1524–1540 *Nucleic Acids Research, 2014, Vol. 42, No. 3* *Published online 29 October 2013*
doi:10.1093/nar/gkt998

Hox in motion: tracking HoxA cluster conformation during differentiation

Mathieu Rousseau[1,2], Jennifer L. Crutchley[1], Hisashi Miura[1], Matthew Suderman[2], Mathieu Blanchette[2] and Josée Dostie[1,*]

[1]Department of Biochemistry and Goodman Cancer Research Center, McGill University, Montréal, Québec, H3G 1Y6, Canada and [2]School of Computer Science and McGill Centre for Bioinformatics, McGill University, Montréal, Québec, H3A 0E9, Canada

Received April 17, 2013; Revised August 28, 2013; Accepted October 2, 2013

ABSTRACT

Three-dimensional genome organization is an important higher order transcription regulation mechanism that can be studied with the chromosome conformation capture techniques. Here, we combined chromatin organization analysis by chromosome conformation capture-carbon copy, computational modeling and epigenomics to achieve the first integrated view, through time, of a connection between chromatin state and its architecture. We used this approach to examine the chromatin dynamics of the *HoxA* cluster in a human myeloid leukemia cell line at various stages of differentiation. We found that cellular differentiation involves a transient activation of the 5′-end *HoxA* genes coinciding with a loss of contacts throughout the cluster, and by specific silencing at the 3′-end with H3K27 methylation. The 3D modeling of the data revealed an extensive reorganization of the cluster between the two previously reported topologically associated domains in differentiated cells. Our results support a model whereby silencing by polycomb group proteins and reconfiguration of CTCF interactions at a topologically associated domain boundary participate in changing the *HoxA* cluster topology, which compartmentalizes the genes following differentiation.

INTRODUCTION

Chromosomes are packaged in a complex hierarchical manner that functionally condenses long chromatin fibers in small nuclear volumes (1,2). Although the mechanisms underlying the formation and regulation of higher order architectures are mostly uncharacterized, it has become apparent that spatial chromatin organization constitutes an important mechanism of gene regulation. For instance, genome organization can bring control DNA elements like promoters, enhancers and insulators close to each other such that they can physically interact even when they lie far apart on the same chromosome or on different ones. Such functional long-range contacts have been found genome-wide and can correlate with either activation or repression of transcription (3–5).

The chromosome conformation capture (3C) technologies are a recently developed class of molecular approaches used to study spatial chromatin organization at high resolution *in vivo*. These include 3C (6), circular chromosome conformation capture (4C) (7–10), 3C-carbon copy (5C) (11), Hi-C (12,13) and the tethered chromosome conformation capture (TCC) methods (14). All 3C-related techniques use proximity-based ligation to capture chromatin contacts in cell populations. The resulting ligation events can be quantified by different methods including quantitative polymerase chain reaction (PCR) and deep sequencing, and the measurements can be used to infer *in vivo* spatial distances, as they are inversely proportional to the measured signals.

The 3C, 4C and 5C were previously used to study the interplay of spatial chromatin organization and gene expression at the *HoxA* gene cluster. This gene locus spans 150 kb on human chromosome 7 and encodes 11 transcription factors that are regulated during development, and whose ectopic expression in adult tissues can lead to diseases (Figure 1A) (15–18). The 3C studies in human cell lines and mouse embryos have uncovered different *HoxA* architectures suggesting that distinct underlying spatial mechanisms likely control these genes in different cell types. For example, we previously found that a silent *HoxA* locus adopts a packaged state with numerous looping contacts within it in human embryonal carcinoma cells (19). Similarly, a study by Noordermeer *et al.* (20) of the corresponding *HoxA* region in non-expressing parts of the mouse embryo, and our 5C study in the human myeloid leukemia cell line THP-1 both revealed that the cluster folds onto itself when genes are repressed (21). In contrast, a high level of interaction was detected within the actively transcribed regions of the

*To whom correspondence should be addressed. Tel: +1 514 398 4975; Fax: +1 514 398 7384; Email: josee.dostie@mcgill.ca

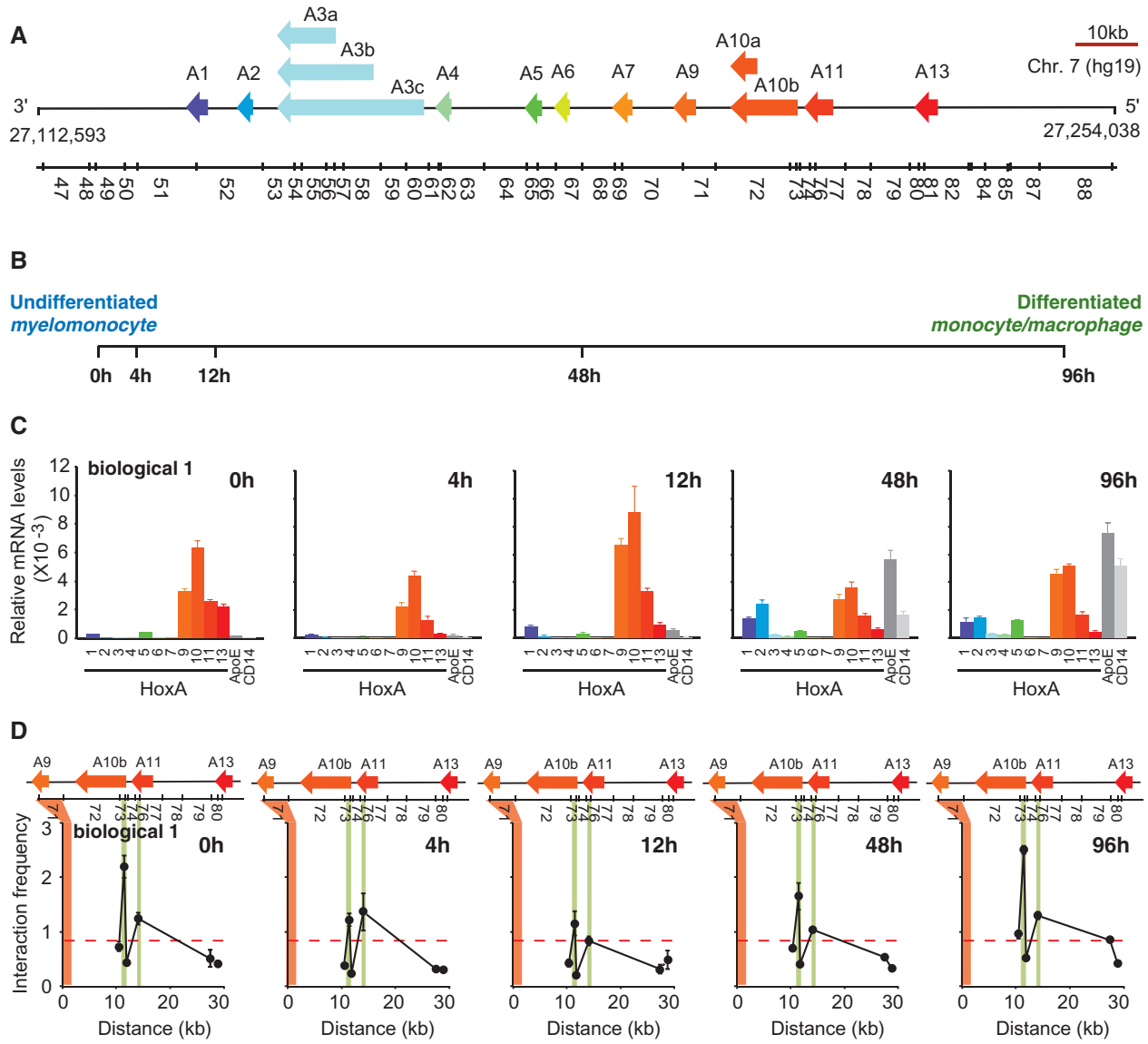

Figure 1. The 5′-end *HoxA* gene expression fluctuates and chromatin conformation varies during THP-1 differentiation. (**A**) Linear schematic representation of the human *HoxA* cluster region is characterized in this study. Genes are shown as left-facing arrows to illustrate transcription direction and to highlight the 3′–5′-end orientation of the cluster. Paralog groups are color-coded and identified above each gene. The color code shown here was used throughout the study. The predicted *Bgl*II restriction fragments of the *HoxA* region characterized in this study are shown below and identified by numbers from left to right. (**B**) THP-1 differentiation time points examined in this study. (**C**) The 5′-end *HoxA* gene expression fluctuates throughout the cellular differentiation time course. The expression of *HoxA* genes was measured by RT-qPCR and normalized relative to actin. The number below each histogram bar identifies the paralog group. The expression of macrophage-specific *ApoE* and *CD14* markers was measured to monitor differentiation. For comparison, the *ApoE* and *CD14* expression levels were divided by 4 and 20, respectively. Each histogram value is the average of at least three PCRs, and error bars represent the standard deviation. (**D**) Chromatin looping changes with gene expression at the *HoxA* 5′-end. Interactions between *HoxA9* and the region containing other transcriptionally regulated 5′-end genes were measured by conventional 3C. The genomic region shown above each graph is to scale. The 'fixed' 3C region is highlighted in orange and the position of known looping contacts is indicated with green vertical lines. Fragments probed were 71 (fixed), 72, 73, 74, 76, 79 and 80. The y-axis shows normalized IFs and the x-axis indicates the end-point distance from the fixed 3C region. Each contact was measured at least three times in cellular and control libraries, and error bars represent the standard error of the mean.

cluster in actively expressing embryonic regions (20). Additionally, Wang *et al.* (22) performed a study in human fibroblasts that used 5C to measure the *HoxA* cluster conformation. Their results indicated that silenced regions marked by H3K27me2/3 adopt an open linear conformation, whereas regions containing transcribed genes exhibit a high level of interaction with each other that was interpreted to form a condensed structure.

Whether *HoxA* displays different architectures because it is controlled by different higher order transcription regulation mechanisms remains unclear. A drawback to using different cell types to study how chromatin conformation relates to gene expression is the difficulty in assessing the actual contribution of architectures to transcription. However, this relationship can be more clearly examined when observed through time along with corresponding

changes in transcription activity and chromatin composition in a given cell system. Here we tracked the *HoxA* cluster organization with 5C throughout differentiation of THP-1 myelomonocytes into monocytes/macrophages. We then used a computational approach to examine how chromatin activity relates to composition and organization, and present an integrated view of the changes occurring at the cluster throughout differentiation.

