Journal of Visualized Experiments

In-nucleus Hi-C in Drosophila Cells

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE62106R1
Full Title:	In-nucleus Hi-C in Drosophila Cells
Corresponding Author:	Mayra Furlan-Magaril MEXICO
Corresponding Author's Institution:	
Corresponding Author E-Mail:	mfurlan@ifc.unam.mx
Order of Authors:	Ayerim Esquivel-Lopez
	Rodrigo Arzate-Mejia
	Rosario Perez-Molina
	Mayra Furlan-Magaril
Additional Information:	
Question	**Response**
Please specify the section of the submitted manuscript.	Genetics
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Mexico City, Mexico
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please provide any comments to the journal here.	

TITLE:

In-nucleus Hi-C in Drosophila Cells

AUTHORS AND AFFILIATIONS:

Ayerim Esquivel-López[1], Rodrigo Arzate-Mejía[1], Rosario Pérez-Molina[1], Mayra Furlan-Magaril[1]*

[1]Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México 04510 Ciudad de México, México

Ayerim Esquivel-López (aesquivel@ifc.unam.mx)
Rodrigo Arzate-Mejía (rarzate89@gmail.com)
Rosario Pérez-Molina (rmolina@ifc.unam.mx)

*Corresponding author
Mayra Furlan-Magaril (mfurlan@ifc.unam.mx)

KEYWORDS:

Chromatin, Genome 3D organization, compartments, topologically associating domains

SUMMARY:

The genome is organized in the nuclear space into different structures that can be revealed through chromosome conformation capture technologies. The in-nucleus Hi-C method provides a genome-wide collection of chromatin interactions in Drosophila cell lines, which generates contact maps that can be explored at megabase resolution at restriction fragment level.

ABSTRACT:

The genome is organized into topologically associating domains (TADs) delimited by boundaries that isolate interactions between domains. In Drosophila, the mechanisms underlying TAD formation and boundaries are still under investigation. The application of the in-nucleus Hi-C method described here helped to dissect the function of architectural protein (AP)-binding sites at TAD boundaries isolating the *Notch* gene. Genetic modification of domain boundaries that cause loss of APs results in TAD fusion, transcriptional defects, and long-range topological alterations. These results provided evidence demonstrating the contribution of genetic elements to domain boundary formation and gene expression control in Drosophila. Here, the in-nucleus Hi-C method has been described in detail, which provides important checkpoints to assess the quality of the experiment along with the protocol. Also shown are the required numbers of sequencing reads and valid Hi-C pairs to analyze genomic interactions at different genomic scales. CRISPR/Cas9-mediated genetic editing of regulatory elements and high-resolution profiling of genomic interactions using this in-nucleus Hi-C protocol could be a powerful combination for the investigation of the structural function of genetic elements.

INTRODUCTION:

In eukaryotes, the genome is partitioned into chromosomes that occupy specific territories in

the nuclear space during interphase[1]. The chromatin forming the chromosomes can be divided into two main states: one of accessible chromatin that is transcriptionally permissive, and the other of compact chromatin that is transcriptionally repressive. These chromatin states segregate and rarely mix in the nuclear space, forming two distinct compartments in the nucleus[2]. At the sub-megabase scale, boundaries separate domains of high-frequency chromatin interactions, called TADs, that mark chromosomal organization[3-5]. In mammals, TAD boundaries are occupied by cohesin and CCCTC-binding factor (CTCF)[6-8]. The cohesin complex extrudes chromatin and halts at CTCF-binding sites that are disposed in a convergent orientation in the genomic sequence to form stable chromatin loops[9,10,13,14]. Genetic disruption of the CTCF DNA-binding site at the boundaries or reduction in CTCF and cohesin protein abundance results in abnormal interactions between regulatory elements, loss of TAD formation, and gene expression deregulation[9-11,13,14].

In Drosophila, the boundaries between TADs are occupied by several APs, including boundary element-associated factor of 32 kDa, Motif 1 binding protein (M1BP), centrosomal protein 190, suppressor of hairy-wing, and CTCF, and are enriched in active histone modifications and Polymerase II[16-18]. It has been suggested that in Drosophila, TADs appear as a consequence of transcription[13,17,19], and the exact role of independent APs in boundary formation and insulation properties is still under investigation. Thus, whether domains in Drosophila are a sole consequence of the aggregation of regions of similar transcriptional states or whether APs, including CTCF, contribute to boundary formation remains to be fully characterized. Exploration of genomic contacts at high resolution has been possible through the development of chromosome conformation capture technologies coupled with next-generation sequencing. The Hi-C protocol was first described with the ligation step performed "in solution"[2] in an attempt to avoid spurious ligation products between chromatin fragments. However, several studies pointed to the realization that the useful signal in the data came from ligation products formed at partially lysed nuclei that were not in solution[20,21].

The protocol was then modified to perform the ligation inside the nucleus as part of the single-cell Hi-C experiment[22]. The in-nucleus Hi-C protocol was subsequently incorporated into cell population Hi-C to yield a more consistent coverage over the full range of genomic distances and produce data with less technical noise[23,24]. The protocol, described in detail here, is based on the population in-nucleus Hi-C protocol[23,24] and was used to investigate the consequences of genetically removing DNA-binding motifs for CTCF and M1BP from a domain boundary at the *Notch* gene locus in Drosophila[25]. The results show that altering the DNA-binding motifs for APs at the boundary has drastic consequences for Notch domain formation, larger topological defects in the regions surrounding the *Notch* locus, and gene expression deregulation. This indicates that genetic elements at domain boundaries are important for the maintenance of genome topology and gene expression in Drosophila[25].

PROTOCOL:

1. Fixation

1.1. Start with 10 million S2 receptor plus (S2R+) cells to prepare 17.5 mL of a cell suspension

in Schneider medium containing 10% fetal bovine serum (FBS) at room temperature (RT).

1.2. Add methanol-free formaldehyde to obtain a final concentration of 2%. Mix and incubate for 10 min at RT, taking care to mix every minute.

NOTE: Formaldehyde is a hazardous chemical. Follow the appropriate health and safety regulations, and work in the fume hood.

1.3. Quench the reaction by adding glycine to achieve a final concentration of 0.125 M and mix. Incubate for 5 min at RT, followed by 15 min on ice.

1.4. Centrifuge for 400 × *g* at RT for 5 min and then for 10 min at 4 °C; discard the supernatant. Resuspend the pellet carefully in 25 mL of cold 1x phosphate-buffered saline.

1.5. Centrifuge at 400 × *g* for 10 min at 4 °C, then discard the supernatant.

NOTE: If continuing with the protocol, go to step 2.1 for lysis; otherwise, flash-freeze the pellet in liquid N_2 and store the pellet at -80 °C.

2. Lysis

2.1. Resuspend the cells in 1 mL of ice-cold lysis buffer (10 mM Tris-HCl, pH 8; 0.2% of non-ionic surfactant (see the **Table of Materials**); 10 mM NaCl; 1x protease inhibitors), and adjust the volume to 10 mL with ice-cold lysis buffer. Adjust the volume to obtain a concentration of 1 × 10^6 cells/mL. Incubate on ice for 30 min, mixing every 2 min by inverting the tubes.

2.2. Centrifuge the nuclei at 300 × *g* for 5 min at 4 °C, and then carefully discard the supernatant. Wash the pellet 1x with 1 mL of cold lysis buffer, and transfer it to a microcentrifuge tube. Wash the pellet 1x with 1 mL of cold 1.25x restriction buffer, and resuspend each cell pellet in 360 µL of 1.25x restriction buffer.

2.3. Add 11 µL of 10% sodium dodecyl sulfate (SDS) per tube (0.3% final concentration), mix carefully by pipetting, and incubate at 37 °C for 45 min, shaking at 700–950 rpm. Pipet up and down to disrupt clumps a few times during incubation.

2.4. Quench the SDS by adding 75 µL of non-ionic surfactant (10% solution, see the **Table of Materials**) per tube (1.6% final concentration), and incubate at 37 °C for 45 min, shaking at 950 rpm. Pipet up and down a few times to disrupt clumps during incubation.

NOTE: If clumps are large and difficult to disrupt, decrease the rotating speed to 400 rpm during SDS and surfactant treatments. If the clumps are difficult to disaggregate by pipetting, split the sample into two; adjust the volumes of restriction buffer, SDS, and surfactant; and proceed with the permeabilization. Next, spin the nuclei at minimum speed (200 × *g*), carefully discard the supernatant, pool the samples together in 450 µL of 1x buffer, and proceed with digestion.

Take a 10 µL aliquot as the undigested sample (UD).

3. Enzymatic digestion

3.1. Digest the chromatin by adding 200 units (U) of Mbo I per tube, and incubate at 37 °C for a period ranging from 4 h to overnight while rotating (950 rpm).

3.2. On the next day, add an additional 50 U of Mbo I per tube, and incubate at 37 °C for 2 h while rotating (950 rpm).

3.3. Inactivate the enzyme by incubating the tubes at 60 °C for 20 min. Place the tubes on ice.

NOTE: Take a 10 µL aliquot as the digested sample (D).

4. Biotinylation of DNA ends

4.1. To fill in the restriction fragment overhangs and label the DNA ends with biotin, add 1.5 µL each of 10 mM dCTP, dGTP, dTTP, 20 µL of 0.4 mM biotin dATP, 17.5 µL of Tris low-EDTA (TLE) buffer [10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0], and 10 µL of 5 U/µL of Klenow (DNA polymerase I large fragment) to all the tubes. Mix carefully and incubate for 75 min at 37 °C. Shake at 700 rpm every 10 s for 30 s. Place all the tubes on ice while preparing the ligation mix.

5. Ligation

5.1. Transfer each digested chromatin mixture to a separate 1 mL tube with ligation mix (100 µL of 10x ligation buffer, 10 µL of 10 mg/mL bovine serum albumin, 15 U of T4 DNA ligase, and 425 µL of double-distilled water [ddH_2O]). Mix thoroughly by gentle pipetting, and incubate overnight at 16 °C.

6. Crosslink reversal and DNA purification

6.1. Degrade proteins by adding 50 µL of 10 mg/mL Proteinase K per tube, and incubate at 37 °C for 2 h. Reverse crosslinks by increasing the temperature to 65 °C and incubate overnight.

6.2. Degrade RNA by adding 20 µL of 10 mg/mL RNase A, and incubate at 37 °C for 1 h.

6.3. Perform phenol:chloroform extraction followed by ethanol precipitation.

6.3.1. Add 1 volume of phenol-chloroform, and mix thoroughly by inversion to obtain a homogeneous white phase.

NOTE: Phenol is a hazardous chemical. Follow the appropriate health and safety regulations. Work in the fume hood.

6.3.2. Centrifuge at 15,000 × *g* for 15 min. Transfer the aqueous phase into a fresh 2 mL microcentrifuge tube. Perform a back-extraction of the lower layer with 100 µL of TLE buffer, and transfer the aqueous phase into the same 2 mL tube.

6.3.3. Precipitate DNA by adding 2 volumes of 100% ethyl alcohol (EtOH), 0.1 volumes of 3 M sodium acetate, and 2 µL of 20 mg/mL glycogen. Incubate at -20 °C for a period ranging from 2 h to overnight.

6.3.4. Spin at 15,000 × *g* for 30 min at 4 °C, and wash the pellets 2x with ice-cold 70% EtOH. Dry the pellets at RT, and resuspend in 100 µL of TLE buffer. Quantify the DNA using a fluorogenic dye that binds selectively to DNA and a fluorometer according to the manufacturer's instructions (**Table of Materials**).

NOTE: The protocol can be paused here. Store ligation products at 4 °C for the short term or at -20 °C for the long term. Use an aliquot (100 ng) of the material as the ligated sample (L) for quality control.

7. Assess Hi-C template quality

7.1. Digestion and ligation qualitative controls

7.1.1. Purify DNA from the UD and D aliquots by reversing the crosslinks, and perform phenol:chloroform extraction and ethanol precipitation as described above.

7.1.2. Load 100 ng of UND, D, and L samples in a 1.5% agarose gel. Look for a smear centered around 500 bp in the D sample versus a high molecular weight band for the L sample (see representative results).

7.2. Digestion efficiency quantitative control

NOTE: To assess the digestion efficiency more accurately, use the UD and D samples as templates to perform quantitative polymerase chain reactions (qPCR) using primers designed as follows.

7.2.1. Design a primer pair that amplifies a DNA fragment containing the DNA restriction site for the enzyme used for digestion (Mbo I in the present protocol), called R in the formula in step 7.2.3.

7.2.2. Design a primer pair that amplifies a control DNA fragment that does not contain the restriction site for the enzyme used for digestion (Mbo I for the present protocol), called C in the formula in step 7.2.3.

7.2.3. Use the cycle threshold values (Ct values) of the amplification to calculate restriction

efficiency according to the formula shown below:

$$\% \text{ Restriction} = 100 - 100/2^{\{(CtR - CtC)D - (CtR - CtC)UD\}}$$

Where CtR refers to the Ct value of fragment R, and CtC refers to the Ct value of the fragment C for sample D and sample UD.

NOTE: The restriction percentage reflects the efficiency of the restriction enzyme cleaving the restricted (R) DNA fragment compared to a control (C) DNA fragment that does not contain the restriction DNA site. A restriction efficiency of ≥ 80% is recommended.

7.3. Detection of known interactions

7.3.1. Perform PCR to amplify an internal ligation control to examine short-range and/or medium- or long-range interactions (see representative results).

7.3.2. Alternatively, design primers to amplify a ligation product in which the primers are in forward-forward or reverse-reverse orientation in adjacent restriction fragments.

7.4. Fill-in and biotin-labeling control

7.4.1. Verify Hi-C marking and ligation efficiency by amplification and digestion of a known interaction or a ligation product between adjacent restriction fragments in the genome, as described above.

NOTE: Successful fill-in and ligation of the Mbo I site (GATC) generates a new site for the restriction enzyme Cla I (ATCGAT) at the ligation junction and regenerates the Mbo I site.

7.4.2. Digest the PCR product with Mbo I, Cla I, or both. After running the samples on a 1.5–2% gel, estimate the relative number of 3C and Hi-C ligation junctions by quantifying the intensity of the cut and uncut bands[26].

