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]

[a]Division of Basic Sciences and [b]Department of Biostatistics and Biomathematics, Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109; [c]Department of Statistics, University of Washington, Seattle, WA 98195; [d]Department of Radiation Oncology, University of Washington School of Medicine, Seattle, WA 98195; and [e]Department 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, identifying 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.

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

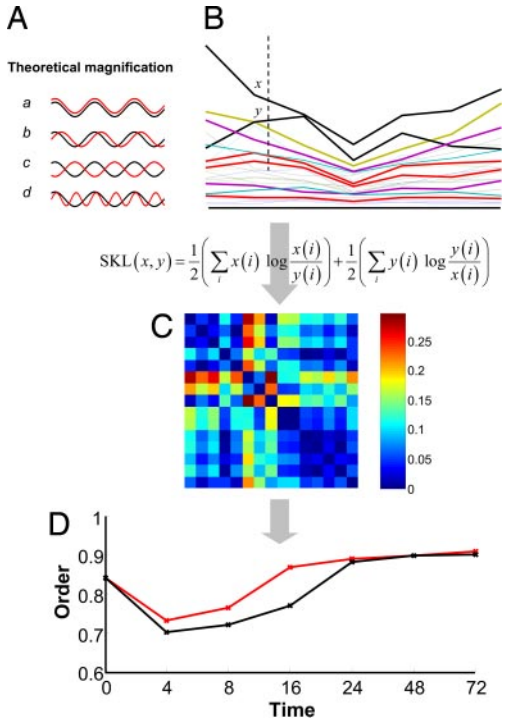

Fig. 1. The order of coregulated gene expression networks increases during differentiation. (*A*) Potential differences between the expression patterns of 2 genes over time. (*B*) Pick any 2 genes (*x* and *y*) and measure the pairwise distance (relative entropy) during differentiation. (*C*) Example of a portion of the **N** × **N** distance matrix consisting of all pairwise comparisons at one time point. (*D*) Order is the fraction of genes at any given time point (0, 4, 8, 16, 24, 48, and 72 h) that are coregulated, thereby residing in a GC that reflects their hierarchical clustering. As genes entrain with or diverge from the expression pattern of the GC during the course of differentiation, the change in order as the system evolves can be determined. In the progenitor state (time 0), the GC corresponds to 84% of coregulated genes, is reduced to as low as 70% at 4 h, and reaches a maximum of 91% in the fully differentiated state. Red and black lines represent erythroid and neutrophil lineages, respectively.

previously reported time course of differentiation from a hematopoietic progenitor (FDCPmixA) to derived erythroid and neutrophil cell types (7 time points, 0, 4, 8, 16, 24, 48, and 72 h) (11, 13). We expressed these individual coregulated gene expression networks as matrices whose cells represent nodes that are linked together with relative strengths (refer to *Dynamics of Networks* in *Methods* for an elaboration on the concept of a network). Changes among nodes over time can occur in many fashions, including differences in periodicities and amplitudes that are the same or inverted (Fig. 1*A*, *a* and *c*). We consider these 2 examples as the basic forms of coregulation in our analysis of dynamical gene expression, representing activation and silencing, respectively. To describe the overall evolution of coregulated gene expression networks during the differentiation of the progenitor to erythroid and neutrophil cell types, we constructed a **N** × **N** distance matrix of the total coregulated gene set for each lineage (N = 1,720 for erythroid, 2,113 for neutrophil, and 3,833—their sum—for the progenitor) and for each time point using the raw expression score with each gene compared with all others. The distance (relative entropy) between gene profiles was established by using symmetrized Kullback–Leibler distances (SKL) (Fig. 1 *B* and *C*) (14–16):

$$\mathrm{SKL}(x,\,y)=\frac{1}{2}\left(\sum_i x(i)\log\frac{x(i)}{y(i)}\right)+\frac{1}{2}\left(\sum_i y(i)\log\frac{y(i)}{x(i)}\right)$$

where $x(i)$ and $y(i)$ are the expression levels of genes x and y at time i. Next, we used hierarchical clustering (17) to find the similarity or dissimilarity for every gene-by-gene comparison, thereby capturing the total organization of the distance matrix. A dendogram of the gene expression at the starting time point—the progenitor state—reveals that there is a very large set of genes that are expressed similarly, creating a giant component (GC). A GC is the core of a network, the connected set of nodes that characterizes its overall behavior or direction. At a given time point, order can therefore be defined as the fraction of genes residing in the GC relative to the total gene set. By following the genes that dynamically entrain with or diverge from the expression pattern represented in the GC, we can determine the change in order as the system evolves, i.e., differentiation occurs. Surprisingly, our analysis indicates that the degree of order represented in the progenitor state decreases as the differentiation program initiates, indicating a phase transition from a defined cell type into one that is amenable to subsequent reorganization (Fig. 1*D*). As the disordered progenitor eventually commits to either the erythroid or neutrophil lineage, there is a concomitant increase in order, eventually stabilizing at a level greater than that of the original multipotent progenitor (Fig. 1*D*). Therefore, we argue that during the differentiation of a progenitor to derived cell types, cellular potential decreases as gene regulatory order increases.

