Unraveling Supra-Nucleosomal Physical Interactions that Govern Global Transcription Activity

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Applied Physics

by

Wenli Wu

EVANSTON, ILLINOIS

March 2019
Abstract

Unraveling Supra-Nucleosomal Physical Interactions that Govern Global Transcription Activity

Wenli Wu

The role of nuclei nanoenvironment in cellular function has been a challenging problem in biology due to the lack of chromatin 3-dimension (3D) structure imaging/capturing techniques and theory connecting physical structure of chromatin to transcription. Recent studies on optical properties measurement on biological sample techniques, nanoscale imaging techniques and chromatin 3D structure techniques unraveled the important role of chromatin structure in regulating cellular functions, especially transcription. Leveraging Molecular Dynamics (MD) simulation, Brownian Dynamics (BD) simulation, system biology and power law scaling nature of chromatin, we developed the computational and theoretical model describing the mechanisms of modulating global patterns in the gene expression through the regulation of chromatin nanoenvironment, e.g. the scaling of chromatin packing (CP) and macromolecular crowding (MC) model. The CP-MC model describes the gene transcription as a function of the physical nanoenvironment, showing a good match with experiment results from RNA sequencing techniques, suggesting the existence of a pathway-independent global chromatin scaling “code” that governs the stochastic variations in gene expression by acting directly on physical factors such as gene accessibility, binding affinities, and transcription factor diffusion. Herein, we studied how various physical factors, including average nuclear density, the scaling of chromatin packing, gene length, and the upper length-scale of chromatin packing influence gene expression. Utilizing the CP-MC model, we
identified a major functional role of the chromatin physical nanoenvironment as a regulator of cellular transcriptional responsiveness, which indicates the potential of a new field of engineering macrogenomic information space. As suggested by the idea of macrogenomic engineering, we developed the procedure of identifying compounds that target the packing of chromatin to enhance chemotherapeutic efficacy – Chromatin Protection Therapies (CPTs). Furthermore, we identify, both computationally and experimentally, the chromatin nanoenvironment as a key regulator of phenotypic plasticity by controlling both intercellular transcriptional heterogeneity and transcriptional malleability. We found that a higher scaling of chromatin packing produces a “tailwind effect” that amplifies the rate of gene transcription in response to external stimuli, possibly augmenting the ability of cells to adapt to cytotoxic stressors. Finally, analyzing transcriptional data from patients with advanced lung, colorectal and breast cancer, we identify an inverse relationship between patient survival and phenotypic plasticity.
Acknowledgements

I deeply appreciate Drs. Vadim Backman and Igal Szleifer for their mentorship during my PhD. It was an honor to be their student and work with them to develop the model and theory in this work. Their intelligence and passion for science have encouraged me to overcome the many challenges I faced during my PhD. Moreover, their support gave me the opportunity to explore different fields and discover what I truly want to do with my future. I would also like to thank Dr. Olvera de la Cruz and Dr. Schatz for being members of my thesis committee, providing crucial advice on my research, especially when I proposed my thesis direction.

Next, I would like to thank my colleagues from both Vadim Backman’s lab and Igal Szleiger’s lab. I would like to thank, in no particular order: Dr. Luay Almassalha, Adam Eshein, Dr. Andrew Radosevich, Dr. The-QuyenNguen, Graham Spicer, Dr. Greta Bauer, David VanDerway, Yue Li, Rongrong Liu, Scott Gladstein, Aya Eid, James Winkelmann, Ranya Virk, Jane Frederick, Dr. John Chandler, Dr. Lusik Cherkezyan, Dr. Di Zhang, Maria Proenca, Xiang Zhou, Andrew Stawarz, Dr. Greg Putzel, Dr. Kai Huang, Dr. Hiroaki Matsuda, Dr. Estefania Gonzalez Solveyra, Anne Shim, Dr. Sung Hyun Park, Dr. Rikkert Nap and Dr. German Picasso. In addition, I would like to thank the outstanding collaborators Dr. Xin Wang and Dr. Zhe Ji.

Last but not least, I am grateful for the opportunity to thank my family and friends who have supported me throughout my PhD. From the bottom of my heart, I could not have completed this work without their continuous understanding and encouragement. I hope this work, that I have been devoted to the past few years, makes them proud.
# Table of Contents

Abstract.................................................................................................................................................. 3

Acknowledgements .................................................................................................................................. 5

List of Tables .......................................................................................................................................... 9

List of Figures ........................................................................................................................................ 10

Chapter 1: Introduction ......................................................................................................................... 13

1.1 Aims and Scope.................................................................................................................................. 13
1.2 Optical properties of cell nuclei indicate the changes of chromatin structure for cells with diseases........................................................................................................................................... 14
1.3 Crowding nanoenviroment regulates transcription and chromatin packing density .............. 15
1.4 Chromatin packing capturing techniques assists the understanding between chromatin nanostructure and the behavior of organisms......................................................... 18
1.5 Global transcription pattern is regulated by chromatin packing .............................................. 19
1.6 Regulates cellular plasticity through engineering global transcription pattern................... 19

Chapter 2: Chromatin structure to optical properties ........................................................................ 21

2.1 Introduction ..................................................................................................................................... 21
2.2 Application of the Born approximation to calculate optical properties ................................ 23
2.3 From TEM Images to Optical Properties ...................................................................................... 27
2.3.1.1 Acquire nuclei TEM images ................................................................................................. 27
2.3.1.2 2D Refractive index autocorrelation function of nuclei TEM image............................. 28
2.3.1.3 Extract 3D autocorrelation function from 2D TEM images ........................................ 30
2.3.2 Calculate optical properties from the autocorrelation function ......................................... 32
2.4 Results ............................................................................................................................................. 33
2.5 Discussion and conclusion ............................................................................................................. 38
Chapter 3: The Crowding Environment in Nucleus Influences the General Transcription Reaction and Chromatin Structure ................................................................. 43

3.1 Introduction .................................................................................................. 43
3.2 The Crowding Environment Influences the General Transcription Reaction .......... 45
3.3 The Crowding Environment Influence on Chromatin Structure .......................... 51

Chapter 4: Chromatin packing capturing techniques ................................................. 58

4.1 Introduction ................................................................................................. 58
4.2 Multi-technique chromatin nanoimaging and computational transcriptomics platform .... 61
4.3 Chromatin is a mass fractal and the fractal dimension is a transcription regulator ......... 63
4.4 Regulating nanoclusters present with similar chromatin packing scaling on single cell level ........................................................................................................ 65
4.5 Characterizing dexamethasone-induced chromatin packing alterations via multi-technique platform in A549 cells ................................................................. 68
4.6 Characterizing dexamethasone-induced time-sensitive chromatin packing alterations via multi-technique platform in BJ cells ...................................................... 72
4.7 Summary ..................................................................................................... 75
4.8 Materials and Methods .................................................................................. 77

Chapter 5: Unraveling the chromatin packing scaling code: physical interactions govern global transcription ................................................................................................. 87

5.1 Introduction ................................................................................................. 87
5.2 Results .......................................................................................................... 90
5.3 Summary ..................................................................................................... 104
5.4 Material and Method .................................................................................... 121

Chapter 6: The role of the chromatin nanoenvironment in regulating phenotypic plasticity ...................................................................................................... 136

6.1 Introduction ................................................................................................. 136
6.2 Chromatin Packing-Macromolecular Crowding (CP-MC) Model ................................ 140
6.3 The scaling of chromatin packing regulates phenotypic plasticity .................................. 145
6.4 Summary ......................................................................................................................... 152

Chapter 7: Discussion, Conclusion and Future Directions ................................................. 189

7.1 Summary ......................................................................................................................... 189
7.2 Future directions and Emerging Questions ...................................................................... 191
    7.2.1 Develop multiscale modeling method to improve prediction by CP-MC model........ 191
    7.2.2 Regulation of chromatin packing scaling ................................................................. 192
    7.2.3 Discover more CPTs with prediction from CP-MC ................................................. 192

Reference ............................................................................................................................ 194
List of Tables

Chapter 6: The role of the chromatin nanoenvironment in regulating phenotypic plasticity

Table 1. Descriptions and values of parameters used in predicting by CP-MC model.

Table 2. Information about patient characteristics.
List of Figures

Chapter 1: Introduction

Fig 1.1 MSA model predicting the relation between nuclei crowding density and transcription rate

Chapter 2: Chromatin structure to optical properties

Fig 2.1 Nuclei in a TEM image.

Fig 2.2 Extracting B6(r:) from 2D images.

Fig 2.3 Results for the autocorrelation function, spectrum of g and the phase function for control sample and AA sample.

Fig 2.4 Results of the optical property calculated from colon cell nuclei TEM images, compared with ISOCT result.

Fig 2.5 The comparison of the phase functions from TEM images (The solid lines) and the HG relation (The dash lines). Both phase functions are calculated at $\lambda = 700$nm

Chapter 3: The Crowding Environment in Nucleus Influences the General Transcription Reaction and Chromatin Structure

Fig 3.1 The comparison of crowding induced PMF in the MAS

Fig 3.2 The 2D contact probability map for RL model (the intensity is under log$_{10}$ scale).

Fig 3.3 Contact probability for RL model and SAW model
Chapter 4: Chromatin packing capturing techniques

Fig 4.1 Examples of the multi-technique nanoimaging and computational transcriptomics platform.

Fig 4.2 Chromatin packing exhibits mass fractal structures in multiple length scales.

Fig 4.3 High resolution imaging platform reveals chromatin nanoclusters on single cell level.

Fig 4.4 Measuring chromatin packing scaling alterations induced by dexamethasone in A549 cells.

Fig 4.5 Measuring chromatin packing scaling alterations induced by dexamethasone in BJ cells.

Chapter 5: Unraveling the chromatin packing scaling code: physical interactions govern global transcription

Fig 5.1 Genomic networks are highly interconnected and decentralized.

Fig 5.2 Genomic interactions depend on a complex physical nanoenvironment.

Fig 5.3 Control of higher-order chromatin packing density scaling allows manipulation of genomic information space.

Fig 5.4 Chemotherapeutic stress increases variations in chromatin packing density.

Fig 5.5 Chromatin protective agents rapidly decrease the spatial variations in chromatin packing density.

Fig 5.6 Regulation of chromatin packing-density scaling modulates transcriptional heterogeneity.

Fig 5.7 Rapid modulation of chromatin packing density scaling by CPT agents greatly enhances chemotherapeutic efficacy.

Chapter 6: The role of the chromatin nanoenvironment in regulating phenotypic plasticity
Fig 6.1 Molecular and physical regulators of transcription.

Fig 6.2 Comparison of the CP-MC model with experimental measurements of gene expression as a function of physical regulators $D_i, L, M_f$ and $\phi_{in}$.

Fig 6.3 CP-MC model revealed the increase of intercellular heterogeneity in the cancer cells as the adaptation to chemotherapy.

Fig 6.4 Tailwind effect unraveled by CP-MC model demonstrated the role of chromatin packing scaling in cancer cell adaptation to stress as facilitating malleability.

Fig 6.5 The relationship between transcriptional divergence (P50/50) and the outcome of patients with stage III and stage IV breast, colon and lung.
Chapter 1: Introduction

1.1 Aims and Scope

Transcription is a complex, multifactorial process happening on different hierarchies of human genome. The understanding of transcriptional activity plays an important role in unraveling complicated epigenetic problems including diseases and cellular plasticity. Previous studies have been focused on finding the connection between the behavior of individual genes and diseases including immune disorders and cancers. However, the studies on human genome project have revealed that the development of diseases and the ability of cells to accommodate environmental perturbations are the products of a sizable fraction of genes, involving multiple biological processes. The regulation of expression for large size of genes simultaneously in cell nuclei remains a challenging problem. Recent studies of diagnosis techniques using optical properties as the signal determining the status of patients, e.g. Inverse Spectroscopic Optical Coherence Tomography and Low-coherence Enhanced Backscattering, have shown the strong correlation between chromatin nanoenvironment and the progressiveness of diseases. The studies using the system biology and the computational biology method indicated the interplay between chromatin nanoenvironment and transcription, which has a major impact on the development of phenotype plasticity and adaptation [1-3]. Moreover, the emerging of chromatin 3D structure capture techniques including HiC [4, 5] and Partial Wave Spectroscopic (PWS)[6] Microscopy and the fast development of gene expression measurements microarray and RNA-seq techniques allow us to monitor the essential interplay between chromatin nanoenvironment and transcription, which makes it possible to reveal the regulation of chromatin packing code on diseases and cellular plasticity. In that regard, we developed a model combing system biology, computational biology
and analytical estimation to unravel the impact of the scaling of chromatin packing (CP) with local macromolecular crowding (MC) on transcription [3]. The statistical properties of total gene expression were evaluated to unravel the impact of chromatin packing on global transcription pattern. The model predicted results were validated using imaging, chromatin structure capture and sequencing techniques including HiC, PWS, microarray and RNA-seq. The mechanisms, through which the cells could regulate their plasticity, resistance and adaptation, indicated by the gene expression statistical properties are discussed to show the potential of engineering the genomic information space by regulating chromatin packing. Furthermore, the chromatin protective therapies (CPTs)[2] were proposed to improve the efficacy of chemotherapy by targeting the physical structure of chromatin based on the macrogenomic engineering mechanisms proposed by the model and experimental evidences. Lastly, the CP-MC model was utilized to discuss the major mechanisms behind phenotype plasticity as a response to the environmental stressors.

1.2 Optical properties of cell nuclei indicate the changes of chromatin structure for cells with diseases

Optical properties in biological media are strongly correlated with the physical structure of the media [7, 8]. Multiple optical instruments measuring the optical properties of tissue or cells have indicated a consistent altering of biological structure in patients with diseases including cancer. Better understanding the connection between optical properties and physical structure of biological samples can serve to facilitate the understanding of the origin of diseases and improve the treatment of diseases. In Chapter 2, we propose a novel analysis technique using Electron
Microscopy (EM) to calculate optical properties of specific biological structures based on the physical structure of cell nuclei imaged by EM [9]. This method is demonstrated with images of human epithelial colon cell nuclei. The spectrum of anisotropy factor $g$, the phase function and the shape factor $D$ of the nuclei are calculated. The results show strong agreement with an independent study. Since the optical properties measured from EM method are directly based on the physical detail inside the nuclei, the consistence between the EM method and previous studies on optical properties in cancer tissue and cells indicates that the change of physical structures in cancer tissue/cells could be a universal independent signal to study the progressiveness of diseases and carcinogenesis.

1.3 Crowding nanoenviroment regulates transcription and chromatin packing density

Nuclei are crowded with macromolecular, e.g. DNA, RNA, proteins and chromatin, which is the largest media in nuclei. As the biggest part within nucleus, it has been shown that the packing density scaling of chromatin $D$ is strongly correlated with the diseases such as cancer [2, 10]. The question remains as how the change of $D$ is related to the change of biological functions. Studies on the macromolecular crowding system using MD simulation have shown that the crowding has significant effect on chromatin folding, chromatin accessible surface area, maintenance of nuclei bodies and gene transcription [11-13]. In crowded media, the biomolecular associations, e.g. self-association of proteins and the binding between DNA and transcription factor, are enhanced. The increasing of the association can be explained by the increasing depletion attraction in crowded biological system [14, 15]. The depletion attraction is the result of the excluded volume effect of hard spheres, which are used as the coarse grain model for the macromolecular crowding in nuclei. The crowding effect on gene transcription is the combination result of increasing depletion
attraction and the decreasing of diffusion coefficient of transcription factor when crowding volume fraction is increasing [16, 17]. Fig. 1(a) shows how the depletion potential is changing with increasing crowding volume fraction. The influence of crowding on diffusion coefficient is shown in Fig. 1(b). The work is done by Hiroaki Matsuda et al [17]. The study on the crowding effect on gene transcription shown in Fig. 1(d) shows a non-monotonic relation between the crowding volume fraction $\phi$ and steady state transcription rate of mRNA. This results suggests the modulating power of crowding on gene transcription, which is tightly related to cancer progressing. This model used by Hiroaki Matsuda et al shows us a promising platform to develop a general Molecular System Approach (MSA), in which an integrating system is developed to study the relation between chromatin structural alternation and transcription in carcinogenesis. However, the real biological system as complicated as nuclei is filled with different interactions including the direct attraction/repulsion between chromatin, transcription factors and crowders. In this simple model from Hiroaki Matsuda et al, only the excluded volume effect driven by crowders is taken into account, which equivalent to the pure repulsion.
Fig. 1.1 The previous result of the general MSA study of transcription in crowding environment by Hiroaki Matsuda et al. (a) shows the dependence of diffusion coefficients on the crowding volume fraction for binding agents with different particle size. (b) is crowding induced free energy change for TF when the crowding volume fraction is 25%. (c) is the example of the molecular system used to calculate diffusion coefficients and free energy in (a-b). The DNA, TF, RNAP, and crowders are represented in green, orange, red and cyan separately. (d) shows the final steady state concentration of mRNA in cytoplasm.

However, the relationship between chromatin packing scaling and the transcriptional outcome has not been fully understood since the previous MSA model only considered the pure repulsive interaction. Therefore, more complicated interaction including non-specific attraction is further
considered in the MSA model and discussed in Chapter 3, finding that the intermolecular attractive potential can regulate both the binding potential and the chromatin packing scaling. Since the binding potential would directly influence the transcription rate of mRNA, as indicated by MSA, the physical interaction within nuclei will significantly affect transcription. Moreover, the interplay between physical interactions, chromatin packing scaling and transcription suggests the important role of chromatin packing scaling in regulating general transcription pattern.

1.4 Chromatin packing capturing techniques assists the understanding between chromatin nanostructure and the behavior of organisms

The optical properties studied in the Chapter 2 has indicated the change of the special organization of macromolecular in nuclei while Chapter 3 has demonstrated the important role of chromatin packing scaling in regulating transcription rate of mRNA. However, due to the lack of imaging techniques capturing the 3 dimensions (3D) structure of chromatin and the floating nucleoplasm, how the chromatin packs itself into a heterogeneous system and where the changes of structure come from during carcinogenesis are still not well studied. The development of chromatin 3D structure techniques ChromEM, HiC, PWS and STORM provide an integrated platform to study the mechanism that the chromatin is packing into an inhomogeneous media [4, 6, 18, 19]. We are going to discuss how to connect homogenous chromatin packing density to the scaling of chromatin packing density $D$ with BD simulation in Chapter 4. The predicted results from the simulation will be validated with HiC and PWS measurements. The study of chromatin packing density scaling $D$ plays an important role in monitoring the change of chromatin structure at real time since $D$ is the major parameter measured in the live cell PWS system, which makes it possible
to engineer the chromatin packing scaling in a preferable direction, as what we are going to discuss in the later chapters.

1.5 *Global transcription pattern is regulated by chromatin packing*

The MSA model unraveled the interplay between chromatin packing density (crowding density) and transcription rate. The BD simulation provides the method to connect the chromatin packing density to the power law scaling of chromatin globule, which can be measured experimentally through HiC and PWS. The combination of the MSA model and the power law scaling property of chromatin makes it possible to study the impact of the heterogeneous chromatin nanoenvironment on global transcription pattern. In Chapter 5, we are going to discuss the Chromatin Packing – Macromolecular Crowding (CP-MC) model, which unravel the essential role of chromatin packing on regulation the genomic information space[2, 3, 20]. In summary, the CP-MC model indicates that the increase of chromatin packing density scaling could lead to the increase of transcription heterogeneity and diversity, which facilitate the cells to explore greater genomic information space and enhance the expression of genes simulated under environmental stress.

1.6 *Regulates cellular plasticity through engineering global transcription pattern*

Cellular plasticity plays a major role in determining cells ability to accommodate environmental perturbations. The unraveling of chromatin packing code presented us with a new way to regulate cell plasticity and adaption, which have been a challenging problem in chemotherapeutics. Utilizing the findings revealed by CP-MC model, we can discover new drugs that target the physical structure of chromatin to regulate the global transcription pattern to improve the efficacy of chemo therapy, i.e. the chromatin protective therapies (CPTs)[2], which is going to be first
discussed in Chapter 5. By modulating the chromatin packing scaling, the CPTs could engineer the macrogenomic information space and facility to increase the efficacy of chemotherapy. In Chapter 6 [3], we are going to discuss the further implication of CP-MC model propose two mechanisms through which the cells develop phenotype plasticity by chromatin packing scaling.
Chapter 2: Chromatin structure to optical properties

2.1 Introduction

Optical properties have been investigated in numerous studies due to the emergence of optical techniques used in diagnosis [7, 8, 21-24], therapy [25, 26] and surgery [27]. Specifically for diagnostic applications, changes in optical properties in tissues and cells can be detected far earlier than any available histological, molecular or genetic markers for many diseases [28-31]. It has been shown that optical properties have the ability to sense sub-diffractional structures down to 30 nm [32, 33]. While conventional optical imaging is constrained by the diffraction limit of light, optical properties are not because they quantify the propagation and scattering of light in tissue. At these small length scales, it is possible to detect the subtle alterations of tissue structure presenting in numerous diseases, which are impossible to observe by traditional optical imaging techniques [10, 34]. The great utility of characterizing optical properties has stimulated the development of a myriad of measurement techniques including integrating sphere [35, 36], frequency-domain photon migration [37, 38], spatially resolved transmittance and spatially resolved reflectance spectroscopy [23, 39-42].

Unfortunately, there are three major limitations among prevalent optical property measurement techniques. First, they do not allow direct nanoscale visualization of the structures they are sensitive to. Second, the optical properties usually average the nanoscale information over a relatively large volume (μm³ to mm³), so conventional optical property measurements do not isolate optical properties to small structures (e.g. organelles) within intact tissue. Lastly, many
optical property extraction methods are based on approximated scattering phase function models [24, 43], which make the measured optical properties vulnerable to the accuracy of the chosen model. These limitations hinder optical property measurement techniques from having insight into the underlying biology presenting the properties being measured. Overcoming these limitations becomes crucial when trying to improve techniques using optical properties as diagnostic markers.

The major technical challenge to overcome these limitations comes from the lack of optical imaging techniques that resolve structures responsible for light scattering (organelles, collagen fibers, chromatin, etc.). Fortunately, the development of nanoscale imaging techniques such as electron microscopy (EM) provides a way to overcome this challenge. EM has the ability to achieve resolution better than 100 pm, and thus can be a great imaging technique to isolate specific structures in biological samples and study the contribution from each structure to the optical properties. While EM has been used extensively in biomedical research, to our knowledge there is no report of using EM to extract optical properties from biological media.

In this work, we present a methodology to use EM to directly extract the optical properties of specific biological structures using the Born approximation. The Born approximation, also known as Rayleigh-Gans-Debye approximation, or weak scattering approximation is a valid predictive model for scattering in tissue and cells due to their weakly scattering properties [39, 44, 45]. We use the Born approximation and the known refractive indexes of cellular components to calculate the optical properties of nuclei from human epithelial cells measured by Transmission Electron Microscopy (TEM), taken from the rectal mucosa of patients with and without precancerous lesions in their colon. In the results, we first show the spectrum of anisotropy factor g and the phase function of colon cell nuclei. Then, the calculated optical properties are compared with the
optical properties of epithelial cells measured by Inverse Spectroscopic Optical Coherence Tomography (ISOCT), a technique capable of measuring the spatially resolved optical properties of cells and tissues [24, 40]. The directly calculated phase function of colon cells is compared with the Henyey-Greenstein (HG) phase function, which is commonly used in biological media [43, 46, 47].

The paper is organized as follows: In section 2 we show the basic theory we used to develop our method. In section 3 we describe the method we used to calculate the optical properties from TEM nuclei images in detail. Section 4 shows the results from human epithelial cells and comparison with an independent study. Finally, in Section 5 we discuss our conclusions and the potential impact of this work.

2.2 Application of the Born approximation to calculate optical properties

The Born approximation, also known as Rayleigh-Gans-Debye approximation, has been used to calculate optical properties in biological media [44, 48]. Here we give a brief review of this theory. A scattering event happens in media when there is a heterogeneous distribution of refractive index $n(\vec{r})$. The refractive index of a medium has a linear relationship with its mass density $\rho(\vec{r})$, which can be expressed by the Gladstone-Dale relationship [38, 49]:

$$n(\vec{r}) = n_{water} + \alpha \rho(\vec{r}) \quad (1)$$

where $n_{water}$ is the refractive index of water and $\alpha$ is the incremental constant for a specific structure. This is the reason why optical scattering phenomena (originating from $n(\vec{r})$) can be related to the structural information of the media (represented by $\rho(\vec{r})$).
Under the Born approximation, the incident electromagnetic wave is the driving field for the scattering event. This approximation is valid in biological media since its refractive index follows the weak scattered condition, which requires $|n/(\langle n \rangle - 1)| \ll 1$ and $2ka|n/(\langle n \rangle - 1| \ll 1$, where $n$ is the refractive index of the media, $\langle n \rangle$ is the averaged refractive index of the background, $k$ is the free-space wavenumber and $a$ is the size of the scattering particle. Under this condition, the scattered field is very small compared to the incident field and its contribution to the total field can be neglected. We will apply the Born approximation in the calculation of the scattering electric field because of its high efficiency compared to more accurate, but more time-consuming computational methods like the Finite-Difference Time-Domain method (FDTD) [50]. Under the Born approximation, the relationship between the differential scattering cross section and autocorrelation function of the refractive index distribution from the sample is simplified to a Fourier Transform [51, 52].

Here, as an example, we will use a cell nucleus to illustrate how to apply this approximation in the scattering calculation, though this method can be expanded to any organelle. Since a scattering event happening in the randomly distributed nucleus is in the weak scattering range, we can consider the nucleus as a particle with a finite size and a random distribution of mass density inside the particle. In this case, the origin of the scattering comes from the relative excess refractive index in the form of $n_\Delta(\vec{r}) = (n_n(\vec{r}) - \langle n_{cyt} \rangle)/\langle n_{cyt} \rangle$, where $n_n(\vec{r})$ is the refractive index distribution inside the nuclei and $\langle n_{cyt} \rangle$ is the averaged refractive index of cytoplasm. The autocorrelation function of the nucleus is represented in Eq. (2):

$$B_n(\vec{r}_d) = \int n_\Delta(\vec{r})n_\Delta(\vec{r} - \vec{r}_d)d\vec{r}^3$$  \hspace{1cm} (2)
By considering that the random media is statistically isotropic, the Fourier Transform of $B_n(\vec{r}_d)$ in 1-D form is:

$$\Phi_s(k_s) = \frac{1}{2\pi^2} \int_0^\infty B_n(r_d) \left[ \frac{\sin(k_s \cdot r_d)}{k_s r_d} \right] dr_d$$  \hspace{1cm} (3)

where $k_s = 2(n_{cyt}) k \sin(\theta/2)$, $\theta$ is the azimuth angle and $\Phi_s(k_s)$ is the power spectral density (PSD) of the refractive index. According to the Born approximation, the differential scattering cross section $\sigma(\theta, \phi, k)$, where $\phi$ is the polar angle in spherical coordinate, is proportional to the PSD. So, we can calculate the differential scattering cross section as:

$$\sigma(\theta, \phi, k) = 2\pi \left( \left\langle n_{cyt} \right\rangle k \right)^4 \left( 1 - \sin^2 \theta \cos^2 \phi \right) \Phi_s(k_s)$$  \hspace{1cm} (4)

Because we assume unpolarized light, $\sigma$ is independent of $\phi$, and the differential scattering cross section can be simplified to its 1-D form:

$$\left( \sigma(k) = 2 \left\langle n_{cyt} \right\rangle k^4 \left( 1 + \cos^2 \right) \Phi_s(k) \right)$$  \hspace{1cm} (5)

With this expression of the differential scattering cross section, we can now calculate the anisotropy factor $g$ as the first moment of $\sigma(\theta, k)$:

$$g(k) = \frac{\int_{-1}^1 \cos \theta \cdot \sigma(\cos \theta, k) d \cos \theta}{\int_{-1}^1 \sigma(\cos \theta, k) d \cos \theta}$$  \hspace{1cm} (6)

As well as the transport mean free path:
\[ l_s^*(k) = \frac{1}{\int_{-1}^{1} \sigma(\cos \theta, k) d \cos \theta \cdot (1 - g(k))} \] (7)

The “shape factor” D is another optical property measured in many studies [39, 52-54], and has been found to have a power law relation with the spectrum of incident light in the biological tissue in these measurements:

\[ l_s^*(k) \propto k^{D-4}, D \leq 4 \] (8)

This power law relation can be derived under a number of analytical autocorrelation function models such as the Whittle-Matern (WM) model [55]. The WM model is an autocorrelation family which actually encompasses many models including the mass-fractal model and HG phase function and is often used to model biological samples [39, 44]:

\[ B_n(r_d) = A_n \left( \frac{r_d}{l_n} \right)^{D-3} \frac{K_{D-3}}{2} \left( \frac{r_d}{l_n} \right) \] (9)

where \( A_n \) is a normalization factor, \( l_n \) is the length scale factor of the modeled sample, \( K_{(D-3)/2}(r_d/l_n) \) is the modified Bessel function of the second kind. The power law relation is only valid when \( kl_n \gg 1 \). The detail of this relation can be found in [53].

Since we are only studying the property of single nuclei, whose size is much smaller than mean free path, we can assume that there is only single scattering in nuclei. Therefore, \( l_s^* \) does not have physical meaning but the relation with D is still valid. To calculate a physically meaningful value of \( l_s^* \), we would also need to measure the number density of nuclei in a volume of tissue, which is not within the scope of this study.
2.3 From TEM Images to Optical Properties

Section 2 shows that optical properties \( g, D \), and any other phase function dependent parameters can be calculated from \( B_n(r_d) \) extracted from a 3D nanoscale image of a sample. However, to our knowledge, there is no well-established direct imaging technique which provides 3D nano-scale (down to 10 nm) visualization of structures in biological samples. Here we will present a method to calculate the optical properties from the \( B_n(r_d) \) extracted from 2D TEM images of cell nuclei.

2.3.1.1 Acquire nuclei TEM images

The biopsy preparation protocol and TEM measurement parameters are described in detail in [52]. Briefly, ten patients who had undergone colonoscopy were recruited into this study. Five patients had no signs of dysplasia or adenomas (precancerous lesions) throughout the colon and were considered a control group. Five patients presented with advanced adenomas (AA group) or adenomas larger than 9mm in diameter. A biopsy was taken from each patient’s colon and was then stained with osmium tetraoxide (OsO₄, which is commonly used to target DNA [56, 57]). TEM images of tissue sections from the biopsies were acquired after excision.

Figure 1(a) shows an example of a TEM image used in this study. The nucleus in the selected region is divided into two parts. The darker part corresponds to heterochromatin, which has denser DNA concentration, and the lighter part corresponds to euchromatin, which has a lower DNA concentration. These two parts are the largest components in nuclei. To isolate nuclei and their
contribution to the optical properties, we cropped each nucleus out of the image and set the background outside the nuclei to be uniform cytoplasm, as shown in Fig. 1(b-c).