NOTE: An efficiency of > 70% is desired (see representative results).

8. Sonication

8.1. Sonicate the samples to obtain 200–500 bp DNA fragments. For the instrument used in this protocol (see the **Table of Materials**), dilute the sample (from 5 to 10 µg) in 130 µL of ddH_2O per tube, and set the instrument to sonicate to 400 bp: fill level: 10; duty factor: 10%; peak incident power (w): 140; cycles per burst: 200; time (s): 80.

9. Biotin removal/end repair

NOTE: The steps shown below are adjusted for 5 µg of Hi-C DNA.

9.1. To perform biotin removal, transfer the sample (130 µL) into a fresh microcentrifuge tube. Add 16 µL of 10x ligation buffer, 2 µL of 10 mM dATP, 5 µL of T4 DNA Polymerase (15U), and 7 µL of ddH_2O (160 µL of total volume). Incubate at 20 °C for 30 min.

9.2. Add 5 µL of 10 mM dNTPs, 4 µL of 10x ligation buffer, 5 µL of T4 polynucleotide kinase (10 U/µL), 1 µL of Klenow, and 25 µL of ddH_2O (200 µL of total volume). Incubate at 20 °C for 30 min.

10. Size selection

10.1. To select fragments mostly in the 250–550 bp size range, perform sequential solid phase reversible immobilization (SPRI) size selection first with 0.6x, followed by 0.90x according to the manufacturer's instructions, and elute the DNA using 100 µL of TLE.

11. Biotin pulldown/A-tailing/adapter ligation

11.1. Perform the washes by resuspending the magnetic beads by vortexing, rotate the samples for 3 min on a rotating wheel, and then briefly spin down the sample and place it on the magnetic stand.

11.2. Allow the beads to stick to the magnet, discard the supernatant, and proceed with the following wash step. Perform the washes at 55 °C on a thermo-block with rotation instead of the rotating wheel.

11.2.1. Make up the final volume to 300 µL per sample with TLE for pull-down. Prepare the bead-washing buffers: 1x Tween Buffer (TB) (TB: 5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween), 0.5x TB, 1x No-Tween buffer (NTB) (5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl), 2x NTB.

11.2.2. Use 150 µL of streptavidin-linked magnetic beads (see the **Table of Materials**) per library. Wash the beads 2x with 400 µL of 1x TB.

11.2.3. Wash the beads with 400 µL of 0.5x TB, and incubate at 55 °C for 3 min, rotating at 750 rpm. Wash the beads in 200 µL of 1x restriction buffer.

11.2.4. Resuspend the beads in 100 µL of dATP tailing mix (5 µL of 10 mM dATP, 10 µL of 10x restriction buffer, 5 µL of Klenow exo-, and 80 µL of ddH_2O). Incubate at 37 °C for 30 min.

11.2.5. Remove the supernatant, wash the beads 2x with 400 µL of 0.5x TB by incubating at 55 °C for 3 min and rotating at 750 rpm.

11.2.6. Wash the beads with 400 µL of 1x NTB and then with 100 µL of 1x ligation buffer.

11.2.7. Resuspend the beads in 50 µL of 1x ligation buffer, and transfer the suspension to a new tube. Add 4 µL of pre-annealed PE adapters (15 µM stock) and 2 µL of T4 ligase (400 U/µL, i.e., 800 U/tube); incubate at RT for 2 h.

NOTE: Pre-anneal the adapters by adding equal volumes of both PE 1.0 and PE 2.0 adapters (30 µM stock) and incubating for 10 min at RT (see the **Table of Materials**).

11.2.8. Recapture the beads by removing the supernatant and washing the beads 2x with 400 µL of TB. Wash the beads with 200 µL of 1x NTB, then with 100 µL of 1x restriction buffer, and resuspend the beads in 40 µL of 1x restriction buffer.

12. PCR amplification

12.1. Set up PCRs of 25 µL volume with 5, 6, 7, and 8 cycles. For each PCR, use 2.5 µL of Hi-C beads, 0.75 µL of 10 µM PE PCR primer 1 (**Table of Materials**), 0.75 µL of 10 µM PE PCR primer 2 (**Table of Materials**), 0.6 µL of 10 mM dNTP, 0.3 µL of DNA polymerase, and 14.65 µL of ddH_2O. Use the following PCR conditions: 98 °C for 30 s, 98 °C for 10 s, 65 °C for 30 s x n cycles,72 °C for 30 s, and 72 °C for 7 min.

12.2. Run the PCR products on a 1.5% agarose gel.

NOTE: A smear is expected around 400–1000 bp. The appropriate number of cycles for the final amplification is the number immediately lower than that at which a smear is just visible, i.e., if the smear is visible with 8 cycles in the test PCR, then use 7 cycles for the final amplification. Perform PCR for as few cycles as possible to avoid PCR duplicates.

13. Final PCR amplification

13.1. Perform final PCR amplification using the number of cycles selected in the previous step. Split the sample using 5 µL as template in 50 µL reactions.

13.2. Collect all PCR reactions and transfer them to a fresh tube. Use the magnet to remove the streptavidin beads and recover the supernatant (PCR products). Transfer the beads to a fresh tube, wash the beads as indicated in step 11.2.8, and store in 1x restriction buffer at 4 °C as a backup.

13.3. Purify the PCR products using 0.85x the volume of the SPRI beads according to the manufacturer´s instructions. Elute with 30 µL of TLE buffer.

13.4. Quantify the Hi-C library using a fluorometric instrument, and confirm the quality of the library by chip-based capillary electrophoresis.

13.5. As a last quality checkpoint, use 1 µL of the Hi-C library as a template to perform a PCR reaction using 10 cycles using the same conditions described in step 12.1. Divide the PCR

product into two microcentrifuge tubes: digest one with Cla I and leave the other one undigested as a control. Run the products in a 1.5–2% agarose gel (see representative results).

13.6. Proceed to 50 bp or 75 bp paired-end sequencing on a suitable sequencing platform.

REPRESENTATIVE RESULTS:

Described below are the results of a successful Hi-C protocol (see a summary of the Hi-C protocol workflow in **Figure 1A**). There are several quality control checkpoints during the in-nucleus Hi-C experiment. Sample aliquots were collected before (UD) and after (D) the chromatin restriction step as well as after ligation (L). Crosslinks were reversed, and DNA was purified and run on an agarose gel. A smear of 200–1000 bp was observed when restriction with Mbo I was successful (**Figure 1B**). The expected size of the molecule depends on the restriction enzyme of choice. If the ligation was successful, a high molecular weight band was seen at the top of the gel (**Figure 1B**). Digestion efficiency can be also confirmed by qPCR as described in detail in the protocol. An acceptable digestion efficiency is 80% or higher (**Figure 1C**).

To assess Hi-C ligation efficiency in detail, primers can be designed to amplify an internal ligation product control in which the primers are in forward-forward or reverse-reverse orientation in adjacent restriction fragments. Alternatively, primers can be designed to amplify known interactions. **Figure 2A** shows the amplification of a known medium-range (300 kb) interaction in Drosophila[25]. Hi-C ligation products (in which the biotin marking, fill-in, and ligation occurred successfully) can be estimated by digestion of the PCR product recovered in the amplification. After fill-in and ligation, Hi-C amplicons will contain a new Cla I restriction site at the original Mbo I site, which is preserved upon blunt-end ligation. If restriction with Cla I is not complete, the fill-in reaction and biotin marking will be inefficient. A digestion efficiency of more than 70% is recommended to avoid having a large proportion of non-useful reads for the libraries after sequencing (**Figure 2A**, compare the Cla I digestion of the 3C versus the Hi-C template).

To determine an adequate number of PCR amplification cycles to amplify the final Hi-C library, PCR reactions were set up using 2.5 µL of a given library on beads, as described in the protocol. The number of PCR cycles for the final amplification is one cycle less than the number of cycles for which the smear is visible (**Figure 2B**). In this case, 4 cycles of PCR amplification were chosen. As a final quality checkpoint, an aliquot of the Hi-C library was re-amplified and digested with Cla I. The level of digestion of the library (a decrease in the smear size) indicates the abundance of valid Hi-C pairs and reflects the proportion of useful reads that will be obtained from the library (**Figure 2C**). A ratio of the upper size range (determined by the size present in the UD sample) and the bottom size range in both UD and D samples should produce a ratio > 1 for the UD and a ratio ≤ 1 for the digested sample if the Cla I digestion is efficient.

After paired-end sequencing, the FASTQ files (**Table 1**) were processed using HiCPro[28] and the generated statistics plotted using MultiQC[28]. An alternative tool to HiCPro is the HiCUP[30] pipeline that yields similar results (not shown). **Figure 3** and **Table 2** show the detailed

statistical information of the sequenced reads. Full read alignment and alignment after trimming are reported. These two categories correspond to successfully aligned reads that will be used in subsequent analysis to find valid Hi-C pairs. The alignment-after-trimming category refers to reads spanning the ligation junction, which were not aligned in the first step and are trimmed at the ligation site to then realign their 5' extremity to the genome[28] (**Figure 3B,C** and **Table 2**). The contact statistics show that the Hi-C library was of high quality with 82.2% valid pairs and 7.6% non-useful reads falling into the same-fragment self-circle, same-fragment dangling-ends, re-ligation, filtered pairs, and dumped pairs categories (**Figure 3A**, **Figure 3D**, and **Table 2**). Moreover, the number of PCR duplicates is very low, indicating that the library complexity is high, and that the PCR cycles introduced minimal artifacts (**Figure 3E** and **Table 2**).

Using the unique valid Hi-C pairs, basic analysis of the pair distribution was performed using HiCPro[27]. This experiment yielded 46.5% unique cis contacts ≤ 20 kbp, 47.1% unique cis contacts > 20 kbp, and 5.8% unique trans contacts (**Figure 3D**). The distribution of cis to trans valid pairs corresponded to the results expected for a successful Hi-C experiment with most of the interactions detected within the same chromosome. A high proportion of trans contacts indicates inefficient fixation. Using the Hi-C valid pairs from HiCPro[27], matrices were normalized at 1 kb by iterative correction and eigenvector decomposition (ICE), and 5 kb bins were made generated using the hicPlotMatrix tool in HiCExplorer[30-32]. Normalized contact matrices at 1 kb and 5 kb resolution are presented for the *Notch* gene locus in Drosophila (**Figure 4A**, **Figure 4C**, and **Figure 4D**). In **Figure 4A**, the *Notch* gene locus can be seen along with the APs, domain I and II, as well as histone modifications along the locus (**Figure 4A** and **Table 1**). The design of the CRISPR-Cas9 deletion involved the motif of CTCF and M1BP (**Figure 4B**).

Upon deletion of the region containing both CTCF and M1BP DNA-binding sites at the 5' boundary of the *Notch* locus (5pN-delta343, **Table 1**), a dramatic change in chromatin contacts can be observed with loss of interactions inside the *Notch* locus and gain of contacts with the upstream TAD compared to the wild type (WT) (**Figure 4C,D**). Finally, **Figure 4E** shows a detailed panorama of WT and mutant interaction profiles at the restriction fragment level from the *Notch* gene 5' UTR, showing a decrease in the proportion of contacts made with the *Notch* gene locus and an increase in contacts with the upstream domain. This virtual 4C view of the *Notch* gene 5' UTR and exon 6 was obtained using the Hi-C other-ends tool available in SeqMonk (SeqMonk (RRID:SCR_001913) http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/. All the results presented in **Figure 4** were obtained by applying the in-nucleus Hi-C protocol in WT and mutant S2R+ Drosophila cells, as described by Arzate-Mejía et al.[25].

FIGURE AND TABLE LEGENDS:

Figure 1: In-nucleus Hi-C digestion and ligation controls. (**A**) Hi-C protocol overview. Cells are cross-linked with formaldehyde, resulting in covalent links between chromatin segments (DNA fragments: pink, purple) and proteins. Chromatin is digested with a restriction enzyme (represented by scissors), in this example, Mbo I. The resulting sticky ends are filled in with nucleotides including a biotinylated dATP (Dark blue circles). DNA is purified, and the biotinylated junctions are enriched using streptavidin-coated magnetic beads (grey circles). Interacting fragments are identified by next-generation, paired-end sequencing. (**B**) Hi-C

digestion and ligation quality controls for two biological replicates (Hi-C 1 and Hi-C 2). Hi-C libraries were resolved on a 1.5% agarose gel. Both digested, D, libraries run as a smear around 600 bp. Ligated samples, Lig, run as a rather tight band larger than 10 kb similar to the undigested UD samples. The differences in signal strength are due to uneven amounts of loaded DNA on the gel. (**C**) Hi-C digestion quantitative control by quantitative polymerase chain reaction for the same two biological replicates as in (**B**) (Hi-C 1 and Hi-C 2) using the cycle threshold values as detailed in the protocol. A successful digestion has ≥ 80% restriction.

Figure 2: In-nucleus Hi-C fill-in and blunt-end ligation controls. (**A**) Fill-in and biotin labeling assessment. A known interaction between fragments located 300 kb apart in chromosome X was used as a control and amplified using the primers indicated with black arrows (see top of the scheme in (**B**), primer 1 (left), primer 2 (right), **Table 1**), generating a 347 bp amplicon. Hi-C ligation products can be distinguished from those produced in a 3C experiment by digestion of the ligation site. Hi-C junctions were digested by Cla I at the original Mbo I site, as this formed upon blunt-end ligation. Hi-C and 3C junctions were digested with Mbo I as the restriction site regenerates upon ligation (left of the gel). In contrast, 3C junctions were not digested by Cla I at the Mbo I site, but only by Mbo I. Compare the digestion profile of the Hi-C and 3C products using Cla I. A 53 bp fragment was obtained by digesting the Hi-C product (due to restriction of the Cla I site formed at the Mbo I site and restriction of a Cla I site already present in the region). This fragment was not observed in the 3C product digestion as the only Cla I site available was the one that was already present in the region. (**B**) After PCR amplification of the Hi-C library using different PCR cycles, the products were run on a 1.5% agarose gel. A smear of 400–1000 bp was expected and observed. The appropriate number cycles for the final amplification PCR should be taken as the number immediately lower than that at which a smear is just visible. (**C**) Final library Cla I digestion. An aliquot of the final library was re-amplified and digested with Cla I. The size reduction of the smear confirmed that a large proportion of the molecules in the library were valid Hi-C pairs. Densitometric analysis of this gel can be performed to obtain a ratio between the UN and D samples, as detailed in the representative results section.