As indicated above, we have previously demonstrated that coregulated gene expression is correlated with the increased proximity of homologous chromosomes, determined in pairwise analysis between chromosomes and the total association of all

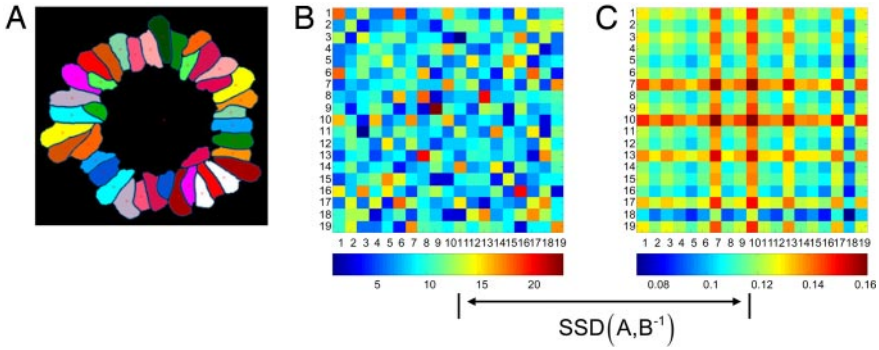

Fig. 2. Coregulated gene expression and chromosomal association networks are related. (*A*) Example of an erythroid rosette with each chromosome identified by using spectral karyotyping (SKY). (*B*) The chromosomal association network represented as a 19 × 19 frequency matrix counting all possible chromosome associations. (*C*) The larger coregulated gene expression network compressed to a 19 × 19 chromosomal association matrix, using the mean of coregulated genes found on each chromosome.

Table 1. Distances between chromosomal association and coregulated gene expression matrices

Comparison	Distance (AB^{-1})	*P*
A_P:B_P	0.064	0.0015
A_E: B_E	0.025	0.0024
A_N:B_N	0.036	0.0052
A_P:B_E	0.316	0.3425
A_P:B_N	0.294	0.2115
A_E:B_P	0.338	0.2463
A_N:B_P	0.341	0.2725
A_E:B_N	0.271	0.2061
A_N:B_E	0.294	0.2134

A and B denote rosette and expression matrices, and subscripts P, N, and E indicate progenitor, neutrophil, and erythroid lineages. We compute the distance between matrix A and B^{-1}, where $B^{-1}$is the inverse of the 19 × 19 transformed coregulated gene expression network. The closer a value is to zero, the greater the shared mutual information. The first 3 rows depict the results for comparisons between matrices derived from the same lineage (intra), whereas the last 6 rows make interlineage comparisons. The *P* values are the result of a permutation test described in the text.

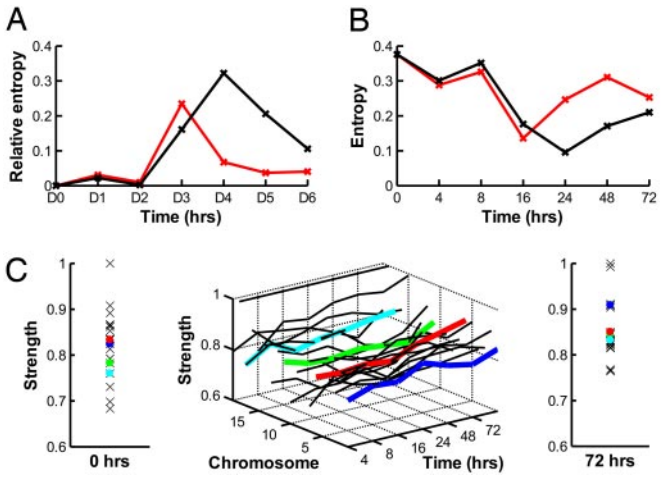

Fig. 3. The organization of chromosomal associations increases during differentiation. (*A*) Graph of the relative entropy of chromosomal associations as computed from the SSDs for the progenitor against successive time points. (*B*) The general entropy (15) for each time point in the differentiation, normalized between 0 and 1. (*C*) Strength of chromosomal contribution to the chromosomal association network during differentiation. (*Left*) The front elevation at 0 h (the progenitor). (*Right*) The elevation from 72 h (erythroid). (*Center*) The total chromosomal contributions over time. Representative chromosomes 1, 6, 11, and 17 are indicated as green, blue, red, and dark blue, respectively.

