Title:
Disordered chromatin packing regulates phenotypic plasticity

Authors:
Ranya K.A. Virk[1¶], Wenli Wu[1¶], Luay M. Almassalha[1,2,3¶], Greta M. Bauer[1], Yue Li[1,4], David VanDerway[1], Jane Frederick[1], Di Zhang[1], Adam Eshein[1], Hemant K. Roy[5], *Igal Szleifer[1,6,7], and *Vadim Backman[1,7]

Affiliations:
[1]Department of Biomedical Engineering, Northwestern University, Evanston, Illinois, 60208, USA
[2]Medical Scientist Training Program, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, 60211, USA
[3]Department of Internal Medicine, Northwestern University, Chicago, Illinois, 60211, USA
[4]Applied Physics Program, Northwestern University, Evanston, Illinois, 60208, USA
[5]Section of Gastroenterology, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, 02118, USA
[6]Department of Chemistry, Northwestern University, Evanston, Illinois, 60208, USA
[7]Chemistry of Life Processes Institute, Northwestern University, Evanston, Illinois, 60208, USA

***Corresponding Authors:**
Vadim Backman, Ph.D.
Walter Dill Scott Professor
Biomedical Engineering Department
Northwestern University
2170 Campus Drive, Silverman Hall 3627, Evanston, IL 60208, USA
Phone: (847) 491-3536
Email: v-backman@northwestern.edu

Igal Szleifer, Ph.D.
Christina Enroth-Cugell Professor
Biomedical Engineering Department
Northwestern University
2170 Campus Drive, Silverman Hall 3627, Evanston, IL 60208, USA
Phone: (847) 467-0674
Email: igalsz@northwestern.edu

¶These authors contributed equally to this work.

One Sentence Summary:
The CPMC model demonstrates the regulatory role of chromatin's physical structure on transcription, with implications for phenotypic plasticity.

Abstract:

Three-dimensional supranucleosomal chromatin packing plays a profound role in modulating gene expression by regulating transcription reactions through mechanisms such as gene accessibility, binding affinities, and molecular diffusion. Herein we employ a computational model that integrates disordered chromatin packing (CP) with local macromolecular crowding (MC) to study how physical factors – including chromatin density, the scaling of chromatin packing, and the size of chromatin packing domains – influence gene expression. We identify computationally and experimentally a major role of these physical factors, specifically chromatin packing scaling, in regulating phenotypic plasticity; determining responsiveness to external stressors by influencing both intercellular transcriptional malleability and heterogeneity. Applying CPMC model predictions to transcriptional data from cancer patients we identify an inverse relationship between patient survival and phenotypic plasticity of tumor cells.

Main Text:

INTRODUCTION

Most perturbations a eukaryotic cell experiences occur at non-replicative time scales. These perturbations are remarkably varied, range in intensity, and can be completely distinct from previously encountered stimuli. Examples exist throughout the human body, including within the skin, the alimentary tract, the immune system, the respiratory tract, the reproductive system, and in malignancy. Consider the epithelial lining of the digestive and respiratory systems. While both systems are constantly renewing their lining, the majority of functional cells within these tissues persist for days to weeks after replication. During their lifespan, these cells are exposed to a wide range of nutrients and toxicants that necessitate modification of gene expression to carry on basic cellular functions across these variable conditions (e.g. appropriately absorb nutrients, regulate ionic homeostasis, maintain a sufficient mucosal barrier, excrete waste products, secrete immunoglobulins). No better example may exist than malignancy, as tumor cells are remarkably adept at acclimating to a broad spectrum of cytotoxic chemotherapies and radiation exposure, while evading detection from the myriad tools present within the immune system. These capabilities evoke a critical question – how do individual cells acclimate to fluctuating or completely novel conditions? Likewise, how do collections of cells, such as an organ or a tumor mass, acclimate in aggregate to a heterogeneous, rapidly evolving environment?

One widely explored mechanism to respond to such varied conditions is to have a level of predetermined functionalization: intermixing specialized cells within an organ to carry out specific roles. Beyond establishing pre-coordinated responses, an intriguing possibility is for cells and cell populations to have an encoded level of phenotypic plasticity in order to acclimate to novel conditions in real time (*1, 2*). In the context of multicellular systems, the level of phenotypic plasticity encoded would be a product of cellular malleability, i.e. the functional responsiveness of cells toward stable end states upon external stimulation, and the level of intercellular heterogeneity, the diversity of stable states that are observed within the same population at a given time.