Fig 2.1. Nuclei in a TEM image. (a) is the original TEM image of the biopsy. (b) is the selected nucleus in TEM image. (c) is the cropped out nucleus. The background of the image is taken as uniform cytoplasm. (d) is the nuclear refractive index distribution $n(x, y)$ after binarization. The refractive index is 1.35 for euchromatin, 1.36 for cytoplasm and 1.39 for heterochromatin. The scale bar in (a) is 2μm, the scale bar in (b-c) is 2.5μm.

2.3.1.2 2D Refractive index autocorrelation function of nuclei TEM image

The TEM images provide the structural distribution information inside the nuclei but not the direct refractive index distribution. However, since the nuclei mainly consist of heterochromatin and euchromatin, dividing the nuclei images into those two parts and assigning the appropriate refractive indexes to each part can provide a good approximation of the refractive index distribution inside nuclei [58]. Note that, this method of approximating refractive index distribution might not be applicable for all the sample types when the refractive indexes of the
components inside the sample are unknown. The refractive index outside nuclei is taken as the averaged refractive index of cytoplasm, which we set to 1.36. This value was taken from direct refractive index measurements [59, 60]. We use 1.35 and 1.39 as the refractive index values for euchromatin and heterochromatin, respectively. These values are based on their respective mass densities and using the Gladstone-Dale relation since, to our knowledge, there is no direct measurement of their refractive indices. The term mass density in this work refers to the mass density only of chromatin (i.e., in the absence of water). Thus, the mass density of euchromatin is 0.1 g/ml and the heterochromatin is 0.3 g/ml [15, 61-63], α in the Gladstone-Dale relationship is 0.2 ml/g [64, 65] for chromatin, and the refractive index for water is 1.33. The final 2D refractive index distribution $n(x, y)$ is shown in Fig. 1 (d).

Using convolution theorem, we calculate the autocorrelation function of $\Delta n(x, y)$, where $\Delta n(x, y) = n(x, y) - n_{cyt}$:

$$
B_n(\Delta x, \Delta y) = \int \int \Delta n(x, y) \cdot \Delta n(x - \Delta x, y - \Delta y) dx dy
$$

$$
= F^{-1}\left\{ \left| F\{\Delta n(x, y)\}\right|^2 \right\}
$$

(10)

By assuming that the samples from the same region of the same patient are statistically isotropic, we convert this 2D autocorrelation function $B_n(\Delta x, \Delta y)$ into the 1D autocorrelation function $B_n(r_d)$:

$$
B_n^{2D}(r_d) = \int B_n(\Delta x, \Delta y) d\theta = \int B_n(r_d \cos \theta, r_d \sin \theta) d\theta
$$

(11)

where $B_n^{2D}(r_d)$ denotes the autocorrelation function from 2D images.
2.3.1.3 Extract 3D autocorrelation function from 2D TEM images

Note that the $B_n^{2D}(r_d)$ obtained in the last subsection is from 2D images. However, scattering happens in 3D and $B_n^{2D}(r_d)$ cannot fully represent the statistics of $B_n(r_d)$. This is because the 2D image is only one slice of the full 3D nucleus. A 3D image of the nucleus would consist of many 2D slices, depending on the thickness of the slices and nucleus. So the statistical information in the 2D image of the sample is not enough to represent the full statistical properties of the 3D sample. The difference between $B_n^{2D}(r_d)$ and $B_n(r_d)$ is shown in Fig. 3(b). The 3D random media (RM) in Fig. 3 is generated using the WM model through a publicly available MATLAB code [66]. We assume biological media is statistically homogeneous, therefore we can average a sufficiently large number (N) of $B_n^{2D}(r_d)$ to represent $B_n(r_d)$. Fig. 3(c) shows the convergence of $B_n^{2D}(r_d)$ and $B_n(r_d)$ with average N=1000 $B_n^{2D}(r_d)$. We present a method to determine the minimum N of $B_n^{2D}(r_d)$ images needed to average over to accurately represent $B_n(r_d)$ of biological samples. The minimum N is dependent on the sample properties and image parameters.

To best model the biological sample, we model tissue using the versatile WM correlation family discussed in Sec. 2 with biologically relevant parameters, $D=3$ and $l_n = 1 \mu m$ [40]. For the image parameters, the resolution and grid size of our TEM images are $dx = 10 nm$ and grid_size= 801. We generated the RM with these four inputs. To determine the minimum N needed to represent $B_n(r_d)$ from a 3D sample, we compared $B_n(r_d)$ from 3D media to $B_n^{2D}(r_d)$ averaged over 1 to 15 2D images of RM. The results from these steps show that after averaging $B_n^{2D}(r_d)$ from 10 RM 2D slides, $B_n^{2D}(r_d)$ converges to $B_n(r_d)$ with $R^2 > 99\%$. Fig. 3(c,e,f) shows these results. Based on these observed results, we divide all the TEM images in each group into subgroups and
each subgroup contains 10 TEM images of nuclei. The $B_n(r_d)$ of each subgroup is averaged over 10 $B_{n}^{2D}(r_d)$ of TEM images in that subgroup.

**Fig 2.1** Extracting $B_n(r_d)$ from 2D images. (a) is an example of numerically generated random media. (b) shows the difference between $B_n(r_d)$ from 3D random media and 2D random media. (c) shows the $B_{n}^{2D}(r_d)$ from 2D slides of random media converge to $B_n(r_d)$ from 3D RM. (d) is one example of RM with biological parameters of the nuclei TEM images. (e) shows the $R^2$ between $B_n(r_d)$ and N averaged $B_{n}^{2D}(r_d)$. This result shows that when $N \geq 10$, $R^2$ between
$B_n(r_d)$ and $B_n(r_d)$ from averaging over $N B_n^{2D}(r_d)$ is larger than 99% and the difference is negligible. (f) is an example of $B_n(r_d)$ after averaging over 10 $B_n^{2D}(r_d)$ of nuclei TEM images (RI represents refractive index).

2.3.2 Calculate optical properties from the autocorrelation function

The differential scattering cross section $\sigma(\theta, k)$ and the spectrum of $g, g(\lambda)$, for the measured sample can be calculated from $B_n(r_d)$ using Eq. (3-6), with the relation $k = 2\pi/\lambda$. The phase function measure from TEM images $P_{TEM}(\theta)$ at a specific wavelength $\lambda = \lambda_0$ will be:

$$P_{TEM}(\theta) = A_0 \sigma(\theta, 2\pi/\lambda_0)$$

(12)

where $A_0$ is a normalization factor so that:

$$\int_0^{2\pi} \left\{ \int_0^\pi P_{TEM}(\theta) \sin(\theta) d\theta \right\} d\phi = 1$$

(13)

The shape factor $D$ can be calculated with the exponential relationship between $l_s^*$ and $D^{-4}$, as shown in Eq. (8). We fit the spectrum of $l_s^*$ in the range of $k$ corresponding to 600 nm to 700 nm to find $D$. To simplify the calculation, we take the log of both sides of Eq. (8) to get Eq. (14).

$$\log(l_s^*(k)) = (D-4) \log k + C$$

(14)

where $C$ is a constant independent of $D$ and $k$. We fit the data with a linear regression and the slope of the line is equal to $D^{-4}$. 

2.4 Results

Using the previously described methodology, we first calculated $B_n(r_d)$ from nuclei TEM images from 5 patients in control data set and 5 in the AA data set separately, as shown in Fig. 3 (a). Figure 3(a) shows that $B_n(r_d)$ from the control set has a lower magnitude at lower length-scales and has a different shape. Then the corresponding $P_{TEM}(\theta)$ when $\lambda_0 = 700\text{nm}$ and $g(\lambda)$ are calculated from $B_n(r_d)$, as shown in Fig. 3(b-c).
Fig 2.3 Results for the autocorrelation function, spectrum of g and the phase function for control sample and AA sample. (a) is the $B_n(r_d)$ from control data set and AA data set. (b) shows the phase functions calculated from TEM images. Phase functions are calculated at $\lambda_0 = 700$nm. (c) shows the spectrum of g for the range of $\lambda$ between 500nm to 700nm.
We then calculated the corresponding optical properties $g$ and $D$ of colon cell nuclei from TEM images from a control set and AA set, where $g$ is calculated at $\lambda_0 = 700\text{nm}$, as shown in Fig. 3. The $g$ and $D$ measured from the EM method is compared with those measured previously with ISOCT by Yi et al [40] as shown in Fig. 4. The ISOCT measurements in Fig. 4 were conducted by an open space Fourier-domain OCT configuration with an illumination wavelength ranging from 650 to 800 nm. The samples measured were colonic mucosa biopsies consisting of a 20 to 30 $\mu$m epithelial cell layer surrounded by the lamina propria from 85 patients with and without colorectal adenomas. The epithelial layer is segmented out of the image to isolate the ISOCT signal from its constituent cells as described by Yi et al [40]. The cell segmentation gives ISOCT the ability to quantify the average optical properties of multiple cells.

Because of the differences in $B_n(r_d)$ shown in Fig. 3(a), the optical properties calculated from each set are significantly different. Both $g$ and $D$ increase from control set to AA set. The comparison in Fig. 4 shows agreement the values of $D$ and $g$ between the two studies with differences less than 5% and 10% respectively. The differences between the two studies likely originate from the fact that ISOCT averages the optical property over whole cells and many organelles. While it lacks the resolution to isolate the nucleus; the largest contribution to the average optical properties of the cell is likely from the nucleus. Assuming most of the OCT signal measured from epithelial cells originates from their nuclei [67, 68], the agreement of the results between this method and ISOCT measurements illustrates the accuracy of our method. The increase in $g$ and $D$ in the AA set originates from the flatter slope of $B_n(r_d)$. Since the rate of change in $B_n(r_d)$ describes how rapidly the structure is changing, the structures sharing similar refractive index should be larger in the AA set. This is explained by the increase in the volume
fraction of heterochromatin in the AA set, as previously observed by Cherkezyan et al [52]. While the increase of heterochromatin implies a decrease of euchromatin, the volume fraction of heterochromatin is smaller compared with euchromatin [52]. This means that the heterochromatin’s volume change would cause more dramatic change to the autocorrelation function.

Fig 2.4 Results of the optical property calculated from colon cell nuclei TEM images, compared with ISOCT result. ** indicates the p-value is smaller than 0.05. The result from TEM image shows significant difference in g and D between control data set and AA set. Compared with ISOCT measurement, the TEM method shows the same trend between control and AA for g and D. In the case of D, the values between two methods match with each other within 5%.

One great advantage of this analytical technique is that it calculates the real phase function from the sample, without fitting to any analytical model. For instance, the HG phase function is one of the commonly used models to represent the scattering phase function for biological samples and has the following form:
\[ P_{HG}(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{[1 + g^2 - 2g \cos(\theta)]^{3/2}} \] (15)

One obvious shortcoming of using the HG phase function model is that the shape factor D is not included in the model, and therefore the HG phase function cannot capture the difference in D shown in Fig. 4. Furthermore, the comparison between the real phase function calculated directly from sample using our analysis technique, and the HG phase function model are shown in Fig. 5. Here, we use g=0.8 for control nuclei and g=0.85 for AA set nuclei, which are the values obtained from our TEM analysis, when \( \lambda_0 = 700\text{nm} \). Both of the phase functions use the normalization factor described in Eq. 13. The results show that the HG phase function gives a nice fit to the calculated phase function of nuclei within the angles between 30 to 60 degrees. For the smaller and larger scattering angles, the HG phase function fails to explain the higher scattering probability at smaller and larger angles, which is consistent with previous investigations of the HG phase function [69]. At small scattering angles, where \( \theta < 5^\circ \), The HG phase function predicts lower scattering probability when \( \theta < 2^\circ \), but higher scattering probability when \( \theta \) is within \( 2^\circ \) and \( 5^\circ \). This difference between HG phase function and the measured phase function is also observed in red blood cells when the angle is smaller than 5 degree [70]. The disability of HG to accurately model the scattering probability at the forward and backward scattering directions demonstrates the great benefit of calculating the true phase function of a sample.
Fig 2.5 The comparison of the phase functions from TEM images (The solid lines) and the HG relation (The dash lines). Both phase functions are calculated at $\lambda_0 = 700\text{nm}$

2.5 Discussion and conclusion

In this work we showed how EM images of tissue sections can be used to calculate the true phase function and optical properties of biological structures. We accomplished this by applying the Born approximation. We showed significant differences in the phase function and optical properties calculated from colon cell nuclei from patients with and without advanced adenomas. These findings were in excellent agreement with an independent study. Furthermore we demonstrated
the advantage of calculating the true phase function from the nuclei, as opposed to fitting to a constrained model.

While this EM image analysis provides a novel method to calculate the autocorrelation function, phase function, and optical properties directly from biological samples, there are two aspects of this technique that can prove to be potential challenges. 1) The application of this EM image analysis requires knowledge of the refractive index distribution of the structures in the image. 2) as discussed in Sec. 3, the theory being employed to calculate the power spectral density function requires calculation of a 3D autocorrelation function from the sample (i.e., it requires a 3D nanoscale image of the sample).

The first aspect was a challenge in this study due to the dying process the tissue sections underwent during sample preparation. This process does not necessarily allow a linear relationship between the refractive index distribution and the image intensity. However, since the refractive index of chromatin is known, the binary process described in Sec. 3 allowed us to overcome this challenge. The binary process will reduce the real fluctuations of the refractive index distribution, and consequently decrease the absolute value of the autocorrelation function. However, these fluctuations are not significant compared with the drastic difference in refractive index between euchromatin and heterochromatin. This binary method can be used in many different biological structures with the caveat that it requires knowledge of the refractive index of each component in the sample image. This means that measuring the optical properties of larger biological samples like whole cells and bulk tissue with this same technique could be challenging because of the lack of experimental measurements of the refractive index of the different components in the samples. For samples with many components with unknown refractive indexes, the recently developed EM
techniques that require no staining might open the door to measuring the refractive index distribution directly of the sample [71, 72]. In these unstained EM images, the intensity of EM image will have an approximately linear relationship with the mass density of the sample. The refractive index distribution of the sample can be easily calculated by applying the Gladstone-Dale relationship.

as discussed in Sec. 3, we overcame the second challenge by using 2D TEM images and averaging them to represent the 3D autocorrelation function. This solution is valid under the assumption that the sample is statistically isotropic. This was necessary due to the lack of 3D nanoscale imaging data sets of biological samples. In the case of measuring a statistically anisotropic sample, the direct 3D nanoscale imaging is needed to calculate the optical properties accurately.

With the possibility of unstained EM techniques, the refractive index distribution of the sample can be measured directly with EM without prior knowledge of the reactive index of each component in the sample. This would allow the EM analysis technique to provide an independent validation of optical property measurements and better understanding of the microscopic origins of the optical properties of cells and tissues in diseased and healthy states. This understanding is crucial for the development, validation and implementation of biomedical optics diagnostic techniques. Specifically, the further application and development of this method can bring us three key benefits: 1) Extraction of the real phase function from any tissue type. 2) The optical properties of each organelle in tissue can be measured independently. 3) Establishing a connection between optical properties and nanoscale structure (e.g. organelles) allows in-depth investigation of the nano-architecture of cells and tissue using optical techniques.
The difference between the real phase function calculated from TEM nuclei images and the HG phase function model indicates that the HG model cannot capture the phase function of the biological sample at small and large angles. This is partially because the HG model does not have a dipole factor, which leads to an observed non-monotonic shape at smaller angles [44, 73, 74]. With the development of experimental methods measuring different optical properties, a method that can retrieve the real phase function is necessary to explain the observed data. There is no phase function model which is capable of precisely modeling all biological samples. To extract the true optical properties of any biological sample, the real phase function is necessary. Extracting the real phase function is also very useful for simulation methods (e.g. Monte Carlo) that use the phase function as an input. These simulations traditionally make assumptions about the shape of the phase function by using models like the HG model [43, 46, 47, 75]. However, inputting the real phase function of a sample from this EM analysis will significantly improve the accuracy of light transport simulations.

The isolation of optical properties of organelles can improve the accuracy of optical diagnostic tools. Many existing diagnostic instruments use optical properties as a marker of disease progression and as the disease progresses this biomarker becomes altered. Our method can isolate the source of this alteration in terms of changes to the organelle structures. This allows researcher to build a direct connection between their optical measurements, and the subcellular alterations in their studies. This can lead to better understanding of disease progression, and improve the development of optical instruments and help improve their ability to target specific structural alterations in disease progression.
Using optical measurements to interpret the nanostructure of biological samples can largely reduce the need for nanoscale imaging. Once the connection between optical properties and the nanostructure of biological samples is built, the repeated measurements of samples by using nanoscale imaging techniques will not be necessary. In other words, the results from optical property measurements can be remapped to the nanostructure information of the sample, which will save time and consumable cost from using nanoscale imaging techniques.

In summary, we developed a methodology of using EM to calculate the optical properties of biological samples from their nano-scale images. The optical properties of cell nuclei measured by this method are validated by an independent optical technique, ISOCT. With further development, we envision this technique could become a standardized method to validate new experimental instruments that measure optical properties. Also, this EM analysis technique provides a way to gain a crucial understanding of the structural origins of optical properties for a given sample. In this work we validated the assumption that changes in chromatin organization are primarily responsible for the optical property changes in colon epithelia observed in a separate study. To investigate other structures responsible for the altered optical properties, we can extend these investigations to other cellular structures such as mitochondria and other organelles.
Chapter 3: The Crowding Environment in Nucleus Influences the General Transcription Reaction and Chromatin Structure

3.1 Introduction

Nuclei are crowded with macromolecular, e.g. DNA, RNA and proteins. The changes of macromolecular crowding density in cancer cell nuclei is one of the most important nuclei architecture alternations during carcinogenesis. Studies on the macromolecular crowding system using MD simulation have shown that the crowding has significant effect on chromatin folding, chromatin accessible surface area [13], maintenance of nuclei bodies [76] and gene transcription [1]. In crowded media, the biomolecular associations, e.g. self-association of proteins and the binding between DNA and transcription factor, are enhanced [77]. The increasing of the association can be explained by the increasing depletion attraction [78] in crowded biological system. The depletion attraction is the result of the excluded volume effect of hard spheres [78], which are used as the coarse grain model for the macromolecular crowding in nuclei. The crowding effect on gene transcription is the combination result of increasing depletion attraction and the decreasing of diffusion coefficient of transcription factor when crowding volume fraction is increasing. Fig.1 (a) shows how the depletion potential is changing with increasing crowding volume fraction. The influence of crowding on diffusion coefficient is shown in Fig. 1(b). The work is done by Hiroaki Matsuda et al [1]. The study on the crowding effect on gene transcription shown in Fig. 1(d) shows a non-monotonic relation between the crowding volume fraction $\phi$ and steady state transcription rate of mRNA. This results suggests the modulating power of crowding on gene transcription, which is tightly related to cancer progressing. This model used by Hiroaki Matsuda et al shows us a promising platform to develop a general Molecular System
Approach (MSA), in which an integrating system is developed to study the relation between chromatin structural alternation and transcription in carcinogenesis. However, the real biological system as complicated as nuclei is filled with different interactions the and the direct attraction/repulsion between chromatin, transcription factors and crowders. In this simple model from Hiroaki Matsuda et al, only the excluded volume effect driven by crowders is taken into account, which equivalent to the pure repulsion. In this first proposed research work, I will use the bases from the model of Hiroaki Matsuda et al to develop a more complicated MSA, in which the crowding system will include the real interaction between chromatin, transcription factors and crowders, depending on the function of particles in the system.

The recent developments in imaging and chromosome captures techniques show that the long-range interaction of chromatin plays an important role in controlling genome function in higher eukaryotes [11, 79-81]. More interestingly, evidence from long-range and cell-specific enhancer-promoter interactions in SHH suggestions that this long-range interaction is also related to the cancer development [81].

These long-range interaction are mostly observed using indirect approaches because the chromatin folding in the interphase nuclei cannot be followed through direct imaging techniques. The most popular methods among these indirect methods are Hi-C measurement and fluorescence in situ hybridization(FISH) [4, 82, 83]. The first Hi-C measurement revealed the a power law relationship in intra-chromosomal contact probabilities associated to a slope of -1.08 in the range of 500 kb to 7 Mbp [5]. The FISH measurement, on the other hand, unraveled the 2 chromatin folding regimes: At genomic distance smaller than 2 Mb, the end to end mean square distance $\langle R^2 \rangle$ increases with the genomic distance and at genomic distance larger than 10 Mb, there is a plateau in $\langle R^2 \rangle$ [84].
Several models have been proposed to explain these observed long-range interaction properties. The fractal globule model is the first model used to explain the power law relation in Hi-C measurement and widely adopted by the other researchers after that [5, 85]. These studies of the fractal globule model suggested that this model can well reproduce the -1 slop of in the power law relation and is compared heavily with another equilibrium model, which predicts the slop to be -3/2 instead. However, neither of these two model can predict the -1 power law relation and plateau in \( \langle R^2 \rangle \) at the same time.

More importantly, how the long-range interaction is regulated under biological environment and how it is related to the cancer development are still unclear. Interestingly, results from recent studies have shown that the cancer development is tightly related to the change of crowding environment in the cell nuclei [10, 52]. These observations have intrigued the interests in studying how the crowding condition is influencing the genetic activities. Previous studies have shown that, the crowding condition has significant influence on chromosome structure and gene transcription [1, 86]. All of these results suggest that the crowding condition is an important regulator in controlling the gene expression and it is reasonable to think that the long-range interaction is another entry regulated by crowding.

3.2 The Crowding Environment Influences the General Transcription Reaction

In this work, we want to understand what the general role of crowding environment on genome activity is with the previously developed MSA. The answer to this question should be based on
the realistic crowded nuclei system, where not only the excluded volume effect is considered but also the non-specific attractive potential between all of the components involved in the crowing system. The main focuses are to answer the following questions: When considering the attractive potential between crowders and crowders, transcription factor and crowders or considering the attraction potential between crowders and crowders, transcription factor and crowders at the same time, how the depletion potential will change, how the final transcription rate will be influenced by these intermolecular potential.

As one of the most important structures in the cell, the major functions of the components of cell nuclei includes transcription, pre-mRNA splicing and ribosome assembly [Functional architecture in the cell nucleus]. Based on the different function of the components, the different interactions are formed between them. In this proposal, I will use the crowded nuclei model to study how the interactions between different components will influence the transcription rate of protein in the cell.

The crowded nuclei model has three components, the DNA chain (the green chain in Fig 1.(c)), the transcription factors (TF) (the red and orange particle in Fig 1.(c)) and the crowders (the cyan particles in Fig 1.(c)). In this system, the chromatin chain is fixed while the transcription factors and crowders are free to move. The change of transcription rate as a result of crowders is the combination result of potential of mean force (PMF) between transcription factors and chromatin and the diffusion rate of transcription factor. In the crowders free environment, the gene expression
rate [mRNA] is determined by the reaction chains shown in the work of Matsuda et al [1]. As a brief introduction, in the crowding system, the non-specific reaction rate $k_0^{ns}$ between DNA and TF, the association rate constant for TF-promoter binding $k_0$ will be modified by the crowders' volume fraction:

$$k^{ns}(\phi) = k_0^{ns} \times \frac{D_{TF}(\phi)}{D_{TF,0}} \times \exp[-\beta \Delta F_{\text{barrier,TF}}(\phi)]$$

$$k(\phi) = k_0 \times \frac{D_{TF}(\phi)}{D_{TF,0}} \times \exp\left[\frac{1}{2} \beta \Delta F_{\text{crowd,TF}}(\phi)\right] \times \exp\left[-\frac{1}{2} \beta \Delta F_{\text{barrier,TF}}(\phi)\right]$$

where $\phi$ is the volume fraction of crowders in the nuclei, $D_{TF}(\phi)$ is the diffusion rate of transcription factor in crowded nuclei, $D_{TF,0}(\phi)$ is the diffusion rate of transcription factor in crowders free system , $\beta = \frac{1}{kT}$, $k$ is the Boltzmann constant, $T$ is the temperature, $\Delta F_{\text{barrier,TF}}(\phi)$ is the free energy barrier for TF to overcome when moving from infinite far away to attach to the chromatin in crowded system and $\Delta F_{\text{crowd,TF}}(\phi)$ is the free energy difference between TF and DNA when the TF is far way from DNA and when TF is attached to DNA, as illustrated in Fig. 1(b).

The DNA chain in the crowding system is consisted by infinite long cylinder with the diameter to be $d_{\text{DNA}} = 2$ nm. The diameter of crowders is $d_c = 5$ nm and the diameter of TF is $d_c = 6$ nm. The $\Delta F_{\text{barrier,TF}}(\phi)$ is achieved through Molecular Dynamic simulation by the collective variables module in LAMMPS. In this study, the interactions considered between DNA chain, TF and
crowders are pure hard sphere repulsion plus the Yukawa potential, which is commonly used to model the attractive potential between proteins and extensively studied \[87, 88\]. The final form of the interaction is:

\[
U(R) = U_h(R) + U_a(R)
\]

\[
U_h(R) = 4\epsilon\left[\frac{\sigma}{(R - r_0)}\right]^{12} - \left[\frac{\sigma}{(R - r_0)}\right]^{6} + \epsilon, \quad (r_0 < R < r_0 + r_c^1)
\]

\[
U_a(R) = -U_0 \left(\frac{\sigma_a}{R}\right) \exp\left(-\frac{R - \sigma_a}{d}\right), \quad (r_0 + r_c^1 < R < r_0 + r_c^2)
\]

where \(U_h(R)\) is the hard sphere potential and \(U_a(R)\) is the Yukawa attractive potential measured in the protein-protein interaction \[87\]. \(r_0 = r_1 + r_2 - 1\text{nm}\), where \(r_1\) and \(r_2\) are the radius of the two components that are interaction with each other. In our study, \(r_1\) and \(r_2\) are the two radius combination from the radius of DNA \(r_{\text{DNA}}\), the radius of crowders \(r_c\) and the radius of TF, \(r_{\text{TF}}\).

The value of \(U_0, d\) are measured in \[87\], where \(U_0 \in (-2.69, -2.63)\text{kT}\) and \(d < \sigma/5\). In this study, we choose \(\sigma_a = r_1 + r_2\), \(U_0 = -2.7\text{kT}\) and \(d = 0.5\text{nm}\).

The diffusion coefficient of TF \(D_{\text{TF}}(\phi)\) in crowding system is achieved by using the over-damped langevin simulation in LAMMPS for the crowding system. The simulation is run in the system without crowders and the interactions, in the same form as the one previously mentioned, are between TF and crowders, crowders and crowders or among three of them. With \(D_{\text{TF}}, F_{\text{barrier,TF}}\) and \(F_{\text{crowder,TF}}\) achieved from the simulation when considering the attractive potential between different components, the steady state transcription rate \([\text{mRNA}]_{\text{cyto}}\) can be then calculated based on the model in \[1\].
The central missing parts to understand how $[\text{mRNA}]_{\text{cyto}}$ is influenced by the attractive potential between components in the crowded nuclei system are the $D_{TF}$, $F_{\text{barrier},TF}$ and $F_{\text{crowder},TF}$. Our preliminary studies have shown that the crowding induced Potential of Mean Force (PMF) behaves very differently when attractive potential exists.

The comparison of PMF between pure repulsive potential and the pure repulsive plus attractive potential in Fig. 2 suggests that when only considering the attractive potential between TF and crowders (e.g. the TF-C system), the $\Delta F_{\text{barrier},TF}$ is higher while the $-\Delta F_{\text{crowd},TF}$ is smaller. The increasing $F_{\text{barrier},TF}$ suggests a decreasing $k^{\text{ns}}(\phi)$, according to Eq (1). The decreasing of $F_{\text{crowd},TF}$, on the other hand, is much higher than the increasing of $F_{\text{barrier},TF}$, which leads to the lower $k(\phi)$ in the end, based on Eq (2). Interestingly, the $\Delta F_{\text{crowd},TF}$ is larger than 0 when only the attractive potential between TF and crowders is considered. This indicates that the excluded volume induced PMF is counteracted by the attractive potential effect, and the association will be destroyed in this case. If we assume the diffusion coefficient will not change too much under attractive potential situation, which is reasonable since the interaction between TF and crowder will be counteracted by same interaction between TF and another crowder, the totally gene expression rate $[\text{mRNA}]_{\text{cyto}}$ might be significantly reduced.
Fig 3.1 The comparison of crowding induced PMF in the MAS when consider the pure repulsion, pure repulsion plus attraction between crowders and TF (TF-C system) and pure repulsion plus attraction between crowders and crowders, crowders and TF (TF-C, C-C system). (a) The crowding induced binding free energy as a function of center to center distance between crowders and TF when the crowding volume fraction $\phi$ is 9$. $\beta\Delta F_{\text{crowd}}$ and $\beta\Delta F_{\text{barrier,TF}}$ are defined as the ones in Fig. 1(b). (b) is the same PMF when $\phi$ of crowders is 32%.

When we look at the comparison between only considering the pure repulsion and considering attraction between TF and crowders and the attraction between crowders and crowders at the same time (e.g. the TF-C, C-C system), the $-\Delta F_{\text{crowd,TF}}$ is still smaller when attraction is considered, same as the precious TF-C attracting system. The comparison of $\Delta F_{\text{barrier,TF}}(\phi)$ shows that with small $\phi$, the $\Delta F_{\text{barrier,TF}}$ is higher in the attractive system, while when $\phi$ gets larger, the $\phi$ gets
smaller in the end, compared with pure repulsion system, which may cause significantly different result compared with the pure repulsion system and the TF-C attracting system.

The results, either from TF-C interacting system or from TF-C and C-C interacting system, the binding free energy induced by crowders behaves differently in the pure repulsion system. This indicates that the specific protein-protein interaction plays an important role in controlling the genome activities. To get the final reaction rate $k^{TS}(\phi)$ and $k(\phi)$, the influence of attractive potential on $D_{TF}(\phi)$ is still needed.

### 3.3 The Crowding Environment Influence on Chromatin Structure

Here, we study the long-range interaction of chromatin under crowding biological environment by using Random loop (RL) model. RL model is the polymer model firstly introduced by Bohn et al [89] and has been shown by Mateos-Langerak et al [84] to have the same plateau relation as in FISH data. Some researchers think that this plateau in the RL model will inevitably lead to a plateau in the contact probability distribution and will be contradict with the Hi-C measurement [85]. However, this idea hasn't been shown through Molecular Dynamic simulation and is challenged by the contradict result from the work of Bohn et al [12, 89] about Diffusion-Driven looping model. We will firstly show the consistent contact probability behavior between Hi-C data and the RL model. To compare with, we also calculate the contact probability of self-avoiding walk (SAW) polymers. The same -1 power law relation is not observed in the SAW model. Then we will show how the contact probability changes is regulated by the crowding biological environment with RL model.
Since the RL model used in this study gives the unified agreement among both the FISH data [84] and Hi-C data [5], we believe that the long-range interaction behavior studied by it can give us reliable prediction of the influence from crowding agency. The results shown in this study suggest a non-monotonic behavior of the slope of the power law relation in contact probability under different crowding volume fraction. This non-monotonic effect is consistent with the chromosome structural changes induced by crowding studied previous by Kim et al [13]. This consistence suggests an universal effect of crowding on chromatin structure and the non-monotonic changing of the long-range contact probability indicates that the cell-specific enhancer-promoter interaction in SHH might be regulated by the crowding volume fraction change during carcinogenesis.

