

Topological domains in mammalian genomes identified by analysis of chromatin interactions

Jesse R. Dixon^{1,2,3}, Siddarth Selvaraj^{1,4}, Feng Yue¹, Audrey Kim¹, Yan Li¹, Yin Shen¹, Ming Hu⁵, Jun S. Liu⁵ & Bing Ren^{1,6}

The spatial organization of the genome is intimately linked to its biological function, yet our understanding of higher order genomic structure is coarse, fragmented and incomplete. In the nucleus of eukaryotic cells, interphase chromosomes occupy distinct chromosome territories, and numerous models have been proposed for how chromosomes fold within chromosome territories¹. These models, however, provide only few mechanistic details about the relationship between higher order chromatin structure and genome function. Recent advances in genomic technologies have led to rapid advances in the study of three-dimensional genome organization. In particular, Hi-C has been introduced as a method for identifying higher order chromatin interactions genome wide². Here we investigate the three-dimensional organization of the human and mouse genomes in embryonic stem cells and terminally differentiated cell types at unprecedented resolution. We identify large, megabase-sized local chromatin interaction domains, which we term ‘topological domains’, as a pervasive structural feature of the genome organization. These domains correlate with regions of the genome that constrain the spread of heterochromatin. The domains are stable across different cell types and highly conserved across species, indicating that topological domains are an inherent property of mammalian genomes. Finally, we find that the boundaries of topological domains are enriched for the insulator binding protein CTCF, housekeeping genes, transfer RNAs and short interspersed element (SINE) retrotransposons, indicating that these factors may have a role in establishing the topological domain structure of the genome.

To study chromatin structure in mammalian cells, we determined genome-wide chromatin interaction frequencies by performing the Hi-C experiment² in mouse embryonic stem (ES) cells, human ES cells, and human IMR90 fibroblasts. Together with Hi-C data for the mouse cortex generated in a separate study (Y. Shen *et al.*, manuscript in preparation), we analysed over 1.7-billion read pairs of Hi-C data corresponding to pluripotent and differentiated cells (Supplementary Table 1). We normalized the Hi-C interactions for biases in the data (Supplementary Figs 1 and 2)³. To validate the quality of our Hi-C data, we compared the data with previous chromosome conformation capture (3C), chromosome conformation capture carbon copy (5C), and fluorescence *in situ* hybridization (FISH) results^{4–6}. Our IMR90 Hi-C data show a high degree of similarity when compared to a previously generated 5C data set from lung fibroblasts (Supplementary Fig. 4). In addition, our mouse ES cell Hi-C data correctly recovered a previously described cell-type-specific interaction at the *Phc1* gene⁵ (Supplementary Fig. 5). Furthermore, the Hi-C interaction frequencies in mouse ES cells are well-correlated with the mean spatial distance separating six loci as measured by two-dimensional FISH⁶ (Supplementary Fig. 6), demonstrating that the normalized Hi-C data can accurately reproduce the expected nuclear distance using an independent method. These results demonstrate that our Hi-C data are of

high quality and accurately capture the higher order chromatin structures in mammalian cells.

We next visualized two-dimensional interaction matrices using a variety of bin sizes to identify interaction patterns revealed as a result of our high sequencing depth (Supplementary Fig. 7). We noticed that at bin sizes less than 100 kilobases (kb), highly self-interacting regions begin to emerge (Fig. 1a and Supplementary Fig. 7, seen as ‘triangles’ on the heat map). These regions, which we term topological domains, are bounded by narrow segments where the chromatin interactions appear to end abruptly. We hypothesized that these abrupt transitions may represent boundary regions in the genome that separate topological domains.