MATERIALS AND METHODS

Cell culture

The THP-1 human cell line was derived from the peripheral blood of a 1-year-old boy with acute monocytic leukemia. Undifferentiated THP-1 cells are myelomonocytic and express the oncogenic MLL-AF9 fusion protein. Dr Yoshihide Hayashizaki provided the THP-1 line used in this study (RIKEN Yokohama Institute, Japan), and is a subclone (THP-1.5) selected for its ability to differentiate homogeneously in response to phorbol 12-myristate 13-acetate (PMA). THP-1 cells were cultured in Roswell Park Memorial Institute medium (RPMI 1640; Invitrogen™, Burlington, ON, Canada) supplemented with 10% fetal bovine serum (HyClone, Logan, UT, USA). The medium was made 'complete' by adding 50 μM of 2-mercaptoethanol (Invitrogen™), 1 mM sodium pyruvate (Invitrogen™), 10 mM HEPES (Invitrogen™) and 1% penicillin–streptomycin (Invitrogen™). The cell cultures were grown in a controlled environment at 37°C in 5% CO_2 atmosphere.

For differentiation, cells were first centrifuged at 1000 rpm for 10 min, resuspended into fresh RPMI (complete) and pooled into a single suspension. Cell density was adjusted to 2×10^5 cells/ml, and the 0-h time point samples were collected. PMA was added to the remaining cell pool at a final concentration of 30 ng/ml, distributed (50 ml) into 15-cm culture dishes and incubated for 4, 12, 48 or 96 h. At the 0-, 4- and 12-h time points, cells remained in suspension and were collected by centrifugation as described later. At the 48- and 96-h time points, the cells were adherent and collected by scraping into 10 ml of fresh RPMI (complete) after one wash with complete RPMI to remove any remaining non-adherent cells. The content of each cell plate was centrifuged separately in 50 ml conical tubes at 2000 rpm for 5 min when samples were collected for formaldehyde fixation, or 10 min when collected for RNA extraction. The media was removed, and pellets were either frozen on dry ice before transferring to −80°C for RNA extraction or resuspended into 22.5 ml of RPMI media containing 1% formaldehyde for fixation. Cells were fixed 10 min with inversion every 2 min, and quenched with glycine at a 200 mM final concentration for 5 min at room temperature followed by 15 min on ice. Samples were centrifuged at 2000 rpm for 5 min at 4°C and washed with 8 ml of cold 1 × phosphate buffered saline. The resulting cell pellets contained approximately 1×10^7 cells and were immediately frozen on dry ice before storage at −80°C. These pellets were used for chromatin immunoprecipitation (ChIP) analysis and for the production of 3C libraries.

Quantitative real-time reverse transcription-polymerase chain reaction

Total THP-1 RNA was extracted with the RNeasy© Mini kit (Qiagen) as described by the manufacturer. Total RNA samples were treated with DNAseI and repurified on Qiagen columns. Two micrograms of total RNA was used for each reverse transcription reaction using the Superscript® III reverse transcriptase kit (Invitrogen™) and oligo$(dT)_{20}$. Real-time PCR quantification was performed with the SsoFast™ EvaGreen© Supermix (Bio-Rad) using a Bio-Rad CFX96™ (C1000™ series) real-time system. Quantification was performed with 2-fold serial dilutions of total complementary DNA, and relative gene expression data were calculated using the 2[-Delta Delta C(T)] method as described previously (23). In all cases, the RT-qPCR data were derived from single complementary DNA samples originating from one RNA preparation. The size and specificity of amplified products was verified with melting temperatures and on agarose gels containing 0.5 μg/ml ethidium bromide. Gels were documented and analyzed with an AlphaImager© HP coupled to a 12-bit digital camera and equipped with the AlphaView® image acquisition and analysis software (version 3.0; Alpha Innotech Corporation). The qPCR primer sequences used in our study were designed to amplify mRNA 3′-ends such that all predicted forms of the *HoxA3* and *HoxA10* transcripts are quantified in our experiments. These primers were previously described (20) and are also available on our Web site (http://Dostielab.biochem.mcgill.ca).

3C analysis

The control *HoxA* 3C library used to correct for differences in primer pair efficiencies contains the following bacterial artificial chromosome (BAC) clones purchased from Invitrogen™: RP11-1132K14, CTD-2508F13, RP11-657H18, RP11-96B9, RP11-197K24, CTD-2594L23 and RP11-1132K14. The control library was generated as described previously (11,24). Briefly, BACs were mixed in equimolar ratio, digested with *Bgl*II and randomly ligated in solution with T4 DNA ligase. This library contains the four human *Hox* clusters and one gene desert region (ENCODE ENr313) used to normalize data from different datasets.

Cellular 3C libraries from the time course were prepared as described previously (11,19,24). Briefly, actively growing cells (0 h) and cells treated with PMA for 4, 12, 48 and 96 h were collected and fixed in the presence of 1% formaldehyde. The 3C libraries were prepared from approximately 2×10^7 cells by digesting overnight with *Bgl*II and ligating with T4 DNA ligase for 2 h to yield 3C products between cross-linked restriction fragments. Purified 3C libraries were titrated by PCR with 3C primers against neighboring restriction fragments in the control gene desert region (ENCODE ENr313). Results from these titrations were used to select 3C library volumes yielding approximately equal levels of 3C products. Contacts of the fixed point *HoxA* 3C analysis were measured in triplicate and repeated as needed to obtain at least three values for each pairwise interaction.

All PCR products were resolved on 1.5% agarose gels containing 0.5 μg/ml ethidium bromide in 1× TBE, and visualized via ultraviolet transillumination at 302 nm. Reactions containing primer dimers or additional bands were excluded, and reactions were repeated as necessary. Image acquisition and documentation was done with the AlphaImager © HP described above, and gel analysis with the AlphaView® software. To calculate interaction frequencies (IFs), we first measured 3C signals and their corresponding background signals taken above or below each lane. Background signals were then subtracted, and cellular 3C signals were divided individually by each of the control 3C measurements. This process, which yields at least three ratios for each 3C measurement, was used to minimize the error originating from variations in the 3C signals measured on gel. This process is prone to error, as background signals tend to fluctuate. The resulting 3C ratios were finally averaged, and the data from each dataset normalized. Normalization between datasets was with a minimum of five different gene desert contacts measured at least in triplicate. *HoxA* normalization factors were estimated from the average log ratio of corresponding gene desert contacts between 3C libraries as previously described (24). The 3C primer sequences used in this analysis were previously described (20) and are also available on our Web site (http://Dostielab.biochem.mcgill.ca).

Preparation of 5C libraries

The *HoxA* 5C library design is as previously reported (20). In summary, alternating forward and reverse 5C primers were designed against consecutive *Bgl*II fragments in the *HoxA* (hg19|chr7:27,112,593-27,254,038) and gene desert (hg19|chr16:62,276,449-62,776,448) regions. The *HoxA Bgl*II fragments characterized by 5C were numbered from 47 to 88 as shown in Figure 1. A total of 29 forward and 28 reverse 5C primers were used to prepare 5C libraries. The 5C primer corresponding to *HoxA* fragment 48 has low complexity and was excluded from library preparation. This design covers 50% of all possible contacts in each region and measures up to 812 different pairwise interactions.

The 5C libraries for each time point and a control library were prepared as previously described (11,20,25,26). Briefly, each 3C library was mixed with salmon testis DNA (Sigma®) to a combined DNA mass of 1.5 μg, before adding the multiplexed 5C primer mix to obtain a final individual primer concentration of ~20 nM in a 10-μl final volume of annealing buffer (20 mM Tris-acetate, pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate and 1 mM dithiothreitol). Samples were denatured for 5 min at 95°C and annealed overnight at 48°C. Annealed 5C primers were ligated by adding 20 μl of ligation buffer containing 10 U of Taq DNA ligase for 1 h at 48°C (New England Biolabs; 25 mM Tris–HCl, pH 7.6, 31.25 mM potassium acetate, 12.5 mM magnesium acetate, 1.25 mM Nicotinamide Adenine Dinucleotide, 12.5 mM dithiothreitol and 0.125% Triton X-100). The samples were then incubated for 10 min at 65°C to terminate the reactions and amplified by PCR with the T7 (TAA TACGACTCACTATAGCC) and T3 (TATTAACCCTC ACTAAAGGGA) primers, which are complementary to the common forward and reverse tail sequences of 5C primers, respectively. The reverse T3 primer was 5′-Cy3-labeled for visualization on microarrays. The MinElute Reaction Cleanup kit (Qiagen®) was used as recommended by the manufacturer to remove unincorporated primers and other contaminants from the amplified 5C libraries. The 5C primer sequences used in this analysis are listed in Supplementary Table S1, were previously described (20) and are also available on our Web site (http://Dostielab.biochem.mcgill.ca).

Microarray analysis of 5C libraries

The Cy3-labeled multiplexed 5C libraries were analyzed on the custom microarrays (Roche Diagnostics) described previously (20). In summary, the arrays contained the sense strand of >45 000 possible 5C ligation products within the four human *Hox* clusters and the ENr313 control gene desert region. Each predicted contact was represented by eight replicates of increasing lengths ranging from 30 to 48 nt centered at the restriction site, which served for quality control and to estimate background signals (see later ''Calculating interaction frequencies with 'IF Calculator 2.0' ''). Approximately 100 ng of 5C libraries were individually hybridized per microarray using the NimbleGen CGH Hybridization kit as recommended by the manufacturer (27–29). Arrays were scanned at 5 μm resolution with a DNA microarray scanner (Agilent Technologies, model G2505). The NimbleScan 2.6 software (NimbleGen Systems, Inc.) was used to extract the data from images, and specific features were extracted with our 'ArrayQC' software as previously described (20).

Calculating 5C IFs with 'IF Calculator 2.0'

The raw microarray fluorescence signals were processed using an updated version of the previously described 'IF Calculator' software (30). The program was updated to include a threshold test for saturated signals and a test to verify the trend of the signal intensity pattern. These tests make use of the replicate features of increasing lengths (30, 36, 38, 40, 42, 44, 46 and 48 nt) on the custom microarrays. The first test removes values that are above a saturation cutoff of 55 000 and below a background cutoff specified as 150% of the 30-nt feature signal. The trend test applied to the remaining background-subtracted values is based on the property of measured signal intensities to monotonically increase with increasing probe length, and selects the longest subsequence of probe lengths that respect this rule. All remaining measurements are then normalized in two ways. Individual cellular signals are first divided by the corresponding BAC signals to correct for primer pair efficiencies, which yields IFs. IF values belonging to the same contact are then averaged and standard deviations are calculated. The second normalization step adjusts *HoxA* IF and standard deviation values between different datasets. The *HoxA* data are adjusted with normalization factors calculated from the control gene desert region using the average log ratio of corresponding gene desert

contacts as previously described (24). Files containing the resulting IFs can then be converted with the 'AnnotateIFs' program into a format compatible with the online my5C tool for visualization (31). The updated 'IF Calculator 2.0' program is available on our Web site (http://Dostielab.biochem.mcgill.ca).