The question we are interested to answer in this part is to improve the MSA where the dynamical chromatin structures are considered to find the relation between the structural alteration of chromatin and carcinogenesis. There are two goals needed to achieve to answer this question: 1) A chromatin model needed to be developed that can reveal the long-range structural properties of chromatin observed in previous experiments; 2) The relation between cancer cell expression and chromatin structure alterations needs to be studies with the experimentally validated chromatin model.

To do so, the RL model is used to model the chromatin, which is previously developed by Kim et al [13]. We will first validate the chromatin model with Hi-C data previously observed by comparing the long range contact probability. Next the the observed increasing crowding density variance in early carcinogenesis [90] will be introduced into the chromatin model to study how the
long-range contact probability of chromatin segments is regulated under different crowding density.

The RL chromatin model consists of two parts. The first part is the polymer backbone that is commonly shared by all the polymer model. The second part, which makes the RL model distinguishing from the traditional polymer model is the fixed link randomly chosen along the polymer.

The polymer backbone in the RL chromatin model used in this study consists 300 monomers. The diameter of each monomer is $5\sigma$, where $\sigma = 6\text{nm}$ is the unit of length in the simulation. The potential between the two monomers next to each other in the polymer backbone is defined as:

$$ U(R) = U_b(R) + U_r(R) $$

$$ U_b(R) = \frac{1}{2} k_b R_b^2 \ln \{1 - [(R - r_0)/R_b]^2 \} $$

$$ U_r(R) = \begin{cases} 
4\epsilon \left\{ \frac{\sigma}{R - r_0} \right\}^{12} - \left\{ \frac{\sigma}{R - r_0} \right\}^{6} + \epsilon, & (r_0 < R < r_0 + r_c) \\
0, & (R \leq r_0 \text{ or } R \geq r_0 + r_c) 
\end{cases} $$

where $U_b(R)$ is the potential connecting the segments together, $U_r(R)$ is the modified repulsive Lennard-Jones potential that accounts for the excluded volume interaction in the polymer, $k_b = 30k_B T/\sigma^2$, $R_b = 1.5\sigma$, $r_0 = 4\sigma$, $R$ is the center to center distance between two segments, $\epsilon = 4k_B T$ is the energy unit, $r_c = 2^{1/6}\sigma$ is the cut-off distance, and $k_B$ is the Boltzmann constant.

To create the fixed loop in the RL model, binding pairs are chosen randomly along the polymer backbone with the uniform looping probability $P_{lp}$. Then the looping pairs are linked together with the same potential as in Eq. (6). In this study, we choose 3 looping probabilities with $P_{lp} = 3 \times$
With these $R_p$, each RL model can have 13, 25 and 37 fixed looping sites and the polymer volume fraction of 0.15, 0.23 and 0.29 separately, which corresponds to the chromatin volume fraction from 0.1 to 0.4 in the cell.

The contact is defined when the center to center distance between two segments are smaller than the threshold $\delta$. This contact site can either be the bond, fixed looping site or the two segments getting close to each other by thermal dynamic movement. $\delta = 5.1\sigma$ is used in the calculation, which makes the center to center distance between two contacting segment is larger than the diameter of the monomer and smaller than $r_0 + r_c$.

The macromolecular crowding effect in the biological environment is included implicitly by introducing the depletion potentials between all the polymer segments. The depletion potentials from excluded volume effect in macromolecular crowding is studied in previous work [14] and modify the free energy between two monomers in the crowding system. It has been shown that this implicit way of considering crowding influence is in a reasonable agreement with doing crowding environment simulation explicitly [14]. The three crowding volume fractions considered in this study 0.1, 0.2 and 0.3, which corresponds to the occupation of macromolecules to be up to 40% in the cell. The detail of the depletion potential for each crowding volume fraction used in this study can be found in [14].

For each $R_p$, the results are averaged over 500 RL model with different looping pairs. For each RL model, 500 independent initial configurations are used to simulate the final output
configurations. In total, there are 150,000 independent final configurations for each $P_{lp}$ in the analysis. The RL simulations are done with GROMACS version 4.0.5 and the SAW simulations are done with LAMMPS.

The distance between each pair of monomers along the polymer model in a crowders free environment is calculated and the contact probability map is calculated and averaged over 150,000 independent configurations. The contact probability distribution $P$ for each $P_{lp}$, which is only the function of the base pair (bp) distance between monomers, is calculated by averaging the result from contact probability map, as shown in Fig. 3. This result reveals two important properties of chromatin in the crowders free situation. First, the power law relation between the bp distance and contact probability $P$ exists in RL model with any $P_{lp}$ and the SAW model for polymer distance in the range of 10Mbp to 100Mbp. In this range, the plateau of $\langle R^2 \rangle$ was also observed in previous works [84]. It is not surprising to see that the slop of the fitted line in power law relation ($k$) is increasing with $P_{lp}$, where SAW model can be considered as the RL model with $P_{lp}=0$, since the chromatin becomes more compact. Second, when $3 \times 10^{-4} \leq P_{lp} \leq 9 \times 10^{-4}$, which correspond to the chromatin volume fraction in the cell, the RL model predicts $k \sim 1$, consisting with the one measured from Hi-C experiments [5]. The power law relation revealed in this result, together with the previous plateau observation, suggests a good agreement between the RL model and the experimental measurements. The consistence between the experiment and RL model and the missing of this consistence in SAW model supports our assumption that the RL model is a good basis to study the chromosome structure in crowded biological environment. It also suggests that the long range interaction in chromatin is driven by chromatin looping.
Fig 3.2 The 2D contact probability map for RL model (the intensity is under $\log_{10}$ scale). (a) is the contact probability map for one set of looping pairs when $P_{lp} = 3 \times 10^{-4}$ in RL, averaged over 300 independent initial configurations. (b) is the contact probability map averaged over all 150,000 independent configurations when $P_{lp} = 3 \times 10^{-4}$. 

[Diagram showing contact probability over distance]
**Fig 3.3** Contact probability for RL model and SAW model in a crowder free environment and the detailed information in the physical range. $k$ is the slope of the fitted line in the physical range for polymer models.

The same contact probability analysis is also performed for the RL model in crowded biological environment, with three different crowding volume fraction $\phi = 10\%$, $20\%$ and $30\%$ separately. The contact probability distributions under different $\phi$ when $P_{lp} = 3 \times 10^{-4}$ in Fig. 4(a) reveals that the slope $k$ is changing non-monotonically with $\phi$ and has the maximum value when $\phi = 20\%$. This non-monotonic dependence indicates that the possibility of the incidence of long-range interaction decreases less rapidly with increasing of crowding volume fraction when $\phi$ is small. At larger $\phi$, the possibility for particles at larger distance to contact decreases faster when the environment become more crowded. The same non-monotonic relation is also observed when $P_{lp}$ increases from $3 \times 10^{-4}$ to $9 \times 10^{-4}$, as shown in Fig. 4(b). The close look at the depletion potential reveals that this non-monotonic relation is the tradeoff between the increasing of the repulsion barrier and the increasing of attraction potential and consists with the previous structural alternation of RL model in crowding environment, where the radius of gyration and the accessible surface area were studied [13]. Thus, we find that the possibility for the chromatin segments at longer distance to contact, relative to the segments at shorter distance, is regulated by the crowding volume fraction.
Chapter 4: Chromatin packing capturing techniques

4.1 Introduction

Cellular phenotype is determined by both their genetic code and the capacity of cells to explore their transcriptional landscape of thousands of genes and create new functional states. Emerging evidence has shown that many human diseases, including cancer, autoimmune and neurological disorders, do not depend on the individual genes, but a complex coordination among ten to thousands of genes in their natural physio-chemical nanoenvironment [91] [3, 11, 92]. While the genetic code has been extensively studied and techniques such as CRISPR-Cas9 has been created to control it, DNA sequences alone do not govern gene expression. In the pursuit of investigating the mechanisms that control the transcription of genes, many studies have provided insights towards regulatory pathways. Regulation of gene expression occurs across a hierarchy of length-scales: (1) at the genetic level through interactions between transcription factors and sequence binding motifs (~ 2 nm); (2) at the nucleosomal level through alterations in local DNA accessibility (~10 nm); and (3) at the level of topologically associated domains (TADs) through gene compartmentalization (~100 nm) [4, 93, 94]. While most studies have focused on molecular modifications that govern the transcription of individual genes, emerging evidence indicates the existence of a yet another level of transcriptional regulation, chromatin packing, that works across all these length scales and affects global patterns in gene expression [3].

Transcriptional interactions are chemical reactions, which depend on the highly dense and complex chromatin nanoenvironment. The nanoenvironment modulates the rate of transcriptional reactions and differentially affects the probability of transcription of thousands of genes simultaneously [3].
In turn, the nanoenvironment depends on chromatin packing, which acts through physical forces and shapes the genomic information space. In particular, it has been shown that the chromatin packing-density heterogeneity at length scales from ~20 nm to 350 nm (kbp to Mbp genomic range) modulates transcriptional diversity (i.e. the dynamic range of gene expression), intercellular transcriptional heterogeneity, and gene network heterogeneity. This chromatin packing code regulates cells’ transcriptional access to their genomic space and is expected to have implications on a wide range of other cellular processes, such as stress adaptation [2, 3, 95].

The characterization of 3D chromatin packing is an active field, and recent evidence suggests that chromatin disordered chains that are packed throughout the nucleus with different densities within TADs in cell ensembles and TADs like nanodomains on single cell level [96]. Independent of the exact configuration of supranucleosomal folding, the 3D chromatin packing can be mathematically described by the mass scaling relation $M(r)$ and the spatial auto-correlation function (ACF). $M(r)$ is the mass of chromatin (M) contained within a sphere of radius r, it dictates how compact chromatin folds from its fundamental unit, the nucleosomes and the nucleotides. The ACF is the quantitative relation between the fundamental unit and larger structures such as TADs like nanodomains and chromosomal territories. Although the exact nature of 3D chromatin architecture is a topic of active debate, most recent evidence with ultrahigh resolution such as ChromSTEM (~3 nm) showed a power-law mass scaling $M \propto r^{D_{MS}}$ up to the Mbp range, suggesting chromatin is a mass fractal with a fractal dimension (scaling exponent) $D_{MS} < 3$. Importantly, the power law mass scaling is commonly found in a variety of polymer systems and is associated with the ACF by $ACF \propto \frac{dM(r)}{dV} \propto r^{D_{ACF}-3}$, where $V$ is the volume. The fractal dimension ($D_{MS}, D_{ACF}$) can be
quantified by ex vivo molecular technique such as small angle neutron scattering and chromatin conformation capture (3C, 5C, and Hi-C), as well as in vitro by nanoscale-sensitive imaging techniques such as electron microscopy (EM) [94], photon localization microscopy (PLM), and partial wave spectroscopic microscopy (PWS) [10, 97, 98]. However, all these independent techniques address a fraction of the length scales of the chromatin packing, and no technologies have been developed that allows for the comprehensive, multi-scale characterization.

Towards this goal, we have developed a new multi-scale multi-technique chromatin nanoimaging platform incorporating Hi-C, ChromEM, and PWS, which can interrogate spatio-temporal changes in chromatin packing, and a computational transcriptomics platform with multi-scale analysis (MSA) to describe gene transcription in the context of realistic chromatin nanoenvironment. Leveraging both imaging and simulation, we can mechanistically interrogate the role of chromatin packing in the regulation of gene expression. We showed from chromEM and Brownian dynamics simulation that the chromatin is indeed a mass-fractal media and validated the inverse relationship between Hi-C contact probability s and fractal dimension D. Leveraging the key quantity D as a proxy of chromatin packing heterogeneity, we further explored the influences of dexamethasone on global chromatin organization. We observed a significant decrease of D after dexamethasone treatment over multiple cell lines across Hi-C, ChromEM, and PWS, which in turn also cross-validated individual component in the nano-imaging platform. Importantly, we observed a continuous change of D during cell cycles by the live-tracking capability in the platform, which opens door for real-time monitoring of chromatin alterations. With the abilities to characterize 3D chromatin organization with ultra-high resolution in all scales, the multi-technique suite can be
employed to understand the role of chromatin organization in gene transcription, to unveil the mechanism of genetic diseases such as cancer, Parkinson’s, and to assist the development of therapies to treat such diseases.

4.2 Multi-technique chromatin nanoimaging and computational transcriptomics platform

Although the genetic and histone codes can be addressed by molecular assays, decoding chromatin packing can only be achieved through “convergent science” that bridges physics-based modeling of transcriptional reactions, nano imaging, and molecular biology. We have developed a new multi-technique platform (Fig 1) including state-of-the-art nanoimaging capabilities which spatio-temporally interrogates chromatin packing from individual DNA strands (ChromEMT with STEM HAADF adaptation) to chromatin compartment visualization (ChromEM with thin section TEM imaging) to single molecular-localization nanoscopy (STORM) to chromatin packing at 20-300 nm length scales in hundreds of live cells in real time (PWS), and a computational platform, which models gene expression in the realistic chromatin nanoenvironment.
Fig 4.1 Examples of the multi-technique nanoimaging and computational transcriptomics platform. The multi-technique platform consists of three parts: nanoimaging to directly visualize chromatin packing at all length scales (5 nm to 200 µm); multi-scale computational analysis to investigate the underlying physics behind chromatin structure; Hi-C characterization to provide big data for high-statistical power analysis across the entire genome over different cells lines and conditions. Top panel, from left to right: 1. Ultra-high-resolution scanning transmission electron tomography at 3 nm resolution for 3D chromatin structure in mammalian cells. Scale bar: 30 nm. 2. High throughput TEM imaging of 50 nm slice of BJ cell nucleus. Scale bar: 1 µm. 3. PWS (red) and super-resolution STORM (green) co-registered imaging of chromatin scaling and Pol II. Scale bar: 3 µm. 4. Label-free PWS image of live HeLa cells. Pseudo-color: D_{PWS} with sensitivity from 20 to 350 nm. Scale bar: 20 µm. Bottom panel, from left to right: 1. Atomic scale MD simulation
of a nucleosome. 2. Brownian Dynamics simulation on an individual chromatin chain. 3. Coarse-grain simulation of macro-molecule crowding. 4. Hi-C contact probability heat map of A549 lung epithelial cancer cell line.

4.3 Chromatin is a mass fractal and the fractal dimension is a transcription regulator

To investigate chromatin packing, direct imaging ChromSTEM HAADF tomography with 2.9 nm voxel resolution and Brownian Dynamics (BD) simulations on individual chromatin chains was performed. From the ChromSTEM HAADF tomography reconstruction, the classical “beads on the string” model with individual nucleosomes and linker DNAs were clearly resolved (Fig 2 (a-c)). Utilizing the HAADF imaging mode with quantitative contrast that is proportional to the mass-thickness of the chromatin, the average ACF of chromatin was calculated directly from all 33 virtual 2D slices to characterize the packing scaling. In Fig 2 (d), the ACF of chromatin was plotted against spatial separation r in log-log scale to visualize the power-law relation between the two. Interestingly, three regions with significantly different power-law scaling $D_{ACF}$ were observed: 1. Nucleosomal region from 0 nm to 11.6 nm. 2. First fractal region from 11.6 nm to 58 nm. 3. Second fractal region from 58 nm to 145 nm. Importantly, the power-law scaling observed in multiple length scales in the chromatin ACF analysis provides a direct experimental evidence to model chromatin as a mass fractal with fractal dimension $D_{MS}$. Using the BD simulations on tension globule chromatin model with different intermonomer interactions, we modeled chains of identical monomers with different mass scaling $D_{MS}$ (Fig 2 (e-f)) and calculated the chromatin contact probability scaling (s) for each chain. From the simulation, we observed an inverse relationship
between s and \( D_{MS} \) (Fig 2 (g)), suggesting that the chromatin packing configurations can regulate gene expression by dictating the contact probability between loci.

**Fig 4.2 Chromatin packing exhibits mass fractal structures in multiple length scales.** (a) A pseudo-2D cross section (2.9 nm in thickness) of A549 cell nucleus chromatin after 3D STEM HAADF tomography reconstruction (contrast inverted). Scale bar: 200 nm. (b-c) The STEM tomography has a nominal resolution of 2.9 nm and is capable of resolving individual nucleosomes and linker DNAs in the pseudo-2D cross sections. Scale bar: 20 nm. (d) Average 2D ACF of 33 chromatin cross sections plotted in log-log scale. Three linear regions were identified on the curve belonging to different topologies in chromatin packing: nucleosome region (0 to 11.6 nm), pure...
mass fractal region (11.6 nm to 58 nm), chromatin interpenetrating region (58 nm to 145 nm). The ACF analysis provided experimental evidence that chromatin is a mass fractal structure in certain length scales. Brownian dynamics simulation on the power law scaling of chromatin packing was performed to investigate the relation between the scaling of contact probability between loci $s$ for tension globule models and the scaling of chromatin packing density $D_{MS}$. Snapshots of tension globule chromatin model with packing density scaling $D_{MS} = 1.58$ (e) and $D = 1.81$ (b) were taken, and an inverse relationship between $s$ and the $D_{MS}$ (g) was observed in the simulation.

4.4 Regulating nanoclusters present with similar chromatin packing scaling on single cell level

We observed the abundance of TAD-like clusters with distinct sizes in chromatin in single cells. Using ChromSTEM HAADF tomography, the 3D chromatin was quantified with 2.9 nm resolution, and the ACF analysis suggested a mass-fractal structure between 11 nm and 58 nm. To capture the fractal dimension within this region, we employed 3D box counting inside a sliding window of 95.7 nm in each side with a stride length of 1 voxel and calculated the slope of the box scaling $D_{3D_{box}}$. **Fig 3 (a)** showed the color coded $D_{3D_{box}}$ mapping of the A549 chromatin, and voxels with similar $D_{3D_{box}}$ values naturally segregated into nanoclusters. The diameter of the nanoclusters was measured using the full width half maximum (FWHM) of the line profile across the cluster. From 30 measurements of different clusters, the average diameter of the nanocluster is calculated to be 90.5 nm, coincide well with the domain boundaries (58 nm in radius) between the first fractal region and the second fractal region observed in the ACF analysis in **Fig 2 (d)**. **Fig 3 (c)** showed the histogram of $D_{3D_{box}}$ for the 1.2 µm by 1.2 µm by 100 nm region. The $D_{3D_{box}}$ ranges from 1.8 to 2.43, and peaks at 2.31.
Notably, we observed clusters with similar chromatin fractal dimension at a larger length scale using PWS, and we investigated their roles in gene transcription via single molecule localization microscopy targeting active mRNA Polymerase II (Pol II). Fig 3 (d) showed results from co-localizing PLM images of Pol II with PWS image of chromatin packing in the nucleus of an ovarian cancer cell (M248). Similar to the D3D_box mapping, we also observed packing domains with similar D values in the PWS map. We segmented the packing domains map by thresholding the PWS D, and quantified the size of over 500 packing domains using the diameter of the minor axis of the domain. The result (Fig 3 (f)) showed a heavy tail distribution of the PWS packing domain size up to 1.5 µm and peaked at around 200 nm. Interestingly, we found that Pol II (green) often localizes in-between the clusters of increased chromatin packing (red), as high density in PWS map means high chromatin packing scaling DPWS (Fig 3 (e)). To quantify the relationship between chromatin packing (DPWS) and transcription activity (Pol II concentration), we divided the nucleus into 520 nm by 520 nm super pixels and calculated the average DPWS and the concentration of Pol II inside the window. As shown in Fig 3 (g) in blue markers, there is a non-monotonic relationship between DPWS and the concentration of Pol II. We speculate that there are two competing forces at play: Increased chromatin packing scaling promotes an increase in transcription by increasing surface accessibility to binding sites [REF]; While at the same time, increased chromatin packing scaling also limits diffusion, hindering the ability of transcription factors and other molecules from reaching binding sites, thus suppressing transcription. CPMC model was performed to predict transcription level at various DPWS (Fig 3 (g) red line) and the model matched perfectly with experiment. Similar trends were observed in colon cancer RKO cell
line as well as the HeLa cervical cancer cell line (SI Fig. XX), indicating the nonmonotonic effect of chromatin packing in regulating gene transcription is a global mechanism.
**Fig 4.3 High resolution imaging platform reveals chromatin nanoclusters on single cell level.**

(a) The fractal dimension from 3D box counting ($D_{3D\_box}$) mapping using the ChromSTEM tomography. The high-resolution 3D mapping showed chromatin with similar $D_{3D\_box}$ values formed local nanoclusters. Scale bar: 200 nm. (b) The histogram of $D_{3D\_box}$ showed that the most probable $D_{3D\_box}$ to be 2.3. (c) The distribution of the diameter of the nanoclusters (quantified by the width half maximum (FWHM)) of the line profile in (a)). The average diameter of the nanoclusters was calculated to be 90.5 nm. (d) A PLM image of a M248 ovarian cancer cell labeled with Pol II (green) overlaid on top of a $D_{PWS}$ map (red). Scale bar: 3 µm. (e) A magnified view of the white square in (a) showed that the Pol II distributed in the periphery of the clusters with similar $D_{PWS}$ values. Scale bar: 500 nm. (f) The size distribution of PWS packing domains. (g) The relationship between $D_{PWS}$ (chromatin packing) and the local concentration of Pol II (gene expression level).

### 4.5 Characterizing dexamethasone-induced chromatin packing alterations via multi-technique platform in A549 cells

The multi-technique platform was employed to capture the chromatin alterations in all length scales induced by Dexamethasone (DXM) treatment in A549 cells (**Fig 4 (a-b)**). In particular, we utilized ChromEM to directly visualize the fine chromatin structure with 5 nm spatial resolution on a 50 nm thick resin section; PWS to examine the mid-range (20 nm to 200 nm) chromatin
packing across hundreds of cells per condition; Hi-C analysis to reveal the chromatin topological domains with size in between kb-Mb (20 nm to 350 nm). Qualitatively, compared to the control group, the cell nucelli in the DXM treatment group showed less variation in the image contrast in the ChromEM micrographs, lower image intensity in the PWS maps, and more homogeneous distribution in the Hi-C heat map. The trend we observed in all three techniques indicated a decrease in the scaling of chromatin packing in A549 cells after DXM treatment. Quantitively, we calculated all three different measures of the chromatin packing scaling: the ACF analysis was performed to obtain $D_{ACF}$ in between 20 nm to 50 nm (Fig 4 (c)), the $D_{PWS}$, and the contact probability analysis was conducted to calculate $s$ (Fig 3 (d)) in the range of 60 kb to 0.3 Mb. As shown in Fig 4 (e-g), in the treated group, we observed a 3.8% decrease in $D_{ACF}$, a 3% decrease in $D_{PWS}$, and a 5% increase in $s$. Note that $s$ is inversely related to the chromatin density mass scaling $D_{MS}$, increment in $s$ indeed indicates decrease in $D_{MS}$. In other words, all three techniques agree with each other and confirmed that there is a significant reduction in the chromatin packing scaling in A549 cells after the DXM treatment. Note that even though chromEM has the highest resolution and is able to directly visualize chromatin structure, it is a low yield method (n ~10) comparing to PWS (n~100) and Hi-C (n~millions). However, in the multi-technique platform, the individual tools cross validate each other, and is capable of rendering conclusions with high statistical power.
Fig 4.4 Measuring chromatin packing scaling alterations induced by dexamethasone in A549 cells. (a) Multi-platform characterization of A549 chromatin with and without DXM treatment. From left to right: TEM images of the chromatin structure with ChromEM staining, scale bar: 1µm. Hi-C contact map of the entire genome. The red square highlighted the regions with contact probability analysis. PWS map of the chromatin packing scaling, the black circle highlighted the nucleus. Qualitatively, from both TEM and PWS, after the DXM treatment, the chromatin packing became more homogeneous. (b) ACF analysis using TEM images of A549 chromatins. The average ACF of the control group (blue) is significantly different from the average ACF of the treated group (red). The shades indicate the standard errors. The chromatin fractal dimension was measured inside the first fractal domain (20 nm to 54 nm) by linear fitting the ACF in log-log scale (dashed lines). (c) Contact probability analysis using the Hi-C contact map between 60 kb and 0.3 Mb bp. The power law scaling s was quantified by linear regression (dotted line). (d–f) Chromatin packing scaling alterations induced by DXM treatment measured using ACF analysis of TEM images, contact probability analysis of Hi-C maps, and PWS. Across the platform, consistent changes were observed in chromatin packing.
4.6 Characterizing dexamethasone-induced time-sensitive chromatin packing alterations via multi-technique platform in BJ cells

Similar to the A549 experiment, the multi-technique platform was employed to capture the chromatin alterations in all length scales induced by DXM treatment in BJ cells (Fig 5 (a-c)). Particularly, we tracked the chromatin changes at two different time points using PWS and Hi-C to investigate the effect of continuous treatment. Likewise, \( D_{ACF} \), \( D_{PWS} \) and \( s \) were calculated to quantify the chromatin packing. Compared to the control group, in the 32 hr treated group, we observed a significant 4% decrease in \( D_{ACF} \), a 4.6 % decrease in \( D_{PWS} \), and a 2% increase in \( s \) (decrease in \( D_{MS} \)). All three aspects of the platform measured the same trend after DXM treatment in BJ cells. We also observed that a less profound reduction in the scaling of chromatin packing (0.6% in \( D_{PWS} \) and 1% in \( s \)) with the intermediate treatment time (16 hr). In addition to measure the chromatin alterations before and after treatment, the multi-technique platform also provides monitoring capabilities to monitor the changing process at different time points.
**Fig 4.5 Measuring chromatin packing scaling alterations induced by dexamethasone in BJ cells.** *(a)* Characterizing the chromatin packing at three different time points (0hr, 16hr, 32hr) after introducing DXM via multi-technique platform. The upper panel showed the Hi-C results and the bottom panel showed the PWS maps. In the PWS maps, the cell nucleus was highlighted by black circle. The PWS signal continued to drop as the we extend the treatment time. The TEM images of the BJ cells in the control group *(b)* also showed more heterogeneous chromatin distribution than the treated group after 32 hours *(c)*. Scale bar: 1µm. *(d)* ACF analysis of the TEM images, the fractal dimension was calculated by linear regression (dashed lines) of the ACF curve in between 20 nm to 54 nm. *(e)* The contact probability analysis of the Hi-C map, the power law scaling $s$ was quantified by linear regression (dashed lines). *(f-h)* Fractal dimension quantified across the multi-technique platform. For ChromEM, only control group and the 32 hr treatment were analyzed. In all modalities, we observed a consistent decreasing of $D$ after the DXM treatment.
4.7 Summary

The multi-scale multi-technique platform bridges the physics-based modeling of transcription reactions, nanoimaging, and molecular biology. The direct visualization of 3D chromatin packing 3D packing at kbp-Mbp scales (20 to 350 nm) in hundreds of live cells revealed diverse intra-cellular chromatin configurations heterogeneity, providing insights into chromatin folding and its role in regulation gene transcription. In this work, we found the chromatin mass-scaling follows a power-law within the domains, indicating that the chromatin folding adopts mass-fractal conformations. The high resolution offered by our platform (ChromSTEM HAADF tomography reconstruction at 2.9 nm resolution) allowed us to identify three chromatin folding patterns: nucleosome (0 to 11.6 nm), single fractal region (11.6 nm to 59 nm), intermixing fractal region (59 nm to 134 nm). The size of fractal domains matches perfectly with published results from other experimental methods. From the BD simulations, we showed an inverse relationship between the contact probability scaling $s$ and the mass scaling $D_{MS}$, revealed the influence of chromatin packing in gene transcription, and established the importance of chromatin packing scaling mapping. From the nanoscale D mapping measure from the multi-technique platform, we discovered the 3D organization of the genome to be a disordered polymer packed into a variety of TADs like domains structures that are challenging to detect by other methods.

In addition to imaging, by combining the high-resolution STORM with live cell PWS in the platform, we are able to associate molecular pathways directly onto chromatin structure. Notably, we observed that the active RNA polymerase (Pol II) has a natural tendency to distribute around these single cell domains in multiple cell lines, and the concentration of the Pol II showed a non-
monotonic relation with D, which matches the molecular dynamics simulation prediction perfectly. As D links the physical and genomic sizes of a chromatin region, D essentially affects the heterogeneity of chromatin packing and its accessibility with the net effect of increasing D being a greater dynamic range of transcription states (transcriptional divergence), intercellular transcriptional heterogeneity, gene network heterogeneity, and the amplification of the rate of change in up/downregulation of gene expression, and a decreasing D lowers the barrier for changes in cellular functional activity.

In the previous work, we found that the increasing in D is one of the earliest events in carcinogenesis, which can be explained by the influence of D in gene expression: cancer cells must keep developing new traits throughout tumor progression, and a large D will benefit the exploring process. One way to reverse the tumor progression is to lower D within cancer-relevant genome domains, and therefore it is crucial to reliably measure the D through drug treatment. We demonstrated the consistency of the multi-technique platform by characterizing the chromatin alterations induced by DXM treatment in one cancer cell line and one normal cell line at multiple time point. Remarkably, we observed agreement among the chromatin packing scaling measured independently across the platform from ChromEM, PWS, and HiC, with emphasis on different chromatin length scales. In addition, PWS Is a noninvasive live cell technique with real-time tracking ability, it gives the platform corresponding time resolution, which is in great compliment with other techniques such as ChromEM and HiC.
Together, our observations of chromatin organization in single cells support the emerging view that the chromatin is a fractal structure with nanometer sizes local domains and the packing scaling serves as an important regulator in gene transcription. The multi-scale multi-technique nanoimaging and computational transcriptomics platform, which provides ultra-high-resolution direct imaging of the chromatin 3D structure as well as real-time live cell chromatin packing for the entire genome over hundreds of cells, can help us better interrogate spatio-temporal changes and mechanistically unveil the underlying physics for chromatin folding and its role in the regulation of gene expression.

4.8 Materials and Methods

Cell culture

M248 cells were cultured in RPMI-1640 Medium (ThermoFisher Scientific, Waltham, MA, #11875127). A549 cells were cultures in Dulbecco's Modified Eagle Medium (ThermoFisher Scientific, Waltham, MA, #11965092). IMR90 and BJ cells were cultured in Minimum Essential Media (ThermoFisher Scientific, Waltham, MA, #11095080). All culture media was supplemented with 10% FBS (ThermoFisher Scientific, Waltham, MA, no. 16000044) and 100 µg/mL Penicillin-Streptomycin (ThermoFisher Scientific, Waltham, MA, # 15140122). All cells were maintained and imaged at physiological conditions (5% CO₂ and 37 °C) for the duration of the experiment. All cell lines were tested for mycoplasma contamination with Hoechst 33342 within the past year. Experiments were performed on cells from passage 5–20.
**ChromEM Sample Prep**

For EM experiment, all the cells were prepared by the ChromEM staining protocol [click EM, Ou,and myself]. The Hank’s balanced salt solution without calcium and magnesium was used to remove the medium in the cell culture. Two-step fixation using EM grade 2.5% glutaraldehyde and 2% paraformaldehyde in 0.1M sodium cacodylate buffer (EMS) was performed: 1. Fixation at room temperature for 10 minutes. 2. Continuous fixation on ice for 1 hour with fresh fixative. The cells were kept cold from this step either on ice or on a cold stage and the solution were chilled prior to use. After fixation, the cells were thoroughly rinsed by 0.1M sodium cacodylate buffer, blocked with potassium cyanide (Sigma Aldrich) blocking buffer for 15 minutes, and stained with DRAQ5™ (Thermo Fisher) with 0.1% saponin (Sigma Aldrich) for 10 minutes. The excessive dye was washed away using blocking buffer. The cells were bathed in 3-3’ diaminobenzidine tetrahydrochloride (DAB) solution (Sigma Aldrich) during photo bleaching.