Interestingly, recent advances in single cell technologies have highlighted the remarkable levels of diversity within multicellular systems for otherwise seemingly identical cells (*3-5*). An extraordinary level of diversity has been demonstrated in the lungs (*3*), breast tissue (*4*), the gastrointestinal tract (*5*), and in malignancy (*6-9*). Furthermore, the cancer state is associated with considerable structural (*10, 11*), transcriptional (*8, 12*), epigenetic (*9, 13*), and mutational heterogeneity (*6, 9*) – all of which have been demonstrated to be independently linked to chemotherapeutic resistance, metastasis, survival and resilience in multiple cancer models. Likewise, transcriptional responsiveness is concomitant with cancer cell survival in response to chemotherapy as well as the functional responsiveness of immune cells to microbes (*14*). To date, these two facets of phenotypic plasticity have largely been viewed as distinct entities as no mechanisms have been identified that simultaneously link both diversity and responsiveness. However, at the level of gene transcription, both the malleability and intercellular heterogeneity of gene expression within cell populations could result from the physical organization of chromatin (*14, 15*).

To test the hypothesis that the physical organization of the genome is a regulator of both transcriptional malleability and intercellular heterogeneity, we utilized multi-scale modeling to describe transcription as a series of chemical reactions occurring in a heterogeneous, crowded environment – a disordered Chromatin Packing Macromolecular Crowding (CPMC) model. Pairing the CPMC model with single-cell RNA sequencing, chromatin electron microscopy tomography (ChromEM) – a DNA-specific staining technique for electron microscopy – and live cell Partial Wave Spectroscopic (PWS) microscopy, we demonstrate that the physical structure of chromatin packing determines both the level of transcriptional malleability and heterogeneity. In particular, the CPMC model predicts that at the supranucleosomal scale (from ~kbp to several Mbp) the scaling behavior of chromatin packing size, which is the relationship between the genomic length of a chromatin chain and its packing size, determines the level of intercellular transcriptional heterogeneity by regulating local variations in chromatin density (*14, 16*). Furthermore, the scaling of chromatin packing regulates the level of transcriptional malleability by regulating both gene accessibility and the free energy of transcription reactions (*17-19*). Finally, applying the CPMC model to interrogate the phenotypic plasticity of cancer cells, we show that increased transcriptional malleability has an impact on cancer mortality. Analyzing gene expression data from The Cancer Genome Atlas (TCGA) (*20*), we demonstrate that transcriptional divergence – a direct measure of the level of transcriptional malleability, which is connected with chromatin packing scaling – is inversely related to patient survival in advanced (Stage 3 and Stage 4) colorectal, breast, and lung cancers. In sum, this work mechanistically links two distinct aspects of phenotypic plasticity, transcriptional malleability and intercellular heterogeneity, with the physical properties representing the structure of disordered chromatin packing. Utilizing the CPMC model, we quantitatively describe the role that physical forces play on gene expression *in vitro* and describe a potential mechanistic relationship between structural alterations of chromatin observed in cancer and prognosis.

MODEL

The CPMC model considers transcription in dilute, *ex vivo* conditions as a series of diffusion limited chemical reactions that utilize DNA, transcription factors (TFs), and RNA polymerase II

(Pol-II) 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**). These molecular factors are well-studied regulators of gene expression *in vitro*. For example, at the scale of nuclear compartments, formation and dissipation of TADs can alter local transcription factor concentrations (*21*). Additionally, post-translational histone modifications alter nucleosomal stability, thereby influencing the rate of polymerase elongation (*22*). Other post-translational modifications of RNA polymerase itself independently control polymerase activity (*23*). Furthermore, gene motifs determine binding affinities of polymerase and transcription factors, resulting in varied dissociation constants of these molecules from their respective target genes (*24*).

Compared to *ex vivo* conditions, the eukaryotic nucleus is a highly crowded, heterogeneous environment (**Fig. 1e**). To model transcription reactions within such an environment requires consideration of the length scales involved. At the smallest scale (within ~20nm of a gene, i.e. an "interaction volume"), macromolecular crowding (ϕ_{in}) influences transcription by affecting the mobility of transcriptional reactants and the dissociation rate of these molecules from DNA (*19, 25, 26*). Additionally, the accessible surface area of chromatin determines the number of DNA binding sites available to transcriptional reactants. The probability of a gene promoter being available for transcription depends on its local accessible surface area. At these small length scales, transcription can be modeled as a network of chemical reactions involving TFs, Pol-II, and DNA. TFs bind to their respective DNA-binding sites and recruit polymerases to gene promoters which, in turn, bind DNA. These series of reactions result in intermediary transcription complexes that stochastically transcribe genes into mRNA. Each reaction coefficient depends on local crowding effects, which can be calculated using Brownian Dynamics (BD) and Monte Carlo (MC) simulations. Gene expression for particular crowding conditions is calculated by solving the steady-state network of equations that models these transcription reactions (*19, 26*). This modeling approach predicts a non-monotonic dependence of transcription on crowding. The non-monotonic behavior is influenced by the molecular factors previously discussed and is due to the opposing effects of macromolecular crowders on chemical reactions. Initially, transcription rates increase with crowding due to an enhanced binding stability of TFs and Pol-II arising from attractive depletion interactions. At higher crowding conditions, however, the crowding-induced reduction of molecular mobility dominates, lowering transcription rates. Notably, the most prevalent macromolecular crowder in the nucleus is chromatin. Thus, local chromatin density within the interaction volume of a gene should have a profound effect on transcription processes. Recent electron microscopy studies have shown that chromatin packing density is highly heterogeneous across the genome. Some genes have interaction volumes with exceedingly high densities (chromatin volume concentration (CVC) up to >60%) while others may be positioned in regions of the nucleus with CVC as low as ~10-20% (*27*). One approach to study the effect of local crowding on transcription in cells would be to experimentally measure the local density of chromatin near every gene using electron microscopy and pair these measurements with *in situ* mRNA levels. This, however, is beyond existing technical capabilities, and an alternate approach is needed.