ChIP-chip

THP-1 cell pellets (1×10^7 cells) were fixed with formaldehyde as described earlier (see Cell culture) and used for ChIP as described previously (http://jura.wi.mit.edu/young_public/hESregulation/Young_Protocol.doc). Briefly, the fixed cells were resuspended in LB1 (lysis buffer 1; 50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 10% glycerol, 0.5% NP-40 and 0.25% Triton X-100). Lysates were centrifuged, and pellets were resuspended in LB2 (10 mM Tris–HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM ethylene glycol tetraacetic acid and protease inhibitors). Lysates were centrifuged again and resuspended in 1 ml of LB3 (10 mM, Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM ethylene glycol tetraacetic acid, 0.1% Na-deoxycholate, 0.5% *N*-laurylsarcosine). The resulting lysate was sonicated for 10 min with a Branson 450D Sonifier equipped with a cooled (4°C) water-jacketed cup horn at 80% duty cycle (30 s ON–30 s OFF). Sonication into fragments of 100–500 bp was verified on agarose gel and the equivalent of two cell pellets was used to ChIP H3K27me2/3 with 10 µg of antibody (Abcam, Ab6002). Samples were incubated overnight at 4°C with magnetic beads pre-bound with the antibody (Protein-G Dynabeads®; Invitrogen™). The precipitated complexes were washed twice each in low salt wash buffer [20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.1% sodium dodecyl sulphate (SDS), 1% Triton X-100 and 2 mM EDTA], medium salt wash buffer (20 mM Tris–HCl, pH 8.0, 250 mM NaCl, 0.1% SDS, 1% Triton X-100 and 2 mM EDTA), LiCl wash buffer (10 mM Tris–HCl, pH8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate and 1 mM EDTA) and $1 \times$ TE buffer. Washed complexes were eluted twice with elution buffer (50 mM Tris–HCl, pH 8.0, 10 mM EDTA and 1% SDS) at 65°C, and cross-links were reversed by incubating samples at 65°C overnight. RNAsed samples were then treated with proteinase K, purified twice by phenol-chloroform and ethanol-precipitated.

Input and ChIPed materials were amplified with the GenomePlex® Complete Whole Genome Amplification kit (Sigma-Aldrich) as recommended by the manufacturer. Samples were fluorescently labeled with the NimbleGen Dual-Color DNA labeling kit (version 01; NimbleGen Systems, Inc.) as recommended by the manufacturer. Input and ChIPed materials were Cy3- and Cy5-labeled, respectively. Approximately 6 µg of each corresponding input and ChIP materials were hybridized per custom tilling array (32) according to the manufacturer's protocol (Roche Diagnostics). Arrays were scanned at 5 µm resolution with a DNA microarray scanner (Agilent Technologies, model G2505). Data from scanned images were extracted with the NimbleScan 2.6 software (NimbleGen Systems, Inc.).

ChIP-chip data analysis

The 'Ringo' package (version 1.16.0) (33) from the 'Bioconductor' project (version 2.8) (34) in '*R*' (version 2.10.1) was used to normalize and analyze the ChIP-chip data. The time course dataset was normalized using the NimbleGen-specific preprocessing method based on a Tukey-biweight scaling procedure to center the mean of the intensity distribution of each sample on 0. This step corrects for technical differences in hybridization efficiency between time points. The processed log-ratio signal-to-input data were then smoothed using a running median computed within a window size of 200 bp. Normalization between the time points was performed by correcting for differences in average signal for probes corresponding to the Xist region on chromosome X (the normalization factors were applied such that the average signal intensity ratios were equal in all datasets). This second normalization step corrects differences of ChIP efficiency between samples. Wiggle format files with a smoothing window value of 8 were generated for each time point and visualized as a continuous-value data track in the University of California, Santa Cruz (UCSC) Genome Browser (35).

ChIP-seq

ChIP was conducted as described earlier (see ChIP-chip section). Two cell pellets were used to ChIP CTCF at each time point with 5 µl of antibody (Millipore, catalog number 07-729). These cell pellets are the only samples of this study that were prepared independently from the time course. The precipitated DNA was used to prepare ChIP-seq libraries with the ChIP-seq DNA sample prep kit (Illumina, catalog number 1003473) according to the manufacturer's protocol. ChIP-seq CTCF libraries were sequenced with an Illumina Genome Analyzer (GAiiX) genome sequencer (36 bp reads) at the Génome Québec Innovation Centre (http://gqinnovationcenter.com/).

ChIP-seq data analysis

Sequence reads (36-mer in Fastq format) were aligned to the human genome (UCSC hg18) using the Bowtie program (36). One mismatch was allowed to the unique mapped reads (option: -v 1 -m 1). The HOMER program (http://biowhat.ucsd.edu/homer/chipseq/), as used for peak calling and WCE data from each time point (0 or 96 h), was used as control. The total mapped tags in each sample were normalized to 'Tags per 10 millions' reads. The CTCF binding regions with 2-fold over in both control (whole cell extract) and local (10 kbp) regions were identified by HOMER program [False Discovery Rate < 0.001 ('findpeak' command)]. The 'LiftOver' tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to convert the data from hg18 to hg19 for display in the hg19 genome version on the UCSC browser.

Generating 3D model ensembles with the 'MCMC5C' program

The 3D model ensembles were generated for each 5C dataset using an updated version of the previously described 'MCMC5C' program (37). Briefly, the calculated IF values were used as input to the MCMC5C program that was updated to extend the available set of moves used to generate candidate structures during the Metropolis–Hastings simulation. In addition to the existing 'single point translation' move, a 'block translation' move and a 'mirroring' move were added. The 'block translation' move randomly selects a contiguous subset of points in the structure and then applies the same randomly chosen translation vector to all of the chosen points, whereas the 'mirror' move similarly selects a random subset of contiguous points and reflects them over a randomly chosen plane passing through at least one of the points in the subset. Using this updated version of MCMC5C, ensembles of 200 3D models were sampled from the posterior distribution (after burn-in) for each of the 5C datasets (MCMC5C parameters as follows: total run length = 10^9, move vector step size = 0.05, IF model exponent = 2.0, single point translation move frequency = 80%, block translation move frequency = 10% and mirror move frequency = 10%). The updated MCMC5C program is available on our Web site (http://Dostielab.biochem.mcgill.ca).

Base density analysis with 'Microcosm 2.0'

The base density along the *HoxA* cluster was measured at each time point using the ensembles of 3D models generated by MCMC5C as input to the previously described 'Microcosm 2.0' program (19,20,30). Briefly, a sphere of fixed radius (1.0) is sequentially centered at each 100th bp along a given 3D model, and the total number of base pairs contained within the volume of the sphere is counted and reported as a base density measurement. Base density profiles are calculated for each 3D model in a given ensemble (or time point), and the average and standard deviation values reported. The resulting base density profiles are output as Wiggle files for visualization in the UCSC Genome Browser (35).

Measuring spatial distances in 3D models with 'StructureAnalyzer'

We developed the 'StructureAnalyzer' program to estimate the average pairwise distance between features in 3D models. The distance values are calculated as Euclidean distances between restriction fragment cut sites and are used as a metric to compare between structures. The program takes as input an ensemble of 3D models generated by the MCMC5C program (here 200) and a list of specified features such as the ones examined in this study [transcriptional start sites (TSSs) of *HoxA* genes, CTCF binding sites or H3K27me2/3 peaks]. From these data, StructureAnalyzer outputs averaged 3D Euclidean pairwise distances in the ensemble across all the possible pairs of features.

3D model annotation with 'StructureAnnotation'

The 3D models generated by the MCMC5C program were annotated with our 'StructureAnnotation' program. This program takes as input a single structure and a list of specified features and outputs a Protein Data Bank file containing the location of the features in the structure that can be superimposed and visualized using a Protein Data Bank-viewing software such as PyMOL (38).

Creating '*Hox* in motion' with the 'MovieMaker' program and PyMOL

Annotated 3D models obtained from the MCMC5C program were used by the 'MovieMaker' program to generate a movie of '*Hox* in motion'. MovieMaker takes as input an ordered set of annotated models and produces an ordered series of frames that are used by PyMOL to create a single movie file as output. Each consecutive pair of input 3D models is successively processed to generate a series of intermediary interpolated structures. These intermediary structures are collected across all successive input structure pairs and are rendered at 23.98 frames per second using PyMOL to obtain a single movie file.

Databases and URLs

As usual, the source code for all of our tools is available through our Web site with specific instructions, and these programs can be run from the command line. The data and complete set of tools created in this study are available at the following address (http://dostielab.biochem.mcgill.ca/tools/Rousseau_2012.tbz).

In an effort to improve on their ease of use, we also integrated these programs into a local instance of the GALAXY (39–41) package that is publicly accessible and provides a full graphical user interface (http://dostielab.biochem.mcgill.ca/galaxy/). This 5C analysis pipeline includes all of the new and updated tools described here in 'Materials and Methods', in addition to others we previously published.

RESULTS

The expression of 5′-end *HoxA* genes fluctuates during differentiation

We previously compared the 3D organization of the *HoxA* cluster in PMA-differentiated THP-1 cells (monocytes/macrophages) and corresponding Dimethyl sulfoxide (DMSO) controls (myelomonocytes). In this study, we found that repression of 5′-end *HoxA* genes correlates with an overall increase in chromatin packaging and in the formation of distinct contacts clustering the genes (20). Likewise, we found that induction of *HoxA* 3′-end genes after differentiation of NT2/D1 cells into neuronal lineages with retinoic acid (RA) is accompanied with loss-of-contacts throughout the entire cluster (19). To better understand the relationship between chromatin organization and transcription at the *HoxA* cluster, we first examined gene expression at the cluster throughout a cellular differentiation time course. THP-1 cells were induced with PMA and collected at five different time

points for *HoxA* chromatin and expression analysis (Figure 1A and B). We measured *HoxA* gene expression by quantitative real-time RT-PCR (RT-qPCR) with a previously validated set of primers (Figure 1C) (19,20). As expected, we found that the *HoxA* 5′-end genes *HoxA9*, *10*, *11* and *13* were highly expressed in undifferentiated (0 h) THP-1 compared with other paralogs. Expression of 5′-end genes slightly decreased at the 4-h time point and were transiently induced after 12 h of PMA treatment. These early events were followed by the progressive appearance of differentiation-specific markers *ApoE* and *CD14* at 48 and 96 h post-induction and by the repression of most 5′-end genes as previously reported (20), and as confirmed in an independent time course (Supplementary Figure S1A).

Although modest, there was a gradual increase in *HoxA* 3′-end gene expression throughout the time course, such that the overall *HoxA* expression pattern was different in undifferentiated and differentiated cells (compare 0 and 96 h). The expression profiles of the *HoxA* genes suggested that the cluster behaved as two independently regulated segments: one corresponding to a 3′-end region including *HoxA1* to *A5* where transcription moderately increases, and the other corresponding to the 5′-end genes where expression fluctuates throughout the time course.

While setting up the experimental conditions of the time course, we noticed that the steady-state *HoxA* mRNA levels can vary significantly between experiments, especially in actively growing undifferentiated samples (0 h; compare Figure 1C, Supplementary Figure S1A–C). This type of variability was also observed in the previously published THP-1 datasets of the FANTOM4 project [Supplementary Figure S1B and (42)]. Thus, chromatin conformation at a given time point in one dataset might not reflect its state in another. This fact highlights the importance of comparing expression and chromatin states in matched samples and for this reason, all the data presented here are from corresponding time course samples, unless replicates are discussed.