Figure 3: HiC-Pro statistics of the Hi-C library. (**A**) Schematic representation of valid Hi-C pairs and the different types of non-valid pairs that can be produced during the experiment and filtered out by HiCPro[27] (**Table 2**). These include reads falling into contiguous sequences, dangling ends, same-fragment, self-circle, re-ligations, and PCR duplicates. (**B**) Mapping statistics. Reads that failed to align are shown (grey), and both fully aligned reads and reads aligned after trimming are shown in blue and light blue, respectively. These two categories represent the useful reads that are considered in subsequent analyses. (**C**) Pairing statistics. Multi Aligned reads (dark orange) represent reads that are aligned in multiple regions in the genome. Uniquely Aligned (dark blue) reads represent the read pairs that are aligned once in the genome, and singletons (light orange) represent read pairs in which just one genomic region was sequenced in both reads. (**D**) Filtering statistics. Valid read pairs (blue) represent successful Hi-C ligation products as described in (**A**). Self-fragment self-circles (light pink) are non-useful reads as they represent the same genomic fragment shown in (**A**). Same-fragment dangling ends (orange) represent reads in which a single restriction fragment was sequenced.

Filtered and dumped pairs (brown) are also non-useful reads that have the wrong size or for which the ligation product could not be reconstructed. Finally, re-ligation reads (red) represent reads in which two adjacent fragments were re-ligated, thus producing non-useful information. (**E**) Valid read pairs contact distribution in the genome. Unique cis contacts (blue) are more frequent than unique trans contacts (green).

Figure 4: Hi-C contact matrices and virtual 4C analysis of WT and mutant S2R+ cells. (**A**) Hi-C normalized heatmap of a 50 kb region at 1-kb resolution centered in the *Notch* gene locus. TAD separation score[32] for the locus is shown, along with the partitioning of the *Notch* locus into two topological domains (Domain 1 and Domain 2). ChIP-seq data for APs, RNA Pol II REF, and histone marks for S2/S2R+ cells[34-37] are shown below the heatmap (**Table 1**). The positions of the *Notch* domain 1 boundaries are highlighted in light green. (**B**) Schematic representation of B1 boundary CRISPR mutant. The green rectangle indicates the deleted 343 bp region. Scissors indicate sgRNAs used for CRISPR-mediated genome editing. Motif-binding sites for APs are shown as boxes for CTCF and M1BP. Peak summits for DNA-binding APs shown in (**A**) are also indicated[35]. (**C**) Hi-C normalized heatmaps at 1 kb resolution covering a 50 kb region centered in *Notch* for the WT and the mutant cells. Left, Hi-C heatmaps of the log2 differences in interaction frequency between WT and mutant cells. (**D**) Hi-C normalized heatmaps covering a 250 kb region centered in *Notch* at 5 kb resolution for WT and mutant cells. Left, Hi-C heatmaps of the log2 differences in interaction frequency between WT and mutant cells. (**E**) Virtual-4C for WT and mutant cells using the 5' UTR of *Notch* as viewpoint. The percentages of interactions between the viewpoint and regions within the upstream *kirre domain-2*, *Notch* domain 1, *Notch* domain 2, and the downstream *dnc* domain for both WT and mutant cells are shown[25]. Abbreviations: WT = wild type; TAD = topologically associated domain; ChIP = chromatin immunoprecipitation; AP = architectural protein; RNA Pol = RNA polymerase; S2R+ = S2 receptor plus; sg RNA = single guide RNA; UTR = untranslated region.

Table 1: GEO accession numbers.

Table 2: HiC-Pro statistics.

DISCUSSION:

The in-nucleus Hi-C method presented here has allowed detailed exploration of Drosophila genome topology at high resolution, providing a view of genomic interactions at different genomic scales—from chromatin loops between regulatory elements, such as promoters, enhancers, and TADs, to large chromatin compartment identification[25]. The same technology has also been efficiently applied to mammalian tissues with some modifications[33]. For example, when processing a tissue instead of a single-cell suspension, the tissue is sieved through a 70 µm filter, and the lysis step is performed while homogenizing the material using a Dounce homogenizer. In addition, as the mammalian genome is 25x larger than the Drosophila genome, the number of valid read pairs needed to build 1–5 kb resolution matrices is greater. The in-nucleus Hi-C method differs from the original Hi-C method[2] in its avoidance of nuclear lysis with 1% SDS at 65 °C prior to ligation, thus preserving the nuclear integrity, and by ligating in 1 mL instead of 7 mL[23,24].

The protocol has some key steps to assure high efficiency. The first step that can introduce digestion and fill-in inefficiencies is the formation of clumps during 0.3% SDS permeabilization and surfactant treatments. If the clumps are large and difficult to disrupt, the rotating speed should be decreased to 400 rpm during SDS and surfactant treatments. If the clumps remain difficult to disaggregate by pipetting, the sample should be split in two by adjusting the volumes with restriction buffer, SDS, and the non-ionic surfactant before proceeding with the permeabilization. Next, the nuclei should be centrifuged at minimum speed (200 × *g*), the supernatant carefully discarded, and the samples pooled together in 450 µL of 1x restriction buffer before proceeding with digestion. Second, the estimation of digestion efficiency is important to provide enough DNA fragments for fill-in and ligation. If upon qualitative assessment, the digestion is found to be inefficient, a second round of digestion should be performed with the restriction enzyme for a period ranging from 4 h to overnight.

Third, the estimation of ligation efficiency is important. If upon qualitative assessment, the ligation is found to be inefficient (i.e., instead of the high molecular weight band, a smear similar to that observed for the digested sample is observed), the ligation should be repeated by centrifuging the nuclei at 200 × *g* and resuspending them in ligation mix using fresh 10x ligation buffer and ligase. Fourth, the percentage of Hi-C valid products should be estimated by digesting a PCR amplicon of an expected interaction with Cla I (for Mbo I original digestion). The efficient amplification and digestion of the amplicon of the expected interaction confirms successful ligation and formation of Hi-C junctions. If amplicon digestion is not efficient, the majority of the molecules will be 3C instead of Hi-C products, and this should be taken into consideration if the library will be sequenced. This can also be confirmed by performing the final library Cla I digestion control, as described in the representative results section. Finally, selection of the lower number of PCR cycles is important to avoid PCR duplicates. If upon sequencing, the percentage of read pair duplicates is found to be high, the number of PCR cycles should be decreased further.

This in-nucleus Hi-C technique has some limitations. First, the protocol described here describes the Hi-C experiment being performed for a cell population. Therefore, the signal of the frequency of genomic contacts represents millions of genomes with variable individual conformations. To obtain the set of genomic contacts from a single genome, a single-cell Hi-C experiment[22] is recommended. Second, Hi-C is based on the ligation of proximal DNA fragments. Thus, if genomic regions are part of a large protein-chromatin complex, the distance between fragments could impede ligation. For example, it has been shown that trans contacts are poorly represented in Hi-C[38]. Moreover, Hi-C finishes with paired-end sequencing, thus retrieving pairs of genomic contacts. However, several DNA fragments can simultaneously interact in the same chromatin complex. To obtain the identity of multiple DNA fragments in a chromatin complex, alternative sequencing methods can be applied to Hi-C[39], or different experimental strategies can be employed in which ligation is not performed[38,40,41]. Finally, although Hi-C measures genomic contacts, it does not reveal the identity of the proteins mediating the interactions. Alternative methods have to be applied to identify the genomic interactions mediated by a particular protein of interest[42] or the identity of the ensemble of

proteins at specific genomic elements[43].

In conclusion, with a high quality Hi-C experiment as the one described here for the Drosophila genome (**Table 2**), matrices can be built at a wide range of resolutions (from 1, 5 kb, 50 kb or lower; see **Figure 4**). Additionally, if a particular region of the genome has to be evaluated at the restriction fragment level, the data can be used to build a virtual 4C landscape of the desired viewpoint (e.g., the *Notch* gene 5' UTR in **Figure 4E**). The Hi-C other-ends tool in SeqMonk is a very user-friendly option that enables the visualization of this landscape. Applying the 4C quantification tool, also a part of SeqMonk, to this landscape can yield statistically significant contacts.

Applying the in-nucleus Hi-C experiment described here to a collection of mutant cell lines with altered AP DNA-binding sites at the TAD boundaries (**Figure 4**) revealed that genetic elements are needed at the boundaries to structure the Drosophila genome in domains and sustain gene expression regulation as fully discussed by Arzate-Mejía et al.[25]. Thus, genetic editing of regulatory elements with the CRISPR/Cas9 system, combined with high-resolution profiling of genomic interactions using the in-nucleus Hi-C protocol described here, can be a powerful strategy to test the structural function of genetic elements.

ACKNOWLEDGMENTS:

This work was supported by UNAM Technology Innovation and Research Support Program (PAPIIT) grant number IN207319 and the Science and Technology National Council (CONACyT-FORDECyT) grant number 303068. A.E.-L. is a master's student supported by the Science and Technology National Council (CONACyT) CVU number 968128.

DISCLOSURES

The authors declare no competing interests.

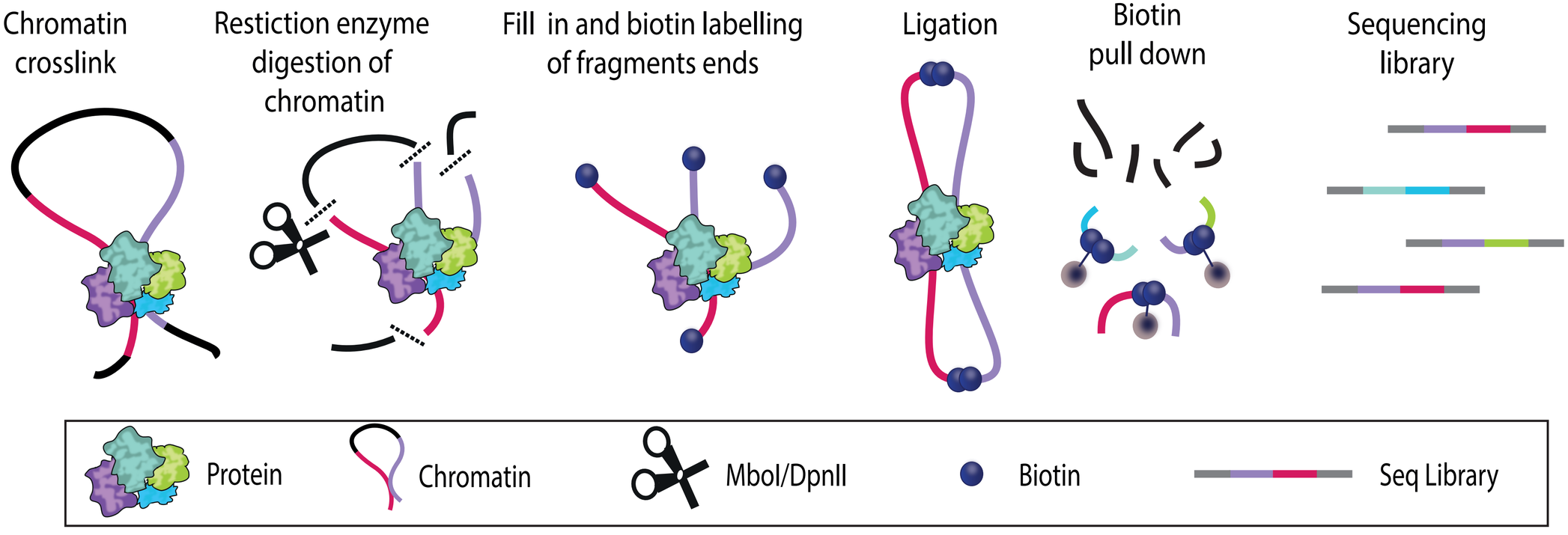
REFERENCES:

1. Cremer, T., Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Review Genetics*. **2**, 292–301 (2001).
2. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
3. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. **485**, 376–380 (2012).
4. Sexton, T. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. **148** (3), 458–472. (2012).
5. Dixon, J. R., Gorkin, D. U., Ren, B. Chromatin domains: the unit of chromosome organization. *Molecular Cell*. **62**, 668–680 (2016).
6. Bonev, B., Cavalli, G. Organization and function of the 3D genome. *Nature Reviews Genetics*. **17**, 661–678 (2016).
7. Lupiáñez, D. G., Spielmann, M., Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends in Genetics*. **32**, 225–237 (2016).
8. Phillips, J. E., Corces, V. G. CTCF: master weaver of the genome. *Cell*. **137** (7), 1194–1211