Instead of experimentally mapping gene expression to locus-specific crowding conditions, the CPMC model probabilistically samples the polymeric properties of chromatin in order to

approximate transcriptional output of an ensemble of genes under similar molecular and varying physical conditions (*14, 28, 29*). A combination of molecular factors influences the relative initial expression levels of these genes (*19*). In this work, we focus on how physical regulators further modulate transcription reactions to produce a final observed transcription rate. The model considers chromatin to be a disordered heteropolymer that is heterogeneously packed in three-dimensional space. The 3D packing of the chromatin polymer determines the volume fraction occupied by chromatin, the number of nucleotides acting together as a grouped polymeric entity (N_d), and the space filling geometry or the scaling behavior of these polymeric entities. N_d can be considered as the number of nucleotides that are contained within a subset of the chromatin polymer that has self-similar, power law scaling properties. As is the case with most other disordered polymers, the power law scaling behavior describes the relationship between the length of a given segment of the chromatin polymer (e.g. the number of nucleotides, N) and the size (R) of the physical space occupied by the segment, $N \propto R^D$ for $N \leq N_d$. The scaling factor D is frequently referred as the fractal dimension of the polymer and is determined by the balance of the free energy of polymer-polymer and polymer-solvent interactions. D of an unconstrained free polymer may range from $D = 5/3$ for an excluded volume polymer to $D = 2$ for an ideal chain polymer in theta solvent and to $D = 3$ for a completely space-filling polymer. A polymer with a uniform chain structure throughout would form a single fractal domain with D determined by the properties of the chain as well as the solvent. Chromatin, on the other hand, is a heterogeneous polymer with variable histone and DNA methylation. This leads to differential interactions between the heterogeneous chromatin subunits and results in chromatin compartmentalization, potentially as a result of liquid-liquid phase separation (*30*). Additional topological constraints induced by chromatin-binding proteins, such as those responsible for the formation of chromatin loops or nuclear lamins, might further influence D within a given chromatin domain or compartment. Indeed, electron microscopy and super-resolution imaging studies have demonstrated the existence of spatially segregated supranucleosomal nanoscale packing domains with a variable size distribution in 3D space (*27, 31*). We have been able to visualize the existence of these packing domains using ChromEM (**Fig. 1e**) and PWS (**Fig. 1f**) as small (100-200 nm in diameter; genomic size between 100 and 400 kb), globular regions of higher chromatin crowding density and D. The CPMC model considers a gene's interaction volume to be located within these packing domains. Accordingly, the local environment of a gene's interaction volume is determined by the encompassing packing domain, each of which may have its own average nuclear crowding density ($\phi_{in,0}$) (**Fig. 1e**), chromatin packing scaling D (**Fig. 1f&g**), and genomic size (N_d) (**Fig. 1h**). These local physical conditions are important determinants of gene expression. In addition, gene length (L) partially influences the size of the interaction volume of a given gene, affecting the range of crowding conditions the gene is probabilistically exposed to. The CPMC model is eminently useful as it uses these measurable physical regulators of chromatin to approximate distributions of mass density and accessibility of chromatin to determine transcription for each gene throughout the entire nucleus, a feat which is currently experimentally infeasible (*17*).

The expected expression rate of a gene *in vitro* is the product of the steady-state mRNA expression rate of that gene (ϵ) and the probability of the gene to be on the accessible surface of the chromatin polymer (p_g). Steady-state expression rate is a function of molecular features surrounding the gene of interest ($\vec{m}$; transcription factor concentration, histone state, enhancer-promoter interactions, etc.) (**Fig. 1b-d**) in the context of local physical conditions (**Fig. 1e-h**) (*14, 18, 19, 25*). The probability of gene accessibility contributes to the likelihood of a gene to interact with

transcriptional components (TFs and Pol-II) *in vitro* (*32*). It is beyond technical capabilities to measure all molecular and physical parameters of the model for specific genes at the single-cell level. Thus, we explore how a given ensemble of genes with similar molecular features $\vec{m}$ (e.g. grouped by their initial expression or associated gene ontologies) would respond to changes in average measurable physical conditions. Specifically, we study how average nuclear crowding density, $\phi_{in,0}$, average chromatin packing scaling, *D*, and genomic size of a packing domain, N_d, change the behavior of global transcription processes. It is critical to stress that the CPMC model does not assume that the chromatin polymer has the same power law scaling behavior or constant density throughout the entire nucleus, but that this is instead an approximation due to existing experimental limitations. The model can further be extended to consider each packing domain has its own chromatin packing scaling *D* as technological cabilities to co-register chromatin packing, molecular, and genomic properties advance. Finally, in this model, nuclear crowding density within each interaction volume, ϕ_{in}, is assumed to be constant relative to the time-scale of transcription (~minutes), in line with recent imaging studies of chromatin mobility (*33*).

