



Published in final edited form as:

Mol Cell. 2012 November 9; 48(3): 327–328. doi:10.1016/j.molcel.2012.10.021.

Order from Chaos in the Nucleus

Zhixiong Xu¹ and Gary Felsenfeld^{1,*}

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

Summary

Although it might appear that chromatin is randomly packed within the nucleus, recent data show that it is organized into defined and functionally important domains marked by preferred intra-domain physical contacts, and with boundaries associated with insulator protein occupancy.

DNA within the nucleus is packaged as chromatin, and folded into a conformation often described by physical chemists as a ‘polymer melt’. But this somewhat chaotic structure conceals hierarchies of organization, which at the molecular level effectively partitions extended regions within chromosomes into structurally and functionally distinguishable domains. Within and between those domains there are further physical interactions that bring together distant regulatory elements and gene promoters, or further divide the larger partition into subdomains. In this issue of *Molecular Cell*, Hou et al. (2012) provide a comprehensive description of the organization of the *Drosophila* genome, which reveals the details of its domain structure at very high resolution.

Information about large scale organization within the nucleus is now coming from work in many laboratories. All of it depends on Hi-C (Lieberman-Aiden et al., 2009) or 5C (Dostie et al., 2006), derivatives of the chromatin conformation capture (3C) method originally described by Dekker and colleagues (Dekker et al., 2002). In all of these methods chromatin in nuclei is fixed, digested with a restriction enzyme, and then re-ligated to form junctions between DNA sequences that, although distant in the linear genome, are close to each other inside the nucleus. Early low resolution studies by Hi-C gave us a glimpse of nuclear order (Lieberman-Aiden et al., 2009). These methods are now capable of much higher resolution because of the greatly increased output capacity of large scale sequencing methods. Thus, Hou et al. (2012) are able to detect statistically significant physical contacts across the entire *Drosophila* genome at resolution of 4 kb for interactions between sites located within ~100 kb of each other. Detailed analysis reveals that interactions can be clustered into about a thousand discrete physical domains of average size 107 kb. While this kind of pattern already had been detected in flies and vertebrates, Hou et al. study the domain properties in more detail. For this purpose they make use of an earlier study (Filion et al., 2010) that decomposed the *Drosophila* genome into five color-coded categories, each associated with a distinct pattern of histone modifications and corresponding gene activity. They show that there is an inverse correlation between the level of activity associated with a given region of chromatin and the average size of the domain in which it is located: domains rich in active chromatin tend to be smaller, whereas the longer domains contain inactive chromatin. But the individual domains are found not to be homogeneous; notably, many are enriched for

*Correspondence: gary.felsenfeld@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

active chromatin at the domain boundaries, and, irrespective of activity, these boundaries are marked by increased gene density, and particularly enriched for genes involved in environmental stress responses. Furthermore, interactions between boundaries occur more frequently than interdomain interactions, perhaps resulting in the formation of clustered highly active domains.

An important goal of such studies is to identify the structures or proteins that are associated with and may be responsible for boundary formation. Hou et al. find that boundary regions ('domain partition sites') tend to be enriched for the known *Drosophila* insulator proteins, BEAF, CTCF, CP190, and to a lesser extent Suppressor of Hairy wing (Su(Hw)). Although individual sites recruit different combinations of these factors, the most abundant configuration involves all four, perhaps generating the strongest insulator activity. The high concentration of these factors at boundaries is correlated with the greater gene density, leading to the suggestion that the combination of insulator sites with high gene activity and density may be necessary for boundary formation.

A paper published earlier this year (Sexton et al., 2012) addresses the same questions using Hi-C, and comes to somewhat different but not necessarily contradictory conclusions. Here the analysis is carried out in *Drosophila* embryos rather than the Kc167 cell line used by Hou et al., and reveals a domain organization similar in many ways. Almost half (42%) of the domain borders identified in the Hou paper overlap with those found by Sexton et al; the difference presumably reflects cell type or developmental specific variation in boundary formation, and deserves further examination. Sexton et al. also report the enrichment of insulator-associated proteins CTCF, CP190 and BEAF at domain boundaries, as well as the mitotic spindle protein Chromator, a novel addition to the list of boundary associated proteins in flies. Domain data are also compared to the color coded deconstruction (Filion et al., 2010), leading to a strong correlation between the locations of domain boundaries and boundaries of color coded domains. This varies somewhat from the conclusion reached by Hou et al. (2012); it probably reflects a difference between the two papers in the way data were analyzed. Whereas Hou et al. categorize domains according to properties of chromatin near the boundaries, Sexton et al. ask about the overall statistical association between the epigenetic profiles and the physical domain patterns. As these authors point out, this can be a powerful tool in predicting domain organization. However this can lead to apparent but not real inconsistencies between the two sets of data. The differences are a reminder that in these exceedingly complex and data-rich systems the conclusions that are reached are partly dependent on the questions that are asked and the statistical constraints that are imposed.

