Deciphering the molecular mechanism of the cancer formation by chromosome structural dynamics

Xiakun Chu¹, and Jin Wang^{1,2,*}

¹ Department of Chemistry

² Department of Physics and Astronomy

State University of New York at Stony Brook, Stony Brook, NY 11794, USA

* jin.wang.1@stonybrook.edu

Abstract

Cancer reflects the dysregulation of the underlying gene network, which is intimately related to the 3D genome organization. Numerous efforts have been spent on experimental characterizations of the structural alterations in cancer genomes. However, there is still a lack of genomic structural-level understanding of the temporal dynamics for cancer initiation and progression. Here, we use a landscape-switching model to investigate the chromosomal structural transition during the cancerization and reversion processes. We find that the chromosome undergoes a non-monotonic structural shape-changing pathway with initial expansion followed by compaction during both of these processes. Furthermore, our analysis reveals that the chromosome with a more expanded structure than those at both the normal and cancer cell during cancerization exhibits a sparse contact pattern, which shows significant structural similarity to the one at the embryonic stem cell in many aspects, including the trend of contact probability declining with the genomic distance, the global structural shape geometry and the spatial distribution of loci on chromosome. We show that cell cancerization and reversion are highly irreversible processes in terms of the chromosomal structural transition pathways, spatial repositioning of chromosomal loci and hysteresis loop of contact evolution analysis. Our model draws a molecular-scale picture of cell cancerization, which contains initial reprogramming towards the stem cell followed by differentiation towards the cancer cell, accompanied by an initial increase and subsequent decrease of cell stemness.

1 INTRODUCTION

Currently, our knowledge of how cancer cells form and proliferate is still quite limited. In general, the development of cancer is controlled by the underlying gene regulatory network within a cell and cell-cell interactions [1], which rely on the molecular interactions between the spatially organized genome and a broad class of proteins, including the transcription factors and chromatin remodelers [2]. As the structural scaffold for the genome function, the 3D genome architecture has been recognized to play an important role in regulating the gene expression [3, 4, 5], leading to the intimate structure-function relationships at the genomic level. Several studies have shown that alterations in chromosome structures and interactions can

make significant contributions to the dysregulation of the gene expressions forming the specific cancer signatures [6, 7, 8, 9]. These findings have fostered a view that the 3D context of the genome is the major player in the development and progression of cancer [10]. Therefore, investigating the disorganization of the 3D genome structure in cancer cells can provide the key to understanding the pathogenesis of cancer.

Although the chromosomal structural variant, which is the major form of the genome instability, has been regarded as a hallmark of almost all human cancers [11, 12, 13], determining the genome structure has been a long-term challenge until the emergence of chromosome conformation capture (3C) techniques nearly two decades ago [14]. As an advanced derivative of the 3C, Hi-C measures the spatial proximity of the chromosomal loci across the entire genome in terms of the contact frequency map and provides an ensemble description of the genome organization within a large number of cells [15, 16]. Recently, Hi-C techniques were applied to characterize the disorganization of the cancer genomes and determine their functional consequences [17]. Taberlay et al. observed the smaller-sized topologically associated domains (TADs) formed in prostate cancer cells than those in normal cells due to the emergence of additional domain boundaries. This further leads to alterations of TP53 tumor suppressor locus [18]. In addition, the effects of the structural disruptions within TADs on leading to various cancer types were established [19, 20, 21]. At the higher hierarchical level, Barutcu et al. found that the significant A/B compartment switching between the normal and breast cancer cell is associated with gene expression changes [22]. These studies have shown that the structural variants in the cancer genome occur throughout the genome sequence, and they have important impacts on the gene expression alterations for inducing cancer formation.

Increasing evidence indicates that cancer development is orchestrated by a small subpopulation of the cancer cells, namely the cancer stem (CS) cells. [23, 24, 25, 26]. CS cells, which possess stem-like properties and functions, are capable of performing self-renewal, proliferation and differentiation. Thus they provide the driving forces for cancer progression [27, 28]. However, experimental identifications on the CS cell and the associated dynamics are non-trivial due to its rare population in the cancer cells [29]. The previous theoretical studies using gene regulatory work, which includes the cancer and development genetic markers as well as the interactions between them, provided a landscape view of how the cancer cell develops and

CS cell forms [30, 31]. However, the work focused on simplified core gene regulatory networks to describe the cancer systems, so the results were restricted in the form of regulation pathways among several marker genes and this description is at the gene network level. Therefore, there is still a lack of a complete molecular chromosomal-structural level understanding of cancer development and CS cell formation. On the other hand, the strategies on inducing the cancer cell to normal cell by reversion [32] or transdifferentiation processes [33, 34, 35] are rapidly developing. These approaches have opened new avenues for cancer treatments, but the underlying mechanism of the cancer reversion is still unclear [36].

Despite the significant achievements made by the Hi-C technique on elucidating the spatial genome disorganization in the cancer cells [17], the temporal dynamic rearrangement of the genome structure during cancer development is still not available. The picture is fundamentally important as it may provide a molecular-level understanding of the cancer mechanism. However, measuring the chromosomal structural dynamics during the cancer cell developmental process is extremely challenging due to the spatial and temporal resolution limits of the current experiments. Here, we aim to address this issue with the computational efforts by investigating the dynamic cancer-related chromosomal structural evolution, which provides the microscopic description of the cancer process.

We developed molecular dynamics models and performed associated simulations to undertake the task. First, we applied the maximum entropy principle to incorporate the Hi-C data of the normal and cancer cells into two independent sets of simulations. These simulations generated two potentials for describing the chromosome structural ensembles in the normal and cancer cells, respectively. In the cell nucleus, the chromosome constantly experiences the non-equilibrium effects even at one cell state, leading to a non-equilibrium system. Theoretically, it has been demonstrated that the dynamics of the non-equilibrium system can be described using the concept of an effective landscape in some circumstances [37, 38]. Recent studies revealed that under an effective equilibrium landscape, the chromosome dynamics reproduces many aspects of kinetic behaviors as observed in experiments, including anomalous diffusion, viscoelasticity, and spatially coherent dynamics [39, 40]. Here, to examine whether the potentials accounting for chromosome structures obtained from maximum entropy principle simulations can also be used for describing the chromosome dynamics, we performed additional simulations under these two potentials and calculated the diffusion behaviors of the chromosome motion in the normal and cancer cell, respectively (Figure S3). We observed sub-diffusivity of chromosome dynamics with the scaling exponents of the mean square displacement in good agreement with multiple experiments [41, 42]. Our results have suggested that these two potentials generated by the maximum entropy principle simulations for chromosomes in the normal and cancer cell can be further regarded as the effective energy landscapes that govern the chromosome conformational dynamics within the normal and cancer cell, respectively. Although each state can be described by an effective equilibrium landscape, the inter-basin dynamics of switching between normal and cancer states often requires a significant amount of energy input [43] and therefore is a non-equilibrium process, which cannot be treated

in an equilibrium way. This is the motivation and target of our current study. Therefore, the cancer-related chromosome structural transitions are further described by the connection between these two effective landscapes during the cell cancerization and reversion processes.

To bridge these two landscapes for describing transitions between the normal and cancer cell, we then used a landscape-switching model, which was developed to simulate the chromosomal structural transition during the cell cycle [44] and the cell developmental processes [45]. The model employed an instantaneous energy excitation to trigger the transition processes followed by the relaxation dynamics (see **"Materials and Methods"**). The switch implementation in the model has a practical correspondence to mimic the roles of the genetic mutations and epigenetic modifications in initiating cancer processes [46, 47]. The rationale of approximating the transitions between the normal and cancer cell to simple switches are based on the following two facts.

1. The cancer process often exhibits switch-like behavior between two steady cell states, in accordance with the landscape-switching model. As a dysregulated cell developmental process [48], cancer progression is often determined by the regulation of the bistable gene expression states [49, 50], which correspond to the normal and cancer cell. Previous theoretical studies using a simple circuitry of the bistable switch between distinct gene expression states to study the cancer development well captured many characteristics of the cancer process [51, 30, 31]. The successes of these simplified models are likely due to the fact that the cell-fate decision-making processes can often be approximated to the transitions of the bistable switch [52, 53, 54].