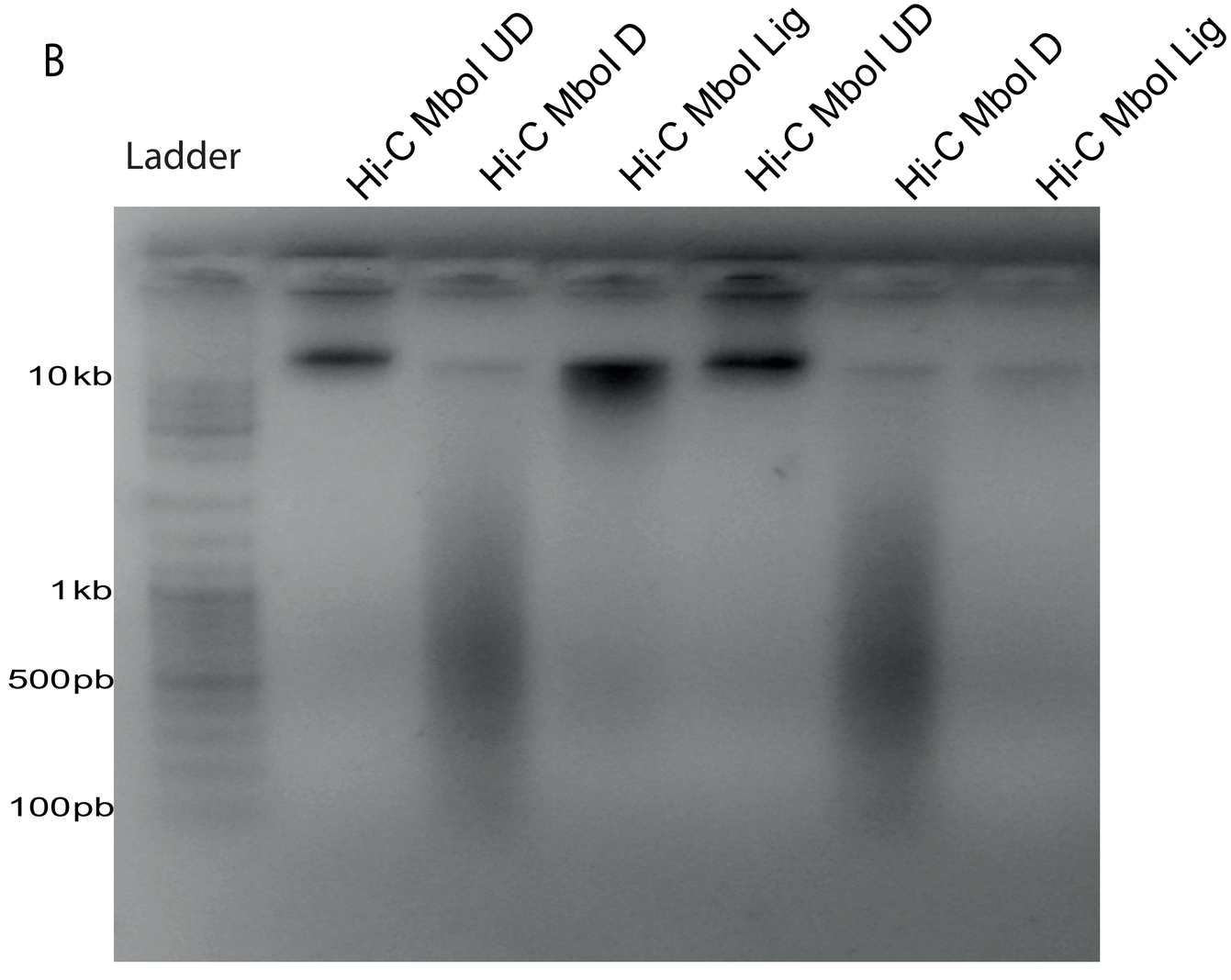
(2009).
9. Hong, S., Kim, D. Computational characterization of chromatin domain boundary-associated genomic elements. *Nucleic Acids Research*. **45**, 10403–10414 (2017).
10. Zuin, J., et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America*. **111** (3), 996–1001 (2014).
11. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*. **162**, 900–910 (2015).
12. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. **161**, 1012–1025 (2015).
13. Van-Steensel, B., Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*. **20**, 327–337 (2019).
14. Merkenschlager, M., Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annual Review of Genomics and Human Genetics*. **17** (1), 17–43 (2016).
15. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*. **6**, 1–10 (2017).
16. Van Bortle, K. et al. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biology*. **15**, R82 (2014).
17. Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications*. **9**, 189 (2018).
18. Ulianov, S. V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Research*. **26**, 70–84 (2016).
19. Rowley, M. J., et al. Evolutionarily conserved principles predict 3D chromatin organization. *Molecular Cell*. **67**, 837–852.e7 (2017).
20. Gavrilov, A. A., Golov, A. K., Razin, S. V. Actual ligation frequencies in the chromosome conformation capture procedure. *PLoS One*. **8**, e60403 (2013).
21. Gavrilov, A. A. et al. Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Research*. **41**, 3563–3575 (2013).
22. Nagano, T., et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. **502**, 59–64 (2013).
23. Rao, S. S. P., et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. **159** (7), 1665–1680 (2014).
24. Nagano, T., et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biology*. **16**, 175 (2015).
25. Arzate-Mejía, R. et al. In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila. *Nature Communications*. **11**, 894 (2020).
26. Schoenfelder, S. et al. Promoter capture Hi-C: high-resolution, genome-wide profiling of promoter interactions. *Journal of Visualized Experiments*. **136**, 57320 (2018).
27. Servant N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C processing. *Genome Biology*. **16**, 259 (2015).
28. Philip, E., Måns, M., Sverker, L., Max, K. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. **10**, 1093 (2016).
29. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000 Research*. **4**, 1310 (2015).

30. Joachim, W. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Research*. **48** (W1), W177–W184 (2020).
31. Joachim, W. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Research*. **46** (W1), W11–W16 (2018).
32. Ramirez, F., et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications*. **9**, 189 (2018).
33. Ando-Kuri, M.et al. The global and promoter-centric 3D genome organization temporally resolved during a circadian cycle. *bioRxiv*. 2020.07.23.217992 (2020).
34. Cuellar-Partida, G. et al. Epigenetic priors for identifying active transcription factor binding sites. *Bioinformatics*. **28** (1), 56–62 (2012).
35. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*. **9**, R137 (2008).
36. Ong, C. -T. et al. Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in Drosophila. *Cell*. **155** (1), 148–159 (2013).
37. Fresán, U. et al. The insulator protein CTCF regulates Drosophila steroidogenesis. *Biology Open*. **4** (7), 852–857 (2015).
38. Quinodoz, S. et al. RNA promotes the formation of spatial compartments in the nucleus. *Cell*. **174**, 744–757.e24 (2018).
39. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature*. **540** (7632), 296–300 (2016).
40. Beagrie, R. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*. **543**, 519–524 (2017).
41. Redolfi, J. et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nature Structural and Molecular Biology*. **26**, 471–480 (2019).
42. Maxwell, R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods*. **13**, 919–922 (2016).
43. Gao, X. et al. C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nature Methods*. **15** (6), 433–436 (2018).

A



B



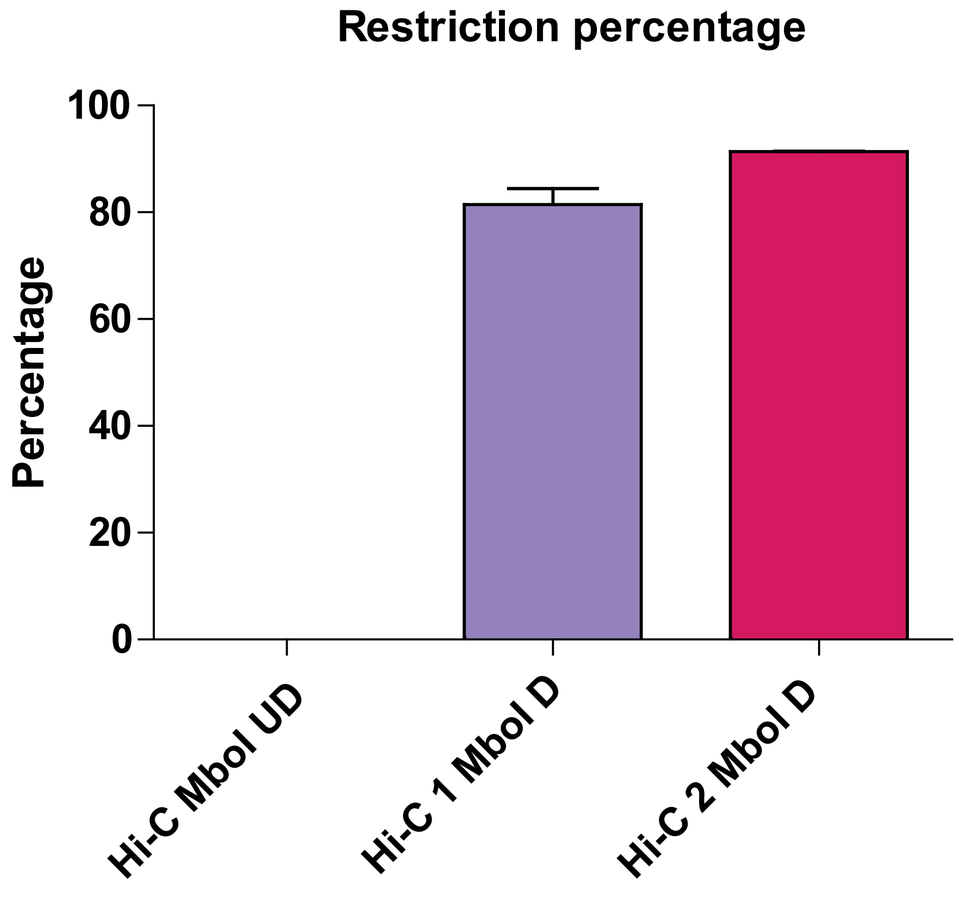
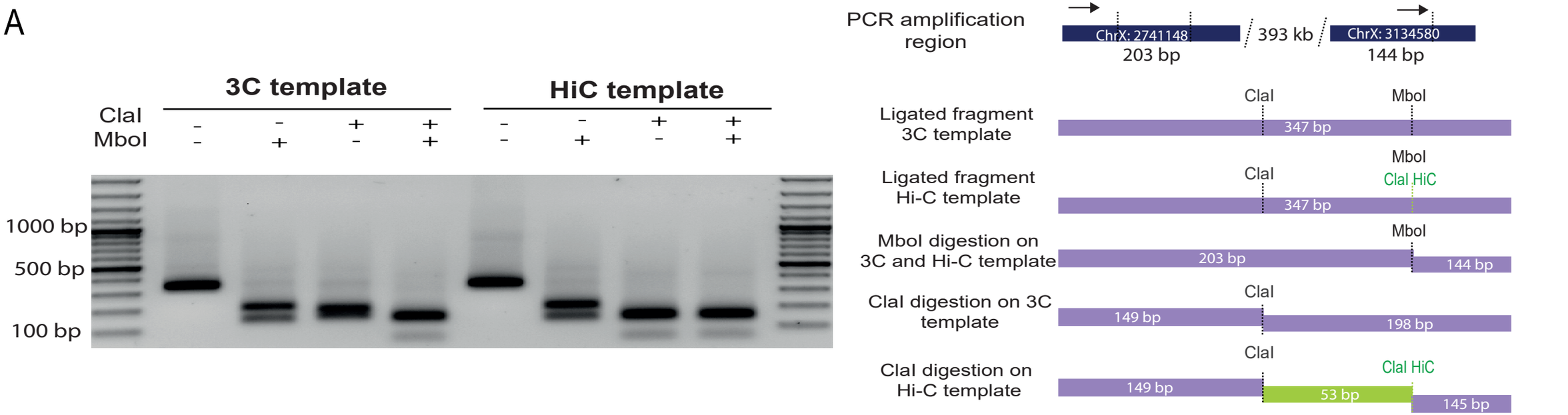
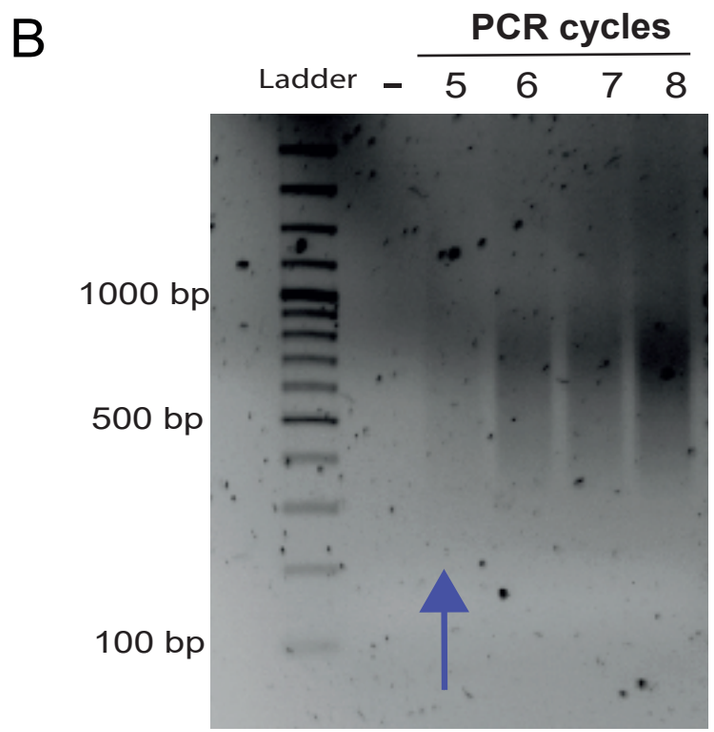