Loss of contacts at the *HoxA* cluster 5′-end accompanies gene activation

To determine whether fluctuation in *HoxA* gene expression is accompanied by changes in chromatin architecture, we first used 3C to profile contacts within the cluster 5′-end region. The 3C libraries were generated from cells collected at each time point and a control library was prepared from BAC clones as described in 'Materials and Methods' section. We used these libraries to measure the local chromatin architecture around *HoxA9*, which we previously found to form discrete contacts with regions containing *HoxA10* and *11* in THP-1 and NT2/D1 cells (Figure 1D) (19,20). Consistent with our previous results, *HoxA9* displayed a similar contact pattern with upstream genes in actively growing cells (Figure 1D; 0 h). The frequency of these interactions decreased with transient 5′-end gene activation at 12 h, and increased on transcription repression at the 48 and 96 h differentiation time points. These results agree with our previous findings that transcription repression correlates with the formation of strong contacts clustering downregulated 5′-end *HoxA* genes (20).

We also observed a comparable relationship between expression and *HoxA9* contacts throughout a biological replicate time course (Supplementary Figure S1D, biological 3), except at 4 h post-induction when transcription transiently decreases. In this biological replicate, the expression of 5′-end *HoxA* genes was comparable at 0 and 12 h and their 3C patterns correlated well. Also, contacts increased at 48 and 96 h when gene expression decreased. The different *HoxA* cluster behaviors at the 4-h time point might stem from a greater variability in the cell population shortly after cells are induced with PMA. Alternatively, this difference could be attributed to limitations in the 3C analysis itself. While 3C is robust at detecting simple chromatin loops between two genomic regions using an anchored scheme (43), it can be less so when the regions engage in many long-range contacts simply because not all regions are examined and thus accounted for. These exploratory 3C data nonetheless suggest that *HoxA* transcription is reflected by lower contacts at the cluster 5′-end, and warranted a more complete investigation of cluster organization with 5C.

Increasing 5′-end *HoxA* expression correlates with lower chromatin contacts that differ on macrophage differentiation

We next quantitatively mapped the entire *HoxA* cluster organization in each 3C library of the differentiation time course with 5C technology (Figure 2). The 5C libraries were generated with a previously validated 5C primer set and hybridized onto custom microarrays as described in 'Materials and Methods' section. To facilitate 5C data analysis, we created a pipeline, which integrates all of our previously published tools and the ones developed for this study. We used this new analysis platform to obtain normalized interaction frequency values for all pairwise interactions in the *HoxA* locus and control gene desert region, and visualized the data with the 'my5C' heatmap tool (31).

Figure 2A shows the *HoxA* 5C maps from our main time course (biological 1) and from a biological replicate (biological 2) in heatmap form where pairwise IFs are color-coded according to the scales under each row. A technical replicate of our main time course is also presented in Supplementary Figure S2. Two main observations can be made from these maps. First, lower IFs are correlated with greater gene expression from 0 to 12 h. This event is particularly visible closer to the diagonal where signals are naturally higher in this cluster. Although contacts appeared weaker throughout the cluster when transcription is activated, the regulated 5′-end region consistently exhibited the greatest loss of IFs from 0 to 12 h in all replicates (Figure 2B).

A second observation drawn from the 5C heatmaps is that the cluster appears to adopt a different 3D architecture on differentiation. The contacts do not simply revert to their original distribution from 12 to 96 h as shown in the bottom panels of each replicate. Instead, contacts appear to be gained, particularly between the cluster

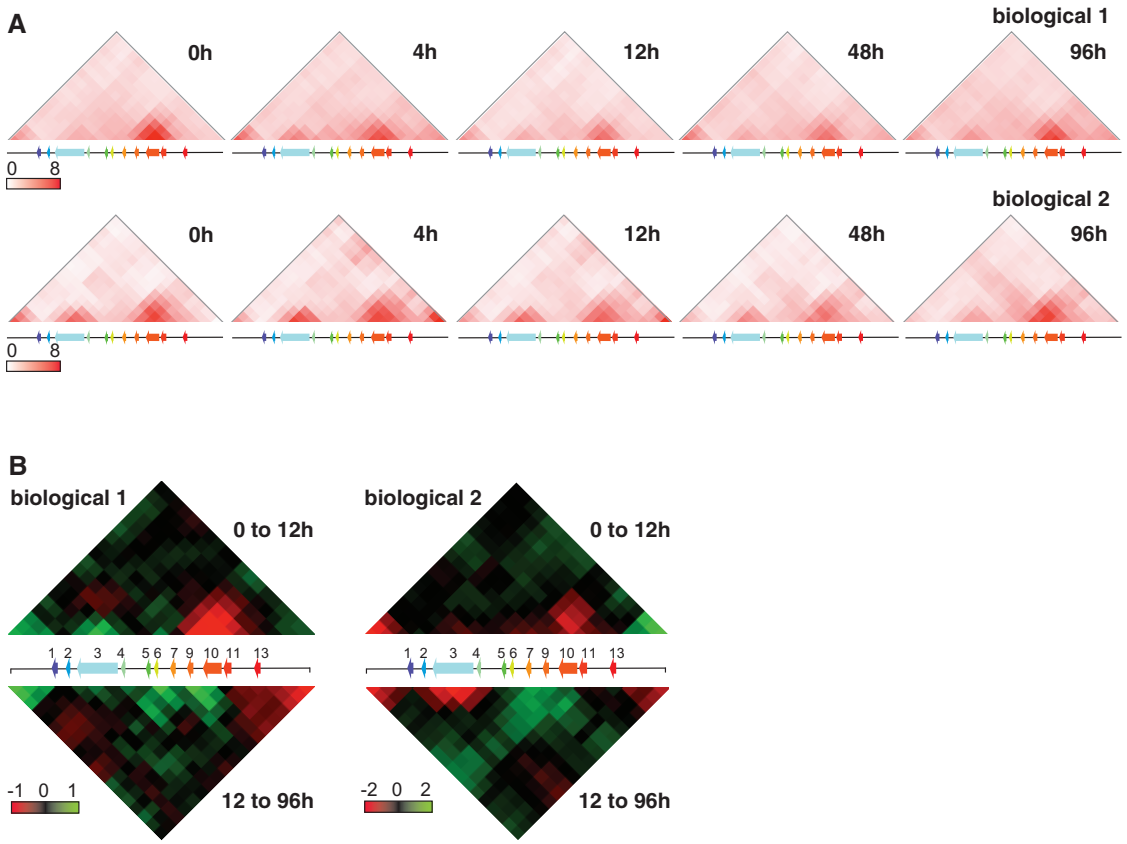

Figure 2. Higher 5′-end *HoxA* expression correlates with less chromatin contacts that differ on macrophage differentiation. (**A**) Pairwise IFs at the *HoxA* cluster vary throughout the differentiation time course. Heatmap representation of chromatin contacts measured with the 5C technology (8 kb bins, 8 kb smoothing). A linear diagram of the *HoxA* region characterized is shown to scale below each heatmap and is according to Figure 1A. DNA contacts are color-coded based on frequency, from low (white) to high (red) as indicated by the color scale below each row. Interaction frequencies were derived from at least three measurements as described in the 'Materials and Methods' section. Heatmaps were produced using the 'my5C' visualization tool (31). (**B**) The 5′-end DNA contacts decrease with activation of 5′-end *HoxA* genes and the interaction profile changes on macrophage differentiation. Changes in chromatin contacts associated with gene activation ('upper panels') and on macrophage differentiation ('lower panels') are shown in heatmap form. Heatmap values (8 kb bins, 8 kb smoothing) represent the difference of IFs between the time points indicated on the right of each panel. Reduced IFs are shown in red, and greater contacts appear in green as indicated by the color scales on the bottom left.

ends (Figure 2, panel B). Together, these results are consistent with a model whereby 5′-end *HoxA* gene activation is accompanied by opening of the cluster, and where differentiation occurs with repression and spatial reorganization of the cluster.

Sequential unfolding and spatial reorganization of the *HoxA* cluster accompany differentiation

This initial 5C data analysis shows that *HoxA* transcription regulation correlates with specific gain and loss of contacts. However, directly comparing 2D heatmaps as in Figure 2 can be noisy and limits observations to large changes of contact frequencies. For this reason, we previously developed the gradient descent '5C3D' modeling program to visualize the spatial distribution of contacts in 3D models (http://Dostielab.biochem.mcgill.ca). We then replaced 5C3D with 'MCMC5C', an improved program that can predict 3D models from the posterior probability distribution of structures using a Markov chain Monte Carlo approach (37). Like 5C3D, MCMC5C considers 5C IFs as inversely correlated with their original physical distances *in vivo,* and integrates all points to predict average models from cell populations.

We optimized MCMC5C as described in 'Materials and Methods' section, and used it to predict 3D models of the *HoxA* cluster and gene desert control from the 5C datasets of our differentiation time course (37). An ensemble of 200 models was generated from each 5C dataset of our time course. Figure 3A shows a superimposition of five randomly selected models visualized using PyMOL's integrated structural alignment utility. Visual inspection of the overlay of models that fit the observed interaction frequency data indicates a considerable degree of flexibility in the cluster, although each overlay points to a similar overall organization at each given time point. By assuming that interaction frequency is inversely proportional to the square of the physical distance *in vivo*, MCMC5C modeling supports our original 5C heatmap interpretation that the cluster is larger when transcription is activated 12 h post-induction. This was reproduced in our other replicates (Supplementary Figure S3), although it does not exclude the possibility that greater structural variability at 12 h contributes to lowering the signal. *HoxA*