A Nikon with Cy5 filter sets were employed for photo-bleaching while the cells were kept cold on a custom-made wet chamber with humidity and temperature control. 15 W Xenon lamp and the red filter was used as the source of epi-illumination. With 100x objective, each spot was photo-bleached for 7 min, and fresh DAB was added to the dish for every 30 minutes. After photo-bleaching, the excessive DAB was washed away by 0.1 M sodium cacodylate buffer, and the cells were stained with reduced osmium (2% osmium tetroxide and 1.5% potassium ferrocyanide, EMS) for 30 minutes on ice to further enhance contrast. Following heavy metal staining, the cells were rinsed by DI, serial ethanol dehydrated, and brought back to room temperature in 100% ethanol.
Standard procedure of infiltration and embedding using Durcupan resin (EMS) was performed. The flat embedded cells were cured at 60°C for 48 hours.

Two kinds of sections were made using an ultramicrotome (UC7, Leica). For the tomography, 100 nm thick resin sections were cut and deposited onto a copper slot grid with carbon/formvar film (EMS). For investigating the chromatin structure difference with and without dexamethasone treatment, 50 nm thin sections were made and deposited onto copper 200 mesh grid with carbon/formvar film (EMS). The grids were plasma-cleaned by a plasma cleaner (Easi-Glow, TED PELLA) prior to use. No post staining was performed, but 10 nm colloidal gold were added to the 100 nm thick samples on both sides as fiducial markers for the tomography.

**ChromEM imaging and tomography reconstruction**

A 200 kV STEM (HD2300, HITACHI) was employed for tomography data collection. High angle annular dark field (HAADF) imaging contrast was used in the tilt series. In order to reduce the missing wedge, tilting – 60° to 60° on two perpendicular axis was performed manually, with 2° step size. The pixel dwell time was kept small (~5 µs) to prevent severe beam damage during imaging. For the thin sections, a TEM (HT7700, HITACHI) was operated at 80 kV in bright field to capture high contrast chromatin data. For statistical analysis, 8 cells were imaged for the control group, and 12 cells for the dexamethasone treated group.
For the STEM HAADF tilt series, the images were aligned using IMOD with fiducial markers. 40 iterations of a penalized maximum likelihood (PML) algorithm with non-negativity constraints in TomoPy was employed for tomography reconstruction for each axis. The two reconstructed tomogram sets were re-combined in IMOD to further suppress the artifacts introduced by missing cone. A nominal voxel size of 2.9 nm was used in the tomography to resolve individual nucleosomes.

**PLM sample preparation**

Cells were grown until approximately 70% confluent on 35 mm glass bottom petri dishes. Cells were washed with phosphate-buffered saline (PBS) for 2 minutes then fixed with a solution of 3% Paraformaldehyde and 0.1% Glutaraldehyde in PBS for 10 minutes. Cells were washed for 5 minutes in PBS, then quenched in 0.1% sodium borohydride in PBS for 7 minutes. Cells were washed 3 times in PBS for 5 minutes each, then permeabilized in blocking buffer (0.2% Triton X-100 and 3% Bovine serum albumin (BSA) in PBS) for 20 minutes. The primary antibody (anti-RNA polymerase II, Abcam) was added to the blocking buffer to a concentration 2.5 µg/mL and incubated for 2 hours. Cells were then washed in washing buffer (0.1% Triton X-100 and 0.2% BSA in PBS) 3 times for 5 minutes. Cells were then incubated with the secondary antibody (Alexa Fluor 546, Thermo Fisher Scientific) at a concentration of 2.5 µg/mL in blocking buffer for 40 minutes. Cells were then washed two time in PBS for 5 minutes each. Cells were imaged in standard imaging buffer with an oxygen scavenging system containing 0.5 mg/mL glucose oxidase (Sigma-Aldrich), 40 µg/mL catalase (Roche or Sigma-Aldrich) and 100 mg/mL glucose in TN buffer (50 mM Tris (pH 8.0) and 10 mM NaCl).
PLM imaging

The SMLM optical instrument is built on a commercial inverted microscope base (Eclipse Ti-U with perfect-focus system, Nikon). The microscope is coupled to two imaging modalities. For SMLM imaging, a 532 nm laser (Obis, Coherent) is collimated through a 100X 1.49 NA objective (SR APO TIRF, Nikon) with an average power at the sample of 5 to 10 kW per cm$^3$. Images are collected via a 100X objectives and sent to an EMCCD (iXon Ultra 888, Andor). At least 8000 frames with a 20 msec acquisition time are collected from each sample. For PWS imaging, samples are illuminated with low NA light (0.5) and images are collected using the same 100X objective and sent through a liquid crystal tunable filter (LCTF, CRI VariSpec) and then to an sCMOS camera (ORCA Flash 4.0, Hamamatsu). The LCTF allows for spectrally resolved imaging. Images are collected between 500 nm and 700 nm with 2 nm spacing. PWS and its sensitivity to the nanoscale structure of chromatin packing (D) has been investigated extensively in previous works [REF].

PWS sample preparation

Before imaging, cells were cultured in 35 mm glass bottom petri dishes until approximately 70% confluent. All cells were given at least 24 hours to re-adhere prior to treatment (for treated cells) and imaging. A549 and BJ cells treated with Dexamethasone (Sigma-Aldrich, St. Louis, MO,
D6645) were treated with a dose of 100 nM. IMR90 cells treated with TNFα (Sigma-Aldrich, St. Louis, MO, SRP3177) were treated with a dose of 10 ng/mL.

**PWS imaging**

The PWS optical instrument is built on a commercial inverted microscope (Leica, Buffalo Grove, IL, DMIRB) using a Hamamatsu Image-EM CCD camera C9100-13 coupled to a liquid crystal tunable filter (CRi Woburn, MA, LCTF) to acquire mono-chromatic spectrally resolved Images are collected between 500 nm and 700 nm with 2 nm spacing. Broad band illumination is provided by an Xcite-120 LED Lamp (Excelitas, Waltham, MA). Cells were imaged live and maintained at physiological conditions (5% CO2 and 37 °C) via a stage top incubator (In Vivo Scientific, Salem, SC, Stage Top Systems). D was calculated as per equations and algorithms described in detail in previous works [REF].

**3D Chromatin packing analysis using ChromSTEM tomography**

**Spatial correlation analysis**

The chromatin density fluctuations were calculated from the grayscale 2D virtual slices of the chromatin. Firstly, the mean gray-scale value was subtracted from each image, and the resulting image is the chromatin density fluctuation ($\rho_\Delta$). Secondly, the Btwo-dimensional autocorrelation function (ACF) was calculated using the Wiener- Khinchine relation as:
\[ B_\rho(x, y) = F^{-1}\{|F(\rho_\Delta(x, y)|^2} \]

Where \( F^{-1} \) and \( F \) are the inverse Fourier and the Fourier transforms, and the \( \rho_\Delta \) is the fluctuating part of the chromatin density. To minimize the noise, a rotational average of \( B_\rho(x, y) \) was taken to obtain the final form of the ACF \( B_\rho(r) \), representing the correlation of chromatin density as a function of spatial separation \( r \). Notice that mathematically, a fractal structure can be characterized by a power law ACF, \( B_\rho(r) \sim r^{D_{ACF}-3} \), with \( D \) being the fractal dimension. For the chromatin reconstructed by ChromSTEM, the mean ACF \( B_\rho(r) \) was averaged over the ACFs of each virtual 2D slices and plotted in log-log scale. Linear regions were identified, and linear regression was performed for each region to obtain the slope. The average fractal dimension for the virtual 2D slices \( D_{ACF} \) was calculated by \( 3 + p \) for each region.

**Fractal dimension mapping (\( D_{3D_{\text{box}}} \)) by 3D box counting**

The fractal dimension of a 2D section of the chromatin can be calculated by the ACF method mentioned above. For the 3D chromatin, the fractal dimension can be obtained by using the box counting method. In this method, we first divided the space into grids with unit size \( r \), then count the number of boxes (\( n \)) needed to cover the whole chromatin structure. For a fractal, the number of boxes and the box size again follows a power law relation \( n \sim r^{D_{3D_{\text{box}}}} \), and the power parameter \( D_{3D_{\text{box}}} \) is the fractal dimension. Prior to box counting, the tomograms were thresholded automatically using Li’s method in FIJI to segment the voxels contain chromatin from the background. In order to obtain the fractal dimension \( D_{3D_{\text{box}}} \) within the domain we observed in the ACF analysis, the box counting analysis was performed for binary chromatin mask inside a 3D
moving window with 95.7 nm (33 pixels) on each side, and a stride size of 2.9 nm (1 pixel). The distribution of $D_{3D,box}$ for each moving window was mapped, and nanoclusters with similar fractal dimension were observed. To quantify the size of the nanoclusters, FWHM of the line profile across 29 nanoclusters in the $D_{3D,box}$ map was calculated.

**Chromatin fractal dimension comparison for A549 cells with dexamethasone treatment**

TEM images of 50 nm thin sections were used in the analysis of chromatin packing alterations induced by the dexamethasone treatment for 32 hours. Unlike STEM HAADF imaging mode, the TEM bright field contrast attenuates following the Beer’s law,

$$I(x, y) = I_0 e^{-\sigma \rho(x,y)t}$$

Where $I(x,y)$ is the TEM image intensity distribution, $I_0$ is the incident beam intensity, $\sigma$ is the absorption coefficient, $\rho(x,y)$ is the density distribution, and $t$ is the section thickness. In our experiment, $I_0$, $\sigma$, $t$ were controlled to be constant for all images, only the chromatin density $\rho(x,y)$ contributes to the final image intensity $I(x,y)$. To obtain the density fluctuation, $\rho_{\Delta}(x,y)$, we took the negative logarithm of all the TEM images directly and subtracted the mean value. At the same time, the incident beam intensity $I_0$ is cancelled out. Each nucleus was carefully segmented manually in FIJI, and the fractal dimension $D_{ACF}$ was calculated through the ACF analysis within the nucleus. The average fractal dimension for the control group ($n = 8$) and treated group ($n = 10$) was compared.
Chromatin-Packing Macromolecular-Crowding (CPMC)

As previously introduced in the CPMC model, at certain fractal dimension $D$, the average expression of a group of genes can be approximated as the product of two components, i.e. the probability of the genes to be on the accessible surface $P_s$ and the average mRNA expression rate of these genes $\bar{\epsilon}$:

$$E = P_s \cdot \bar{\epsilon}$$

Based on the power law mass scaling (fractal property) of the chromatin, $P_s$ is determined by the mass of fractal $M_f$, the mass of the elementary particle in the chromatin fractal and the fractal dimension $D$ as:

$$P_s = \left( \frac{M_f}{M_{\text{min}}} \right)^{-1/D}$$

The average expression rate of mRNA can be evaluated using the mRNA expression rate $\epsilon(\phi)$ under certain crowding density $\phi$ from the systems biology method developed for genes under crowded environment by Hiroaki et al. This systems biology method incorporates the simulation results from Brownian Dynamics and Monte Carlo into complex network of reactions involved in transcription to unravel the relationship between the local physical nanoenvironment and gene expression. If the probability distribution function of $\phi$ is $f(\phi)$, the average mRNA expression rate is therefore:

$$\bar{\epsilon} = \int \epsilon(\phi)f(\phi)d\phi$$
Without the loss of generality, by assuming a Gaussian distribution for crowding density, \( f(\phi) \) can be approximated as 
\[
f(\phi) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\phi - \bar{\phi})^2}{2\sigma^2}},
\]
where \( \sigma^2 \) is the variance of crowding density and \( \bar{\phi} \) is the average crowding density of the entire nuclear. The \( \sigma^2 \) is determined by D as: 
\[
\sigma^2 = \bar{\phi}(1 - \bar{\phi})(d_{\text{min}}/L_{\text{in}})^{3-D},
\]
where \( d_{\text{min}} \) is the diameter of the elementary particle in the chromatin (the nucleosome in this case) and \( L_{\text{in}} \) is length of the interaction volume whose nanoenvironment can affect the transcription of a single gene.

**Data analysis for SMLM studies**

SMLM images were reconstructed using the ThunderSTORM plugin for FIJI [REF]. Maps of chromatin packing from raw PWS images were created using a custom analysis script in MATLAB that has been described in detail in previous works. Colocalization between PWS and SMLM images was achieved through alignment of widefield reflectance images collected on each separate imaging arm. Errors caused by sample drift during imaging sequence were first corrected in ThunderSTORM; any additional corrections were applied manually as needed. The average chromatin packing (D) is calculated from the PWS data (in red), and average local Pol II concentration was calculated from SMLM data (in green). Data points with similar D are grouped together (D within 0.025). The circles represent the means, and error bars are standard error between regions.
Chapter 5: Unraveling the chromatin packing scaling code: physical interactions govern global transcription

5.1 Introduction

Fully sequencing the human genome has allowed unprecedented exploration of the roles that genes play in diseases, including neurological and autoimmune disorders, heart disease, and cancer. One of the most significant findings from the human genome project was the fact that, in many cases, these diseases do not depend on the behavior of individual genes, but on the complex interplay between tens to thousands of genes over long periods of time[99-101]. Despite this emerging understanding that many human diseases are thus inherently multifactorial and genomic in nature, no technologies have been developed that allow for the simultaneous, predictable engineering of multidimensional transcriptional interactions.

In this regard, many studies have explored the mechanisms that control the transcription of genes and provided new understanding of the epigenetics that govern diseases. At the level of histones and nucleosomal organization (~10 nm), it is widely recognized that the physical structure of chromatin plays an important role in governing gene expression[102, 103]. In a disease where particular genes or their mutational variants produce a hierarchal, central network topology, this information can be leveraged to control many genes simultaneously by targeting a key node (Figure 1A). These insights have greatly expanded our knowledge of the basis of genetic diseases, but they face the limitation that genetic changes to transcription-factor binding sites, nucleosomal remodeling, or specific gene compartments[104, 105] determine the activity of a few genes only; yet most genetic networks are inherently decentralized or diffuse (Figure 1B). Therefore, methods
to manipulate the expression patterns of many genes simultaneously for non-hierarchal diseases, including most cancers and many diseases of aging (Figure 1C), are largely missing. To address this need, here we present a method to target supra-nucleosomal (>10 nm) chromatin physical structure as a means to predictably modulate global patterns in gene transcription. Starting from first-principle physical considerations and using a combination of Brownian dynamics and Monte Carlo simulations paired with systems modeling, we develop a model that explains the role of supra-nucleosomal chromatin organization on gene expression at the level of physiochemical interactions. Testing the predictions from this model with experimental results obtained from nanoscale measurements of chromatin structure using partial wave spectroscopic (PWS) microscopy, which measures nanoscopic alterations in the scaling of chromatin packing density with a sensitivity to chromatin organization between 20 and 350nm, in live cells and from measurements of gene expression using mRNA microarrays, we show that altering the packing-density scaling of chromatin produces predictable changes in gene expression and that one of the main functional roles of the physical organization of chromatin is controlling the genomic information space as well as intercellular transcriptional heterogeneity. Accordingly, whereas existing understanding of transcriptional regulation has focused on means that regulate individual genes (modifying transcription-factor binding domains, performing nucleosomal post-translational modifications, or localizing genes to chromatin compartments), the approach introduced in this work modulates the physical structure of chromatin for global transcriptional modulation (Figure 2A-D).
Owing to this role of physical structure of chromatin as a major regulator of the genomic information space through shaping the physiochemical nanoenvironment, we then apply the predictions in our model to overcome chemo evasion produced at the level of transcriptional heterogeneity[106, 107]. One of the main challenges in cancer therapy is the innate ability of cells to adaptively sample their genome in order to develop mechanisms of chemotherapeutic evasion[108]. To reverse this process, we employ our model to show that the use of agents that reduce intranuclear variations in chromatin packing density, termed chromatin protective therapies (CPTs), should reduce this transcriptional heterogeneity, and therefore that CPT agents would be effective adjuvant compounds to prevent the emergence of chemo resistance by limiting the replicative adaptability of cancer cells[106]. Using live cell PWS microscopy[6], we test this hypothesis on two potential adjuvants, celecoxib and digoxin, to see if they can act to reduce transcriptional heterogeneity and lead to significant amplification of cancer-cell lethality induced by chemotherapeutic agents. We validate this hypothesis in four models of aggressive gynecological tumors and show that the level of chromatin modulation directly corresponds to the predicted in vitro efficacy.

Hence, we show that global patterns in gene transcription can be controlled by manipulating the physical nanoenvironment within the nucleus. Although we apply the functional consequence of such macrogenomic engineering to the screening of chemotherapeutic adjuvants, the approach paves the way for the study and treatment of diseases — such as Parkinson’s disease, atherosclerosis, and autoimmune disorders — that are governed by the complex interplay of dozens of genes.
5.2 Results

In chromatin, chemical reactions such as gene transcription take place in a highly dense and heterogeneous physical nanoenvironment, the consequence of which is not yet fully understood. Here, we present and use a predictive model to leverage the effects of this physical nanoenvironment on gene transcription in order to control global patterns in gene expression. This model achieves this by considering that these chemical reactions depend on (1) the accessibility of the genome[13], (2) the intrinsic molecular characteristics of the gene (binding affinities, local concentration of reactant species, and reaction rates, among others)[17], and (3) the effects of physical interactions on these chemical reactions[16, 17, 109]. Notably, the supra-nucleosomal physical organization of chromatin and the associated spatial fluctuations of concentration and the overall accessibility of chromatin have a role in determining the behavior of these chemical reactions, as explored below.

To begin, we consider the elements that control the physical nanoenvironment within the eukaryotic nucleus and that influence many genes simultaneously. Chromatin (DNA and DNA-associated proteins, such as histones, polymerases and transcription-factors) is the predominant macromolecular assembly within the nucleus[110]. Thus, the nuclear physical nanoenvironment with respect to multiple genes is shaped by the assembly of chromatin packed into supra-nucleosomal structures (>10 nm) (Figure 2A-D). Recent evidence suggests that chromatin polymer is a disordered chain that is packed together at different densities throughout the nucleus[18]. Independent of the exact configuration and of the kinetics of supra-nucleosomal
folding, 3D chromatin packing density can be described statistically by the scaling of the mass of chromatin \( (M) \) contained within a sphere of radius \( r \), \( M(r) \), and by the associated auto-correlation function (ACF) of chromatin packing density. The ACF is the quantitative relationship between smaller components (e.g. nucleotides/nucleosomes) and larger structures (up to chromosomal territories). Although the exact nature of chromatin organization is a topic of active debate, most of the recent evidence suggests that for supra-nucleosomal length-scales up to the Mbp range, chromatin can be characterized as a mass fractal (power-law scaling) media with a fractal dimension (scaling exponent) \( D < 3 \), a property which is commonly found in a variety of polymer systems[5, 19, 85, 111-117]: \( M \propto r^D \) and \( \text{ACF} \propto \frac{dM(r)}{dV} \propto r^{D-3} \), where \( V \) is volume[118-120]. It will be of consequence to the discussion below that ACF also quantifies the intranuclear heterogeneity of chromatin packing density, and thus the scaling of chromatin packing density and packing-density heterogeneity are inherently linked. Experimentally, the power-law scaling of chromatin packing density was found using both \( \text{ex vivo} \) molecular techniques such as neutron scattering and chromatin conformation capture (3C, 5C, and Hi-C)[4, 96, 121] as well as \( \text{in vitro} \) by nanoscale sensitive imaging techniques such as transmission electron microscopy (TEM), PWS microscopy, fluorescence correlation spectroscopy, and photon localization microscopy (PLM)[9, 10, 63, 122].

By using this evidence, and without loss of generality, we consider the mass and ACF of chromatin structure to follow a power-law scaling form with fractal dimension, \( D \). To extend this work to non-fractal conditions, one should note that \( D \) is proportional to the molecular correlation distance regardless of the exact form of the correlation function (for example, when \( D>3 \))[51, 90, 98]. In
cases with a non-fractal ACF, $D$ instead statistically quantifies the fraction of large versus small length-scale structures in chromatin (a larger $D$ implies a greater fraction of larger length-scales)\cite{51, 123}, and the physical properties of chromatin (including the accessible surface area and the local variations in density, which are two critical quantities that play a major role in global transcriptional regulation) can still be quantified using the ACF. Consequently, in the context of this work, the scaling of chromatin packing density refers to the polymeric properties of chromatin as it relates to the scaling dimension (or scaling exponent), $D$, of its mass, as a function of distance from a reference point. Clearly, multiple chromatin conformations may potentially correspond to the same scaling dimension. However, as it occurs with other polymers, modifying the scaling dimension will produce critical differences in both the accessibility of chromatin regions and the distribution of chromatin packing density.

To understand the link between chromatin packing-density scaling and transcription, we consider the average mRNA transcription rate, $E$, for any given group of genes ($n>1$) as the product of the expectation of the fraction of chromatin that corresponds to the accessible surface ($ASA$) and the average rate of transcription of genes associated with the accessible chromatin surface, $\bar{\epsilon}$,

$$E = ASA \times \bar{\epsilon}(\bar{m}, \phi),$$

where $\epsilon(\bar{m}, \phi)$ is the expression rate of any given gene within the group as a function of its molecular features, $\bar{m}$, for a given local molecular crowding density, $\phi$. $\bar{m}$ accounts for molecular regulators such as the local concentration of transcription-factors, their binding affinities, and the transcription rate of RNA polymerases, which in turn depend on histone regulation, genome
compartmentalization (such as gene positioning with respect to A/B domains and transcription-factories),[5, 124] and other molecular regulatory processes.

We first consider the influence that packing-density scaling has on the accessible surface area of chromatin. In a fractal medium, the relation between ASA and $D$ is analytically quantified following the law of dimensional coadditivity as:

$$\text{(2) } \text{ASA} = \left(\frac{M_f}{M_{\text{min}}}\right)^{-1/D},$$

where $M_{\text{min}}$ is the mass of the basic unit of the chromatin chain (a single nucleotide) and $M_f$ is the total mass of the chromatin domain within which the power-law scaling holds with the number of base pairs $\frac{M_f}{M_{\text{min}}}$. Therefore, in a power-law medium such as chromatin, increases in $D$ directly elevate ASA, and without other considerations, would elevate transcription overall. In principle, $D$ can vary throughout the nucleus and, as a result, ASA could vary from gene to gene. However, global increases or decreases in packing-density scaling would be expected to change the accessibility of genes on average. It is important to note that these conclusions also apply to a medium with a non-fractal ACF, as an increase in the correlation distance (higher $D$) would increase the ASA (see Supplementary Eq. 10-14 for details).

In addition to altering ASA, changes in $D$ also have an effect on $\epsilon$ through the heterogeneity of chromatin packing density. Changes in the local mass density (that is, changes in crowding) will non-monotonically alter expression depending on the molecular features of a given gene. This
complexity and molecular dependence is rooted in the competition between the effect of local crowding on molecular diffusion and the stability of binding reactions[17]. As crowding increases from dilute conditions, the initial effect on chemical reactions is a gain in stabilization of the intermediary complexes, as the entire system gains entropy by decreasing the excluded volume of the reactant species. As the volume fraction continues to increase, the gain in entropy is eventually overcome by the decreased mobility that results from crowding. This produces a non-monotonic relationship between the local density and the rate of transcription that disproportionately influences under-expressed genes (Figure 3A). To quantify this relationship as a function of various molecular features for a homogenous media, we used Brownian Dynamics and Monte Carlo simulations[17]. However, chromatin is an inherently heterogeneous environment, and therefore predicting the effects of changes in density requires information on the distribution of mass density. To estimate these changes, we analytically calculated the local variations in crowding as a function of $D$ on the basis of the properties of a polymeric assembly. Of note, the effects of the spatial distribution of mass density on gene expression are applicable both to genes localized within and outside of higher-order compartments, as crowding has a generalized effect on chemical reactions (Figure 2B&C). Therefore, although compartmentalization has itself been shown to control individual genes by modulating the distribution of cis regulatory factors, in the context of the transcriptional modulation by chromatin packing density, the compartmentalization effects will be similar to those of genomic or histone modifications: they alter the underlying molecular predisposition of individual genes for gene expression, which is further modulated by the local chromatin packing density.
With respect to chemical reactions, it is reasonable to assume that the physical environment within a small ‘interaction volume’ of radius \( L_{\text{in}} \) neighboring any given gene is homogenous with a local crowding density \( \phi \) (Figure 2C&D). Further, outside of this interaction volume, crowding conditions have no effect on transcription reactions of the given gene, as crowding density has a negligible influence on the free energy of reactions[17]. Therefore, the effect of \( D \) on transcriptional reactions is mediated by the variations in mass density (\( \phi \)) throughout the nucleus.

Analytically, the variance of local density within the nucleus, \( \sigma_{\phi_{\text{in}}}^2 \), can be derived as a function of packing-density scaling, \( D \), and is calculated as:

\[
(3) \quad \sigma_{\phi_{\text{in}}}^2 = \bar{\phi}(1 - \bar{\phi}) \left( \frac{r_{\text{min}}}{L_{\text{in}}} \right)^2 \frac{\bar{\phi}}{D},
\]

where \( r_{\text{min}} \) is the size of the unit element (here a single base pair of DNA) and \( \bar{\phi} \) is the average of \( \phi \) across the entire nucleus. As \( D \) increases, so too do the local variations in density. As the location of genes within a group (e.g. genes with similar molecular features) can vary throughout the nucleus, the ensemble expression is given by the expectation of gene expression as a function of the local crowding conditions. Therefore \( \bar{\epsilon} \) can be approximated by:

\[
(4) \quad \bar{\epsilon} \approx \epsilon(\bar{\phi}) \left( 1 + \frac{1}{2} \sigma_{\phi_{\text{in}}}^2 \frac{\partial \epsilon^2(\bar{\phi})}{\partial \phi^2} \left| \frac{\bar{\phi}}{\epsilon(\bar{\phi})} \right| \right),
\]

where \( \epsilon(\bar{\phi}) \) is the rate of gene expression that would be observed in the absence of chromatin packing-density heterogeneity. Using Eq. 4, the effect of \( D \) on \( \bar{\epsilon} \) is analytically computed as a function of the local crowding conditions with \( \epsilon(\bar{\phi}) \) obtained from the simulations (Figure 3A). Since \( \epsilon(\bar{\phi}) \) is non-monotonic and, for the range of \( \bar{\phi} \) typically found in cell nuclei (~30-45%).
\( \varepsilon(\overline{\phi}) \) is close to its maximum, \( \frac{\partial \varepsilon^2(\overline{m},\phi)}{\partial \phi^2} \bigg|_{\overline{\phi}} < 0 \). Therefore, an increase in \( D \) reduces \( \varepsilon \), as genes become exposed to a wider range of local crowding conditions for which \( \varepsilon \) is not at its maximum.

As Figure 3B illustrates, this effect depends on \( \varepsilon(\overline{m}) \): the absolute value of \( \frac{\partial \varepsilon^2(\overline{m},\phi)}{\partial \phi^2} \bigg|_{\overline{\phi}} / \varepsilon(\overline{m},\overline{\phi}) \) decreases as a function of the rate of \( \varepsilon(\overline{m},\overline{\phi}) \) (Supplementary Eq.19). This is because highly expressed genes are already optimized by molecular modifications (such as histone interactions or gene positioning within chromatin domains) and crowding has a lesser effect on the stabilization of their intermediary complexes. Therefore, the net result of an increase in \( D \) is greater suppression of initially under-expressed genes in comparison to over-expressed genes. Integrated with the overall upregulation of gene expression probabilities due to an increase in ASA as a function of \( D \), the resulting behavior of increases in \( D \) on transcription would be asymmetric and monotonic.

What is practically significant is to predict how gene expression for a given functional group of genes that share a common characteristic (such as similar initial expression, length, or other attributes) changes in response to a change in chromatin packing-density scaling as quantified by \( D \). To characterize this relationship, we define the relative change of gene expression as a function of the change in \( D \) (‘gene expression sensitivity’), \( Se = \frac{\partial \ln(E)}{\partial \ln(D)} \). Combining equations (1)–(4), this can be directly quantified as:

\[
Se = \left[ 1 - \frac{1}{g(\overline{\varepsilon},D)} \right] \cdot \left[ D \ln \left( \frac{L_{in}}{r_{min}} \right) + \frac{3-D}{D} \frac{r_{min}}{L_{in}} L^{1/D} \ln(L) \right] + \frac{1}{D} \ln \left( \frac{M_f}{M_{min}} \right).
\]
Here, $L$ is the sequence length of the gene, and function $g$ is used as a change of variables to transform $\epsilon(\vec{m}, \vec{\phi})$ into a function of $\bar{\epsilon}$, and can be computed either numerically from simulations or analytically (Supplementary Eq.15-20).

To validate this model, we tested these predictions against experimental data. Cells were exposed to multiple perturbations, including stimulation with serum, epidermal growth factor (EGF), and phorbol 12-myristate 13-acetate (PMA) (Figure 3C). The changes in the scaling of chromatin packing density due to a perturbation were assessed using PWS microscopy on live cells while the consequential changes in gene transcription were evaluated using microarray mRNA sequencing. Here, PWS microscopy was used to quantify the nanoscale (20-350nm) heterogeneity of chromatin packing density averaged throughout the nucleus. This range of structural length-scales corresponds to genomic distances from the Kbp to the Mbp range. The produced signal, $\Sigma$, is a result of the heterogeneity of the spatial variations in chromatin density within each diffraction-limited voxel, and is proportional to $D$: $\Sigma \propto (D - D_0)$, where $D_0 \approx 1.50$.\cite{125} As shown by Eq. 5, the influence of the change in the packing-density scaling of chromatin on gene expression depends on (1) the average initial (i.e. preceding the perturbation) expression rate $\bar{\epsilon}$ determined by $\bar{m}$, (2) the initial $D$, (3) the upper length-scale of packing-density scaling of chromatin ($\frac{M_f}{M_{\min}}$), (4) gene length $L$, and (5) the size of the interaction volume (see Supplementary Section 1, Macrogenomics model and analysis, for the derivation). These parameters were obtained experimentally and from the simulations: initial gene expression and $D$ for each condition were obtained by microarray and PWS microscopy measurements respectively\cite{6, 20}. $\frac{M_f}{M_{\min}}$ was
assumed to correspond to the average size of a single chromosome, $L$ was calculated as the average gene length for genes in the microarray dataset, and the radius of the interaction volume was determined from the depletion distance between DNA and transcription-factors from simulations (Supplementary Eq. 7)[14, 17].

