

Insights into the molecular mechanisms of cell fate decision making processes from 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

May 9, 2021

Abstract

Cell state transitions or cell fate decision making processes, such as cell development and cell pathological transformation, are believed to be determined by the regulatory network of genes, which intimately depend on the structures of chromosomes in the cell nucleus. The high temporal resolution picture of how chromosome reorganizes its 3D structure during the cell state transitions is the key to understanding the mechanisms of these fundamental cellular processes. However, this picture is still challenging to acquire at present. Here, we studied the chromosome structural dynamics during the cell state transitions among the pluripotent embryonic stem cell (ESC), the terminally differentiated normal cell and the cancer cell using landscape-switching model implemented in the molecular dynamics simulation. We considered up to 6 transitions, including differentiation, reprogramming, cancer formation and reversion. We found that the pathways can merge at certain stages during the transitions for the two processes having the same destination as the ESC or the normal cell. Before reaching the merging point, the two pathways are cell-type-specific. The chromosomes at the merging points show high structural similarity to the ones at the final cell states in terms of the contact maps, TADs and compartments. The post-merging processes correspond to the adaption of the chromosome global shape geometry through the chromosome compaction without significantly disrupting the contact formation. On the other hand, our detailed analysis showed no merging point for the two cancer formation processes initialized from the ESC and the normal cell, implying that cancer progression is a complex process and may be associated with multiple pathways. Our results draw a complete molecular picture of cell development and cancer at the dynamical chromosome structural level, and help our understanding of the molecular mechanisms of cell fate decision making processes.

opment of a tumor. The transition processes between distinct cell phenotypic states are believed to be governed by the underlying gene regulatory networks [1, 2]. As the scaffold for the genome function, the chromosome folds into a 3D organization to underpin the gene expressions, which are cell-type-specific. During the transition, the cell undergoes phenotypic switching determined by the complex gene network, which is supported by the large-scale chromosome structural rearrangement in favor of the specific gene expression pattern towards the destined cell state [3]. It has been well recognized that the 3D chromosome structure plays important roles in regulating cell development [4] and cancer [5]. However, our understanding of the abovementioned cell fate decision making processes at the molecular level through chromosome structural dynamics is restricted by the limited experimental approaches, which are often confronted with difficulties in covering the large spatiotemporal scales of the chromosome systems in the cell nucleus.

Currently, the chromosome structural determination is largely performed by the Hi-C techniques and the results are generally presented in terms of an ensemble-averaged contact map [6, 7]. Hi-C experiments measure the frequency of contact formed by all chromosomal loci pairs across the genome, so a high (low) contact frequency formed by two chromosomal loci corresponds to a close (far away) spatial distance and implies a strong (weak) physical interaction between these two loci. Careful analyses on the Hi-C contact map further revealed that the chromosome folds into a hierarchical organization from the multi-megabase compartments [6, 7] to the sub-megabase topologically associating domains (TADs) [8, 9]. These two chromosome structural features have been recognized to have significant impacts on gene expressions [10, 11, 12, 13]. Our knowledge of the chromosome 3D structure and its roles in gene regulation has been significantly increased by the Hi-C techniques. However, Hi-C data only provides the static description of the chromosome organization within one cell state, the dynamical changes of the chromosome structure during the transition between the two cell states are still inaccessible.

Here, we applied the landscape-switching model to study the chromosome structural dynamics during the cell state transitions among the embryonic stem cell (ESC), the normal and cancer cell (See "Materials and Methods"). Overall, there are 6 transition processes, including the ESC differentiation

1 Introduction

Cell state transition or cell fate decision making, is one of the fundamental cell events in forming multicellular organisms. The transformation of the normal cell to the cancer cell, which is a typical cell fate transition in pathology, leads to the devel-