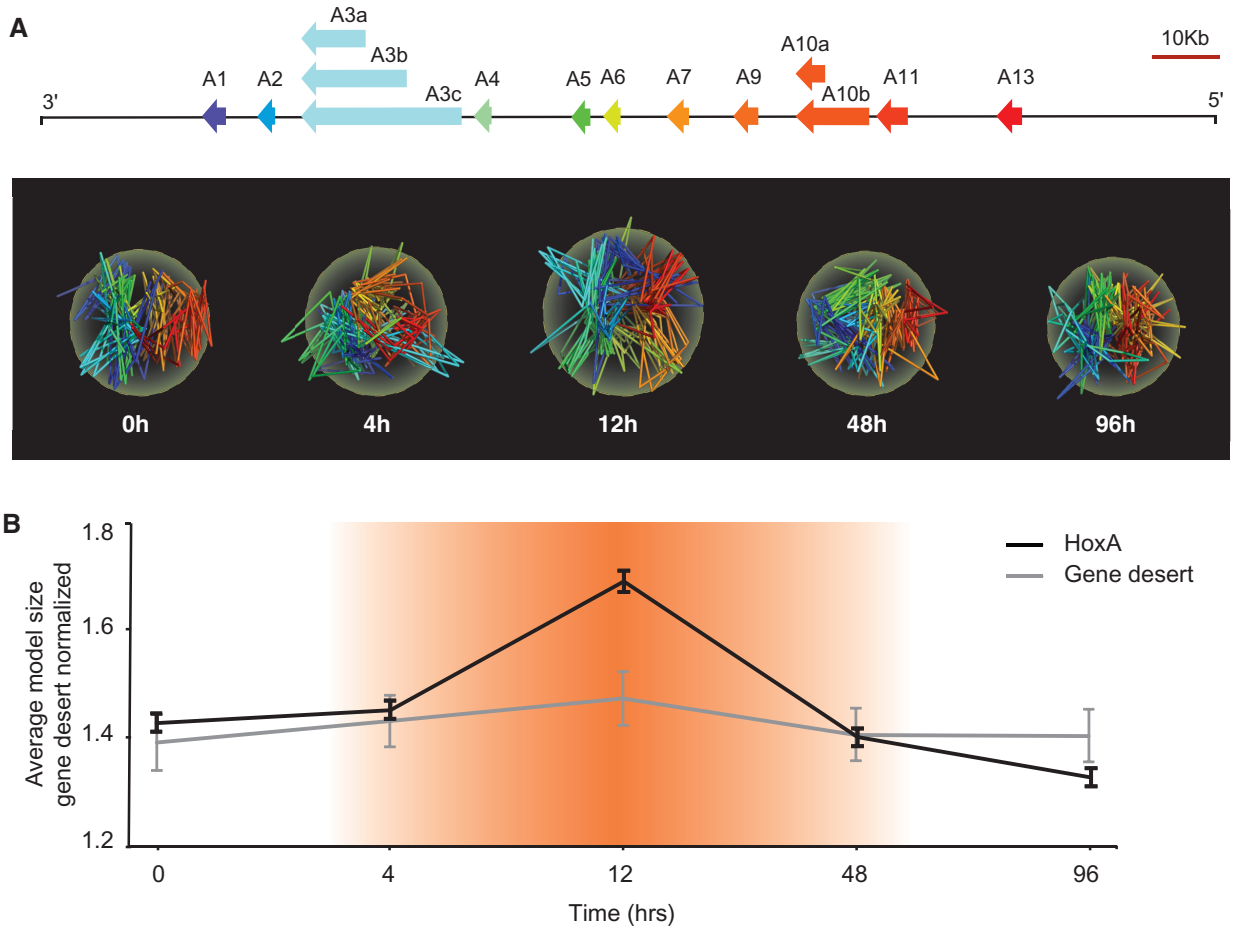

Figure 3. The 3D modeling reveals the transient unfolding of the *HoxA* cluster with gene activation and its reorganization on macrophage differentiation. (**A**) Spatial modeling of *HoxA* 5C datasets produces similar 3D *HoxA* models at individual differentiation time points. Overlay of five 3D *HoxA* models selected randomly from a pool of 200 variants generated with the MCMC5C program and visualized using PyMOL (37,38). Models are color-coded as shown in the linear *HoxA* diagram above, and aligned in the same 3′–5′-end orientation. The average model sizes estimated in (**B**) are highlighted with transparent yellow spheres. (B) The average *HoxA* cluster size is greater when 5′-end genes are activated. The average *HoxA* and gene desert model sizes were calculated at each time point across an ensemble of 200 models generated with the MCMC5C program. Model sizes were calculated as the average distance between all pairs of restriction fragments. Error bars represent the standard error of the mean.

modeling also revealed a different organization in myelomonocytes and in macrophages, although it appeared largely folded into two opposing domains in both cell types (Figure 3C). Specifically, the cluster 3′- and 5′-ends seemed closer to each other after differentiation. Thus, MCMC5C modeling is consistent with our earlier 3C and 5C analyses and supports a model whereby the cluster is first activated and unfolds before adopting a different organization on differentiation.

Extensive cluster unfolding precedes spatial reconfiguration of the *HoxA* middle domain in monocytes/macrophages

We next examined the ensembles of *HoxA* cluster models to identify specific conformational changes associated with either transcription regulation or cellular differentiation. Given the different transcriptional behavior of each half of the cluster, we wondered whether this response translated into divergent chromatin conformation changes. We first compared the average interaction frequency around the 3′- and 5′-ends but found no significant difference except at the 0-h time point where slightly fewer contacts were observed at the 3′ region (Figure 4A). We next developed the 'StructureAnalyser' program to measure 3D Euclidean pairwise distances between any given features in MCMC5C ensembles as described in the 'Materials and Methods' section. Using this program, we could not find any notable difference of average model size between both halves of the cluster (Figure 4B, and Supplementary Figure S3A and B). When we measured the average promoter distance between genes at the 3′- and 5′-ends in our main time course, the distance between 5′-end genes appeared to follow the overall transcription dynamics (Figure 4C). They were closer to each other when repressed (e.g. 4 h) and farther when active (12 h). The *HoxA* 3′-end genes did not follow this pattern, and changes were not as pronounced as for the 5′ genes. However, the rather large variability associated with this type of measurement prevented us from observing this behavior in our replicates (Supplementary Figure S3C and D), and suggests

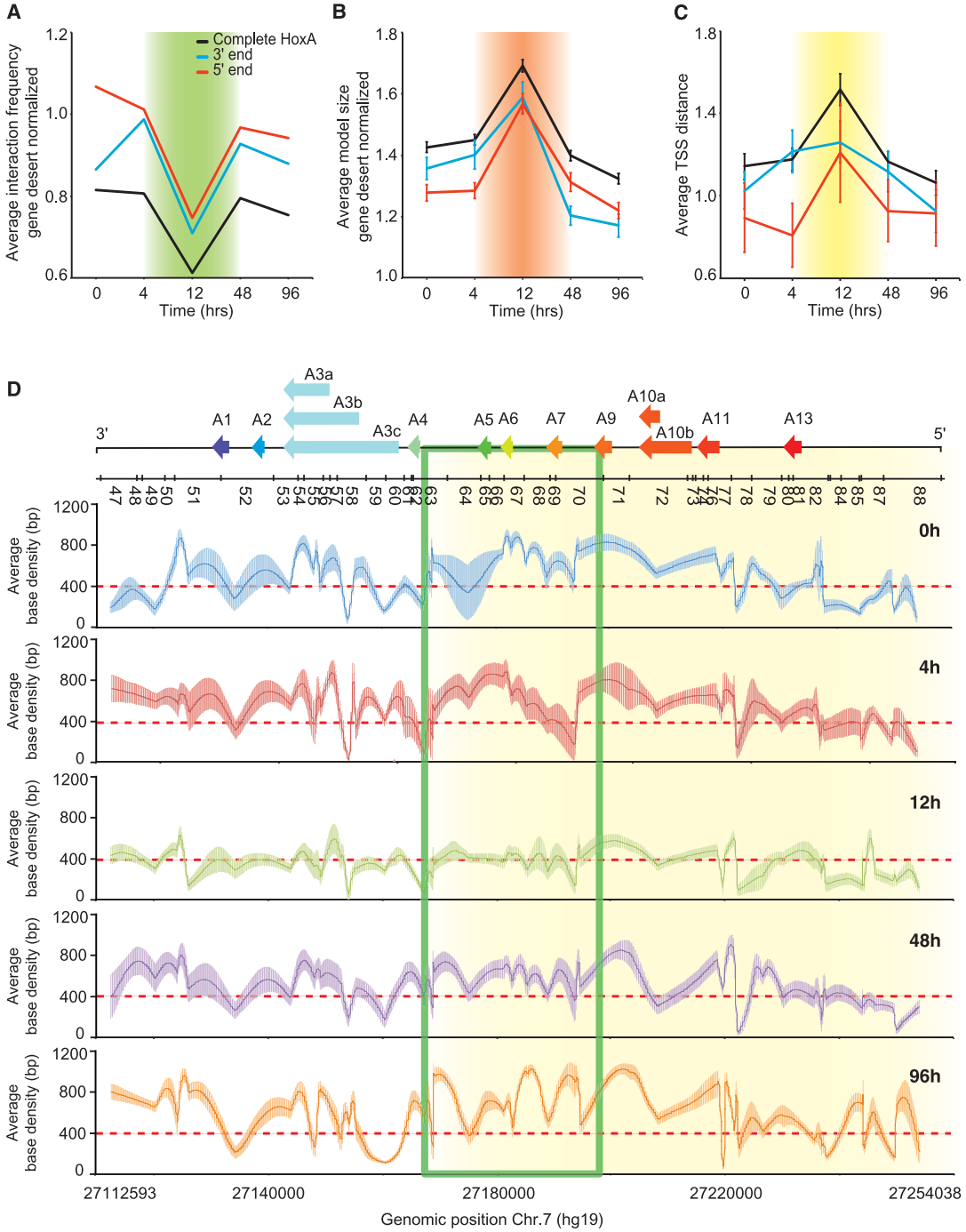

Figure 4. Complete cluster unfolding on 5′-end gene activation precedes spatial reconfiguration of the *HoxA* middle domain in monocytes/macrophages. (**A**) The *HoxA* 3′- and 5′-ends show similar changes of average IFs throughout cellular differentiation. The average 3′- and 5′-end IFs were calculated from the normalized pairwise IFs of the main time course (biological 1) shown in Figure 2A. *HoxA* cluster 3′-end values include all interactions between fragments 47 and 64, and the 5′-end measurements encompass contacts between fragments 65 and 88. The 5′-end region is highlighted in yellow in (**D**). (**B**) The *HoxA* cluster 3′- and 5′-ends show similar changes in size throughout the differentiation time course. The *HoxA* cluster 3′- and 5′-end regions are as described in (A). The 3′- and 5′-end sizes were calculated from 200 models. (**C**) The spatial distance between transcriptionally regulated 5′-end genes correlates with gene expression changes. *HoxA1* to *A7* represent the 3′-end genes, and 5′-end genes include *HoxA9* to *A13*. Average distances were derived from 200 models. Error bars in (B) and (C) represent the standard error of the mean. (D) The average *HoxA* local base density decreases across the cluster when 5′-end gene expression increases, and adopts a different profile on macrophage differentiation. Local base densities (*y-axis*) were estimated every 100 bp along the *HoxA* cluster region (*x-axis*) in 200 models with the Microcosm 2.0 program (19). The width of the trace corresponds to the standard deviation with sharper areas corresponding to smaller deviations. The green box highlights the middle *HoxA* region showing greatest base density changes throughout differentiation. Red dashed lines are used as guides to facilitate comparison.

that the genes do not play an important role in controlling how the cluster is organized.

