUSSION

This study was aimed to analyse DNA targets for CTCF130 and CTCF180, together with the transcriptomes, in 226LDM cells successfully used in our previous studies (25). In the absence of a specific anti-CTCF180 antibody, it was rational to use a cell model in which a switch from CTCF130 to CTCF180 can be induced and validated anti-CTCF antibodies recognizing both CTCF130 and CTCF180 applied to immunoprecipitate individual CTCF population to identify DNA targets specific for each form. Following optimization of hydroxyurea and nocodazole concentrations (26), it was possible to obtain cells with CTCF180 only ("treated") whereas CTCF130 was predominantly present in proliferating cells ("control"). The generation of CTCF180 in response to the drugs can be explained by initiation of checkpoint signalling cascades, leading to activation of PARP enzymes and subsequent PARylation of CTCF. Indeed, the nocodazole - (51) and hydroxyurea-induced (52) cell cycle arrests have been linked to the activation of the PARP-signalling pathways (53,54). Global changes in gene expression profiles were consistent with the changes in the biological states of the cells, revealing up-regulation of genes involved in cell cycle arrest, development, differentiation and energy reserve metabolic processes and down-regulation of genes associated with metabolic and cell signaling pathways, ion transport and cell adhesion (Supplemental Figures S4 and S5).

The analysis of the ChIP-seq confirmed for the first time that CTCF180 has well-defined genomics targets, paving the way for further research into the specifics of this binding in different conditions, cell lines or tissues. The number of sites occupied by CTCF180 was found to be much smaller than in control cells (n=2271 vs n=9986, respectively). This may be due to reduction of total CTCF and/or re-localization of at least some of the CTCF180 molecules into the cytoplasm after cell cycle arrest (Supplemental Figures S2 and S9). Such CTCF distribution pattern has previously been reported in normal breast tissues where only CTCF180 is detected (25,55). The presence of smaller number of

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CTCF sites in the genome of the treated cells may indicate that, individually, these sites organize and regulate larger chromatin domains. Moreover, the molecular basis of these networks is likely to be different because of the particular nature of the binding sites of CTCF180. These aspects will need to be explored in the future.

This study provides new insights on DNA-binding and gene regulatory properties of CTCF180 (summarized in Figure 7). Common and lost sites contain the classical CTCF motif, although the former are more GC-rich at the summit and in the background around the motif, whereas the latter are embedded into more AT-rich sequences. The effect of common CTCF sites residing in more GC-rich areas has been previously reported in our study of mouse embryonic stem cell differentiation (41) and it seems to be a general effect. Such properties of these sequences may be linked to the strength of CTCF binding, which is higher in the common group than in the lost group (Figure 4D), thereby defining the behaviour of CTCF after PARylation. In both groups, the increase in chromatin densities was observed in the regions overlapping with CTCF sites. Changes in chromatin compaction resulting from loss of CTCF have been previously reported (46,56,57) and it could be one of the mechanisms explaining the effects of CTCF180. Alternatively, although not mutually exclusive, epigenetic modifications of chromatin caused by the treatment, might be responsible for the release of CTCF180. It is in agreement with down-regulation of the majority of the genes within these regions, however considerable number of up-regulated genes were also observed thus implying the involvement of additional regulatory factors/mechanisms (57-59). The complexity of this regulation is further illustrated by the observations that changes of CTCF binding at promoters were associated with larger chromatin rearrangements in the neighbouring locations not overlapping with the CTCF binding site (Supplemental Figures S7 and S8).

No CTCF binding motif was observed in the gained group suggesting that CTCF180 may be binding to rare non-canonical CTCF sites. It is also possible that CTCF180 interacts with these regions in a DNA-independent manner, directly or through recruitment by other proteins. At least one of histone modifications redistributing at common and lost (but not gained) CTCF binding sites was found to be H3K9me3, however it is important that a wider range of chromatin marks is tested in future studies.

The lost CTCF sites localized in promoter regions are associated with the highest numbers of genes whose expression is affected (80.1%), followed by the gained (15.3%) and common (4.6%) sites. Unchanged expression of a significant number of genes in the lost and gained groups indicates that the regulation of these genes does not depend on CTCF binding. This may not be the case for the common group in which CTCF may be important to sustain the optimal level of expression of certain genes needed for survival of cells in different functional states. According to our study, this effect is especially relevant for housekeeping genes (Figures 1D and 3).

The importance of CTCF modification in the biological processes is supported by changes in expression profiles of genes associated with CTCF (Figures 1E and 7, far right). These changes

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involve down-regulation if genes involved in cell cycle and cell migration, and up-regulation of genes involved in differentiation thereby adequately reflecting the biological situation, i.e. transition from proliferating to arrested cells. Furthermore, some of the affected genes appear to be characteristic for particular groups of CTCF sites, for example, genes responsible for cell cycle regulation are down-regulated in the group of genes where CTCF is lost. It is tempting to speculate that such preference may be due to the change of behaviour of PARylated CTCF at the particular type of CTCF sites.

It should be noted that in this report we suggest how CTCF PARylation may control its DNA binding properties and, subsequently, changes in local chromatin and gene expression. It was beyond our scope to consider in this experimental model global effects of CTCF on higher order chromatin structures, which can be expected from CTCF as an architectural protein (9,57-61).