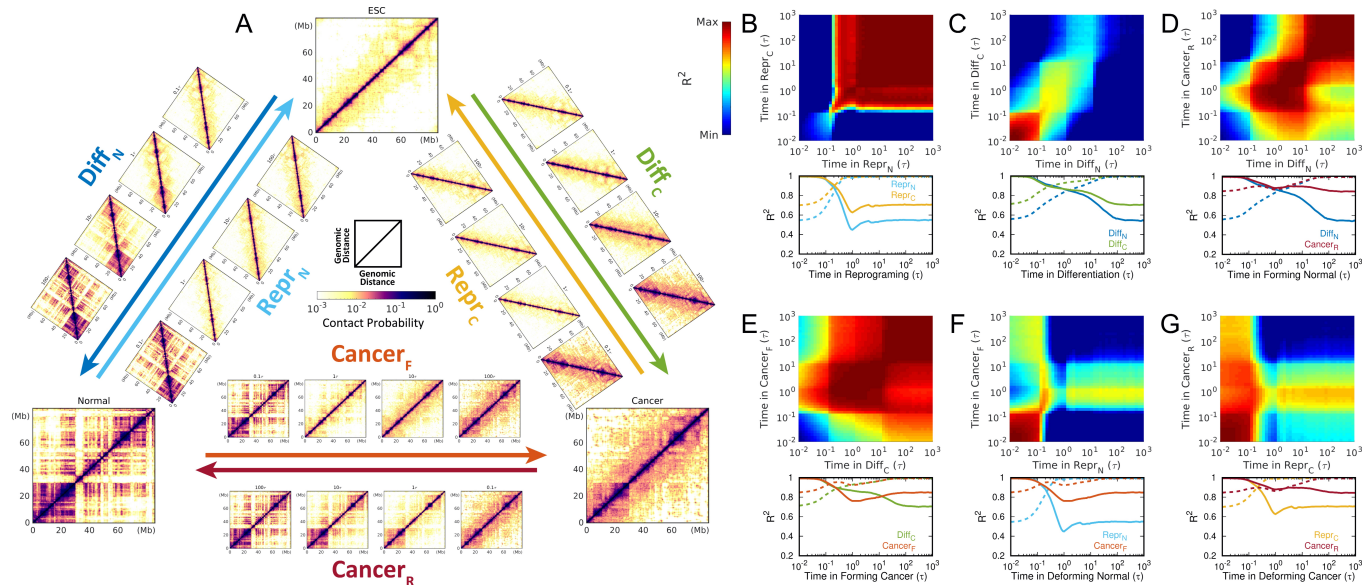


Figure 1: The chromosome structural rearrangements during cell fate decision making among the ESC, the normal and cancer cell. (A) The 6 cell state transition processes associated with chromosome structural dynamics. The chromosome Hi-C contact maps of the ESC, the normal and cancer cell, as well as the cell states at the typical time points (0.1τ , 1τ , 10τ , 100τ) during each transition are shown. The correlation maps between the time-course Hi-C contact maps of two transitions during (B) reprogramming from the normal (Repr_N) and cancer cell (Repr_C) to the ESC processes, (C) the ESC differentiation to the normal (Diff_N) and cancer cell (Diff_C) processes, (D) forming the normal cell from the ESC (differentiation, Diff_N) and the cancer cell (cancer reversion, Cancer_R) processes, (E) forming the cancer cell from the ESC (differentiation, Diff_C) and the cancer cell (cancer formation, Cancer_F) processes, (F) deforming the normal cell to the ESC (Diff_N) and the cancer cell (Cancer_F) processes, (G) deforming the cancer cell to the ESC (Diff_C) and the normal cell (Cancer_R) processes. The comparison between process "A" at time $t^A = I$ and process "B" at time $t^B = J$ is made by calculating the coefficient of determination $R^2(I, J)$ between the contact probability P_{ij} formed by chromosomal loci "i" and "j": $R^2(I, J) = 1 - \frac{\sum_{i,j} [P_{ij}(t^A = I) - \langle P_{ij}(t^A = I) \rangle]^2}{\sum_{i,j} [P_{ij}(t^A = I) - \langle P_{ij}(t^A = I) \rangle]^2 + \sum_{i,j} [P_{ij}(t^B = J) - \langle P_{ij}(t^B = J) \rangle]^2}$. $R^2(I, J)$, which measures the similarity of the two Hi-C contact map at time $t^A = I$ of process A and $t^B = J$ of process B ($R^2(I, J) = 1$ corresponds to the identical Hi-C contact maps with $P_{ij}(t^A = I) = P_{ij}(t^B = J)$ and the deviation of $R^2(I, J)$ from 1 indicates the degree of difference between these two contact maps.), is shown by 2D plots colored from blue (minimum) to red (maximum) at the upper panels in (A-G). At the lower panels in (A-G), the cell states "J" are selected to be either the initial states (solid lines) or the end states (dashed lines).

to the normal and cancer cell, the normal and cancer cell reprogramming to the ESC, and the cancer cell formation from the normal cell and reversion to the normal cell. The system presents a complete circuitry of cell fate decision making processes, including the normal and pathological developments of cells. The experimental Hi-C contact maps have revealed significant differences in chromosome structures among the ESC, the normal and cancer cell [8, 14]. This feature implies that there are large-scale chromosome structural rearrangements during the cell state transition among these 3 cell states. In our simulations, we constantly monitored the chromosome structures evolving along the time during the cell state transitions and quantified the transition pathways in terms of Hi-C contact map, chromosome structural shape geometry and the spatial distribution of loci on the chromosome.

The quantified pathways of chromosome structural dynamics draw a molecular-level picture of how the cell state transitions or cell fate decision making processes occur. We observed early bifurcation events in two of the cell state transition processes when the initial cell states are the same, such as the ESC differentiation and the processes of deforming the normal and cancer cells. For the two processes with the same destination, the pathways can merge at certain stages, after which the chromosome has to undergo further structural com-

paction to reach the final state. Detailed structural analyses revealed that the two transition processes for forming cancer cell never overlap, though they can share a similar chromosome structural compaction route. This implies the complex and multiple-route characteristics of cancer formation. Our results suggested that the cell transition processes with the same cell state as the destination for the ESC and normal cell, are made up of an initial cell-type-specific stage followed by a common-route stage. Our study provides a comprehensive understanding of the cell fate decision making processes through chromosome structural dynamics.

2 Results

2.1 Dynamical chromosome structural rearrangements during differentiation, reprogramming and cancer formation

We calculated the simulated Hi-C contact map evolving in time. We observed that the chromosome dynamically rearranges its structure during the cell state transitions in terms of gradually adapting the Hi-C contact map from the initial cell state towards the final cell state. Interestingly, it appears that the evolutions of the contact maps do not follow

the same routes for the forward and reverse transitions between any two cells. This indicates that both the differentiation and reprogramming processes, as well as the cancer formation and reversion processes, are irreversible from the chromosome structural perspective. The irreversibility at the chromosome structural-level can further influence the irreversible gene regulation pathways that determine the non-equilibrium cell state decision-making processes [15, 16].

The differentiation processes starting from the ESC should bifurcate during the transitions towards the normal and the cancer cell, while the normal and the cancer cell reprogramming processes should merge during the transitions towards the ESC. To gain insights into how the bifurcation and merging of the pathways occur from the chromosome structural dynamics, we calculated the correlation coefficient of determination R^2 between the two contact probability matrices in time (Figure 1B and 1C). For reprogramming, it appears that the pathways of the normal and cancer cell reprogramming to the ESC (Repr_N and Repr_C) start to merge quite early at $\sim 0.1\tau$ with high values of R^2 (Figure 1B, *upper panel*). At $\sim 0.1\tau$, the chromosomes have very different contact formations from either the ESC and final state (Figure 1B, *lower panel*). Interestingly, we observed over-expanded chromosome structures with respect to the one at the ESC formed during the normal and the cancer cell reprogramming processes. Besides, we see that the pathways of reprogramming start to merge before forming the over-expanding chromosome conformation. For differentiation, the two transitions towards the normal and cancer cell (Diff_N and Diff_C) bifurcate at $\sim 2\tau$ (Figure 1C, *upper panel*), where the chromosome has different structures from that at the ESC and is also deviated from those at the final cell states (Figure 1C, *lower panel*).

Since the destinations of differentiation Diff_N and cancer reversion to the normal cell (Cancer_R) are the same, it implies that these two pathways should merge at certain stages of these two transitions. From Figure 1D, we see that the chromosomes during Diff_N and Cancer_R transitions become structurally similar at $\sim 1\tau$, where the chromosome has accomplished significant structural rearrangements for forming the normal cell (Figure 1D, *lower panel*). However, a considerable number of the chromosome contact changes after 1τ is still seen during Diff_N to make the chromosome gradually deviate from the ESC. A similar observation can be found in forming the cancer cell by differentiation Diff_C and cancer formation process (Cancer_F) (Figure 1E). The contact maps of the two pathways become similar at $\sim 1\tau$, where the chromosome with different structures from that at the ESC still has to undergo a certain degree of structural arrangements in order to form the structure at the cancer cell.

In addition to the ESC differentiation, we also studied the bifurcation of the processes with the initial cell states being the normal cell (Repr_N and Cancer_F, Figure 1F) and cancer cell (Repr_C and Cancer_R, Figure 1G), respectively. The two pathways in both of these processes exhibit bifurcation at the early stages of the transitions around 0.2τ , which is later than the bifurcation time observed at the ESC differentiation ($\sim 2\tau$). It is worth noting that during Cancer_F and Cancer_R, the chromosome at $0.1-1\tau$ shares a certain degree of structural similarity with that after 1τ during the reprogramming towards the ESC. This feature has led to an implication that the cell during the

transitions between the normal and cancer states increases the degree of the "stemness" from the chromosome structural perspective [17].

2.2 Chromosome structural shape changes during differentiation, reprogramming and cancer formation

Using molecular dynamics simulations, we can have the 3D chromosome structural information in addition to the 2D contact map. Here, we introduced three reaction coordinates to describe the geometry of the chromosome structural shape. The radius of gyration R_g is used to describe the chromosome compaction and the aspheric shape parameter Δ is used to measure the extent of anisotropy with deviation from 0 as an indication of deviation from a perfect sphere [18]. The lengths of chromosome structure along the longest principal axis (PA1) and shortest principal axis (PA3) are used to describe the global geometric shape of the chromosome structure.

We observed the over-expanded chromosome structures with respect to the one at the ESC during both the normal and cancer cells reprogramming to the ESC processes (Figure 2A). During both of the reprogramming processes, the chromosome expanding stages are isotropic and reach the over-expanded phase with a similar shape at similar time points (1τ). Then, the chromosome undergoes slight compaction to form the chromosome shape at the ESC through the merging pathways with negligible contact probability adaption (Figure 1B). From the chromosome structural perspective, these two reprogramming processes use different routes to form the same over-expanded chromosome structures, implying that the reprogramming pathways are cell-type-specific at the early stage but conserved for different cell types at the late stage.

