A variably occupied CTCF binding site in the *Ultrabithorax* gene in the *Drosophila* Bithorax Complex

Jose Paolo Magbanua[a], Estelle Runneburger[a], Steven Russell[b] and Robert White[a#]

Department of Physiology, Development and Neuroscience[a] and Department of Genetics[b], University of Cambridge, Cambridge, UK.

Running Head: Variably occupied CTCF site in *Ultrabithorax* gene

#Address correspondence to Rob White, rw108@cam.ac.uk.

Word Count Materials and Methods: 1,651

Word Count Introduction, Results, Discussion: 3,902

Abstract

Although the majority of genomic binding sites for the insulator protein CTCF are constitutively occupied, a subset show variably occupancy. Such variable sites provide an opportunity to assess context-specific CTCF functions in gene regulation. Here we have identified a variably occupied CTCF site in the *Drosophila Ultrabithorax (Ubx)* gene. This site is occupied in tissues where *Ubx* is active (third thoracic leg imaginal disc) but is not bound in tissues where the *Ubx* gene is repressed (first thoracic leg imaginal disc). Using chromatin conformation capture we show that this site preferentially interacts with the *Ubx* promoter region in the active state. The site lies close to *Ubx* enhancer elements and is also close to the locations of several *gypsy* transposon insertions that disrupt *Ubx* expression, leading to the *bx* mutant phenotype. *Gypsy* insertions carry the Su(Hw)-dependent *gypsy* insulator and were found to affect both CTCF binding at the variable site and the chromatin topology. This suggests that insertion of the *gypsy* insulator in this region interferes with CTCF function and supports a model for the normal function of the variable CTCF site as a chromatin loop facilitator, promoting interaction between *Ubx* enhancers and the *Ubx* transcription start site.

Introduction

There is considerable evidence indicating a major role for the multi-Zn finger protein CCCTC-binding factor (CTCF) in genome organisation (reviewed in 1, 2). CTCF binds to insulator elements and is required for their function in blocking interactions between enhancers and promoters (3). It has been shown to be involved in the formation of chromatin loops (4) and CTCF binding is enriched at the boundaries of topological chromatin domains (5–8). However, it is remains to be determined how much of CTCF function is linked to a specifically architectural role in genome organisation and how much is more directly involved in the control of gene expression.

CTCF was originally identified as a transcription factor (9). Subsequent genome-wide mapping of CTCF binding revealed that 20% of binding sites are within 2.5kb upstream of transcription start sites (10) and CTCF sites are enriched at gene promoters (11, 12). A current unifying hypothesis is that the molecular function of CTCF is to mediate chromosomal loop formation and this may give rise to a variety of context-dependent roles; in some contexts loop formation may serve an architectural purpose and in others it may be more intimately associated with gene regulation. One way to partition CTCF binding sites into possible functional classes is to differentiate between sites that are constantly occupied and sites that show variable occupancy. The first comparisons between whole genome maps of CTCF binding in different cell lines indicated that the majority of sites are constitutively bound (10, 13, 14). However more recent studies have revealed higher proportions of variable

sites (15, 16) and interestingly the variable sites are preferentially associated with enhancers (12). However, very few individual variable CTCF sites have yet been analysed and more examples are required to build an understanding of their association with gene regulation.

The classical example of a variable CTCF site is at the imprinted control region (ICR) of the mammalian insulin-like growth factor 2 (*Igf2)/H19* locus, where CTCF binding is regulated by DNA methylation of the binding sites. On the maternal chromosome CTCF binds the unmethylated ICR and the enhancer-blocking action of CTCF prevents *Igf2* expression. However, on the paternal chromosome, methylation of the ICR prevents CTCF binding and the lack of insulator function enables *Igf2* expression (17–20). A second example involves a CTCF site in the chicken lysozyme locus where CTCF binding is regulated by chromatin structure. Activation of the lysozyme gene is linked to eviction of CTCF and this is mediated through transcription of a noncoding RNA, chromosome remodeling and repositioning of a nucleosome over the CTCF binding site (21). Recently, in *Drosophila*, Wood et al provided evidence for two classes of regulated insulator (22). In one class, the occupancy of DNA-binding insulator proteins (e.g. BEAF, CTCF, Su(Hw)) at insulator sites is regulated. In a second class, the DNA-binding insulator proteins are constitutively bound, but the insulators are regulated by the variable recruitment of other components (e.g. CP190) required to build a functional insulator complex.

Here we present an analysis of a variably occupied CTCF site in the *Drosophila* Bithorax complex (BX-C). The BX-C contains three Hox genes *Ultrabithorax* (*Ubx*), *abdominal A* (*abd-A*) and *Abdominal B* (*Abd-B*) and has a clear regulatory domain structure with independent regulatory elements controlling gene expression in the parasegmental (PS) units along the anteroposterior axis of the developing embryo (reviewed in 23). The regulatory domains are separated by boundaries that constrain the activation of PS-specific enhancers. Genetic deletion of boundaries leads to inappropriate enhancer activation and ectopic expression of Hox genes. CTCF binding is associated with BX-C boundaries and CTCF mutations cause mis-expression of *Abd-B* (24–26). The CTCF binding at boundary elements appears to be constitutive and this may fit with an architectural role for these sites. Here we report the identification of a variable CTCF site within the *Ubx* gene that preferentially binds CTCF when the *Ubx* gene is active and is associated with a different chromatin topology in active and inactive states. We present a model where CTCF has a role facilitating the interaction between *Ubx* enhancers and the *Ubx* promoter.

Materials and Methods

Fly lines

The wild type *Drosophila melanogaster* strain Oregon R was used in the ChIP-Array, ChIP-qPCR and 3C experiments. In addition, homozygous bx^{83Ka} mutants (27) from the strain bx^{83Ka} / *TM6B* were used in ChIP-PCR and 3C experiments.

Antibodies

The following antibodies were used in the ChIP experiments: anti-CTCF-C antiserum (24), anti-CP190 antiserum (28), anti-RNA Pol II (affinity purified IgG 0.9 mg/ml, Abcam, ab5131) and anti-GAGA Factor (0.2 mg IgG/ml, Santa Cruz Biotechnology, SC-98263).

Chromatin preparation

Dissected head segments of late 3rd instar larvae were inverted and fixed with 2% formaldehyde in PBS for 20 min at room temperature. These were washed with twice with PBS/125 mM Glycine/0.01% Triton X-100 followed by a single wash with PBS and then with PBS containing 1% protease inhibitor cocktail (Sigma, P8340). The T1 and T3 leg imaginal discs were then dissected, snap frozen in liquid nitrogen and stored at -80°C prior to use. Approximately 150 leg discs were combined in PBS/0.01% Triton X-100 and centrifuged in a microfuge at 1200 rpm for 1 min. The discs were resuspended in 20 µl cell lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 0.5% NP-40) containing 1% protease inhibitor cocktail and homogenised using a motorised pestle at 2 min intervals for 8 min. After a brief microfuge centrifugation (13,200 rpm, 10 sec), the pellet was resuspended in 300 µl Nuclear Lysis Buffer (50 mM Tris.HCl pH 8.1, 10 mM EDTA.Na_2, 1% SDS) with protease inhibitors and incubated for 20 min at room temperature. The extracts were sonicated in a Bioruptor Standard (Diagenode) at high setting for 4 min 15 sec (30 sec “ON”, 30 sec “OFF” cycle), producing 0.5 to 3.0 kb

sized fragments. 100 µl aliquots of chromatin extracts were flash frozen in liquid nitrogen and then stored at -80°C prior to use.

Chromatin immunopurification

Chromatin immunopurification was performed as described by Birch-Machin (29). 100 µl aliquots of chromatin were pre-cleared with 13 µl blocked *S. aureus* cells (SAC) and mixed with 200 µl of IP dilution buffer (16.7mM Tris.HCl pH 8, 167mM NaCl, 1% EDTA, 1.1% Triton X-100, 0.01% SDS) with protease inhibitors. 2 µl of antibody was added and incubated on a roller overnight at 4°C. Then 13 µl of SAC was added to each IP reaction and the samples were incubated for 35 min at 4°C on a roller. The mixture was centrifuged in a microfuge at 13,200 rpm at room temperature and the pellets were washed successively with 1 ml each of Low Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA. Na_2 pH 8.0, 20 mM Tris.HCl pH 8, 150 mM NaCl), High Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA.Na_2 pH8, 20 mM Tris.HCl pH 8, 500mM NaCl), LiCl Buffer (0.25 M LiCl, 1% NP40, 1% NaDeoxycholate, 1 mM EDTA. Na_2 pH 8.0, 10 mM Tris.HCl pH 8.0) and twice with TE buffer pH 8.0, for 5 min at 4°C on roller for each solution. The immune-precipitated chromatin was then eluted twice from the SAC pellet with 300 µl of IP elution buffer (50 mM $NaHCO_3$, 1% SDS) by vigorously vortexing for 15 min at room temperature. One µl of RNase A (Sigma, R4642) and 24.3 µl of 4M NaCl (0.3 M final concentration) were then added to the eluate and the mixture was incubated for 4h at 65°C, to reverse the cross-linking. The DNA was then precipitated by adding 812 µl of 100% ethanol and incubating overnight at -20°C. The samples were centrifuged in a microfuge at 4°C for 20

min and the pellets were air dried for 1 h at room temperature. The pellets were resuspended in 100 µl TE buffer followed by the addition of 25 µl of 5X PK buffer (50 mM Tris HCl pH 7.5, 25 mM EDTA.Na_2 pH 8, 1.25% SDS) and 1.5 µl of 20 mg/mL Proteinase K, incubated at 45°C for 2 h and purified using the QIAquick PCR Purification Kit (Qiagen, 28104). The DNA was eluted in 30 µl of buffer EB and stored at -20°C until use.

CTCF-ChIP Array

5 µl each of CTCF-ChIP and control ChIP DNA from T1 and T3 leg discs obtained from Oregon R larvae were amplified using GenomePlex Single Cell Whole Genome Amplification Kit (Sigma-Aldrich WGA4) according to manufacturer's instructions. The samples were amplified for 21 cycles and the amplified DNA purified using the QIAquick PCR Purification Kit. 1 µg each of amplified ChIP and control DNA was labelled with Cy5 and Cy3 in the presence of Cy3- or Cy5-dCTP (GE Healthcare) using the BioPrime DNA Labeling Kit (Invitrogen) and hybridised onto Nimblegen ChIP-chip 2.1M Whole-Genome Tiling Arrays according to the manufacturer's instructions.

Microarray data processing

Two biological replicates were prepared for each sample with a Cy3/Cy5 dye swap for one biological replicate of each sample. ChIP DNA prepared with pre-immune serum was used as the reference control to assay ChIP enrichment in the array experiments. Arrays were scanned and processed as previously described (30). The enrichment profiles were visualised using the Integrated Genome Browser (http://bioviz.org/igb/index.html). Patser position-

specific weight matrix analysis was as described (24). The ChIP-array data have been submitted to GEO under accession number GSE62234. Analysis of conservation used the PhastCons multiple alignment data available from http://genome.ucsc.edu.