To identify systematically all such topological domains in the genome, we devised a simple statistic termed the directionality index to quantify the degree of upstream or downstream interaction bias for a genomic region, which varies considerably at the periphery of the topological domains (Fig. 1b; see Supplementary Methods for details). The directionality index was reproducible (Supplementary Table 2) and pervasive, with 52% of the genome having a directionality index that was not expected by random chance (Fig. 1c, false discovery rate = 1%). We then used a Hidden Markov model (HMM) based on the directionality index to identify biased ‘states’ and therefore infer the locations of topological domains in the genome (Fig. 1a; see Supplementary Methods for details). The domains defined by HMM were reproducible between replicates (Supplementary Fig. 8). Therefore, we combined the data from the HindIII replicates and identified 2,200 topological domains in mouse ES cells with a median size of 880 kb that occupy ~91% of the genome (Supplementary Fig. 9). As expected, the frequency of intra-domain interactions is higher than inter-domain interactions (Fig. 1d, e). Similarly, FISH probes⁶ in the same topological domain (Fig. 1f) are closer in nuclear space than probes in different topological domains (Fig. 1g), despite similar genomic distances between probe pairs (Fig. 1h, i). These findings are best explained by a model of the organization of genomic DNA into spatial modules linked by short chromatin segments. We define the genomic regions between topological domains as either ‘topological boundary regions’ or ‘unorganized chromatin’, depending on their sizes (Supplementary Fig. 9).

We next investigated the relationship between the topological domains and the transcriptional control process. The *Hoxa* locus is separated into two compartments by an experimentally validated insulator^{4,7,8}, which we observed corresponds to a topological domain boundary in both mouse (Fig. 1a) and human (Fig. 2a). Therefore, we hypothesized that the boundaries of the topological domains might correspond to insulator or barrier elements.

Many known insulator or barrier elements are bound by the zinc-finger-containing protein CTCF (refs 9–11). We see a strong enrichment of CTCF at the topological boundary regions (Fig. 2b and Supplementary Fig. 10), indicating that topological boundary regions

¹Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093, USA. ²Medical Scientist Training Program, University of California, San Diego, La Jolla, California 92093, USA.

³Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, California 92093, USA. ⁴Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, California 92093, USA. ⁵Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA. ⁶University of California, San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, UCSD Moores Cancer Center, 9500 Gilman Drive, La Jolla, California 92093, USA.

