

The emergence of lineage-specific chromosomal topologies from coordinate gene regulation

Indika Rajapakse^{a,b}, Michael D. Perlman^c, David Scalzo^a, Charles Kooperberg^b, Mark Groudine^{a,d,1}, and Steven T. Kosak^{e,1}

^aDivision of Basic Sciences and ^bDepartment of Biostatistics and Biomathematics, Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109; ^cDepartment of Statistics, University of Washington, Seattle, WA 98195; ^dDepartment of Radiation Oncology, University of Washington School of Medicine, Seattle, WA 98195; and ^eDepartment of Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, 303 East Chicago Avenue, Ward 8-132, Chicago, IL 60611

Contributed by Mark Groudine, January 30, 2009 (sent for review December 15, 2008)

Although the importance of chromosome organization during mitosis is clear, it remains to be determined whether the nucleus assumes other functionally relevant chromosomal topologies. We have previously shown that homologous chromosomes have a tendency to associate during hematopoiesis according to their distribution of coregulated genes, suggesting cell-specific nuclear organization. Here, using the mathematical approaches of distance matrices and coupled oscillators, we model the dynamic relationship between gene expression and chromosomal associations during the differentiation of a multipotential hematopoietic progenitor. Our analysis reveals dramatic changes in total genomic order: Commitment of the progenitor results in an initial increase in entropy at both the level of gene coregulation and chromosomal organization, which we suggest represents a phase transition, followed by a progressive decline in entropy during differentiation. The stabilization of a highly ordered state in the differentiated cell types results in lineage-specific chromosomal topologies and is related to the emergence of coherence—or self-organization—between chromosomal associations and coordinate gene regulation. We discuss how these observations may be generally relevant to cell fate decisions encountered by progenitor/stem cells.

cellular differentiation | chromosomal organization | coregulated gene expression | distance matrices | networks

The convergence of biological questions and mathematical approaches has encouraged the characterization of complex cellular processes. Our understanding of the regulation of gene activity, for example, has been aided by *in silico* modeling—exploring genetic associations from the standpoint of Boolean networks or cellular automata (1, 2). Similarly, more recent developments in graph theory—particularly the description of small-world and scale-free networks—have uncovered a scarcity of randomness in biological systems (3). The shared insight from these different approaches is that biological processes are inclined to self-organize, in which a network of localized interactions yields an emergent structure that subsequently feeds back on and strengthens the original network (4). With this conceptual framework, cellular organelles can be viewed as the spatial organization of dynamic cellular tasks. Therefore, it is not surprising that perturbing the function of an organelle—such as the Golgi apparatus' processing of polypeptides—results in the loss of its 4-dimensional form (5, 6).

Current evidence indicates that genes and chromosomes are nonrandomly localized within the nucleus. For example, several gene loci have been shown to be positioned at the periphery when inactive and then relocalized to the nuclear center upon their developmentally regulated activation (7, 8). Also, various chromosomal attributes, including gene density, size (base pair length), and coregulated gene activity, have been indicated in their organization in mammalian nuclei (9–11). Given the growing evidence for deterministic nuclear organization, iden-

tifying its origin and structure has become increasingly important. We have previously shown that during the *in vitro* differentiation of a murine hematopoietic progenitor to derived erythroid and neutrophil cell types, there is a correlation between the nonrandom organization of the genome and coordinate gene regulation (11). We observed that genes that are coregulated during differentiation have a significant tendency to be proximally distributed along chromosomes. In turn, we found that the frequency at which homologous chromosomes associate is related to the number of coregulated genes they possess. We have therefore suggested that coordinate gene regulation during cellular differentiation may yield lineage-specific nuclear topologies that facilitate gene coregulation (7). Moreover, we have hypothesized that the process of self-organization is responsible for the emergence of these topologies (12).

Here, we directly test our idea that chromosomes self-organize during differentiation according to coordinate gene regulation. To do so, we first made the assumption that a full understanding of nuclear organization necessitates the view that it is a dynamical system. For example, capturing a snapshot of genomic gene expression at any one point in time during differentiation reveals a static regulatory network; however, a genetic network actually evolves over time, with groups of genes coupling or decoupling their expression to the overall coregulated gene set. Furthermore, based on our previous analysis, we propose that this evolving regulatory network is manifested spatially at the level of chromosomal organization, with all chromosomes—both homologs and heterologs—associating according to their overall coregulation. Our analysis focuses on determining the collective similarity between gene regulatory and chromosomal association networks by expressing them as matrices. To construct the matrices, we measure the relative entropy—or “distance”—among nodes within networks as well as between networks during differentiation, allowing us to assay shared global properties and the emergence of lineage-specific relationships. In addition, using the theoretical concept of coupled oscillators, we can determine whether the evolving nature of these relationships is reflected in an increase in coherence, *i.e.*, self-organization. Our analysis demonstrates that the networks of coregulated gene expression and chromosomal association are indeed mutually related during differentiation, resulting in the self-organization of lineage-specific chromosomal topologies.