We next measured the average local base density along the *HoxA* cluster with 'Microcosm 2.0', a program that estimates DNA concentration around chromatin features in a given ensemble of 3D models (19). The local base density scans generated by Microcosm 2.0 revealed that the entire cluster unfolds, as 5′-end genes are transiently activated 12 h post-induction (Figure 4D, compare density values at 0 and 12 h). We also found that the general base density pattern along the cluster does not significantly change except at a middle region (boxed in green), where it completely changed between myelomonocytes and monocytes/macrophages. This result was reproduced in our biological replicate (Supplementary Figure S4A), and is interesting partly because such changes could only be found by probing chromatin organization in 3D models. Differences of IFs are difficult to observe in 5 C heatmaps when they extend over several pairwise interactions, especially here, as contacts are not particularly strong (see Figure 2A and B). This observation is also interesting because reconfiguration occurs at the junction between the two cluster segments, which we observed to behave as though independent of each other transcriptionally (Figure 1C). This region corresponds to a boundary between two 'topologically associated domains' (TADs) observed in human IMR90 and mouse embryonic stem cells (44). TADs are recently identified basic units of chromatin organization that are detectable at the megabase-scale and which are thought largely conserved between cell types and across species. The boundaries of TADs were suggested to participate in establishing the topological domain structure of genomes and were found enriched in retrotransposons, transfer RNA, housekeeping genes and for the insulator binding protein CTCF. We found that CTCF binds to this region at sites corresponding to each of the four base density peaks at 96 h, suggesting that it might participate in compartmentalizing the cluster as described later in Figure 6. Together, these results indicate that complete unfolding of the cluster followed by chromatin reorganization at the middle domain accompanies THP-1 differentiation.

3′ enrichment and 5′ depletion of H3K27me2/3 at the *HoxA* cluster correlates with selective 5′-end gene activation

Because the entire *HoxA* cluster appears to unfold at the 12-h time point, we wondered how 5′-end genes could specifically be induced. To address this question, we measured the level of methylation on histone H3 at lysine 27 (H3K27me2/3) in the five samples of our time course (Figure 5A). *Hox* clusters are important targets of polycomb group (PcG) complexes, which silence transcription in part by modifying histones epigenetically with marks that include H3K27me2/3 and H2A ubiquitination at lysine 119 (45). We found that selective transcription activation correlates with the progressive enrichment of H3K27me2/3 throughout the cluster except over the region spanning the activated *HoxA9*, *10*, *11* and *13* genes where it was severely depleted. Repression by H3K27me2/3 encroached over the 5′-end region at 48 and 96 h post-induction, as the cells differentiated and the genes became repressed. The resulting pattern and levels of H3K27me2/3 along *HoxA* in monocyte/macrophages were different compared with undifferentiated myelomonocytes and were consistent with transcriptional silencing and packaging of the cluster following differentiation (compare 0 and 96 h in Figure 4D).

To obtain an integrated view of the epigenomic and structural changes occurring at the cluster throughout differentiation, we created the 'StructureAnnotation' program, which can superimpose specified chromatin features over given 3D models. StructureAnnotation outputs files that can be viewed with open-source software such as PyMOL (38). Figure 5B illustrates the position of H3K27me2/3 relative to gene TSSs in randomly selected MCMC5C 3D *HoxA* models. In this analysis, the *HoxA* TSSs are first observed sequentially from left to right in actively growing cells (0 h). The position of genes is then slightly reorganized, as the cluster progresses to the 12-h time point where all the inactive genes spatially localize within a domain repressed by H3K27me2/3, whereas those actively transcribed fall to one side devoid of the repression mark. Transition from the 12- to 96-h time point is marked by spreading of H3K27me2/3 and corresponds to an overall folding and reorganization of the cluster. These results, which illustrate a possible mutual influence between histone modification and chromatin conformation, support our analysis of the gene expression and 5C data presented earlier.

These data also result in questioning what might be responsible for maintaining the contacts that are lost on transient 5′ gene activation and for reorganizing the chromatin at the middle *HoxA* region in monocytes/macrophages. We previously examined the role of PcG proteins in maintaining the *HoxA* cluster organization in NT2/D1 cells and found that they were only partially required for looping and silencing (19). PcG proteins were shown to mediate long-range contacts in *Drosophila BX-C*, and to form higher order chromosome conformations in human cells (46–48). These proteins were also found to induce chromatin compaction in a non-catalytic manner making them attractive regulatory candidates at the *HoxA* cluster (49–51). For instance, PcG proteins might participate in structurally maintaining the silent genes together within a repressed domain as it appears at 12 h and following differentiation (Figure 5B). However, here, as in NT2/D1 cells, PcG complexes are not likely entirely responsible for the chromatin conformation changes for at least two reasons. First, the fact that H3K27me2/3 increases throughout the cluster except where the genes are induced is not consistent with the specific reconfiguration observed in the middle HoxA region. Second, H3K27me2/3 distribution along the cluster is broad and does not correlate with any particular structural *HoxA* feature, which is inconsistent with a particular spatial function. Proteins involved in *HoxA* conformation changes therefore remain unaccounted for.

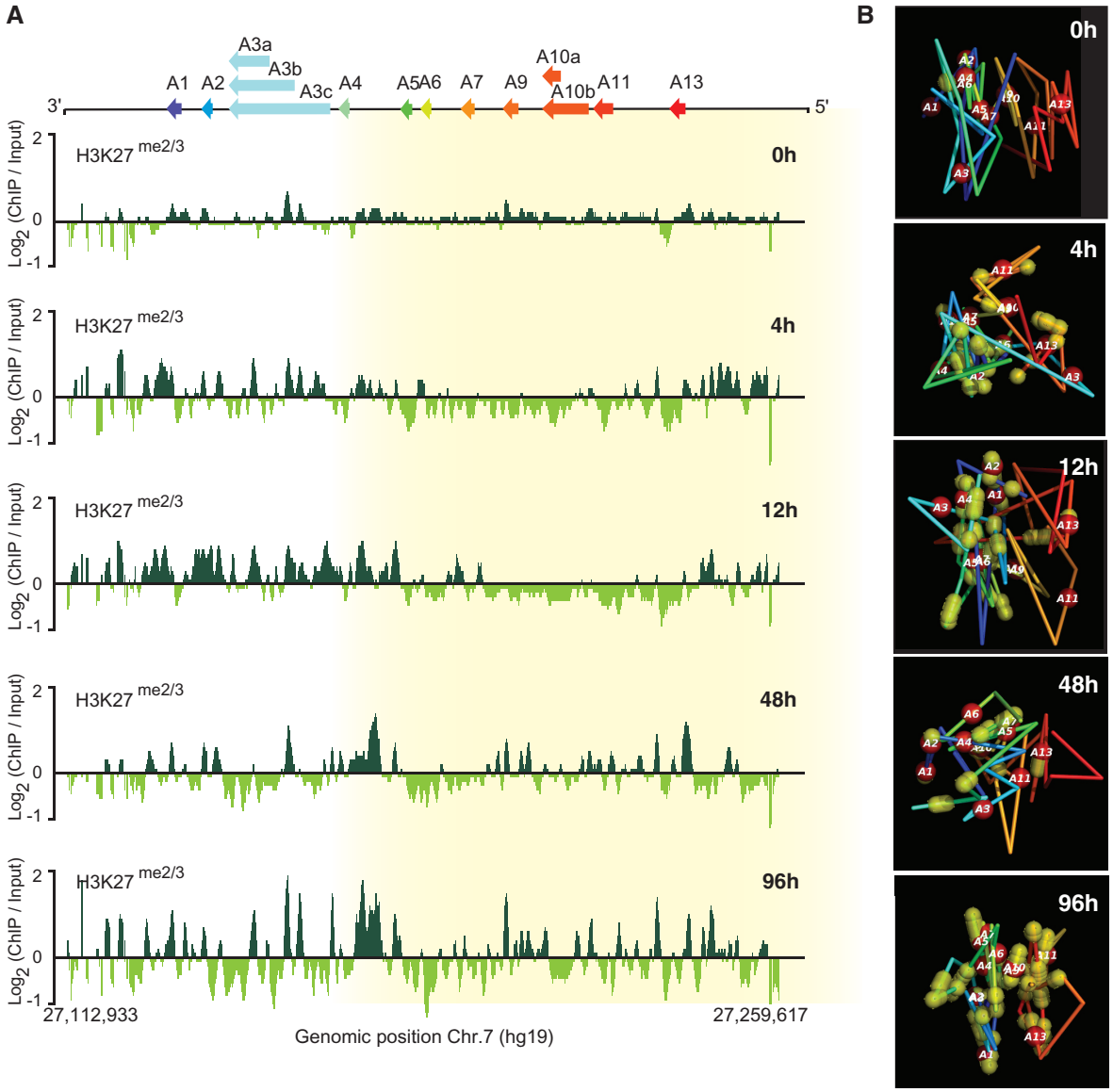

Figure 5. The 3′ enrichment and 5′ depletion of H3K27me2/3 at the *HoxA* cluster correlates with selective 5′-end gene activation. (**A**) Selective 5′-end gene activation correlates with the local depletion of H3K27me2/3 and its enrichment at the cluster 3′-end. H3K27me2/3 distribution along the *HoxA* cluster throughout the THP-1 differentiation time course. ChIPed material and input were hybridized onto custom tilling microarrays and normalized as described in 'Materials and Methods' section. (**B**) H3K27me2/3 marks cover the repressed *HoxA* cluster 3′-end and are depleted from the transcriptionally induced 5′-region in 3D models. The annotated 3D *HoxA* models are shown for each time point as described in Figure 3. Models were generated with MCMC5C, annotated with the *HoxA* TSSs (TSSs) and repression H3K27me2/3 marks and visualized using PyMOL. Red and yellow spheres represent TSSs and H3K27me2/3 marks, respectively. Only H3K27me2/3 signals above a threshold with a cutoff = 0.8 are shown.

Spatial proximity changes between CTCF binding sites are consistent with a role for insulator looping in regulating the *HoxA* cluster architecture

We recently used 3D modeling to identify candidate proteins that mediate *HoxA* chromatin loops in NT2/D1 cells (19). In this study, 3D models generated with the 5C3D program were analyzed with Microcosm 2.0 and local base density scans were aligned against available genome-wide histone modification and protein-binding datasets. With this approach, we identified CTCF as a candidate mediator of chromatin loops in the silent *HoxA* cluster in NT2/D1 cells (19). To determine whether CTCF might be involved in *HoxA* conformation changes during THP-1 differentiation, we first mapped CTCF binding in undifferentiated (0 h) and monocyte/macrophage-differentiated (96 h) THP-1 cells (Figure 6A). CTCF binding is highly conserved between cell types, and accordingly, the seven binding sites we identified in THP-1 did not change between 0 and 96 h, and have been observed in other cell types (19,52). Except for site CTCF7, at least one CTCF binding motif could be predicted at each site (53). We retained CTCF7 because it is found in all cell lines characterized so far.