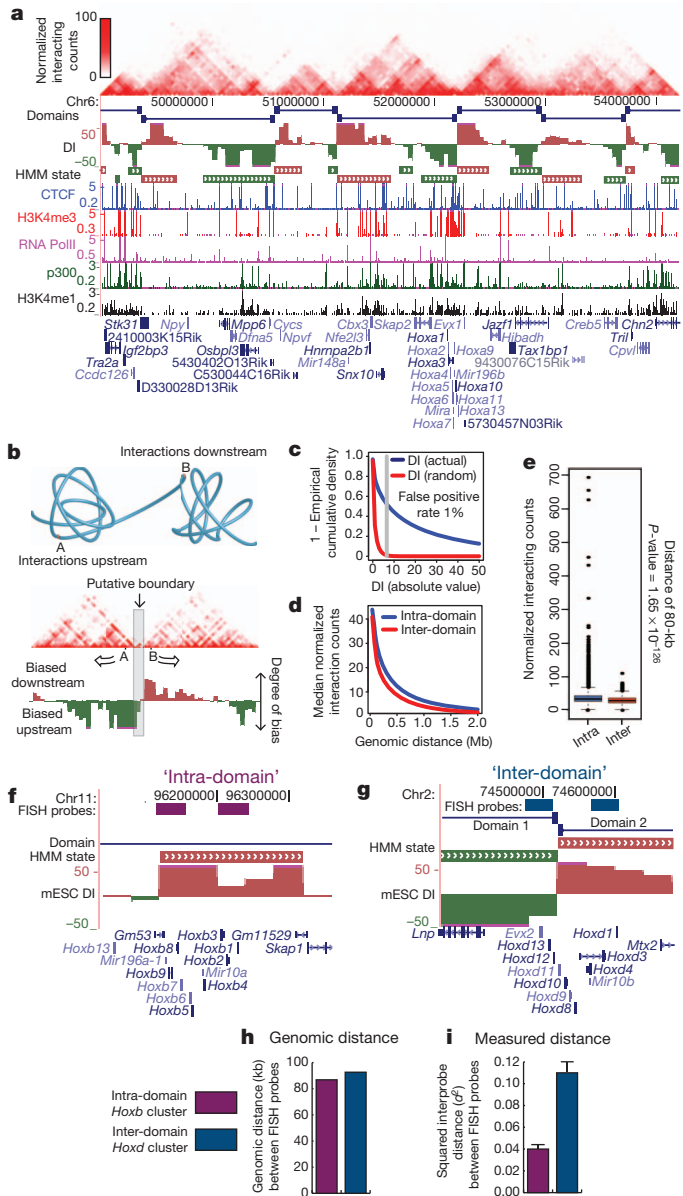


Figure 1 | Topological domains in the mouse ES cell genome. **a**, Normalized Hi-C interaction frequencies displayed as a two-dimensional heat map overlaid on ChIP-seq data (from Y. Shen *et al.*, manuscript in preparation), directionality index (DI), HMM bias state calls, and domains. For both directional index and HMM bias state calls, downstream bias (red) and upstream bias (green) are indicated. **b**, Schematic illustrating topological domains and resulting directional bias. **c**, Distribution of the directionality index (absolute value, in blue) compared to random (red). **d**, Mean interaction frequencies at all genomic distances between 40 kb to 2 Mb. Above 40 kb, the intra- versus inter-domain interaction frequencies are significantly different ($P < 0.005$, Wilcoxon test). **e**, Box plot of all interaction frequencies at 80-kb distance. Intra-domain interactions are enriched for high-frequency interactions. **f–i**, Diagram of intra-domain (**f**) and inter-domain FISH probes (**g**) and the genomic distance between pairs (**h**). **i**, Bar chart of the squared inter-probe distance (from ref. 6) FISH probe pairs. mESC, mouse ES cell. Error bars indicate standard error ($n = 100$ for each probe pair).

share this feature of classical insulators. A classical boundary element is also known to stop the spread of heterochromatin. Therefore, we examined the distribution of the heterochromatin mark H3K9me3 in humans and mice in relation to the topological domains^{12,13}. Indeed, we observe a clear segregation of H3K9me3 at the boundary regions that occurs predominantly in differentiated cells (Fig. 2d, e and Supplementary Fig. 11). As the boundaries that we analysed in