2. From the physical perspective, cell cancerization and reversion are non-equilibrium non-adiabatic processes, which can be described by the landscape-switching model. The cancer progression requires extensive energy supplies that break the detailed balance of the system, leading to non-equilibrium dynamics. Furthermore, the transitions between the normal and cancer cell are impossible to occur spontaneously, so the waiting time of the inter-landscape hopping is expected to be much longer than the timescale of the intra-landscape dynamics. This feature of relatively slower inter-landscape dynamics compared to intra-landscape dynamics signifies a non-adiabatic process [55, 56]. In analogous to the surface hopping method [57], we separated the simulations of adiabatic chromosome dynamics at one cell state and non-equilibrium non-adiabatic inter-state switching dynamics, giving rise to the landscape-switching model.

We used the landscape-switching model to study the chromosome structural dynamics during the transitions between the human normal and cancer lung cell. We predicted that the chromosome can form more expanded structures than those at the normal and cancer cell during both of these processes but with different structural characteristics. Further analyses revealed that the chromosome at the transient intermediate state with a more expanded structure than those at the normal and cancer cell during the cancerization process share significant structural similarity to the one at the embryonic stem (ES) cell. This feature implies forming a cell state with the features of stemness or the CS cell during the cancerization process. We observed the high irreversibility for cancerization and rever-

sion through the quantified structural transition pathways, the spatial repositioning of chromosomal loci and the hysteresis loop of establishing the chromosomal contacts. These findings underlined distinct mechanisms for these two cancer cellular processes. The prediction of forming a stem-like intermediate cell state during cancerization can be tested by the future time-course Hi-C experiments designed for the cancer developmental processes.

2 RESULTS

2.1 Chromosomal structural transition during cancerization and reversion

We used the landscape-switching model to investigate the chromosome dynamics during the cell cancerization and reversion processes. The implementation of the model is briefly summarized as follows. First, we iteratively fit a generic polymer model to reproduce the experimental Hi-C data through the maximum entropy principle simulation for the terminally differentiated human lung fibroblast cell (IMR90) and lung cancer cell (A549) (Figure S1), separately [59]. Previous studies have shown that the resulting potential not only captures the thermodynamics of the chromosome (i. e., Hi-C) [60, 61], but also describes the correct kinetic properties of the chromosomal loci diffusion within one cell state (one phase in cell cycle or one cell state in cell differentiation/reprogramming) [39, 40]. Here, we term the potential $V(\mathbf{r}|S)$ as the effective energy landscape for describing the chromosome dynamics in the normal or cancer cell, where $\mathbf{r}$ is the coordinate of the system at the cell state S (IMR90 or A549). Next, the simulation was set up with the chromosome exploring the structural dynamics under either the energy landscape of the normal ($V(\mathbf{r}|IMR90)$) or cancer ($V(\mathbf{r}|A549)$) cell. Then, the energy landscape underwent a switch from normal to cancer cell ($V(\mathbf{r}|IMR90) \rightarrow V(\mathbf{r}|A549)$) or cancer to normal cell ($V(\mathbf{r}|A549) \rightarrow V(\mathbf{r}|IMR90)$) to trigger the chromosomal structural transition during cancerization or reversion, respectively. Finally, the chromosome dynamics was governed by the post-switched energy landscape of the cancer cell ($V(\mathbf{r}|A549)$) or normal cell ($V(\mathbf{r}|IMR90)$). The model allowed us to observe the chromosome transformation during the cancerization and reversion processes with affordable computational resources (Figure 1).

We performed hundreds of independent landscape-switching simulations starting from different chromosome structures in the normal (and cancer) cell state to investigate the chromosomal structural transitions during the cancerization (and reversion) process (See "Materials and Methods"). By comparing the similarity of the contact probability formed through the pairwise chromosomal loci between time-series frames during the cancerization process ($\Delta P^{I,J}$, where P is contact probability and I and J are the time points), the transition can be reduced into two clusters according to the hierarchical clustering of $\Delta P^{I,J}$ along with the processing time (Figure 1A). However, the clustering dendrograms for the local and non-local chromosomal contacts are very different, indicating distinct behaviors of structural rearrangements at the local and non-local ranges. The separation of the two

clusters at the local contacts occurs early at $\sim 0.1\tau$, where τ is the time unit in the simulation. In contrast, the non-local chromosomal contacts at the beginning of the transition are relatively similar to those at the late stages, and they are separated by the second cluster at $\sim 0.5-10\tau$. This is also reflected by the contact probability $P(s)$ versus the genomic distance s between a pair of chromosomal loci (Figure 1C), where an apparent decrease of $P(s)$ is observed at time $\sim 0.5-10\tau$.

By examining the clustering dendrogram patterns of the local and non-local contact formations in the chromosome, we further reduced the cancerization process into 8 stages (Figure 1B). As a result, chromosomes within one stage are structurally similar in terms of the local and non-local contacts. The contact maps of the 8 stages show a picture of how chromosomes perform the structural rearrangements in the cancer formation process (Figure 1D). Interestingly, we found that as the cancerization proceeds, the chromosome heat maps show lower probabilities (more sparse contacts in heat maps) than those in both normal and cancer cell. The result implies that the chromosome structure is expanding during cancerization. Besides, we observed a complex manner of the chromosome in organizing the compartment change during cancerization shown by the colored bands in the circle plots (Figure 1E). From the stage "C1" to "C3", the populations of the chromosomal loci in compartment A increase associated with an increasing quantity of contacts (lines in each circle plot) formed within compartment A. At the stage "C4", the chromosomal loci are roughly assigned to compartment A and B by sequence and the long-range contacts are rare. This implies the extensive breaking of the contacts formed in the chromosome at the normal cell. As cancerization further proceeds, a significant compartment-switching occurs during the stage "C4" to "C5". From the stage "C5" to "C8", the chromosomal compartment profiles and contact patterns gradually adapt to those at the cancer cell. Finally, we see that the chromosome in the cancer cell has more chromosomal loci in compartment A and forms more sparse contacts than the one in the normal cell. In the ES cells, chromosomes possess high populations of euchromatin, which maps with compartment A, to benefit the cell pluripotency and differentiation [62, 63, 64]. Our observation of the chromosome with populated expressive loci and open structure in the cancer cell resonates with the experimental evidence that the cancer cells possess certain contents of the "stemness" in favor of the self-renewal, differentiation, and proliferation [23, 65, 66].

Then we applied similar analyses into the reversion process. We found that the chromosomal structural transition during the reversion process is not the simple reversal of the cancerization. From the chromosomal contact formation perspective, the reversion process can also be grouped into two clusters (Figure 1F). However, these two clusters are separated by the time in sequence from the dendrogram plots. In detail, the time for separating the two clusters at local contact ($\sim 0.1\tau$) occurs much earlier than that at non-local contacts ($\sim 10\tau$). This indicates that chromosome structurally evolves faster at local than non-local range during reversion. Interestingly, the time evolution of $P(s) \sim s$ also shows a chromosomal structural expansion at the time $\sim 0.5-10\tau$ (Figure 1H). However, the expansion appears to be less significant than that observed

during cancerization. The chromosomal contact maps show distinct evolving ways deviated from the reversal of that in cancerization (Figure 1I). In particular, there is no significant compartment-switching during the reversion process, as the compartment undergoes gradual and moderate change, associated with reducing the contacts formed within compartment A (Figure 1J). Overall, we found that the chromosomal structural transitions undergo initial expansions followed by compactations during both the cancerization and the reversion processes but through irreversible paths.

2.2 Chromosome structural expansion during cancerization and reversion

To examine the chromosomal structural expansion during cancerization and reversion, we monitored the contact probability $P(s)$ versus the genomic distance s . $P(s)$ provides information on how the chromosome organizes its structure along the genomic sequence, thus it dictates the polymer state of the chromosome [67, 68]. Chromosomes in both the normal and cancer cells show higher $P(s)$ plots, corresponding to more compacted structures, than the one in the ES cell [64] (Figure 2A). This can also be inferred by the slope in the logarithmic relation of $P(s) \sim s$ (Figure 2B). As the normal cell converts to the cancer cell, the profile of $P(s)$ (colored by time) initially goes down and approaches the one at the ES cell (blue line), then increases to the cancer cell (purple line). The result draws a picture of the chromosome expanding its structure followed by compaction during cancerization. To quantitative measure this trend, we calculated the slope in the logarithmic relation of $P(s) \sim s$. We see that the slope at $\sim 2\tau$ has the lowest value and approximates the value found at the ES cell. This implies that the chromosome with the maximum structural expansion possesses a similar contact formation scaling versus the genomic sequence with the one at the ES cell. In other words, the chromosome may form the chromosomal structural characteristics of the ES cell during the cancerization. Besides, it is worth noting that the chromosome in the cancer cell is more expanded than it is in the normal cell, as the $P(s)$ plot for the cancer cell is slightly below the normal cell.