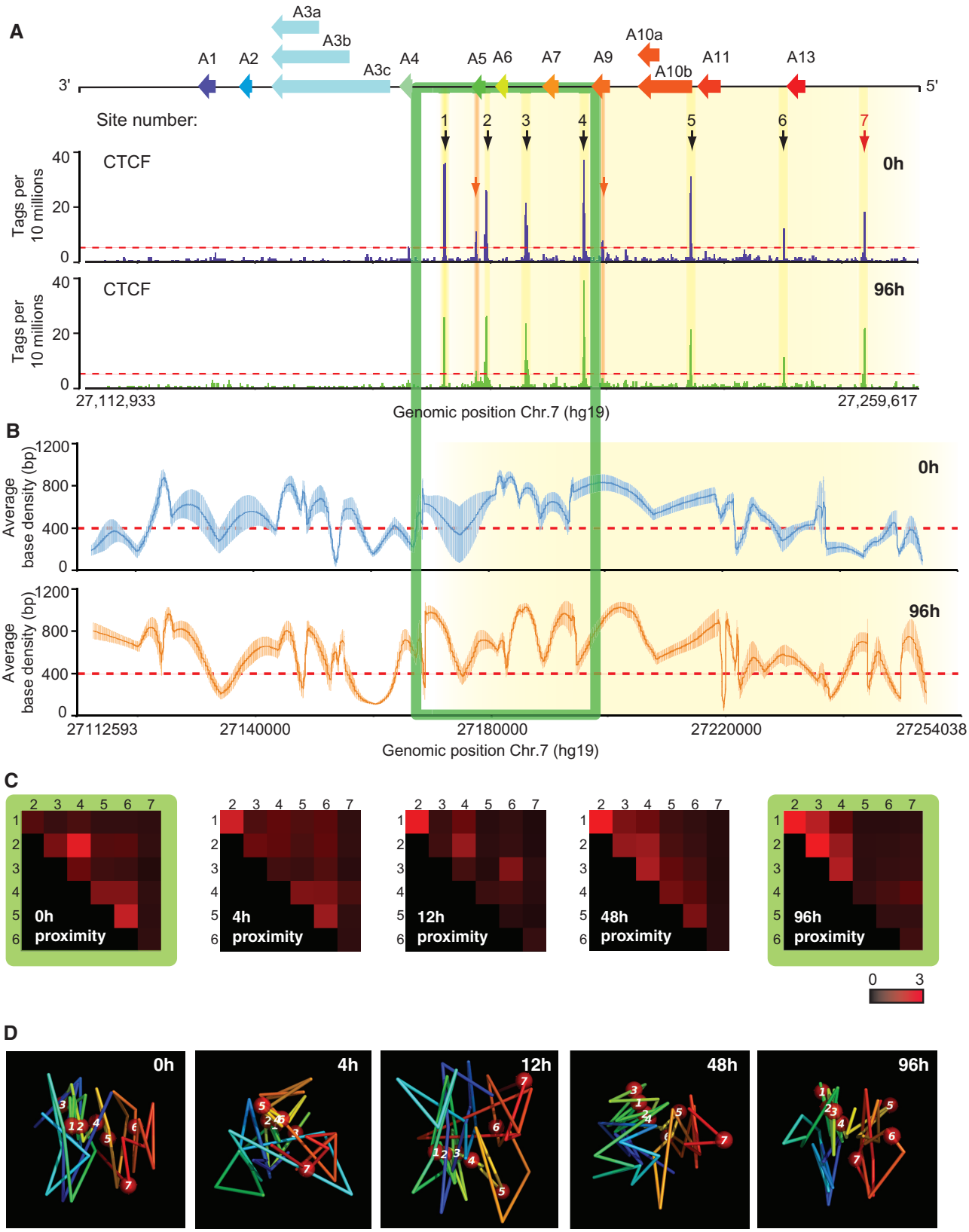

Figure 6. Spatial proximity changes between CTCF binding sites are consistent with a role for insulator looping in regulating the *HoxA* cluster architecture. (**A**) CTCF binds at the *HoxA* cluster 5′-end in myelomonocytes and macrophages. CTCF was ChIPed before (0 h) and after (96 h) differentiation of THP-1 cells. ChIPed material was sequenced and analyzed as described in 'Materials and Methods' section. The y-axis shows the number of CTCF 'Tags per 10 millions' after normalization against input, across the region characterized (*x-axis*). CTCF peaks are numbered from left to right (CTCF1 to 7). CTCF7 lacks a consensus binding sequence and is highlighted in red. The two orange arrows point to two minor peaks

(continued)

As mentioned above, some CTCF sites overlap with peaks in the local base density scans (Figure 6B). These include CTCF 1 to 4 located in the middle domain of *HoxA* boxed in green where extensive spatial reorganization occurs on differentiation. This observation suggests that CTCF dimerization or multimerization might assist conformational changes at the *HoxA* cluster. We measured the average CTCF pairwise distances in *HoxA* MCMC5C ensembles at 0 and 96 h, and also predicted their spatial distances at the remaining three connecting time points (Figure 6C and D, Supplementary Table S2). We found that sites 2/4 in the middle domain, and 4/5/6 tended to be clustered at 0 h. Transcription activation at 12 h correlated with an overall reduction in CTCF proximity with the exception of CTCF 2/4, which appeared to stay close to each other. These early proximity changes were followed by progressive clustering of CTCF sites, particularly 1–4 at 48 and 96 h post-induction, which was also observed in the biological replicate (Supplementary Figure S4B). Together, these analyses support a role for CTCF in regulating the architecture of the *HoxA* cluster in THP-1 cells.

Hox in motion

To visualize the spatial connectivity of CTCF binding sites throughout our differentiation time course, we created 'MovieMaker' to produce movies from annotated models. 'MovieMaker' simply concatenates MCMC5C models annotated with StructureAnnotation, and connects each time point by inserting intermediary models inferred from the shortest path between them. We generated a movie that integrates the *HoxA* conformation (5C), H3K27me2/3 (ChIP-chip) and CTCF binding (ChIP-seq) datasets generated for our time course (Supplementary Movie S1). In this movie, CTCF binding sites are first observed to localize consecutively from left to right in myelomonocytes (0 h) except for CTCF7, which is located outside of the cluster. CTCF1 to 6 are then grouped at the 4-h time point when transcription is repressed and when H3K27me2/3 levels start to rise.

Interestingly, CTCF1 to 4 were observed to stay together in a domain enriched in H3K27me2/3 at 12 h, whereas CTCF5/6/7 extended outside of the repressed region where transiently activated 5′ genes are located. The CTCF1 to 4 sites remained together on one side of the structure as cells transited from 12 to 96 h post-induction and the repression mark spread throughout the cluster until reaching its final configuration in differentiated monocytes/macrophages. We also generated a movie illustrating the position of *HoxA* TSSs relative to the experimental conformation (5C) and H3K27me2/3 (ChIP-chip) datasets (Supplementary Movie S2). This movie illustrates the position of genes relative to dynamic changes in chromatin structure and supports a model whereby *HoxA* gene induction is associated with the loss of chromatin contacts, possibly between CTCF-bound insulators.

DISCUSSION

In this study, we provide the first integrated view of a relationship between chromatin activity, epigenomics and architecture, which is involved in regulating *HoxA* gene expression during cellular differentiation. We found that the 5′-end *HoxA* genes are transiently induced 12 h after differentiation is triggered, and that higher expression correlates with lower chromatin contacts. By abiding to the current notion that IFs proportionally reflect physical distances between chromatin fragments *in vivo*, we interpreted this loss of contacts as unfolding of the chromatin and the appearance of a larger cluster. However, our data do not exclude the possible contribution of a greater structural variability in diluting the 3C and 5C signals, as was previously reported at the inactive X chromosome (54). Differential fixation, cutting or ligation at the 12-h time point may also alter the frequency at which contacts are captured, although the fact that a pattern of base densities resembling the ones observed at 0 and 4 h is maintained at 12 h with lower valleys and smaller peaks (Figure 4D) argues against these factors being largely responsible for lowering the signals.

Our finding that the entire *HoxA* cluster seems to unfold while only the 5′-end genes are considerably induced can be explained by the specific enrichment of H3K27me2/3 repression marks at the 3′-end of the cluster. The progressive appearance of H3K27me2/3 from 0 to 12 h is interesting given that the modification is deposited and bound by PcG proteins, which are also known to play a structural role by mediating long-range chromatin interactions and regulating compaction (46–51). Therefore, this result suggests that recruitment of PcG proteins at the 3′-end might serve the dual role of repressing transcription and compartmentalizing silenced genes into an inactive domain held together by PcG. This would point to a direct link between epigenomics and chromatin architecture whereby the epigenomic state of chromatin and its conformation could mutually affect each other. Such concept has recently been put forward as the driving force in the establishment, maintenance and propagation of genome organization (55).

We found that the cluster adopts a different conformation in differentiated cells where the 3′- and 5′-ends are

Figure 6. Continued
without predicted consensus binding sites. The green box indicates the middle *HoxA* region with greatest base density changes. (**B**) *HoxA* local base density scans at the 0 and 96-h time points from Figure 4 are shown to situate CTCF binding and base density. (**C**) The predicted spatial proximity of neighboring CTCF binding sites in the *HoxA* middle domain correlates with the base density changes observed on cellular differentiation. Predicted pairwise spatial distances between the seven CTCF binding sites identified in (A) were estimated from 200 structures, converted into proximity (1/distance), and represented in heatmap form. Values of spatial proximity are color-coded as indicated at the bottom left of the heatmap set. (**D**) The four CTCF binding sites at the *HoxA* middle domain are spatially clustered in macrophages. Annotated 3D *HoxA* models show the predicted position of CTCF binding sites during the differentiation time course. Red spheres represent CTCF binding sites.

closer to each other. A computational analysis of the conformational changes taking place during the time course identified the middle domain of the *HoxA* cluster as the main reorganized region. This region corresponds to a boundary between two TADs described previously in other cell lines (44), and like TAD boundaries in general, the region was enriched in sites bound by CTCF. The fact that this boundary separates the 3′- and 5′-end genes, which behaved independently of each other transcriptionally, and localizes at the edge of the 3′-end region enriched in H3K27me2/3 at the 12-h time point suggests that CTCF sites within this region might function as insulators. This model is supported by the recent report that CTCF is required for selective gene activation and heterochromatin partitioning at the *HoxA* cluster (56).

Interestingly, the CTCF sites at the boundary became closer to each other as the cells differentiated (Figure 6C), suggesting that looping between CTCF-bound insulator sequences may contribute to shaping the cluster, perhaps by inducing a kink at the boundary region that effectively brings the two TADs closer to each other. Therefore, this result indicates that TAD boundaries can undergo significant reorganization during cellular differentiation, and that, in general, regulated looping between CTCF-bound sites within these regions might alter the proximity between TADs, thereby changing the topology of genomes, which are otherwise largely conserved between cell types and across species at this resolution.