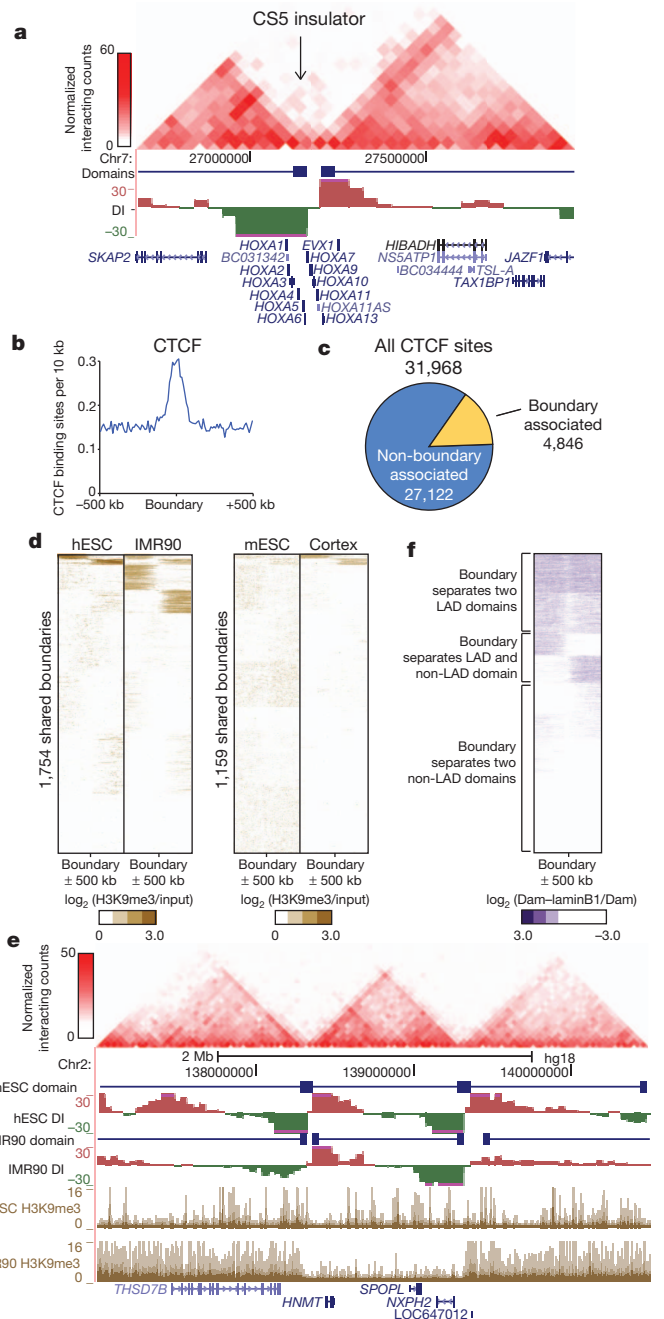


Figure 2 | Topological boundaries demonstrate classical insulator or barrier element features. **a**, Two-dimensional heat map surrounding the *Hoxa* locus and CS5 insulator in IMR90 cells. **b**, Enrichment of CTCF at boundary regions. **c**, The portion of CTCF binding sites that are considered ‘associated’ with a boundary (within ± 20 -kb window is used as the expected uncertainty due to 40-kb binning). **d**, Heat maps of H3K9me3 at boundary sites in human and mouse. **e**, UCSC Genome Browser shot showing heterochromatin spreading in the human ES cells (hESC) and IMR90 cells. The two-dimensional heat map shows the interaction frequency in human ES cells. **f**, Heat map of LADs (from ref. 14) surrounding the boundary regions. Scale is the \log_2 ratio of DNA adenosine methylation (Dam)–lamin B1 fusion over Dam alone (Dam–lamin B1/Dam).

Fig. 2d are present in both pluripotent cells and their differentiated progeny, the topological domains and boundaries appear to pre-mark the end points of heterochromatic spreading. Therefore, the domains do not seem to be a consequence of the formation of heterochromatin. Taken together, the above observations strongly suggest that the topological domain boundaries correlate with regions of the genome displaying classical insulator and barrier element activity, thus revealing a

potential link between the topological domains and transcriptional control in the mammalian genome.

We compared the topological domains with previously described domain-like organizations of the genome, specifically with the A and B compartments described by ref. 2, with lamina-associated domains (LADs)^{10,14}, replication time zones^{15,16}, and large organized chromatin K9 modification (LOCK) domains¹⁷. In all cases, we can see that topological domains are related to, but independent from, each of these previously described domain-like structures (Supplementary Figs 12–15). Notably, a subset of the domain boundaries we identify appear to mark the transition between either LAD and non-LAD regions of the genome (Fig. 2f and Supplementary Fig. 12), the A and B compartments (Supplementary Fig. 13, 14), and early and late replicating chromatin (Supplementary Fig. 14). Lastly, we can also confirm the previously reported similarities between the A and B compartments and early and late replication time zone (Supplementary Fig. 16)¹⁶.

We next compared the locations of topological boundaries identified in both replicates of mouse ES cells and cortex, or between both replicates of human ES cells and IMR90 cells. In both human and mouse, most of the boundary regions are shared between cell types (Fig. 3a and Supplementary Fig. 17a), suggesting that the overall domain structure between cell types is largely unchanged. At the boundaries called in only one cell type, we noticed that trend of upstream and downstream bias in the directionality index is still readily apparent and highly reproducible between replicates (Supplementary Fig. 17b, c). We cannot determine if the differences in domain calls between cell types is due to noise in the data or to biological phenomena, such as a change in the strength of the boundary region between cell types¹⁸. Regardless, these results indicate that the domain boundaries are largely invariant between cell types. Lastly, only a small fraction of the boundaries show clear differences between two cell types, suggesting that a relatively rare subset of boundaries may actually differ between cell types (Supplementary Fig. 18).