Finally, this study issues an important cautionary note concerning design and interpretation of any experiments using cells and tissues where CTCF180 may be present and can go undetected since not all antibodies can recognize this form of CTCF. The 226LDM cells as a model for the switch from CTCF130 to CTCF180 provided us with a unique opportunity to develop an experimental framework to study CTCF180. This approach can be used to investigate the role of CTCF180 in cell lines and tissues, normal and tumor, where either both forms or exclusively CTCF180 are present. The screening of the existing antibodies for their ability to recognise either both forms of CTCF or CTCF130 only will be necessary to enable to subtract the targets recognised by CTCF130 from the combined CTCF130/CTCF180.

ACKNOWLEDGEMENTS

We thank Parmjit Jat and Michael O'Hare for 226LDM cells and Mike Hubank and colleagues from the University College London (UCL) Genomics Centre for sequencing and initial bioinformatics analysis of the data. We also thank Adele Angel for excellent technical assistance and Svetlana Gretton, Theodoros Giakoumis and Yemane Tedros for their help with ChIP sample preparations. We thank members of Vladimir Teif's, Elena Klenova's and Greg Brooke's laboratories for experimental advice and many helpful discussions.

Author contributions: F.D., I.P. and E.K. conceived the project, I.P., F.D. and E.K. performed experiments. I.P., I.C., V.T. and E.K. analysed the data, E.K., I.P. and V.T. wrote the manuscript.

FUNDING

This work was supported by Medical Research Council (G0401088 to E.K.); Breast Cancer Campaign (2004Nov45 to F.D. and E. K.), Helen Rollason Cancer Charity and Faculty of Medical Sciences, Postgraduate Medical Institute, Anglia Ruskin University (to F. D. and I.P.); University of

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Essex (to E.K. and I.P); Wellcome Trust (200733/Z/16/Z to VT). Funding for open access charge: Medical Research Council.