For cell differentiation, the chromosome structural pathways with shape changes share the same route towards the normal and cancer cells till $\sim 10\tau$ (Figure 2B), when a large degree of contact maps in the destined cell states are established (Figure 1C). From the chromosome contact formation perspective, we observed that the bifurcations of the two differentiation processes start at an earlier time of 1τ than that based on the chromosome structural shape-changing pathways (Figure 1C). These features indicate that the geometric chromosome compaction during the ESC differentiation to the normal and cancer cell share the same route for a long duration at the first stage of differentiation, even when the chromosome contact maps within these two transitions have deviated from each other already.

From the chromosome structural shape-changing pathways, we can see that the two processes of forming the normal cell by the ESC differentiation and cancer reversion merge at $\sim 10\tau$ (Figure 2C), where the chromosome has a more open structure than that at both the normal and cancer cell. Thus, the merging pathways correspond to chromosome compaction processes associated with forming the chromosome contact maps at the normal cell. It appears that the chromosome contact maps become similar at the very early stages of Diff_N and Cancer_R around 1τ (Figure 1D), so the chromosomes within these two processes can form similar structural shapes. These findings imply that the chromosome contact changes during $1-10\tau$ con-

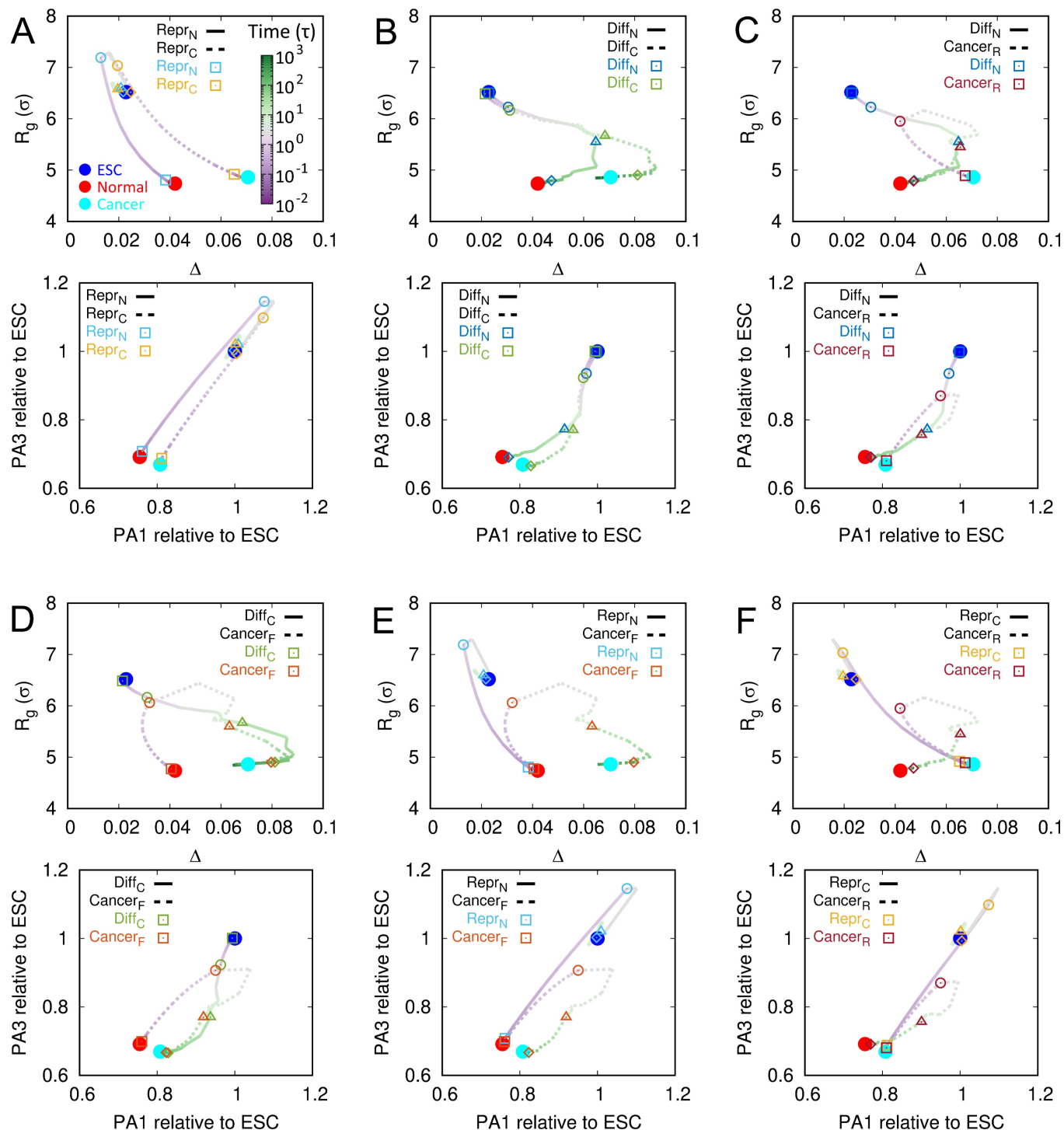


Figure 2: Quantified averaged pathways for chromosome structural shape changes during cell fate decision making among the ESC, the normal and cancer cell. The pathways are shown for (A) differentiation, (B) reprogramming, (C) forming the normal cell, (D) forming the cancer cell, (E) deforming the normal cell and (F) deforming the cancer cell. The reaction coordinates are the radius of gyration (R_g) and aspheric shape parameter (Δ) at the upper panels in (A-E), and the length of structural geometric extensions at the longest and shortest principal axes (PA1 and PA3) at the lower panels in (A-E). In each panel, the typical time points in the trajectories are additionally shown as squares (\square , $t=0.1\tau$), circles (\circ , $t=1\tau$), triangles (\triangle , $t=10\tau$) and diamonds (\diamond , $t=100\tau$). Time is in the logarithmic scale.

tribute to the increase of the similarity of the chromosome structural shapes during these two processes. A similar observation can be observed for forming the cancer cell by Diff_C and Cancer_F . Overall, we found that the transitions between the normal and cancer cells are made up of an initial chromosome structure expansion followed by compaction. The late

stages share similar routes with the chromosome geometric compaction pathways during the ESC differentiation.

For the cell state transitions with the same initial states at the normal and cancer cell, the pathways projected onto the chromosome shape parameters bifurcate at the very beginning around 0.1τ . In order words, although the chromosome has to

undergo an initial structural expansion during the transitions between the normal and cancer cells, the pathways deviate from the ones during reprogramming. The early bifurcation in the chromosome shape-changing pathways is coincident with what was observed from the contact probability perspective (Figure 1F and 1G). This indicates that the arrangements of the contact maps at the early stage of deforming the normal and cancer cell contribute significantly to the specific adaptations of the chromosome structural shape towards the destined cells.

2.3 Spatial rearrangement of chromosomal loci during differentiation, reprogramming and cancer formation

The spatial distribution of the loci in chromosome has significant impacts on the transcriptional activity [19, 20]. We calculated the changes of the radial density of the chromosome during the 6 transitions ($\rho(r, t)$) and further classified them into the descriptions of the whole loci, loci in compartment A and compartment B (Figure S3-S5). Then we collected $\rho(r, t)$ for all the processes and performed the principal component analysis (PCA) on all sets of $\rho(r, t)$. We projected the chromosome structural dynamics during differentiation, reprogramming and cancer onto the first and second principal components (PC1 and PC2) and obtained the trajectories in terms of dynamically arranging the spatial distribution of the chromosome loci (Figure 3).