Critically, the theoretical prediction of the model strongly matches the experimentally observed changes in expression (Figure 3D). Thus, the combined effect of increases in $D$ on transcription is to warp the genomic landscape toward a state where over-expressed genes are further upregulated and under-expressed genes are suppressed, which can be referred as ‘transcriptional divergence’. Furthermore, over 90% of the change in gene expression is captured by the model as the number of genes per group increases (>50 genes), indicating that the scaling of chromatin packing density plays the predominant role in the collective response of tens to hundreds of genes (Figure 3E). In relation to the diffuse network hierarchy that is present for most genes, this provides a key feature of macrogenomic engineering via the modulation of the physical structure of chromatin because it allows predictable regulation of gene expression for multiple genes simultaneously. As uncovered by the model, whereas individual genes still retain the capacity to respond to specific stimuli, the collective global behavior of genes is dominated by alterations to the physical nanoenvironment.

To leverage the predictive capabilities of our model on the expression patterns of hundreds of genes, we next explored if controlling chromatin packing-density scaling could be used to
modulate the available genomic information space. In brief, this genomic information space is the cumulative functional capacities present within an individual cell (intra-network heterogeneity or transcriptional divergence) or a population of cells (intercellular transcriptional heterogeneity). Analytically, our model shows that the variations in chromatin packing density determine intercellular transcriptional heterogeneity, $H$, which is defined as the standard deviation of the transcription rate of the same gene across a cell population (see Supplementary Section 1.3 for detail):

$$(6) \quad H(D) \approx \frac{1}{\sqrt{2}} ASA\left| \frac{\partial \epsilon^2(m,\phi)}{\partial \phi^2} \right| \sigma_{\phi_{in}}^2.$$  

Since $ASA$ and $\sigma_{\phi_{in}}^2$ increase monotonically with $D$, $H$ also increases with $D$ (Supplementary Figure 4). The relative intercellular heterogeneity between two cell populations with different chromatin packing-density scaling properties, $D_1$ and $D_2$, is therefore $\frac{H(D_2)}{H(D_1)} \approx \left( \frac{M_f}{M_{\text{min}}} \right)^{-\frac{1}{D_2}+\frac{1}{D_1}} \left( \frac{L_{\text{min}}}{L_{\text{in}}} \right)^{-D_2+D_1}$ (Supplementary Eq. 26). This indicates that chromatin packing-density scaling is directly coupled to intercellular transcriptional heterogeneity and to the divergence in gene expression for critical processes, including metabolic regulation, proliferation, apoptosis, and differentiation, which is in agreement with the microarray data (Figures 3F and G). Finally, the coefficient of variation of intercellular transcriptional heterogeneity $CV_E \approx \sqrt{2} \left( \frac{1}{\delta(\epsilon_s D)} - 1 \right)$ also increases with $D$ (Supplementary Eq. 27-28 and Supplementary Figure 5). In sum, elevation of $D$ augments both intracellular transcriptional divergence and intercellular transcriptional heterogeneity.
In this context, carcinogenesis may present a test bed for macrogenomic regulation. Indeed, increased chromatin packing-density fluctuations are a near universal hallmark of early oncogenesis. The model described above would suggest that chromatin packing-density heterogeneity could facilitate tumorigenesis by expanding the genomic information space available to neoplastic cells to stabilize otherwise deleterious states[86, 126-135]. By extension, this principle would also be expected to apply to the cellular response to cytotoxic chemotherapeutic stress, because increasing intra-network and intercellular transcriptional heterogeneity for functional processes (such as stress response, proliferation, and cell-cycle maintenance) would allow cells to stochastically develop chemotherapeutic resistance in real time[106].

In brief, cytotoxic stressors can be overcome by numerous means and can depend on the capacity of cells to sample their genome to acclimate to the stressful environmental conditions. Since apoptotic and non-apoptotic decisions occur over a concomitant but separated timescale that spans several hours[136], decreasing the accessible genomic information space is expected to shift behavior toward the initial stress-response activity, that is, apoptosis. Conversely, rapidly increasing the information space provides a means for cells to arrive at one of a number of successful evasive mechanisms. Experimentally, intercellular transcriptional heterogeneity has so far been observed as a critical determinant of chemoevasion without a clear mechanistic basis[137]. However, as demonstrated by our model, increased scaling of chromatin packing density and the resulting density fluctuations can produce the observed transcriptional heterogeneity. In principle, this allows for a direct application of macrogenomic engineering to develop adjuvant agents as
CPTs that would decrease the scaling of chromatin packing density and would enhance the efficacy of cytotoxic chemotherapy by limiting information sampling encoded in the genome and reducing the fitness of cells during their response to cytotoxic chemotherapy.

To explore this hypothesis and the applicability of macrogenomic engineering for selecting CPT adjuvant agents, we tested if cytotoxic chemotherapeutic intervention did indeed produce increased chromatin fluctuations coupled to an expanded genomic information space (that is, increased intra-network and intercellular heterogeneity). In particular, we tested the effect of cytotoxic treatment on chromatin organization and transcriptional heterogeneity in five cell-line models of three gynecological tumors by using a wide range of compounds: microtubule depolymerization inhibitors (paclitaxel or docetaxel), DNA intercalating agents (oxaliplatin), and nucleoside analogs (5-fluorouracil or gemcitabine). The three tumors were chosen based on their clinical aggressiveness: uterine leiomyosarcoma (MES-SA and mitoxantrone resistant MES-SA.MX2), ovarian carcinoma (A2780 and TP53 mutant clone A2780.M248), and triple negative breast cancer (MDA-MB-231). As expected, cytotoxic intervention increases the heterogeneity of chromatin density $\Sigma$ — and thus $D$ — within 48 hours independent of the cell-line model or of the mechanism of the chemotherapeutic agent (Figure 4A-D and Supplementary Figures 1&2).

By taking advantage of the knowledge that chromatin $D$ is indeed predictably and directly coupled to the chemotherapeutic response, we explored whether CPT compounds that can rapidly (<30 min) reduce chromatin packing-density fluctuations would act as adjuvant agents for
chemotherapeutic efficacy. This short time point was chosen to avoid potential confounding from protein translation on chromatin structure. To test the macrogenomic engineering CPT approach, we selected two compounds that act on biological processes that were transformed by paclitaxel treatment: celecoxib (stress response) and digoxin (ion homeostasis). Utilizing live cell PWS microscopy, we measured the transformation in chromatin-scaling within 30 minutes for A2780, A2780.M248 (M248), MES-SA, and MES-SA.MX2 (MX2) cells treated with either digoxin or celecoxib. We focused on uterine leiomyosarcoma and ovarian carcinoma, as we had both a resistant and sensitive subclone for each model. Notably, each cell-type’s response to these compounds varied, but a substantive response was identifiable (Figure 5A-D). In view of the differential response between CPT and chemotherapeutic agents at the level of variations in chromatin packing-density scaling, we next tested if these observations would extend into transcriptional heterogeneity.

Critically, analysis of single-cell RNA-seq data of MDA-MB-231 cells treated with paclitaxel in comparison to control cells[108] shows the expected shift towards increased intercellular transcriptional heterogeneity, as well as towards intra-network transcriptional heterogeneity, owing to chemotherapeutic intervention. The shift affects numerous biological processes, as it includes genes involved in proliferation, apoptosis, oxidation/reduction, ion transport, and nucleosome assembly (Figure 6A-C). Furthermore, analysis of RNA-seq data of digoxin-treated cells shows that decreases in the chromatin packing-density fluctuations through CPT agents likewise correlated with decreases in intercellular and intra-network transcriptional heterogeneity (Figure 6A-C). This differential response between chemotherapeutic agents (taxols) increasing $D$
and CPT agents (digoxin) decreasing $D$ (Figure 6A) paired with the expected changes to gene expression (Figure 6B&C) supports our finding that controlling chromatin packing-density scaling can be used to modulate the genomic information space. By extension, we hypothesize that if the genomic information space is critical for chemotherapeutic evasion, we would observe that differential changes in chromatin packing-density scaling would extend to the adjuvant efficacy in vitro.

This is indeed the case. Under normal growth conditions, untreated ovarian A2780 cells rapidly grew into colonies and covered over 90% of the imaging field (Figure 7A). As expected, 48 hour mono-treatment with IC-50 concentration of paclitaxel resulted in cellular coverage of ~50% of the imaging field over the same growth period as the controls (Figure 7B). Combination treatment of paclitaxel with a CPT agent (celecoxib) greatly enhanced the efficacy of chemotherapeutic intervention, with clearance approaching 100% (Figure 7C), even though CPT agents on their own did not induce apoptosis (Supplementary Figure 3). Furthermore, this effect extends across all the investigated cell lines, showing an increased efficacy even in models with intrinsic resistance such as the M248 and MX2 models (Figure 7D). These effects were model- and adjuvant-independent, with the effective clearance centering on the total modification to chromatin packing-density scaling. Importantly, the magnitude of the decrease in chromatin packing-density heterogeneity by the CPT agents as measured by PWS microscopy strongly correlates with their increased efficacy to induce cancer-cell death when administered in combination with the chemotherapeutic compounds (Figure 7E), resulting in highly linear behavior that directly
matches the level of decrease in chromatin packing-density scaling with the level of cellular inhibition ($R^2>0.99$, Figure 7E).

5.3 Summary

Supra-nucleosomal chromatin can have profound effects on gene expression by acting on accessibility, mobility, and the binding affinities between reactant molecules. Previous investigations of the interaction between molecular behavior and physical organization have focused on the regulators of local compaction/decompaction in the context of the expression of individual genes[102, 103]. Whereas previous work has demonstrated the role of genetic modifications, histone post-translational modifications and genomic compartments in the regulation of the expression of individual genes, the role of the physical environment within the nucleus on the broad regulation of gene expression has not been previously explored. Here, we have described a physiochemical framework that maps the collective behavior of multiple genes simultaneously on the basis of chromatin’s physical nanoenvironment (Figure 2A-D). These capabilities are derived from experimental evidence indicating that chromatin is the dominant crowder within the nucleus. In this context, modulating the packing-density scaling of chromatin ($D$) is one mechanism to shape the nuclear physical nanoenvironment and alter global patterns in gene expression. In particular, we have demonstrated that macrogenomic engineering can control the transcriptional activity of many genes simultaneously and can be applied to the selection of adjuvant compounds to increase the efficacy of chemotherapeutic agents in vitro. Physiochemical modulation of the chromatin nanoenvironment influences patterns in gene expression owing to the sensitivity of genes to changes in the local physical conditions.
Although previous work investigating supra-nucleosomal organization has shown that gene expression depends on gene localization into a compartment or outside of it, we have shown that both genes within compartments and those outside of them respond to the physical forces produced by the physical nanoenvironment (Figure 2B&C). We integrated Brownian dynamics and Monte Carlo simulations of the chemical reactions governing transcription with analytical predictions of the change in global accessible surface area and of the variations in local density of chromatin packing. The combined model allows the analytical prediction of transcriptional consequences of changes in the power-law scaling of chromatin packing density. Although we consider chromatin as a power-law media because of recent experimental evidence, our predictions can be extended for any known auto-correlation function describing the structural relationship between smaller and larger structures within the nucleus. Even in non-fractal conditions, both the ASA and variations in density monotonically increase as a function of $D$ (Supplementary Figure 4). Critically, the results from the model are in strong agreement with experimental results obtained through a combination of microarray measurements of gene expression and of live-cell PWS microscopy of the cell’s physical structure. The model appears to be best equipped to explain collective patterns in gene expression, and ultimately becomes the dominant predictor of expression patterns for larger groups of genes (>50 Figure 3E). At the level of transcription, the observed collective behavior is anisotropic and monotonic, with highly expressed genes benefiting from increased variations in chromatin packing density and under-expressed genes responding conversely.
A major functional consequence of this asymmetric response is the transformation of the genomic information space, as the level of intercellular transcriptional heterogeneity (Figure 3F), genomic divergence (Figure 3D&G), and intra-network transcriptional heterogeneity (Figure 3G) relate directly to chromatin packing-density scaling (Figure 3C&F). As this intercellular transcriptional heterogeneity is a major factor in chemotherapeutic resistance, we predicted that (1) cytotoxic chemotherapeutic intervention would produce increased variations in chromatin packing density and (2) agents could be predictably selected as adjuvants based on their capacity to reverse this effect. As predicted, treatment with cytotoxic chemotherapeutic compounds selected for cells with increased chromatin packing-density heterogeneity independent of the cell line model (Ovarian, Breast, and Sarcoma) and of the mechanism of the chemotherapeutic agent (DNA intercalators, microtubule assembly inhibitors, and DNA analogs). Furthermore, the transformation of chromatin towards increased packing-density fluctuations corresponded with increased intra-network and intercellular transcriptional heterogeneity as demonstrated by single-cell RNA sequencing. As these findings are in strong agreement with our model predictions, we therefore hypothesized that compounds that could reverse this process (decrease chromatin packing-density heterogeneity) at short time scales (<30min) would increase the efficacy of existing chemotherapies.

We tested this hypothesis by examining the effects of two compounds that act on processes that demonstrated increased intercellular and intra-network transcriptional heterogeneity during paclitaxel treatment: stress response (celecoxib) and ion homeostasis (digoxin). Both of these compounds have some anti-neoplastic inhibitory effects; however, our results suggest that they
also modulate chromatin packing density (Figure 5). Indeed, we observed that the efficacy of these agents as adjuvants depends in large part on their capability to modulate chromatin packing-density heterogeneity (Figure 7D&E). For example, the ovarian carcinoma cells, A2780 and M248, had a robust decrease in the intranuclear variations in chromatin packing density for both digoxin and celecoxib, and showed a marked enhancement in clearance for both adjuvants. Critically, the adjuvant efficacy is directly linked ($R^2>0.99$) to the effect on chromatin, with the level of decrease in the variations of chromatin packing density linearly matching cellular death. Although these results are strongly in agreement with the model, it is impossible to rule out the presence of secondary mechanisms that could produce the observed adjuvant efficacy. However, the robust agreement between the observed changes in gene expression and the predictions of our model (Figures 3), the effect of chemotherapeutics on transcriptional and chromatin packing-density heterogeneity (Figures 4), the effects of CPT agents on decreasing both transcriptional and chromatin packing-density heterogeneities (Figures 5&6), and a robust agreement between the ability of CPT agents to reduce chromatin packing-density heterogeneity and the synergistic lethality imparted by these agents when administered in combination with cytotoxic chemotherapy (Figure 7) support the overall potential of macrogenomic engineering for modulating chromatin packing density.
**Fig 5.1 Genomic networks are highly interconnected and decentralized.** A) Classically, the role of critical genes such as MYC, BRCA1, and YAP has been viewed in the context of a hub-spoke model, where these genes form the critical link between the elements in the system. B) However, evidence has shown that the full mapping of the interactions that occur for all genes within a given interaction network shows a diffuse plurality of connections and broad network redundancy. C) Mathematically, the divergence in these models can be represented by the number of connections each gene shares. In the classical hub-spoke system, most genes are anchored only by the central elements (such as BRCA1, c-MYC and YAP). In most genetic networks, however, this is a major oversimplification. Indeed, most genes share direct interactions with at least 5 other genes within the network necessitating a strategy to target the overall regulators of gene transcription.
Fig 5.2 Genomic interactions depend on a complex physical nanoenvironment. A) One universally shared feature of all genes is the physical nanoenvironment that is determined by the supra-nucleosomal (>10nm) packing density of chromatin within the nucleus. B) While previous work has shown that localizing genes into or out of compartments will influence their expression, both genes within compartments (Genes A-C) and outside of compartments (Genes X-Z) will respond to the physical forces produced by their differential packing density. C) As a consequence, while genes are regulated by distinct molecular characteristics (transcription-factor binding affinity, compartment concentrations of factors, nucleosomal modifications) that predispose them toward a preferred expression state (overexpressed, intermediate, underexpressed) the
transcription of these genes into mRNA will also depend on local physical forces. Thus, regardless of the determinant of expression, overexpressed genes (A, X) will differentially respond to local physical organization produced by chromatin packing when compared to intermediately expressed (B, Y) or under expressed (C, Z) genes. To integrate these effects, we consider the power-law scaling of chromatin packing density through fractal dimension, $D$. Increased $D$ produces increased variations in chromatin packing density whereas decreased $D$ does the opposite. D) Ultimately, the physical geometry of chromatin (scaling) determines accessible surface area as well as local crowding conditions that will influence the chemical reactions governing transcription by altering gene accessibility, molecular mobility of reactant species, and the free energy of the transcriptional reactions.
Fig 5.3 Control of higher-order chromatin packing density scaling allows manipulation of genomic information space. A) Local macromolecular crowding density ($\phi$) non-monotonically regulates gene expression. The rate of expression ($\epsilon(m, \phi)$) relative to that for the average crowding that would be observed in the absence of chromatin packing-density heterogeneity ($\epsilon(m, \bar{\phi})$ with $\bar{\phi}=40\%$) is a non-monotonic function of ($\phi$) and also depends on $\epsilon(m, \bar{\phi})$. In turn, $\epsilon(m, \bar{\phi})$ is determined by molecular factors $m$ including transcription-factor concentrations, binding affinities, and the rate of transcription among others. Expression of: suppressed genes is 0.01 fold of the average and enhanced genes is 10 fold the average. B) The result of this non-monotonic relationship between macromolecular crowding and gene expression is an anisotropic response of the rate of expression to changes in crowding ($\frac{\partial^2 \epsilon(m, \phi)}{\partial \phi^2} / \epsilon(m, \bar{\phi})$) as a function of the rate of expression $\epsilon(m, \bar{\phi})$ where $\bar{\epsilon}$ is the average rate of expression. C) Differential PWS microscopy of the variations in chromatin packing density and RNA microarray experiment to measure the relation between changes in chromatin packing-density scaling and transformation of global gene expression using stimulation with serum, epidermal growth factor (EGF), or phorbal 12-myristate-13-acetate (PMA). Scale bars are 15μm. Pseudo-color: heterogeneity of chromatin packing density ($\Sigma$). Arrows: cell nuclei. D) Comparison of the analytical macrogenomic model predicting the changes in gene expression in response to changes in chromatin packing-density scaling (fractal dimension) $D$ (blue curve, Eq. 5) with experimental microarray results (purple markers) obtained from (C). Each experimental data point represents the average of 100 genes. $\bar{E}$ is the average expression of all genes. Error bars are the standard errors of $Se$ for microarray data in each subgroup. E) The accuracy of the macrogenomic model (Eq. 5) increases as a function of the number of genes in each group. For gene groups with more than 50 genes, over 90% of the
A major functional role of the regulation of chromatin packing-density scaling is the modulation of the genomic information. Increases in the variations in chromatin packing density are directly linked to increased intercellular transcriptional heterogeneity (F) and transcriptional divergence (G). F) Comparison of the analytical macrogenomic model predicting intercellular transcriptional heterogeneity (H) as a function of D (blue curve, Eq. 6) with experimental microarray results (purple markers). Error bars represent the standard errors of the heterogeneity of 1000 genes for each condition. Genes were selected such that their expression was within 1 standard deviation of gene expression of the mean expression for all conditions. G) Processes where transcriptional divergence occurs include but are not limited to (1) metabolic regulators, (2) proliferation, (3) apoptosis, and (4) developmental regulation.
Fig 5.4 Chemotherapeutic stress increases variations in chromatin packing density. (A) Representative PWS microscopy images of cell nuclei before and 72 (5-FU) or 48 (paclitaxel and oxaliplatin) hours after their exposure to cytotoxic chemotherapy for A2780 and MDA-MB-231 (M231) cells. Scale bars are 15μm. Pseudo-color: heterogeneity of chromatin packing density (Σ). Arrows: cell nuclei. B-D) Treatment of ovarian A2780 cells (p=2.5×10^{-4}, 1.9×10^{-7}, and 2.8×10^{-28}) (B), uterine leiomyosarcoma MES-SA cells (p=2.1×10^{-6} and 1.1×10^{-19}) (C), and triple negative breast cancer MDA-MB-231 cells (p=6.8×10^{-2}, 1.6×10^{-4}, and 5.6×10^{-5}) (D), with cytotoxic chemotherapeutic agents (5-FU, paclitaxel, and oxaliplatin) produces an increase in the intranuclear chromatin packing-density heterogeneity (Σ) independent of the mechanism of cytotoxic action. Significance was determined using Student’s t-test with unpaired, unequal variance on the average nuclear Σ normalized by the average Σ of the accompanying
control group between the conditions. Box represents the 25-75% range of values and whisker represents the 10-90% range around the mean for N=(823 control, 145 5-FU, 132 paclitaxel, and 101 oxaliplatin) A2780 cells, N=(836 control, 102 docetaxel, and 69 gemcitabine) MES-SA cells, and N=(264 control, 81 5-FU, 36 paclitaxel, and 59 oxaliplatin) MDA-MB-231 cells (p-value <0.001 ***, p-value <0.05 *).
Fig 5.5 Chromatin protective agents rapidly decrease the spatial variations in chromatin packing density. A-D) Representative PWS images (left) and quantification (right) of the effects of CPT agents on the variations of chromatin packing density for MES-SA (A), MES-SA.MX2 (MX2) (B), A2780 (C), and A2780.M248 (M248) (D) cells. Notably, variations of chromatin
packing density for each cell line model have a differential response to CPT agents celecoxib (p=3.9×10^{-34}, 1.5×10^{-33}, 1.5×10^{-30}, and 1.3×10^{-3} for MES-SA, MX2, A2780, and M248 cells respectively) and digoxin (p=2.7×10^{-8}, 7.6×10^{-69}, 3.1×10^{-36}, and 6.2×10^{-9} for MES-SA, MX2, A2780, and M248 cells respectively). Significance was determined using Student’s t-test with unpaired, unequal variance on the average nuclear Σ normalized by the average Σ of the accompanying control group between the conditions. Box represents the 25-75% range and whisker represents the 10-90% range of values around the mean for N=(836 control, 275 celecoxib, and 342 digoxin) MES-SA cells, N=(558 control, 216 celecoxib, and 252 digoxin) MX2 cells, N=(823 control, 132 celecoxib, and 130 digoxin) A2780 cells, and N=(525 control, 36 celecoxib, and 91 digoxin) M248 cells. Scale bars are 15μm. Pseudo-color: heterogeneity of chromatin packing density (Σ). Arrows: cell nuclei. (p-value <0.001 ***, p-value <0.05 *).
A) Heterogeneity ($\Delta \Sigma$, %)

B) Chemotherapy (Paclitaxel) vs. CPT (Digoxin)

C) Chemotherapy (Paclitaxel) vs. CPT (Digoxin)
Fig 5.6 Regulation of chromatin packing-density scaling modulates transcriptional heterogeneity. A) Comparison of the alterations in the variations of chromatin packing density due to taxol treatment (paclitaxel or docetaxel) in contrast to CPT agent digoxin for five cell line models (A2780, M248, MDA-MB-231, MES-SA, MX2). Notably, chemotherapeutic intervention produces increased variations in chromatin packing density whereas a CPT agent (digoxin) decreases variations in chromatin packing density. Box represents the 25-75% range and whisker represents the 10-90% range of values around the mean for N=401 taxol treated cells (132 A2780, 25 M248, 102 MES-SA, 106 MX2, and 36 MDA-MB-231) and N=815 digoxin treated cells (130 A2780, 91 M248, 342 MES-SA, and 252 MX2). B&C) As expected, intercellular (B) and intranetwork (C) transcriptional heterogeneity increases in cells treated with the chemotherapy agent and decreases in cells treated with the CPT agent for critical biological processes including (1) cell cycle, (2) apoptosis, (3) proliferation, (4) transcription, (5) signaling, (6) differentiation, (7) glycolysis, (8) translation, (9) ion transport, (10) metabolism, (11) oxidation/reduction, (12) stress response, and (13) nucleosome assembly. Circle size represents the number of each genes belonging to a functional network/process and thickness the number of shared genes. Color intensity represents the % change in transcriptional heterogeneity in paclitaxel treated vs. controls and in digoxin treated cells vs. control (see RNA-Seq transcriptional analysis, Intranetwork transcriptional heterogeneity, and Intercellular transcriptional heterogeneity in the methods section for calculation.)
**Fig 5.7** Rapid modulation of chromatin packing density scaling by CPT agents greatly enhances chemotherapeutic efficacy. A) Representative image of untreated A2780 cells grown for the same duration as (B) cells treated with 5μ paclitaxel and (C) cells co-treated with paclitaxel and celecoxib. Scale bars are 100μm. D) Reduction of the scaling of chromatin packing density by CPT agents is directly linked to chemotherapeutic efficacy independent of cell line model and the primary molecular mechanism of action of the chemotherapy and the CPT compounds. Key: Docetaxel (D), Docetaxel + Digoxin (DD), Docetaxel + Celecoxib (DC), Paclitaxel (P), Paclitaxel + Celecoxib (PC), Paclitaxel + Digoxin (PD). Mean was calculated from N=(45 D, 45 DD, and 45 DC) MES-SA, N=(30 D, 30 DD, and 30 DC) MX2, N=(60 P, 30 PC, and 30 PD) A2780, and N=(60 P, 30 PC, and 30 PD) M248 individual measurements of cell density per low power field.
(410μm²) for each condition normalized by the average cell density per low power field of the accompanying control group. Box represents the 25-75% range and whisker represents the 10-90% range of values around the mean. E) Relative elimination of cancer cells due to the co-treatment with chemotherapy and adjuvant CPT compounds versus the chemotherapy mono-treatment (Relative Inhibition) is strongly correlated to the efficacy of the CPT compounds to reduce chromatin packing-density scaling (Chromatin Modification) (R²>0.99). Relative inhibition was calculated by measuring the effective difference between the two CPT agents when paired with chemotherapy normalized by the therapeutic efficacy of chemotherapy alone (see Viability Analysis in the methods section for details).

5.4 Material and Method

5.4.1.1 Chromatin-Packing Macromolecular-Crowding Model

The average expression rate of a group of genes, \( E \), with similar molecular characteristics, \( \bar{m} \), and gene length, \( L \), can be approximated as:

(1) \[ E = ASA \cdot \bar{\epsilon} \]

where \( ASA \) is the probability of a gene being on an accessible surface and \( \bar{\epsilon} \) is the mRNA rate of expression of the accessible genes.

To integrate these effects into a unified model, the chromatin-packing macromolecular-crowding (CPMC) model, we consider the fact that the fractal nature of chromatin governs both the overall accessible area as well as the distribution of mass density. The morphology of a fractal can be characterized by the fractal dimension \( D \), which is defined by[138]:

\[ D = \frac{\log N}{\log \frac{1}{r}} \]
(2) \( P \propto s^{D-3-d} \)

where \( s \) is the scale, \( P \) is the value of the measured property at scale \( s \), and \( d \) is the topological dimension (\( d=3 \) for mass and \( d=2 \) for surface area). This produces the following relationships for the fractal:

\[
(3) \quad \frac{M_f}{M_{\text{min}}} = \left( \frac{r_f}{r_{\text{min}}} \right)^D \\
(4) \quad \frac{S_f}{S_{\text{min}}} = \left( \frac{r_f}{r_{\text{min}}} \right)^{D-1}
\]

where \( M_f, M_{\text{min}}, S_f, S_{\text{min}}, r_f \) and \( r_{\text{min}} \) are the mass of the fractal, the mass of the elementary structure of the fractal, the accessible surface area of the fractal, the accessible surface area of the elementary structure of the fractal, the radius of fractal, and the radius of the elementary structure in the fractal respectively. In chromatin, \( M_f, S_f, r_f \) are the mass, accessible surface area, and radius of the region of a chromosome for which a power-law scaling holds, and \( M_{\text{min}}, S_{\text{min}}, r_{\text{min}} \) are the mass, accessible surface area, and radius of a single double stranded DNA basepair (\( r_{\text{min}} \sim 1\text{nm} \)). Therefore, the expectation of the fraction of chromatin that corresponds to accessible surface area (ASA) can be approximated as:

\[
(5) \quad ASA = \left( \frac{S_f}{S_{\text{min}}} \right) / \left( \frac{M_f}{M_{\text{min}}} \right) = \left( \frac{M_f}{M_{\text{min}}} \right)^{-1/D}
\]

We note that ASA increases with \( D \) if \( M_f \) is conserved.
The average mRNA expression rate $\bar{\epsilon}$ depends on the mRNA expression rate at constant chromatin mass density and the distribution of DNA and DNA associated proteins (i.e. chromatin), the second of which can be considered as the predominant macromolecular crowder within the nucleus:

\begin{equation}
\bar{\epsilon} = \int \epsilon(m, \phi) f(\phi) d\phi
\end{equation}

where $\epsilon(m, \phi)$ is the rate of expression of genes as a function of the set of molecular regulatory features $m$ and the average crowding density within the transcriptional interaction volume $\phi$, and $f(\phi)$ is the probability distribution function (p.d.f.) of the crowding densities to which the genes within the gene group can be exposed. The transcriptional interaction volume is the space within which transcriptional reactions are influenced by the local macromolecular density. The radius $L_{in}$ of this volume is determined by the distance at which macromolecular crowding no longer influences the binding affinity of polymerases and transcription-factors during chemical reactions. Notably, the size of the interaction volume is also dependent on the size of gene $L$. Considering the fractal nature of chromatin, the relation between $L_{in}$ and $L$ can be approximated as:

\begin{equation}
L_{in} = L_{in}^0 + L^{1/D} r_{min}
\end{equation}

where $L_{in}^0$ is the radius of the interaction volume for a single base pair of a gene and is approximated as 15 nm based on the Monte Carlo simulation of crowding effect[17]. Previously, the study of the effect of macromolecular crowding on transcription showed a non-monotonic dependence of $\epsilon(m, \phi)$ on $\phi$. [17] In relation to chemical reactions (i.e. transcription), macromolecular crowding has a two-fold effect: (1) it decreases the mobility (diffusion) of the reactant species involved in forming transcriptional complexes, and (2) it increases the binding affinity of the reactant species due to excluded volume interactions. The fractal nature of chromatin
determines the local mass density distribution of macromolecular crowders. The primary structures of chromatin are packed into clusters along a hierarchy of length-scales, and these clusters are considered to be the main crowders in the nucleus. The fractal dimension of chromatin determines the size distribution of these crowders, and \( \bar{\epsilon} \) can be approximately evaluated from Eq.6 by expanding \( \epsilon(\vec{m},\phi) \) into the Taylor series and then integrating each of the terms:

\[
(8) \quad \bar{\epsilon} \approx \int \left[ \epsilon(\vec{m},\bar{\phi}) + \frac{\partial \epsilon(\vec{m},\phi)}{\partial \phi} \bigg|_{\bar{\phi}} (\phi - \bar{\phi}) + \frac{1}{2} \frac{\partial^2 \epsilon(\vec{m},\phi)}{\partial \phi^2} \bigg|_{\bar{\phi}} (\phi - \bar{\phi})^2 \right] f(\phi) d\phi
\]

\[
= \epsilon(\vec{m},\bar{\phi}) + \frac{1}{2} \sigma^2_{\phi_{\text{in}}} \right|_{\bar{\phi}}
\]

where \( \bar{\phi} \) is the average crowding density in the nucleus and \( \sigma^2_{\phi_{\text{in}}} \) is the variance of the average intra-interaction volume density across multiple interaction volumes. \( \sigma^2_{\phi_{\text{in}}} \) can be estimated by considering the power-law scaling property of chromatin.

\[
(9) \quad \sigma^2_{\phi_{\text{in}}} = B(\vec{r} = 0) = \int B_c(\vec{r}) ACF_{\text{in}}(\vec{r}) d\vec{r} \approx \bar{\phi}(1 - \bar{\phi}) \left( \frac{r_{\text{min}}}{L_{\text{in}}} \right)^{3-D}
\]

where \( B(\vec{r}) \) is the auto-correlation function of \( \phi \), \( B_c(\vec{r}) \) is the auto-correlation function of crowding density at each point within the nucleus, and \( ACF_{\text{in}}(\vec{r}) \) is the auto-correlation function for the shape function of the interaction volume \( SH(\vec{r}) \). Eq. (9) holds independent of the assumption of \( SH(\vec{r}) \) (e.g. Gaussian function or binary function). The first term in Eq. (8) describes the effect of macromolecular crowding on transcription in the absence of chromatin packing-density variations and the second term in Eq. (8) reveals the impact of the variations of chromatin packing density, which are in turn due to the power-law scaling of chromatin packing density.
Here, we assume that $\bar{\phi}$ is within the physiological range (30-45%). Earlier Monte Carlo and Brownian Dynamics simulations predicted that this range of $\bar{\phi}$ maximizes $\epsilon(\bar{m}, \phi)$. In other words, on average the packing density of chromatin is near the optimal configuration. Since the effects of macromolecular crowding on $\epsilon$ cannot be experimentally obtained in vitro, $\left. \frac{\partial \epsilon^2(\bar{m}, \phi)}{\partial \phi^2} \right|_{\bar{\phi}}$ in Eq.8 was estimated from the Monte Carlo and Brownian Dynamics simulations in silico[17]. Thus, the integration of the fractal nature of chromatin and the macromolecular crowding model gives us the estimation of the expression rate for mRNA.