Figure 2





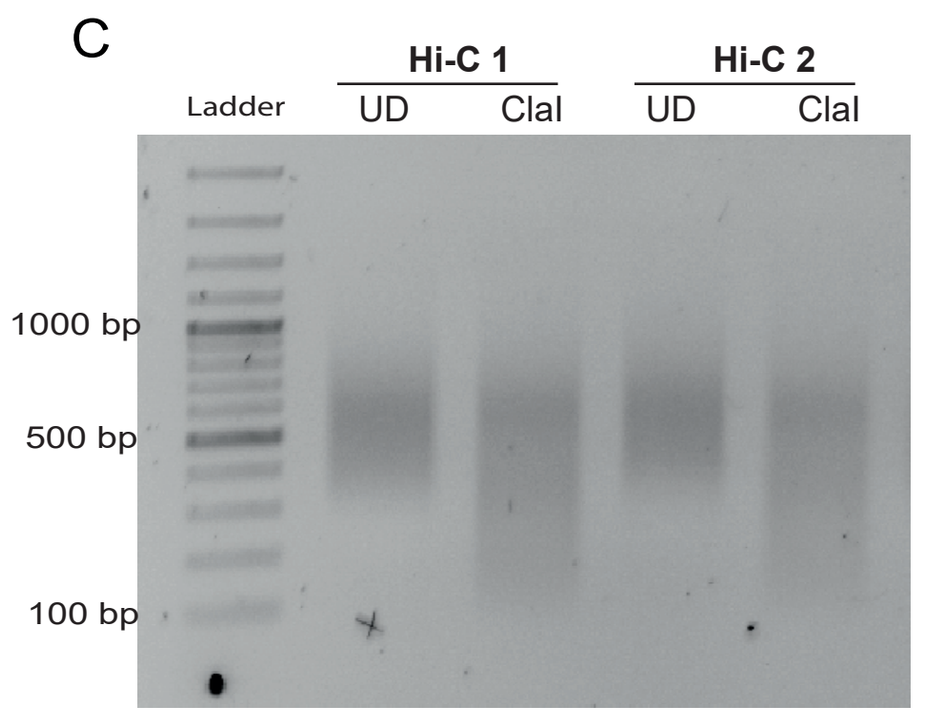


Figure 3

Click here to access/download;Figure;figure_3_aael_vers4.pdf

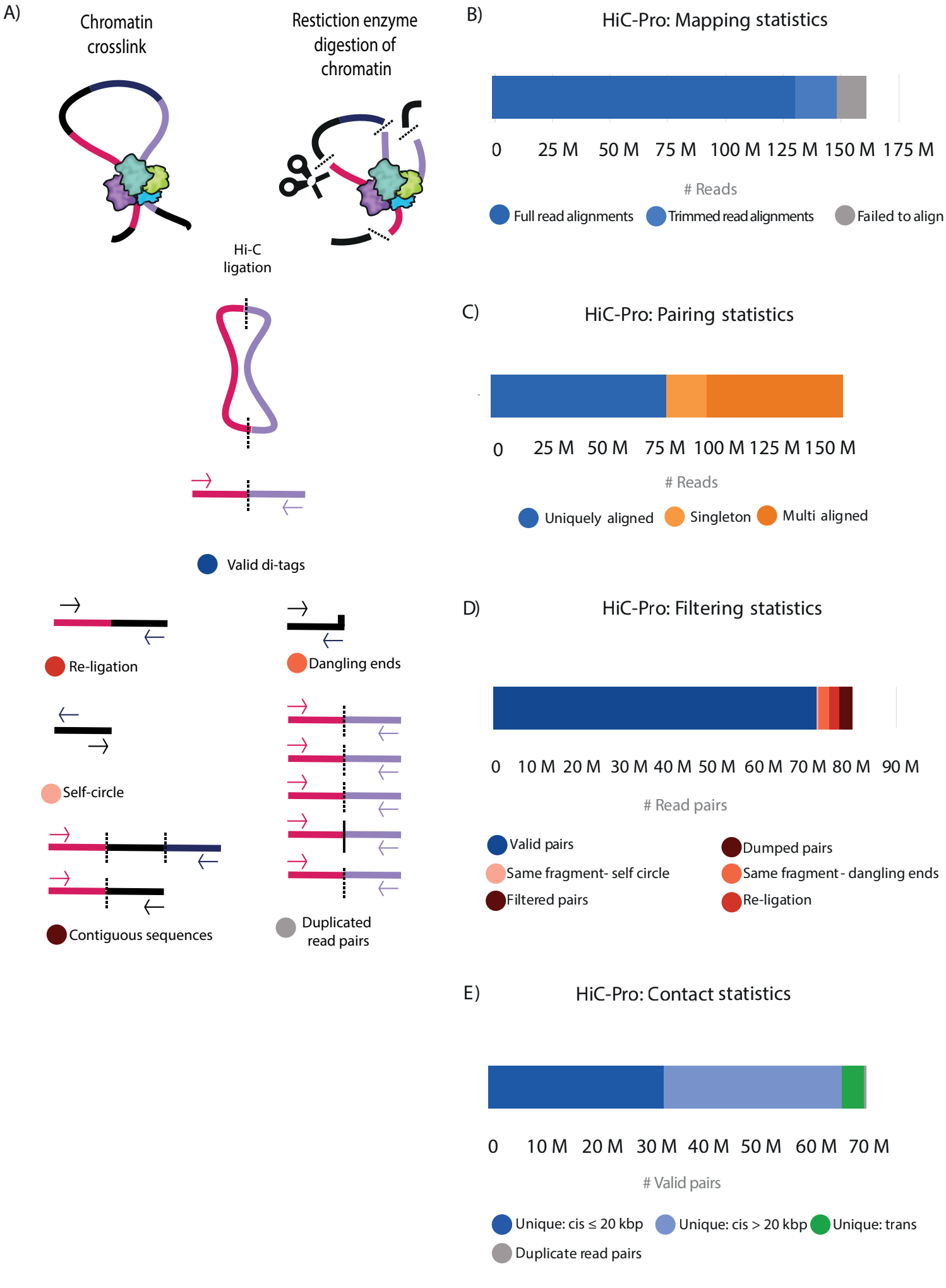


Figure 4

Click here to access/download;Figure;Figure_4_ael_vers3.pdf

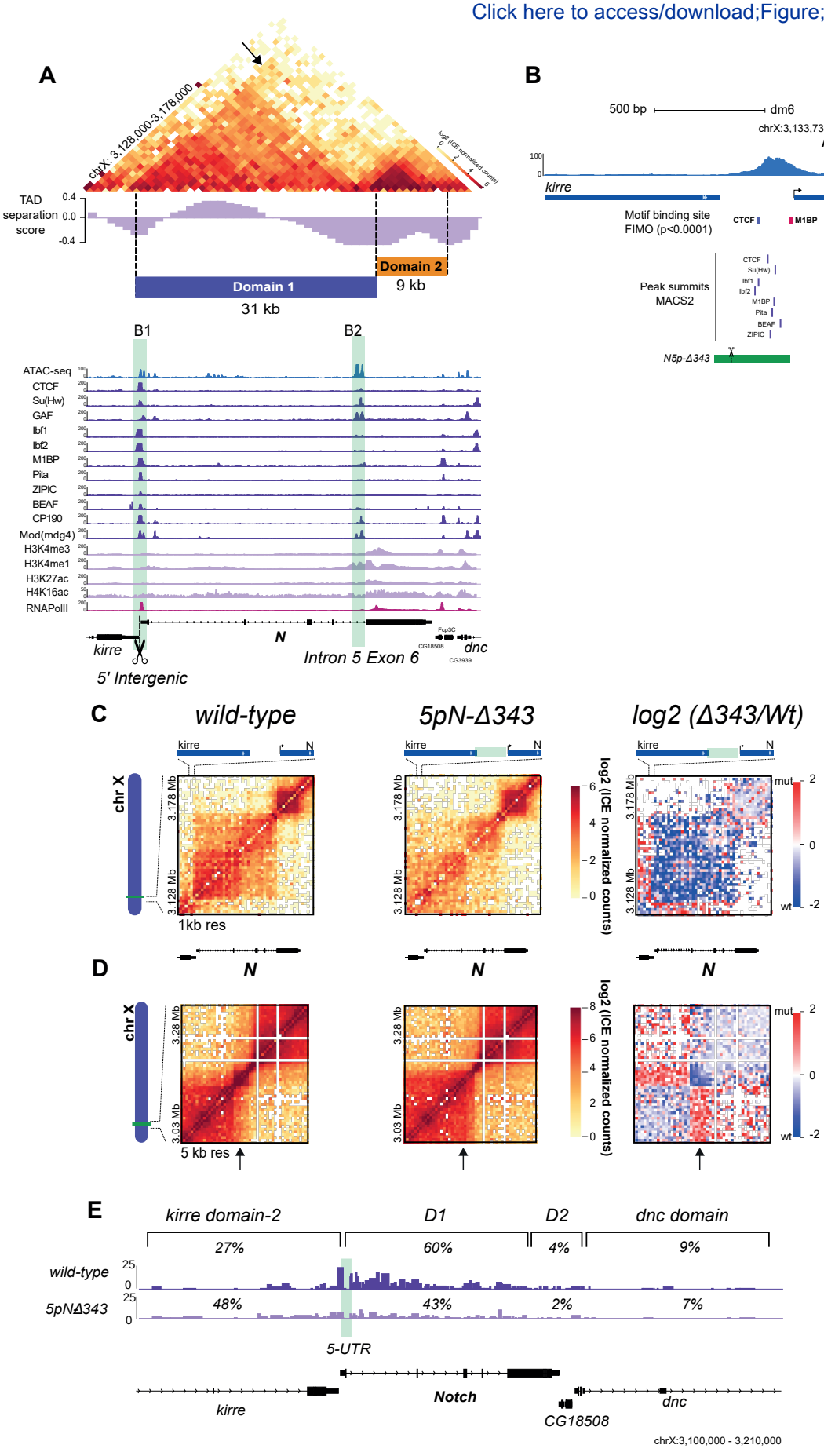


Table 1

Click here to access/download;Table;Table_1_GEO_accession_numbers.xlsx

Experiment	Sample	GEO Accession number
ChIP	CP190	GSM1015404
ChIP	SuHW	GSM1015406
ChIP	Mod(mdg4)	GSM1015408
ChIP	CTCF	GSM1015410
ChIP	Ibf1	GSM1133264
ChIP	Ibf2	GSM1133265
ChIP	BEAF32	GSM1278639
ChIP	Pita	GSM1313420
ChIP	ZIPIC	GSM1313421
ChIP	RNA PolII	GSM2259975
ChIP	H3K4me1	GSM2259983
ChIP	H3K4me3	GSM2259985
ChIP	H3K27ac	GSM2259987
ChIP	MSL2	GSM2469507
ChIP	H4K16ac	GSM2469508
ChIP	M1BP	GSM2706055
ChIP	GAF	GSM2860390
ChIP	Input	GSM1015412
Hi-C	S2R+ WT cells	GSE136137
Hi-C	S2R+ 5pN-delta343 cells	GSE136137

Table 2

Click here to access/download;Table;Table_2_HiCPro_statistics (2).xlsx

HiC-Pro statistics	Reads	Percentage
Mapping Statistics		
Full read Alignments	131515921	82.20%
Trimmed read Alignments	16408309	10.30%
Failed to align	12110964	7.60%
Pairing Statistics		
Uniquely Aligned	79428455	50.90%
Singleton	19063418	12.20%
Multi Aligned	57700021	36.90%
Filtering Statistics		
Valid Pairs	71373989	90.12%
Same-Fragment: Self-circle	2340697	2.90%
Same-Fragment: Dangling Ends	2578783	3.20%
Filtered Pairs	2773043	3.50%
Dumped pairs	196565	0.20%
Contact Statistics		
Unique: cis ≤ 20 kbp	33108815	46.50%
Unique: cis > 20 kbp	33539888	47.10%
Unique: trans	4133602	5.80%
Duplicate read pairs	398400	0.60%

Table of Materials

Click here to access/download;Table of Materials;Table_of_materials.xlsx

Materials	Company
16% (vol/vol) paraformaldehyde solution	Agar Scientific
Biotin-14-dATP	Invitrogen
ClaI enzyme	NEB
COVARIS Ultrasonicator	Covaris
Cut Smart	NEB
Dulbecco's Modified Eagle Medium (DMEM) 1x	Life Technologies
Dynabeads MyOne Streptabidin C1	Invitrogen
Fetal bovine serum (FBS) sterile filtered	Sigma
Klenow Dna PolI large fragment	NEB
Klenow exo(-)	NEB
Ligation Buffer	NEB
MboI enzyme	NEB
NP40-Igepal	SIGMA
PE adapter 1.0	Illumina
PE adapter 2.0	Illumina
PE PCR primer 1.0	Illumina
PE PCR primer 2.0	Illumina
Phenol: Chloroform:Isoamyl Alcohol 25:24:1	SIGMA
Primer 1 (known interaction, **Figure 2A**)	Sigma
Primer 2 (known interactions, **Figure 2A**)	Sigma
Protease inhibitor cocktail tablet	Roche
Proteinase K	Roche
Qubit	ThermoFisher
RNAse	Roche
SPRI Beads	Beckman

T4 DNA ligase	Invitrogen
T4 DNA polymerase	NEB
T4 polynucleotide kinase (PNK)	NEB
TaqPhusion	NEB

Triton X-100

Catalog number	Comments
R1026	
CA1524-016	
R0197S	
LE220-M220	
B72002S	
41965-039	
65002	
F9665	
M0210L	
M0210S	
B020S	
R0147M	
CA-420	Non-ionic surfactant for addition in lysis buffer
5'-P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'	
5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'	
5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'	
P2069	
5'-TCGCGGTAATTTTGCGTTTGA-3'	
5'-CCTCCCTGCCAAAACGTTTT-3'	
4693132001	
3115879001	
Q33327	
10109142001	
B23318	

15224-025	
M0203S	
M0201L	
M0530S	DNA polymerase
	Non-ionic surfactant for quenching of SDS

Dear Vineeta Bajaj, Ph.D.
Review Editor
JoVE

Thank you for your interest in our manuscript JoVE62106 entitled **"In nucleus Hi-C in Drosophila cells"** and the opportunity to revise it and address the reviewers concerns.

Please find a point-by-point response to the editorial office and reviewer's comments below.

Editorial comments:

Changes to be made by the Author(s):

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. Please define all abbreviations at first use.

We have checked spelling and grammar issues and defined abbreviations at fist use.

2. Please provide an email address for each author.

We have provided the email address for each author.

3. Please refer to the instructions for authors and JoVE's style guide to organize your manuscript in this order: Title, author names and affiliations (with email addresses)+corresponding author information; Summary; Abstract; Introduction; Protocol; Representative Results, Figure and Table Legends; Discussion; Acknowledgments; Disclosure; References; Figures (uploaded individually without legends); Tables (uploaded individually without legends); Table of Materials if applicable; Supplementary material

We have organized the manuscript according to JoVE's style guide.

4. Please revise the following lines to avoid overlap with previously published work: 197-204

We have modified the text in the manuscript to avoid overlap with previously published work as follows and also cited previous work:

"*Fill-in and biotin labeling control*

Verify Hi-C marking and ligation efficiency by amplifying a known interaction or a ligation product between adjacent restriction fragments in the genome. Successful fill-in and ligation of a Mbo I site (GATC) generates a new site for the restriction enzyme Cla I (ATCGAT) at the ligation junction and regenerates the Mbo I site. Digest the PCR product with MboI, ClaI or both. After running the samples on a 1.5-2% gel, the relative number of 3C and Hi-C ligation junctions can be estimated by quantifying the intensity of the cut and uncut bands[27]. An efficiency above 70% is desired (see representative results)"

5. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. For example: Qubit; COVARIS (LE220-M220);

We have now remove all commercial products from the manuscript and included them in Table 1.