Quantitative PCR

Quantitative real-time PCR experiments were performed with LightCycler 480 II (Roche Diagnostics) in 10 µl reactions using SYBR Green PCR Master Mix (Roche, Cat. 04707516001). Each reaction consisted of 5 µl SYBR Green PCR Master mix, 3 µl water, 1 µl 10 µM primer mix and 1 µl DNA. Amplifications was carried using the following conditions: 1 cycle at 95°C, 15 min; 45 cycles of 95°C, 10 sec; 58°C, 10 sec; and 72°C 10 sec. The primer pairs used for the amplification are listed in Table 1. Serial dilutions of *Drosophila* genomic DNA (100 – 0.01 ng/µl) were used as standards for quantification.

Preparation of 3C DNA from T1 and T3 leg discs

Approximately 450 each of T1 and T3 leg discs from 3rd instar larvae were dissected and frozen as described above. The discs were thawed on ice and transferred to a 1.5 ml microcentrifuge tube. The pooled discs were briefly centrifuged at 13,200 rpm for 10 sec. The excess liquid was discarded and the discs were resuspended in 20 µl lysis buffer (31) containing 10 mM Tris-Cl pH 8.0, 10 mM NaCl, 0.2% Igepal CA360 (Sigma, I8896) and 10 µl/ml of protease inhibitor (Sigma). The discs were homogenised using a plastic motorised pestle at 2 min intervals for a total of 8 min. After a brief

centrifugation, 500 µl lysis buffer with 50 µl of protease inhibitor was added to the homogenate and the suspension was centrifuged at 5,000 rpm for 5 min at room temperature.

The 3C DNA was prepared based on the protocol described by Hagege et al. (32). The leg disc lysate pellet was washed twice with ice-cold 1.2x NEBuffer 3 (New England Biolabs, B7003S) at 5,000 rpm for 5 min at room temperature. The pellet was then resuspended in 500 µl 1.2x NEBuffer 3 and 7.5 µl 20% SDS. The mixture was incubated at 37°C, 900 rpm for 1 h in a Thermomixer (Eppendorf, Cat. 5355000038). Then 50 µl 20% Triton X-100 was added and the mixture further incubated at 37°C, 900 rpm for 1 h. The lysate was then digested with 400U of *DpnII*, at 37°C, 900 rpm overnight. The enzyme was inactivated by heat treatment at 65 °C for 20 min and the mixture was ligated at 16°C for 16 hours in a 10 ml reaction with 10,000U of T4 DNA ligase (New England Biolabs). The ligated chromatin digest was then de-crosslinked and purified as described by Hagege et al. (32). The purified 3C DNA was resuspended in 50µl TLE Buffer (10 mM Tris-Cl, pH 8.0, 0.1 mM EDTA) and DNA concentration was measured by using the Qubit dsDNA HS Assay kit (Invitrogen, Q32854). 3C DNA samples were stored at -20° until use.

PCR amplification of 3C DNA

3C interactions were determined according to the protocol by Dekker et al. (33). To investigate the chromatin conformation and interactions in the *Ubx* region in T1 and T3 leg discs, 29 primers spanning Chr 3R:12400341..12695484 were designed based on the expected fragments

generated by *Dpn*II digestion (Table 2). In addition, primer pairs located in *Dpn*II fragments containing the CTCF differential peak in *Ubx*, the *Ubx* promoter and the *Mcp* region were also designed to serve as anchor fragment internal primers (Table 2).

For each anchor fragment investigated, individual 10 µM primer mixes composed of the anchor fragment internal primers and individual anchor primer / target primer pairs were prepared. The 3C PCR reactions were carried out in a 25µl mixture using Thermo-Start Taq DNA Polymerase Kit (Thermo Scientific, AB-1057). Each reaction contained 18.3 µl water, 2.5 µl 10X PCR Buffer, 1.5 µl 25 mM $MgCl_2$, 0.5 µl 10 mM dNTP mix, 0.2 µl Taq DNA polymerase, 1 µl 10 µM primer mix and 1 µl (1 ng/µl) of 3 C DNA sample. Amplification was carried out in an iCycler 582BR Thermal Cycler (BioRad) using a touchdown protocol with 1 cycle at 95°C for 15 min and then 10 cycles at 95°C for 30 sec; annealing from 69 to 59°C for 30 sec and 72°C for 30 sec. This was followed by 30 cycles at 95°C, 30 sec; 59°C, 30 sec and 72°C, 30 sec followed by a final extension at 72°C for 10 min. PCR products were then subjected to electrophoresis on a 2% agarose gel in 0.5X TBE.

Quantification of 3C PCR products

Gel images were digitised and the bands were quantified using ImageJ software (http://imagej.nih.gov/ij). The relative interaction between the different primer pairs was then expressed as the ratio of the signal strength between the anchor/target 3C PCR product and the anchor fragment PCR product. Relative interaction between the 3C primer pairs and each specific anchor fragment was plotted to visualise interactions.

Results

Identification of a variably occupied CTCF site in the *Ubx* gene

The individual Hox genes of the BX-C are expressed in different segments along the anteroposterior axis (23), presenting a useful experimental system for the isolation of *in vivo* tissues with different states of gene expression in sufficient quantities for genomic analysis. Here we have used the imaginal discs from *Drosophila* larvae to compare the genome-wide CTCF binding profile in leg imaginal discs from the 1st thoracic segment (T1) with leg discs from the 3rd thoracic segment (T3). The Hox gene *Ubx* is not expressed in T1 but is active in T3. The other two genes of the BX-C, *abd-A* and *Abd-B*, are inactive in both T1 and T3. The activity state of these BX-C genes is regulated by Polycomb (Pc) silencing which imposes a repressive chromatin state on inactive genes. Comparing the T1-leg disc with T3-leg disc CTCF ChIP-array profiles, we find the profiles are generally extremely similar with very few clear differential peaks found, however we identified a clear differential CTCF binding peak in the *Ubx* gene (Figure 1A). There is strong CTCF binding at this position in the T3-leg disc where *Ubx* is expressed but we find little binding at this site in the T1-leg disc where the *Ubx* gene is repressed. In contrast, the binding of CTCF in the repressed *abd-A* and *Abd-B* regions is very similar in both discs.

The variably occupied CTCF site lies in an intron within the *Ubx* transcription unit. Motif analysis with the CTCF position-weight-matrix revealed a strong sequence match at this position (Figure 1B). It has been proposed that CTCF

sites serving different functions may be identifiable at the sequence level and subfamilies of CTCF binding sites have been identified. We examined the variable site for sequence features that might place it in a defined subfamily. In general the variable site has features associated with high occupancy having, in addition to the strong match to the core motif (Patser score =12.3), the conserved T of Module #1 described by Rhee and Pugh and the CC motif (Figure 1C) that are both associated with higher levels of CTCF binding (34, 35). The variable site is on the edge of a sequence block highly conserved across 15 insect genomes (Figure 1C) and CTCF binding at this site is clearly identified in pupal-stage chromatin from four *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. yakuba* and *D. pseudoobscura*) covering a range of evolutionary divergence of up to 25 million years (36).

We validated the differential CTCF binding at this site using quantitative PCR with a set of primer pairs spanning the CTCF peak (Figure 1B and 1D). We see clearly enriched CTCF binding in T3 versus T1 leg disc chromatin specifically at this CTCF site.

Protein complex formation at the variable CTCF site

To investigate whether the DNA binding protein CTCF is involved in building a protein complex together with other insulator proteins or transcription factors at this site, we analysed the binding of other protein components (Figure 2). Centrosomal Protein 190 (CP190) does not bind DNA directly but associates with CTCF (and other DNA-binding insulator components such as Su(Hw)) through a BTB domain interaction and is required for the enhancer-blocking

function of insulator complexes (25, 37, 38) and for looping interactions of CTCF insulators (22). We find no evidence for CP190 association with the variable CTCF site in T1 leg-disc chromatin, but CP190 is significantly associated with this site in T3 leg disc chromatin. This suggests that differential binding of CTCF in T3 enables the formation of a protein complex involving proteins associated with insulator function.

GAGA-Factor (GAF) appears to participate in a diverse range of transcriptional processes and is required for the activity of some insulators (39–41). GAF does not bind at the variable CTCF site but there is substantial binding in the region of the primer pair "1" that lies about 1 kb away from the CTCF site (Figure 2). This strong GAF binding is similar in both T1 and T3 leg imaginal disc chromatin. We also examined the binding of the insulator components Su(Hw), mod(mdg4 isoform N) and BEAF32 but found no evidence for binding in the region of the variable CTCF site in leg discs (data not shown).

Intronic CTCF sites have been implicated in splicing regulation and PolII pausing (42). We examined the binding profile of PolII across the region spanning the variable CTCF site and at the *Ubx* promoter using an antibody that recognises the Ser5 phosphorylated PolII (Figure 2). PolII-Ser5P is found preferentially bound across the region in T3 versus T1 discs which fits with the specific *Ubx* expression in T3, however there is no pronounced peak at the CTCF site and thus we see no evidence of PolII pausing at this site. At the promoter, PolII-Ser5P shows strong binding in T3 and no binding in T1

indicating the engagement of PolII with the active promoter and a lack of paused PolII when the *Ubx* promoter is inactive.

Chromatin topology in the active and inactive states

We next investigated whether the variable CTCF-dependent protein complex that assembles on the active *Ubx* gene is associated with alteration in chromosomal topology between the inactive and active states of *Ubx* transcription. We used Chromosome Conformation Capture (3C; 33) to analyse interactions from the viewpoint of the variable CTCF site as an anchor fragment and 28 nearby target sites including the *Ubx* promoter, the *abd-A* promoter and CTCF sites across the *Ubx* and *abd-A* regions. The overall interaction profiles are shown in Figure 3A and the interaction scores for selected primers closest to particular features, e.g. the *Ubx* promoter and the *abd-A* promoter, are detailed in Figure 3B. We find that the variable CTCF site shows a marked preferential interaction with the *Ubx* promoter in the *Ubx* active (T3) state (*Ubx* 5' primers in Figure 3B Anchor 1). In contrast, the interaction of the variable CTCF site with the repressed *abd-A* promoter shows the reverse preference; in T3 there is no interaction but in the *Ubx* inactive state (T1) the variable CTCF site is associated with the repressed *abd-A* promoter (abd-A 5' primers in Figure 3B Anchor 1).

As using the variable CTCF site as the 3C anchor indicated a specific preferential interaction with the *Ubx* promoter in the active state, we next examined interaction from the viewpoint of a 3C anchor at the *Ubx* promoter. This confirmed the preferential interaction between the variable CTCF site

and the *Ubx* promoter in the active (T3) state (CTCF site primers in Figure 3B Anchor 2). In contrast, in T1 the repressed *Ubx* promoter shows evidence of a preferential interaction with the repressed *abd-A* promoter.

We also examined a third viewpoint using a 3C anchor at the *Mcp* boundary element, which contains a CTCF binding site and is in the repressed *abd-A* domain in both T1 and T3. The *Mcp* anchor shows a peak of interaction with the *abd-A* promoter in both T1 and T3 but shows a preferential interaction with the *Ubx* promoter and the variable CTCF site in the inactive (T1) state (Figure 3B Anchor 3). Since there is little CTCF associated with the variable site in the inactive state, these interactions may involve the nearby Polycomb Response Element (*bx*-PRE; Figure 4).

Overall, the 3C analysis indicates that the *Ubx* region adopts a different chromatin topology in the active versus inactive state. The active (T3) state is characterised by increased interaction between the variable CTCF site and the *Ubx* promoter and decreased association of both the variable CTCF site and the *Ubx* promoter with repressed regions, specifically the *abd-A* promoter and the *Mcp* boundary element.