In order to evaluate the behavior of chromatin in non-fractal conditions, we utilized the Whittle-Matérn (WM) functional family to model the auto-correlation function (ACF) of chromatin. The WM functions are a convenient and versatile choice as they cover a broad range of functional forms, including: fractal ($D < 3$), stretched exponential ($D \in (3,4)$), exponential ($D=4$), and Gaussian ($D \rightarrow \infty$).[51, 123] In order to estimate how the total surface area of chromatin changes with $D$, we modeled the media as a combination of particles with a size distribution that would match the shape of the ACF. The probability density function of the size distribution of these particles can be approximated as:

$$\text{(10)} \quad N(s, D, lc) = \frac{1}{s^2} \int \text{ACF}(s, D, lc) ds,$$

where $s$ is the particle size and the WM ACF is given by

$$\text{(11)} \quad \text{ACF}(r, D, lc) \propto \frac{2^{\frac{5-D}{2}} \Gamma \left( \frac{D-3}{2} \right) \frac{r}{lc}^{\frac{D-3}{2}} \Gamma \left( \frac{D-3}{2} \right) \Gamma \left( \frac{D-3}{2} \right) \frac{r}{lc} \Gamma \left( \frac{D-3}{2} \right)}{\left[ \Gamma \left( \frac{D-3}{2} \right) \right]} \ .$$
where \( l_c \) is the length-scale of the ACF and \( K \) is the Bessel function of the second kind and \( \Gamma() \) is the gamma function. The physical meaning of \( l_c \) depends on \( D \): for the fractal regime \((D<3)\), \( l_c \) is proportional to the upper length-scale of self-similarity, whereas for \( D=4 \), it is the \( 1/e \) correlation distance.

Using this probability density function to characterize the distribution of mass density, the probability of access to the genes and the local variation in mass density can be analytically calculated as a function of the WM ACF as:

\[
\text{(12) } \text{ASA} = A \int_0^{R_{\text{max}}} N(s, D, l_c) \ast s^2 \, ds,
\]

\[
\text{(13) } \sigma_{\phi_{\text{in}}}^2 = \int_0^{R_{\text{max}}} ACF(\vec{r}, D, l_c) \ast ACF_{\text{in}}(\vec{r}) \, d\vec{r},
\]

where \( A=\text{const} \) depends on the shape of the particles and a fixed total mass of chromatin \( M \) and \( R_{\text{max}} \) is the maximum particle size. Owing to the fact that the total mass of a chromatin region can be calculated by

\[
\text{(14) } M = \rho B \int_0^{R_{\text{max}}} N(s, D, l_c) \ast s^3 \, ds \approx \frac{1}{2} \rho B \alpha(D) l_c R_{\text{max}}^2,
\]

where \( \rho \) is the internal density of the particles, constant \( B \) depends on the particle shape, and \( \alpha(D) = \lim_{l_c \to \infty} \left( N(s, D, l_c) \ast s^2 \right) \), we can approximate \( R_{\text{max}} \approx \sqrt{\frac{2 \ast M}{\rho B \alpha(D) \ast l_c}} \). By substituting \( R_{\text{max}} \) into equations (12) and (13), we can estimate how \( \text{ASA} \) and \( \sigma_{\phi_{\text{in}}}^2 \) depend on \( D \). We find that both the accessible surface area and the local variations in mass density increase in non-fractal \((D>3)\) conditions (Supplementary Figure 4).
In summation, the CPMC model predicts the effects of the scaling of chromatin packing density on the rate of gene expression by considering the inseparable tension between increased accessible surface area and variations in local crowding that depend on $D$. On one hand, as accessible surface area increases so too would the number of available binding sites throughout the genome. Thus, increased accessibility has a universally enhancive effect on transcription. The effective contribution from increased variations in local crowding are more nuanced and depend on the interplay between diffusivity and the stabilization of transcriptional complexes. As a function of crowding, transcription is a non-monotonic function that depends intrinsically on the initial expression state of a gene and peaks within the physiological range of 30-45% [17]. This non-monotonic dependence of mRNA synthesis on molecular crowding is due to the ‘competition’ between two key physical effects that influence the probability of expression of a gene: molecular diffusion (e.g. transcriptional factors and other molecules that eventually form a transcription complex), which is suppressed by crowding, and the binding constants of formation of transcription complexes, which is enhanced by crowding. Under-expressed genes are far more sensitive to changes in crowding than over-expressed genes. The formation of the transcription complexes for these genes is the rate limiting step, and the increased binding probability due to crowding increases the probability of mRNA synthesis for these genes significantly more so than for already highly expressed genes with optimized binding probabilities. Consequently, the net effect of an increase in $D$ is the gene expression enhancement for the majority of genes due to the greater surface area of chromatin and the concomitant suppression due to the greater variations in chromatin packing density within the interaction volume with the latter effect being more pronounced for the already under-expressed genes. In other words, this is a further enhancement of active genes and further suppression of partially suppressed genes.
5.4.1.2. Sensitivity

We analyze the sensitivity, \( S_e \), the response of a group of genes with different initial molecular and physical conditions to alterations in chromatin packing-density scaling as:

\[
(15) \quad S_e = \frac{\partial \ln(E)}{\partial \ln(D)}.
\]

Since the average expression rate for genes with similar characteristics is \( E = ASA \cdot \bar{\epsilon} \), \( S_e \) can be evaluated as:

\[
(16) \quad S_e = D \left( \frac{\partial \bar{\epsilon}}{\partial D} + \frac{\partial ASA}{\partial D} \right).
\]

Substituting equations (5) and (8) into equation (16), we obtained:

\[
(17) \quad S_e = \left[ 1 - \frac{1}{g(\bar{\epsilon}, D)} \right] \cdot \left[ D \ln \left( \frac{L_{in}}{r_{min}} \right) + \frac{3 - D}{L_{in}} \frac{r_{min}}{D} L^{1/D} \ln(L) \right] + \frac{1}{D} \ln \left( \frac{M_f}{M_{min}} \right)
\]

where the function \( g(.) \) is the variable transformation function defined as:

\[
(18) \quad \bar{\epsilon}(\bar{m}, \bar{\phi}) \approx \bar{\epsilon} - \frac{1}{2} \sigma_{\bar{\phi}}^2 \left. \frac{\partial \bar{\epsilon}^2(\bar{m}, \bar{\phi})}{\partial \bar{\phi}^2} \right|_{\bar{\phi}} = \bar{\epsilon} / g(\bar{\epsilon}, D).
\]

The Monte Carlo and Brownian Dynamics simulations of the effect of macromolecular crowding on gene transcription show that the relation between \( \bar{\epsilon} \) and \( \left. \frac{\partial \bar{\epsilon}^2(\bar{m}, \bar{\phi})}{\partial \bar{\phi}^2} \right|_{\bar{\phi}} \) can be approximated as (Figure 3B in the main text)[17]:

\[
(19) \quad \left. \frac{\partial \bar{\epsilon}^2(\bar{m}, \bar{\phi})}{\partial \bar{\phi}^2} \right|_{\bar{\phi}} \approx - \sqrt{\frac{\kappa}{\bar{\epsilon}(\bar{m}, \bar{\phi})}}.
\]
Therefore this function $g(\cdot)$ can be analytically approximated as:

$$
(20) \quad g(\bar{\epsilon}, D) = \frac{1}{1 + \frac{1}{\bar{\epsilon}^2}(\sigma_{\phi_{in}}^2)^2 \left(1 + \sqrt{1 + \frac{1}{\bar{\epsilon}^2}(\sigma_{\phi_{in}}^2)^2 \kappa}ight)},
$$

where $\kappa = 22.6$ nM/s, is the critical rate of expression such that for $\bar{\epsilon} < \kappa$ crowding has a significant effect. Because $\kappa$ exceeds the physiological range of the rate of transcription, crowding is expected to have a significant effect in gene transcription.

### 5.4.1.3. Intercellular Heterogeneity

Intercellular transcriptional heterogeneity ($H$) is the standard deviation of the transcription rate of a gene across a population of cells with the same $D$. The cell population-average transcription for a gene can be written as:

$$
(21) \quad E_s = ASA \cdot \bar{\epsilon}_s
$$

where $\bar{\epsilon}_s$ is the population-average transcription rate of the gene that is accessible for transcription,

$$
(22) \quad \bar{\epsilon}_s = \int \epsilon(m, \phi_s)f(\phi_s)d\phi_s
$$

where $\epsilon(m, \phi_s)$ is the expression rate of the gene as a function of the interaction volume-averaged crowding $\phi_s$, and $f(\phi_s)$ is the probability distribution function of $\phi_s$ across different cells within the cell population. The variance of $\epsilon_s$, $\text{Var}_{\epsilon_s}$, is then:

$$
(23) \quad \text{Var}_{\epsilon_s} \approx \frac{1}{4} \epsilon''(\bar{m}, \bar{\phi}_s)^2 \left(E[(\phi_s - \bar{\phi}_s)^4] - E[(\phi_s - \bar{\phi}_s)^2]^2\right)
$$
where $\overline{\phi_s}$ is the mean of $\phi_s$, $E[(\phi_s - \overline{\phi_s})^4]$ is the expectation of $(\phi_s - \overline{\phi_s})^4$, $\frac{\partial \epsilon(m, \phi_s)}{\partial \phi_s}|_{\phi_s} = 0$, and, to simplify notations, $\epsilon''(m, \overline{\phi_s}) \equiv \frac{\partial^2 \epsilon(m, \phi_s)}{\partial \phi_s^2}|_{\phi_s}$. If $f$ is a normal distributed, $E[(\phi_s - \overline{\phi_s})^4] = 3E[(\phi_s - \overline{\phi_s})^2]^2$. Thus:

\[
(24) \text{Var}_{\epsilon_s} \approx \frac{1}{2} \epsilon''(m, \overline{\phi_s})^2 \sigma_{\phi_{in,s}}^4
\]

where $\sigma_{\phi_{in,s}}$ is the standard deviation of $\phi_s$ across the cell population. Therefore, the intercellular transcriptional heterogeneity can then be written as:

\[
(25) H(D) = \text{Var}_{\epsilon_s} \frac{1}{2} ASA * \sigma_{\phi_{in,s}}^2 \frac{1}{\sqrt{2}} |\epsilon''(m, \overline{\phi_s})| \approx \frac{1}{\sqrt{2}} ASA * \sigma_{\phi_{in,s}}^2 |\epsilon''(m, \overline{\phi_s})|,
\]

where $\text{Var}_{\epsilon_s}$ is the variance of $\epsilon_s$.

Since $H$ is difficult to measure experimentally, from a practical standpoint, there are two metrics of interest that can be derived from $H$. The first metric is the relative intercellular transcriptional heterogeneity between two cell states or populations with fractal dimensions $D_1$ and $D_2$, respectively:

\[
(26) \frac{H(D_2)}{H(D_1)} = \frac{\sigma_{\phi_{in,s}}^2(D_2)ASA(D_2)}{\sigma_{\phi_{in,s}}^2(D_1)ASA(D_1)} = \left( \frac{M_f}{M_{\min}} \right)^{-\frac{1}{D_2} + \frac{1}{D_1}} \left( \frac{L_{\min}}{L_{in}} \right)^{-D_2 + D_1}.
\]

The second metric is the coefficient of variation of intercellular transcriptional heterogeneity, $CV_E$. $CV_E$ can be found using the Taylor expansion of $\epsilon_s$:

\[
(27) CV_E = \frac{\text{Var}_{\epsilon_s}^{\frac{1}{2}}}{\epsilon_s} \approx \frac{1}{\sqrt{2}} \frac{\sigma_{\phi_{in,s}}^2|\epsilon''(m, \overline{\phi_s})|}{\epsilon_s(m, \overline{\phi_s})} + \frac{1}{\sqrt{2}} \sigma_{\phi_{in,s}} \frac{e''(m, \overline{\phi_s})}{\phi_{in,s}} \epsilon''(m, \overline{\phi_s}),
\]
and substituting Eq. (18) into (27), $CV_E$ is simplified:

$$
(28) \quad CV_E \approx \sqrt{2} \left( 1 - \frac{1}{g(\bar{\epsilon}, D)} \right) \cdot sgn[\epsilon''(\bar{m}, \bar{\phi}_s)] = \sqrt{2} \left( \frac{1}{g(\bar{\epsilon}, D)} - 1 \right)
$$

where $sgn[\epsilon''(\bar{m}, \bar{\phi}_s)]$ is the sign of $\epsilon''(\bar{m}, \bar{\phi}_s)$ and can be treated as -1. Since $\sigma_{\phi_{in}}^2$ increases with $D$ (Supplementary Figure 4B) and $g(.)$ decreases with $\sigma_{\phi_{in}}^2$ (Eq. 20), the increase of $D$ leads to the increase of $CV_E$. This trend was confirmed experimentally. (Supplementary Figure 5).

5.4.2. Supplementary Figures

SI Fig. 1) Leiomyosarcoma cancer-cells treated with chemotherapy. A) Representative images of chromatin packing-density heterogeneity of the leiomyosarcoma MES-SA and mitoxantrone
resistant derivative MESSA.MX2 (MX2) cells treated with gemcitabine and docetaxel for 48 hours.

**B)** Chromatin heterogeneity is increased in MX2 derivative cells exposed to docetaxel (p=6.6×10^{-12}) or gemcitabine (p=3.6×10^{-13}) for 48 hours. Significance was determined using Student’s t-test with unpaired, unequal variance on the average nuclear Σ normalized by the average Σ of the accompanying control group between the conditions. Box represents the 25-75% range and whisker represents the 10-90% range of values around the mean for N=558 control, 106 docetaxel treated, and 103 gemcitabine treated MX2 cells.

**SI Fig. 2** Ovarian A2780.M248 (M248) cancer-cells treated with chemotherapy. **A)** Chromatin packing-density heterogeneity is increased in M248 derivative cells treated for 72 hours with 5-FU (p=2.2×10^{-3}), or 48 hours with paclitaxel (p=6.6×10^{-7}) or oxaliplatin (p=5.0×10^{-15}). Significance was determined using Student’s t-test with unpaired, unequal variance on the average nuclear Σ normalized by the average Σ of the accompanying control group between the conditions. Box represents the 25-75% range and whisker represents the 10-90% range of values around the mean for N=525 control, 100 5-FU treated, 45 paclitaxel treated, and 85 oxaliplatin treated A2780
cells. B) Representative images of chromatin heterogeneity of the ovarian carcinoma M248 cells treated with 5-FU, paclitaxel, or oxaliplatin for 48 hours.

*SI Fig. 3*) Monotreatment with chromatin protective therapies does not induce Caspase-3/7 activation. (A-C) Representative flow cytometry data of (A) control untreated, (B) 48 hour Digoxin treated, and (C) 48 hour Celecoxib treated A2780 cells. D) Quantification of the
percentage of Caspase (Cas)-3/7 positive cells comparing untreated controls to 48 hour celecoxib or 48 hour digoxin treated cells (p=n.s., n=2). Error bars represent uncertainty based on ±10% change in gating thresholds.

**SI Fig. 4)** *Extension of the model in non-fractal conditions.* A) Fold change in accessible surface area in non-fractal conditions (*D*>3). ASA was scaled relative to the ASA for *D*<3. B) Variance of the interaction volume-averaged chromatin packing density across interaction volumes (σ²_{φ_{in}}) in non-fractal (*D*>3) conditions. σ²_{φ_{in}} is normalized by the variance of chromatin crowding density at each point in the nucleus, σ². Even in non-fractal conditions, both the ASA and the local variations in density increase monotonically as a function of *D*. 
**SI Fig. 5** Standard deviation of intercellular gene expression. Model-predicted standard deviation for intercellular gene expression as a function of $D$ (blue curve) and the standard deviation for intercellular gene expression calculated based on the experimental microarray data (red dots). The result is normalized by the standard deviation of cells with the lowest $D$. Error bars are the standard error of the standard deviation.
Chapter 6: The role of the chromatin nanoenvironment in regulating phenotypic plasticity

6.1 Introduction

Phenotypic plasticity plays a major role in determining cells ability to accommodate environmental perturbations [139, 140]. In particular, we consider how two components of transcriptional malleability phenotypic plasticity assist multicellular systems to respond to stressors: malleability, which herein will be defined as the functional responsiveness of cells toward stable end states, and heterogeneity, which refers to the number of stable states that can be accommodated within a population, assists multicellular systems to respond to exogenous stressors. In the context of human disease, phenotypic plasticity in oncology plays a critical role in patient survival by determining the responsiveness of cancer cells to immunotherapy or chemotherapy and in potentiating metastatic transformation. Indeed, structural, transcriptional, and mutational heterogeneity have been demonstrated to independently predict chemotherapeutic resistance in multiple cancer models. Indeed, the responsiveness of cells to stress at the level of transcriptional malleability of genes, dynamic changes in expression have been shown to determine the survival of cancer cells in response to chemotherapy [3].

It has become increasingly evident that transcription is highly dependent not only on transcription factor behaviors but on the hierarchical organization of the genome [3, 11, 92]. Therefore, within this work, we focus on exploring the regulation of mRNA transcription due to the organization of the genome in the context of controlling phenotypic plasticity. While considerable effort has been put into understanding the mechanisms governing cellular plasticity at the level of gene expression, it still remains unclear how cells simultaneously regulate the expression of a sizable fraction of their genome [141, 142]. Previous work has demonstrated that enhancers, transcription factor
proteins which promiscuously bind to regions of the genome to promote transcription, play a role in mediating global transcriptional patterns [143-145]. Furthermore, localization and accessibility of enhancers to genes appears to be in part mediated by higher order organization of chromatin (supra-nucleosomal) through the formation and dissolution of topological associated domains (TADs) [96]. Beyond the interplay between TADs and enhancers, however, are several mechanisms through which the physical organization of the genome can modulate gene expression and therefore be a key determinant of phenotypic plasticity [3, 94].

Often, the interplay between gene transcription and the physical organization of chromatin is described in the context of understanding the accessibility of genes (e.g. compaction vs. decompaction) [4, 96, 146, 147]. However, the physical conditions within the nucleus, shaped by the organization of chromatin, can also act on transcription by altering enzyme structure, molecular mobility, and the free energy of transcription reactions [3, 17, 148]. To better understand the role of chromatin structure in the regulation of transcription requires modeling these interactions as we describe within this work using the Chromatin Packing – Macromolecular Crowding (CP-MC) model [3]. Briefly, the CP-MC model considers transcriptional regulation occurs across a hierarchy of length-scales: (1) at the genetic level through the interactions between transcription factors and sequence binding motifs (the genomic code, ~1nm), (2) at the nucleosomal level through alterations in DNA binding affinity (the histone code, ~10 nm), (3) at the level of topologically associated domains (TADs) and genomic compartmentalization (the “compartmentalization code”, ~100nm), and explored herein in greater detail, (4) across supra-nucleosomal length-scales due to the effects of the physical nanoenvironment (the “chromatin scaling code”, >10nm) [4, 93, 94].
These various regulatory mechanisms can be each used to adjust gene expression. To begin, we consider transcription in dilute ex vivo experiments as a series of diffusion limited chemical reactions that utilize DNA, transcription factors, and a polymerase to produce mRNA (Fig 1a). The total production of mRNA in these conditions will depend on the concentration of reactants ([C]_tot, Fig 1b), the rate of polymerase elongation (k_m, Fig 1c), and (3) the dissociation rates of transcription factors and polymerase from DNA (K_D, Fig 1d). The CP-MC model takes these ex vivo considerations and models their behavior in the highly crowded, heterogeneous environment of the cell nucleus (Fig 1e). In vitro, these properties constitute well studied molecular regulators of gene expression. For example, post-translational modification of histones alters nucleosomal stability to control the rate of mRNA synthesis by controlling the rate of polymerase elongation (Fig 1c) [144, 145]. Likewise, gene motifs determine polymerase and transcription factor binding affinities resulting in varied dissociation constants of these molecules from target genes (Fig 1d) [4]. At the compartment scale, formation and dissipation of TADs can alter local TF concentrations (Fig 1b) [96, 147]. Utilizing Monte Carlo and Brownian Dynamic simulations of macromolecular crowding, the CP-MC model then integrates these considerations with physical factors that regulate molecular mobility, the free energy of transcription reactions, and gene accessibility: crowding density (Fig 1f), the scaling of chromatin packing (Fig 1g), the upper length scale of self-similarity of chromatin (Fig 1h), and gene length (Fig 1i).

These physical factors within the model were measured experimentally as follows: crowding density was calculated using ChromEM [94], the scaling of chromatin packing by Partial Wave
Spectroscopic (PWS) Microscopy [10, 97, 98], upper length-scale of self-similarity by high throughput chromatin conformation capture, and gene length was obtained from Mathematica GenomeData function. Predictions from the CP-MC model as a function of these physical factors were tested against measured changes in expression obtained by a combination of mRNA microarrays [149, 150], RNA-seq [151, 152], and single cell RNA-seq [153, 154]. A strong agreement between the CP-MC model and the experimental measurements was found for all of these factors, suggesting a critical role for physical mechanisms to regulate gene transcription. The presence of these physical regulators next led us to ask if they play an important role in controlling the phenotypic plasticity of eukaryotic cells. With respect to phenotypic plasticity, the CP-MC model reveals that physical conditions within the nucleus are critical in determining both the level of transcriptional malleability and heterogeneity. In particular, it shows that higher variations in chromatin packing density, a result of increased chromatin scaling, is associated with an increase in intercellular transcriptional heterogeneity […]. Furthermore, the scaling of chromatin packing produces a “tail-wind” effect on cells, increasing transcriptional malleability in response to exogenous stressors such as chemotherapy. Finally, to see if this increased malleability would have an impact on human disease, we utilized gene expression data from The Cancer Genome Atlas (TCGA) [155] and found that transcriptional divergence (a direct product of the tail-wind effect produced by the scaling of chromatin packing) is inversely related to patient survival in advanced (Stage 3 and Stage 4) colorectal, breast, and lung cancers. In sum, the CP-MC model quantitatively describes the role of physical forces on gene expression in vitro, which functionally regulates phenotypic plasticity, and potentially mechanistically describes how structural alterations of chromatin observed in cancer result in poor prognosis.
6.2 Chromatin Packing-Macromolecular Crowding (CP-MC) Model

The CP-MC model combines theoretical modeling of chromatin as a power-law scaling (a mass fractal – which should not be confused with a fractal/equilibrium globule model) medium with mathematical modeling of the chemical reactions governing gene transcription utilizing inputs from Monte Carlo and Brownian Dynamics simulations to calculate the effects of local physical conditions on each reaction. Utilizing this approach, the CP-MC model can calculate the expected ensemble average expression, $E$, of a group of genes sharing certain similar characteristics (e.g. initial expression). $E$ is modeled as the product of the probability of these genes to be on the accessible surface of the chromatin polymer, $p_g$, which governs the probability of a gene interacting with transcriptional components (TFs and Pol-II) \textit{in vitro} [156], and the ensemble average steady state rate of mRNA expression for that group of genes, $\bar{\epsilon}$, which is a function of intrinsic molecular features of transcription reactions (Fig 1b-d) and local crowding conditions (Fig 1f&g) [3, 17, 148, 157]:

$$E = \bar{\epsilon} \cdot p_g$$

In the model, transcription depends on a number of molecular features, $\vec{m}$, that regulate expression: (1) the transcription rate of RNA polymerases, $k_m$, (Fig 1c) (2) the total concentrations of molecular reactant species involved in the reaction, $[C]_{tot}$ (RNA polymerase, transcription factor, DNA binding sites, Fig 1b), and (3) the dissociation constant $K_{D, Pol-II}$ of the specific binding of RNA
polymerase (Fig 1d). Expression of genes as a function of these molecular features is modified by a number of physical regulators determined by the chromatin nanoenvironment: (1) scaling of chromatin packing, $D$ (Fig 1g), (2) the average crowding volume fraction, $\phi_{in,0}$ (Fig 1f), (3) linear size of gene, $L$ (Fig 1i), and (4) the total mass of chromatin at the upper length scale of self-similarity, $M_f$ (Fig 1h). While the average macromolecular crowding density $\phi_{in,0}$ throughout the nucleus is assumed to be constant across a population, local crowding density, $\phi_{in}$, vary for each gene’s interaction volume as a function of $D$. The size of a gene’s interaction volume is determined by the distance at which external macromolecular crowding levels no longer influences the binding affinity of polymerases and transcription factors during chemical reactions. Finally, in this model, $\phi_{in}$ in each interaction volume is assumed to be constant relative to the time-scale of transcription.

Since genes with similar molecular and physical features may be located at different positions throughout the nucleus, they will therefore be exposed to different crowding density $\phi_{in}$, with probability distribution function $f(\phi_{in})$. This gives rise to the form of $\bar{\varepsilon}$ as:

$$\bar{\varepsilon} = \int \varepsilon(m, \phi_{in}) f(\phi_{in}) d\phi_{in}$$

where $\varepsilon$ is the transcription rate of single gene. Assuming a normal distribution of $\phi_{in}$, the variance of $\phi_{in}$, $\sigma_{\phi_{in}}^2$, would be $\sigma_{\phi_{in}}^2 \approx \phi_{in,0}(1 - \phi_{in,0})(r_{min}/r_{in})^{3-D}$, where $r_{min}$ is the radius of the elementary particle in chromatin (i.e. a nucleotide) and $r_{in}$ is the radius of the interaction volume (See SI) […]. The accessibility of any portion of these genes, $p_g$, to transcription factors or
polymerases to allow for facilitated diffusion to the promoter will then depend on the upper length-scale of self-similarity, $M_f/M_{\text{min}}$ and chromatin scaling as [115, 158]:

$$p_g = \left( \frac{M_f}{M_{\text{min}}} \right)^{-1/D}$$

(3)

In relation to prior work on higher order chromatin organization, $M_f$ could extend into millions of basepairs, representing organization ranging between TADs and whole chromosomes [93, 122, 159]. To quantitatively analyze the effect of changing $D$ in vivo on gene expression, we calculated the sensitivity of gene expression as a function of $D$ predicted by the CP-MC model. Sensitivity ($Se$) is the measurement of how a dependent variable (i.e. gene expression) will change as a function of a perturbation to an independent variable (i.e. $D$). In particular, the $Se$ of gene expression for any group of genes to changes in chromatin packing is defined as:

$$Se = \frac{\partial \ln(E)}{\partial \ln(D)} \bigg|_{E=E_i, D=D_i}$$

(4)

where $E_i$ is the initial average expression of the group of genes sharing similar molecular features $\bar{m}$ and gene length $L$ whereas $D_i$ is the initial fractal dimension of the chromatin polymer. A positive $Se$ for a given group of genes indicates that an increase in the scaling of chromatin packing ($D\uparrow$), on average, enhances their expression, and vice versa.
To solve Eq. 4, we utilized the Taylor expansion of $\bar{\varepsilon}$ around $\phi_{in,0}$ which results in $\bar{\varepsilon} \approx \varepsilon(\bar{m}, \phi_{in,0}) + \frac{1}{2} \sigma^2_{\phi_{in}} \frac{\partial \varepsilon^2(\bar{m}, \phi_{in})}{\partial \phi_{in}^2} |_{\phi_{in}=\phi_{in,0}} [...], where $\varepsilon(\bar{m}, \phi_{in})$ is a non-monotonic function of $\phi_{in}$ due to the competition between excluded volume and diffusion effects (Fig. 2b) and $\frac{\partial \varepsilon^2(\bar{m}, \phi_{in})}{\partial \phi_{in}^2} |_{\phi_{in}=\phi_{in,0}}$ is the sensitivity of the gene expression to changes in local density. Integrating Eq. 1, 3 and 5-7 [...], the $Se$ of expression becomes:

\[ Se(\bar{\varepsilon}) = \left[1 - \frac{1}{g(\bar{\varepsilon})}\right] \cdot \left[D_i \ln\left(\frac{r_{in}}{r_{min}}\right) + \frac{3 - D_i r_{min}}{D_i r_{in}} L^{1/D_i} \ln(L)\right] + \frac{1}{D_i} \ln\left(\frac{M_f}{M_{min}}\right) \]

Here, $g(\bar{\varepsilon})$ is used as a change of variables to transform $\varepsilon(\bar{m}, \phi_{in,0})$ into a function of $\bar{\varepsilon}$, by $\varepsilon(\bar{m}, \phi_{in,0}) = \bar{\varepsilon} / g(\bar{\varepsilon})$, with the results of this transformation estimated by Eq. 4 using inputs derived from MC and BD simulations [3].

To test the CP-MC model predictions in vitro, we utilized PWS microscopy (Fig 2a&b) to measure changes in $D$ paired with mRNA microarrays and RNA-Seq to measure gene expression [97, 150,
We utilized publically available DNA sequencing information to obtain gene length, Hi-C to obtain $M_f$ [161], and performed ChromEM to measure $\phi_{in,0}$ (Fig 2c&d) [94]. As predicted by the CP-MC model, experimental measurements of the $Se$ of gene expression to the change of $D$ show a bidirectional, monotonic responsiveness to $D$ as a function of initial expression in HT-29 cells ($\phi_{in,0} \sim 39\%$) (Fig 2e&f). To test the role of initial $D_i$, we performed RNAi knockdown of the chromatin remodeling enzyme, Arid-1a (A-Kd) which produced a lower $D_i$ compared to wild-type (WT) HT-29 cells [150]. With respect to the model, $D_i$ predominantly changes the responsiveness of under-expressed genes, a result found to be in strong agreement with experimental results (Fig 2e&f).