The stability of the domains between cell types is surprising given previous evidence showing cell-type-specific chromatin interactions and conformations^{5,7}. To reconcile these results, we identified cell-type-specific chromatin interactions between mouse ES cell and mouse cortex. We identified 9,888 dynamic interacting regions in the mouse genome based on 20-kb binning using a binomial test with an empirical false discovery rate of <1% based on random permutation of the replicate data. These dynamic interacting regions are enriched for differentially expressed genes (Fig. 3b–d, Supplementary Fig. 19 and Supplementary Table 5). In fact, 20% of all genes that undergo a four-fold change in gene expression are found at dynamic interacting loci. This is probably an underestimate, because by binning the genome at 20 kb, any dynamic regulatory interaction less than 20 kb will be missed. Lastly, >96% of dynamic interacting regions occur in the same domain (Fig. 3e). Therefore, we favour a model where the domain organization is stable between cell types, but the regions within each domain may be dynamic, potentially taking part in cell-type-specific regulatory events.

The stability of the domains between cell types prompted us to investigate if the domain structure is also conserved across evolution. To address this, we compared the domain boundaries between mouse ES cells and human ES cells using the UCSC liftover tool. Most of the boundaries appear to be shared across evolution (53.8% of human boundaries are boundaries in mouse and 75.9% of mouse boundaries are boundaries in humans, compared to 21.0% and 29.0% at random, P value $< 2.2 \times 10^{-16}$, Fisher's exact test; Fig. 3f). The syntenic regions in mouse and human in particular share a high degree of similarity in their higher order chromatin structure (Fig. 3g, h), indicating that there is conservation of genomic structure beyond the primary sequence of DNA.

We explored what factors may contribute to the formation of topological boundary regions in the genome. Although most topological boundaries are enriched for the binding of CTCF, only 15% of CTCF