Given these considerations, in a population of cells, each gene will be exposed to different crowding densities ϕ_{in}. Each ϕ_{in} will be sampled from the probability distribution function $f(\phi_{in})$, which is assumed to follow a normal distribution with mean $\phi_{in,0}$ and variance $\sigma^2_{\phi_{in}} \approx \phi_{in,0}(1-\phi_{in,0})({}^{r_{min}}/_{r_{in}})^{3-D}$, where r_{min} is the radius of the elementary unit of chromatin (e.g. a base pair) and r_{in} is the radius of the interaction volume (**Supplementary Text**) (*14*). Due to the mass-fractal nature of chromatin, $r_{in} = L_{in}^0 + L^{1/D} r_{min}$ for a gene of length *L*, where L_{in}^0 is the radius of the interaction volume for a single base pair and is approximated from previous MC simulations of crowding effects (*14, 19*). Thus, the expected range of crowding densities each gene is exposed to is dependent on the statistical properties of the packing domain where the gene is located, including *D* and $\phi_{in,0}$, and is further influenced by length *L* of the gene. The transcription rate ϵ itself is assumed to depend on molecular features $\vec{m}$ as well as on local crowding density ϕ_{in}. We calculate all expression rates under the assumption that molecular features $\vec{m}$ remain constant throughout the population, with physiologically relevant values used in previous Monte Carlo and Brownian Dynamics crowding simulations (**Table S1**) (*19*). This gives rise to the form of $\bar{\epsilon}$, the average expression rate for an ensemble of genes that share a given $\vec{m}$ as:

$$\bar{\epsilon} = \int \epsilon(\vec{m}, \phi_{in}) f(\phi_{in}) d\phi_{in} \quad (1)$$

Likewise, a power law model of chromatin packing scaling allows the CPMC model to calculate the probability of a unit of DNA (e.g. a gene promoter) to be on the accessible surface of chromatin, p_g, (*28, 29*):

$$p_g \propto N_d^{\ -1/D} \quad (2)$$

Finally, merging accessibility with steady-state expression rate for a group of genes, the ensemble expression rate is:

$$E = \bar{\epsilon} \cdot p_g \quad (3)$$

To quantitatively analyze the effect of *D* on gene expression, we calculate the sensitivity of gene expression as a function of *D* as predicted by the CPMC 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*). *Se* of expression rate for any group of genes to changes in chromatin packing is defined as:

$$Se = \frac{\partial \ln(E)}{\partial \ln(D)}\Big|_{E=E_i,\ D=D_i} \tag{4}$$

where E_i is the initial average expression rate of the group of genes sharing similar molecular features $\vec{m}$ and gene length *L*, and D_i is the initial average packing scaling 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 collective expression rate. Importantly, the CPMC model predicts the output of transcription reactions that occur within the nucleus. Assuming that the half-life of mRNA transcripts is dictated by cytoplasmic conditions, structural changes in chromatin that alter chromatin packing scaling *D* are not considered to alter the degradation rate of mRNA. Thus, sensitivity should be directly related to the number of transcripts produced for any group of genes in the nucleus.

To solve **Eq. 4**, we utilized a Taylor series approximation of $\bar{\epsilon}$ around $\phi_{in,0}$:

$$\bar{\epsilon} \approx \epsilon\left(\vec{m}, \phi_{in,0}\right) + \frac{1}{2}\sigma_{\phi_{in}}^2 \frac{\partial \epsilon^2(\vec{m}, \phi_{in})}{\partial \phi_{in}^2}\Big|_{\phi_{in}=\phi_{in,0}} \tag{5}$$