6. Please revise the text, especially in the protocol, to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).

We have revised the text and avoid the use of personal pronouns

7. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “Note.” However, notes should be concise and used sparingly. Please include all safety procedures and use of hoods, etc.

We have avoided phrases such as “should be” “could be” and revised that all actions are in the imperative tense. We have included safety procedures where needed.

8. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Please add more details to your protocol steps. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action (e.g., digestion and ligation qualitative controls section). Please add more specific details (e.g., button clicks for software actions, numerical values for settings, etc) to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol.

We have now described in more detail the steps needed to perform the “digestion and ligation quantitative control” and added references to previous material when needed, for instance in lines 225-262 describing the Fill-in and biotin-labelling control. We have also added more detailed explanations in the Figure legends. All numerical values for centrifugation and rotation steps as well as temperatures are presented. We have also modified the sections highlighted in yellow to try to make a more comprehensive script for the video.

9. Please consider providing reaction set-ups and solution composition as Tables in separate .xls or .xlsx files uploaded to your Editorial Manager account. These tables can then be referenced in the protocol text.

We have incorporated some of the solution compositions into the text as we have done for other steps in the protocol. We find this is actually more useful for people following the protocol than having to go to an additional table file if the list of reagents needed to make the solutions is short. We have added an additional Table (Table 3) with GEO accession numbers for the data sets used in the manuscript.

10. Please include a one line space between each protocol step and then highlight up to 3 pages of protocol text for inclusion in the protocol section of the video.

Done

11. As we are a methods journal, please add to the Discussion the following in detail with citations:

a) Any modifications and troubleshooting of the technique

b) Any limitations of the technique

We have now complemented the Discussion section of the manuscript as follows:

“*Trouble shooting for key steps of the protocol*

During the protocol there are key steps to assure high efficiency as presented. The first step that can introduce digestion and fill-in inefficiencies is the formation of clumps during 0.3% SDS permeabilization and triton treatments. As specified in NOTE 2. If clumps are large and hard to disrupt decrease the rotating speed to 400rpm during SDS and triton treatments. If still the clumps are hard to disaggregate by pipetting, split the sample in two adjust the volumes of restriction buffer, SDS and triton and proceed with the permeabilization. Next, spin the nuclei at minimum speed (200 x g) carefully discard supernatant and pool the samples together in 450 µl of 1 X NEB2 buffer. Then carry on with digestion. Second and linked to the previous point, digestion efficiency estimation is important to provide enough DNA fragments for fill-in and ligation. If upon qualitative assessment the digestion was inefficient perform a second round of digestion with the restriction enzyme for 4 hrs to overnight. Third, estimation of ligation efficiency is recommended. If upon qualitative assessment the ligation was inefficient (i.e. the high molecular weight band is not recovered but a smear similar to the digested sample is observed), repeat the ligation step by centrifuging the nuclei at 200 x g and resuspending in ligation mix using fresh 10 X ligation buffer and ligase. Fourth, estimation of the percentage of Hi-C valid products is recommended by digesting a PCR amplicon of an expected interaction with Cla I (for Mbo I original digestion). The efficient amplification and digestion of the amplicon of the expected interaction corroborates successful ligation and formation of Hi-C junctions. If amplicon digestion is not efficient the majority of the molecules are 3C instead of Hi-C products and this should be taken into consideration if the library will be sequenced. This can also be confirmed performing the final library Cla I digestion control as described in the "representative results" section. Finally, selection of the less number of PCR cycles as possible is important to avoid PCR duplicates. If upon sequencing the percentage of read pair duplicates is high, consider decreasing the PCR cycles even more as described during the protocol based on the PCR cycle test.

Limitations of the Hi-C experiment

Some of the limitations of the in nucleus Hi-C technique are: 1. The protocol described here corresponds to the Hi-C experiment performed in a cell population. Therefore, the signal of the frequency of genomic contacts obtained represents what is observed for millions of genomes whit variable individual conformations. To obtain the set of genomic contacts from a single genome, the single cell Hi-C experiment[23] is recommended. 2. Hi-C is based on ligation of proximal DNA fragment. Thus, if genomic regions are part of a large protein-chromatin complex the distance between fragments could impede ligation. For example, it has been shown that trans contacts are poorly represented in Hi-C[39]. Also, Hi-C finishes with paired-end sequencing thus retrieving pairs of genomic contacts. However, several DNA fragments can be simultaneously interacting in the same chromatin complex. To obtain the identity of multiple DNA fragments in a chromatin complex alternative sequencing methods can be applied to Hi-C[40] or, different experimental strategies can be performed in which ligation is not performed[39,41,42]. Finally, Hi-C measures genomic contacts but does not reveal the identity of the proteins mediating the interactions. Alternative methods can be applied

to resolve the genomic interactions mediated by a particular protein of interest[43] or the identity of the ensemble of proteins at particular genomic elements[44].”

12. Please ensure that the references appear as the following: [Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage–LastPage (YEAR).] For more than 6 authors, list only the first author then et al. Please include volume and issue numbers for all references and do not abbreviate journal names.

Done

13. Please remove the embedded Table from the manuscript. All tables should be uploaded separately to your Editorial Manager account in the form of an .xls or .xlsx file. The table legend or caption (title and description) should appear in the Figure and Table Legends section after the Representative Results in the manuscript text.

We have incorporated some of the solution compositions into the text as we have done for other steps in the protocol. The tables provided correspond to the Table of Materials (Table 1), The HiC-Pro statistics (Table 2) and the GEO accession numbers for the Hi-C data and ChIP data presented (Table 3).

Reviewer #1:

Manuscript Summary:

Esquivel-López et al. described a protocol of Hi-C experiment based on cell population in nucleus and applied it on Drosophila to verify the consequences of removing DNA binding sites of CTCF and M1BP at the Notch gene locus. They described their method in a detail way that will be helpful to other scientists in this field.

1. We thank the reviewer for his/her comment.

Major Concerns:

In the introduction (lines 72-76), the authors introduced their finding for the genetically removing of the binding sites of two insulator proteins (CTCF and M1BP) would result in dramatic change of TADs thus highlight the genetic elements are significant in 3D genome and expression. But, I don't see this result in their manuscript. In Figure4, only two Hi-C contact maps and a track described the gene structure TAD annotation and a bar plot of hic contact frequency. If you say a change here, you should provide the background unchanged status for your comparison. A status for normal CTCF and M1BP binding sites and the other is the result after removing these binding sites. Additionally, CTCF does not suggested as a direct insulator protein for TAD formation in Drosophila, the author should discuss carefully here. And the author should also clarify whether the two biding sites are overlap with each other or they are distinct sites when they do genetic removing.

2. We thank the reviewer for his/her comments. Indeed, as the reviewer is mentioning this was indeed very confusing, as we were not showing results related to the effects of removing binding sites for CTCF and M1BP at a TAD boundary of the *Notch* gene locus in Drosophila (characterized in[26] Arzate-Mejía et al., Nat Comm 2020) in figure 4, but instead we were displaying examples of a different region altogether, just as examples

of the Hi-C matrices that can be built from the data generated with the in nucleus Hi-C protocol.

To be more coherent and consistent with our text and as suggested by the reviewer´s helpful comments we have now modified Figure 4. We have included the Hi-C contact map of the *Notch* locus at different resolutions (Figure 4a,c and d), indicated where the removed binding sites are located at the *Notch* 5´ TAD boundary (Figure 4 b), and show the effect of this genetic deletion in the 3D architecture of the locus at 1kb, 5kb and fragment resolution levels (Figure 4b, c and e).

We have modified the results section of the text as follows to fully describe and explain the new figure 4:

"Using the Hi-C valid pairs from HiCPro[28], ICE normalized matrices at 1 kb, 5 kb bin resolutions were made using hicPlotMatrix tool in HiCExplorer [31-33]. An example of normalized contact matrix at 1kb and 5kb resolution are presented for the *Notch* gene locus in Drosophila (Figure 4A, C and D). In figure 4A the *Notch* gene locus can be seen in detail together with the Aps, Pl II and histone modifications occupancy along the locus (Figure 4A, Table 3). The design of the CRISPR-Cas9 deletion shown involved de motif of CTCF and M1BP as detailed in figure 4B. Upon deletion of the region containing both CTCF and M1BP DNA binding sites at the 5' boundary of the *Notch* locus (5pN-delta343, Table 3), a dramatic change in chromatin contacts can be observed with loss of interactions inside the *Notch* locus and gain of contacts with the upstream TAD compared to the WT (Figure 4C, D). Finally a detailed panorama of WT and mutant interaction profiles at the restriction fragment level from *Notch* gene 5'UTR is presented again, showing a decrease in the proportion of contacts made with the *Notch* gene locus and an increase of contacts with the upstream domain (Figure 4E). This virtual 4C from the *Notch* gene 5'UTR and exon 6 was obtained using the Hi-C other-ends tool implemented on SeqMonk (SeqMonk (RRID:SCR_001913) http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/. All of the results presented in Figure 4 were obtained applying the in nucleus Hi-C protocol presented here in WT and mutant S2R+ Drosophila cells as described in detail in[26]."

We have also modified the discussion section as follows:
"With a high quality Hi-C experiment as the one described here for the Drosophila genome sequenced at the presented depth (Table 1) one can build matrices at a wide range of resolutions (from 1, 5kbs, 50kb or lower, see Figure 4). Also, if a particular region of the genome wants to be evaluated at the restriction fragment level, the data can be used to build a virtual 4C landscape of the desired viewpoint (for instance the *Notch* gene 5' UTR in Figure 4E). The Hi-C other ends tool in SeqMonk is very a friendly option to do so and statistically significant contacts can be obtained applying the 4C quantification tool also incorporated in SeqMonk.

Applying the in nucleus Hi-C experiment described here to a collection of mutant cell lines with altered AP DNA binding sites at the TAD boundaries (Figure 4), we demonstrated that genetic elements at boundaries are needed to structure the Drosophila genome in domains and sustain gene expression regulation as fully discussed in[26]. Thus, genetic edition of regulatory elements with CRISPR/Cas9 system combined with high-resolution profiling of genomic interactions through the in nucleus Hi-C protocol described here, provide a powerful strategy to test the structural function of genetic elements."

We have modified the Figure 4 legend as well as follows:
"Figure 4. Hi-C contact matrices and virtual 4C analysis of WT and mutant S2R+ cells. A) Hi-C normalized heatmap of a 50 kb region at 1-kb resolution centered in the *Notch* gene locus. TAD separation score[33] for the locus is shown. The *Notch* locus is partitioned into two topological domains (Domain 1 and Domain 2). ChIP-seq data for Architectural Proteins (APs), RNA Pol II REF, and histone marks for S2/S2R+ cells[35-38] is shown below the heatmap (Table 3). The position of the *Notch* domain 1 boundaries, are highlighted in light green. B) Schematic representation of B1 boundary CRISPR mutant. The green rectangle indicates the deleted 343 bp region. Scissors indicate sgRNAs used for CRISPR mediated genome editing. Motif binding sites for APs are shown as boxes for CTCF and M1BP. Peak summits for DNA-binding APs shown in A are also indicated[36] C) Hi–C normalized heatmaps at 1-kb resolution covering a 50-kb region centered in *Notch* for the wilde-type (WT) and the mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells. D) Hi–C normalized heatmaps covering a 250-kb region centered in *Notch* at 5-kb resolution for WT and mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells E) Virtual-4C for WT and mutant cells using the 5' UTR of *Notch* as viewpoint. The percent of interactions between the viewpoint and regions within the upstream *kirre domain-2*, the *Notch* domain 1, the *Notch* domain 2, and the downstream *dnc* domain for both WT and mutant cells are shown[26]."

Finally, as this manuscript is centred on the details and controls of the in nucleus Hi-C protocol instead of the controversy of the function of CTCF in structuring the genome in Drosophila, we have emphasized in the text that this discussion was thoroughly made in[26] Arzate-Mejía et al., Nat Comm, 2020.

Minor Concerns:
Figure 4B. the y axis make sure you plot right , for 7,10 to 20Mb?

3. This concern refers to the original Figure 4

Reviewer #2:
Manuscript Summary:

The manuscript describes a protocol for performing in nucleus Hi-C in Drosophila cells. However, there are several details unclear or not well elaborated.

Starting from the introduction the previous literature is not presented in a clear context. For example
- The manuscript never cited the Sexton et al Cell 2012 article that first reported TADs in Drosophila. Please note the reviewer is NOT an author in that article.
- The protocol starts from Rao et al for in situ Hi-C but this is not cited as the first one proposing "in situ/in nucleus" Hi-C. Please note the reviewer is NOT an author in that article.

1. We thank the reviewer for his/her comment. This was a mistake on our part as, of course, we were supposed to cite these two key papers and thought we have done so. We have now cited them in all appropriate sections of the introduction and discussion.

Major Concerns:
- The abstract and the discussion talk about
"mutant cell lines with altered AP DNA binding sites at the TAD boundaries"
But such results are not presented in the figures. If they are, it's not clear what is the mutant and what is the wild type profile as there are no labels about this in the figures.
Even the methods section does not explain what is the mutation the authors are referring to.

2. We thank the reviewer for his/her comments. The protocol described here is the in nucleus Hi-C experiment to analyse genomic interactions. However to be more coherent and consistent with our text and as suggested by the reviewer´s helpful comments we have now modified Figure 4. We have included the Hi-C contact map of the *Notch* gene locus at different resolutions (Figure 4A, C, D, E), indicated where the APs are located in the locus (Figure 4A) and detailed the locations of the removed binding sites located at the *Notch* 5´ TAD boundary (Figure 4B). Also we show the effect of the genetic elimination of CTCF and MBP1 binding sites in the 3D architecture of the locus (Figure 4C, D and E).