REFERENCES

1. Ong, C.T. and Corces, V.G. (2014) CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, **15**, 234-246.
2. Ohlsson, R., Lobanenkov, V. and Klenova, E. (2010) Does CTCF mediate between nuclear organization and gene expression? *BioEssays : news and reviews in molecular, cellular and developmental biology*, **32**, 37-50.
3. Bonora, G., Plath, K. and Denholtz, M. (2014) A mechanistic link between gene regulation and genome architecture in mammalian development. *Current opinion in genetics & development*, **27**, 92-101.
4. Ali, T., Renkawitz, R. and Bartkuhn, M. (2016) Insulators and domains of gene expression. *Curr Opin Genet Dev*, **37**, 17-26.
5. Merkenschlager, M. and Odom, D.T. (2013) CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*, **152**, 1285-1297.
6. Holwerda, S.J. and de Laat, W. (2013) CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*, **368**, 20120369.
7. Angelini, C. and Costa, V. (2014) Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. *Front Cell Dev Biol*, **2**, 51.
8. Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A. and Tegner, J. (2014) Data integration in the era of omics: current and future challenges. *BMC Syst Biol*, **8 Suppl 2**, I1.
9. Lu, Y., Shan, G., Xue, J., Chen, C. and Zhang, C. (2016) Defining the multivalent functions of CTCF from chromatin state and three-dimensional chromatin interactions. *Nucleic acids research*, **44**, 6200-6212.
10. Kraus, W.L. and Hottiger, M.O. (2013) PARP-1 and gene regulation: progress and puzzles. *Molecular aspects of medicine*, **34**, 1109-1123.
11. Thomas, C. and Tulin, A.V. (2013) Poly-ADP-ribose polymerase: machinery for nuclear processes. *Molecular aspects of medicine*, **34**, 1124-1137.
12. Golia, B., Singh, H.R. and Timinszky, G. (2015) Poly-ADP-ribosylation signaling during DNA damage repair. *Front Biosci (Landmark Ed)*, **20**, 440-457.
13. Robert, I., Karicheva, O., Reina San Martin, B., Schreiber, V. and Dantzer, F. (2013) Functional aspects of PARylation in induced and programmed DNA repair processes: preserving genome integrity and modulating physiological events. *Molecular aspects of medicine*, **34**, 1138-1152.
14. Burkle, A. (2000) Poly(ADP-ribosyl)ation: a posttranslational protein modification linked with genome protection and mammalian longevity. *Biogerontology*, **1**, 41-46.
15. Li, M. and Yu, X. (2015) The role of poly(ADP-ribosyl)ation in DNA damage response and cancer chemotherapy. *Oncogene*, **34**, 3349-3356.
16. Caiafa, P. and Zlatanova, J. (2009) CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *Journal of cellular physiology*, **219**, 265-270.
17. Klenova, E. and Ohlsson, R. (2005) Poly(ADP-ribosyl)ation and epigenetics. Is CTCF PARt of the plot? *Cell cycle*, **4**, 96-101.
18. Han, D., Chen, Q., Shi, J., Zhang, F. and Yu, X. (2017) CTCF participates in DNA damage response via poly(ADP-ribosyl)ation. *Sci Rep*, **7**, 43530.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19. Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A. and Klenova, E. (2010) Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. *Mol Cell Biol*, **30**, 1199-1216.
20. Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L. and Caiafa, P. (2008) CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem*, **283**, 21873-21880.
21. Nalabothula, N., Al-Jumaily, T., Eteleeb, A.M., Flight, R.M., Xiaorong, S., Moseley, H., Rouchka, E.C. and Fondufe-Mittendorf, Y.N. (2015) Genome-Wide Profiling of PARP1 Reveals an Interplay with Gene Regulatory Regions and DNA Methylation. *PloS one*, **10**, e0135410.
22. Zhao, H., Sifakis, E.G., Sumida, N., Millan-Arino, L., Scholz, B.A., Svensson, J.P., Chen, X., Ronnegren, A.L., Mallet de Lima, C.D., Varnoosfaderani, F.S. *et al.* (2015) PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Mol Cell*, **59**, 984-997.
23. Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C. *et al.* (2004) Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet*, **36**, 1105-1110.
24. Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M.T., Renkawitz, R. *et al.* (2006) Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J Cell Sci*, **119**, 1746-1759.
25. Docquier, F., Kita, G.X., Farrar, D., Jat, P., O'Hare, M., Chernukhin, I., Gretton, S., Mandal, A., Alldridge, L. and Klenova, E. (2009) Decreased poly(ADP-ribosyl)ation of CTCF, a transcription factor, is associated with breast cancer phenotype and cell proliferation. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **15**, 5762-5771.
26. Pavlaki, I. (2016) PhD Thesis, University of Essex, UK, Colchester, Essex, UK.
27. Chrambach, A. and Rodbard, D. (1971) Polyacrylamide gel electrophoresis. *Science*, **172**, 440-451.
28. Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*, **76**, 4350-4354.
29. Kaboord, B. and Perr, M. (2008) Isolation of proteins and protein complexes by immunoprecipitation. *Methods in molecular biology*, **424**, 349-364.
30. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.
31. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.
32. Vainshtein, Y., Rippe, K. and Teif, V.B. (2017) NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data. *BMC Genomics*, **18**, 158.
33. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576-589.
34. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, **4**, P3.
35. Eisenberg, E. and Levanon, E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet*, **29**, 569-574.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36. Teif, V.B., Erdel, F., Beshnova, D.A., Vainshtein, Y., Mallm, J.P. and Rippe, K. (2013) Taking into account nucleosomes for predicting gene expression. *Methods*, **62**, 26-38.
37. Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V. and Ren, B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, **128**, 1231-1245.
38. Ziebarth, J.D., Bhattacharya, A. and Cui, Y. (2013) CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res*, **41**, D188-194.
39. Plasschaert, R.N., Vigneau, S., Tempera, I., Gupta, R., Maksimoska, J., Everett, L., Davuluri, R., Mamorstein, R., Lieberman, P.M., Schultz, D. *et al.* (2014) CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic acids research*, **42**, 774-789.
40. Essien, K., Vigneau, S., Apreleva, S., Singh, L.N., Bartolomei, M.S. and Hannenhalli, S. (2009) CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol*, **10**, R131.
41. Teif, V.B., Beshnova, D.A., Vainshtein, Y., Marth, C., Mallm, J.P., Hofer, T. and Rippe, K. (2014) Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res*, **24**, 1285-1295.
42. Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K. and Zhao, K. (2009) Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, **19**, 24-32.
43. Kremsky, I., Bellora, N. and Eyras, E. (2015) A Quantitative Profiling Tool for Diverse Genomic Data Types Reveals Potential Associations between Chromatin and Pre-mRNA Processing. *PLoS One*, **10**, e0132448.
44. Schwartz, Y.B., Kahn, T.G. and Pirrotta, V. (2005) Characteristic low density and shear sensitivity of cross-linked chromatin containing polycomb complexes. *Molecular and cellular biology*, **25**, 432-439.
45. Teif, V.B., Beshnova, D.A., Vainshtein, Y., Marth, C., Mallm, J.P., Höfer, T. and Rippe, K. (2014) Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Research*, **24**, 1285-1295.
46. Teĭf, V.B., Shkrobkov, A.V., Egorova, V.P. and Krot, V.I. (2012) [Nucleosomes in gene regulation: theoretical approaches]. *Mol Biol (Mosk)*, **46**, 3-13.
47. Beshnova, D.A., Cherstvy, A.G., Vainshtein, Y. and Teif, V.B. (2014) Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions. *PLoS Comput Biol*, **10**, e1003698.
48. Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrancois, P., Struhl, K., Gerstein, M. and Snyder, M. (2009) Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci U S A*, **106**, 14926-14931.
49. Teytelman, L., Thurtle, D.M., Rine, J. and van Oudenaarden, A. (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A*, **110**, 18602-18607.
50. Teytelman, L., Ozaydin, B., Zill, O., Lefrancois, P., Snyder, M., Rine, J. and Eisen, M.B. (2009) Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One*, **4**, e6700.
51. Hayashi, M.T. and Karlseder, J. (2013) DNA damage associated with mitosis and cytokinesis failure. *Oncogene*, **32**, 4593-4601.
52. Chaudhury, I. and Koepp, D.M. (2016) Recovery from the DNA Replication Checkpoint. *Genes*, **7**.
53. Bryant, H.E., Petermann, E., Schultz, N., Jemth, A.S., Loseva, O., Issaeva, N., Johansson, F., Fernandez, S., McGlynn, P. and Helleday, T. (2009) PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *The EMBO journal*, **28**, 2601-2615.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