Overall, we see that the rearrangements of the spatial distribution of the whole loci share similar trends of the loci in compartment B for all the transitions. During the normal cell reprogramming process, the changes of $\rho(r)$ of the whole loci and loci in compartment B appear to go through that at the cancer cell state (Figure 3A), while the $\rho(r)$ of the loci in compartment A can only have overlap with that during cancer cell reprogramming starting at the over-expanding stages. This implies that the distinct reprogramming routes from different cells at the early stage strongly depend on the spatial arrangement of the loci in compartment A. For the ESC differentiation, we see that the bifurcations of the two pathways to the normal and cancer cells start early for both $\rho(r)$ of the loci in compartment A and compartment B (Figure 3B). This indicates a unique evolution route to the destined terminally differentiated cell for the ESC differentiation.

For forming the normal cell, the pathways of $\rho(r)$ for differentiation and cancer reversion processes merge (Figure 3C). For the whole loci and loci in compartment B, the merging appears when there is still a long way to form the final $\rho(r)$ at the normal cell. Interestingly, we found that the merging point for $\rho(r)$ of loci in compartment A is located near to the position of $\rho(r)$ at the ESC. Similar trends can be observed in forming the cancer cell from the ESC and the normal cell (Figure 3D). However, there is a slight discrepancy between the $\rho(r)$ pathways of loci in compartment B for Diff_C and Cancer_F, implying that these two processes for forming the cancer cell may not use the same routine to the spatial repositioning of the chromosomal loci.

We also observed the bifurcation of the pathways from the $\rho(r)$ between Repr_N and Cancer_F (Figure 3E), and between

Repr_C and Cancer_R (Figure 3F), similar to that from the chromosome structural shape perspective. However, the elements leading to the bifurcation of the two pathways in deforming the normal cell and the two pathways in deforming the cancer cell are different. For Repr_N and Cancer_F, we see an overlap between these two pathways at the very beginning with $\rho(r)$ of the whole loci and loci in compartment B. Thus, the evolution of $\rho(r)$ of the loci in compartment A contributes to the differences of these two pathways. In contrast, for Repr_C and Cancer_R, $\rho(r)$ of the loci in compartment A shares the same route towards forming the $\rho(r)$ at the ESC, so the bifurcation of these two pathways is led by the changing $\rho(r)$ of the loci in compartment B.

2.4 Chromosome structures in transient states during differentiation, reprogramming and cancer formation

Based on the above pathways analysis, we can see that the transitions among the differentiation, reprogramming and cancer can merge at certain points. Identifying these points can provide a full picture of how these processes occur from the chromosome structural perspective. To determine these merging points, we first chose a series of time points located on the pathways projected on the chromosome structural shape parameters and PCs of ρ , when the two pathways (Repr_N and Repr_C, Diff_N and Cancer_R, Diff_C and Cancer_F) start to merge. Then we calculated the R^2 of the contact maps between the two time points on these two pathways and selected the highest R^2 as the time when the pathways emerge (Figure S6). There are 6 transient states determined as the merging points for reprogramming to the ESC (Repr_N^S and Repr_C^S), forming the normal cell (Diff_N^S and Cancer_R^S) and forming the cancer cell (Diff_C^S and Cancer_F^S). In Figure 4, we can see one transient state is overlapped with the other for determining the merging point. However, it is worth noting that Diff_C^S and Cancer_F^S have an apparent separation on pathways of $\rho(r)$ of loci in compartment B. It implies that the chromosomes at Diff_C^S and Cancer_F^S should be structurally different.

The direct comparison of the contact maps between the two transient states at the merging points indicates high similarities of chromosome structural formation with R^2 all equal to 0.99 (Figure 5(A-F)). To examine the similarity of the chromosome structures in terms of the TADs and compartments, we further calculated the signals of the insulation score and compartment profile. We found that the formations of TAD boundaries of the chromosome at the two transient states for one merging point are very similar and close to the ones at the destined cell states (Figure 5G and Figure S8). It is worth noting that chromosomes at the merging point for forming normal cell (Diff_N^S and Cancer_R^S) and for forming cancer cell (Diff_C^S and Cancer_F^S) have similar structural shape, but the signals of the TAD boundaries are not strongly correlated. During our simulations, we did not detect significant changes in TAD boundaries (Figure S9), resonating with the experimental evidence that TAD features are mostly conserved in mammals across different cell types and species [8, 21]. For compartments, we found strong correlations of the signals between transient states at the merging point of reprogramming (Repr_N and Repr_C) and forming nor-