where $\epsilon(\vec{m}, \phi_{in})$ is a non-monotonic function of ϕ_{in} due the competing effects of crowding on depletion interactions and molecular diffusion, and $\frac{\partial \epsilon^2(\vec{m},\phi_{in})}{\partial \phi_{in}^2}\Big|_{\phi_{in}=\phi_{in,0}} \approx -\sqrt{\epsilon(\vec{m}, \bar{\phi})\kappa}$ quantifies gene expression as a function of crowding within a transcriptional interaction volume. Expression rate $\kappa = 22.6\ \mathrm{nM/s}$ is derived from a steady-state solution of rate equations that model transcription and whose crowding-dependent rates were determined from BD and MC simulations as described previously (*14, 19*). Of note, the function $\frac{\partial \epsilon^2(\vec{m},\phi_{in})}{\partial \phi_{in}^2}\Big|_{\phi_{in}=\phi_{in,0}}(\epsilon(\vec{m}, \bar{\phi}))$ can be simulated by varying any or several of the components of $\vec{m}$. Although, in principle, the exact form of $\frac{\partial \epsilon^2(\vec{m},\phi_{in})}{\partial \phi_{in}^2}\Big|_{\phi_{in}=\phi_{in,0}}$ as a function of $\epsilon(\vec{m}, \bar{\phi})$ may depend on which component of $\vec{m}$ is being varied, i.e. $\kappa = \kappa(\vec{m})$, in practice κ is only weakly dependent on $\vec{m}$. In other words, $\frac{\partial \epsilon^2(\vec{m},\phi_{in})}{\partial \phi_{in}^2}\Big|_{\phi_{in}=\phi_{in,0}}$ depends on $\vec{m}$ primarily through $\epsilon(\vec{m}, \bar{\phi})$, with the average expression rate as the "common dominator" of multiple molecular factors. Thus, predictions of the CPMC model regarding the effects of physical regulators on ensemble gene expression should be robust to changes in molecular factors. Integrating **Eq. 1-5** the *Se* of expression rate becomes:

$$Se(\bar{\epsilon}, D_i) \cong \frac{1}{D_i}\ln N_d - \frac{1}{8}\frac{\kappa}{\bar{\epsilon}}(\sigma_{\phi_{in}}^2)^2\left(1 + \sqrt{1 + \frac{16}{(\sigma_{\phi_{in}}^2)^2}\frac{\bar{\epsilon}}{\kappa}}\right) \cdot \left[D_i\ \ln\left(\frac{r_{in}}{r_{min}}\right) + \frac{3 - D_i}{D_i}\frac{r_{min}}{r_{in}}L^{1/D_i}\ln(L)\right] \tag{6}$$

RESULTS

Physical factors of chromatin structure regulate sensitivity of gene expression to changes in chromatin packing scaling.

To first test the CPMC model predictions *in vitro*, we employed live cell PWS microscopy to measure D (**Fig. 2a&b**) (*38,39*) and ChromEM to measure $\phi_{in,0}$ (**Fig. 2c&d**) (*27*) paired with mRNA microarrays, RNA-Seq, and single-cell RNA sequencing (scRNA-seq) to measure gene expression of cell populations under different conditions. Specifically, average D was calculated by first averaging D values from PWS measurements within each cell nucleus and then averaging these measurements over the entire cell population for each treatment condition. Utilizing ChromEM, average chromatin density was measured within each nucleus with ~3 nm resolution. As $\phi_{in,0}$ represents the crowding contributions from both chromatin and mobile crowders within the nucleus, we added to CVC measured by ChromEM an additional 5% contribution from unbound macromolecules (as described in the **Materials and Methods** section). In addition, we utilized publicly available DNA sequencing information to obtain gene length and high-throughput chromatin conformation capture (Hi-C) data to approximate N_d from the size of topologically associating domains (TADs) (*35*). In relation to prior work on higher order chromatin organization, N_d could extend from tens of thousands to millions of basepairs. While N_d might not necessarily represent the organization observed in TADs, TAD size was utilized as an approximate measure of N_d as these domains have been shown to obey self-similar organization (*36*), as evidenced by power law scaling properties of contact probability within TADs (*37*). Combining these methods, we then tested the CPMC model's predictions of Se of gene expression against *in vitro* measurements for each identified physical regulator of gene expression.

To test the role of initial D_i, we performed an RNAi knockdown of the chromatin remodeling enzyme, Arid-1a (A-Kd) in human colon carcinoma HT-29 cells, which resulted in a lower D_i compared to wild-type (WT) cells (*17*). Next, we measured changes in chromatin packing scaling D in serum starved WT and A-Kd HT-29 cells before and 30 minutes after stimulation with 10% fetal bovine serum, 100nM epidermal growth factor (EGF), and 100nM phorbol 12-myristate 13-acetate (PMA) (*14*). In parallel, we measured gene expression for these conditions at 5 hours utilizing mRNA microarrays. Genes were grouped for WT and A-Kd cells separately based on their relative initial expression during serum starvation, and the experimentally measured sensitivity $\Delta lnE/\Delta lnD$ was calculated for each group of genes. As predicted by the CPMC model, experimental measurements of the Se of gene expression shows a bidirectional, monotonic responsiveness to D as a function of initial expression in HT-29 cells ($\phi_{in,0}$ ~ 39%, approximated by dividing chromosome copy number by nuclear volume). In addition, we found that D_i predominantly changes the responsiveness of initially under-expressed genes (**Fig. 2e&f**). These results indicate that populations of cells with a higher D would have a higher level of transcriptional divergence (the difference between highly and low expressed genes) than low D cells. Cancer cells across most malignancies, stem cells, and, especially, cancer stem cells, are all examples of types of cell populations that have elevated chromatin packing scaling (*2, 38*).