To measure the similarity of the chromosomal contact formation with that in the ES cell during cancerization, we calculated the average of the contact probability similarity $\langle \Delta P(t) \rangle$ between the processing state at time point t during cancerization and the normal, cancer and ES cell at the local and non-local ranges (Figure 2C and 2D). We found that the chromosome at $\sim 0.5-10\tau$, when the chromosome appears to be more expanded than those at the normal and cancer cell, as indicated in Figure 2B, $\langle \Delta P(t) \rangle$ of the normal and cancer increase significantly while $\langle \Delta P(t) \rangle$ of the ES cell decreases. This implies that the contact formation in the chromosome with more expanded structure than those at the normal and cancer cell, becomes similar to the one at the ES cell. The results indicate that the chromosome may go through stem-like structures to accomplish the cancerization process.

In contrast, the chromosome expansion in the reversion process from the cancer to the normal cell is not significant (Figure 2E). When the chromosome has the most significant structural expansion with the lowest $P(s)$, the profile of $P(s)$ still visu-

Figure 2: The time evolution of the chromosomal contact probability during cell cancerization and reversion. (A) The time evolution of contact probability $P(s)$ versus genomic distance s during cancerization. The Hi-C data of IMR90, A549, and ES cells are also plotted. (B) The time evolution of the slope in the logarithmic relation of $P(s) \sim s$ at the range of 0.5-7 Mb during cancerization. The differences in (C) local and (D) non-local $\langle \Delta P(t) \rangle$ relative to the Hi-C data of IMR90, A549, and ES cells. (E-H) are the same with (A-D) except for the reversion process.

ally deviates from the one in the ES cell. The steepest slope in $P(s) \sim s$ shows a slight decrease from that in the cancer cell and occurs at the time $\sim 1-3 \tau$. Besides, we can still observe that the similarity of the chromosomal contact formation between the processing state and the ES cell increases, but it is not as much as that observed during cancerization (Figure 2G and 2H). Therefore, the chromosome during the reversion process may possess less structural properties at the ES cell than during cancerization.

2.3 Quantified chromosomal structural transition pathways during cancerization and reversion

Figure 3: The pathways of chromosomal structural transitions during cell cancerization and reversion. The pathways are projected onto the order parameters of the chromosome with all trajectories presented for cancerization (*Left*) and reversion (*Middle*). The 8 stages are indicated by triangles ("C1"- "C8" in cancerization, *Left*) and diamonds ("R1"- "R8" in reversion, *Middle*) on the averaged pathways of the cancerization (solid line) and reversion (dashed line), respectively. The averaged pathways are additionally shown and colored by time (*Right*). The quantities of the chromosome in the IMR90, A549, and ES cells are correspondingly placed as brown, purple and blue points, respectively. The pathways are projected onto (A) the extensions of the longest and the shortest principal axes (PA1 and PA3), (B) the radius of gyration (R_g) and the aspheric quantity (Δ), contact similarity in terms of the fraction of native contact Q to the IMR90 and A549 at (C) local range (0-2Mb), and (D) long-range (10-15Mb). The data of the ES cell are obtained from our previous work [45].

In order to obtain a quantitative picture of how the chromosomal structural transition occurs during the cancerization and the reversion processes, we projected all the transition trajectories as well as the averages onto several order parameters, which describe the shapes of the chromosome structure and

contact formation at various ranges (Figure 3 and Figure S4). We first used the extension lengths of the chromosome structure along the longest and shortest principal axes (PA1 and PA3) [61] (Figure 3A). The individual pathways show high stochasticity, reminiscent of the stochastic dynamics in cell development [69]. During both the cancerization and reversion processes, there is an increase followed by the decrease of shape extension on the chromosome structure along with both PAs. However, from the average pathways, we see that the transitions of these two processes do not follow the same routes. This is also observed by projecting the pathways onto the radius of gyration (R_g) and aspheric quantity (Δ). Δ measures the asphericity of the chromosome structure, and the deviation of Δ from 0 measures the deviation from the perfect sphere [70]. Besides, we can see that the shape of the chromosome with the most significant expanded structure is geometrically different from the one in the ES cell. However, the chromosome at the stage "C4" in cancerization is closer to the ES cell than the stage "R4" in reversion.

The chromosome structural expansion during the cancerization and reversion can also be reflected by the local and non-local contact formations (Figure 3C and 3D). Here we used the fraction of "native" contact Q , which was widely applied in protein folding [71], with references being the average pairwise distances between loci in chromosome ensembles at the normal and cancer cells. There are different pathways for local chromosome structural formations during these two processes, which appear to deviate significantly from the chromosome in the ES cell. For non-local contacts, the pathways become overlapped but still show differences at the most expanded chromosome structure. The results imply that the highly irreversible chromosomal structural transition occurs universally at both the local and non-local range.

2.4 Spatial rearrangement of chromosomal loci during cancerization and reversion

It has been recognized that the spatial distribution of the chromosomal loci strongly influences the transcriptional activity [72, 73]. To see how the chromosome rearranges the spatial distribution of the chromosomal loci along with the compartment formation during cancerization and reversion, we calculated the radial density $\rho(r)$ of chromosomal loci and further classified it into the compartment A and B. We found that the normal cell, which is a terminally differentiated cell, tends to locate the chromosomal loci of compartment A and B towards the chromosome's surface and interior, respectively (Figure 4A). The finding is in line with previous simulations [74, 40, 75] and experiment [76]. In contrast, the chromosome in the ES cell exhibits a roughly uniform distribution of the chromosomal loci regardless of the compartment states. For the cancer cell, $\rho(r)$ appears to be the intermediate between those at the normal and ES cell. We further calculated the radial density similarity $\Delta\rho(t)$ between the processing state at time point t and the normal, cancer and ES cell for the total loci and the loci in compartment A and B (Figure 4B). When the normal cell converts to the cancer cell, $\Delta\rho(t)$ of the ES cell decreases and reaches the minimum at time $\sim 0.9-3\tau$, which are varied by the total loci and the loci in the compartment A

Figure 4: The change of the radial density in the chromosome during cell cancerization and reversion. (A) The change of the radial density profile $p(r)$ for the whole loci (*Left*), loci in compartment A (*Middle*) and loci in compartment B (*Right*) in the chromosome during cancerization. The profiles of the IMR90, A549 and ES cells are colored brown, purple and blue, respectively. (B) The difference of the radial density, calculated by $\Delta p(t) = \int_0^{R_0} |\rho(r,t) - \rho(r,Ref)| dr/R_0$, from the processing time (t) to the reference cells (Ref). (C) Structural illustrations of the chromosome with loci colored by compartment states during cancerization. (D-F) are the same with (A-C) except for the reversion process.

and B (Figure 4B). We note that this period corresponds to the stage when chromosome forms more expanded structure than those at the normal and cancer cell (Figure 2B), so the results indicate that the spatial distribution of the loci in chromosome during this period is similar to that in the ES cell. Further proceeding cancerization decreases $\Delta\rho(t)$ of the cancer cell. Overall, we found a non-monotonic spatial repositioning of the chromosome loci during cancerization. In the beginning, the chromosome segregates the active loci located in compartment A and inactive loci located in compartment B towards the surface and interior, respectively (Figure 4C). Then, the compartment segregation of the chromosome is still effective, but it forms a stem-like pattern with a uniform radial distribution of genomic loci regardless of the compartment states. Finally, the spatial distribution of the genomic loci reaches the one at the cancer cell.