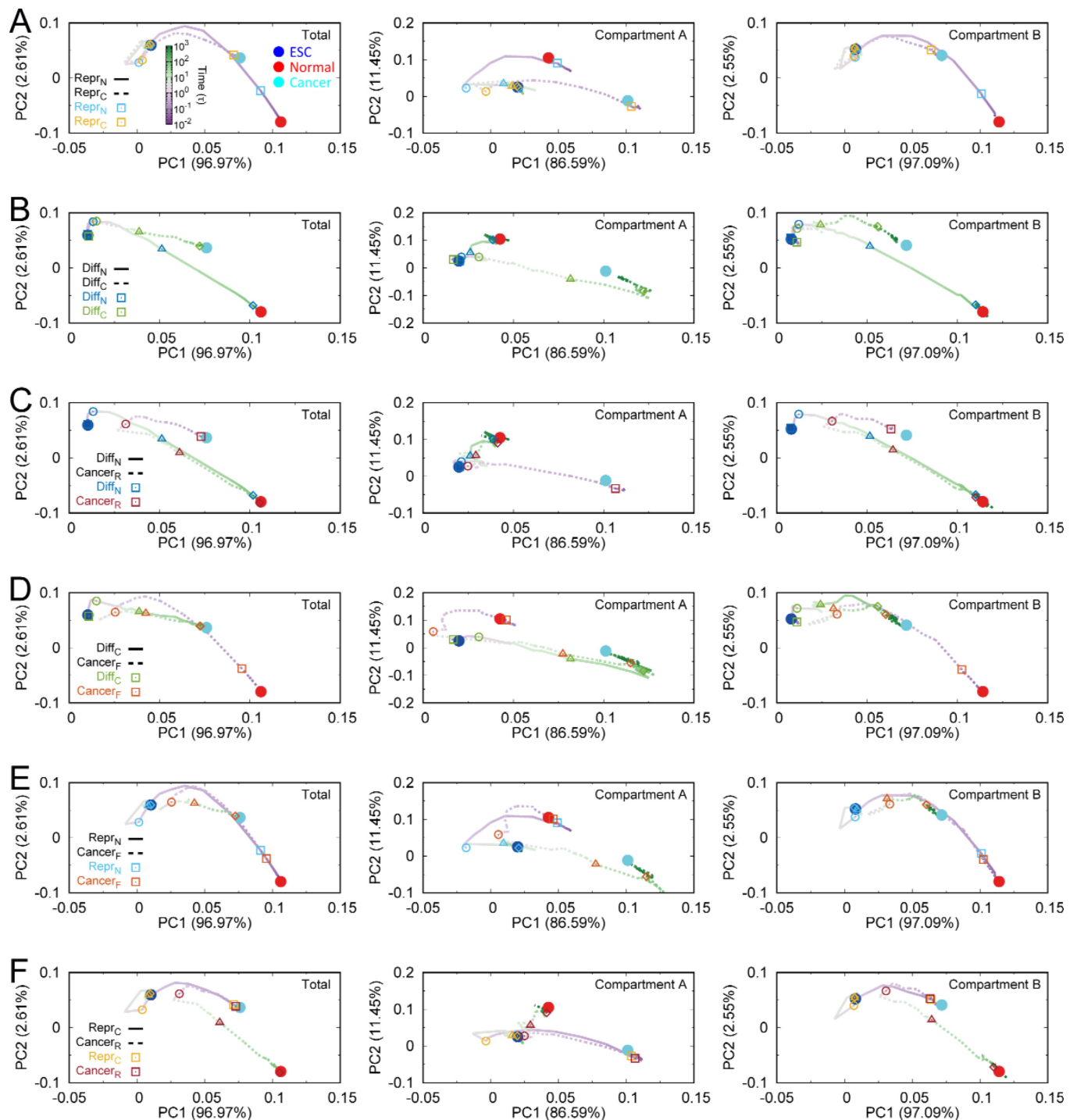


Figure 3: Quantified averaged pathways for the spatial distribution of chromosomal loci changing during cell fate decision making among the ESC, the normal and cancer cell. The pathways are shown for (A) differentiation, (B) reprogramming, (C) forming the normal cell, (D) forming the cancer cell, (E) deforming the normal cell and (F) deforming the cancer cell. The trajectories are projected on the the first and second principal components (PC1 and PC2) of the spatial distribution of chromosomal loci changes for the whole loci (left panel), loci in compartment A (center panel) and loci in compartment B (right panel). In each panel, typical time points in the trajectories are additionally shown as squares (\square , $t=0.1\tau$), circles (\circ , $t=1\tau$), triangles (\triangle , $t=10\tau$) and diamonds (\diamond , $t=100\tau$). Time is in the logarithmic scale.

mal cell (Diff_N^S and Cancer_R^S), while the compartment profiles are quite different at Diff_C^S and Cancer_F^S for the merging point in the cancer formation process (Figure 5H and Figure S10). Thus, we can state that the chromosomes at Diff_C^S and Cancer_F^S are structurally different, leading to the different compartment formation. The pathways from the ESC and normal cell in forming the cancer cell are not merged based on the chromo-

some structural formation at the compartment level.

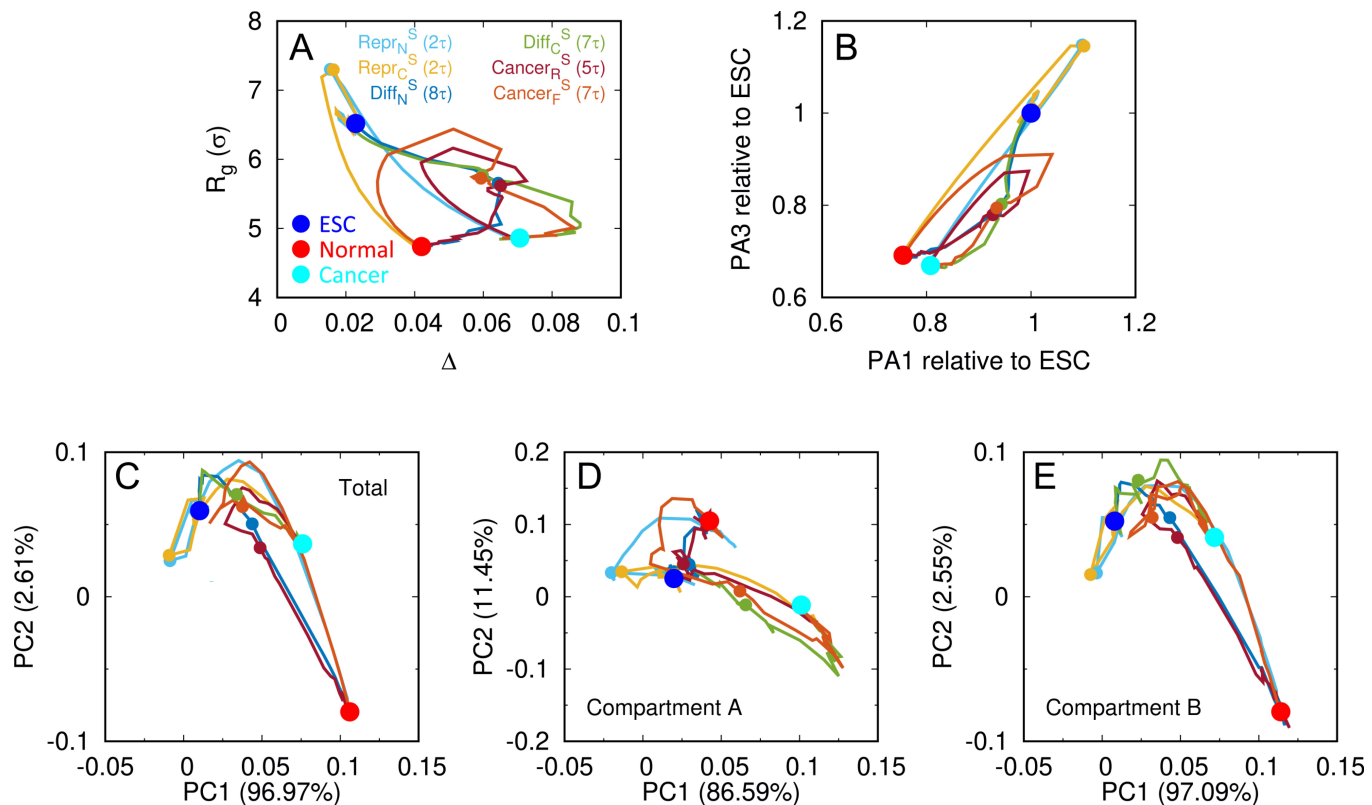


Figure 4: All the averaged trajectories of the 6 processes for differentiation, reprogramming and cancer involved with the transient states projected onto (A) R_g and Δ , (B) extension lengths at PA1 and PA3, and PC1 and PC2 of spatial distribution of chromosome loci for (C) the whole loci, (D) loci in compartment A and (E) loci in compartment B.

3 Discussion and conclusions

In this work, we studied the chromosome structural dynamics during the transitions between the two different cell states in cell fate decision making among the ESC, the normal and cancer cell. At the chromosomal structural-level, we showed that the forward and reverse cell state transitions between two cell states, such as the differentiation and reprogramming, the cancer formation and reversion, use different routes, leading to irreversible processes. More importantly, we found that the chromosome structural dynamical pathways for the transition towards the same cell can merge prior to the destination. The chromosomes at the merging points are structurally different from those at the corresponding final cell state. Our findings indicate that further chromosome structural arrangements, which are necessary for accomplishing the transition, undergo the same pathway for different processes after passing the merging point.