54. Yelamos, J., Farres, J., Llacuna, L., Ampurdanes, C. and Martin-Caballero, J. (2011) PARP-1 and PARP-2: New players in tumour development. *American journal of cancer research*, **1**, 328-346.
55. Docquier, F., Farrar, D., D'Arcy, V., Chernukhin, I., Robinson, A., Loukinov, D., Vatolin, S., Pack, S., Mackay, A., Harris, R. *et al.* (2005) Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer research*, **65**, 5112 - 5122.
56. Tark-Dame, M., Jerabek, H., Manders, E.M., van der Wateren, I.M., Heermann, D.W. and van Driel, R. (2014) Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput Biol*, **10**, e1003877.
57. MacQuarrie, K.L., Fong, A.P., Morse, R.H. and Tapscott, S.J. (2011) Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet*, **27**, 141-148.
58. Lopes Novo, C. and Rugg-Gunn, P.J. (2016) Chromatin organization in pluripotent cells: emerging approaches to study and disrupt function. *Briefings in functional genomics*, **15**, 305-314.
59. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. and Bruneau, B.G. (2017) Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, **169**, 930-944 e922.
60. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015) CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, **162**, 900-910.
61. de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H. and de Laat, W. (2015) CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*, **60**, 676-684.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES LEGENDS

Figure 1. Analysis of CTCF binding and gene expression profiles in control and treated 226LDM cells. (A) Schematic illustration of three groups of CTCF sites detected by ChIP-seq in control and treated cells: common sites are present in both cell states, lost sites are present only in control cells, and gained sites appear only in treated cells. (B) A pie chart showing the numbers of common, lost and gained CTCF sites. (C) Association of gene expression patterns in control and treated cells with the three groups of CTCF binding sites present within +/- 10,000 bp from TSS. Numbers of genes up-regulated, down-regulated and unchanged is shown for all promoters with CTCF sites (left). Percentages and numbers of genes with different expression patterns for each group of CTCF-binding sites are shown in the middle and right panels, respectively. (D) Association of gene expression patterns in control and treated cells with the three groups of CTCF binding sites present within +/- 10,000 bp from TSS of housekeeping genes. Numbers of genes up- regulated, down-regulated and unchanged is shown for all promoters with CTCF sites (left). Percentages and numbers of genes with different expression patterns for each group of CTCF-binding sites are shown in the right panel. (E) Gene ontology terms enriched for promoters within the interval [-10,000, +10,000] from CTCF sites. The ensuing list of genes that were associated with each of our binding sites was divided into upregulated/downregulated/no-change based on the RNA-seq data. Subsequently each of these sub-lists was used with the functional annotation clustering / GO analysis tool in DAVID (https://david.ncifcrf.gov/). Genes are ordered by expression fold change. Red colour corresponds to up-regulation, green colour – down-regulation.

Figure 2. Correlation of CTCF binding within +/-10,000 bp from TSS with gene expression from corresponding genes, in control and treated cells. (A) Genes containing CTCF at their promoters in control cells sorted by expression in control cells. (B) Genes containing CTCF at their promoters in treated cells sorted by expression in treated cells. (C) Genes containing CTCF at their promoters in control (black) or in treated cells (red), sorted by expression fold change between treated and control cells. Each vertical bar corresponds to one gene.

Figure 3. Correlation of gene expression after cell treatment and CTCF occupancy change at the corresponding promoters (+/-1000 bp from TSS) for the common (A), lost (B) and gained (C) CTCF sites. Promoters with common and lost CTCF sites did not show statistically significant correlation of the change of gene expression with CTCF occupancy change upon cell treatment.

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4. Nucleotide frequencies as a function of distance from the summit of CTCF ChIP-seq peak for the subsets of common (A), lost (B) and gained sites (C). The consensus motifs for common, lost and gained CTCF sites are shown in the inserts in the corresponding panels. Most common CTCF sites contain classical CTCF recognition motif; part of lost CTCF site also contain the classical CTCF recognition motif; gained CTCF sites do not contain CTCF binding motif, suggesting nonspecific binding to open chromatin regions defined by other TFs. D) The strength of CTCF binding reflected by the heights of ChIP-seq peaks for different classes of CTCF sites.

Figure 5. Average profiles of CTCF occupancy and chromatin density at common, lost and gained CTCF sites. Black – CTCF in control cells, red – CTCF in treated cells, green – chromatin accessibility in control cells, blue – chromatin accessibility in treated cells.

Figure 6. Average profiles of H3K9me3 enrichment at common, lost and gained CTCF sites. Black – control cells, red – treated cells.