Chromatin topology in the *bx^{83Ka}* mutation

The variable CTCF site lies close to the *bx*-PRE (43), the BRE embryonic enhancers (44) and the *abx* enhancers (45)) which are active in both the embryo and in imaginal discs (Figure 4). This arrangement, together with the interaction between the variable CTCF site and the *Ubx* promoter suggests a

model where the variable CTCF site may play a role in facilitating interaction between the *abx/bx* enhancers and the *Ubx* promoter. Deletion of a 9.5kb region that includes the variable CTCF site gives a *bx* phenotype ($bx^{34e\text{-}prv}$; 27) caused by decreased *Ubx* expression in T3 discs and it is intriguing that the variable CTCF site lies in the heart of the region defined by the cluster of *bx* mutations. There is a strong connection between *bx* mutations and insulator function since, of the ten *bx* mutations, seven are caused by the insertion of *gypsy* transposable elements (27, 46) which carry a cluster of binding sites for the Su(Hw) insulator protein, the most studied insulator in *Drosophila* (47). These *gypsy*-induced *bx* alleles are all suppressed in a *su(Hw)* mutant background (27, 46), indicating that it is not simply the presence of the 7.5kb *gypsy* element but rather the binding of the Su(Hw) insulator protein that causes the *bx* mutant phenotype. This suggests that this region is topologically sensitive and that the *gypsy* insertions may interfere with interactions between the *abx/bx* enhancers and the *Ubx* promoter. Specifically in terms of the above model for the function of the variable CTCF site, insertion of a second topological regulator, Su(Hw), in this region may interfere with the interaction between the CTCF-variable site and the *Ubx* promoter.

To test this hypothesis we examined the effect of a *bx* mutation on chromatin topology carrying out 3C analysis on homozygous bx^{83Ka} T1 and T3 leg discs. The phenotype of *bx* mutations is a loss of *Ubx* expression in the anterior compartment of the T3 imaginal discs, haltere and T3 leg (Figure 4B and C; 48). In the anterior compartment, *Ubx* expression may depend on interactions

between the promoter and the downstream enhancers, *abx* and *bx*, whereas in the posterior compartment the *Ubx* promoter may contact the upstream *pbx* region. This fits with the presence of both upstream and downstream preferential interactions with the *Ubx* promoter in the active state that we observed in the 3C analysis (Figure 3; small arrows). The *bx* mutations might be expected to specifically interfere with the downstream interaction.

In the 3C analysis, we find that the mutation has several effects on chromatin topology in the *Ubx* region (Figure 5). First, contrary to the expectations of the model, the *gypsy* insertion enhances interaction between the variable CTCF site and the *Ubx* promoter. This enhancement is seen in both T1 and T3, although the interaction remains stronger in T3 (Figure 5B, *Ubx*5' primers Anchor 1 and CTCF site primers Anchor 2). Second, fitting the predictions of the model, the preferential interaction seen in the active state (T3) between the downstream *abx* enhancer region and the variable CTCF site is lost in the mutant (*abx* primer in Figure 5B Anchor 1). Similarly, for the interaction between the *abx* enhancer region and the *Ubx* promoter (*abx* primer in Figure 5B Anchor 2) there is evidence for stronger interaction in T3 versus T1 in the wild type and this differential is lost in the mutant. Also, fitting the model, in contrast to the *abx* region, the *pbx* region preferentially interacts with the *Ubx* promoter in the active state (T3) in the bx^{83Ka} mutant (*pbx* primer in Figure 5B Anchor 2).

Overall, although some predictions of the model are borne out, it appears that the effects of the *gypsy* insertion are more complex than simply blocking interactions between the variable CTCF site and the *Ubx* promoter

The *bx*^{83Ka} insertion affects protein binding in flanking regions

To investigate this further, we examined protein binding in the region of the variable CTCF site in homozygous *bx*^{83Ka} T1 and T3 leg discs (Figure 6). Strikingly we find that, in the mutant, CTCF is strongly associated with the site, not only in T3, but also in T1. In addition, we find that the *gypsy* insertion in the *bx*^{83Ka} mutation also strongly affects GAF binding; compared to the wild type it is markedly reduced in both T1 and T3. PolII binding shows, as expected, clear occupancy in the T3 discs, where *Ubx* is expressed in posterior compartment cells.

Overall, perhaps the most striking effect of the *bx*^{83Ka} insertion is the increase in CTCF binding at the variable CTCF site, particularly in T1. This indicates that the *gypsy* insulator can affect the loading of insulator proteins onto a nearby site and this fits with an increased association between the variable insulator site and the *Ubx* promoter. It is possible that this interaction may exclude the *abx* regulatory region since the preferential contact between the *abx* regulatory region and the variable CTCF site seen in the active state in the wild type is lost in the mutant.

Discussion

We have identified a variably occupied CTCF binding site in the *Ubx* gene in the *Drosophila* BX-C. This site lies close to characterised *Ubx* regulatory elements and we find that CTCF occupancy is associated with a specific interaction between the variable site and the *Ubx* promoter in the transcriptionally active state. These observations suggest a model that CTCF binding at this site facilitates interaction between the regulatory elements and the *Ubx* promoter.

This model is supported by our studies on the bx^{83Ka} mutation where the insertion of a *gypsy* insulator close to the variable CTCF site disrupts the chromatin topology. One explanation for the effect of the *gypsy* insertion on *Ubx* expression is that the *gypsy* insulator acts as an enhancer-blocker, preventing interactions between the *Ubx* promoter and regulatory elements (e.g. *abx*) lying beyond the insulator insertion site (49). However a simple enhancer blocking model does not fit with the enhanced interaction we see between the variable CTCF site and the *Ubx* promoter in the bx^{83Ka} mutant, nor does it explain the tight clustering of *gypsy* insertions with a *bx* phenotype within a specific 11kb region centred on the variable CTCF site. Our analysis shows that the bx^{83Ka} insertion does not simply introduce an insulator but also has effects on flanking regions. In particular, the bx^{83Ka} insertion affects the binding of CTCF at the variable CTCF site leading to clearly enhanced CTCF occupancy in both T1 and T3 discs. In the case of bx^{83Ka} the *gypsy* insertion also lies close to a GAF ChIP binding peak and results in loss of GAF binding in both T1 and T3 discs. This effect on GAF binding is difficult to interpret functionally; GAF has a role in *Ubx* expression as the GAF gene *Trl* interacts

with *Ubx* alleles (50), however *Trl* mutant clones in imaginal discs do not appear to affect *Ubx* expression (51, 52). The topological changes associated with the bx^{83Ka} insertion include enhanced interactions between the variable CTCF site and the *Ubx* promoter in both T1 and T3, and loss of the preferential interaction between the variable CTCF site and the distant *abx* regulatory region in T3. This suggests that the insertion of a *gypsy* insulator may stabilise CTCF binding and promote interactions with the *Ubx* promoter but in a manner that excludes interactions with distant regulatory elements. Hence the *gypsy* Su(Hw) insulator element may indeed act as an enhancer blocker, but it may do so in collaboration with a CTCF complex. We speculate that the involvement of CTCF in the mechanism that generates the mutant phenotype explains the observed clustering of *gypsy* insertions with *bx* phenotypes around the variable CTCF site.

Although our observations indicate a likely role for CTCF in facilitating enhancer-promoter interaction in *Ubx* regulation, functional studies will be required to confirm the role of CTCF and its importance for *Ubx* expression. In this regard we have looked for genetic interaction between *CTCF* and *Ubx.* As null *CTCF* mutants are lethal, we investigated whether the *Ubx* haplo-insufficent phenotype is enhanced by heterozygosity for *CTCF.* We have not seen clear enhancement in this situation and further work will be required to test the proposed CTCF role.

Why are some CTCF binding sites constitutive and others variably occupied? The occupancy of CTCF sites across the BX-C sheds light on this issue but

initially presents a puzzle. CTCF sites within the *abd-A* and *Abd-B* domains are occupied even when these domains are silenced by Pc-mediated repression, whereas the variable CTCF site in the *Ubx* gene is only occupied when the *Ubx* domain is de-repressed. This raises questions about the ability of CTCF to access its binding site in different chromatin states. There is evidence that CTCF binding is sensitive to chromatin configuration. In particular CTCF binding is affected by nucleosome positioning and CTCF is unable to bind if its target site is covered by a nucleosome (21, 53). Examination of chromatin accessibility within the repressed *abd-A* and *Abd-B* domains by DNase1 sensitivity, reveals that CTCF sites generally correspond to small regions of DNase1 accessibility within the repressed domains (Figure 7A), indicating that CTCF is bound at sites of open, potentially nucleosome-free, chromatin. Interestingly, these sites are bound by other factors, for example Yki and GAF, so it is unclear which factor or factors are responsible for initiating and establishing open chromatin at these positions. Importantly, the presence of other factors indicates that CTCF is not necessarily responsible for pioneering binding at these sites in repressed chromatin. The variable CTCF site in *Ubx* supports the idea that CTCF on its own may not be able to bind to repressed chromatin and it is intriguing that in this particular case the adjacent DNase1 site, occupied by Yki, GAF and Pho, does not extend over the CTCF site (Figure 7B). Occupancy of the variable site may be dependent on Pc-derepression of the *Ubx* domain enabling nucleosome remodeling to expose the CTCF site for binding. A different perspective is given by the finding that, although CTCF does not bind to the variable site in the repressed *Ubx* domain in T1 in the wild type, it does bind in the context of

the *bx*^{83Ka} mutant. The insertion of the *gypsy* transposon carrying the Su(Hw)-dependent *gypsy* insulator may stabilise CTCF binding at the variable binding site perhaps through a general function of insulator complexes to facilitate loading of insulator components at nearby sites. Overall, our studies point to a view of CTCF binding where CTCF is in competition with nucleosomes for site occupancy. In the repressed state in T1, the nucleosome is dominant and there is very little CTCF binding to the variable site. CTCF binding may be enhanced either by decreasing nucleosome occupancy, associated with the opening of the *Ubx* domain in T3, or by local interactions between insulator complexes stabilising CTCF binding.

Our data also provide a view of the *in vivo* 3D organisation of the BX-C comparing the situation in T1, where all three BX-C genes are inactive, with T3 where *Ubx* is active and *abd-A* and *Abd-B* are inactive. In the active *Ubx* state both the variable CTCF site and the *Ubx* promoter engage in long-range interactions over a range of about 100kb, but the interactions we see are nevertheless confined to the *Ubx* domain. In the repressed state, the variable CTCF site and the *Ubx* promoter show more association with distant repressed regions outside the *Ubx* domain (Figure 3). This fits with previous studies both in *Drosophila* (54, 55) and in the mammalian Hox complexes (56–60) which support the idea of regulatory domains as dynamic topological structures where repressed domains cluster together and where expressed domains are segregated into a separate compartment.

Acknowledgments

The work was supported by the Wellcome Trust [089834/Z/09/Z] and ER was supported by the Erasmus Programme.

We thank Bettina Fischer in the FlyChIP-genomic facility for support and Sarah Bray for comments on the manuscript.