chromosomes depicted in prometaphase rosettes (Fig. 2*A*; see *Methods* for a description of our measurement of chromosomal associations) (11). By expressing the frequency of chromosome associations in rosettes as a 19 × 19 symmetric matrix (sex chromosomes are not included in our analysis), denoted A (*Methods*) (Fig. 2*B*), we can now analyze the chromosomal association network in its entirety, looking at the relationship among all homologs and heterologs. To make the comparison between chromosomal associations and gene expression among the progenitor, erythroid, and neutrophil cell types, we must first convert each of the larger coregulated gene expression matrices to a comparable 19 × 19 matrix, denoted B, by mapping the coregulated genes onto their respective chromosomes (Fig. 2*C*). We define an element (b_{kl}) in B where *k*,*l* represent any 2 of the 19 autosomal chromosomes as follows:

$$b_{kl} = \underset{x,y}{\text{mean}}\ (\text{SKL}(x, y))$$

where (x, y) ranges over all gene pairs. If $k \neq l$, gene x is on chromosome k and gene y is on chromosome l, whereas if $k = l$, gene x and gene y are on the same chromosome. If gene coregulation is related to the overall proximity of chromosomes, then the 2 types of matrices should be inversely related: The rosette data are an actual frequency count with a larger number indicating greater overall physical proximity; the expression data are a measure of the distance between genes' expression profiles, with smaller values revealing less distance and therefore indicating coregulation. To test this relationship we computed the symmetrized Stein distance (SSD) (*Methods*) (15) between A and the inverse of B as follows:

$$\text{SSD}(\text{A}, \text{B}^{-1}) = \text{trace}(\text{AB}) + \text{trace}(\text{A}^{-1}\text{B}^{-1}) - 2d$$

where $d = 19$ is the size of the matrices (*Methods*).

$$\text{SSD}(\text{A}, \text{B}^{-1}) = 0$$

implies that A and B^{-1} are identical, having maximum mutual information. For the comparison in each lineage, the distance between matrices is close to zero, indicating that gene coregulation is in fact manifested in overall chromosomal organization (Table 1). In support of the lineage specificity of these relationships, comparing the chromosomal matrix (A) from one cell type to the gene expression (B) of another yields SSD values closer to 1 (Table 1). To determine the statistical significance of these results, we performed a permutation test, keeping the chromosomal association matrix constant and permuting the rows and columns of the coregulated gene expression matrix elements while maintaining their symmetry. After each permutation of the coregulated gene expression matrix (10,000×) we computed the $\text{SSD}(\text{A},\text{B}^{-1})$. We then compared the observed value of $\text{SSD}(\text{A},\text{B}^{-1})$ to this permutation distribution to obtain a *P* value for its significance. Each of the 3 lineages demonstrates a significant difference from the permutation, whereas the interlineage comparisons do not (Table 1). Therefore, the overall organization of chromosomes at the onset and conclusion of differentiation is a reflection of the unique expression pattern of each cell type.

Having demonstrated that the coregulated gene expression data can be visualized as rosette-based chromosomal association matrices, we next asked how overall chromosomal organization evolves during differentiation. To do this, we used the coregulated gene expression matrices from the remaining 5 time points (Fig. 1) and converted each to a 19 × 19 matrix (B), as we did for the progenitor and derived cell types above, and then transformed these matrices to a corresponding chromosomal association matrix (A) (where $\text{A} \simeq \text{B}^{-1}$). With these distance matrices based on gene regulation, we determined the change in relative entropy (SSD) among chromosomes, comparing the matrix of the progenitor to each successive time point (e.g., D6 is the difference in entropy between the progenitor and the differentiated cell types) (*Methods*) (Fig. 3*A*). This analysis reveals that there is decreased entropy in chromosomal organization at defined stages of differentiation to the erythroid and neutrophil lineages (Fig. 3*A*). Importantly, the initial decrease in entropy coincides with the increasing reordering of gene expression (compare with Fig. 1*D*). We also determined the overall entropy (15) at each time point, which supports the observation that cellular commitment leads to decreased entropy (Fig. 3*B*). In both types of analysis, the overall order of chromosomes is greater in the differentiated cell types than the progenitor. Therefore, chromosomes can be viewed as reorganizing accord-

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 ijth 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. 3*C*). In Fig. 3*C 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. 3*C 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