Functionally, this suggests that D can act as a means to optimize transcriptional responses as is explored in subsequent sections.

Next, we tested the effect of average nuclear crowding density, $\phi_{in,0}$, on gene expression sensitivity to changes in the chromatin packing scaling D. ChromEM was employed to measure average chromatin density for both human lung adenocarcinoma A549 cells and differentiated BJ fibroblast cells, which had mean chromatin volume concentration (CVC) of 0.35 and 0.30, respectively (**Fig. 2c&d**, distribution of CVC values shown in **Fig. S3**). Approximating for an additional space filling contribution from mobile crowders, estimates of $\phi_{in,0}$ were 40% in A549 and 35% in BJ cells. Each cell line was treated with 100nM dexamethasone (DXM) to modulate D, which was measured by PWS microscopy. Gene expression of both cell lines with and without DXM treatment was measured by RNA-seq. Sensitivity of gene expression was measured as described above for each cell line. Interestingly, the CPMC model predicts cells with a lower $\phi_{in,0}$ would had an attenuated bidirectional Se, an effect confirmed experimentally in the lower chromatin density BJ cells (**Fig. 2g**). In contrast, the higher chromatin density A549 cells (**Fig. 2h**) retain an asymmetric response to altered chromatin packing scaling. This suggests that cells with smaller nuclear volume, such as immune cells, or cells with increased chromosome copy number, such as malignant cancer cells, would be predisposed to produce a more pronounced bidirectional response in gene expression to stimuli that alter whole nuclear chromatin structure compared to cells with lower chromatin density. These results demonstrate the net effect of increasing D and $\phi_{in,0}$ is an increased transcriptional divergence between initially over- and under-expressed genes.

Finally, we tested the roles of N_d and gene length on Se. From our model, N_d determines the probability of genes being on an exposed surface to allow transcription reactions to occur, a relationship which depends non-linearly on D (**Eq. 2**). Consequently, the CPMC model predicts that (1) genes in larger packing domains (e.g. N_d>2Mbp) would be relatively under-expressed in comparison to those within smaller N_d domains (<50Kbp) and (2) genes within large N_d domains would be more likely to become enhanced as a function of increasing D (+Se). To test these predictions experimentally, we utilized the Arrowhead function in Juicer tools to measure TAD sizes from Hi-C data of untreated and DXM treated A549 cells (*39*). As the dissociation and formation of TADs has previously been shown to alter gene expression, for our analysis we only selected TADs that were unaltered with DXM treatment. The top 20% largest (~2Mbp) and bottom 30% smallest (~50Kbp) of these TADs were chosen to produce roughly equal sized groups of genes (~130 genes in each group). Using RNA-seq to measure gene expression and PWS microscopy to measure the change in D before and after DXM treatment, we analyzed the sensitivity of expression for genes localized to smaller 50Kbp TADs compared to larger 2Mbp TADs (**Fig. 2i**). As predicted from the CPMC model, *in vitro* results demonstrate that genes within larger 2Mbp TADs have an overall higher sensitivity to changes in D (**Fig. 2i**) while simultaneously having lower initial expression compared to those within smaller 50Kbp TADs. Consequently, these findings suggest a regulatory role of spatially confining genes into self-similar structures, such as those found in TADs, in determining the probability of a gene being exposed to transcriptional reactants. Given the recent work indicating significant variability in TADs from cell to cell, this would suggest yet another mechanism that cells can use to regulate their functional diversity within a population.

In addition, we tested the role of gene length on sensitivity of two fold under-expressed (low) and two fold over-expressed (high) genes in the serum starved WT HT-29 cells described above. Using the built-in *Mathematica* function, GenomeData, to obtain sequence length of genes, the sensitivity of gene expression to D was then calculated as a function of their length. The model predicts shorter genes have a smaller interaction volume, increasing the variance of crowding conditions these genes are exposed to. Consequently, an increase in D should further increase fluctuations in crowding concentrations surrounding these shorter genes, causing initially under-expressed genes to further reduce their expression in proportion to decreasing gene length L. However, genes with an initially higher expression level will be relatively unaffected by changes in gene length due to more optimal molecular characteristics (e.g. high TF and Pol-II concentration) and initial crowding conditions these genes are exposed to. In line with the CPMC model, our experimental microarray data demonstrates that shorter, initially under-expressed genes become disproportionately under-expressed as a function of increasing D, whereas length minimally influences initially over-expressed genes (**Fig. 2j**).