References

1. **Phillips-Cremins JE, Corces VG**. 2013. Chromatin insulators: linking genome organization to cellular function. Mol. Cell **50**:461–474.
2. **Phillips JE, Corces VG**. 2009. CTCF: master weaver of the genome. Cell **137**:1194–1211.
3. **Bell AC, West AG, Felsenfeld G**. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell **98**:387–396.
4. **Splinter E, Heath H, Kooren J, Palstra R-J, Klous P, Grosveld F, Galjart N, De Laat W**. 2006. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. **20**:2349–2354.
5. **Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B**. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**:376–380.
6. **Hou C, Li L, Qin ZS, Corces VG**. 2012. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell **48**:471–484.
7. **Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N,**

592 **Dekker J, Heard E**. 2012. Spatial partitioning of the regulatory landscape of
593 the X-inactivation centre. Nature **485**:381–385.
594 8. **Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M,**
595 **Parrinello H, Tanay A, Cavalli G**. 2012. Three-dimensional folding and
596 functional organization principles of the Drosophila genome. Cell **148**:458–
597 472.
598 9. **Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM,**
599 **Polotskaja AV, Goodwin GH**. 1990. A novel sequence-specific DNA binding
600 protein which interacts with three regularly spaced direct repeats of the
601 CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene.
602 Oncogene **5**:1743–1753.
603 10. **Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD,**
604 **Zhang MQ, Lobanenkov VV, Ren B**. 2007. Analysis of the vertebrate
605 insulator protein CTCF-binding sites in the human genome. Cell **128**:1231–
606 1245.
607 11. **Nègre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG,**
608 **Feng X, Ahmad K, Russell S, White RAH, Stein L, Henikoff S, Kellis M,**
609 **White KP**. 2010. A Comprehensive Map of Insulator Elements for the
610 Drosophila Genome. PLoS Genet **6**:e1000814.
611 12. **Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J,**
612 **Lee L, Lobanenkov VV, Ren B**. 2012. A map of the cis-regulatory sequences
613 in the mouse genome. Nature **488**:116–120.
614 13. **Cuddapah S, Jothi R, Schones DE, Roh T-Y, Cui K, Zhao K**. 2009. Global
615 analysis of the insulator binding protein CTCF in chromatin barrier regions

616 reveals demarcation of active and repressive domains. Genome Res. **19**:24–
617 32.
618 14. **Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF,**
619 **Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE,**
620 **Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA,**
621 **Crawford GE, Kellis M, Ren B**. 2009. Histone modifications at human
622 enhancers reflect global cell-type-specific gene expression. Nature
623 **459**:108–112.
624 15. **Chen H, Tian Y, Shu W, Bo X, Wang S**. 2012. Comprehensive identification
625 and annotation of cell type-specific and ubiquitous CTCF-binding sites in the
626 human genome. PloS One **7**:e41374.
627 16. **Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield**
628 **T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Myers RM,**
629 **Stamatoyannopoulos JA**. 2012. Widespread plasticity in CTCF occupancy
630 linked to DNA methylation. Genome Res. **22**:1680–1688.
631 17. **Bell AC, Felsenfeld G**. 2000. Methylation of a CTCF-dependent boundary
632 controls imprinted expression of the Igf2 gene. Nature **405**:482–485.
633 18. **Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM**.
634 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at
635 the H19/Igf2 locus. Nature **405**:486–489.
636 19. **Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi C-F, Wolffe A, Ohlsson**
637 **R, Lobanenkov VV**. 2000. Functional association of CTCF with the insulator
638 upstream of the H19 gene is parent of origin-specific and methylation-
639 sensitive. Curr. Biol. **10**:853–856.

640 20. **Szabó PE, Tang S-HE, Rentsendorj A, Pfeifer GP, Mann JR**. 2000.
641 Maternal-specific footprints at putative CTCF sites in the H19 imprinting
642 control region give evidence for insulator function. Curr. Biol. **10**:607–610.
643 21. **Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C**. 2008. The LPS-
644 induced transcriptional upregulation of the chicken lysozyme locus involves
645 CTCF eviction and noncoding RNA transcription. Mol. Cell **32**:129–139.
646 22. **Wood AM, Van Bortle K, Ramos E, Takenaka N, Rohrbaugh M, Jones BC,**
647 **Jones KC, Corces VG**. 2011. Regulation of chromatin organization and
648 inducible gene expression by a Drosophila insulator. Mol. Cell **44**:29–38.
649 23. **Maeda RK, Karch F**. 2006. The ABC of the BX-C: the bithorax complex
650 explained. Dev. Camb. Engl. **133**:1413–1422.
651 24. **Holohan EE, Kwong C, Adryan B, Bartkuhn M, Herold M, Renkawitz R,**
652 **Russell S, White R**. 2007. CTCF genomic binding sites in Drosophila and the
653 organisation of the bithorax complex. PLoS Genet. **3**:e112.
654 25. **Mohan M, Bartkuhn M, Herold M, Philippen A, Heinl N, Bardenhagen I,**
655 **Leers J, White RAH, Renkawitz-Pohl R, Saumweber H, Renkawitz R**.
656 2007. The Drosophila insulator proteins CTCF and CP190 link enhancer
657 blocking to body patterning. EMBO J. **26**:4203–4214.
658 26. **Moon H, Filippova G, Loukinov D, Pugacheva E, Chen Q, Smith ST,**
659 **Munhall A, Grewe B, Bartkuhn M, Arnold R, Burke LJ, Renkawitz-Pohl**
660 **R, Ohlsson R, Zhou J, Renkawitz R, Lobanenkov V**. 2005. CTCF is
661 conserved from Drosophila to humans and confers enhancer blocking of the
662 Fab-8 insulator. EMBO Rep. **6**:165–170.
663 27. **Peifer M, Bender W**. 1986. The anterobithorax and bithorax mutations of
664 the bithorax complex. EMBO J. **5**:2293–2303.

665 28. **Whitfield WG, Millar SE, Saumweber H, Frasch M, Glover DM**. 1988.
666 Cloning of a gene encoding an antigen associated with the centrosome in
667 Drosophila. J. Cell Sci. **89 (Pt 4)**:467–480.
668 29. **Birch-Machin I, Gao S, Huen D, McGirr R, White RAH, Russell S**. 2005.
669 Genomic analysis of heat-shock factor targets in Drosophila. Genome Biol.
670 **6**:R63.
671 30. **El-Sharnouby S, Redhouse J, White RAH**. 2013. Genome-wide and cell-
672 specific epigenetic analysis challenges the role of polycomb in Drosophila
673 spermatogenesis. PLoS Genet. **9**:e1003842.
674 31. **Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T,**
675 **Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R,**
676 **Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J,**
677 **Mirny LA, Lander ES, Dekker J**. 2009. Comprehensive mapping of long-
678 range interactions reveals folding principles of the human genome. Science
679 **326**:289–293.
680 32. **Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W,**
681 **Forné T**. 2007. Quantitative analysis of chromosome conformation capture
682 assays (3C-qPCR). Nat. Protoc. **2**:1722–1733.
683 33. **Dekker J, Rippe K, Dekker M, Kleckner N**. 2002. Capturing chromosome
684 conformation. Science **295**:1306–1311.
685 34. **Essien K, Vigneau S, Apreleva S, Singh LN, Bartolomei MS, Hannenhalli**
686 **S**. 2009. CTCF binding site classes exhibit distinct evolutionary, genomic,
687 epigenomic and transcriptomic features. Genome Biol. **10**:R131.
688 35. **Rhee HS, Pugh BF**. 2011. Comprehensive genome-wide protein-DNA
689 interactions detected at single-nucleotide resolution. Cell **147**:1408–1419.

690 36. **Ni X, Zhang YE, Nègre N, Chen S, Long M, White KP**. 2012. Adaptive
691 evolution and the birth of CTCF binding sites in the Drosophila genome.
692 PLoS Biol. **10**:e1001420.
693 37. **Gerasimova TI, Lei EP, Bushey AM, Corces VG**. 2007. Coordinated control
694 of dCTCF and gypsy chromatin insulators in Drosophila. Mol. Cell **28**:761–
695 772.
696 38. **Pai C-Y, Lei EP, Ghosh D, Corces VG**. 2004. The centrosomal protein CP190
697 is a component of the gypsy chromatin insulator. Mol. Cell **16**:737–748.
698 39. **Belozerov VE, Majumder P, Shen P, Cai HN**. 2003. A novel boundary
699 element may facilitate independent gene regulation in the Antennapedia
700 complex of Drosophila. EMBO J. **22**:3113–3121.
701 40. **Ohtsuki S, Levine M**. 1998. GAGA mediates the enhancer blocking activity
702 of the eve promoter in the Drosophila embryo. Genes Dev. **12**:3325–3330.
703 41. **Schweinsberg S, Hagstrom K, Gohl D, Schedl P, Kumar RP, Mishra R,**
704 **Karch F**. 2004. The enhancer-blocking activity of the Fab-7 boundary from
705 the Drosophila bithorax complex requires GAGA-factor-binding sites.
706 Genetics **168**:1371–1384.
707 42. **Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M,**
708 **Oberdoerffer P, Sandberg R, Oberdoerffer S**. 2011. CTCF-promoted RNA
709 polymerase II pausing links DNA methylation to splicing. Nature **479**:74–
710 79.
711 43. **Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R**. 1998. Binding of
712 trithorax and Polycomb proteins to the bithorax complex: dynamic changes
713 during early Drosophila embryogenesis. EMBO J. **17**:5141–5150.

714 44. **Qian S, Capovilla M, Pirrotta V**. 1991. The bx region enhancer, a distant
715 cis-control element of the Drosophila Ubx gene and its regulation by
716 hunchback and other segmentation genes. EMBO J. **10**:1415–1425.
717 45. **Simon J, Peifer M, Bender W, O'Connor M**. 1990. Regulatory elements of
718 the bithorax complex that control expression along the anterior-posterior
719 axis. EMBO J. **9**:3945–3956.
720 46. **Modolell J, Bender W, Meselson M**. 1983. Drosophila melanogaster
721 mutations suppressible by the suppressor of Hairy-wing are insertions of a
722 7.3-kilobase mobile element. Proc. Natl. Acad. Sci. U. S. A. **80**:1678–1682.
723 47. **Spana C, Harrison DA, Corces VG**. 1988. The Drosophila melanogaster
724 suppressor of Hairy-wing protein binds to specific sequences of the gypsy
725 retrotransposon. Genes Dev. **2**:1414–1423.
726 48. **White RAH, Wilcox M**. 1985. Regulation of the distribution of Ultrabithorax
727 proteins in Drosophila. Nature **318**:563–567.
728 49. **Pirrotta V, Chan CS, McCabe D, Qian S**. 1995. Distinct parasegmental and
729 imaginal enhancers and the establishment of the expression pattern of the
730 Ubx gene. Genetics **141**:1439–1450.
731 50. **Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F**. 1994. The
732 Trithorax-like gene encodes the Drosophila GAGA factor. Nature **371**:806–
733 808.
734 51. **Bejarano F, Busturia A**. 2004. Function of the Trithorax-like gene during
735 Drosophila development. Dev. Biol. **268**:327–341.
736 52. **Brown JL, Fritsch C, Mueller J, Kassis JA**. 2003. The Drosophila pho-like
737 gene encodes a YY1-related DNA binding protein that is redundant with
738 pleiohomeotic in homeotic gene silencing. Development **130**:285–294.