Accordingly we have modified the text incorporating the new Figure 4 as follows:

Representative results section
"Using the Hi-C valid pairs from HiCPro[28], ICE normalized matrices at 1 kb, 5 kb bin resolutions were made using hicPlotMatrix tool in HiCExplorer [31-33]. An example of normalized contact matrix at 1kb and 5kb resolution are presented for the *Notch* gene locus in Drosophila (Figure 4A, C and D). In figure 4A the *Notch* gene locus can be seen in detail together with the Aps, Pl II and histone modifications occupancy along the locus (Figure 4A, Table 3). The design of the CRISPR-Cas9 deletion shown involved de motif of CTCF and M1BP as detailed in figure 4B. Upon deletion of the region containing both CTCF and M1BP DNA binding sites at the 5' boundary of the *Notch* locus (5pN-delta343,

Table 3), a dramatic change in chromatin contacts can be observed with loss of interactions inside the *Notch* locus and gain of contacts with the upstream TAD compared to the WT (Figure 4C, D). Finally a detailed panorama of WT and mutant interaction profiles at the restriction fragment level from *Notch* gene 5'UTR is presented again, showing a decrease in the proportion of contacts made with the *Notch* gene locus and an increase of contacts with the upstream domain (Figure 4E). This virtual 4C from the *Notch* gene 5'UTR and exon 6 was obtained using the Hi-C other-ends tool implemented on SeqMonk (SeqMonk (RRID:SCR_001913) http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/. All of the results presented in Figure 4 were obtained applying the in nucleus Hi-C protocol presented here in WT and mutant S2R+ Drosophila cells as described in detail in[26]."

We have also modified the discussion section as follows:
"With a high quality Hi-C experiment as the one described here for the Drosophila genome sequenced at the presented depth (Table 1) one can build matrices at a wide range of resolutions (from 1, 5kbs, 50kb or lower, see Figure 4). Also, if a particular region of the genome wants to be evaluated at the restriction fragment level, the data can be used to build a virtual 4C landscape of the desired viewpoint (for instance the *Notch* gene 5' UTR in Figure 4E). The Hi-C other ends tool in SeqMonk is very a friendly option to do so and statistically significant contacts can be obtained applying the 4C quantification tool also incorporated in SeqMonk.

Applying the in nucleus Hi-C experiment described here to a collection of mutant cell lines with altered AP DNA binding sites at the TAD boundaries (Figure 4), we demonstrated that genetic elements at boundaries are needed to structure the Drosophila genome in domains and sustain gene expression regulation as fully discussed in[26]. Thus, genetic edition of regulatory elements with CRISPR/Cas9 system combined with high-resolution profiling of genomic interactions through the in nucleus Hi-C protocol described here, provide a powerful strategy to test the structural function of genetic elements."
We have modified the Figure 4 legend as well as follows:
"Figure 4. Hi-C contact matrices and virtual 4C analysis of WT and mutant S2R+ cells. A) Hi-C normalized heatmap of a 50 kb region at 1-kb resolution centered in the *Notch* gene locus. TAD separation score[33] for the locus is shown. The *Notch* locus is partitioned into two topological domains (Domain 1 and Domain 2). ChIP-seq data for Architectural Proteins (APs), RNA Pol II REF, and histone marks for S2/S2R+ cells[35-38] is shown below the heatmap (Table 3). The position of the *Notch* domain 1 boundaries, are highlighted in light green. B) Schematic representation of B1 boundary CRISPR mutant. The green rectangle indicates the deleted 343 bp region. Scissors indicate sgRNAs used for CRISPR mediated genome editing. Motif binding sites for APs are shown as boxes for CTCF and M1BP. Peak summits for DNA-binding APs shown in A are also indicated[36] C) Hi–C normalized heatmaps at 1-kb resolution covering a 50-kb region centered in *Notch* for the wilde-type (WT) and the mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells. D) Hi–C normalized heatmaps

covering a 250-kb region centered in *Notch* at 5-kb resolution for WT and mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells E) Virtual-4C for WT and mutant cells using the 5' UTR of *Notch* as viewpoint. The percent of interactions between the viewpoint and regions within the upstream *kirre domain-2*, the *Notch* domain 1, the *Notch* domain 2, and the downstream *dnc* domain for both WT and mutant cells are shown[26]."

- The authors claim that with the presented dataset they can reach 1Kb resolution, but more than a billion reads were used instead in previous literature to reach that resolution in Drosophila. e.g see:
Wang Q, Sun Q, Czajkowsky DM, Shao Z (2018) Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. Nat Commun 9:331. https://doi.org/10.1038/s41467-017-02526-9

3. Wang et al., generated 353 million valid pair-end reads to build contact maps of ~200 bp resolution in Drosophila as stated in their paper: ...*"Consequently, for all further analysis, we combined both data sets to finally obtain 353 million pair-end reads with a maximal estimated "map resolution" of ~200 bp, as calculated following Rao et al."*
Using the in nucleus Hi-C protocol described here we generated 73,371,778 valid pair-end reads (4.81 times less than in Wang et al., 2018) and therefore were able to effectively build 1kb and 5 kb resolution matrices as shown in figure 4 A, C of this manuscript and on Arzate-Mejia et al., Nat Comm, 2020.

- All "RESULTS" are only quality controls. All of them reported without specific comments on their meaning and interpretation.

4. We thank the reviewer for his/her comment. As this is a manuscript centred on the in nucleus Hi-C protocol the results and discussion are primarily focusing on the quality controls and checkpoints described in detailed in the protocol as well as describing representative results from it, as demanded by the JoVE "instruction for authors" section. We have added more details to interpret this representative results section of the protocol. Also, as stated in answer 2 we have added a full description of the results presented in the new Figure 4 and indicated that a detailed discussion of the results can be found in Arzate-Mejía et al., Nat Comm 2020. Finally we have also incorporated a "Trouble shooting for key steps of the Hi-C protocol" and a "Limitations of the Hi-C experiment" sections in the Discussion.

- Figure 1B has only one size marker, thus it's not possible to understand the range of sizes --> the same applies to other gel images in figure 2
There seems to duplicated labels. Are these replicates? it's never clear if the authors are presenting replicates or different samples.
Abreviations not defined in the legend. e.g. "UD" undigested? and the same for "D" and "Lig" which are also abbreviations not defined in the figure legend

5. We thank the reviewer for his/her suggestions. We have now added a second and third label to the size marker in addition to the 100bp already presented. We have also defined all abbreviations in the figure legend and text and specified that the samples are two biological replicates.

-Figure 1C very confusing pseudo mathematical equation written at the bottom of Figure 1C, without any explanation in the legend. The explanation in the methods is also not clear.

6. We have added a more detailed and clear description of the equation in the main text under the "Digestion efficiency quantitative control" section of the protocol:

"Digestion efficiency quantitative control

To assess the digestion efficiency more accurately the UD (undigested) and D (digested) samples (see notes 3 and 4 on the protocol) are used as templates to perform qPCR reactions using primers designed as follows:

1. **Design a primer pair that amplifies a DNA fragment that contains the DNA restriction site for the enzyme used for digestion (Mbo I for the present protocol), called R in the formula.**
2. **Design a primer pair that amplifies a control DNA fragment that does not contain the restriction site for the enzyme used for digestion (Mbo I for the present protocol), called C in the formula.**
3. **Use the Cycle Thresholds values (Ct values) of the qPCR amplification to calculate the restriction efficiency according to the equation below:**

$$\%\text{Restriction} = 100 - 100/2^{\{(CtR - CtC)D - (CtR - CtC)UD\}}$$

Where CtR refers to the Ct value of fragment R and CtC refers to the Ct value of the fragment C for the Digested (D) sample and the Undigested (UD) samples.

The restriction percentage reflects the efficiency of the restriction enzyme cutting the restricted (R) DNA fragment compared to a control (C) DNA fragment that does not contains the restriction DNA site. A restriction efficiency equal or above 80% is recommended."

We have also decided to remove it from Figure 1 to avoid duplication of the information fully detailed in the protocol.

- Figure 2A data are not clearly explained. What band in the gel on the left correspond to each of the expected fragments in the schema on the right?

My understanding based on the main text is that in Hi-C MboI is the first cutter and ClaI the second cutter. If this is the case, why there are two bands in the gel for Hi-C with ClaI only cut, i.e. the same as in the next lane with MboI+ClaI cuts?

7. We thank the reviewer for his/her comments. We have now added a schematic of the interaction detected in figure 2A to show where the amplified fragments are located in the genome as well as the primers used. We have also explained the expected digestion patterns in more detail in both the corresponding representative results section and the figure 2A legend as follows:

"To assess Hi-C ligation efficiency in more detail, primers can be designed to amplify an internal ligation product control in which the primers are in forward-forward or reverse-reverse orientation in adjacent restriction fragments. Alternatively primers can be designed to amplify known interactions. In Figure 2A the amplification of a known medium-range (300kb) interaction in Drosophila[26] is presented. Hi-C ligation products (in which the biotin marking, fill-in and ligation occurred successfully) can be estimated by digestion of the PCR product recovered in the amplification. After fill-in and ligation Hi-C amplicons will contain a new Cla I restriction site at the original Mbo I site, which is preserved upon blunt-end ligation. If restriction with Cla I is not complete, the fill-in reaction and biotin marking was inefficient. A digestion efficiency of more than 70% is recommended to avoid having a large proportion of non-useful reads on the libraries after sequencing (Figure 2A, compare the Cla I digestion of the 3C versus the Hi-C template)."

"Figure 2. In nucleus Hi-C fill-in and blunt-end ligation controls. A) Fill-in and biotin labeling assessment. A known interaction between fragments located 300kb apart in chromosome X was used as a control and amplified using the primers indicated in black arrows (see top of the scheme, primer 1 (left), primer 2 (right), Table 3) generating a 347 bp amplicon. Hi-C ligation products can be distinguished from those produced in a 3C experiment by digestion of the ligation site. Hi-C junctions are digested by Cla I at the original Mbo I site, as this forms upon blunt-end ligation. Hi-C and 3C junctions will be digested with Mbo I as the restriction site regenerates upon ligation (see scheme at the left of the gel). 3C junctions in contrast are not digested by Cla I at the Mbo I site but only by Mbo I. Compare the digestion profile of the Hi-C and 3C products using Cla I. A 53bp fragment is obtained digesting the Hi-C product (due to restriction of the Cla I site formed at the Mbo I site and restriction of a Cla I site already present in the region). This fragment is not observed in the 3C product digestion as the only Cla I site available is the one that was already present in the region. B) After PCR amplification of the Hi-C library using different PCR cycles products are ran on a 1.5% Agarose gel. A smear of 400-1000pb is expected. The appropriate number cycles for the final amplification PCR should be taken as the number immediately below that at which a smear is just visible. C) Final library Cla I digestion. An aliquot of the final library is re-amplified and digested with Cla I. The size reduction of the smear confirms that a large proportion of the molecules in the library are valid Hi-C pairs. Densitometry analysis of this gel can be

performed to obtain a ratio between the UN and D samples as detailed in the "Representative results" section."

- Figure 2C: are these 2 replicates? unclear lebelling

8. We have now specified this are two biological replicates in the figure legend

- Trimmed reads alignment is mentioned (but not explained) in the tables of results. Why not using chimeric reads alignment strategies (e.g. iterative mapping) as adopted in many pipelines for Hi-C?

9. We have added an explanation for the trimmed reads alignment on the representative results section as follows:

"After paired-end sequencing the fastq files (Table 3) were processed using HiCPro[28] and the generated statistics plotted using MultiQC[29]. An alternative tool to HiCPro is the HiCUP[30] pipeline obtaining similar results (not shown). In Figure 3 and Table 2 the detailed statistical information of the sequenced reads is shown. Full read alignment and alignment after trimming are reported. These two categories correspond to successfully aligned reads that will be used in subsequent analysis to find valid Hi-C pairs. The alignment after trimming category refers to reads spanning the ligation junction, which were not aligned in the first step and are trimmed at the ligation site to then realign their 5'extremity to the genome [29]. (Figure 3B and C, Table 2). As observed the contact statistics show that the Hi-C library was of high quality with 82.2% valid pairs and 7.6% non-useful reads falling into the same fragment-self-circle, same fragments-dangling-ends, re-ligation, filtered pairs and dumped pairs categories (Figure 3A and D, Table 2). Also, the amount of PCR duplicates is very low, which shows the library complexity is high and the PCR cycles introduced minimal artifacts (Figure 3E, Table 2)."

To analyse the Hi-C data presented here mapping, filtering, correction and generation of Hi-C matrices were done using HiC-Pro[29] which is a broadly used pipeline for Hi-C analysis and more importantly was the pipeline used by us in[26] Arzate-Mejia et al., Nat Comm 2020. The use of chimeric reads by iterative mapping is of course a great suggestion and will increase the amount of valid read pairs. However our protocol yielded 90.2% valid read pairs using HiC-Pro (Figure 3B), thus iterative mapping in this case will not add much more valid pairs and will not change the presented results.

- Figure 4 is pixelated.
Its legend says ". Significant contacts are marked by asterisks." but there are no asterisks. and anyway it's not explained how significant contacts are called
There is no explanation for figure 4C. e.g. what is the different height of TADs "triangles" representing?

10. We thank the reviewer for his/her comments. Figure 4 has been modified as explained in response number 2.

Minor Concerns:
- some lines in the manuscript PDF are highlighted in yellow.

11. JoVE asks authors to highlight the text in the protocol that includes the steps that will be recorded by JoVE in the video.

- Figure 1A "Chromatin restriction" --> "Restriction enzyme digestion of DNA"

12. We have now modified the label to "Restriction enzyme digestion of chromatin" as the digestion takes place inside the nucleus on cross-linked chromatin and not nude DNA.

Reviewer #3:
Manuscript Summary:
The current manuscript describe an inside nucleus or in situ Hi-C protocol in Drosophila cells for the analysis of 3D genomic organization. In this protocol, the DNA-DNA proximity ligation to capture chromosomal interactions is performed in intact nuclei, leading to a better quantification of ligation junctions and the generation of dense Hi-C maps.
In general, the protocol is well written, clear and easy to follow step by step. The efficacy of the protocol is well demonstrated, with several important quality control checkpoints indicated all along the protocol, and which are well illustrated. The protocol would be of great values for beginners in this field, including the indication of tools for analysis. Although I have several major and minor points that need to be addressed by the authors, I would certainly recommend this manuscript for publication.