Interestingly, we observed a similar trend of spatial repositioning of the chromosomal loci during the reversion process (Figure 4(D-F)). In other words, $\rho(r)$ during reversion also goes through processing states sharing similarity with that in the ES cell. In detail, we found that $\Delta\rho(t)$ of the ES cell decreases more for the loci in compartment A than the ones in compartment B at time $\sim 1-3\tau$ (Figure 4E), corresponding to the stage when chromosome forms more expanded structures than those at the normal and cancer cell (Figure 2F). This is different from that observed in the cancerization process, where spatial repositioning of the chromosomal loci in compartment B is dominant to be the stem-like, as decreasing $\Delta\rho(t)$ of the ES is more significant for the loci in compartment B than the ones in compartment A. The combined results suggest that during both cancerization and reversion processes, the chromosome dynamically rearranges the spatial positions of the chromosomal loci and the distribution can resemble the one at the ES cell.

2.5 Stemness and irreversibility of chromosomal structural transitions during cancerization and reversion

Our results have implied that the chromosome during the cancerization and reversion may adopt the chromosome structure in the ES cell. To quantitatively assess this, we divided the contacts by different ranges and then performed the PCA on the contact probability evolving trajectories. Since the contacts at long-ranges are usually formed with very low probabilities, it may introduce large uncertainty and imprecision when directly applying to PCA. To resolve this issue, we instead used the matrix P_{obs}/P_{exp} , which captures the essence of the compartment formation [15]. We also note that the boundaries of TADs are mostly conserved during the transitions (Figure S5). This indicates that the chromosomal structural arrangement during cancerization and reversion should mostly rely on the contacts at the long ranges, which can be captured by the compartment.

We projected the transitions onto the first and second most weighted PCs (Figure 5). In this respect, at the local range (< 2 Mb), we see that the ES cell is not on either of the two pathways. As the contact range increases (2-5 and 5-10 Mb), the location of the ES cell becomes close to the cancerization pathway but is far from the reversion pathway. Further increasing

the contact range to 10-15 and 15-20 Mb shows that the ES cell is close to the cancer cell. When the contact range is very long (20-40 Mb and > 40 Mb), the ES cell locates far from both of these two pathways. Our results showed that the chromosome during the cancerization process may explore the structures formed in the ES cell at the moderate contact range (2-20 Mb). In contrast, the chromosome during the reversion process seems to be structurally different from the one in the ES cell, though the chromosome is prone to adopt many of the structural properties in the ES cell, such as the contact probability scaling, structural shape geometry and spatial distribution of the chromosome loci.

Interestingly, we found that the two pathways do not overlap, resulting in the hysteresis phenomenon. Hysteresis indicates a toggle-like switching behavior and was often described in the context of ferromagnetism. In biology, hysteresis was found in the cell cycle and demonstrated as the driving force for the irreversible cell-cycle transitions [77, 78]. In this respect, the hysteresis loop area provides a quantitative measure of the degree of the irreversibility, thus it was calculated here (Figure 5H). We found that the largest hysteresis area, which corresponds to the most significant irreversibility of the chromosomal structural transitions during cancerization and reversion, is at the moderate contact range. Combining the results, we showed that the chromosomal structural transitions during the cancerization and reversion processes are highly irreversible, and the chromosome tends to form stem-like structures during the cancerization.

Figure 5: The pathways of the chromosomal structural transition during cell cancerization and reversion. The principal component analysis (PCA) of the contact matrix P_{obs}/P_{exp} , which is further divided into different ranges (A-G). The reference points from the IMR90, A549 and ES cells are plotted as brown, purple and blue points, respectively. The hysteresis loop (ω) is colored grey in (A-G), and the area is calculated for different contact ranges (H).

3 DISCUSSION AND CONCLUSIONS

We used the landscape-switching model to explore the chromosomal structural transition during the cell cancerization and reversion processes. More than two hundred independent simulations (226 simulations for cancerization and 246 simulations for reversion, see **Materials and Methods**) were performed for a long time under the potential designated for describing the chromosome dynamics in the normal (or cancer) cell state. We found that these simulations with the initial chromosome structures chosen based on the clustering method recapitulate the structural ensembles in terms of the contact maps, TADs and compartments (Figure S6). The feature suggests that the selected chromosome structures can represent the chromosome structural distributions at the original cell states and are statistically sufficient to initialize the landscape-switching simulations.

Our structural analyses showed that the chromosome with the highest degree of structural expansion during cell cancerization and reversion share significant structural similarity to the chromosome at the ES cell in many aspects, including the trend of contact probability declining with the genomic distance, the global structural geometry and the spatial distribution of chromosome loci. However, there are differences between the chromosomes with expanded structures during these two processes. As shown by the enhanced contact probability evolving trajectories (Figure 5), the chromosome during can-

cerization can form a stem-like chromosome contact pattern up to the range of 20 Mb. In contrast, this is not observed during reversion. The results suggested a potential role of the cell state with stemness during cancerization in guiding the process.

Figure 6: The Waddington landscape of the cell cancerization (solid line) and reversion (dashed line) processes from the chromosomal structural transition perspective.

Based on our simulation results, we can propose a pictorial Waddingtons landscape from the chromosomal structural perspective to understand the cancerization and reversion (Figure 6). In the context of Waddingtons epigenetic landscape [79], the cell, metaphorically referred to as a ball, rolls down from the ES cell at the top, which possesses the highest degree of

stemness, to the terminally differentiated cell at the basin of the landscape to accomplish cell differentiation. As the reverse process, cell reprogramming transforms the differentiated cell to the ES cell by gaining the stemness [80, 81]. Hi-C data revealed that the chromosome at the cancer cell has more loci in compartment A than compartment B (Figure 1), corresponding to more populated active, open euchromatin than that of inactive, closed heterochromatin. This feature suggests that the cancer cell, which exhibits the characteristics of the stem cell, should be located at a landscape layer, higher than the normal cell, which is terminally differentiated with a large proportion of the chromosomal loci in repressive compartment B (Figure 1). Transforming the normal cell to cancer cell increases the stemness. However, our simulations showed that cancerization is a non-monotonic process, during which the chromosome with the highest degree of structural expansion exhibits significant structural similarity with the one at the ES cell. In the landscape view, the cell at the first stage of cancerization climbs up on the landscape approaching the ES cell, reminiscent of cell reprogramming. The second stage of cell cancerization corresponds to a rolling down process toward the cancer cell, reminiscent of cell differentiation. On the other hand, the chromosome during the reversion process also expands and reorganizes its structure towards the one at the ES as an implication of gaining the stemness. However, the structural difference of the chromosome to the one at the ES during reversion is more significant than that during the cancerization. This leads to a relatively weak reprogramming process at the first stage of the reversion process. Thus the reversion pathway should appear to be under the cancerization pathway (Figure 6).

Our simulation showed that cancerization and reversion are irreversible processes. Previously, it was assumed that the forward and reverse processes of cell state transitions may pass through the same intermediate cell states, referring to as the metastable attractors on the Waddingtons landscape [82, 83]. Theoretical studies at the gene network level have shown that irreversibility is a universal feature for cell state transition due to the non-equilibrium effects [84, 85]. Our simulation results provided a molecular-level description that the irreversibility of cell cancerization and reversion can be reflected by the chromosome structural dynamics. Besides, we proposed a quantitative way to measure the irreversibility in cancerization and reversion using the hysteresis loop analysis of the chromosome contact formation. Our findings also resonate with a recent experiment, which used a combination of Hi-C and replication timing analyses to characterize the irreversible non-overlapped pathways between the cell state transition, namely the differentiation and reprogramming [86].

The stemness-related genetic markers and interactions in cells are controlled by the underlying gene regulatory network [87], which uses chromosome as the structural scaffold for functionalities. Besides, the CS cells were found to have similar marker expression profiles with normal stem cells [88]. These features support that our approach for inferring the stemness and connecting it to the partial formation of the CS cell is reasonable and based on the intimate structure-function relationship at the genomic level [72, 4, 5]. Therefore, we speculate that the stem-like intermediate cell state during cell cancerization may contribute to the formation of the CS cells, of

which the terminally differentiated somatic cell has been suggested as one potential origin [89]. Due to the lack of the Hi-C data at the CS cells, the precise and direct comparisons of the intermediate state to the CS cells were not performed. A strict assessment of our statement can be made when the experimental data are available.

Our predictions of the transient intermediate state with stemness during cell cancerization may have implications for cancer treatments. The increase of the stemness during cancerization may promote the cell self-renewal and differentiation processes, similar to what the CS cells do. Thus, decreasing the stemness of the cell by deviating the chromosome structures from the one at the ES cell can help to repress the cancer initiation and progression. Further efforts can be focused on developing effective genome structure engineering methods for modulating and disrupting the specific chromosomal interactions similar to those formed in the ES cells [90, 91, 92, 93, 94], serving as a potential therapeutic approach.