Next, we tested the role of gene length $L$ and $M_f$, which from our model determine the probability of genes being on an exposed surface to allow transcription reactions to occur. From the CP-MC model, shorter under-expressed genes become disproportionately under-expressed as a function of increasing $D$ (Fig 2g). This effect on under-expressed, shorter genes owes to their reduced total surface area which decreases the binding sites for reactants to move to the transcription start site via facilitated diffusion (Fig 1i). In contrast to length, the role of $M_f$ on gene accessibility is more complex as it inversely tied to changes in $D$ (Eq. 3). At $D=2.56$, for example, a gene within a $M_f$ of 50Kbp has a ~1.7 times higher probability of being on an accessible surface area than if it resided in an 2Mbp $M_f$. However, if $D$ increases to 2.66, a gene within the 2Mbp experiences a 23% increase in $p_g$ whereas 50KBp would only experience a 17% increase. Consequently, CP-MC model would predict that (1) genes in 2Mbp $M_f$ would be relatively underexpressed in comparison to those within a 50KBp $M_f$ and (2) genes within the 2Mbp $M_f$ have a higher sensitivity than those
in the 50Kbp $M_f$ to an increase in $D$. To test this experimentally, we utilized Hi-C and RNA-seq data for A549 lung adenocarcinoma cells treated with dexamethasone 100nm (DXM) for 12 hours. The $Se$ of expression for genes as a function of $D$ within small $M_f$ (TAD size of ~50Kbp) were compared to those within large $M_f$ (TAD size of ~2Mbp) (Fig 2h). As predicted from the CP-MC model, _in vitro_ results demonstrate that genes within 2Mbp TADs have a higher positive sensitive to changes in $D$ (Fig 2h) while simultaneously having lower initial expression compared to those within ~50Kbp TADs.

Finally, the effect of $\phi_{in,0}$ was tested for A549 and BJ cells using ChromEM to measure the chromatin macromolecular crowding density. As $\phi_{in,0}$ is the sum of both the chromatin and mobile protein fractions within the nucleus and ChromEM only provides the chromatin contribution, we considered an additional 5-10% contribution from mobile proteins resulting in a $\phi_{in,0}$ of up to 40% in A549 and 36% in BJ cells (Fig 2i&j). In line with the predictions of the CP-MC model, as $\phi_{in,0}$ decreases, the bidirectional behavior of $Se$ attenuates in BJ cells (Fig 2i) but is maintained in the more dense A549 cells (Fig 2j). For full details of PWS microscopy, ChromEM experiments, HiC data analysis and RNA-seq data analysis.

6.3 The scaling of chromatin packing regulates phenotypic plasticity

Owing to the role of physical regulators in shaping gene expression as described above, the CP-MC could provide a mechanistic framework to link physical properties of chromatin with
phenotypic plasticity. In particular, as the scaling of chromatin packing can change on the order of minutes in eukaryotic cells [3, 97, 150], it could act to assist cells to rapidly respond to environmental stressors both by increasing cellular heterogeneity and transcriptional malleability.

We focus here on cancer responsiveness to chemotherapy as a model eukaryotic system to study the integration between physical structure and phenotypic plasticity given (1) the ubiquitous microscopic and nanoscopic structural changes observed at all stages in malignancy, (2) the emergent role of cellular heterogeneity in determining chemotherapeutic responsiveness, and (3) the capability of cancer cells to rapidly alter gene expression to overcome cytotoxic stressors [2, 3, 95].

With respect to transcriptional heterogeneity, the CP-MC model indicates that $D$ directly regulates transcriptional heterogeneity. The variations in $\epsilon$ across any given population, $\text{Var}_\epsilon$, are:

$$\text{Var}_\epsilon \approx \frac{1}{2} \left( \frac{\partial \epsilon^2 (\bar{m}, \phi_{\text{in}})}{\partial \phi_{\text{in}}^2} \right)^2 \sigma_{\phi_{\text{in}}}^4$$

Consequently, intercellular transcriptional heterogeneity would be:

$$H(D) = p_g \cdot \text{Var}_\epsilon^{1/2} \approx \frac{1}{\sqrt{2}} p_g \cdot \sigma_{\phi_{\text{in}}}^2 \left| \frac{\partial \epsilon^2 (\bar{m}, \phi_{\text{in}})}{\partial \phi_{\text{in}}^2} \right|_{\phi_{\text{in}}=\phi_{\text{in},0}}$$

To explore whether this effect translated into increased population diversity in cellular response to chemotherapy due to changes in $D$, we performed concurrent single cell RNA sequencing and live cell PWS microscopy of ovarian adenocarcinoma A2780 cells. A2780 cells were treated with paclitaxel for 16 (16hr pac) or 48 hours (48hr pac), celecoxib for 16 hours, and 16 hours celecoxib
plus paclitaxel treatment (16hr combo). Single cell RNA-seq measurements of these conditions were conducted as previously described and sequenced using Illumina NextSeq500 [...]. The raw reads were aligned, mapped and used to calculated transcripts per million (TPM) under each condition using bowtie2 and RSEM [152]. 8275 genes were identified across all conditions and principle component analysis analysis using t-Distributed Stochastic Neighbor Embedding (t-SNE) mapped each cell population into a 3-dimensional projection. Genomic diversity in each population was quantified as the radius of genomic space by the average radius of the cluster, \( R_c \), where \( R_c = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r_i - r_{mean})^2} \). \( r_i \) is the position of each cell in the t-SNE cluster, \( N \) is the total number of cells in each treatment group, and \( r_{mean} = \frac{1}{N} \sum_{i=1}^{N} r_i \).

Consistent with predictions from the CP-MC model, we found that genomic diversity increases with \( D \) and in response to treatment with paclitaxel by 16 hours (Fig 3a,b&f), remaining comparably elevated at 48 hours (Fig 3a,c&f) in comparison to untreated controls (Fig 3a&f). Notably, after 48 hours of paclitaxel treatment, the population of surviving cells have a higher \( D \) compared to untreated controls. In contrast, cells treated with celecoxib, a nonsteroidal anti-inflammatory agent, had a lower \( D \) and less genomic diversity than controls (Fig 3d&f) at 16 hours. Likewise, cells treated with a combination of celecoxib and paclitaxel at 16 hours had both a lower \( D \) and a restriction in the genomic diversity comparable to celecoxib monotreatment(Fig 3e&f). As we have previously shown that that celecoxib decreases \( D \) within 30 minutes in A2780 cells, this suggests that the effect of lower \( D \) at early timescales predominates in determining genomic diversity (Fig 3f).
Next, we hypothesized that the scaling of chromatin packing also impacts phenotypic plasticity at the level of transcriptional malleability by determining the responsiveness of genes to upregulation by transcription factors through a chromatin “tail-wind” effect. As previous work has demonstrated that chemotherapeutic induction of apoptosis is determined by the rate of change in expression of critical genes (e.g. p53) and not their steady-state levels, mechanisms which accelerate the upregulation of genes would facilitate the development of cellular resilience to stressors. Theoretically, we consider two otherwise identical cells within a population that have a baseline difference in their initial $D$, cell $a$ and cell $b$ with $D_b > D_a$. These cells are then exposed to an exogenous stressor, such as chemotherapy, and a series of stress signaling pathways are activated. At baseline, these stress response genes were not activated (underexpressed) and their survival now depends on the increased expression of these genes within a critical time frame (Fig 4a). Quantitatively, this effect of $D$ reflects the trajectory of gene expression along the sensitivity curves (Fig 4a) and can be calculated as follows:

The expression of mRNA for protein $x$ in cell $a$ is $E_{1,a}$ at initial chromatin packing state $D_a$ and the expression of $x$ becomes $E_{2,a}$ after treatment with a cytotoxic compound (paclitaxel). Likewise, for cell $b$ with $D_b$, the expression of $x$ without treatment becomes $E_{1,b}$ and after treatment becomes $E$. Recalling from the previous definition of sensitivity ($S_e$), $S_e = \frac{\Delta E}{\Delta D} = \frac{\Delta \beta}{\Delta D}$, where $\beta$ is the expression rate and $\Delta \beta$ is the change of expression rate, then the expression of $\beta_{1,b}$ and $\beta_{2,b}$ can be expressed in terms $\beta_{1,a}$ and as $\beta_{2,a}$:
\[
\frac{d\beta}{\beta} = \frac{Se(D)}{D} \, dD
\]

\[
\beta_{1,b} = \beta_{1,a} \exp \left[ \int_{D_a}^{D_b} \frac{Se_1(D')}{D'} \, dD' \right]
\]

\[
\beta_{2,b} = \beta_{2,a} \exp \left[ \int_{D_a}^{D_b} \frac{Se_2(D')}{D'} \, dD' \right]
\]

where \(Se_1(D)\) and \(Se_2(D)\) are the sensitivity at \(\beta_{1,a}\) and \(\beta_{2,a}\) states correspondingly. Notably, if the stimulation of cytotoxic compound change the expression of protein x from initially underexpressed \(\beta_{1,a}\) to overexpressed \(\beta_{2,a}\) relative to the genome population, then along the \(Se(\beta)\) curve, \(\beta_{1,a}\) would be greater than \(\beta_{1,b}\) but \(\beta_{2,a}\) would be less than \(\beta_{2,b}\). The effective contribution of \(D\) on expression is therefore the ratio \(\delta = \left( \frac{\beta_{2,b}}{\beta_{1,b}} \right) / \left( \frac{\beta_{2,a}}{\beta_{1,a}} \right)\), which represents the enhancement of expression due to chromatin.

Experimentally, we have found that celecoxib reduces \(D\) within 30 minutes in A2780 cells by at least 8% compared to controls. Utilizing the CP-MC model with an input of a 8% difference in \(D\) between theoretical cells \(a\) and \(b\), the model predicts up to a 4-fold \(\delta\) for initially underexpressed genes that become activated in response to stimulation (Fig 4b) in the favor of cells with higher initial \(D\) (cell \(b\)). As Paek et. al. demonstrated that the crucial window for response to chemotherapy frequently occurs within 24 hours, we compared the change in expression for A2780
cells mono-treated with paclitaxel to combination treatment with paclitaxel and celecoxib for 16 hours. As expected from the CP-MC model, $\delta$ of suppressed genes for cells with a lower $D$ ($\beta_{1,a}$, celecoxib treated) in response to chemotherapy ($\beta_{2,a}$, 16hr combo) is lower than that found in those with an initially higher $D$ ($\beta_{1,b}$, control) treated with chemotherapy ($\beta_{2,b}$, 16hr pac) (Fig 4c). Likewise, a similar but smaller effect is also observed in initially over-expressed genes (Fig 4c) in strong agreement with the model predictions. Moreover, the relative expression for cells with a higher $D$ ($\beta_{2,b}$, 16hr pac) is higher than that in cells with a lower $D$ ($\beta_{2,a}$, 16hr combo), qualitatively in agreement with the model (Fig S4d).

To explore if these predicted results are influencing the responsiveness of stress response pathways, we utilized RNA-seq of A2780 cells treated with paclitaxel for 48 hours to identify genes that increase their expression by at least 2 fold after paclitaxel treatment as compared with control. Ontology analysis of the biological processes activated was consistent with the expectation that multiple stress response pathways become activated during treatment with paclitaxel including DNA repair, autophagy, cell cycle arrest, and apoptotic response (P-value < 0.05, Fig 4e. Full list of biological processes that are up-regulated can be found in SI). As with the change observed to other up-regulated genes, genes belonging to these stress response pathways shown in Fig 4e are similarly influenced by $D$, showing $\delta$ of ~4 at 16 hours (Fig 4f).

**Transcriptional divergence is inversely associated with patient survival**

As described above, $D$ determines the cellular responsiveness to chemotherapy *in vitro* through a chromatin mediated tailwind effect on gene expression, suggesting that physical factors in cancer
cell nuclei may play a role in tumor responsiveness to chemotherapy *in vivo*. To test if such a relationship exists, we analyzed publically available RNA-seq data collected by the TCGA Research Network [...] for the three most prevalent malignancies in the U.S: lung, colorectal, and breast cancers. As the model predicts cellular responsiveness to chemotherapy, we focus on patients presenting with Stage III and Stage IV tumors as systemic therapy is the standard of care. Using the *R* package, *TCGAbiolinks*, we quantified gene expression in units of fragments per kilobase million (FPKM) for each patient. Since this data lacks initial control measurements of cancer cells prior to initiation of systemic therapy, tailwind analysis of transcriptional malleability cannot be measured directly for each patient. However, the essence of the chromatin tailwind effect on transcriptional malleability is that elevated $D$ maximizes a gene’s transcriptional state as determined by molecular factors (Fig 1b-d): over-expressed genes are enhanced whereas under-expressed genes are suppressed as $D$ increases (Fig 2). Consequently, from the sensitivity curves, the combination of an elevated $D$ and $\phi_{in,0}$ widens the distribution of gene expression resulting in transcriptional divergence (Fig 5a). Transcriptional divergence can be quantified by the ratio of expression between the top 50% of genes to the bottom 50% of genes ($P50/P50$) as:

$$\frac{P50}{P50} = \frac{\sum_{j=0}^{N/2} E_j}{\sum_{k=0}^{N} E_k}$$

where $E_j$ is expression of the top 50% and $E_k$ is the expression of the bottom 50% of genes in each patient, respectively, and $N$ is the total number of measurably transcribed genes. *In vitro*, both our model and the gene expression show that as $D$ increases, so too does the $P50/P50$ ratio (Fig S8).
As age is also a major predictor of all-cause mortality, we restrict analysis to patients under 70 years of age at time of diagnosis. Since tumor robustness would be inversely related to patient survival, we hypothesized that patients with an elevated P50/P50 would have a shorter duration of survival. To test this, we compared the P50/P50 ratio measured at time of diagnosis in patients with survival duration below the median versus those above the median for each respective malignancy (~2 years) (Fig 5b). As expected, there was an inverse and statistically significant difference in the P50/P50 ratio between breast (Fig 5b, p-val <0.0001 ) and colon cancers (Fig 5b, p-val= 0.003) and a nearly significant difference for lung cancer (Fig 5b, p = 0.165) with tumors in patients surviving for less than the median survival time having a higher P50/P50. Next, we wanted to test if P50/P50 correlated with the duration of survival. As patient survival depends on a multitude of factors, some of which were not available within the TCGA dataset (e.g. attempted therapeutic intervention, comorbidities), a fixed moving window average was applied to the data (See SI for details). Notably, we found a continuous inverse relationship between transcriptional divergence and patient survival for all three malignancies (Fig 5c, S9a-c). These results suggest a mechanistic link between chromatin packing, phenotypic plasticity, and to patient survival (Fig 5).

6.4 Summary

As a result of the development of new imaging and conformation capture techniques, it is becoming exceedingly accepted that the physical structure of chromatin has an influence on the expression of genes(citation). However, there remains a need for modeling methods that can quantify the effects of the physical properties of chromatin on global patterns in gene transcription.
While there have been in-depth investigations into the role of molecular modifications on both gene expression and local compaction (e.g. promoter methylation, histone acetylation state, chromatin compartments), it has mainly demonstrated a regulatory role in the transcription of individual or few genes. These past studies, however, have not accounted for the role of the physical chromatin environment on all chemical reactions, including those occurring during transcription. Given the highly dense environment within the eukaryotic nucleus, transcription of any gene will depend both on the relatively well studied molecular regulators as well as the less explored physical regulators. The CP-MC model aims to meet this unmet need by integrating the molecular and physical interactions that govern gene expression (Fig 1). The model considers gene transcription as a series of chemical reactions that depend on multiple molecular characteristics (e.g. transcription rate, reactant concentrations, binding affinities, dissociation constants) which are modulated by physical properties within the cell nucleus (average density, scaling of chromatin packing, the upper length-scale of self-similarity).

To do so, the CP-MC model considers the packing of chromatin as a self-similar, power-law medium with scaling dimension, $D$. Utilizing the relationship between $D$, the accessible surface area, and variations in density, the CP-MC model then utilizes Monte Carlo and Brownian Dynamics simulations to calculate the effect of density variations on the chemical reactions governing transcription. From these interactions, the following are identified to play a key role in regulating gene expression: gene length, the upper length scale of self-similarity, average density, and the scaling of chromatin packing. Their effects are as follows: (1) gene length determines the probability of a gene being accessible to transcription factors and RNA polymerase with longer
genes having a greater likelihood of containing an exposed surface. Once bound, these reactants would then be able to move via facilitated diffusion to the transcription start site. (2) The upper length-scale of self similarity; (3) the average density and (4) scaling of chromatin packing.

While we primarily focused on the effect of $D$ on expression, our model shows that changes in the average volume fraction and the limit of self-similarity can also globally influence gene expression. For instance, shifting the average volume fraction through changes in osmolarity, nuclear size, or through aneuploidy could result in a global dysregulation of expression. In conditions that shift homeostasis toward a low baseline volume fraction (e.g. chromosomal loss) an increase in $D$ would produce an effect quite distinct from the $Se(E_i)$ trend observed for more crowded cells, with nearly all genes becoming relatively enhanced. Further, genome compartmentalization may potentially lead to a differential deployment of the chromatin packing regulation within the same genome: Wang et al [146] have proposed the existence of two packing domains with potentially different upper length scales of self-similarity and fractal dimension. A change in $D$ would manifest differently in these domains. One intriguing possibility with respect to evolutionary dynamics of multicellular systems (such as mammalian tissue) is that a heterogeneous cellular population with variable physical traits (nuclear size, texture, density; of note, these are the histological hallmarks of anaplastic or neoplastic transformation) could have a much larger transcriptional range than a homogeneous population.
To test this model experimentally, different platforms and cell lines were leveraged together to compare with the predicted results. In particular, the PLM measurements were conducted to evaluate the transcription with different scaling of chromatin packing density, which showed the same non-monotonic trend between $D$ and transcription as predicted in the model. The divergence trend in transcription while increasing $D$ was testified in the cell line models stimulated by a variety of transcriptional modulators and by knockdown of the SWI/SNF chromatin remodeler, Arid-1a. Changes in the higher-order physical structure of chromatin were monitored in live cells using PWS microscopy, and the transcriptional consequences were measured using mRNA microarrays (Fig. 3). The model predictions, including increased transcriptional divergence as a function of $D$, was observed independently of the underlying molecular pathways of the transcriptional modulation. For example, transformation in chromatin packing within 30 minutes during stimulation acting through vastly different pathways (e.g. differential PMA, EGF, FBS) produced predictable patterns in expression based on changes in $D$, gene size, and initial expression. The increasing size of genomic landscape and the tailwind effect of chromatin packing were validated with the single cell RNA-seq measurement on A2780 cells treated with paclitaxel as stimulator and celecoxib as agent changing $D$ (Fig 5).

The functional consequence of a physical regulatory framework for gene expression was explored in relation to chemotherapeutic resistance. In this work that the chromatin packing code has evolved in order to help cells explore their genomic landscape at time scales much shorter than that of gene mutations, which might be critical for cell/organism acclimation and adaptation to external stimuli and, especially, stresses, thus augmenting the fitness of a cell/organism. Chromatin
heterogenization does not compel cells to change their genome in any specific way; it simply modulates the barrier for functional changes to occur. Although distinct from the gene mutation-dependent adaptation, tailwind effect of chromatin packing on the transcriptional divergence-dependent adaptation indirectly facilitate gene mutations by allowing cells to amplify their survive favorable stimuli, e.g. DNA repair, above the critical level of expression and live long enough for the mutations to occur. These mechanism were confirmed in this work using single cell RNA-seq results. In this context, higher order chromatin packing is a common denominator that controls the overall genomic information space, i.e. transcriptional heterogeneity and divergence. Although the specific biological processes involved in pro-survival processes might be vary from cancer to cancer, the regulation of chromatin packing on genomic information space, which amplifies the chemotherapy evasion. Therefore, the reverse application of the regulation of chromatin packing in genomic information space through CPTs agents would facilitate the efficacy in chemotherapy by limiting the size of genomic landscape the cells could explore and the amplification of pro-survival biological processes through reducing the chromatin packing density variance. Moreover, multiple lines of evidence have shown that chromatin structure is nearly universally transformed in malignancy. At the microscale, chromatin alterations consistent with an increase in $D$ are the gold standard of histopathological diagnosis of dysplasia and malignancy. At the nanoscale, an increase in $D$ has been previously reported to occur at the pre-dysplastic stages of lung, colon, esophageal, ovarian, and pancreatic cancers; the severity of the chromatin transformation has been shown to be an accurate indicator of the tumor aggressiveness. Increased $D$ in a pre-malignant epithelium viewed in relation to the chromatin packing code would explain oncogenic transcriptional inequality, where the high-$D$ cells would be able to explore a greater genomic information space.
In summary, we first found that the model was confirmed to correctly predict the non-monotonic relationship between chromatin packing scaling and the transcription. Secondly, previously observed bidirectional monotonic trend in transcription was found to be depended on both of their initial expression state and the gene length: enhancing the expression of initially up-regulated long genes while simultaneously suppressing initially down-regulated short genes leads to the transcriptional divergence. This transcriptional divergence was experimentally validated \textit{in vitro} with two different cell models (Fig 3). Lastly, the CP-MC model previously revealed the regulation of genomic information space of chromatin packing and we reported another mechanism, the tailwind effect of chromatin packing. In the tailwind effect, the amplification of overexpressed genes in cells with higher chromatin packing density variance was predicted by the model. This prediction indicated that if any genes are up-regulated with certain drug, the up-regulation of these genes would be further amplified as the result of higher chromatin packing density variance. The reverse of this prediction explained the mechanism behind the CPTs, which increases the efficacy of chemotherapy by decreasing chromatin packing density variance, i.e. the overexpression of the genes that are stimulated in chemotherapy would be counteracted with the drug agents that can decrease the chromatin packing density variance.

\textbf{Figures}
Fig 6.1 Molecular and physical regulators of transcription. The regulators influencing transcriptional reactions can be generally divided into two categories: *ex vivo* molecular regulators
(km, KD and [C]tot) (a-d) and in vivo physical regulators (D, φin,0, Mf and L) (d-g). (a) The main molecular components considered in transcription ex vivo are the concentration of transcriptional compounds [C]tot (TF, Pol-II, transcription factor binding site), the transcription rate km, and the disassociation rate of Pol-II KD. (b) The total concentration of available transcription factors [C]tot can change the sensitivity of genes to local crowding conditions. (c) The transcription rate km of Pol-II (yellow ellipse) along DNA (black string) depends on a number of factors, including molecular modification of histones (cyan ellipses) and their density along DNA. (d) The disassociation rate of Pol-II from the transcription start site (TSS) is an inherent property of the chemical reaction that is modulated by local crowding conditions. (e) When considering the reactions in vivo, the physical environment inside nucleus will work together with the molecular regulator to affect transcription, which can be summarized into four physical regulators as: (f) The ChromEM measurement of the average volume fraction (φin,0) of nucleus in A549 lung adenocarcinoma cells with low φin,0 (left) and high φin,0 (right). (g) Chromatin is differentially folded into clusters in the nucleus along a hierarchy of length scales. The nucleus with lower D (left) has a chromatin cluster size distribution decaying faster than the nucleus with higher D (right) resulting in a more homogeneous chromatin topology. Red sub-region represents an interaction volume with a size of 2rin defining the interaction volume within which crowding influences transcription of a gene. (h&i) The linear size of a gene (in bp), L, alters the physical volume the gene occupies upon packing into larger assemblies with a characteristic number of base pairs within the range ~Mf/Mmin where Mf governs the length-scale of chromatin self-similarity. The total mass or genomic length of each shown assembly is the same, but the upper length scale of the self-similarity defines the scaling properties of chromatin domains. For example, a small Mf (left)
produces more independent domains (coded with different colors) than chromatin with a large $M_f$ (right).
Fig 6.2 Comparison of the CP-MC model with experimental measurements of gene expression as a function of physical regulators $D_i, L, M_f$ and $\phi_{in,0}$. (a&b) Representative live cell PWS microscopy images of nuclear ultrastructure with $\Sigma$ scaled between 0.01-0.05 for control A549 (a) and A549 treated with DXM for 12 hours (b) cells. (c&d) Representative ChromEM images of cell nuclei crowding density $\phi_{in,0}$ on A549 (c) and BJ (d) cell nuclei at 0 hour control condition. The zoomed-in regions from each cell line show a high $\phi_{in,0}$ in A549 cell than in BJ cell. (e&f) Validate the model predicted effect of initial chromatin packing scaling $D_i$ on $Se$ with two experimental conditions – wild type HT29 cells (e) and shRNA knockdown Arid-1a HT29 cells (A-KD), a member of the chromatin remodeling SWI/SNF family (f). $D_i$ was approximated using PWS microscopy as $D_i=2.7$ for wild type HT29 and $D_i=2.5$ for Arid-1a knockdown. For the microarray data, each data point is an average of the expression of 100 genes grouped based on their initial expression $E_i$. (g) Comparison of the CP-MC model predictions for $Se$ as a function of gene length $L$ in live control cells. All the other parameters used in the model predictions are the same as the ones in (e&f). To account for initial expression, two representative curves are shown with the different $E_i$: the red curve represents up-regulated genes and the blue curve represents down-regulated genes. Results from the model are in robust agreement with experimental observations for $Se$ as a function of both initial expression and gene size. (h) Validation of the CP-MC model prediction of $Se$ under different $M_f$ in live cells. Hi-C data measuring the contacting domains of each gene in A549 were used to categorize genes in live cells into high $M_f$ (blue dots) and low $M_f$ (red dots) groups (See SI). $M_f$ used in the model for high/low $M_f$ groups was approximated as the average $M_f$ of the genes in each experimental group. PWS measurements on
D and RNA-seq measurements on mRNA expression for A549 treated with 0 hour and 12 hour DXM were used to compare with the model prediction. (i&j) The effect of $\phi_{in,0}$ on $Se$ was predicted by the model and validated with BJ cells (i) and A549 cells (j). The $\phi_{in,0}$ of each cell type was measured using ChromEM with most probable chromatin volume concentration (CVC) of A549 to be 35% and BJ to be 26% (See SI for details). 0-10% of crowding density was added to each CVC to count for the density the floating proteins in nuclei, which gives A549 cell a $\phi_{in,0}$~40% and BJ cell a $\phi_{in,0}$~36%.
Fig 6.3 CP-MC model revealed the increase of intercellular heterogeneity in the cancer cells as the adaptation to chemotherapy. (a-e) The 3D t-SNE analysis on the TPM values of 8275 genes measurement with single cell RNA-seq experiments for 5 conditions, i.e. control cells (a), cells treated with paclitaxel for 16 hours (16hr Pac) (b) and 48 hours (48hr Pac) (c), celecoxib for 16 hours (16hr Cele) (d) and celecoxib and paclitaxel for 16 hours (16hr Combo) (e). The size of the cluster indicates the genomic information space covered by the genome of the surviving cells after each treatment. (f) The radius of genomic space $R_c$ (the radius of clusters through a-e) increases with the increase of the chromatin packing scaling $D$. $D$ is measured by live cell PWS system. Cells treated with paclitaxel (higher $D$) showed an increase in genomic information spaces while cells treated with celecoxib (lower $D$) showed a significant decrease in the size of information space. The increasing size of the genomic information space as a function of $D$ when cells were exposed to chemo agrees with the increase of intercellular heterogeneity in the cancer cells under stress predicted by CP-MC model (red curve and the y-axis on the right).
Fig 6.4 Tailwind effect unraveled by CP-MC model demonstrated the role of chromatin packing scaling in cancer cell adaptation to stress as facilitating malleability. (a) The tailwind
effect amplifies the adaptation of cancer cells under stressor. If the expression of pro-adaptation gene $x$ in the cells is initially $\beta_{1,a}$ at low $D$ state. When the cell is exposed to a drug or a stressor, its response is to upregulate the expression of gene $x$ to $\beta_{2,a}$. If the threshold of the expression of gene $x$ for the cell to survive under stressor is above $\beta_{2,a}$ initially, the cell would commit to apoptosis. On the other hand, if the cell is under higher $D$ state (initial expression without stressor of $\beta_{1,b}$), with everything thing else being equal, the high-$D$ cell will on average reach a higher level of expression of gene $x \beta_{2,b}$ relative to the initial expression $\beta_{1,b}$ than the low-$D$ cell, which increase the possibility of $\beta_{2,b}$ to be above the critical level of expression for survive. (b) Tailwind effect predicted by the model. $\beta_{1,a}$ is the expression rate of genes at $D_a$ without drug stimulation, $\beta_{2,a}$ is the expression rate of the genes after stimulation with drug and $\beta_{1,b}, \beta_{2,b}$ are the expression rate of genes at $D_b$ with and without drug stimulation, where $D_b > D_a$ (c) Tailwind effect validated with A2780 single cell RNA-seq experiment data for initially underexpressed genes (blue dots) and initially overexpressed genes (orange dots). Both of the model predicted results (solid curves) and experimental results showed a strong enhancement of stimulation for the initially underexpressed genes if their expressions are up-regulated with drug treatment. (d) The gene ontology (GO) study of the biological processes involved by the up-regulated genes (2 fold increase of expression) for cells treated with 48 hours paclitaxel. The biological processes shown in (d) were selected out of 20 significant biological processes (P-value<0.05) as the relevant biological processes in cell survival response to chemo (refer to SI for all 20 biological processes). (e) The tailwind effect for up-regulated genes (blue dots) and the genes involved in the relevant biological processes shown in (d) (red dots).
Fig 6.5 The relationship between transcriptional divergence (P50/50) and the outcome of patients with stage III and stage IV breast, colon and lung. (a) As predicted in Se curve by CP-MC model, initially overexpressed genes get further overexpressed when $D$ increases (vice versa), resulting in a broader probability distribution in gene expression at each tail. (b) The comparison
of transcriptional divergence at the time of diagnosis for patients with survival time below versus exceeding the median survival time (2 years) for each cancer type. (c) The relationship between divergence and the survival time in days for patients with breast invasive carcinoma, colon adenocarcinoma, and lung adenocarcinoma. The y-axis is the transcriptional divergence at the time of diagnosis, and the x-axis is the survival time (days from the time of diagnosis and enrollment in the TCGA until death). The results include 38 patients for breast cancer, 33 patients for colon cancer and 47 patients for lung cancer after excluding patients older than 70 years old at diagnosis. All the patients from different cancer types were pooled together to perform a moving window average with window size of 15 samples over P50/P50, illustrating an inverse relationship between the divergence and the ability for the patient to survival in cancer.