Results

To test our hypothesis of dynamical genome organization, we first analyzed how coordinate gene expression evolves along a

Author contributions: I.R., M.D.P., M.G., and S.T.K. designed research; I.R., D.S., and S.T.K. performed research; M.D.P. and C.K. contributed new reagents/analytic tools; I.R., D.S., M.G., and S.T.K. analyzed data; and I.R., M.G., and S.T.K. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. E-mail: markg@fhcrc.org or s-kosak@northwestern.edu.

ing to differential gene expression, creating a new higher-order state or attractor than that in the progenitor.

To determine the evolving role of each individual chromosome to the overall organization of the network during differentiation, we computed a centrality measure for each chromosome at each time point with the vertex strength s_i defined as

$$s_i = \sum_{j=1}^N a_{ij}$$

(18), where a_{ij} is the ij th entry of the matrix A . This quantity measures the importance of vertices (i.e., chromosomes) in terms of the total weight of their connections, and it is therefore a natural measure of the importance or centrality of a given chromosome i in the network. Our analysis demonstrates that chromosomes contribute differently over the time course of differentiation (Fig. 3C). In Fig. 3C *Center*, we have depicted the relative strengths for each chromosome during the differentiation, arbitrarily highlighting 4 chromosomes (1, 6, 11, and 17) for elucidation. In the Fig. 3C *Left* and *Right* sides of the total time course, we focus on the progenitor and erythroid lineages, respectively, illustrating the change in contribution of the 4 chromosomes from the beginning and end states. The erythroid state also reveals that many of the chromosomes (indicated by x) have collapsed onto each other in terms of their relative strengths, representing increased coherence. Moreover, because the centrality measure of all chromosomes is greater at the end than the beginning, the overall organization of chromosomes increases during differentiation. Therefore, paralleling our results with gene expression, cellular potential is lost as chromosomal order is gained.

The analysis above suggests that coregulated gene expression and chromosome association networks are mutually related during cellular differentiation. Genomic organization proceeds from an unstable ordered to a disordered state during the commitment of the progenitor and ultimately reaches a highly ordered state in the differentiated cell types. This evolving nature of genomic order prompted us to use a mathematical framework to model the interrelated behaviors of gene regulation and chromosome association as a dynamical system. The cooperative phenomenon of mutual entrainment is well described by the classical Kuramoto model (4, 19–21), in which a collection of globally coupled phase oscillators exhibit a transition from incoherence to a coherent state as the coupling strength increases past a critical threshold. Therefore, the Kuramoto framework is an ideal mathematical model with which to demonstrate the organized behavior of the complex dynamical reorganization of the genome during differentiation. A generalized form of this mathematical model can be written as

$$\frac{d(\theta_i)}{dt} = \omega_i + \left(\sum_{j=1}^N A_{ij}(t) \sin(\theta_j - \theta_i) \right), \quad i, j = 1, 2, \dots, N.$$

Here, θ_i and ω_i denote, respectively, the phase and intrinsic frequency of oscillator i and N is the number of oscillators (19). The matrix $A_{ij}(t)$ is the time-dependent network architecture of oscillators, assumed to be symmetric. To characterize the degree of self-organization in the network, we used a global order parameter (R) that ranges between 0 and 1, with 0 meaning uniform incoherence and 1 meaning complete self-organization (19, 22). To apply this framework to the dynamic mutual relationship between chromosomal organization and coregulated gene expression during differentiation, we rendered each of the 19 chromosomes as an oscillator by first mapping the coregulated gene set onto their respective chromosomal positions. Therefore, we will be measuring how each chromosome