The scaling behavior of chromatin packing regulates phenotypic plasticity through transcriptional divergence and malleability

A major implication of the CPMC model is the role physical chromatin structure plays in shaping gene expression. Thus, the model could provide a mechanistic link between two aspects of phenotypic plasticity of a population of cells: transcriptional malleability and intercellular transcriptional heterogeneity. In this case, we can consider transcriptional malleability to be the average change in expression of a gene in response to an external stimulus, while transcriptional heterogeneity can be thought of as the range in expression levels of each gene across a cell population. While there is likely to be increased complexity that results from the variations from cell to cell in average density and D, we herein test how heterogeneity and malleability are influenced by the measurable features of disordered chromatin packing within a cell population. An ideal testbed for this mechanistic integration is cancer. Multiple lines of evidence have shown that chromatin structure is nearly universally transformed in malignancy (*40-43*). Microscale structural alterations in chromatin are currently the gold standard for histopathological diagnosis of dysplasia and malignancy (*40*). At the nanoscale, an increase in D has been previously reported to occur at pre-dysplastic stages of lung, colon, esophageal, ovarian, liver, prostate, and pancreatic cancers, while the severity of the chromatin transformation has been shown to be an accurate indicator of the tumor aggressiveness (*41*, *43*). Since (1) elevated D is a hallmark of malignancy, (2) there is an emergent role of intercellular heterogeneity in determining chemotherapeutic responsiveness and (3) cancer cells rapidly alter their gene expression to overcome cytotoxic stressors (*14, 16, 44*), we hypothesized that cancer cells could leverage physical transformation within the nucleus to gain survival advantages. Therefore, we wanted to test if cells could utilize the scaling of chromatin packing as a regulator of both transcriptional malleability and heterogeneity to achieve a rapid response to external stressors.

According to the CPMC model, the dependence of transcriptional malleability on chromatin packing scaling results from the observed asymmetric response of upregulated and downregulated

genes to changes in *D* (**Fig. 2**), which we denote as transcriptional divergence. Here, we focus on changes in gene expression due to an external stimulus. A transcriptional response of a cell to a chemotherapeutic stress provides a case in point. Chemotherapeutic induction of apoptosis has been shown to depend on the *rate of change* in expression of critical genes (e.g. p53) and not their steady-state levels alone (*45*). Accordingly, mechanisms which increase the rate of upregulation of these critical genes would facilitate the development of cellular resilience to stressors. Consider two populations of cells that have a baseline difference in their initial *D.* These two populations are then exposed to the same exogenous stressor and a series of stress signaling pathways are activated in an attempt to overcome the perturbation. The cells' survival now depends, in part, on the increased expression of these genes within a critical time frame. The CPMC model predicts that the population of cells with initially higher *D* will be more likely to upregulate these critical genes and remain viable (**Fig. 3a**).

To quantify the effect of initial *D* on transcriptional responsiveness, let $mRNA_{1,a}$ be the initial expression (the number of mRNA transcripts) for a given gene in cell *a* with chromatin packing state D_a. At time point $t = 0$, a stimulus produces an increase in the gene's rate of expression from $E_{1,a}$ to $E_{2,a}$. Without loss of generality, we first assume that both expression rate $E_{2,a}$ remains stable and that the rate of mRNA degradation, *v*, remains constant post stimulation. The relative change in expression at time *t* is $\left(mRNA_a(t) - mRNA_{1,a}\right)/mRNA_{1,a} = \left(E_{2,a}/E_{1,a} - 1\right)(1 - e^{-\nu t})$, where $mRNA_{1,a} = E_{1,a}/\nu$ is the pre-stimulation steady-state expression. This relative change in expression increases with the ratio $E_{2,a}/E_{1,a}$, which is itself a function of both molecular features and the chromatin packing state surrounding the gene. This can be illustrated by comparing the response of an individual gene to an exogenous stressor in two cells *a* and *b*. Let the same gene in both cells be associated with similar molecular features ($\vec{m}_{i,a} = \vec{m}_{i,b}, i = 1,2$) but different chromatin packing states D_a and D_b, with $D_b > D_a$. From **Eq.4**, $\frac{\mathrm{d}E}{E} = \frac{\mathrm{Se}(D)}{D}\mathrm{d}D$, it follows that:

$$\mathrm{E}_{i,b} = \mathrm{E}_{i,a} \exp\left[\int_{D_a}^{D_b} \frac{Se_i(D')}{D'} dD'\right],\ i = 1,2 \tag{7}$$

where $Se_i(D)$ is the sensitivity of expression state $\mathrm{E}_{i,a}$. In this situation, the effect of *D* on relative changes in transcription in cell *b* compared to cell *a* would be defined as:

$$\delta = \left(\frac{\mathrm{E}_{2,b}}{\mathrm{E}_{1,b}}\right)\Big/\left(\frac{\mathrm{E}_{2,a}}{\mathrm{E}_{1,a}}\right) = \exp\left[\int_{D_a}^{D_b} \frac{\mathrm{Se}_2(\mathrm{D}\prime) - \mathrm{Se}_1(\mathrm{D}\prime)}{D\prime}\mathrm{d}\mathrm{D}'\right]. \tag{8}$$