Our predictions can be tested by future experiments. In this regard, the time-course Hi-C experiments appear to be promising and competent in monitoring the chromosomal structural evolution during the cancerization and reversion processes [95]. It is expected that the Hi-C data at the high temporal resolution can provide a dynamic picture of how the chromosome structure rearranges and offer useful clues as to the formation of the stem-like intermediate state during cell state transition using advanced data-analysis methods [96, 97]. Meanwhile, mounting theoretical and experimental studies confirm the existence of intermediates during the epithelial-mesenchymal transition (EMT) [98, 99, 100, 101], which contributes to cancer metastasis and tumor relapse. The EMT intermediates often exhibit a gain of stemness compared with the epithelial and mesenchymal cells [102, 103, 104, 105], leading to a non-monotonic increasing-followed-by-decreasing trend in stemness during the EMT [106]. These properties of the EMT process are similar to what we have observed in our simulations of the transition from the normal cell to the cancer cell. Therefore, the fruitful approaches developed for uncovering the EMT intermediate and studying the EMT pathways can be adapted to investigate the cancerization processes [107]. Ultimately, the combined efforts on the direct cancerization and the EMT processes will help us to complete the understanding of how cancer initializes and metastasizes.

In summary, we presented a chromosomal structural-level study of illuminating the molecular mechanisms of cancer formation and its reversion processes. The targeted cells belong to the human lung organism and should be regarded as one example of cancer. Up to now, more than a hundred distinct types of cancer have been uncovered. Different types of cancer may vary substantially in their behaviors. However, all types of cancer are caused by the common mechanism of uncontrolled cell division, reflecting the dysregulation of cell self-renewal. Therefore, we speculate that our observation of the stem-like intermediate state during cancer formation may likely be observed by the other types of cancer. Recent experimental observations that the chromosome opening its structure in the early stages of tumor formation is regardless of cancer types and shares a similar trait of chromosome organization with the stem cell provide evidence of our theoretical findings and speculations our speculation [108, 109]. The open chromosome

structure in the stem-like intermediate state may contribute to the increased genomic instability and active transcription for cells to gain plasticity that has been demonstrated as the universal key to facilitate cancer progression [110]. Future studies on elucidating the cancer formation pathways and the stem-like intermediate state can help to decipher the fundamental mechanisms that underlie the different types of cancer.

Materials and Methods

Chromosome polymer model

The basic background of our polymer model is a generic polymer model with homogeneously weighted bonded and non-bonded potentials:

$$V_{\text{Polymer}} = V_{\text{Bond}} + V_{\text{Nonbond}}.$$

The bonded potential contains the finitely extensible nonlinear elastic (FENE) bond stretching interactions, linear-promoting angle bending interactions, and neighboring beads non-overlap interactions with the expression as follows [111]:

$$V_{\text{Bond}} = V_{\text{FENE}} + V_{\text{Angle}} + V_{\text{hc}},$$

where the FENE potential for the bond ($r_{i,i+1}$) between neighboring beads ($i, i+1$) is expressed as:

$$V_{\text{FENE}} = -0.5K_b R_0^2 \ln[1 - (\frac{r_{i,i+1}}{R_0})^2],$$

and the angle potential is applied to the angle (θ_i) by the three adjacent beads with the following form:

$$V_{\text{Angle}} = K_a [1 - \cos(\theta_i - \theta_0)],$$

and the non-overlap interaction between ($i, i+1$) is described by a hard-core Lennard-Jones potential (V_{LJ}):

$$V_{\text{hc}} = \begin{cases} V_{LJ}(r_{i,i+1}), & r_{i,i+1} \leq \sigma 2^{1/6}, \\ 0, & r_{i,i+1} > \sigma 2^{1/6}, \end{cases}$$

where

$$V_{LJ}(r_{i,j}) = 4\epsilon \left[\left(\frac{\sigma}{r_{i,j}} \right)^{12} - \left(\frac{\sigma}{r_{i,j}} \right)^6 \right] + \epsilon.$$

The nonbonded potential for every bead pair (i,j) uses the soft-core interactions that allow the chain-crossing to mimic the effects of topoisomerase [112].

$$V_{\text{Nonbond}} = \begin{cases} 2(1 + \tanh[0.5V_{LJ}(r_{i,j}) - 1]), & r \leq r_0, \\ V_{LJ}(r_{i,j}), & r_0 < r_{i,j} \leq \sigma 2^{1/6}, \\ 0, & r_{i,j} > \sigma 2^{1/6}, \end{cases}$$

In addition, a spherical confinement was used to mimic the volume fraction of the chromosome in the cell nucleus at 10%, same as used previously [111, 60]. With only the generic polymer potential, the model generates an ensemble that resembles the equilibrium globule [45].

Reduced units were used throughout simulations. The energy unit was $\epsilon = 1.0$. The bond length σ was set to be the length unit. In FENE potential, $R_0 = 1.5\sigma$ allows the bonds to stretch flexibly and $K_b = 30.0/\sigma^2$. The angle potential has a strength of $K_a = 2.0$ and angle of $\theta_0 = \pi$. In the soft-core potential, we set $r_0 = \sigma/((1 + \sqrt{2})/2)^{1/6}$, so $V_{LJ}(r_0) = 2.0$ [60]. The temperature in all simulations was set to 1.0 in energy units by multiplying by the Boltzmann constant. Langevin stochastic dynamics was applied with a time step of 0.001τ and a friction coefficient of $10.0\tau^{-1}$, where τ is the reduced time unit.

Maximum entropy principle simulation

The potential $V(\mathbf{r})$ in the maximum entropy principle simulations is made up of the generic polymer potential V_{Polymer} and the Hi-C restraint potential $V_{\text{Hi-C}}$:

$$V(\mathbf{r}) = V_{\text{Polymer}} + V_{\text{Hi-C}}.$$

The entropy of system under $V(\mathbf{r})$ relative to a given prior distribution $\rho_0(\mathbf{r})$ under polymer potential V_{Polymer} is defined as [59]:

$$\mathbb{S}[\rho||\rho_0] = - \int d\mathbf{r} \rho(\mathbf{r}) \ln \frac{\rho(\mathbf{r})}{\rho_0(\mathbf{r})}.$$

This entropy should be maximized subject to the following constraints in order to be compatible with observations (contact probability $P_{i,j}$ and experimental Hi-C data $f_{i,j}$):

$$\begin{cases} \int d\mathbf{r} \rho(\mathbf{r}) P_{i,j} = \langle P_{i,j} \rangle = f_{i,j}, \\ \int d\mathbf{r} \rho(\mathbf{r}) = 1. \end{cases}$$

The maximization of the entropy can be obtained using the method of Lagrangian multipliers by searching for the stationary points of the Lagrange function:

$$\mathcal{L} = \mathbb{S}[\rho||\rho_0] + \sum_{i,j} \alpha_{i,j} (d\mathbf{r} \rho(\mathbf{r}) P_{i,j} - f_{i,j}) + \beta (\int d\mathbf{r} \rho(\mathbf{r}) - 1),$$

where $\alpha_{i,j}$ and β are Lagrangian multipliers. By setting $\frac{\delta \mathcal{L}}{\delta \rho(\mathbf{r})} = 0$ and neglecting the normalization factor, the posterior distribution has the following expression:

$$\rho(\mathbf{r}) \propto e^{-\sum_{i,j} \alpha_{i,j} P_{i,j}} \cdot \rho_0(\mathbf{r}).$$

Therefore, the potential $V(\mathbf{r})$ with maximum entropy principle, is expressed as follows:

$$V(\mathbf{r}|S) = V_{\text{Polymer}} + \sum_{i,j} \alpha_{i,j} P_{i,j},$$

where $V_{\text{Hi-C}}$ is in a linear form of contact probabilities and S represents the cell state (IMR90 or A549). In practice, $P_{i,j}$ is the calculated contact probability between chromosomal i and j using step function and $\alpha_{i,j}$ is the corresponding contact strength, which is iteratively adapted by simulations. During the iteration process, the contact probability $P_{i,j}$ is restrained by the Hi-C data $f_{i,j}$. In the end, the chromosome under the potential $V(\mathbf{r}|S)$ generates an ensemble that can reproduce the Hi-C maps at the normal and cancer cell, separately (Figure S1). Details of the maximum entropy principle simulation can be found in previous studies [60, 61]. The maximum entropy principles simulations for the ES and IMR90 cells were done in our previous work [45].