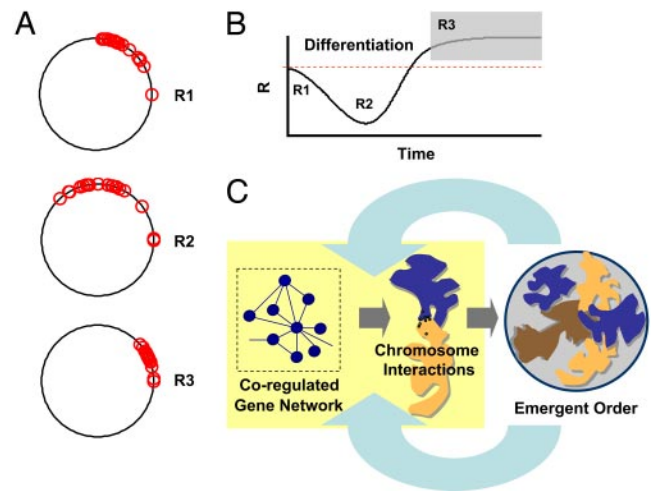


Fig. 4. Chromosomes self-organize according to coordinate gene regulation. (A) Geometric interpretation of the chromosomal network, where the phases of θ_i are plotted on the unit circle. Three snapshots during differentiation are represented, where R1 is the distribution of the network representing the progenitor state, R2 the phase transition, and R3 the erythroid differentiated cell type. (B) Schematic illustration of the evolution of the order parameter (R) seen in numerical simulations of the chromosomal oscillator model for a specific coupling strength parameter (19). (C) Schematic illustration for the mechanics of self-organization, with local interactions (gene coregulation) leading to chromosomal associations that emerge cooperatively in a cell-specific organization of the nucleus, which in turn feeds back to strengthen the local associations.

associates with all other chromosomes (i.e., oscillates) in the rosettes as a function of its share of the coregulated gene set. The initial conditions for parameters θ_i and ω_i of the oscillators were obtained from s_i as previously defined for the progenitor state (see Fig. 1), and $A_{ij}(0)$ is derived from the progenitor state chromosomal association matrix A (see Fig. 2). By following the relationship through time (the differentiation), we can capture whether coordinate gene regulation and chromosomal organization are integrated. In support of this hypothesis, the numerical simulations show that the behavior of R initially decreases in order, after which a highly ordered state emerges [Fig. 4 A (R1–R3) and B]. Furthermore, the chromosome association network of the differentiated state is similar to the model predicted network structure. Therefore, these simulation results reveal that coregulated gene expression and chromosome association networks are mutually related and lead to deterministic nuclear self-organization (Fig. 4C). We believe our approach provides a simple mathematical framework for further investigation of the dynamics of genome organization during cellular differentiation and other cell fate decisions.

Discussion

The nucleus is compartmentalized according to the various functions it performs. For example, a subnuclear body, such as the nucleolus, represents the spatial localization of the components necessary to carry out its particular activities, rDNA transcription and ribosomal biogenesis (5, 7). The process of self-organization provides an attractive model by which to understand the relationship between nuclear form and function (4). The key feature of a self-organizing system is the emergence of a structure that both results from and subsequently supports localized interactions (Fig. 4C). Evidence indicates a role for this phenomenon in the genesis of nuclear bodies. For example, it has been shown that the DNA damage response can be initiated in the absence of damage by the localization of components of the repair machinery to a particular genomic site (23). Moreover, a

which is invariant under both matrix scale transformations and matrix inversion. Note especially that $SSD(X, Y) = 0$ if $X = Y$ (16). The distance between X and Y^{-1} is:

$$SSD(X, Y^{-1}) = \text{trace}(XY) + \text{trace}(X^{-1}, Y^{-1}) - 2d$$

This distance measure is designed to test our main hypothesis: If gene coregulation is related to the overall proximity of chromosomes, then the 2 types of matrices should be inversely related. The relative entropy difference is standardized with respect to the first time point and is defined as follows:

$$\begin{aligned} D_0 &= SSD((B_0)^{-1}, (B_1)^{-1}), \\ D_1 &= SSD((B_0)^{-1}, (B_2)^{-1}), \dots, \\ D_6 &= SSD((B_0)^{-1}, (B_7)^{-1}), \end{aligned}$$

where D_6 is the difference in entropy between the progenitor and the differentiated cell state. We performed all calculations using the MATLAB software package.

Dynamics of Networks. The complexity of a network depends on topological structure, network evolution, node connectivity and diversity, and/or dynamical evolution. The evolving nature of a network is determined by both the dynamical rules governing the nodes and the flow occurring along each link. Mathematically, a network can be represented by a graph, recalling that a graph is an ordered pair of disjoint sets (V, E) such that E is a subset of the set of unordered pairs of V . V is the set of vertices and E is the set of edges. Most computations of graph properties are accomplished by representing the graph in the form of a matrix, called the adjacency matrix A ; A is a $n \times n$

symmetric matrix, where n is the number of vertices in the network. The matrix A has the elements a_{ij} : $a_{ij} = 1$ if the node i is connected to the node j and $a_{ij} = 0$ otherwise. The A is symmetric if there is an edge between i and j there is also an edge between j and i (3, 18). For a weighted network, a_{ij} has a numerical value, which represents the weight on the edge connecting the nodes i and j . Thus, a weighted network can be represented mathematically by an adjacency matrix with entries that are not simply 0 or 1, but are equal instead to the weights on the edges. Note that distance matrices are related to adjacency matrices, with the difference that an entry of the distance matrix is smaller if 2 elements are closer, whereas "close" (connected) vertices yield larger entries in an adjacency matrix. The evolving nature of the network is given by:

$$\frac{d(x_i)}{dt} = f_i(x_i) + C \left(\sum_{j=1}^n a_{ij} h_j(x_j) \right)$$

where $x_i = [x_i, y_i, z_i, \dots]^T \in \mathbb{R}^N$ is the state vector of node i for $i = 1, 2, \dots, n$ and describes the node equations as $f: \mathbb{R}^N \rightarrow \mathbb{R}^N$. If we assume that the first components of each node are connected to each other, then $h_j(x_j): \mathbb{R} \rightarrow \mathbb{R}$ is the output of the node j , $C = [1, 0, \dots, 0]^T$, and a_{ij} is as described above (22). The above model becomes the Kuramoto model when it is used to describe a network of phase oscillators.

ACKNOWLEDGMENTS. We thank Lindsey Muir and William Hazelton for critical reading of the manuscript. This work was supported by National Institutes of Health (NIH) Interdisciplinary Training Grant T32 CA80416 (to I.R.), NIH Grants R01 CA 074841 and P01 CA53996 (to C.K.) and R37 DK44746 and RO1 HL65440 (to M.G.), and a Burroughs Wellcome Fund Career Award in the Biomedical Sciences (to S.T.K.).

1. Kauffman SA (1984) Emergent properties in random complex automata. *Physica D* 10(1-2):145-156.
2. Langton CG (1990) Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D* 42(1-3):12-37.
3. Newman MEJ, Barabasi AL, Watts DJ (2006) *The Structure and Dynamics of Networks* (Princeton Univ Press, Princeton).
4. Strogatz SH (2003) *Syn: The Emerging Science of Spontaneous Order* (Hyperion, New York).
5. Misteli T (2001) The concept of self-organization in cellular architecture. *J Cell Biol* 155(2):181-186.
6. Misteli T (2007) Beyond the sequence: Cellular organization of genome function. *Cell* 128(4):787-800.
7. Kosak ST, Groudine M (2004) Form follows function: The genomic organization of cellular differentiation. *Genes Dev* 18(12):1371-1384.
8. Takizawa T, Meaburn KJ, Misteli T (2008) The meaning of gene positioning. *Cell* 135(1):9-13.
9. Croft JA, et al. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145(6):1119-1131.
10. Bolzer A, et al. (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3(5):e157.
11. Kosak ST, et al. (2007) Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol* 5(11):e309.
12. Kosak ST, Groudine M (2004) Gene order and dynamic domains. *Science* 306(5696):644-647.
13. Bruno L, et al. (2004) Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. *Mol Cell Biol* 24(2):741-756.
14. Frigiyik BA, Srivastava S, Gupta MR (2008) Functional Bregman divergence and Bayesian estimation of distributions. *IEEE Trans Inf Theor* 54(11):5130-5139.
15. Cover T, Thomas J (2006) *Elements of Information Theory* (Wiley Interscience, New York), 2nd Ed.
16. Kullback S (1959) *Information Theory and Statistics* (Wiley, New York).
17. Martinez W, Martinez A (2007) *Exploratory Data Analysis with MATLAB* (Taylor and Francis, New York).
18. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167-256.
19. Strogatz SH (2000) From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143(1-4):1-20.
20. Strogatz SH (2001) Exploring complex networks. *Nature* 410(6825):268-276.
21. Strogatz SH, Abrams DM, McRobie A, Eckhardt B, Ott E (2005) Theoretical mechanics: Crowd synchrony on the Millennium Bridge. *Nature* 438(7064):43-44.
22. Arenas A, Diaz-Guilera A, Kurths J, Moreno Y, Zhou C (2008) Synchronization in complex networks. *Phys Rep* 469(3):93-153.
23. Soutoglou E, Misteli T (2008) Activation of the cellular DNA damage response in the absence of DNA lesions. *Science* 320(5882):1507-1510.
24. Kaiser TE, Intine RV, Dundr M (2008) De novo formation of a subnuclear body. *Science* 322(5908):1713-1717.
25. Gan Q, Yoshida T, McDonald OG, Owens GK (2007) Concise review: Epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells. *Stem Cells* 25(1):2-9.
26. Bernstein BE, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2):315-326.
27. Yu J, Thomson JA (2008) Pluripotent stem cell lines. *Genes Dev* 22(15):1987-1997.
28. Takahashi K, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5):861-872.