6.5 Materials and Methods

Chromatin Packing-Macromolecular Crowding (CP-MC) Model

Variance of chromatin packing density

The variance of chromatin packing density in each interaction volume \( \phi_{in} \) of nucleus, \( \sigma_{\phi_{in}}^2 \), can be expressed as a function of \( \sigma^2 \), the variance of chromatin packing density for the nucleus as continuous media \( \phi \), through \( D \). Considering the fact that the variance of any \( x \) with probability distribution function \( h(x) \) can be calculated from the autocorrelation function \( H(x) \) of \( h(x) \) as \( \text{Var}(x) = H(x=0) \), the relationship between \( \sigma_{\phi_{in}}^2 \) and the autocorrelation function of \( \phi_{in} \), \( B_{in}(\vec{r}) \), is:

\[
\sigma_{\phi_{in}}^2 = B_{in}(\vec{r} = 0)
\]  

(1)

According to the definition of autocorrelation function, \( B_{in}(\vec{r}) \) can be calculated by:
\( B_{\text{in}}(\vec{r}) = \int [\phi_{\text{in}}(\vec{r}' + \vec{r}) - \bar{\phi}] [\phi_{\text{in}}(\vec{r}') - \bar{\phi}] d\vec{r}' = [\phi_{\text{in}}(\vec{r}') - \bar{\phi}] \ast [\phi_{\text{in}}(\vec{r}') - \bar{\phi}] = [\int [\phi_{\text{in}}(\vec{r}') - \bar{\phi}] d\vec{r}' \ast [\phi_{\text{in}}(\vec{r}) - \bar{\phi}] \ast A_{\text{in}}(\vec{r}') - \bar{\phi} A_{\text{in}}(\vec{r}')] = [\phi_{\text{in}}(\vec{r}') - \bar{\phi}] \ast [\phi_{\text{in}}(\vec{r}) - \bar{\phi}] \ast A_{\text{in}}(\vec{r}) \]

\[ = [\phi_{\text{in}}(\vec{r}) - \bar{\phi}] \ast [\phi_{\text{in}}(\vec{r}) - \bar{\phi}] \ast A_{\text{in}}(\vec{r}) \ast A_{\text{in}}(\vec{r}) = B(\vec{r}) \ast A_{\text{in}}(\vec{r}) = \int B(\vec{r}') A_{\text{in}}(\vec{r}') d\vec{r}' \]

(2)

where ‘\( \ast \)’ represents the convolution operation, \( B(\vec{r}) \) is the autocorrelation function of \( \phi(\vec{r}) \) and \( ACF_{\text{in}}(\vec{r}) \) is the autocorrelation function of \( A_{\text{in}}(\vec{r}) \). As a result, Eq. simplifies to:

\[ \sigma^2_{\phi_{\text{in}}} = \int B(\vec{r}) ACF_{\text{in}}(\vec{r}) d\vec{r}' \]

(3)

Because of the self-similar scaling of chromatin, the autocorrelation function of the nuclear crowding density distribution for the chromatin packing model is defined as:

\[ \frac{B(r)}{a^2} = \begin{cases} 
1, r < r_{\text{min}} \\
(r_{\text{min}})^{3-D}, r_{\text{min}} \leq r \leq r_f \\
0, r > r_f 
\end{cases} \]

(4)

Here we use the 1D form of \( B(\vec{r}) \) because of power-law symmetry.

If we assume a Gaussian distribution of the shape of interaction volume \( A_{\text{in}}(\vec{r}) \) with a radius of \( r_{\text{in}} \), the autocorrelation function \( ACF_{\text{in}}(\vec{r}) \) of \( A_{\text{in}}(\vec{r}) \) is also a Gaussian distribution:
\[ ACF_{in}(\vec{r}) = \alpha e^{-\frac{r^2}{2r_{in}}} \]  
(5)

where \( \alpha \) is a constant.

When \( r_{in} \ll r_{min} \), or equivalently taking the limit \( r_{in} \to 0 \), \( \sigma_{\phi_{in}}^2 \) should converge to \( \sigma^2 \), which gives us:

\[ \lim_{r_{in} \to 0} \frac{\sigma_{\phi_{in}}^2}{\sigma^2} = 4\pi \alpha \int_0^{\infty} r^2 e^{-\frac{r^2}{2r_{in}}} \, dr = 4\pi \alpha \frac{\pi}{2} r_{in}^3 \equiv 1 \]  
(6)

In turn, Eq. \( \lim_{r_{in} \to 0} \sigma_{\phi_{in}}^2 = 4\pi \alpha \int_0^{\infty} r^2 e^{-\frac{r^2}{2r_{in}}} \, dr = 4\pi \alpha \frac{\pi}{2} r_{in}^3 \equiv 1 \)

(6) allows us to solve for alpha:

\[ \alpha = \frac{1}{4\pi} \left( \frac{2}{\pi r_{in}^3} \right)^{\frac{1}{2}} \]  
(7)

Therefore, Eq. 5 becomes:

\[ \sigma_{\phi_{in}}^2 = \sigma^2 \frac{1}{4\pi} \sqrt{\frac{2}{\pi r_{in}^3}} \left[ \int_0^{r_{min}} r^2 e^{-\frac{r^2}{2r_{in}}} \, dr + \int_{r_{min}}^{\infty} r^2 e^{-\frac{r^2}{2r_{in}}} \left( \frac{r_{min}}{r} \right)^{3-D} \, dr \right] \approx \sigma^2 \frac{2}{\pi r_{in}^3} \left[ \frac{1}{3} r_{min}^3 + \right] \]

\[ r_{min}^{3-D} \frac{\partial}{\partial r_{min}^{3-D}} \int_{r_{min}}^{\infty} x^{\frac{D}{2}} e^{-x} \, dx \approx \sigma^2 \frac{2}{\pi r_{in}^3} \left[ \frac{1}{3} r_{min}^3 \right] \]

\[ \sigma^2 \left( \frac{r_{min}}{r_{in}} \right)^{3-D} \]  
(8)
where $\Gamma\left(\frac{D}{2}\right)$ is the Gamma function. Here, we assume that $r_{in} \gg r_{min}$ and $v(D) = \sqrt{\frac{2}{\pi}} \frac{D}{2^{D-1}} \Gamma\left(\frac{D}{2}\right) \approx 1$ when $D$ is between 2 and 3. The actual form of $v(D)$ will depend on our assumptions of the interactions that occur within the interaction volume, $A_{in}(\mathcal{F})$. If it has a uniform distribution, $v(D) = \frac{12}{D(D+1)} \approx 1$, which also gives us the same expression of $\sigma_{\phi_{in}}^2$. Next, if we assume a binary distribution of chromatin crowding density (assuming the hard sphere property of chromatin), the variance of the crowding density in nuclei can be approximated as:

$$\sigma^2 = \phi_{in,0}(1 - \phi_{in,0})$$

(9)

So the final expression of $\sigma_{\phi_{in}}^2$ is:

$$\sigma_{\phi_{in}}^2 = \phi_{in,0}(1 - \phi_{in,0}) \left(\frac{r_{min}}{r_{in}}\right)^{3-D}$$

(10)

The equation shown above indicates that the variance of local crowding increases with the increase of $D$. This effect of $D$ on $\sigma_{\phi_{in}}^2$ derived analytically was also observed in simulations of random clusters (Fig S1) and random media (Fig S2 and the “Random media simulation” section), which indicates that the effect of chromatin packing on the variance of local crowding is independent of the choice of chromatin model.
**Fig S1. Relationship between $D$ and $\sigma_{\phi_{in}}^2$.** A media with clusters size distribution $\sim 1/r_s^{4-D}$, where $r_s$ is the radius of cluster, is randomly simulated. The mass density in each interaction volume $\phi_{in}$ is calculated by dividing the media into separate boxes. Then, the standard derivation of $\phi_{in}$, $\sigma_{\phi_{in}}$, is calculated with two different $D$. The blue dash in the figure belongs to the media with low $D$ and the red dash in the figure belongs to the media with high $D$, showing a high $\sigma_{\phi_{in}}$ for the media with higher $D$.

**Random Media Simulation**

To test the relationship between $\sigma^2$, $D$ and $\sigma_{\phi_{in}}^2$ derived based on the chromatin packing scaling model shown in Eq.10, we calculate the variations in density from a numerically generated random media model. The random medium is the medium generated from an autocorrelation function (ACF) with random noise. Of note, the mass density $\rho$ in this calculation can be transformed into
crowding volume fraction, or crowding density $\phi$ through $\rho = \phi \rho_c$, where $\rho_c$ is the dry mass density of macromolecular crowders. First, an autocorrelation function was generated with density variance $\sigma_\rho^2=0.2g/cm^3$, fractal range (1 to 100 nm) and fractal dimension D (1.2 or 2.5), based on the following equation:

$$ACF = \sigma_\rho^2 \exp \left[ \frac{-r/r_{max}}{1+(r/r_{min})^{3-D}} \right]$$

(11)

where $r_{min}=1$ nm, $r_{max}=100$ nm. The fractal dimension in this autocorrelation function can be understood as the power low scaling factor controlling the shape of the ACF. In general, with the other form of ACF of an arbitrary medium (not necessary a fractal), D can be larger than 3. Then the random media were generated using this ACF (Fig S2). The random media generated through this method had on average the same total mass density variance $\sigma_\rho^2=0.2g/cm^3$. For each individual voxel, the mass density was averaged within the interaction volume with radius $r_{in}=20nm$ to calculate $\rho_{in}$. Next, the local variance of each interaction volume is calculated to determine $\sigma_{\rho_{in}}^2$ of each random media to compare low and high D cases. The results from these simulations are as shown in Fig S2. As we can see, the random media generated through ACF with higher D ($D=2.5$) has a larger $\sigma_{\rho_{in}}^2$ compared with the $\sigma_{\rho_{in}}^2$ of the random media generated with lower D ($D=1.25$) when $\sigma_\rho^2$, $r_{min}$ and $r_{max}$ are the same. This confirmation of a direct relationship between D and $\sigma_\rho^2$ shows both the capacity to generate analytical estimates for the variations in density as a function of the ACF and confirms the analytical relationship we derived from the fractal chromatin model in Eq. 10.
Fig S2 Random media simulation for the relationship between $D$ and $\sigma_{\phi_{in}}^2$.

Gene expression analysis

**mRNA Microarray for HT-29 cells**

HT-29 cells were serum deprived for 5 hours as described above prior to treatment with 10% FBS v/v (SE), 100ng/ml epidermal growth factor (EGF), or 100ng/ml phorbol 12-myristate 13-acetate (PMA). mRNA for these treatment groups was collected by TRIzol® isolation (Life Technologies, Carlsbad California) from 10mL petri dishes and analyzed using Illumina human HG12-T microarray chips. The R Bioconductor package, lumi, was used for quality control analysis by the Northwestern Genomics Core to assess probe level processing from the Illumina microarray data. Downstream analysis of gene expression patterns (intra-network transcriptional divergence) was
performed using the 2445 differentially expressed genes using Mathematical® v10 using inbuilt GenomeDatal® to match the annotated genes with their respective processes. Transcriptional divergence for each process was measured analyzing the relative gene expression for each gene in reference to the expression of the serum deprived cells [150].

RNA-seq for A549 and BJ

RNA-seq data on A549 and BJ cells were downloaded from ENCODE and GEO with access code ENCSR897XFT for A549 cells and GSE81087 for BJ cells [160, 161]. 4 replicates are included in both control and 12 hour DXM treated A549 cells. The processed gene quantifications results using featureCounts [162] were downloaded from GEO for A549 cells. The length and the counts for each replicates from featureCounts outputs were then changed into Transcripts Per Kilobase Million (TPM) using \(TPM_i = 10^6 \frac{cts_i}{L_i} / \sum_i \left( \frac{cts_i}{L_i} \right)\), where \(TPM_i\), \(cts_i\) and \(L_i\) is the TPM value, the count and the length of gene \(i\). The differential expression (DE) analysis for A549 cells were performed using the \textit{DESeq2} packages in R [163]. 2292 differentially expressed genes after 12 hours 100 nM DXM treatment were selected from A549 cells using \(p\)-value<0.01. 3 replicates are included in this analysis for BJ cells. The processed fragments per kilobase of transcript per million mapped reads (FPKM) results from 3 replicated for BJ cells from control cells and the cells treated with 100 nM DXM for 32 hours were downloaded from GEO and transformed into TPM unit using \(TPM_i = 10^6 FPKM_i / \sum_i FPKM_i\), where \(FPKM_i\) is the FPKM value of gene \(i\). The same DE method was used on BJ cells and 7601 genes were selected after DE.
Single cell RNA-seq for A2780

The single cell RNA-seq experiments on A2780 were conducted under Illumina NextSeq 500 platform by University of Illinois at Chicago Research Resources Center Cores using Smart-seq protocol [164]. The paired FASTQ reads with four technical replicates of each cell were aligned to mapped and used to hg38 using bowtie2. The gene expression levels, transcripts per million (TPM), under each condition were estimated using software package RSEM [152]. 46 out of 57 control cells, 55 out of 58 16hr pac cells, 53 out of 62 48hr pac cells, 62 out of 67 16hr cele cells and 59 out of 59 16hr combo cells were selected after quality control (cells with less than 4000 genes expressed were excluded). The quality control using the expression level housekeeping genes [165] were performed afterwards, which didn’t exclude extra cells. 8415 genes in each cell were kept after removing the genes expressed in less than 20% of the total cell population. To quantify the size of genomic information space at different chromatin packing conditions, 8276 important genes (average fold changes relative to control are larger than 1.5 or smaller than 2/3) were selected to do a 3-dimension (3D) t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis [166]. The t-SNE analysis was done using ‘Rtsne’ package [166-168] in R with initial PCA step performed.

HiC data processing for A549 to quantify the upper limit of chromatin packing self-similarity

The total mass of chromatin at the upper length scale of self-similarity $M_f$ of genes in the 3D space was estimated using the publically available Hi-C data on A549 cells (GEO access code: GSE92819 for control cells and GSE92811 for cells treated with DXM for 12 hours). $M_f$ was approximated as the size the topologically associated domains (TADs) measured from Hi-C. The processed TADs in A549 cells from the GEO data sets were used to determine the size of TADs and the TADs occupied by each gene. The
genes within the same TAD were assigned with the same $M_f$. The TADs that are the same before and after DXM treatments were selected to ensure that the $M_f$ are consistent. Genes within these consistent TADs were divided into high $M_f$ group and low $M_f$ group with ~130 genes in each group and the average $M_f$ for each group were approximated as 50Kbp for low $M_f$ group and 2Mbp for high $M_f$ group. The average expression for gene in high $M_f$ group is three times higher than the average expression in low $M_f$ group. Genes with top 5% and bottom 5% $M_f$ were removed from each groups to exclude the extreme cases. The $Se$ of each gene group was then calculated by the RNA-seq results and PWS measurements on A549 following the same method described in cell culture section.

Cell Culture and live cell PWS imaging

HT-29 cells

HT-29 Cells (ATCC, Manassas Virginia) were grown in Gibco® formulated McCoys-5A Media (Life Technologies, Carlsbad California) supplemented with 10% v/v FBS (Sigma Aldrich, St. Louis Missouri) and grown at 37°C and 5% $CO_2$. All of the cells in this study were maintained between passage 5 and 25. Transient HT-29 Arid-1a shRNA knockdown line (A-Kd) was produced using a lipofectamine vector. qRT-PCR was used to assess for knockdown: imaging and microarrays were performed on clones that demonstrated at least an 80% reduction in ARID-1a expression compared to the control vector.
Prior to imaging, cells were cultured in 35mm glass bottom petri dishes (Cellvis, Mountain View, CA) until at least 50% confluent. Cells were given at least 24 hours to re-adhere prior to 5 hours of serum deprivation. For serum deprivation, cells were grown in fresh McCoy’s 5A (Life Technologies) without serum supplementation and maintained at 37°C with 5% CO₂.

**A2780 cells**

Ovarian A2780 cells were a gift from Dr. Chia-Peng Huang Yang and obtained from the lab of Dr. Elizabeth de Vries at Albert Einstein College of Medicine. They were cultured in RPMI-1640 Medium (ThermoFisher Scientific, Waltham, MA # 11875127). All culture media was supplemented with 10% FBS (ThermoFisher Scientific, Waltham, MA #16000044).

Cells were cultured in 35mm 6-well glass bottom plates (Cellvis, Mountain View, CA) until 60-85% confluent. All cells were given at least 24 hours to re-adhere prior to pharmacological treatment. Cells were treated with 75uM celecoxib (2 hrs, 16 hrs), 5nM paclitaxel (16 hrs, 48 hrs), or combination celecoxib and paclitaxel (16 hrs) prior to trypsinization and being resuspended in growth media. Cell sorting was performed on a Fluidigm - C1 Single-Cell Capture instrument. Single cell sequencing of the sorted cells was performed by staff researchers at the University of Illinois Chicago Genomics Core.

**A549 and BJ cells**

A549 cells were cultures in Dulbecco’s Modified Eagle Medium (ThermoFisher Scientific, Waltham, MA, #11965092). BJ cells were cultured in Minimum Essential Media (ThermoFisher Scientific, Waltham, MA, #11095080). All culture media was supplemented with 10% FBS (ThermoFisher Scientific, Waltham, MA,
no. 1600044) and 100 µg/mL Penicillin-Streptomycin (ThermoFisher Scientific, Waltham, MA, # 15140122). All cells were maintained and imaged at physiological conditions (5% CO\textsubscript{2} and 37 °C) for the duration of the experiment. All cell lines were tested for mycoplasma contamination with Hoechst 33342 within the past year. Experiments were performed on cells from passage 5–20.

Before imaging, cells were cultured in 35 mm glass bottom petri dishes until approximately 70% confluent. All cells were given at least 24 hours to re-adhere prior to treatment (for treated cells) and imaging. A549 and BJ cells treated with Dexamethasone (Sigma-Aldrich, St. Louis, MO, D6645) were treated with a dose of 100 nM.

**Live-cell PWS measurements**

PWS measurements were performed on a commercial inverted microscope (Leica DMIRB) using a Hamamatsu Image-EM CCD camera C9100-13 coupled to a liquid crystal tunable filter (LCTF; CRi Woburn, MA) to acquire mono-chromatic spectrally resolved images that range from 500-700nm at 1nm intervals produced by a broad band illumination provided by an Xcite-120 LED Lamp (Excelitas, Waltham, Massachusetts) as previously described [97]. These spectrally resolved images were normalized by the incident light scattering produced from the glass-media interface by taking an independent reference from a field of view without cells. A low-pass Butterworth filter was applied to reduce spectral noise prior to the calculation of Σ at each pixel. The measured change in structure, quantified by ΔΣ, was obtained by analyzing nuclei from over 50 cells per condition group [150].
ChromEM measurement on crowding density

Cell culture

Two cell lines were used in this work: adenocarcinomic human lung epithelial cell line (A549), and human cellosaurus cell line (BJ). The A549s were grown in DMEM with 10% FBS. The BJ cells were grown in MEM with 10% FBS and 1x non-essential amino acids (NEAA). All cells were cultured on 35 mm MatTek dishes (MatTek Corp) at 37°C at 5% CO₂. Confluency of around 60% were reached for all experiments.

EM sample preparation and EM data collection

For EM experiment, all the cells were prepared by the ChromEM staining protocol and embedded in Durcupan resin (EMS) [94]. After curing, 40 nm thin sections were made and deposited onto copper 200 mesh grid with carbon/formvar film (EMS). The grids were plasma-cleaned by a plasma cleaner (Easi-Glow, TED PELLA) prior to use. A HT7700 (HITACHI) transmission electron microscopy was employed to record TEM images of cell sections at 80 kV with a pixel size of 2.5 nm.

Chromatin packing analysis using TEM images of thin sections

Prior to processing, the negative logarithm of the TEM image intensity were calculated to convert the image contrast into mass-thickness distribution based on the Beer’s law. The moving-window average DNA concentration were calculated for the whole nucleus, and the window size was chosen to be 100 nm³ after taking the thickness of the sections into consideration. The nucleus segmentation was conducted manually in FIJI. We then normalized the corresponding nuclear CVC so that it has the same range as the CVC distribution in previously published work [2], and the nominal minimum and maximum from the TEM images of thin sections were defined as the CVC values that accounts for 0.05% and 99.95% of the total data respectively.
Fig S3. The relative expression for cells with a higher $D$ ($\beta_{2,b}$, 16hr pac) is higher than that in cells with a lower $D$ ($\beta_{2,a}$, 16hr combo). The CP-MC model (blue curve) qualitatively predicts the similar trend as it is measured with RNA-seq results.

Gene Oncology analysis
To perform the gene oncology analysis, the average TPM values of each gene across all the A2780 cells under each condition were normalized to the mean TPM values of control. The top 10% expressed genes after normalizing with 48hr pac treatment were selected (841 genes) to conduct the gene ontology analysis using DAVID [169, 170]. 20 biological processes were shown to be significantly involved by these overexpressed genes (Fig S4). Out of the 20 upregulated biological processes, 11 of them are related to DNA repair (Fig 4d)
Fig S4. Full list of biological processes involved by upregulated genes.

**TCGA patients data analysis**

The P50/P50 ratio of each patient’s gene expression is calculated by:
\[
\frac{P_{50}}{P_{50}} = \frac{\sum_{j=0}^{[N/2]} FPKM_j}{\sum_{k=0}^{[N/2]} FPKM_k}
\]  
(12)

where \( FPKM_j \) is the FKPM value of the top 50% transcribed genes in each patient, \( FPKM_k \) is the FKPM value of the bottom 50% transcribed genes in each patient and \( N \) is the total number of measurably transcribed genes. Only transcribed genes (FKPM \( \neq 0 \)) are considered using RNA-seq obtained from the TCGC database. Patients from breast, colon and lung cancers in stage III and stage IV are divided into survival over/below 2 years based on their vital status and their days to death at diagnosis. The Average box whisker plots of \( P50/P50 \) for these two groups for each cancer type are plotted in Fig 5b. The number of patients in each group can be found in Table S2. Then, the \( P50/P50 \) values of patients from all three cancer types (breast, colon and lung cancers) are pulled together to apply a fixed moving window average (MWA) with 15 patients per group to visualize if an overall trend exists between \( P50/P50 \) and survival time (days) Fig 5c. This analysis is applicable to inherently noisy data or for datasets where important co-variates are not completely available (e.g. chemotherapeutic/radiation therapy status or comorbidities were not considered). A linear regression analysis using survival time as response, \( P50/P50 \) and stages as predictors is conducted using Python showing a significant prediction of \( P50/P50 \) (p-value<0.05) with negative coefficient. But no evidence exists showing a strong prediction power of stages (p-value>0.05), indicating that \( P50/P50 \) is an independent predictor of survival, instead of the marker of stages. The relationships between survival time and \( P50/P50 \) for the patients in each cancer are shown separately in Fig S5.
Fig S5. Relationship between survival time (days) and P50/P50 for breast (a), colon (b) and lung (c) cancers. A fixed MWA with window size 5 is applied to visualize any trend in the data. The red curves are the fitted curve based on the trend in the data.
Tables:

Table 1. Descriptions and values of parameters used in predicting by CP-MC model.
<table>
<thead>
<tr>
<th>FIXED PARAMETERS</th>
<th>DESCRIPTION</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{D_{Pol-II}}$</td>
<td>Dissociation rate of Pol-II in the absence of crowders</td>
<td>1nM</td>
</tr>
<tr>
<td>$k_m$</td>
<td>Transcription rate of Pol-II in the absence of crowders</td>
<td>$1 \mu M^{-3} s^{-1}$</td>
</tr>
<tr>
<td>$r_{min}$</td>
<td>Lower length scale of chromatin self-similarity</td>
<td>1nm</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Variance of continues crowding density $\phi$</td>
<td>$\phi_{in,0}(1 - \phi_{in,0})$</td>
</tr>
<tr>
<td>$L$</td>
<td>Average number of base pairs in each gene</td>
<td>6Kbp</td>
</tr>
<tr>
<td>$r_{in}^0$</td>
<td>Radius of interaction volume for single base pair</td>
<td>15nm</td>
</tr>
<tr>
<td>$M_{min}$</td>
<td>Mass of lower length scale of chromatin self-similarity</td>
<td>1bp</td>
</tr>
<tr>
<td>$M_f$</td>
<td>Total mass of upper length scale of chromatin self-similarity</td>
<td>Average for all cell types:~1Mbp</td>
</tr>
<tr>
<td>$\phi_{in,0}$</td>
<td>Average crowding density</td>
<td>HT29 cells: 39% v/v</td>
</tr>
<tr>
<td>$D_i$</td>
<td>Initial chromatin fractal dimension</td>
<td>Wild-type HT-29 cell: 2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UNFIXED PARAMETERS</th>
<th>DESCRIPTION</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |
|                   |             |       |</p>
<table>
<thead>
<tr>
<th>Tumor</th>
<th>Tissue</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Invasive Carcinoma</td>
<td>Breast</td>
<td>Alive: 163, Dead: 24</td>
</tr>
<tr>
<td>Colon Adenocarcinoma</td>
<td>Colorectal</td>
<td>Alive: 40, Dead: 31</td>
</tr>
<tr>
<td>Lung Adenocarcinoma</td>
<td>Lung</td>
<td>Alive: 34, Dead: 37</td>
</tr>
</tbody>
</table>

Table 2. Information about patient characteristics.
Chapter 7: Discussion, Conclusion and Future Directions

7.1 Summary

Epigenetic biological activities within the cells including the cellular plasticity, adaption and malleability are not the consequence of the change of individual gene or network. To better understand the development of diseases we need to find the connection between the changes of a large group of genes, in terms of their transcriptional outcome, with the alteration of cells, which usually happens within nanoscale at early stage of emergence. With the development of diagnosis tools measuring optical properties in diseases, many effort has been put into studying the nanoscale physical structure of chromatin since it has been known that the optical properties of biological samples can reflect the physical structure of the sample with nanoscale resolution. To fully understand the mechanisms behind the diseases including cancer, a method is needed to interpret the regulation of cellular function by a group of genes. Due to the development of experimental techniques that measuring the statistical properties of nuclei structure and RNA expression quantification techniques, it became promising to connect the physical structure of cell nuclei at supra-nucleosomal level to whole genome transcription pattern. With this connection, we can investigate how the cells regulate the global transcription pattern through altering the physical structure. The interpretation of the functional roles of chromatin physical structure in transcription suggests a new method to study the epigenetic cell behavior regulation through exploring the engineering of macrogenomic information. In this work, we combined the diagnoses techniques, chromatin 3D structure capture techniques, molecular dynamics simulations, mathematical calculation and mRNA expression quantification methods to build an integrated framework to understand the regulation of chromatin packing scaling on cellular plasticity.
To integrate different components in the framework, we first developed a method to calculate optical properties by using EM to measure the nanoscale physical structural details within cell nuclei (Chapter 2). The agreement between the optical properties measured based on the structural information from the EM method and the ones measured from independent studies provides the method to visualize the change of optical properties in the chromatin structure. Since the changes of optical properties of nuclei have been frequently reported in multiple diseases, the understanding of the connection between physical structure and its functional outcomes plays an important role in understanding the epigenetic evolution of diseases. Based on this, the Chapter 3 is focused on using MD and BD simulation method to study how the transcription is changed under different chromatin packing density. Incorporated with the system biology method, the simulation results indicate different inter-molecule interactions will lead to a different binding free energy between transcriptional molecules including transcription factors and RNAp. Since the previous chapter revealed that the chromatin packing density scaling can be regulated with different intermolecular interaction, the interplay between the binding free energy and the intermolecular potential indicates that the change of transcription can be reflected by the change of chromatin packing density scaling, based on previous work by Hiroaki Matsuda et al [17], which unravel the relation between the binding free energy and mRNA transcription rate. Therefore, in Chapter 4 we developed an integrated experimental platform that can be used to measure the 3D physical structure of chromatin at different length scale by combining ChromEM, HiC, PWS and STORM. The BD simulation is used to connect the chromatin packing density scaling $D$ measured by PWS with the 3D chromatin structure technique HiC measuring the contact probability scaling $s$ between base pairs showing that both $D$ and $s$ can be altered through the intermolecular interaction within nuclei. Finally, we combined the effect of chromatin packing...
density on transcription with the power law scaling properties of chromatin to describe the chromatin physical structure as the scaling of chromatin packing (CP) with local macromolecular crowding (MC) in Chapter 5. The CP-MC model suggests the existence of a pathway-independent global chromatin scaling “code” that governs the stochastic variations in gene expression by acting directly on physical factors such as gene accessibility, binding affinities, and transcription factor diffusion. The unraveling of the chromatin packing code suggest a macrogenomic engineering mechanisms that the global transcriptional malleability and heterogeneity are critically impacted by the physical condition within nuclei. Since the malleability and heterogeneity play a paramount role in regulating the cell resistance under environmental stress, we proposed the chromatin protective therapies (CPTs) to improve the efficacy of chemotherapy by targeting the physical structure of chromatin. All the proposed mechanisms have been validated with mRNA expression abundance analysis from multiple cell lines.

7.2 Future directions and Emerging Questions

7.2.1 Develop multiscale modeling method to improve prediction by CP-MC model

The CP-MC model discussed in Chapter 5 is mainly based on the coarse grained simulation with beads on the string model as chromatin and the symmetric sphere as free macromolecules in the nuclei, which can represent the essential roles of chromatin packing nano-environment in transcription. However, the real nuclei environment is more complicated than this. Therefore, a more sophisticated model is needed to better predict the transcriptional outcome under the effect of chromatin packing. For example, the loop extrusion chromatin model has been proposed to well represent the 3D structure of chromatin packing, which can be incorporated into the CP-MC model when we consider the packing of chromatin within length scale ~ 10 – 100 nm. Moreover, the free
crowders in nuclei can be modeled with structural detail in MD simulation at the length scale < 10 nm. With this application of multi-scale modeling, we can develop a more accurate nuclei model that can be used to in macrogenomic engineering in the future.

7.2.2 Regulation of chromatin packing scaling

The $D$ of a chromatin polymer depends on the relative strength of chromatin-chromatin and chromatin-nucleoplasm interactions as revealed in Chapter 3. For instance, the change of ionic concentration can modulate the chromatin-chromatin interaction. By utilizing the ChromEM, PWS and STORM, we can monitor the change of chromatin structure at different length scale. The CP-MC model can be used to predict the effects of the nuclear environment on chromatin interactions from atomistic detail to the Mbp scale, and on the resulting chromatin packing and the intertwined chromatin nanoenvironment. The chromatin nanoenvironment will in turn affect the reaction constants for the chromatin packing-dependent regulation of transcription

7.2.3 Discover more CPTs with prediction from CP-MC

While we primarily focused on the effect of $D$ on expression with the CP-MC model as one of way to engineer the macrogenomic expression, the parameter study in Chapter 6 showed that other physical regulators, especially $\phi_{in,0}$, also have a significant impact on the genomic landscape in cells. With the development of the experimental techniques that can fast and accurately measure $\phi_{in,0}$, we can potentially discover drugs regulate the density of chromatin as CPTs to prevent the emergence of chemoresistance. For instance, the simulation shown in Chapter 3 indicates that the change of intermolecular attractive potential can change the chromatin packing density as well as the scaling of chromatin packing density. Herein, if we can regulate the force field in nuclei,
through regulating the ionic environment for instance, the chromatin packing density can be engineered in a preferable way. Drugs, e.g. Aquaporin-4, can therefore be further tested to evaluate their efficacy in terms of regulating the chromatin structure as well as the genomic landscape to be potential CPTs.
Reference


[66] A.J. Radosevich, Generate random media according to the Whittle Matern model in Matlab, in, Matlab functions, Biophotonics Laboratory at Northwestern University, Matlab functions, Biophotonics Laboratory at Northwestern University, pp. Matlab functions, Biophotonics Laboratory at Northwestern University.


[82] (!!! INVALID CITATION !!).


[155] TCGA, in.