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

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

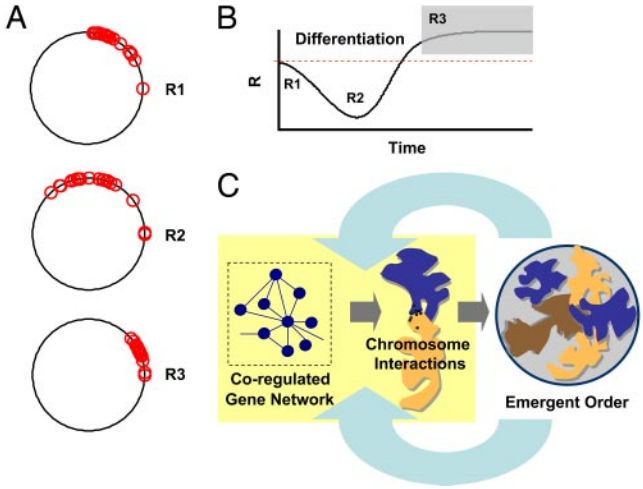

Fig. 4. Chromosomes self-organize according to coordinate gene regulation. (*A*) Geometric interpretation of the chromosomal network, where the phases of θ_i are plotted on the unit circle. Three snapshots during differentiation are represented, where R1 is the distribution of the network representing 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 equation (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. 4*C*). 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. 4*C*). 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

recent study has directly demonstrated features of self-organization in the de novo formation of Cajal bodies, involved in the processing of small nuclear ribonucleoproteins, by the nonhierarchical localization of constituent proteins (24).

Beyond the organization of specific activities into nuclear bodies, the nucleus also demonstrates a general compartmentalization of transcriptional activity. One example of this is gene loci assuming developmentally regulated positions within the nucleus (7). We have reasoned that if this behavior was extended over the total complement of genes that are regulated during the differentiation of a specific cell type, then lineage-specific nuclear topologies may occur (12). In testing this idea, we have previously shown that the genome—in both the linear distribution of genes and the association of homologous chromosomes—is organized according to coregulated gene expression during the in vitro differentiation of a hematopoietic progenitor (FDCPmixA cells) to derived erythroid and neutrophil cell types (11). We have therefore suggested that the genome—at the level of chromosomes—may self-organize to facilitate coordinate gene regulation during cellular differentiation. Here, we present a test of this hypothesis, applying the mathematical approaches of distance matrices and coupled oscillators to our datasets of gene expression and chromosomal associations (in the form of rosettes) from the differentiation of the progenitor to the erythroid and neutrophil lineages.

A major difficulty in comparing total coordinate gene expression with total chromosomal association patterns is finding similarities embedded in high-dimensional datasets. In our current study, we have found a solution to this problem by depicting the networks of gene expression and chromosomal associations as matrices, allowing us to compare them in their entirety to uncover shared features. We believe this approach may be useful for a broad range of biological questions dealing with information-rich datasets. Our analysis determined that coordinate gene expression undergoes a phase transition—characterized by an increase in entropy—upon commitment of the progenitor. As differentiation continues, there is a gradual loss of entropy, culminating in a highly ordered state in the differentiated cell types (Fig. 1). The coregulated gene sets of the semiordered progenitor and ordered erythroid and neutrophil lineages are significantly correlated with lineage-specific chromosomal association patterns (Fig. 2 and Table 1). Furthermore, by transforming the gene expression networks along the time course to corresponding chromosomal association matrices, we demonstrated that chromosomal topologies change dynamically during differentiation but, as with gene expression, result in a more highly ordered state in the differentiated cell types (Fig. 3). Finally, we modeled this relationship by employing nonlinear dynamical systems theories (i.e., the concept of coupled oscillators), revealing the mutual entrainment—or self-organization—of coordinate gene regulation and chromosomal associations (Fig. 4).

That the genome self-organizes according to gene coregulation may be broadly relevant for cell fate decisions, in particular those of progenitor/stem cells. The phase transition we observe upon commitment of the hematopoietic progenitor suggests 2 salient features of a multipotent state: (*i*) a progenitor/stem cell is a defined cell type in terms of its genomic organization; and (*ii*) this semiordered state must be disrupted for a subsequent identity or higher-order state to be established. Additionally, our demonstration that terminal differentiation coincides with the stabilization of this highly ordered state indicates that cellular potential is inversely related to genomic order. A hallmark of stem cells is their ability to undergo asymmetric division, allowing both self-renewal and commitment to a differentiation pathway. We suggest that this process may be related to our observation of a phase transition: A daughter cell that commits to a program of differentiation will have a marked increase in genomic entropy, whereas the self-renewing cell will not. Recent studies have demonstrated bivalent chromatin marks underscoring the "primed" expression profile of stem cells, in particular ES cells (25, 26). The resolution of this bivalency and subsequent differentiation may require a brief increase in disorder for the establishment of a pattern of chromatin modifications that potentiates a given differentiation pathway. Based on our results, we would further suggest that this reordering would be manifested in a discernable chromosome topology. Future work comparing gene expression analyses and chromatin compartments both in silico and in situ will be necessary to validate these claims.