1. We appreciate the reviewer´s comments.

Major Concerns:
- The Rao et al. in situ Hi-C paper, Cell 2014 should be cited. This paper really pioneers the usage of the in situ Hi-C protocol in many laboratories and in different organisms.

2. We thank the reviewer for his/her comment. This was a mistake on our part as, of course, we were supposed to cite these two key papers and thought we have done so. We have now cited them in all appropriate sections of the introduction and discussion.

- Figure 1B shows two examples of HiC MboI Digested and HiC MboI Ligated. Although the smear is similar in both MboI D lanes, the signal is much weaker in the second Hi-C MboI Lig (far right lane). The author should explain why such difference between the two samples and whether it affects further processing.

3. We thank the reviewer for his/her comment. The differences in signal in this gel are due to different amounts of DNA loaded on the gel and do not interfere with further processing of the samples. We have now added the following to the Figure 1B legend: "The differences in signal strength are due to uneven amounts of loaded DNA on the gel."

- Also on the graph 1C, not sure to what samples of Fig. 1B, the HiC MboI Digest bars correspond.

4. We thank the reviewer for his/her comment. We have now added clearer labels to figure 1B and C and added text to the figure legend to specify that the graphs in figure 1C correspond to the to Hi-C 1 and Hi-C2 samples presented in figure 1B as follows:

"C) Hi-C digestion quantitative control by qPCR for the same two biological replicates as in B (Hi-C 1 and Hi-C 2) using the cycle threshold values as detailed in the "Digestion efficiency quantitative control" step of the protocol. A successful digestion has > 80% restriction."

- For Figure 2, what distinguish their 3C and HiC fragments? The author should better precise in the figure legend (or even in the text) the difference between 3C and HiC generated fragments; I presume that 3C fragments are also ligated at MboI site, but no blunt end ligation, meaning no ClaI site formed in this case. Not sure that it is trivial for everyone.

4. We thank the reviewer for his/her comment. We have now modified the representative results section as well as the figure legend of figure 2 to be more thorough in the explanation of the expected digestion patterns:

"To assess Hi-C ligation efficiency in more detail, primers can be designed to amplify an internal ligation product control in which the primers are in forward-forward or reverse-reverse orientation in adjacent restriction fragments. Alternatively primers can be designed to amplify known interactions. In Figure 2A the amplification of a known medium-range (300kb) interaction in Drosophila[26] is presented. Hi-C ligation products (in which the biotin marking, fill-in and ligation occurred successfully) can be estimated by digestion of the PCR product recovered in the amplification. After fill-in and ligation Hi-C amplicons will contain a new Cla I restriction site at the original Mbo I site, which is preserved upon blunt-end ligation. If restriction with Cla I is not complete, the fill-in reaction and biotin marking was inefficient. A digestion efficiency of more than 70% is recommended to avoid having a large proportion of non-useful reads on the libraries after sequencing (Figure 2A, compare the Cla I digestion of the 3C versus the Hi-C template)."

"Figure 2. In nucleus Hi-C fill-in and blunt-end ligation controls. A) Fill-in and biotin labeling assessment. A known interaction between fragments located 300kb apart in

chromosome X was used as a control and amplified using the primers indicated in black arrows (see top of the scheme, primer 1 (left), primer 2 (right), Table 3) generating a 347 bp amplicon. Hi-C ligation products can be distinguished from those produced in a 3C experiment by digestion of the ligation site. Hi-C junctions are digested by Cla I at the original Mbo I site, as this forms upon blunt-end ligation. Hi-C and 3C junctions will be digested with Mbo I as the restriction site regenerates upon ligation (see scheme at the left of the gel). 3C junctions in contrast are not digested by Cla I at the Mbo I site but only by Mbo I. Compare the digestion profile of the Hi-C and 3C products using Cla I. A 53bp fragment is obtained digesting the Hi-C product (due to restriction of the Cla I site formed at the Mbo I site and restriction of a Cla I site already present in the region). This fragment is not observed in the 3C product digestion as the only Cla I site available is the one that was already present in the region. B) After PCR amplification of the Hi-C library using different PCR cycles products are ran on a 1.5% Agarose gel. A smear of 400-1000pb is expected. The appropriate number cycles for the final amplification PCR should be taken as the number immediately below that at which a smear is just visible. C) Final library Cla I digestion. An aliquot of the final library is re-amplified and digested with Cla I. The size reduction of the smear confirms that a large proportion of the molecules in the library are valid Hi-C pairs. Densitometry analysis of this gel can be performed to obtain a ratio between the UN and D samples as detailed in the "Representative results" section."
We have also added a schematic of the amplified genomic interaction in Figure 2A and indicated the primers used to do so.

- The authors should also mention to what correspond « a ligation junction formed by two nearby fragments ». Are the fragment juxtaposed or closed but separated. Also, is this fragment corresponds to a FF, RR, RF or FR pair? In other words, they should be a bit more precise in the figure legend on the origin and definition of this ligated fragment.

5. We thank the reviewer for his/her comment. See response number 4.

-Also in this figure, the authors should draw the ClaI HiC site closer to the MboI site. If not, it can be confusing and we don't follow well the digestion pattern presented on the left. Also, the digested fragments respective to the original Ligation Fragment and the restrictions sites need to be better aligned. Finally, what means « Kirre » in this panel?

6. We thank the reviewer for his/her comments. We have modified the figure 2A to align the Cla I site generated through Hi-C at the Mbo I site and distinguish it from a Cla I site already present before Hi-C ligation in the genomic fragment and also replaced the "Kirre" legend for "ligated fragment".

- Is there a way to quantify the size reduction of the smear in Fig 2C, indicative of valid HiC pairs? A ratio could maybe help.

7. We thank the reviewer for her/his comment. Indeed a ratio between the UD and D samples could be useful. We have now added the following text to the representative results section of the protocol. "A ratio between the upper size range (determined by the size present in the UD sample) and the bottom size range in both UD and D samples should produce a ratio above 1 for the UD and a ratio of 1 or lower for the DIG sample if the Cla I digestion was efficient."

- In the discussion, the authors should evaluate whether this rather straightforward protocol could be readily applied to mammalian cells. And if yes, what could be the differences, if any, with the Drosophila cells used in this protocol. Such paragraph could be highly informative for the readers.

8. We thank the reviewer for her/his comment. We have now added the following paragraph in the discussion section:
"The same technology has also been efficiently applied in mammalian tissues with some modifications[34]. Mainly when processing a tissue instead of a single cell suspension, the tissue is sieved through a 70µM filter and the lysis step is performed while homogenizing the material using a dounce homogenizer. In addition, as the mammalian genome is 25 larger than the Drosophila genome the amount of valid read pairs needed to build a 1-5kb resolution matrices is greater."

Minor Concerns:
- In the Table of Materials, the ClaI reference is missing. Also Proteinase K is missing. Why Protein Kinase is mentioned here? I don't think it is used in this protocol.

9. We have added the reference for the Cla I enzyme and corrected Protein Kinase to Proteinase K.

- Is « SPRI » the equivalent of « SpriBeads ». Define and explain, and make it consistent throughout the text.

10. In the "Size selection" section of the protocol we define SPRI beads as it is the first time we mentioned the acronym for Solid Phase Reversible Immobilization. We have now consistently referred to SPRI beads throughout the text.

- For Figure 3C, it would have been nice to have an illustration of the filtered valid pairs (FF, RR, RF and FR pairs) versus the non-valid pairs (self-circle ligation, dangling-ends, re-ligation…). This could help beginners to visualize valid versus non-valid reads.

11. We have now added an extra panel at the top of figure 3 (A) to illustrate the different types of non-valid pairs commonly produced in the experiment as suggested by the reviewer and described them in the figure legend:

"Figure 3. HiC-Pro statistics of the Hi-C library. A) Schematic representation of valid Hi-C pairs and the different types of non-valid pairs that can be produced during the experiment and filtered out by HiCPro[28] (Table 2). These include reads falling into contiguous sequences, dandling ends, same fragment, self circle, re-ligations, and PCR duplicates. B) Mapping statistics. Reads that failed to align are shown (grey) and both full aligned reads and reads aligned after trimming are shown in blue and light blue respectively. These two categories represent the useful reads that are considered in the subsequent analysis. C) Pairing statistics. Multi Aligned reads (dark orange) represents reads that are aligned in multiple regions in the genome. Uniquely Aligned (dark blue) represents the read pairs that are aligned once in the genome and singletons (light orange) represent read pairs in which just one genomic region was sequenced in both reads. D) Filtering statistics. Valid read pairs (blue) represents successful Hi-C ligation products as described in A. Self fragment self-circles (light pink) are non useful reads as they represent the same genomic fragment as shown in A. Same fragment dandling end (orange) represents reads in which a single restriction fragment was sequenced. Filtered and dumped pairs (brown) are also unuseful reads that have the wrong size or that the ligation product could not be reconstructed. Finally re-ligation reads (red) represents reads in which two adjacent fragments where re-ligated thus producing unuseful information. E) Valid read pairs contact distribution in the genome. Unique cis contacts (blue) are most frequent than unique trans contacts (green)."

For simplicity we have also grouped the different types of valid-pairs into one category as the different combinations (i.e. FF, RR, RF, FR) represent that genomic fragments can be ligated in Forward or Reverse orientation.

- The Antennapedia complex should be abbreviated « ANT-C », « Antp » corresponding to the Antennapedia gene. What means « Btx-C » in Figure 4A? If it means the Drosophila Bithorax complex, it should be written « BX-C ». If not, precise.

- In the same Figure, the square in panel A should correspond to the square zoomed in panel B.

- In Figure 4C, for the virtual 4C, the significant contacts are not indicated, nor the scale of the figure.

12. We have modified Figure 4 to answer concerns by other reviewers, thus these comments do not longer apply. The modifications to figure 4 changed its description in the representative results section, discussion section and figure legend as detailed below. We hope the reviewer find this new figure more coherent with the rest of the text.

Representative results section

"Using the Hi-C valid pairs from HiCPro[28], ICE normalized matrices at 1 kb, 5 kb bin resolutions were made using hicPlotMatrix tool in HiCExplorer [31-33]. An example of

normalized contact matrix at 1kb and 5kb resolution are presented for the *Notch* gene locus in Drosophila (Figure 4A, C and D). In figure 4A the *Notch* gene locus can be seen in detail together with the Aps, Pl II and histone modifications occupancy along the locus (Figure 4A, Table 3). The design of the CRISPR-Cas9 deletion shown involved de motif of CTCF and M1BP as detailed in figure 4B. Upon deletion of the region containing both CTCF and M1BP DNA binding sites at the 5' boundary of the *Notch* locus (5pN-delta343, Table 3), a dramatic change in chromatin contacts can be observed with loss of interactions inside the *Notch* locus and gain of contacts with the upstream TAD compared to the WT (Figure 4C, D). Finally a detailed panorama of WT and mutant interaction profiles at the restriction fragment level from *Notch* gene 5'UTR is presented again, showing a decrease in the proportion of contacts made with the *Notch* gene locus and an increase of contacts with the upstream domain (Figure 4E). This virtual 4C from the *Notch* gene 5'UTR and exon 6 was obtained using the Hi-C other-ends tool implemented on SeqMonk (SeqMonk (RRID:SCR_001913) http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/. All of the results presented in Figure 4 were obtained applying the in nucleus Hi-C protocol presented here in WT and mutant S2R+ Drosophila cells as described in detail in[26]."

Discussion section

"With a high quality Hi-C experiment as the one described here for the Drosophila genome sequenced at the presented depth (Table 1) one can build matrices at a wide range of resolutions (from 1, 5kbs, 50kb or lower, see Figure 4). Also, if a particular region of the genome wants to be evaluated at the restriction fragment level, the data can be used to build a virtual 4C landscape of the desired viewpoint (for instance the *Notch* gene 5' UTR in Figure 4E). The Hi-C other ends tool in SeqMonk is very a friendly option to do so and statistically significant contacts can be obtained applying the 4C quantification tool also incorporated in SeqMonk.

Applying the in nucleus Hi-C experiment described here to a collection of mutant cell lines with altered AP DNA binding sites at the TAD boundaries (Figure 4), we demonstrated that genetic elements at boundaries are needed to structure the Drosophila genome in domains and sustain gene expression regulation as fully discussed in[26]. Thus, genetic edition of regulatory elements with CRISPR/Cas9 system combined with high-resolution profiling of genomic interactions through the in nucleus Hi-C protocol described here, provide a powerful strategy to test the structural function of genetic elements."

Figure 4 legend

"Figure 4. Hi-C contact matrices and virtual 4C analysis of WT and mutant S2R+ cells. A) Hi-C normalized heatmap of a 50 kb region at 1-kb resolution centered in the *Notch* gene locus. TAD separation score[33] for the locus is shown. The *Notch* locus is partitioned into two topological domains (Domain 1 and Domain 2). ChIP-seq data for Architectural Proteins (APs), RNA Pol II REF, and histone marks for S2/S2R+ cells[35-38] is shown below

the heatmap (Table 3). The position of the *Notch* domain 1 boundaries, are highlighted in light green. B) Schematic representation of B1 boundary CRISPR mutant. The green rectangle indicates the deleted 343 bp region. Scissors indicate sgRNAs used for CRISPR mediated genome editing. Motif binding sites for APs are shown as boxes for CTCF and M1BP. Peak summits for DNA-binding APs shown in A are also indicated[36] C) Hi–C normalized heatmaps at 1-kb resolution covering a 50-kb region centered in *Notch* for the wilde-type (WT) and the mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells. D) Hi–C normalized heatmaps covering a 250-kb region centered in *Notch* at 5-kb resolution for WT and mutant cells. Left, Hi–C heatmaps of the log2 differences in interaction frequency between WT and mutant cells E) Virtual-4C for WT and mutant cells using the 5′ UTR of *Notch* as viewpoint. The percent of interactions between the viewpoint and regions within the upstream *kirre domain-2*, the *Notch* domain 1, the *Notch* domain 2, and the downstream *dnc* domain for both WT and mutant cells are shown[26].”