739 53. **Kanduri M, Kanduri C, Mariano P, Vostrov AA, Quitschke W,**
740 **Lobanenkov V, Ohlsson R**. 2002. Multiple nucleosome positioning sites
741 regulate the CTCF-mediated insulator function of the H19 imprinting
742 control region. Mol. Cell. Biol. **22**:3339–3344.
743 54. **Cléard F, Moshkin Y, Karch F, Maeda RK**. 2006. Probing long-distance
744 regulatory interactions in the Drosophila melanogaster bithorax complex
745 using Dam identification. Nat. Genet. **38**:931–935.
746 55. **Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V**. 2007. Polycomb
747 response elements mediate the formation of chromosome higher-order
748 structures in the bithorax complex. Nat. Cell Biol. **9**:1167–1174.
749 56. **Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu**
750 **M, Trono D, Spitz F, Duboule D**. 2013. A switch between topological
751 domains underlies HoxD genes collinearity in mouse limbs. Science
752 **340**:1234167.
753 57. **Chambeyron S, Bickmore WA**. 2004. Chromatin decondensation and
754 nuclear reorganization of the HoxB locus upon induction of transcription.
755 Genes Dev. **18**:1119–1130.
756 58. **Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N,**
757 **Fan Y, Skoultchi AI, Wutz A, Bickmore WA**. 2010. Ring1B compacts
758 chromatin structure and represses gene expression independent of histone
759 ubiquitination. Mol. Cell **38**:452–464.
760 59. **Ferraiuolo MA, Rousseau M, Miyamoto C, Shenker S, Wang XQD, Nadler**
761 **M, Blanchette M, Dostie J**. 2010. The three-dimensional architecture of
762 Hox cluster silencing. Nucleic Acids Res. **38**:7472–7484.

60. **Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D**. 2011. The dynamic architecture of Hox gene clusters. Science **334**:222–225.

Figure Legends

Figure 1 A variably occupied CTCF site in the *Ubx* gene. (A) CTCF binding profiles from T1 (*Ubx* inactive; blue) and T3 (*Ubx* active; green) leg imaginal discs. The arrow indicates the variably occupied CTCF site. *Ubx*, *abd-A* and *Abd-B* are transcribed from right to left. (B) The CTCF ChIP peak aligns with a match to the CTCF position-specific weight matrix. The positions of the PCR primers used in (D) are shown. (C) Phastcons conservation plot across 15 insect species (http://genome.ucsc.edu). The sequence at the variable CTCF site is compared with the *Drosophila* consensus (red; 36). The conserved CC motif (34) and conserved T in module #1 of Rhee and Pugh (35) are indicated blue. (D) ChIP-PCR confirming the differential binding of CTCF at the variable site. *Ubx*P is at the *Ubx* promoter, for -ve and +ve primers see Table 1.

Figure 2 ChIP-PCR analysis of binding of CP190, GAF and RNAPolII (Ser 5) in the region of the variably occupied CTCF site. RNAPolII (Ser5) refers to the Ser5-phosphorylated form of RNAPolII. T1 chromatin in blue, T3 chromatin in green. Primers as in Figure 1. * p-value = 0.02 (t-test).

Figure 3 Chromatin Interactions in the BX-C in T1 and T3. (A) 3C interactions at 29 sites in the BX-C. The top panel provides an overview of the BX-C showing the T3 CTCF ChIP profile with 3C anchor positions highlighted in grey. The lower panels show the 3C profiles; T1 is in blue and T3 is in green. Anchor 1 (primer 589) is at the variable CTCF site, anchor 2 (primer 675) is at the *Ubx* promoter and anchor 3 (primer 983) is at the *Mcp* boundary. Anchor positions are indicated by red shaded bars, orange shaded bars indicate positions detailed in B. Small arrows in anchor 2 panel indicate interactions of the *Ubx* promoter with sites in the *abx* (left) and *pbx* (right) regulatory regions. The grey dotted vertical line indicates the boundary between the *Ubx* and *abd-A* regulatory domains (60). Primers are listed in Table 2. (B) T1 versus T3 comparisons focussing on selected primers that are closest to key genomic features; for the interactions between Anchors and the variable CTCF site we show data for primers 9 and 10; for the *Ubx* promoter: primers 12 and 13 and for the *abd-A* promoter primers 24 and 25. Error bars are standard error of the mean. T1, blue; T3, green.

Figure 4 *Ubx* regulation and *bx* mutations. (A) Map of the *Ubx* regulatory region. Enhancers in green, PREs in red, regulatory regions defined by mutation in blue. Black rectangle on *gypsy* transposable element indicates Su(Hw) binding sites. Coordinates: *abx* enhancer "abx20" (45), *bx* and *bxd* PREs (61), *pbx* and *bxd* mutations (60), abx^{1} and *bx* alleles (27), BRE (44). The *gypsy* insertion in bx^{83Ka} was mapped by sequencing: the insertion is at chr3R: 12,528,835 with a 6bp duplication of the target site 12,528,830-12,528,835. In addition to the indicated cluster of *bx* alleles there is also an

outlier, *bx*^{F31}, associated with an *I* element insertion at approximately 12,516,500 (27). (B) Immunofluorescence labelling of Ubx expression in wild type T3 leg imaginal disc. (C) Immunofluorescence labelling of Ubx expression in *bx*^{83Ka} T3 leg imaginal disc. The discs in B and C are oriented with anterior to the left; in C Ubx expression is strongly reduced in the anterior compartment.

Figure 5 Chromatin Interactions in the BX-C in T1 and T3 comparing wild type and *bx*^{83Ka} mutant. (A) 3C interactions at 29 sites in the BX-C. The top panel provides an overview of the BX-C showing the T3 CTCF ChIP profile with 3C anchor positions highlighted in grey. The positions of the *abx* and *pbx* regulatory regions are indicated, corresponding to *abx*^{1} deletion (27) and pbx deletions (60). The lower panels show the 3C profiles; T1 is in blue and T3 is in green. Top two 3C profiles: anchor at variable CTCF site (primer 590). Bottom two profiles: anchor at *Ubx* promoter (primer 675). Anchor positions are indicated by red shaded bars, orange shaded bars indicate positions detailed in B. The grey dotted vertical line indicates the boundary between the *Ubx* and *abd-A* regulatory domains (60). Primers are listed in Table 2. (B) Comparisons of interactions at specific sites focussing on selected primers that are closest to key genomic features; for the interactions between Anchors and the *abx region* we show primer 8; for the variable CTCF site: primers 9 and 10; for the *Ubx* promoter: primers 12 and13; for the *pbx* region: primer 17 and for the *abd-A* promoter: primers 24 and 25. Error bars are standard error of the mean. T1 wildtype, blue; T1 *bx*^{83Ka}, light blue T3 wildtype, green; T3 *bx*^{83Ka}, light green.

Figure 6 Binding of CTCF, GAF and RNAPolII (Ser 5) in the region of the variably occupied CTCF site in *bx^{83Ka}* mutant. (A) ChIP-PCR analysis in T1 and T3 in *bx^{83Ka}* mutant. RNAPol (Ser5) refers to the Ser5-phosphorylated form of RNAPolII. T1 chromatin in light blue, T3 chromatin in light green. Primers as in Figure 1. (B) Comparison of wild type versus *bx^{83Ka}* at T1 and T3 for the CTCF peak (primer 3) and for the GAF peak (primer1). Error bars are standard error of the mean. T1 wildtype, blue; T1 *bx^{83Ka}*, light blue T3 wildtype, green; T3 *bx^{83Ka}*, light green. The GAF binding interval is from (62).

Figure 7 Chromatin accessibility and protein binding at CTCF sites in the BX-C. (A) In the repressed BX-C in Kc cells, DNase1 profiling reveals specific accessible sites in the repressed chromatin. Thirteen CTCF sites, bound in T3 chromatin, are numbered; 11 of the 13 are associated with DNase1 sensitivity peaks. (B) Close up of selected sites; the binding peaks of several regulators align with the DNAse1 sites. The variable CTCF site (Site 1) is offset from this alignment whereas other, constitutive, CTCF sites are more closely aligned with the DNase1 sites. Data from: CTCF T1 and T3 leg: this paper; Pho (63); Yki and GAF (62); DNase1 Kc (64); CTCF Kc: ModENCODE DCC ID 908.

Table 1 ChIP-qPCR primers

ID in Figs 1,2& 6	Primer Name	Chr	Forward		Reverse	
			Start	End	Start	End
0	Neg	3R	CCTAAATGGCAGAGGATTGG		AAATTCAGGATGCAGGATGC	
			12526683	12526702	12526792	12526773
1	R1	3R	ATCAGCAGCCGTTGAGTAGG		ATTCCTCAGCGACAAAGAGC	
			12528866	12528885	12528971	12528952
2	R2	3R	GAGTTGCCATAAAGCACTCG		TTCTCTTCGCAGCCTATTCC	
			12529660	12529679	12529764	12529745
3	R3	3R	TTACAGCCGACACCTCATCA		CTGGCTTGACACTGGGCTAC	
			12529861	12529880	12529987	12529968
4	R4	3R	CTCGCTGGTTCCTAATATGATATAC		GTGCCTTTCGGTGACTTC	
			12530745	12530769	12530863	12530846
5	R5	3R	GCACAGATTCCGTTGAGC		CCTTCTATGCTCTGCTCTCG	
			12531112	12531129	12531253	12531234
+ve	BXC-49	3R	ATCGATAAAAAGCGCCAACA		GCTCTTACTGCCCGATTCTG	
			12760726	12760707	12760565	12760584
-ve	SuVar 3-9	3R	AGCCGCTACTATTGCTTGGA		GCAGCGACAGCAGTATGAAA	
			11087377	11087396	11087573	11087554
Ubx-P	F-675	3R	AATACTTGGATTGCGCTTGC		TTTCCACTAGATTGGCGTCC	
			12559800	12559819	12560001	12559982