Although asymmetric division of a stem cell was once thought to be unidirectional, resulting in the irreversible commitment of a daughter cell, recent evidence indicates that it is possible to induce pluripotent stem (iPS) cells from differentiated cell types (27, 28). Because these strategies rely heavily on the expression of transcription factors to induce pluripotency, it is likely that the changes leading to iPS cells are reflected in nuclear organization. Our demonstration of dynamic genomic reorganization through a phase transition may provide a means to refine iPS strategies for a wide variety of differentiated cell types. For example, by identifying the GC of gene expression during the differentiation of a particular cell lineage, we can determine the gene set that drives the phase transition and thereby reverse-engineer the genetic network to that critical juncture. Knowing the GC of a given differentiation program permits the identification of the transcriptional regulatory factors significant in the process, either by analysis of functionally characterized regulatory elements and/or the alignment and description of shared domains. These factors could then be ectopically expressed to drive a differentiated cell back to the disordered state of the phase transition wherein it is amenable to multiple pathways. Beyond a possible role in the asymmetric cell division of stem cells or iPS cell strategies, our demonstration of nuclear self-organization may be of importance to any feature of the nucleus representing a dynamical system.

Methods

Symmetrized Kullback-Leibler (SKL) Divergence. For any two points x and y of $\Delta \subseteq \mathbb{R}^d$, the Bregman divergence $B_f(.,.)$: $\Delta \mapsto \mathbb{R}$ of x to y, corresponding to a strictly convex and differentiable function F, is defined as

$$B_F(x, y) = F(x) - F(y) - \langle \nabla F(y), (x - y) \rangle$$

where ∇F is the gradient of F and $\langle .,. \rangle$ the inner product. SKL divergence can be derived by substituting the convex function: $F(x) = \sum_i x(i) \log x(i)$ (14, 16). We define SKL $(x,y) = 1/2\ B_F(x,y) + 1/2\ B_F(y,x)$, then

$$\mathrm{SKL}(x, y) = \frac{1}{2}\left(\sum_i x(i) \log \frac{x(i)}{y(i)} \right) + \frac{1}{2}\left(\sum_i y(i) \log \frac{y(i)}{x(i)} \right).$$

Note that x and y should be nonnegative. It can be shown that SKL(x,y) is always nonnegative, $0 \leq \mathrm{SKL}(x,y)$, and $\mathrm{SKL}(x,y) = 0$ iff $x = y$ (16).

Chromosomal Association Analysis. As reported previously (11), total chromosomal organization was analyzed in prometaphase rosettes from progenitor, erythroid, and neutrophil cells. At least 30 rosettes were examined for each cell type. Briefly, chromosome identity was determined through spectral karyotyping (SKY) and a pattern recognition algorithm was used to assay chromosomal associations in the rosette. The relative proximity of each chromosome to all others was measured. Chromosomes were considered to be associated if they were removed by no more than 2 chromosomes. Results from rosettes were verified in the interphase nucleus (11).

SSD.

$$\begin{aligned} \mathrm{SSD(X, Y)} &= \mathrm{SD(X, Y)} + \mathrm{SD(Y, X)} \\ &= \mathrm{trace(XY^{-1})} + \mathrm{trace(X^{-1}Y)} - 2d \end{aligned}$$

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:

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

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:

$$\mathrm{D}_0 = \mathrm{SSD}((\mathrm{B}_0)^{-1},(\mathrm{B}_1)^{-1}),$$

$$\mathrm{D}_1 = \mathrm{SSD}((\mathrm{B}_0)^{-1},\ (\mathrm{B}_2)^{-1}),\ \ldots\ldots,$$

$$\mathrm{D}_6 = \mathrm{SSD}\ ((\mathrm{B}_0)^{-1},\ (\mathrm{B}_7)^{-1}),$$

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 $x_i = [x_i, y_i, z_i,\ldots]^{\mathrm{T}} \in \mathbb{R}^N$ is the state vector of node i for $i = 1,2,\ldots,n$ 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, $C = [1,0,\ldots,0]^{\mathrm{T}}$, 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) *Sync: 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. Frigyik 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, Díaz-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.