Based on the chromosome structural dynamical pathways from our analyses, we can draw a full picture of the cell state transitions among the ESC, the normal and cancer cell with more molecular-level details than the one that only contains 6 direct transitions between two cell states at the gene expression level (Figure 6). For reprogramming, the normal and cancer cells are initially converted to the merging point (Repr_N^S and Repr_C^S), followed by the further transition to the ESC using the same pathway. At the merging point, the chromosomes have a high degree of similarity of the contact probability map with that at the ESC, and slight over-expanded structures compared to the one at the ESC. In other words, the normal and cancer

cells have to undergo a long journey of cell-type-specific reprogramming before reaching the merging point, where most of the chromosome interactions at the ESC have been established with an open chromosome structure. We note that the experimental identifications on the transcriptional dynamics trajectories of different somatic cells reprogramming to the iPSC are strongly dependent on the cell types [22]. This appears to be consistent with the first stages of reprogramming processes to the merging point observed in our simulations. A recent integrative transcriptional and epigenomic analysis on the human reprogramming process with a high temporal resolution unprecedentedly uncovered that the gene network established at the late reprogramming process has a signature of pre-implantation stages [23]. During pre-implantation stages, the chromosome undergoes the structural reorganization in establishing the transition from the totipotency to pluripotency state [24]. The nuclear volumes in many mammalian embryos have been found to gradually decrease prior to the blastocyst stages during early embryonic development, leading to a universal chromosome structural compaction process during the pre-implantation stages [25, 26]. Our theoretical predictions on the late reprogramming stage after passing the merging point at the Repr_N^S and Repr_C^S with the over-expanded chromosome conformation are reminiscent of the pre-implantation stages of the reprogramming process [23]. We also found that the cell states, which originate from the normal and cancer cell at the merging point, and the following pathways towards the ESC are conserved regardless of the cell types. This echoes the conserved pre-implantation stages, where the cell has not

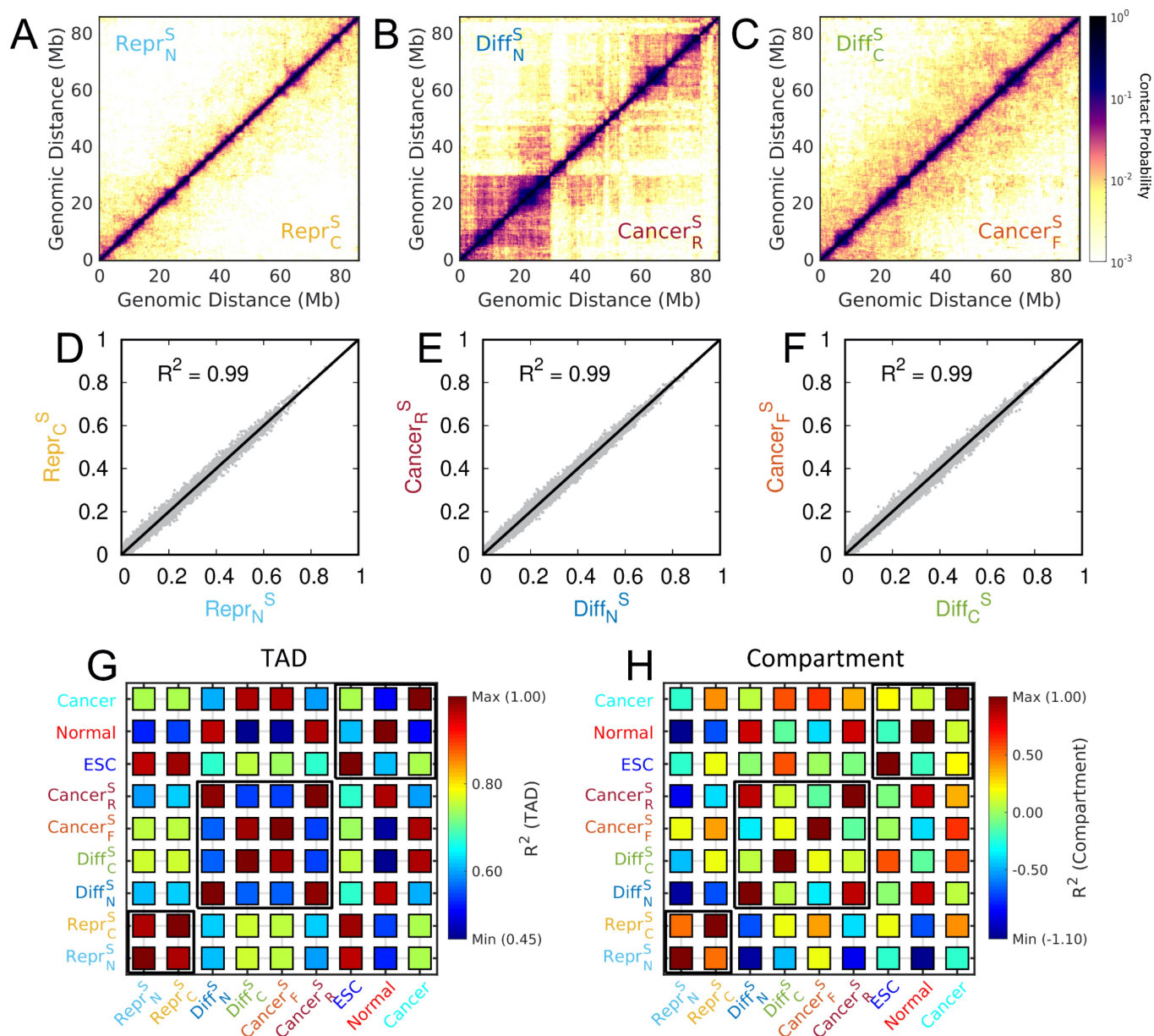


Figure 5: Correlations of chromosome structures between the transient states during cell fate decision making among the ESC, the normal and cancer cell. The contact probability maps for the transient states formed during (A) reprogramming, (B) forming the normal cell and (C) forming the cancer cell. The coefficient of determination R^2 s are shown under the corresponding Hi-C contact maps in (D-F). Chromosome structural similarities among the transient states, ESC, normal and cancer cells in terms of (G) TAD and (H) compartment. For TAD, R^2 is calculated based on the insulation score. For compartment, R^2 is calculated based on the compartment profile. The calculation procedures of TAD insulation score and compartment profile can be found in "Materials and Methods".

started the cell-specific differentiation.

We also found that the chromosome structural dynamics pathways of the differentiation and cancer reversion processes towards the normal cell are conserved at the late stages of cell state transitions. The chromosomes during these two transitions at the merging point of the pathways (Diff_N^S and Cancer_R^S) show high structural similarity. In addition, the TAD boundaries and compartments in chromosomes at the Diff_N^S and Cancer_R^S are similar to those at the normal cell. Further compaction of the chromosome from the cell state at the merging point towards the normal cell requires the spatial repositioning of the chromosome loci, while the local chromosome structures and compartment status of the chromosomal loci remain

largely unaltered. As the TAD structures and compartment segregations in the chromosome contribute significantly to the gene expressions that determine the cell fate [13, 27, 28], we speculate that the chromosomes at the merging point of the differentiation and cancer reversion transition have accomplished most of the structural arrangements for the functional purposes at the final cell state. In other words, the pathways for forming the normal cell merge at the very late stages of the cell state transition processes.