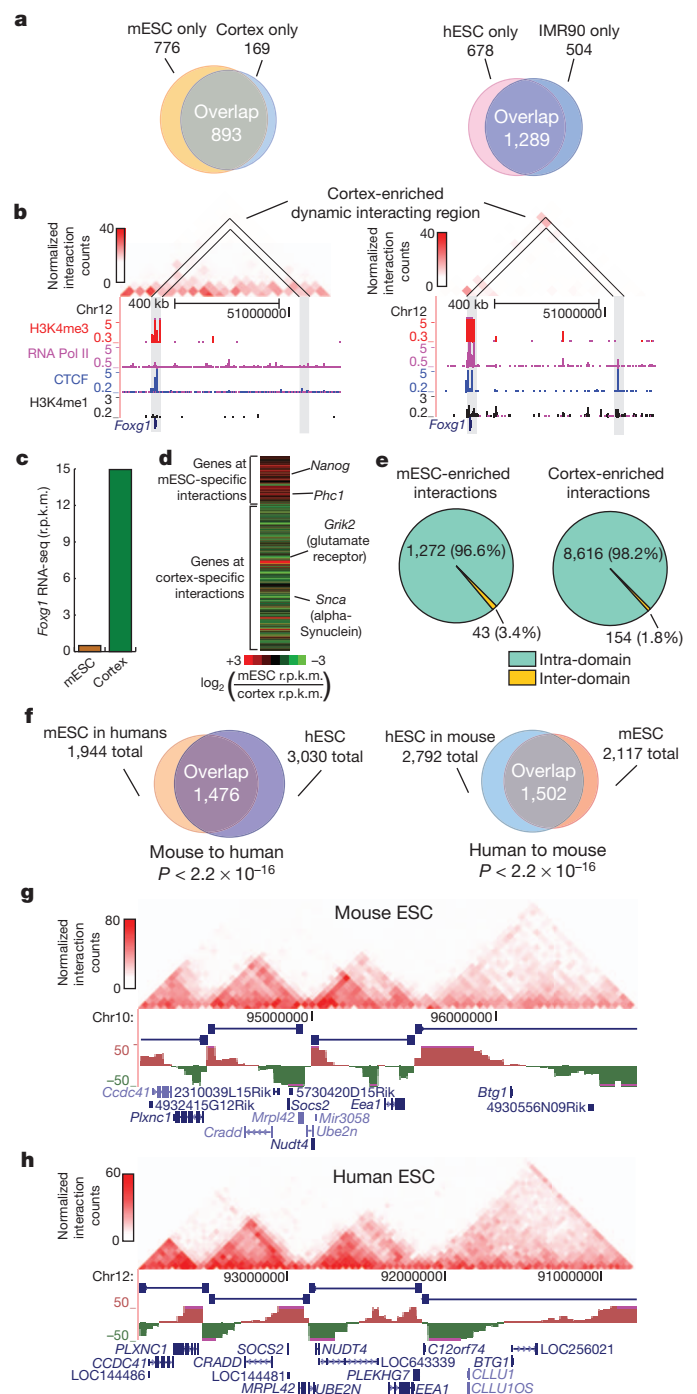


Figure 3 | Boundaries are shared across cell types and conserved in evolution. **a**, Overlap of boundaries between cell types. **b**, Genome browser shot of a cortex enriched dynamic interacting region that overlaps with the *Foxg1* gene. **c**, *Foxg1* expression in reads per kilobase per million reads sequenced (r.p.k.m.) in mouse ES cells and cortex as measured by RNA-seq. **d**, Heat map of the gene expression ratio between mouse ES cell and cortex of genes at dynamic interactions. **e**, Pie chart of inter- and intra-domain dynamic interactions. **f**, Overlap of boundaries between syntenic mouse and human sequences ($P < 2.2 \times 10^{-16}$ compared to random, Fisher's exact test). **g, h**, Genome browser shots showing domain structure over a syntenic region in the mouse (**g**) and human (**h**) ES cells. Note: the region in humans has been inverted from its normal UCSC coordinates for proper display purposes.

binding sites are located within boundary regions (Fig. 2c). Thus, CTCF binding alone is insufficient to demarcate domain boundaries. We reasoned that additional factors might be associated with topological boundary regions. By examining the enrichment of a variety of

histone modifications, chromatin binding proteins and transcription factors around topological boundary regions in mouse ES cells, we observed that factors associated with active promoters and gene bodies are enriched at boundaries in both mouse and humans (Fig. 4a and Supplementary Figs 20–23)^{19,20}. In contrast, non-promoter-associated marks, such as H3K4me1 (associated with enhancers) and H3K9me3, were not enriched or were specifically depleted at boundary regions (Fig. 4a). Furthermore, transcription start sites (TSS) and global run on sequencing (GRO-seq)²¹ signal were also enriched around topological boundaries (Fig. 4a). We found that housekeeping genes were particularly strongly enriched near topological boundary regions (Fig. 4b–d; see Supplementary Table 7 for complete GO terms enrichment). Additionally, the tRNA genes, which have the potential to function as boundary elements^{22,23}, are also enriched at boundaries (P value < 0.05 , Fisher's exact test; Fig. 4b). These results suggest that high levels of transcription activity may also contribute to boundary formation. In support of this, we can see examples of dynamic changes in H3K4me3 at or near some cell-type-specific boundaries that are cell-type specific (Supplementary Fig. 24). Indeed, boundaries associated with both CTCF and a housekeeping gene account for nearly one-third of all topological boundaries in the genome (Fig. 4e and Supplementary Fig. 24).