Landscape-switching model

We used the resulting potentials $V(\mathbf{r}|S)$ of the maximum entropy principle simulations for the normal and cancer cells to represent the effective energy landscapes. We performed the hierarchical clustering on the chromosome ensembles generated by the maximum entropy principle simulations. Two chromosome structures in each cluster, which has a population higher than 0.3%, were picked out as the initial structures for performing the landscape-switching simulations (Figure S2). The purpose of choosing the chromosome structures in the populated clusters rather than using the whole sets of the structures in the ensembles is to generate a limited number of structures that can sufficiently represent the ensemble in order to significantly reduce the computational expenses. This led to 226 and 246 trajectories for the cancerization and reversion processes, respectively.

The landscape-switching simulations were performed as follows. First, the simulations were run under the potential $V(\mathbf{r}|IMR90)$ (cancerization) or $V(\mathbf{r}|A549)$ (reversion) for 5000τ , where τ is the time unit of the simulations. Then a sudden switch of potential in terms of $V(\mathbf{r}|IMR90) \rightarrow V(\mathbf{r}|A549)$ (cancerization) or $V(\mathbf{r}|A549) \rightarrow V(\mathbf{r}|IMR90)$ (reversion) was implemented. Finally, the simulations were under the new potential $V(\mathbf{r}|A549)$ (cancerization) or $V(\mathbf{r}|IMR90)$ (reversion) for 10000τ . All the trajectories were collected and combined for the analyses to generate the results present in this study.

Acknowledgement

We acknowledge the support from the National Science Foundation PHY-76066. The authors would like to thank Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system, which was made possible by a \$1.4M National Science Foundation grant (#1531492).

References

- [1] Dimitra Repana, Joel Nulsen, Lisa Dressler, Michele Bortolomeazzi, Santhilata Kuppili Venkata, Aikaterini Tournas, Anna Yakovleva, Tommaso Palmieri, and Francesca D Ciccarelli. The network of cancer genes (ncg): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome biology*, 20(1):1, 2019.
- [2] Carlo M Croce. Oncogenes and cancer. *New England journal of medicine*, 358(5):502–511, 2008.
- [3] Fulai Jin, Yan Li, Jesse R Dixon, Siddarth Selvaraj, Zhen Ye, Ah Young Lee, Chia-An Yen, Anthony D Schmitt, Celso A Espinoza, and Bing Ren. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, 503(7475):290–294, 2013.
- [4] Giacomo Cavalli and Tom Misteli. Functional implications of genome topology. *Nature structural & molecular biology*, 20(3):290, 2013.
- [5] Boyan Bonev and Giacomo Cavalli. Organization and function of the 3d genome. *Nature Reviews Genetics*, 17(11):661, 2016.
- [6] Batool Akhtar-Zaidi, Richard Cowper-Sal, Olivia Corradin, Alina Saiakhova, Cynthia F Bartels, Dheepa Balasubramanian, Lois Myeroff, James Lutterbaugh, Awad Jarrar, Matthew F Kalady, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*, 336(6082):736–739, 2012.
- [7] Ken J Kron, Swneke D Bailey, and Mathieu Lupien. Enhancer alterations in cancer: a source for a cell identity crisis. *Genome medicine*, 6(9):77, 2014.
- [8] Wen Fong Ooi, Manjie Xing, Chang Xu, Xiaosai Yao, Muhammad Khairul Ramlee, Mei Chee Lim, Fan Cao, Kevin Lim, Deepak Babu, Lai-Fong Poon, et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nature communications*, 7(1):1–17, 2016.
- [9] William A Flavahan, Yotam Drier, Brian B Liao, Shawn M Gillespie, Andrew S Venteicher, Anat O Stemmer-Rachamimov, Mario L Suvà, and Bradley E Bernstein. Insulator dysfunction and oncogene activation in idh mutant gliomas. *Nature*, 529(7584):110–114, 2016.
- [10] M Ryan Corces and Victor G Corces. The three-dimensional cancer genome. *Current opinion in genetics & development*, 36:1–7, 2016.
- [11] P Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman, and Michael R Stratton. A census of human cancer genes. *Nature reviews cancer*, 4(3):177–183, 2004.
- [12] Simona Negrini, Vassilis G Gorgoulis, and Thanos D Halazonetis. Genomic instability: an evolving hallmark of cancer. *Nature reviews Molecular cell biology*, 11(3):220–228, 2010.
- [13] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.
- [14] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *science*, 295(5558):1306–1311, 2002.
- [15] Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.
- [16] Suhas SP Rao, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, Ido Machol, Arina D Omer, Eric S Lander, et al. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.
- [17] Jesse R Dixon, Jie Xu, Vishnu Dileep, Ye Zhan, Fan Song, Victoria T Le, Galip Gürkan Yardımcı, Abhijit Chakraborty, Darrin V Bann, Yanli Wang, et al. Integrative detection and analysis of structural variation in cancer genomes. *Nature genetics*, 50(10):1388, 2018.
- [18] Phillippa C Taberlay, Joanna Achinger-Kawecka, Aaron TL Lun, Fabian A Buske, Kenneth Sabir, Cathryn M Gould, Elena Zotenko, Saul A Bert, Katherine A Giles, Denis C Bauer, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome research*, 26(6):719–731, 2016.
- [19] Lifei Li, Nicolai KH Barth, Christian Pilarsky, and Leila Taher. Cancer is associated with alterations in the three-dimensional organization of the genome. *Cancers*, 11(12):1886, 2019.
- [20] Kadir C Akdemir, Victoria T Le, Sahaana Chandran, Yilong Li, Roel G Verhaak, Rameen Beroukhim, Peter J Campbell, Lynda Chin, Jesse R Dixon, and P Andrew Futreal. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nature genetics*, 52(3):294–305, 2020.