Interestingly, the two processes of forming the cancer cell from the ESC and normal cell do not share the chromosome structural dynamics pathways, in contrast with either the reprogramming processes or the processes of forming the nor-

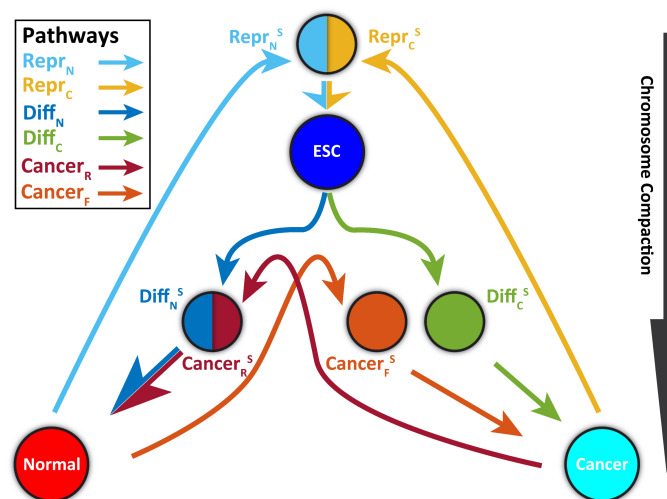


Figure 6: Scheme illustrating the 6 cell state transition processes among the ESC, the normal and cancer cell involved with the transient states from the chromosome structural dynamics perspective. The vertical arrow from the top to the bottom indicates the degree of chromosome compaction.

mal cell. Despite showing similar degrees of global chromosome compaction and the overall chromosome contact map patterns, the chromosomes at the Diff_C^S and Cancer_F^S with similar TAD structures exhibit significant differences at the compartment formation. Further changing the compartment status with spatial repositioning of the chromosome loci from Diff_C^S and Cancer_F^S towards the cancer cell is indispensable and undergoes different pathways for the differentiation and cancer formation processes. Therefore, our results show multiple routes for forming the cancer cell from the chromosome structural perspective. The results here focused on the chromosome structural dynamics at the molecular level resonate with our recent theoretical work using the gene regulatory network, which revealed the global mechanisms of cancer with multiple states and paths, underlining the complexity of cancer at the gene level [29].

Our theoretical predictions can be assessed by future biophysical experiments. In this respect, the recently developed time-course single-cell Hi-C experiments can be implemented to measure the chromosome structural rearrangements for the differentiation, reprogramming and cancer processes [30, 31]. The additional Hi-C data on the intermediate states in the processes can also be incorporated into the landscape-switching model, leading to stepwise switching transitions. It is expected that the multistep landscape-switching model can produce more realistic results as more experimental information is incorporated. In summary, the picture drawn here provides a comprehensive and physical understanding of the cell state transition processes among the pluripotent cell, the terminally differentiated cell and the cancer cell at the chromosomal, molecular, and structural levels.

4 Materials and Methods

4.1 Hi-C data

The Hi-C data of the ESC and the normal cell (IMR90, human fetal lung cell) were downloaded from the Gene Expression Omnibus (GEO) repository archives with accession number GSE35156 [8]. The Hi-C data of the cancer cell (A549, human lung carcinoma cell) was obtained from the ENCODE project with GEO accession number GSE105600 [14]. All the replicas in each dataset of Hi-C data were used and proceeded to the Hi-C Pro software following the standard pipeline [32]. All the Hi-C contact maps were built at 100kb resolution after iterative correction and eigenvector decomposition (ICE) normalization. Here, we focused on the long arm of chromosome 14 (20.5-101.Mb).

4.2 Polymer model

The polymer simulation is based on a "beads-on-a-string" model. Each bead, representing a 100kb chromosome segment, is connected by pseudo bonds to the neighboring beads. Thus, the system has a total of 857 beads. We used the finitely extensible nonlinear elastic (FENE) potential to describe the bond stretching [33]. The additional repulsive potential was added between the neighboring beads to avoid the spatial overlap [34]. The angle potential for the consecutive three beads and long-range soft-core potential for non-bonded beads were implemented as suggested previously [35]. Spherical confinement was added to mimic the volume fraction of the chromosome inside the cell nucleus at 10% [36]. The bonded and non-bonded potentials together lead to a typical homopolymer potential and finally generates an equilibrium globule chromosome ensemble, which has no TAD and compartment formation [37]. Details of the polymer model can be found in our previous work [38].

4.3 Maximum entropy principle simulation

Hi-C data was additionally incorporated into the polymer potential as the experimental restraint through the maximum entropy principle [39]. The maximum entropy principle generates a chromosome ensemble that can reproduce the experimental Hi-C data with a minimal bias by maximizing the entropy of the system. We performed the maximum entropy principle simulations for the ESC, normal and cancer cell, respectively. The simulations lead to three data-driven potentials, under which the chromosomes form three heterogeneous structural ensembles, which exhibit the experimentally consistent Hi-C contact maps for the ESC, normal and cancer cell, respectively [37, 17]. The potential was further surveyed to be capable of generating the chromosome dynamics that is in good agreement with multiple experiments in many aspects [40, 41], including anomalous diffusion, viscoelasticity, and spatially coherent dynamics [42, 43]. These features suggest that the potential can be regarded as an effective equilibrium landscape that governs both the chromosome structure and dynamics in one cell state.

4.4 Landscape-switching model

To explore the chromosome dynamics during the cell state transitions, we used the landscape-switching model, which was developed in our previous work [38, 37, 17]. To connect the landscapes of two cells among the ESC, normal and cancer cell, we performed two sets of molecular dynamics simulations of chromosomes under the potentials of two distinct cell states, separately. After a long duration of simulations when the systems become steady, we switched these two potentials (landscapes), which occurs abruptly to recapitulate the bistable switch between the two gene expression states of two cell states in cell development [44, 45, 46]. From a physical perspective, this implementation resembles an instantaneous energy excitation that drives the system out-of-equilibrium. Then, the chromosomes relaxed on the post-switching potential landscapes and the relaxation processes were referred to as the chromosome dynamics during the transitions between these two cell states. The landscape-switching model has led to non-equilibrium non-adiabatic dynamics, thus it can be used to capture the essence of cell state transitions for the inter-landscape crossing events (transition between the steady stable cell states) [47, 48]. In this study, the landscape-switching model leads to 6 transitions indicated in Figure 1A, including the ESC differentiation to the normal cell and cancer cell, the reprogramming from the normal cell and cancer cell, the cancer formation from the normal cell and the cancer reversion to the normal cell. The trajectories of the relaxation processes in the landscape-switching model were extracted and analyzed for describing the chromosome structural dynamics during the cell state transition processes. Details of maximum entropy principle simulation and landscape-switching model can be found in our previous work [38, 37, 17].

4.5 Identification of TADs and compartments

The TAD structures were described by the insulation score, which was previously used to determine the boundaries of the TADs [49]. The sliding space (5×5 beads) was used to calculate the insulation score from the contact map at each cell state. The compartment profiles were calculated by contact probability $P_{\text{obs}}/P_{\text{exp}}$, which is the ratio between the observed contact probability P_{obs} and expected contact probability P_{exp} at the 1Mb resolution. Then the principal component analysis (PCA) was performed on the enhanced contact probability map after the ICE normalization. Finally, the first principle component (PC1) was used to represent the compartment profile. The compartment status of the normal cell was determined by the direction of the PC1 values, which was assigned with the gene density (positive to gene-rich and compartment A, negative to gene-poor and compartment B). The direction of the PC1 values of other cell states was determined by calculating the correlation coefficient to the compartment profile (PC1 with the correct direction) of the normal cell.