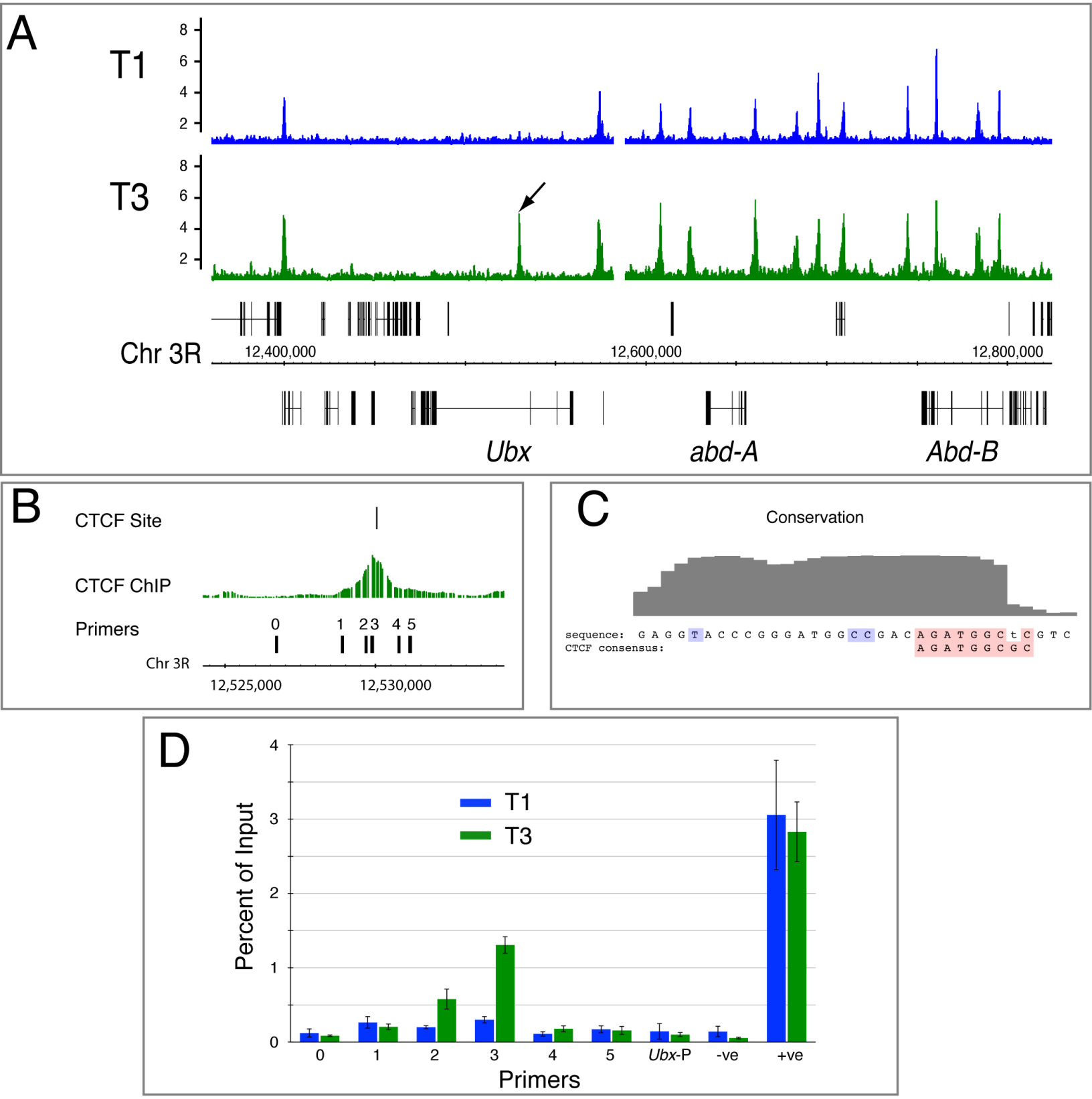
Table 2. 3C Primers

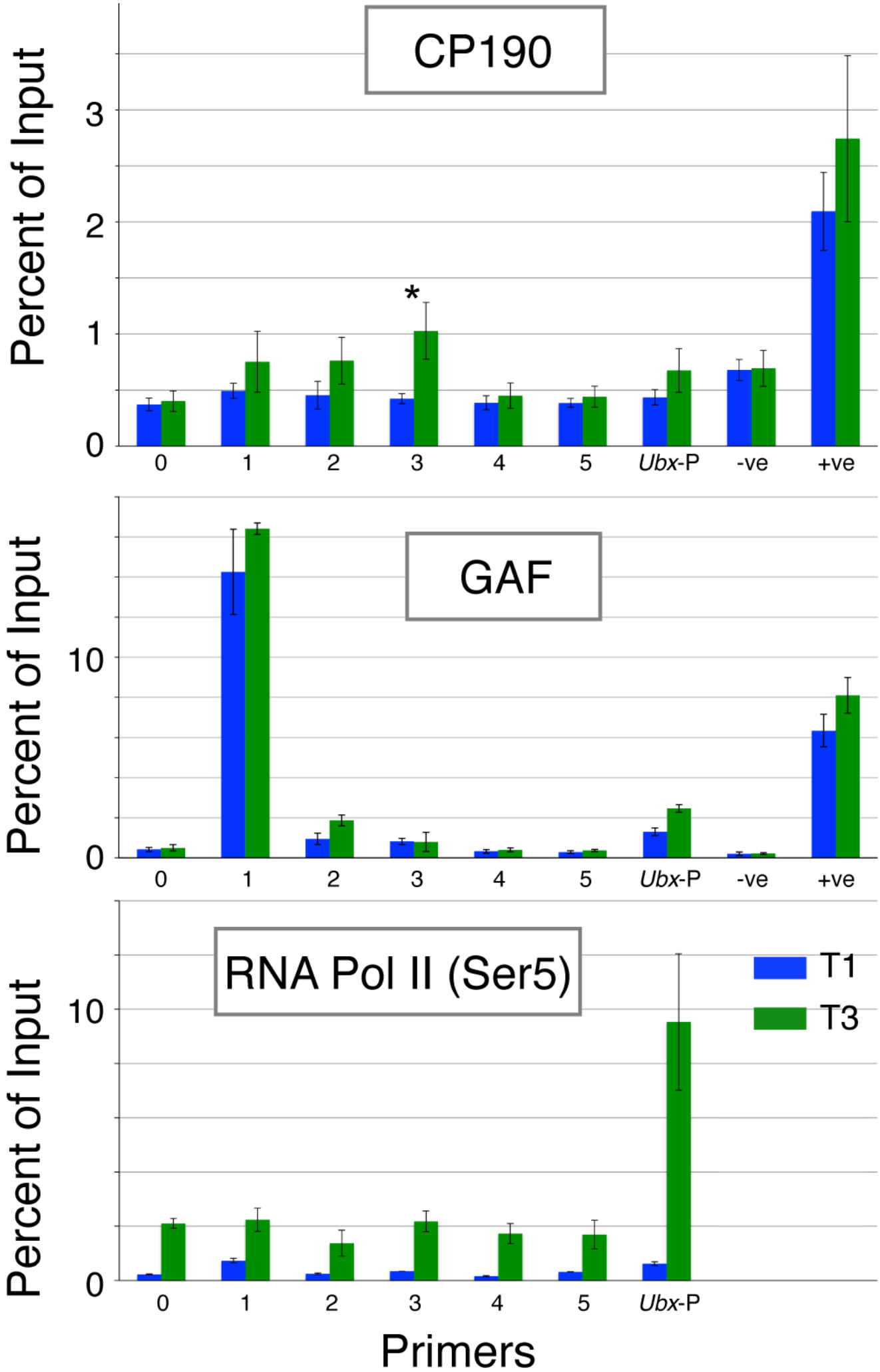
Anchor fragment internal primers

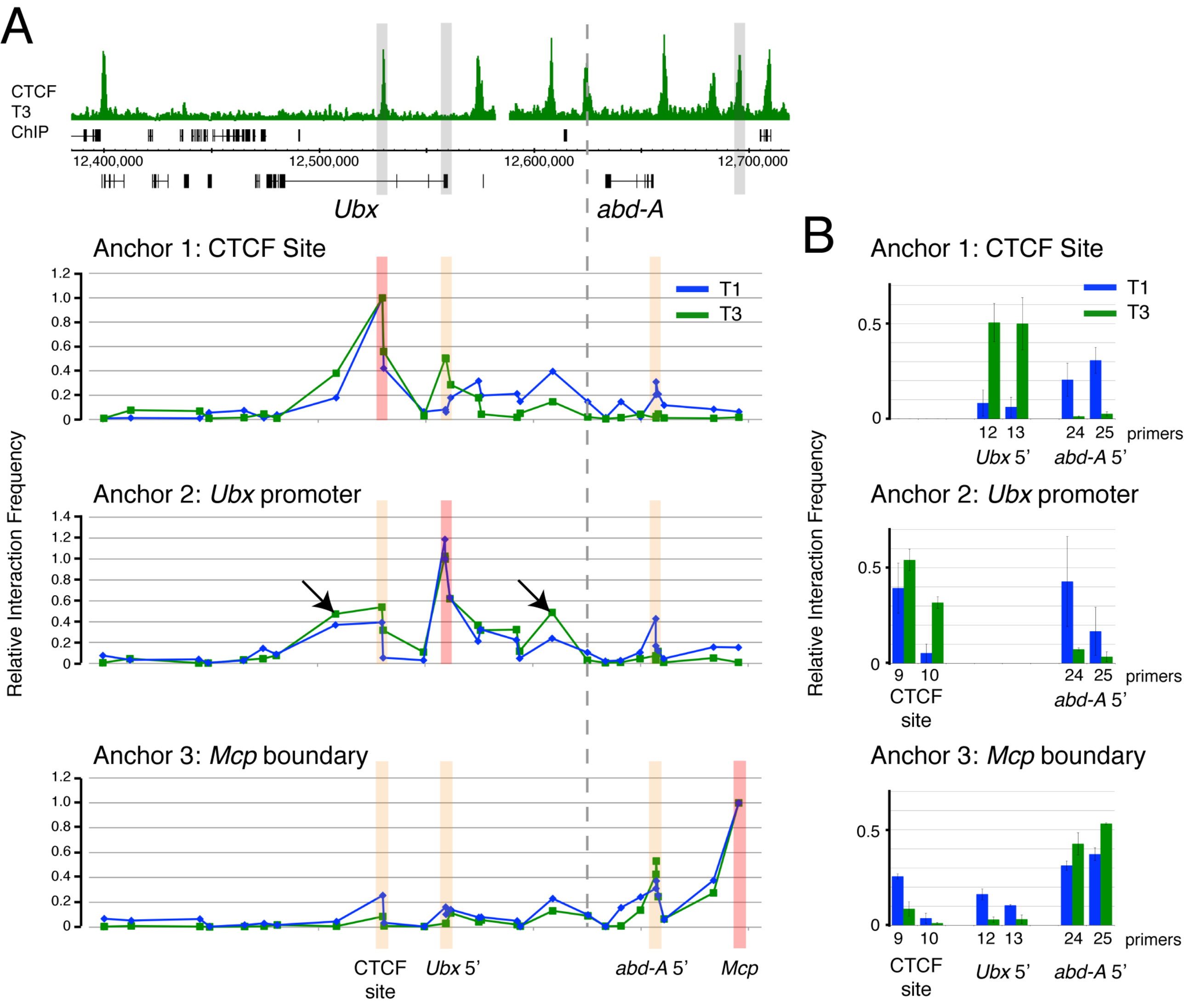
Anchor Position	Fragment ID	Chr	Forward		Reverse	
			Start	End	Start	End
Ubx Promoter	675	3R	AATACTTGGATTGCGCTTGC		TTTCCACTAGATTGGCGTCC	
			12559800	12559819	12560001	12559982
Variable CTCF Site_1	589	3R	TTACAGCCGACACCTCATCA		CTGGCTTGACACTGGGCTAC	
			12529861	12529880	12529987	12529968
Variable CTCF Site_2	590	3R	AGGGTTAATTCGTTCATCGC		CTGATGATGACGCTGTTGTG	
			12530221	12530240	12530362	12530343
Mcp	983	3R	ATTGTATGTATCCGCTCCGC		AAGCCCTTATTTGCAGACCC	
			12694755	12694774	12694917	1269898