- [21] Kadir C Akdemir, Victoria T Le, Justin M Kim, Sarah Killcoyne, Devin A King, Ya-Ping Lin, Yanyan Tian, Akira Inoue, Samirkumar B Amin, Frederick S Robinson, et al. Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nature Genetics*, pages 1–11, 2020.
- [22] A Rasim Barutcu, Bryan R Lajoie, Rachel P McCord, Coralee E Tye, Deli Hong, Terri L Messier, Gillian Browne, Andre J van Wijnen, Jane B Lian, Janet L Stein, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome biology*, 16(1):214, 2015.
- [23] Tannishtha Reya, Sean J Morrison, Michael F Clarke, and Irving L Weissman. Stem cells, cancer, and cancer stem cells. *nature*, 414(6859):105–111, 2001.
- [24] Sheila K Singh, Ian D Clarke, Mizuhiko Terasaki, Victoria E Bonn, Cynthia Hawkins, Jeremy Squire, and Peter B Dirks. Identification of a cancer stem cell in human brain tumors. *Cancer research*, 63(18):5821–5828, 2003.
- [25] Sheila K Singh, Cynthia Hawkins, Ian D Clarke, Jeremy A Squire, Jane Bayani, Takuichiro Hide, R Mark Henkelman, Michael D Cusimano, and Peter B Dirks. Identification of human brain tumour initiating cells. *nature*, 432(7015):396–401, 2004.
- [26] Rossella Galli, Elena Binda, Ugo Orfanelli, Barbara Cipelletti, Angela Gritti, Simona De Vitis, Roberta Fiocco, Chiara Foroni, Francesco Dimeco, and Angelo Vescovi. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer research*, 64(19):7011–7021, 2004.
- [27] Ain Zubaidah Ayob and Thamil Selvee Ramasamy. Cancer stem cells as key drivers of tumour progression. *Journal of biomedical science*, 25(1):1–18, 2018.
- [28] Jane E Visvader and Geoffrey J Lindeman. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature reviews cancer*, 8(10):755–768, 2008.
- [29] Tobias Schatton, Natasha Y Frank, and Markus H Frank. Identification and targeting of cancer stem cells. *Bioessays*, 31(10):1038–1049, 2009.
- [30] Chunhe Li and Jin Wang. Quantifying the underlying landscape and paths of cancer. *Journal of The Royal Society Interface*, 11(100):20140774, 2014.
- [31] Chunhe Li and Jin Wang. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer research*, 75(13):2607–2618, 2015.
- [32] Scott Powers and Robert E Pollack. Inducing stable reversion to achieve cancer control. *Nature Reviews Cancer*, 16(4):266, 2016.
- [33] Francesca Rapino, Eloy F Robles, Jose A Richter-Larrea, Eric M Kallin, Jose A Martinez-Climent, and Thomas Graf. C/ebp α induces highly efficient macrophage transdifferentiation of b lymphoma and leukemia cell lines and impairs their tumorigenicity. *Cell reports*, 3(4):1153–1163, 2013.
- [34] Dana Ishay-Ronen, Maren Diepenbruck, Ravi Kiran Reddy Kalathur, Nami Sugiyama, Stefanie Tiede, Robert Ivanek, Glenn Bantug, Marco Francesco Morini, Junrong Wang, Christoph Hess, et al. Gain fatlose metastasis: converting invasive breast cancer cells into adipocytes inhibits cancer metastasis. *Cancer Cell*, 35(1):17–32, 2019.
- [35] Zhuo Cheng, Zhiying He, Yongchao Cai, Cheng Zhang, Gongbo Fu, Hengyu Li, Wen Sun, Changcheng Liu, Xiu-liang Cui, Beifang Ning, et al. Conversion of hepatoma cells to hepatocyte-like cells by defined hepatocyte nuclear factors. *Cell research*, 29(2):124–135, 2019.
- [36] Kwang-Hyun Cho, Soobeom Lee, Dongsan Kim, Dongkwan Shin, Jae Il Joo, and Sang-Min Park. Cancer reversion, a renewed challenge in systems biology. *Current Opinion in Systems Biology*, 2:49–58, 2017.
- [37] Shenshen Wang and Peter G Wolynes. Communication: Effective temperature and glassy dynamics of active matter. *Journal of Chemical Physics*, 135(5):051101–051101, 2011.
- [38] S Wang and PG Wolynes. Tensegrity and motor-driven effective interactions in a model cytoskeleton. *The Journal of chemical physics*, 136(14):145102, 2012.
- [39] Michele Di Pierro, Davit A Potoyan, Peter G Wolynes, and José N Onuchic. Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proceedings of the National Academy of Sciences*, 115(30):7753–7758, 2018.
- [40] Lei Liu, Guang Shi, D Thirumalai, and Changbong Hyeon. Chain organization of human interphase chromosome determines the spatiotemporal dynamics of chromatin loci. *PLoS computational biology*, 14(12):e1006617, 2018.
- [41] Irena Bronstein, Y Israel, Eldad Kepten, Sabine Mai, Yaron Shav-Tal, E Barkai, and Yuval Garini. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Physical review letters*, 103(1):018102, 2009.
- [42] Soya Shinkai, Tadasu Nozaki, Kazuhiro Maeshima, and Yuichi Togashi. Dynamic nucleosome movement provides structural information of topological chromatin domains in living human cells. *PLoS computational biology*, 12(10):e1005136, 2016.
- [43] Ralph J DeBerardinis and Navdeep S Chandel. Fundamentals of cancer metabolism. *Science advances*, 2(5):e1600200, 2016.
- [44] Xiakun Chu and Jin Wang. Conformational state switching and pathways of chromosome dynamics in cell cycle. *Applied Physics Reviews*, 7(3):031403, 2020.
- [45] Xiakun Chu and Jin Wang. Microscopic chromosomal structural and dynamical origin of cell differentiation and reprogramming. *bioRxiv*, 2020.
- [46] R Kanwal and S Gupta. Epigenetic modifications in cancer. *Clinical genetics*, 81(4):303–311, 2012.
- [47] Leigh Ellis, Peter W Atadja, and Ricky W Johnstone. Epigenetics in cancer: targeting chromatin modifications. *Molecular cancer therapeutics*, 8(6):1409–1420, 2009.
- [48] Dongya Jia, Mohit Kumar Jolly, Prakash Kulkarni, and Herbert Levine. Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers*, 9(7):70, 2017.
- [49] D Kim, O Rath, W Kolch, and Kwang-Hyun Cho. A hidden oncogenic positive feedback loop caused by crosstalk between wnt and erk pathways. *Oncogene*, 26(31):4571–4579, 2007.
- [50] Ming Wu, Li Liu, and Christina Chan. Identification of novel targets for breast cancer by exploring gene switches on a genome scale. *BMC genomics*, 12(1):547, 2011.
- [51] Tetsuya Shiraishi, Shinako Matsuyama, and Hiroaki Kitano. Large-scale analysis of network bistability for human cancers. *PLoS Comput Biol*, 6(7):e1000851, 2010.
- [52] James E Ferrell and Eric M Machleder. The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science*, 280(5365):895–898, 1998.

- [53] Wen Xiong and James E Ferrell. A positive-feedback-based bistable memory module that governs a cell fate decision. *Nature*, 426(6965):460–465, 2003.
- [54] Sui Huang, Gabriel Eichler, Yaneer Bar-Yam, and Donald E Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, 94(12):128701, 2005.
- [55] Kun Zhang, Masaki Sasai, and Jin Wang. Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. *Proceedings of the National Academy of Sciences*, 110(37):14930–14935, 2013.
- [56] Jin Wang. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Advances in Physics*, 64(1):1–137, 2015.
- [57] John C Tully. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics*, 93(2):1061–1071, 1990.
- [58] Martin Krzywinski, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra. Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9):1639–1645, 2009.
- [59] Andrea Cesari, Sabine Reißer, and Giovanni Bussi. Using the maximum entropy principle to combine simulations and solution experiments. *Computation*, 6(1):15, 2018.
- [60] Bin Zhang and Peter G Wolynes. Topology, structures, and energy landscapes of human chromosomes. *Proceedings of the National Academy of Sciences*, 112(19):6062–6067, 2015.
- [61] Bin Zhang and Peter G Wolynes. Shape transitions and chiral symmetry breaking in the energy landscape of the mitotic chromosome. *Physical review letters*, 116(24):248101, 2016.
- [62] R David Hawkins, Gary C Hon, Leonard K Lee, QueMinh Ngo, Ryan Lister, Mattia Pelizzola, Lee E Edsall, Samantha Kuan, Ying Luu, Sarit Klugman, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell*, 6(5):479–491, 2010.
- [63] Wei Xie, Matthew D Schultz, Ryan Lister, Zhonggang Hou, Nisha Rajagopal, Pradipta Ray, John W Whitaker, Shulan Tian, R David Hawkins, Danny Leung, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*, 153(5):1134–1148, 2013.
- [64] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539):331, 2015.
- [65] Michael F Clarke and Andrew T Hass. Cancer stem cells. *Reviews in Cell Biology and Molecular Medicine*, 2006.
- [66] Long V Nguyen, Robert Vanner, Peter Dirks, and Connie J Eaves. Cancer stem cells: an evolving concept. *Nature Reviews Cancer*, 12(2):133–143, 2012.
- [67] Leonid A Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19(1):37–51, 2011.
- [68] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of chromosomal domains by loop extrusion. *Cell reports*, 15(9):2038–2049, 2016.
- [69] Richard Losick and Claude Desplan. Stochasticity and cell fate. *science*, 320(5872):65–68, 2008.
- [70] Ruxandra I Dima and D Thirumalai. Asymmetry in the shapes of folded and denatured states of proteins. *The Journal of Physical Chemistry B*, 108(21):6564–6570, 2004.
- [71] Samuel S Cho, Yaakov Levy, and Peter G Wolynes. P versus q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proceedings of the National Academy of Sciences*, 103(3):586–591, 2006.
- [72] Tom Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800, 2007.
- [73] Wendy A Bickmore and Bas van Steensel. Genome architecture: domain organization of interphase chromosomes. *Cell*, 152(6):1270–1284, 2013.
- [74] Michele Di Pierro, Bin Zhang, Erez Lieberman Aiden, Peter G Wolynes, and José N Onuchic. Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016.
- [75] Yifeng Qi and Bin Zhang. Predicting three-dimensional genome organization with chromatin states. *PLoS computational biology*, 15(6):e1007024, 2019.
- [76] Alistair N Boettiger, Bogdan Bintu, Jeffrey R Moffitt, Siyuan Wang, Brian J Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Chao-ting Wu, and Xiaowei Zhuang. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418, 2016.
- [77] Wei Sha, Jonathan Moore, Katherine Chen, Antonio D Las-saletta, Chung-Seon Yi, John J Tyson, and Jill C Sible. Hysteresis drives cell-cycle transitions in xenopus laevis egg extracts. *Proceedings of the National Academy of Sciences*, 100(3):975–980, 2003.
- [78] Joseph R Pomeroy, Eduardo D Sontag, and James E Ferrell. Building a cell cycle oscillator: hysteresis and bistability in the activation of cdc2. *Nature cell biology*, 5(4):346–351, 2003.
- [79] Conrad Hal Waddington. *The strategy of the genes*. Routledge, 1957.
- [80] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126(4):663–676, 2006.
- [81] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872, 2007.
- [82] Kazutoshi Takahashi. Cellular reprogramming—lowering gravity on waddington’s epigenetic landscape. *Journal of cell science*, 125(11):2553–2560, 2012.
- [83] Joseph G Crompton, David Clever, Raul Vizcardo, Mahendra Rao, and Nicholas P Restifo. Reprogramming antitumor immunity. *Trends in immunology*, 35(4):178–185, 2014.
- [84] Jin Wang, Kun Zhang, Li Xu, and Erkang Wang. Quantifying the waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, 108(20):8257–8262, 2011.
- [85] Xiaona Fang and Jin Wang. Nonequilibrium thermodynamics in cell biology: Extending equilibrium formalism to cover living systems. *Annual review of biophysics*, 49:227–246, 2020.
- [86] Hisashi Miura, Saori Takahashi, Rawin Poonperm, Akie Tanigawa, Shin-ichiro Takebayashi, and Ichiro Hiratani. Single-cell dna replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. *Nature genetics*, 51(9):1356–1368, 2019.