Figure 7. A schematic model illustrating the events observed in control and treated 226LDM cells in which transition from CTCF130 to CTCF180 takes place. Following treatment, cells change morphologically from adherent and flat to suspended and rounded. PARylated CTCF180 in treated cells is largely redistributed from the cell nucleus into cytoplasm (depicted on top of the Figure). More GC-rich stronger common sites retain CTCF180 (with smaller strengths). CTCF180 is evacuated from weaker ("lost") sites. "Gained" sites characterised by the absence of the CTCF motif acquire CTCF180 after treatment possibly due to interaction with additional proteins or directly. Chromatin density associated with the higher levels of the H3K9me3 is increased in the regions overlapping with CTCF sites in common and lost groups. Molecular changes within regions containing these CTCF sites result in alterations in gene expression patterns (illustrated with the examples of processes linked to these changes as in Figure 1E).

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

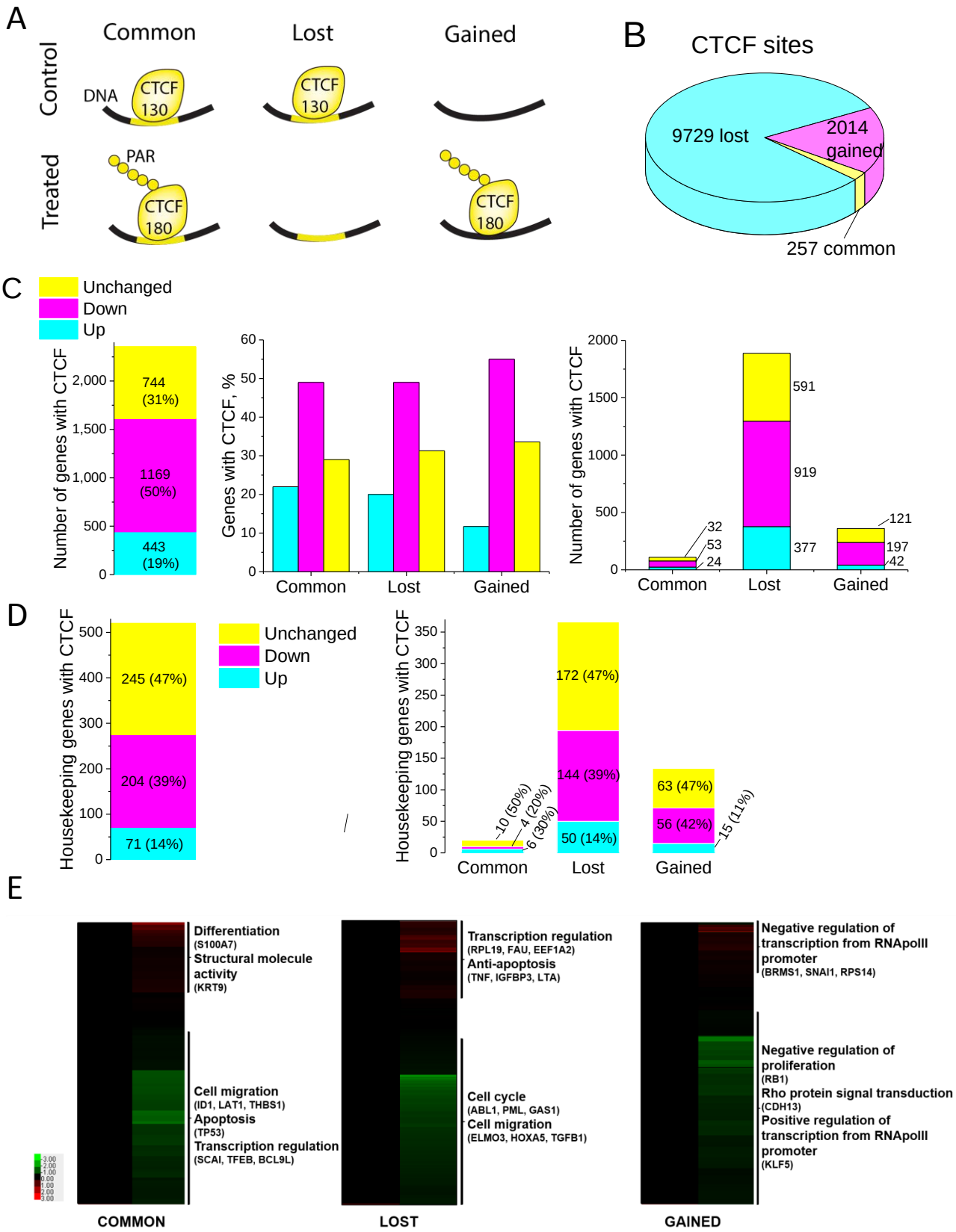

Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

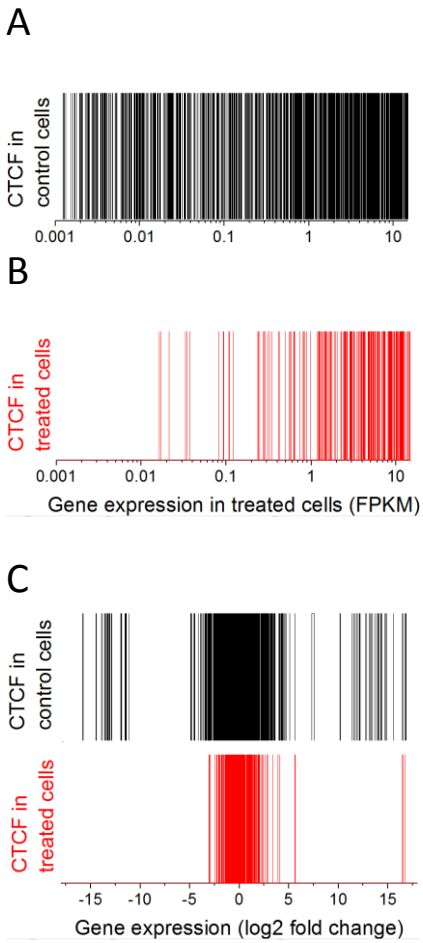

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

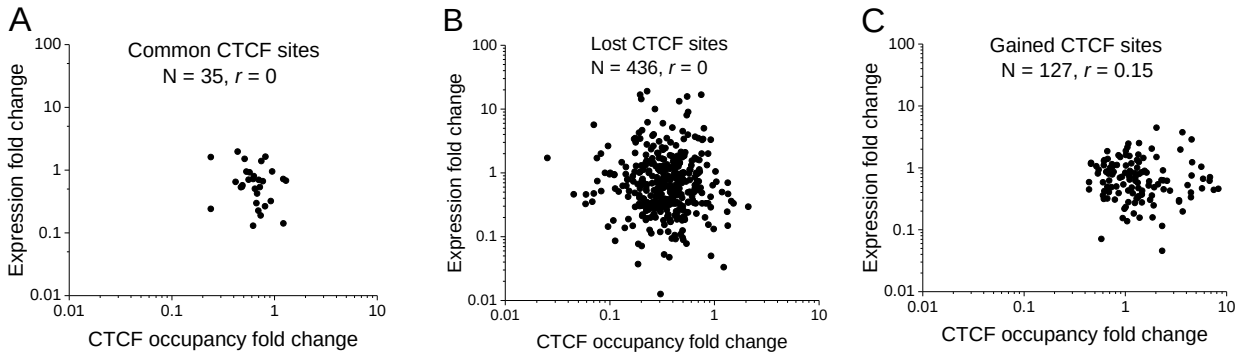

Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

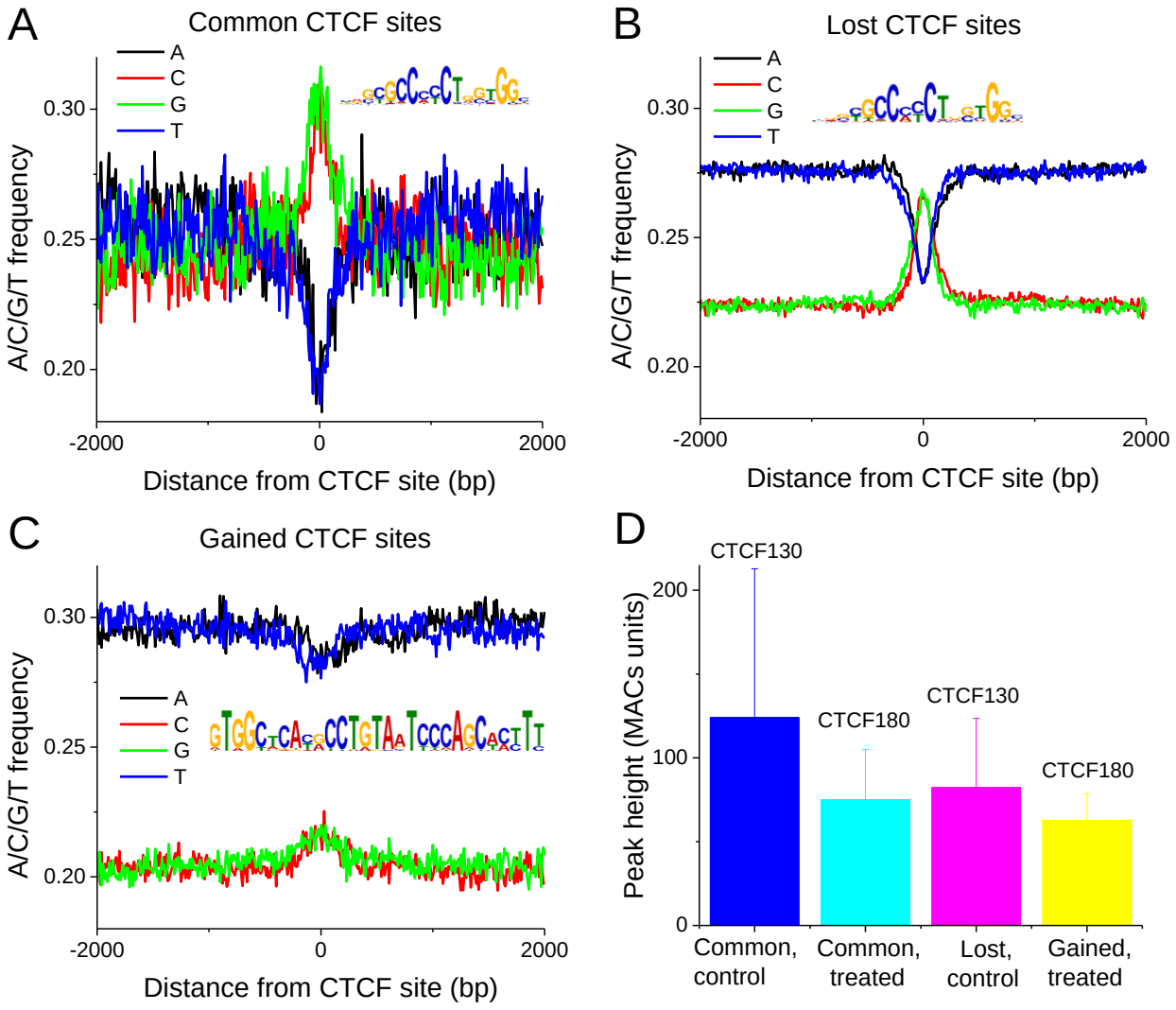

Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

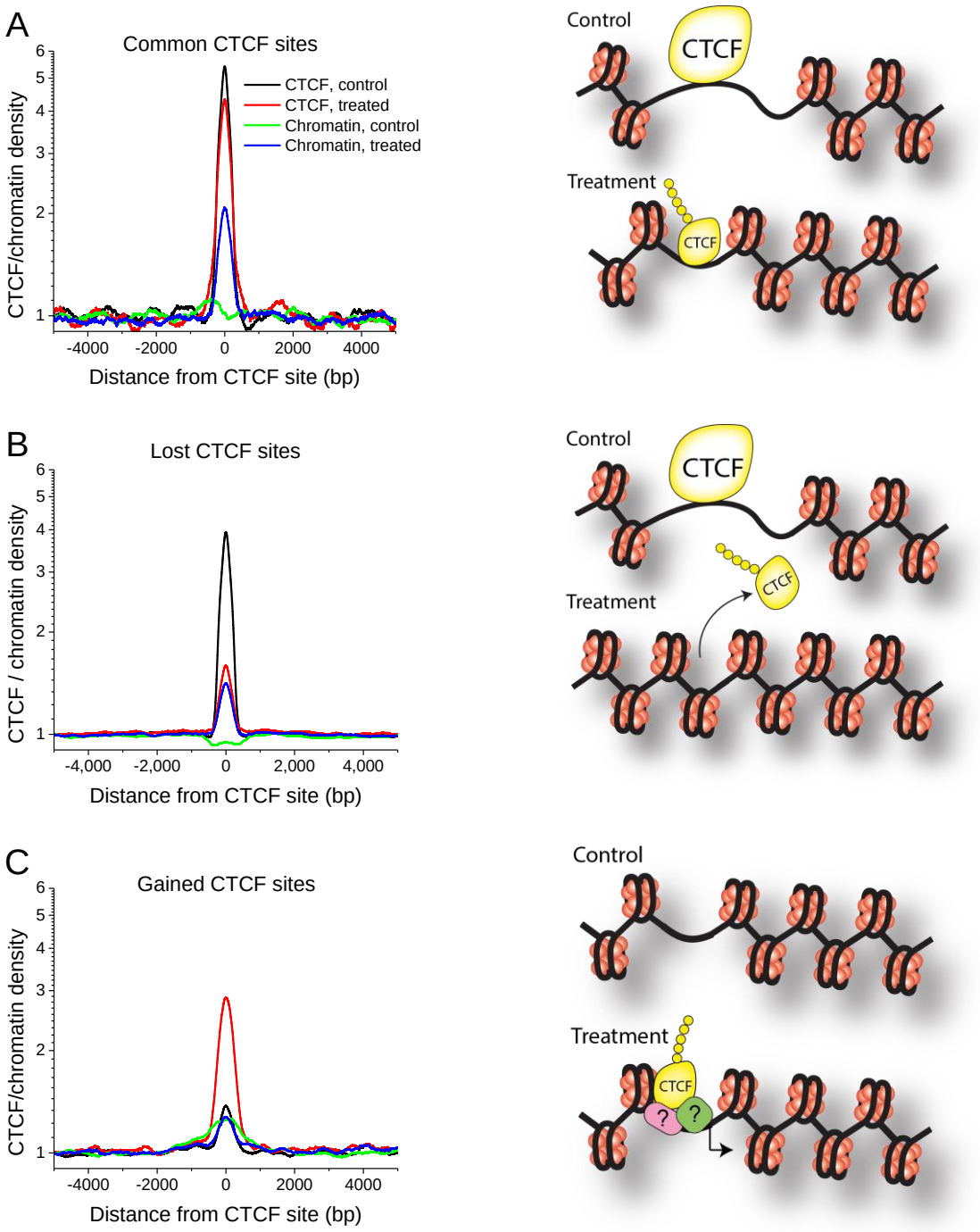

Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

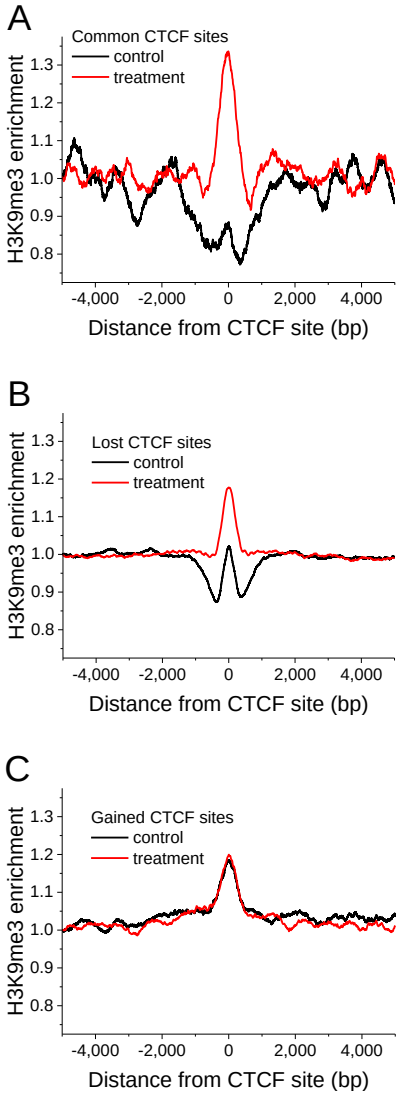

Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/175448; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

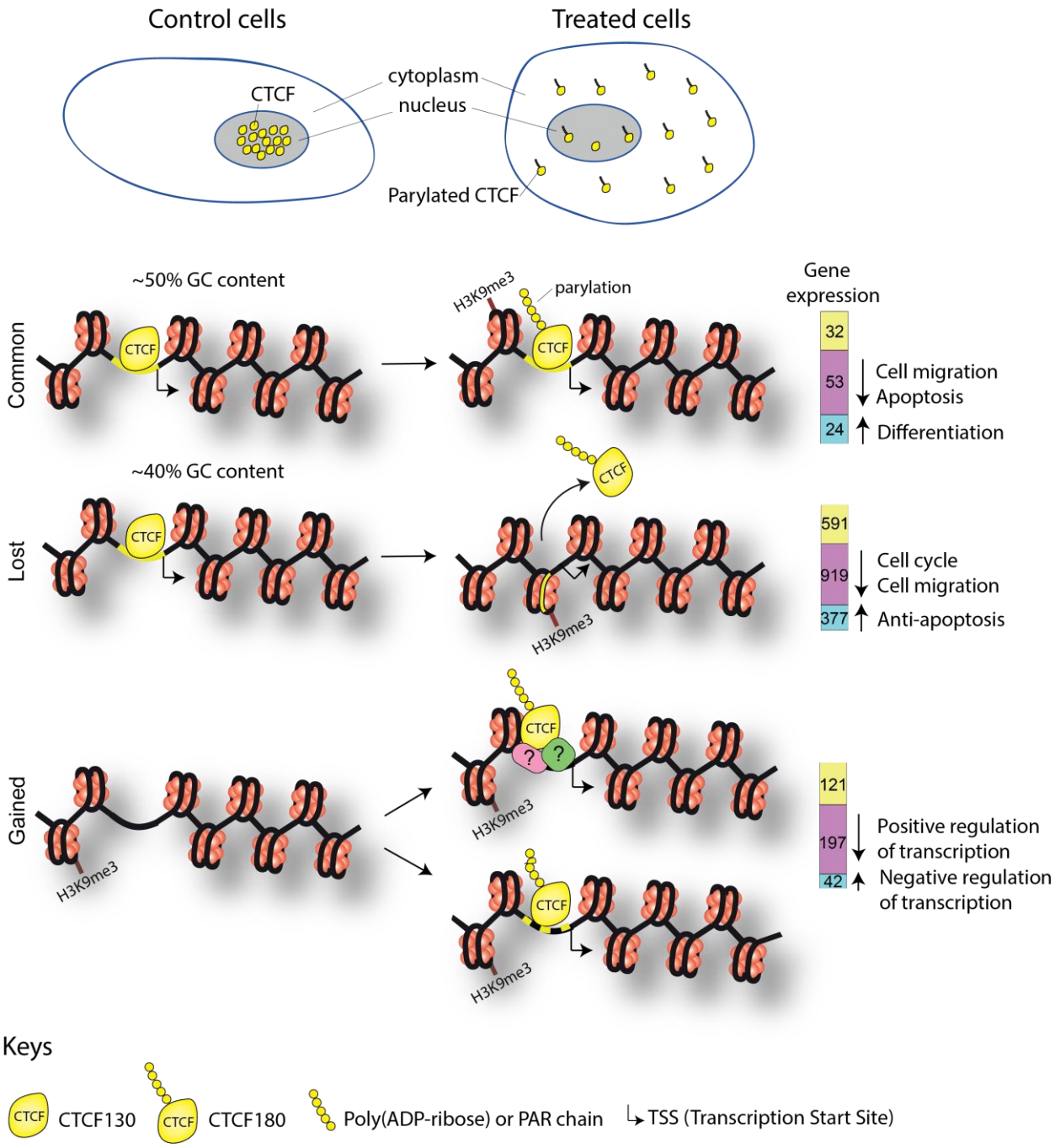

Figure 7