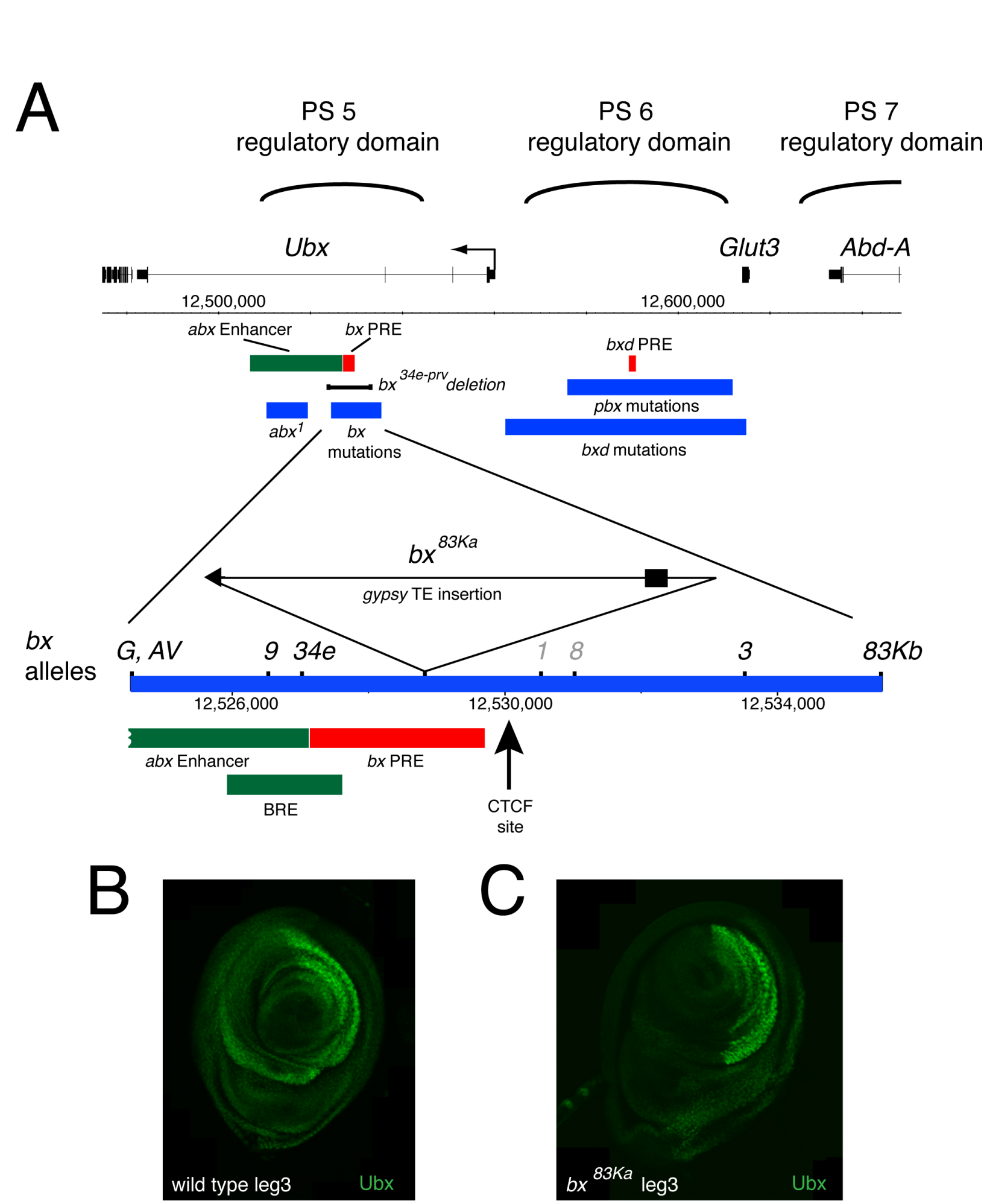
3C Primers

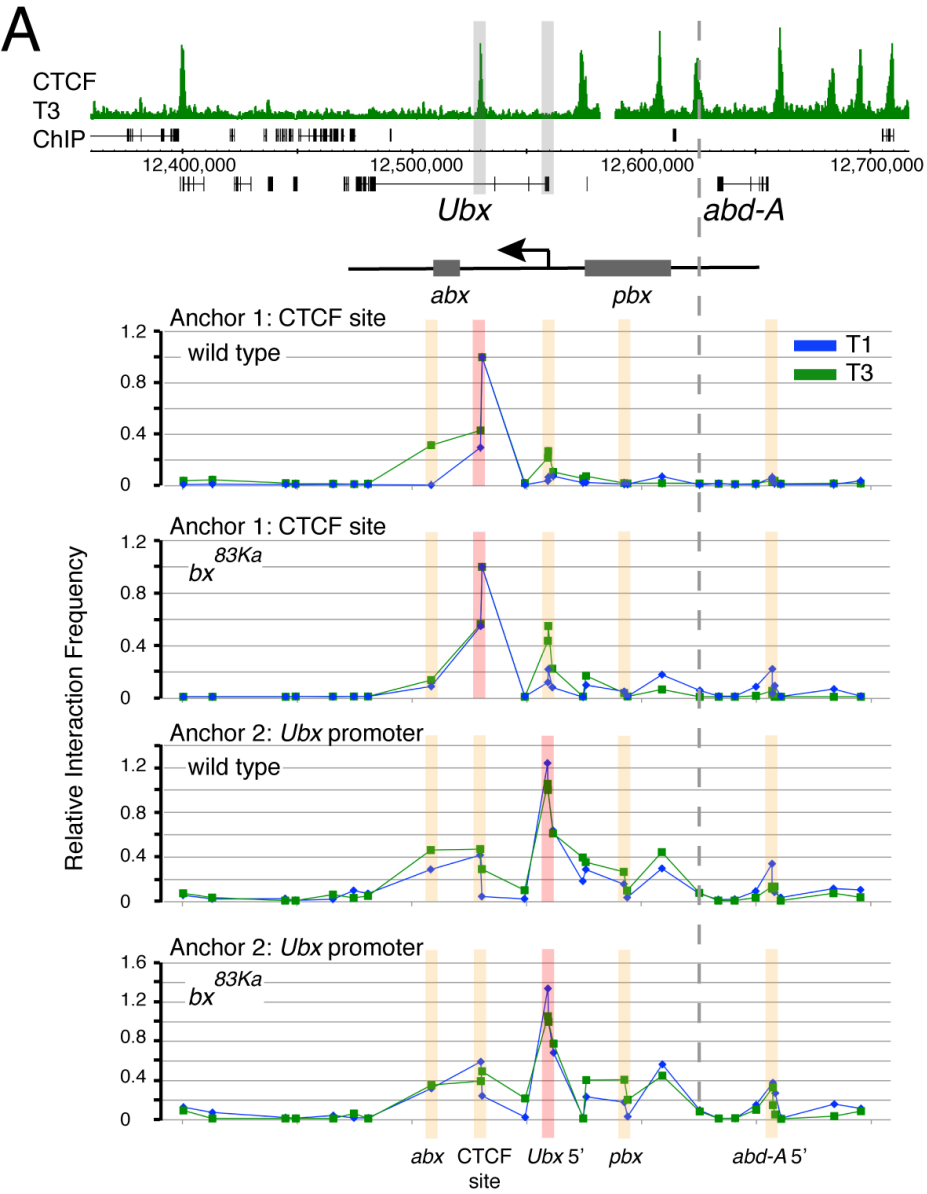
ID	Primer	Chr	Start	End	Primer Sequence
1	223	Chr3R	12400341	12400360	GCGAGACGATAAACGACGAC
2	237	Chr3R	12412997	12413016	AAGAAGTGGTAAAGTGGCGG
3	372	Chr3R	12444906	12444925	CTGTGCATCTCCACCACATC
4	396	Chr3R	12449306	12449325	CAGAAGCTGCCTCTCGTAGG
5	444	Chr3R	12465581	12465600	CAAAGCCACCTTCCTGAAAC
6	478	Chr3R	12474725	12474744	ATCTCGCCCAGCACTATTTG
7	504	Chr3R	12480871	12480890	TTTGAGTGGGTTAAGCTGCC
8	559	Chr3R	12508313	12508332	TAAATACGAAGTGCATGCGG
9	589	Chr3R	12529861	12529880	TTACAGCCGACACCTCATCA
10	590	Chr3R	12530474	12530494	GGAACACGCATATAGCATTGG
11	636	Chr3R	12549178	12549196	TTTGAAATGCAAACACGGC
12	674	Chr3R	12559159	12559178	GGAGGCCTGTTCAAAGTACG
13	675	Chr3R	12559351	12559332	CAAAGGAGGCAAAGGAACAG
14	677	Chr3R	12561570	12561589	CGAGAAGACCCAGAGCAAAG
15	698	Chr3R	12574489	12574509	AAGAAATATGCGTTTCCCACC
16	699	Chr3R	12575770	12575788	CGCCAGACAATGGAAACTG
17	745	Chr3R	12592412	12592433	GTGCTATCAACTCGCTTTCTTG
18	751	Chr3R	12593896	12593915	CTCTTTGTTAGCGGAGGCAG
19	789	Chr3R	12608923	12608942	TAAGCGAGTGCGTGTCATTC
20	842	Chr3R	12625282	12625303	TCATCTGGAACTGGTTCTATCG
21	858	Chr3R	12633588	12633607	AATCCGGTTGTGAAACAAGG
22	875	Chr3R	12640691	12640710	TCAGTCTCACAGCCATTTCG
23	899	Chr3R	12649777	12649797	GCATGTGCATTTAAGGAGTGG
24	918	Chr3R	12657009	12657031	CCAGTTAATGTGCTTCCTACCTG
25	918	Chr3R	12657020	12657043	GCTTCCTACCTGTCTATTTGTTGG
26	919	Chr3R	12658026	12658046	GTGTCGAGTTTCGGTTGAGTC
27	923	Chr3R	12660715	12660734	AAATGTTTGGACGGGAAATG
30	961	Chr3R	12683796	12683817	GCTTTAACTTTAACCTCTGGCG
31	983	Chr3R	12695484	12695507	CTGCTCTGCTTATCAGTTTATTGG

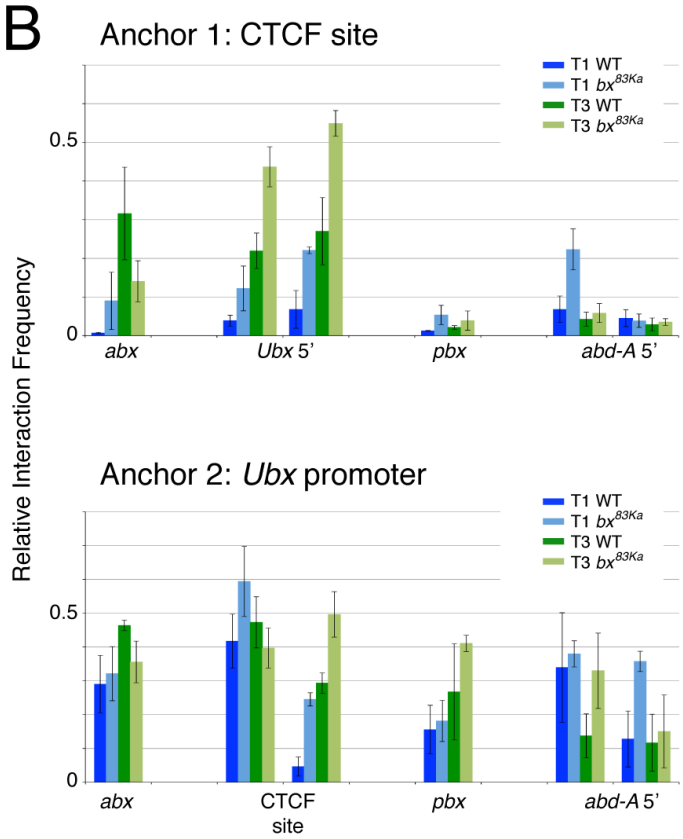

A
T1
T3
8
6
4
2
Chr 3R
12,400,000
12,600,000
12,800,000
Ubx
abd-A
Abd-B
B
CTCF Site
CTCF ChIP
Primers
0
1
2 3
4 5
Chr 3R
12,525,000
12,530,000
C
Conservation
sequence: G A G G T A C C C G G G A T G G C C G A C A G A T G G C t C G T C
CTCF consensus: A G A T G G C G C
D
T1
T3
Percent of Input
0
1
2
3
4
0
1
2
3
4
5
Ubx-P
-ve
+ve
Primers