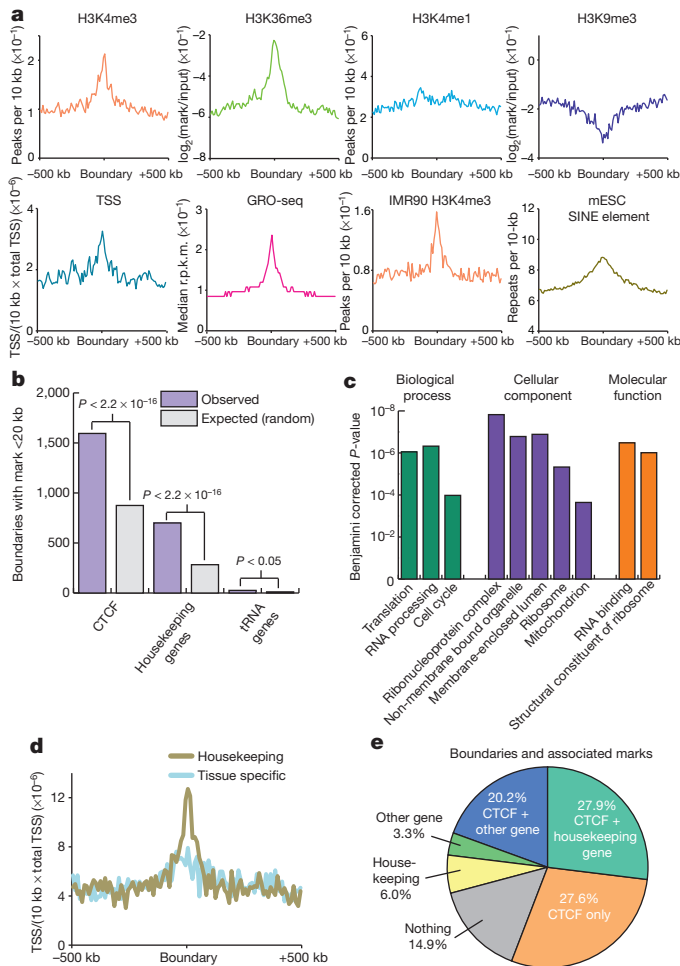


Figure 4 | Boundary regions are enriched for housekeeping genes.

a, Chromatin modifications, TSS, GRO-seq and SINE elements surrounding boundary regions in mouse ES cells or IMR90 cells. **b**, Boundaries associated with a CTCF binding site, housekeeping gene, or tRNA gene (purple) compared to expected at random (grey). **c**, Gene Ontology P -value chart. **d**, Enrichment of housekeeping genes (gold) and tissue-specific genes (blue) as defined by Shannon entropy scores near boundaries normalized for the number of genes in each class (TSS/10 kb/total TSS). **e**, Percentage of boundaries with a given mark within 20 kb of the boundaries.

Finally, we analysed the enrichment of repeat classes around boundary elements. We observed that Alu/B1 and B2 SINE retrotransposons in mouse and Alu SINE elements in humans are enriched at boundary regions (Fig. 4a and Supplementary Figs 24 and 25). In light of recent reports indicating that a SINE B2 element functions as a boundary in mice²⁴, and SINE element retrotransposition may alter CTCF binding sites during evolution²⁵, we believe that this contributes to a growing body of evidence indicating a role for SINE elements in the organization of the genome.

In summary, we show that the mammalian chromosomes are segmented into megabase-sized topological domains, consistent with some previous models of the higher order chromatin structure^{1,26,27}. Such spatial organization seems to be a general property of the genome: it is pervasive throughout the genome, stable across different cell types and highly conserved between mice and humans.

We have identified multiple factors that are associated with the boundary regions separating topological domains, including the insulator binding factor CTCF, housekeeping genes and SINE elements. The association of housekeeping genes with boundary regions extends previous studies in yeast and insects and suggests that non-CTCF factors may also be involved in insulator/barrier functions in mammalian cells²⁸.

The topological domains we identified are well conserved between mice and humans. This indicates that the sequence elements and mechanisms that are responsible for establishing higher order structures in the genome may be relatively ancient in evolution. A similar partitioning of the genome into physical domains has also been observed in *Drosophila* embryos²⁹ and in high-resolution studies of the X-inactivation centre in mice (termed topologically associated domains or TADs)³⁰, indicating that topological domains may be a fundamental organizing principle of metazoan genomes.