Vertebrate genomes have been the subject of similar analysis, but this is more demanding because the greater size of the genome requires large amounts of Hi-C data. Dixon et al. (Dixon et al. (2012) recently use over 1.7 billion read pairs to identify domain boundaries in human and mouse ES cells, a human fibroblast line and mouse cortex by searching directly for sites of minimum interaction with neighboring sequences. In mouse ES cells about 91% of the genome is resolved into 2,200 domains of average size 880 kb, and thus almost an order of magnitude larger than that in flies. About 75% of boundaries are bound by CTCF, but only 15% of all bound CTCF is located at boundaries, consistent with the other roles that CTCF probably plays in vertebrates, including stabilization of shorter range intradomain interactions (Wallace and Felsenfeld, 2007). Comparison of different cell types indicates that the boundaries are determined largely independently of the cell type or histone modification state. Boundaries are enriched in transcription start sites and factors associated with active promoters and gene bodies, reminiscent of observations by Hou et al. in flies. Enrichment is also observed for housekeeping genes and, notably, for tRNA genes and SINE elements, both recently implicated in domain organization in vertebrates.

Although genome-wide studies have been illuminating, other insights into mechanisms can be obtained by focusing on a single region. Nora et al. (Nora et al., 2012) study the organization of the mouse X-inactivation center (*Xic*) by 5C, which allows them to obtain detailed conformational information over a 4.5-megabase (Mb) region on the X chromosome that includes *Xist* and its antisense repressor, *Tsix*. The organization of the region into domains 0.2 - 1 Mb long places the promoters of these two genes in separated domains. Deletion of the boundary between them results in ectopic chromatin folding as well as mis-regulated transcription, confirming the importance of this kind of organization. Furthermore, loci that are coordinately regulated tend to be located in the same domain, suggesting that such arrangements either exploit higher order structure or were instrumental in its evolution. The authors point out that the importance of the entire domain may explain why even the largest tested transgenes do not direct correct patterns of *Xist* expression.

The principal focus of all of these papers has been to determine the way in which the fly or vertebrate genome is organized into large-scale domains. As the work on the *Xic* (Nora et al., 2012) and the observation of increased transcription at the domain borders genome-wide (Hou et al., 2012) have made clear, there are important consequences of such organization for gene regulation. It seems likely that the domain structure will also be involved in other processes such as DNA replication and recombination. Indeed, Zhang et al. (Zhang et al., 2012) showed recently that higher-order organization of the mouse genome contributes to recurrent chromosomal translocation. Despite differences between flies and vertebrates in genome size, extent of heterochromatic domains, and detailed identity of boundary factors, the broad features of organization among species and cell types are rather similar, and suggest that these structures are a central feature of nuclear architecture. It should be kept in mind, however, that data in these papers also reveals less frequent cell-type specific contacts that are likely to be significant for the establishment of cellular identities and may follow rules that remain to be worked out. Both between domains and within them there are still other levels of interactions, some more stable than others, but all likely to play roles in bringing order to what might at first glance into the nucleus have looked like chaos.

Acknowledgments

This work was supported by the intramural research program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases.

REFERENCES

- Dekker J, Rippe K, Dekker M, Kleckner N. Science. 2002; 295:1306–1311. [PubMed: 11847345]
- Dixon JR, Siddarth S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Nature. 2012; 485:376–380. [PubMed: 22495300]
- Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. Genome Res. 2006; 16:1299–1309. [PubMed: 16954542]
- Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, et al. Cell. 2010; 143:212–224. [PubMed: 20888037]
- Hou CH, Li L, Qin ZS, Corces VG. 2012; 48 this issue.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Science. 2009; 326:289–293. [PubMed: 19815776]
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Nature. 2012; 485:381–385. [PubMed: 22495304]
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Cell. 2012; 148:458–472. [PubMed: 22265598]
- Wallace JA, Felsenfeld G. Curr. Opin. Genet. Dev. 2007; 17:400–407. [PubMed: 17913488]

Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J.
Cell. 2012; 148:908–921. [PubMed: 22341456]

\$watermark-text

\$watermark-text

\$watermark-text