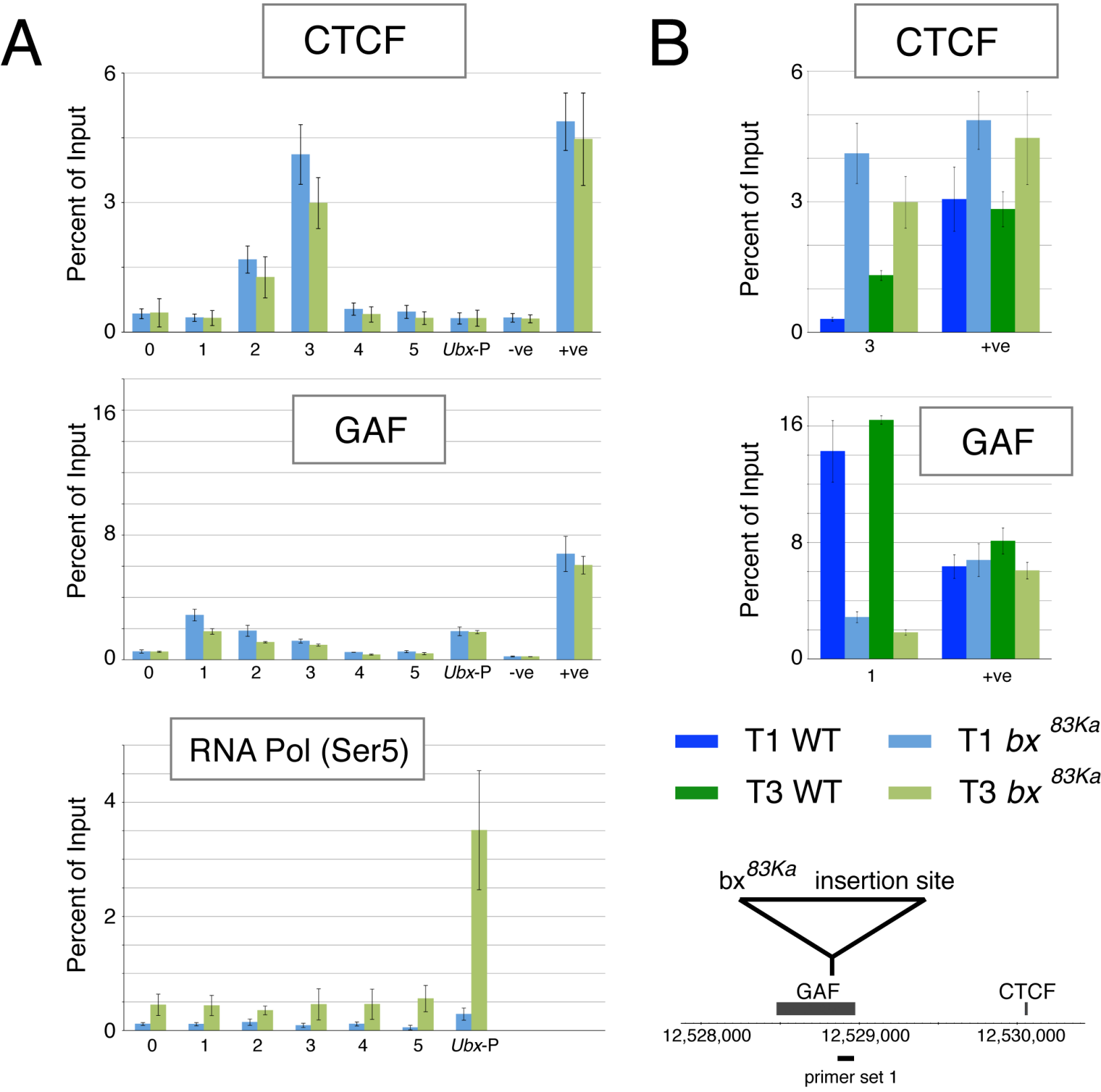

CP190
Percent of Input
0
1
2
3
*
0
1
2
3
4
5
Ubx-P
-ve
+ve
GAF
Percent of Input
0
10
0
1
2
3
4
5
Ubx-P
-ve
+ve
RNA Pol II (Ser5)
T1
T3
Percent of Input
0
10
0
1
2
3
4
5
Ubx-P
Primers

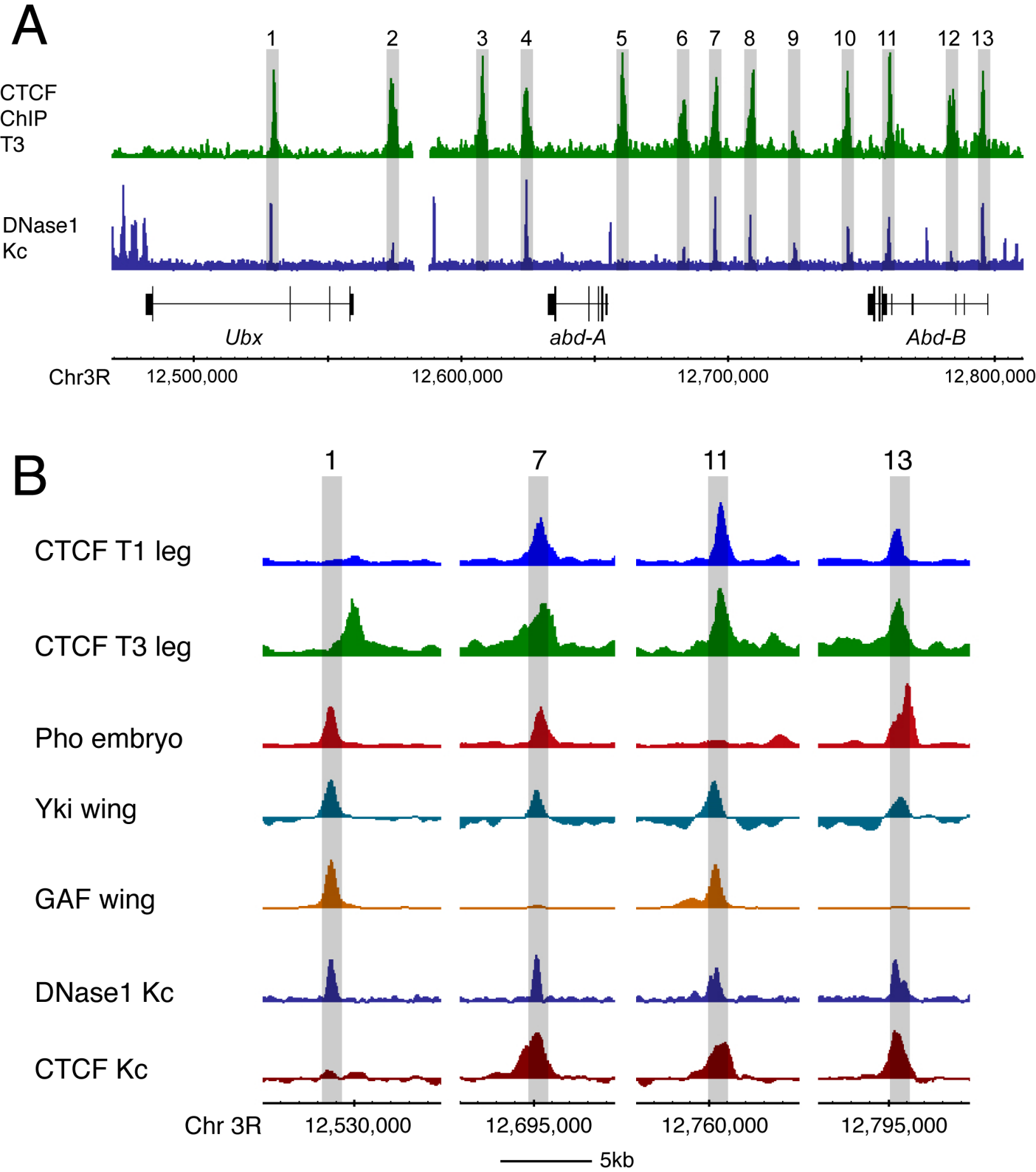

A
CTCF
T3
ChIP
12,400,000
12,500,000
12,600,000
12,700,000
Ubx
abd-A
Anchor 1: CTCF Site
T1
T3
Relative Interaction Frequency
Anchor 2: Ubx promoter
Anchor 3: Mcp boundary
CTCF site
Ubx 5'
abd-A 5'
Mcp
B
Anchor 1: CTCF Site
T1
T3
12 13 24 25 primers
Ubx 5' abd-A 5'
Relative Interaction Frequency
Anchor 2: Ubx promoter
9 10 24 25 primers
CTCF site
abd-A 5'
Anchor 3: Mcp boundary
9 10 12 13 24 25 primers
CTCF site
Ubx 5'
abd-A 5'


A
PS 5
regulatory domain
PS 6
regulatory domain
PS 7
regulatory domain
Ubx
Glut3
Abd-A
12,500,000
12,600,000
abx Enhancer
bx PRE
bxd PRE
bx 34e-prv deletion
pbx mutations
abx1
bx
mutations
bxd mutations
bx 83Ka
gypsy TE insertion
bx
alleles
G, AV
9
34e
1
8
3
83Kb
12,526,000
12,530,000
12,534,000
abx Enhancer
bx PRE
BRE
CTCF
site
B
wild type leg3
Ubx
C
bx 83Ka leg3
Ubx


A
CTCF
T3
ChIP
12,400,000
12,500,000
12,600,000
12,700,000
Ubx
abd-A
abx
pbx
Anchor 1: CTCF site
wild type
T1
T3
Anchor 1: CTCF site
bx83Ka
Anchor 2: Ubx promoter
wild type
Anchor 2: Ubx promoter
bx83Ka
Relative Interaction Frequency
abx
CTCF site
Ubx 5'
pbx
abd-A 5'


B
Anchor 1: CTCF site
T1 WT
T1 bx83Ka
T3 WT
T3 bx83Ka
Relative Interaction Frequency
abx
Ubx 5'
pbx
abd-A 5'
Anchor 2: Ubx promoter
CTCF site


A
CTCF
Percent of Input
6
3
0
0
1
2
3
4
5
Ubx-P
-ve
+ve
GAF
16
8
RNA Pol (Ser5)
4
2
B
T1 WT
T1 bx 83Ka
T3 WT
T3 bx 83Ka
bx 83Ka
insertion site
GAF
CTCF
12,528,000
12,529,000
12,530,000
primer set 1


A
1
2
3
4
5
6
7
8
9
10
11
12
13
CTCF ChIP T3
DNase1 Kc
Ubx
abd-A
Abd-B
Chr3R
12,500,000
12,600,000
12,700,000
12,800,000
B
1
7
11
13
CTCF T1 leg
CTCF T3 leg
Pho embryo
Yki wing
GAF wing
DNase1 Kc
CTCF Kc
Chr 3R
12,530,000
12,695,000
12,760,000
12,795,000
5kb