METHODS SUMMARY

Cell culture and Hi-C experiments. J1 mouse ES cells were grown on gamma-irradiated mouse embryonic fibroblasts cells under standard conditions (85% high glucose DMEM, 15% HyClone FBS, 0.1 mM non-essential amino acids, 0.1 mM β -mercaptoethanol, 1 mM glutamine, LIF 500 U ml⁻¹, 1 \times Gibco penicillin/streptomycin). Before collecting for Hi-C, J1 mouse ES cells were passaged onto feeder free 0.2% gelatin-coated plates for at least two passages to rid the culture of feeder cells. H1 human ES cells and IMR90 fibroblasts were grown as previously described¹³. Collecting the cells for Hi-C was performed as previously described, with the only modification being that the adherent cell cultures were dissociated with trypsin before fixation.

Sequencing and mapping of data. Hi-C analysis and paired-end libraries were prepared as previously described² and sequenced on the Illumina Hi-Seq2000 platform. Reads were mapped to reference human (hg18) or mouse genomes (mm9), and non-mapping reads and PCR duplicates were removed. Two-dimensional heat maps were generated as previously described².

Data analysis. For detailed descriptions of the data analysis, including descriptions of the directionality index, hidden Markov models, dynamic interactions identification, and boundary overlap between cells and across species, see Supplementary Methods.

Received 26 September 2011; accepted 27 March 2012.

Published online 11 April 2012.

- Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
- Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genet.* **43**, 1059–1065 (2011).
- Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
- Noordermeer, D. *et al.* The dynamic architecture of Hox gene clusters. *Science* **334**, 222–225 (2011).

8. Kim, Y. J., Cecchini, K. R. & Kim, T. H. 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 (2011).
9. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
10. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
11. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet.* **43**, 630–638 (2011).
12. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012).
13. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
14. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
15. Hiratani, I. *et al.* Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. *Genome Res.* **20**, 155–169 (2010).
16. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
17. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genet.* **41**, 246–250 (2009).
18. Scott, K. C., Taubman, A. D. & Geyer, P. K. Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* **153**, 787–798 (1999).
19. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
20. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
21. Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).
22. Donze, D. & Kamakaka, R. T. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. *EMBO J.* **20**, 520–531 (2001).
23. Ebersole, T. *et al.* tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. *Cell Cycle* **10**, 2779–2791 (2011).
24. Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248–251 (2007).
25. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348 (2012).
26. Jhunjhunwala, S. *et al.* The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**, 265–279 (2008).
27. Capelson, M. & Corces, V. G. Boundary elements and nuclear organization. *Biol. Cell* **96**, 617–629 (2004).
28. Amouyal, M. Gene insulation. Part I: natural strategies in yeast and *Drosophila*. *Biochem. Cell Biol.* **88**, 875–884 (2010).
29. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
30. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* <http://dx.doi.org/10.1038/nature11049> (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for the comments from and discussions with Z. Qin, A. Desai and members of the Ren laboratory during the course of the study. We also thank W. Bickmore and R. Eskeland for sharing the FISH data generated in mouse ES cells. This work was supported by funding from the Ludwig Institute for Cancer Research, California Institute for Regenerative Medicine (CIRM, RN2-00905-1) (to B.R.) and NIH (B.R. R01GH003991). J.R.D. is funded by a pre-doctoral training grant from CIRM. Y.S. is supported by a postdoctoral fellowship from the Rett Syndrome Research Foundation.

Author Contributions J.R.D. and B.R. designed the studies. J.R.D., A.K., Y.L. and Y.S. conducted the Hi-C experiments; J.R.D., S.S. and F.Y. carried out the data analysis; J.S.L. and M.H. provided insight for analysis; F.Y. built the supporting website; J.R.D. and B.R. prepared the manuscript.

Author Information All Hi-C data described in this study have been deposited in the GEO under accession number GSE35156. We have developed a web-based Java tool to visualize the high-resolution Hi-C data at a genomic region of interest that is available at <http://chromosome.sdsc.edu/mouse/hi-c/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.R. (biren@ucsd.edu).