- [87] Thomas Klonisch, Emilia Wiechec, Sabine Hombach-Klonisch, Sudharsana R Ande, Sebastian Wesselborg, Klaus Schulze-Osthoff, and Marek Los. Cancer stem cell markers in common cancers—therapeutic implications. *Trends in molecular medicine*, 14(10):450–460, 2008.
- [88] Mark Shackleton. Normal stem cells and cancer stem cells: similar and different. In *Seminars in cancer biology*, volume 20, pages 85–92. Elsevier, 2010.
- [89] Rolf Bjerkvig, Berit B Tysnes, Karen S Aboody, Joseph Nabjauer, and AJA Terzis. The origin of the cancer stem cell: current controversies and new insights. *Nature Reviews Cancer*, 5(11):899–904, 2005.
- [90] Richard Kostriken and Cathy J Wedeen. Engineered interphase chromosome loops guide intrachromosomal recombination. *The EMBO journal*, 20(11):2907–2913, 2001.
- [91] Wulan Deng, Jongjoo Lee, Hongxin Wang, Jeff Miller, Andreas Reik, Philip D Gregory, Ann Dean, and Gerd A Blobel. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, 149(6):1233–1244, 2012.
- [92] Stefanie L Morgan, Natasha C Mariano, Abel Bermudez, Nicole L Arruda, Fangting Wu, Yunhai Luo, Gautam Shankar, Lin Jia, Huiling Chen, Ji-Fan Hu, et al. Manipulation of nuclear architecture through crispr-mediated chromosomal looping. *Nature communications*, 8(1):1–9, 2017.
- [93] Ji Hun Kim, Mayuri Rege, Jacqueline Valeri, Margaret C Dunagin, Aryeh Metzger, Katelyn R Titus, Thomas G Gilgenast, Wanfeng Gong, Jonathan A Beagan, Arjun Raj, et al. Ladd: light-activated dynamic looping for endogenous gene expression control. *Nature methods*, 16(7):633, 2019.
- [94] Manyu Du, Fan Zou, Yujie Yan, and Lu Bai. Chemically induced chromosomal interaction (cici): A new tool to study chromosome dynamics and its biological roles. *bioRxiv*, 2020.
- [95] Job Dekker, Andrew S Belmont, Mitchell Guttman, Victor O Leshyk, John T Lis, Stavros Lomvardas, Leonid A Mirny, Clodagh C O’Shea, Peter J Park, Bing Ren, et al. The 4d nucleome project. *Nature*, 549(7671):219–226, 2017.
- [96] Kristin Abramo, Anne-Laure Valton, Sergey V Venev, Hakan Ozadam, A Nicole Fox, and Job Dekker. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase. *NATURE CELL BIOLOGY*, 21(11):1393–1402, 2019.
- [97] Ning Qing Liu and Elzo de Wit. A transient absence of smc-mediated loops after mitosis. *Nature cell biology*, 21(11):1303–1304, 2019.
- [98] Ievgenia Pastushenko, Audrey Brisebarre, Alejandro Sifrim, Marco Fioramonti, Tatiana Revenco, Soufiane Boumahdi, Alexandra Van Keymeulen, Daniel Brown, Virginie Moers, Sophie Lemaire, et al. Identification of the tumour transition states occurring during emt. *Nature*, 556(7702):463–468, 2018.
- [99] M Angela Nieto, Ruby Yun-Ju Huang, Rebecca A Jackson, and Jean Paul Thiery. Emt: 2016. *Cell*, 166(1):21–45, 2016.
- [100] Mingyang Lu, Mohit Kumar Jolly, Herbert Levine, José N Onuchic, and Eshel Ben-Jacob. MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proceedings of the National Academy of Sciences*, 110(45):18144–18149, 2013.
- [101] Federico Bocci, Mohit Kumar Jolly, Jason Thomas George, Herbert Levine, and José Nelson Onuchic. A mechanism-based computational model to capture the interconnections among epithelial–mesenchymal transition, cancer stem cells and notch-jagged signaling. *Oncotarget*, 9(52):29906, 2018.
- [102] Robert Strauss, Zong-Yi Li, Ying Liu, Ines Beyer, Jonas Persson, Pavel Sova, Thomas Möller, Sari Pesonen, Akseli Hemminki, Petra Hamerlik, et al. Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PloS one*, 6(1):e16186, 2011.
- [103] Mohit Kumar Jolly, Bin Huang, Mingyang Lu, Sendurai A Mani, Herbert Levine, and Eshel Ben-Jacob. Towards elucidating the connection between epithelial–mesenchymal transitions and stemness. *Journal of the Royal Society Interface*, 11(101):20140962, 2014.
- [104] Shubham Tripathi, Mohit Kumar Jolly, Wendy A Woodward, Herbert Levine, and Michael W Deem. Analysis of hierarchical organization in gene expression networks reveals underlying principles of collective tumor cell dissemination and metastatic aggressiveness of inflammatory breast cancer. *Frontiers in oncology*, 8:244, 2018.
- [105] Marcus Ruscetti, Bill Quach, Eman L Dadashian, David J Mulholland, and Hong Wu. Tracking and functional characterization of epithelial–mesenchymal transition and mesenchymal tumor cells during prostate cancer metastasis. *Cancer research*, 75(13):2749–2759, 2015.
- [106] Luigi Ombrato and Ilaria Malanchi. The emt universe: space between cancer cell dissemination and metastasis initiation. *Critical Reviews in Oncogenesis*, 19(5), 2014.
- [107] Ievgenia Pastushenko and Cédric Blanpain. Emt transition states during tumor progression and metastasis. *Trends in cell biology*, 29(3):212–226, 2019.
- [108] Maria Aurelia Ricci, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160(6):1145–1158, 2015.
- [109] Jianquan Xu, Hongqiang Ma, Hongbin Ma, Wei Jiang, Christopher A Mela, Meihan Duan, Shimei Zhao, Chenxi Gao, Eun-Ryeong Hahm, Santana M Lardo, et al. Super-resolution imaging reveals the evolution of higher-order chromatin folding in early carcinogenesis. *Nature communications*, 11(1):1–17, 2020.
- [110] Andrew P Feinberg, Michael A Koldobskiy, and Anita Göndör. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews Genetics*, 17(5):284–299, 2016.
- [111] Angelo Rosa and Ralf Everaers. Structure and dynamics of interphase chromosomes. *PLoS computational biology*, 4(8):e1000153, 2008.
- [112] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R Lajoie, Leonid A Mirny, and Job Dekker. Organization of the mitotic chromosome. *Science*, 342(6161):948–953, 2013.