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] 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.
- [2] Sui Huang, Ingemar Ernberg, and Stuart Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. In *Seminars in cell & developmental biology*, volume 20, pages 869–876. Elsevier, 2009.
- [3] Jiang Zhu, Mazhar Adli, James Y Zou, Griet Verstappen, Michael Coyne, Xiaolan Zhang, Timothy Durham, Mohammad Miri, Vikram Deshpande, Philip L De Jager, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, 152(3):642–654, 2013.
- [4] Hui Zheng and Wei Xie. The role of 3d genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology*, 20(9):535–550, 2019.
- [5] M Ryan Corces and Victor G Corces. The three-dimensional cancer genome. *Current opinion in genetics & development*, 36:1–7, 2016.
- [6] 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.
- [7] Suhas SP Rao, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sarnborn, 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.
- [8] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376, 2012.
- [9] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, et al. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381, 2012.
- [10] David L Spector. The dynamics of chromosome organization and gene regulation. *Annual review of biochemistry*, 72(1):573–608, 2003.
- [11] Robert Schneider and Rudolf Grosschedl. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes & development*, 21(23):3027–3043, 2007.
- [12] Andrea Smallwood and Bing Ren. Genome organization and long-range regulation of gene expression by enhancers. *Current opinion in cell biology*, 25(3):387–394, 2013.
- [13] David U Gorkin, Danny Leung, and Bing Ren. The 3d genome in transcriptional regulation and pluripotency. *Cell stem cell*, 14(6):762–775, 2014.
- [14] ENCODE Project Consortium et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74, 2012.

- [15] 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.
- [16] Xiaona Fang, Karsten Kruse, Ting Lu, and Jin Wang. Nonequilibrium physics in biology. *Reviews of Modern Physics*, 91(4):045004, 2019.
- [17] Xiakun Chu and Jin Wang. Deciphering the molecular mechanism of cancer formation by chromosome structural dynamics. *bioRxiv*, pages 2021–2, 2021.
- [18] 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.
- [19] Tom Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800, 2007.
- [20] Wendy A Bickmore and Bas van Steensel. Genome architecture: domain organization of interphase chromosomes. *Cell*, 152(6):1270–1284, 2013.
- [21] Matteo Vietri Rudan, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T Odom, Amos Tanay, and Suzana Hadjur. Comparative hi-c reveals that ctcf underlies evolution of chromosomal domain architecture. *Cell reports*, 10(8):1297–1309, 2015.
- [22] Christian M Nefzger, Fernando J Rossello, Joseph Chen, Xiaodong Liu, Anja S Knaupp, Jaber Firas, Jacob M Paynter, Jahnvi Pflueger, Sam Buckberry, Sue Mei Lim, et al. Cell type of origin dictates the route to pluripotency. *Cell reports*, 21(10):2649–2660, 2017.
- [23] Davide Cacchiarelli, Cole Trapnell, Michael J Ziller, Magali Soumilion, Marcella Cesana, Rahul Karnik, Julie Donaghey, Zachary D Smith, Suthera Ratanasirintrao, Xiaolan Zhang, et al. Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell*, 162(2):412–424, 2015.
- [24] Aline V Probst and Geneviève Almouzni. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends in Genetics*, 27(5):177–185, 2011.
- [25] Tiphaine Aguirre-Lavin, Pierre Adenot, Amélie Bonnet-Garnier, Gaëtan Lehmann, Renaud Fleurot, Claire Boulesteix, Pascale Debey, and Nathalie Beaujean. 3d-fish analysis of embryonic nuclei in mouse highlights several abrupt changes of nuclear organization during preimplantation development. *BMC developmental biology*, 12(1):1–20, 2012.
- [26] Jens Popken, Volker J Schmid, Axel Strauss, Tuna Guengoer, Eckhard Wolf, and Valeri Zakhartchenko. Stage-dependent remodeling of the nuclear envelope and lamina during rabbit early embryonic development. *Journal of Reproduction and Development*, 2015.
- [27] Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and François Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome research*, 24(3):390–400, 2014.
- [28] Boyan Bonev and Giacomo Cavalli. Organization and function of the 3d genome. *Nature Reviews Genetics*, 17(11):661, 2016.
- [29] Chong Yu, Qiong Liu, Cong Chen, and Jin Wang. Quantification of the underlying mechanisms and relationships among cancer, metastasis, and differentiation and development. *Frontiers in Genetics*, 10:1388, 2020.
- [30] Vishnu Dileep, Korey A Wilson, Claire Marchal, Xiaowen Lyu, Peiyao A Zhao, Ben Li, Axel Poulet, Daniel A Bartlett, Juan Carlos Rivera-Mulia, Zhaohui S Qin, et al. Rapid irreversible transcriptional reprogramming in human stem cells accompanied by discordance between replication timing and chromatin compartment. *Stem cell reports*, 13(1):193–206, 2019.
- [31] Samuel Collombet, Noémie Ranisavljevic, Takashi Nagano, Csilla Varnai, Tarak Shisode, Wing Leung, Tristan Piolot, Rafael Galupa, Maud Borensztein, Nicolas Servant, et al. Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature*, 580(7801):142–146, 2020.
- [32] Nicolas Servant, Nelle Varoquaux, Bryan R Lajoie, Eric Viara, Chong-Jian Chen, Jean-Philippe Vert, Edith Heard, Job Dekker, and Emmanuel Barillot. Hic-pro: an optimized and flexible pipeline for hi-c data processing. *Genome biology*, 16(1):259, 2015.
- [33] Harold R Warner Jr. Kinetic theory and rheology of dilute suspensions of finitely extendible dumbbells. *Industrial & Engineering Chemistry Fundamentals*, 11(3):379–387, 1972.
- [34] Naoko Tokuda, Tomoki P Terada, and Masaki Sasai. Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophysical journal*, 102(2):296–304, 2012.
- [35] Angelo Rosa and Ralf Everaers. Structure and dynamics of interphase chromosomes. *PLoS computational biology*, 4(8):e1000153, 2008.
- [36] 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.
- [37] Xiakun Chu and Jin Wang. Microscopic chromosomal structural and dynamical origin of cell differentiation and reprogramming. *Advanced Science*, 7(20):2001572, 2020.
- [38] Xiakun Chu and Jin Wang. Conformational state switching and pathways of chromosome dynamics in cell cycle. *Applied Physics Reviews*, 7(3):031403, 2020.
- [39] Andrea Cesari, Sabine Reißer, and Giovanni Bussi. Using the maximum entropy principle to combine simulations and solution experiments. *Computation*, 6(1):15, 2018.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.

- 777 [46] Hannah H Chang, Martin Hemberg, Mauricio Barahona, Don-
778 ald E Ingber, and Sui Huang. Transcriptome-wide noise con-
779 trols lineage choice in mammalian progenitor cells. *Nature*,
780 453(7194):544–547, 2008.
- 781 [47] Kun Zhang, Masaki Sasai, and Jin Wang. Eddy current and cou-
782 pled landscapes for nonadiabatic and nonequilibrium complex
783 system dynamics. *Proceedings of the National Academy of Sci-*
784 *ences*, 110(37):14930–14935, 2013.
- 785 [48] Jin Wang. Landscape and flux theory of non-equilibrium dy-
786 namical systems with application to biology. *Advances in*
787 *Physics*, 64(1):1–137, 2015.
- 788 [49] Emily Crane, Qian Bian, Rachel Patton McCord, Bryan R La-
789 joie, Bayly S Wheeler, Edward J Ralston, Satoru Uzawa, Job
790 Dekker, and Barbara J Meyer. Condensin-driven remodelling of
791 x chromosome topology during dosage compensation. *Nature*,
792 523(7559):240, 2015.