An important question remaining relates to why the 5′-end *HoxA* genes are transiently induced during differentiation. One possible explanation is that the transcription factors are required at that time to take part in a network regulating genes important in the differentiation process. Alternatively, a short transcription pulse might participate in changing the cluster organization and how CTCF-bound sites interact with each other. Under this scenario, the induction of 5′-end genes would trigger extensive unfolding of the cluster, whereas repression of the 3′-end genes would recruit PcG proteins and form an inactive compartment distinct from the 5′-end. Unfolding of the cluster would involve a loss of CTCF interactions, and spreading of the H3K27me2/3 beyond the boundary domain would be prevented by CTCF's insulator activity. On inactivation of the 5′-end genes beyond the 12-h time point, the cluster would fold into a different configuration driven in part by the realignment of CTCF interactions, and the previous segregation of the 3′-end into a PcG-coated domain. The new conformation adopted by the cluster in differentiated cells might make it more or less prone to regulation by certain types of signals, and to its aberrant activation, which can lead to human disease such as cancer.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of our laboratory for insightful discussions. We are indebted to M. A. Ferraiuolo for excellent technical assistance. We are grateful to Dr. J. Teodoro for critical reading of this manuscript.

FUNDING

This work was supported by the Canadian Institutes of Health Research [CIHR MOP-86716 to J.D.], and the Canadian Cancer Society Research Institute [The Terry Fox Foundation; CCSRI 019252]; Supported by a scholarship from the National Sciences and Engineering Research Council (NSERC) (to M.R.); CIHR New Investigator and FRSQ Research Scholar (to J.D.). Funding for open access charge: Canadian Institutes of Health Research Grant [MOP-86716] and NSERC Discovery grant (to M.B.).

Conflict of interest statement. None declared.

REFERENCES

1. Fraser,P. and Bickmore,W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413–417.
2. Babu,M.M., Janga,S.C., de Santiago,I. and Pombo,A. (2008) Eukaryotic gene regulation in three dimensions and its impact on genome evolution. *Curr. Opin. Genet. Dev.*, **18**, 571–582.
3. West,A.G. and Fraser,P. (2005) Remote control of gene transcription. *Hum. Mol. Genet.*, **14**, R101–R111.
4. Woodcock,C.L. (2006) Chromatin architecture. *Curr. Opin. Struct. Biol.*, **16**, 213–220.
5. Gondor,A. and Ohlsson,R. (2009) Chromosome crosstalk in three dimensions. *Nature*, **461**, 212–217.
6. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing Chromosome Conformation. *Science*, **295**, 1306–1311.
7. Simonis,M., Klous,P., Splinter,E., Moshkin,Y., Willemsen,R., de Wit,E., van Steensel,B. and de Laat,W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.*, **38**, 1348–1354.
8. Zhao,Z., Tavoosidana,G., Sjolinder,M., Gondor,A., Mariano,P., Wang,S., Kanduri,C., Lezcano,M., Singh Sandhu,K., Singh,U. *et al.* (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.*, **38**, 1341–1347.
9. Ling,J.Q., Li,T., Hu,J.F., Vu,T.H., Chen,H.L., Qiu,X.W., Cherry,A.M. and Hoffman,A.R. (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, **312**, 269–272.
10. Würtele,H. and Chartrand,P. (2006) Genome-wide scanning of HoxB1- associated loci in mouse ES cells using an open-ended chromosome conformation capture methodology. *Chromosome Res.*, **14**, 477–495.
11. Dostie,J. and Dekker,J. (2007) Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.*, **2**, 988–1002.
12. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
13. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y.J., Lee,C., Shendure,J., Fields,S., Blau,C.A. and Noble,W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
14. Kalhor,R., Tjong,H., Jayathilaka,N., Alber,F. and Chen,L. (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.*, **30**, 90–98.

15. Calvo,R., West,J., Franklin,W., Erickson,P., Bemis,L., Li,E., Helfrich,B., Bunn,P., Roche,J., Brambilla,E. *et al.* (2000) Altered HOX and WNT7A expression in human lung cancer. *Proc. Natl Acad. Sci. USA*, **97**, 12776–12781.
16. Makiyama,K., Hamada,J., Takada,M., Murakawa,K., Takahashi,Y., Tada,M., Tamoto,E., Shindo,G., Matsunaga,A., Teramoto,K. *et al.* (2005) Aberrant expression of HOX genes in human invasive breast carcinoma. *Oncol. Rep.*, **13**, 673–679.
17. Maeda,K., Hamada,J., Takahashi,Y., Tada,M., Yamamoto,Y., Sugihara,T. and Moriuchi,T. (2005) Altered expressions of HOX genes in human cutaneous malignant melanoma. *Int. J. Cancer*, **114**, 436–441.
18. Eklund,E.A. (2007) The role of HOX genes in malignant myeloid disease. *Curr. Opin. Hematol.*, **14**, 85–89.
19. Ferraiuolo,M.A., Rousseau,M., Miyamoto,C., Shenker,S., Wang,X.Q., Nadler,M., Blanchette,M. and Dostie,J. (2010) The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.*, **38**, 7472–7484.
20. Noordermeer,D., Leleu,M., Splinter,E., Rougemont,J., De Laat,W. and Duboule,D. (2011) The dynamic architecture of Hox gene clusters. *Science*, **334**, 222–225.
21. Fraser,J., Rousseau,M., Shenker,S., Ferraiuolo,M.A., Hayashizaki,Y., Blanchette,M. and Dostie,J. (2009) Chromatin conformation signatures of cellular differentiation. *Genome Biol.*, **10**, R37.
22. Wang,K.C., Yang,Y.W., Liu,B., Sanyal,A., Corces-Zimmerman,R., Chen,Y., Lajoie,B.R., Protacio,A., Flynn,R.A., Gupta,R.A. *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, **472**, 120–124.
23. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
24. Miele,A., Gheldof,N., Tabuchi,T.M., Dostie,J. and Dekker,J. (2006) In: Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*, Vol. Supplement 74. John Wiley & Sons, Hoboken, N.J., pp. 21.11.21–21.11–20.
25. Dostie,J., Richmond,T.A., Arnaout,R.A., Selzer,R.R., Lee,W.L., Honan,T.A., Rubio,E.D., Krumm,A., Lamb,J., Nusbaum,C. *et al.* (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, **16**, 1299–1309.
26. Dostie,J., Zhan,Y. and Dekker,J. (2007) Chromosome conformation capture carbon copy technology. *Curr. Protoc. Mol. Biol.*, **Chapter 21**, Unit 21 14.
27. Nuwaysir,E.F., Huang,W., Albert,T.J., Singh,J., Nuwaysir,K., Pitas,A., Richmond,T., Gorski,T., Berg,J.P., Ballin,J. *et al.* (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.*, **12**, 1749–1755.
28. Kim,T.H., Barrera,L.O., Zheng,M., Qu,C., Singer,M.A., Richmond,T.A., Wu,Y., Green,R.D. and Ren,B. (2005) A high-resolution map of active promoters in the human genome. *Nature.*, **436**, 876–880.
29. Selzer,R.R., Richmond,T.A., Pofahl,N.J., Green,R.D., Eis,P.S., Nair,P., Brothman,A.R. and Stallings,R.L. (2005) Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer*, **44**, 305–319.
30. Fraser,J., Rousseau,M., Blanchette,M. and Dostie,J. (2010) Computing chromosome conformation. *Methods Mol. Biol.*, **674**, 251–268.
31. Lajoie,B.R., van Berkum,N.L., Sanyal,A. and Dekker,J. (2009) My5C: web tools for chromosome conformation capture studies. *Nat. Methods*, **6**, 690–691.
32. Rinn,J.L., Kertesz,M., Wang,J.K., Squazzo,S.L., Xu,X., Brugmann,S.A., Goodnough,L.H., Helms,J.A., Farnham,P.J., Segal,E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.
33. Toedling,J., Skylar,O., Krueger,T., Fischer,J.J., Sperling,S. and Huber,W. (2007) Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics*, **8**, 221.
34. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
35. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
36. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
37. Rousseau,M., Fraser,J., Ferraiuolo,M.A., Dostie,J. and Blanchette,M. (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics*, **12**, 414.
38. The PyMOL Molecular Graphics System, Version 1.5.0.4. Schrödinger, LLC.
39. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
40. Goecks,J., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
41. Blankenberg,D., Von Kuster,G., Coraor,N., Ananda,G., Lazarus,R., Mangan,M., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.*, **Chapter 19**, Unit 19 10 11–21.
42. Suzuki,H., Forrest,A.R., van Nimwegen,E., Daub,C.O., Balwierz,P.J., Irvine,K.M., Lassmann,T., Ravasi,T., Hasegawa,Y., de Hoon,M.J. *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
43. Ferraiuolo,M.A., Sanyal,A., Naumova,N., Dekker,J. and Dostie,J. (2012) From cells to chromatin: capturing snapshots of genome organization with 5C technology. *Methods*, **58**, 255–267.
44. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
45. Bantignies,F. and Cavalli,G. (2011) Polycomb group proteins: repression in 3D. *Trends Genet.*, **27**, 454–464.
46. Lanzuolo,C., Roure,V., Dekker,J., Bantignies,F. and Orlando,V. (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.*, **9**, 1167–1174.
47. Tiwari,V.K., Cope,L., McGarvey,K.M., Ohm,J.E. and Baylin,S.B. (2008) A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.*, **18**, 1171–1179.
48. Tiwari,V.K., McGarvey,K.M., Licchesi,J.D., Ohm,J.E., Herman,J.G., Schubeler,D. and Baylin,S.B. (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.*, **6**, 2911–2927.
49. Eskeland,R., Leeb,M., Grimes,G.R., Kress,C., Boyle,S., Sproul,D., Gilbert,N., Fan,Y., Skoultchi,A.I., Wutz,A. *et al.* (2010) Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell*, **38**, 452–464.
50. Simon,J.A. and Kingston,R.E. (2009) Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*, **10**, 697–708.
51. Francis,N.J., Kingston,R.E. and Woodcock,C.L. (2004) Chromatin compaction by a polycomb group protein complex. *Science*, **306**, 1574–1577.
52. Schmidt,D., Schwalie,P.C., Ross-Innes,C.S., Hurtado,A., Brown,G.D., Carroll,J.S., Flicek,P. and Odom,D.T. (2010) A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.*, **20**, 578–588.
53. Bao,L., Zhou,M. and Cui,Y. (2008) CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Res.*, **36**, D83–D87.
54. Splinter,E., de Wit,E., Nora,E.P., Klous,P., van de Werken,H.J., Zhu,Y., Kaaij,L.J., van Ijcken,W., Gribnau,J., Heard,E. *et al.*

(2011) The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.*, **25**, 1371–1383.
55. Cavalli,G. and Misteli,T. (2013) Functional implications of genome topology. *Nat. Struct. Mol. Biol.*, **20**, 290–299.
56. Kim,Y.J., Cecchini,K.R. and Kim,T.H. (2011) Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl Acad. Sci. USA*, **108**, 7391–7396.