Within the physiological range of transcription, *Se* is an increasing function of *E* (**Fig. 2**) and, as $E_2 > E_1$ for both cells, $\delta > 1$. Consequently, the same stimulus will result in enhanced upregulation of the same gene in cell *b* compared to cell *a*, driven by the differences in chromatin packing scaling between the two cells. This effect is expected to be particularly pronounced for initially under-expressed genes with $Se_1 < 0$ that undergo a significant amplification ($Se_2 > 0$) upon stimulation. We see that δ is directly related to the transcriptional divergence and the shape of the function *Se*(*E*) (**Fig. 2**). A faster rise of *Se* as a function of *E* results in a higher δ. For cells *a* and

b with similar D, $\delta \approx 1 + (\mathrm{Se}_2 - \mathrm{Se}_1)(D_b - D_a)/D_a$. This implies that factors that tend to increase transcriptional divergence (e.g. high D, crowding, small N_d) would be expected to result in a higher transcriptional malleability.

The functional significance of the relative transcriptional malleability coefficient δ is twofold. First, for highly amplified genes ($E_2/E_1 \gg 1$) the relative increase in transcription at any given time after the stimulation is proportional to δ:

$$\left([mRNA]_b(t) - [mRNA]_{1,b}\right)/[mRNA]_{1,b} \approx \delta\left([mRNA]_a(t) - [mRNA]_{1,a}\right)/[mRNA]_{1,a}. \quad (9)$$

Second, the time τ required to reach a given level of expression is dependent on chromatin packing scaling and inversely proportional to δ, i.e. $\tau_b/\tau_a \approx \delta^{-1}$. This conclusion is applicable to genes that are both upregulated as well as those that are downregulated in response to a stimulus, an effect that might be especially consequential if decisions regarding cell fates must be made within a limited time period after the introduction of the stressor (*45*).

To experimentally explore the relationship between D and phenotypic plasticity, we performed concurrent single cell RNA sequencing and live cell PWS microscopy experiments on A2780 ovarian adenocarcinoma cells in response to treatment conditions that modulate chromatin packing scaling. We first tested whether chemotherapy treatment of cancer cells resulted in a pre-selection of high D cells. We measured changes in D using live cell PWS in A2780 ovarian adenocarcinoma cells before and after treatment with a chemotherapeutic agent, 5nM paclitaxel, for 48 hours. We also monitored cell coverage, which represents survival of a cell population. Defining high D cells as those that fall within the top 25th percentile of D in the cell population prior to the PAC treatment (D=2.47), we then measured the percentage of cells with high D at 48 hours after paclitaxel treatment. We observed that the percentage of high D cells increased in paclitaxel-treated cells compared to the control population (**Fig. 3b**). In combination with coverage measurements, which demonstrated significant cell death after 48 hours of paclitaxel treatment, our results indicate that high D cells have an increased survival rate when exposed to paclitaxel treatment (**Fig. 3b&c**).

We then compared the transcriptional malleability of populations of cells with differential D. As a model system, we relied on chemically-induced modulation in D. To reduce D, we treated A2780 cells with 75 μM celecoxib (CBX), a nonsteroidal anti-inflammatory agent for 16 hours. Previously, we have found that celecoxib reduces D within 30 minutes of treatment in A2780 ovarian carcinoma cells by at least 8% compared to untreated cells (*14*). As a model of high-D cells, we used untreated A2780 cells. Both CBX-treated cells (low D) and untreated cells (high D) were then exposed to a chemotherapeutic agent, 5nM paclitaxel (PAC) for 16 or 48 hours. Single cell RNA sequencing was conducted using Illumina NextSeq500. Raw reads were aligned, mapped and used to calculate transcripts per million (TPM) for each condition using bowtie2 (*46*) and RSEM (*47*). Thus, as a model system, we measured transcriptional perturbation induced by a cytotoxic chemotherapy stressor in a lower D (celecoxib-treated) versus higher D (not treated by celecoxib) A2780 cell populations.

Inputting the experimentally observed difference in D into the CPMC model, we estimated $\delta > 4$ for initially under-expressed genes that become activated (**Fig. 3d**, blue manifold) and a smaller

increase in δ for initially over-expressed genes that are upregulated in response to stimulation (**Fig. 3d**, red manifold). We then tested if these predicted trends are observed experimentally using single-cell RNA sequencing. Importantly, the crucial window for response to chemotherapy frequently is thought to occur within 24 hours (*45*). Thus, we compared changes in gene expression in A2780 cells with initially higher *D* to initially lower *D* after stimulation by paclitaxel treatment for 16 hours. In agreement with the CPMC model, the stimulation of initially under-expressed genes by chemotherapy treatment in initially higher *D* cells (upregulation of expression rate from control rate $E_{1,b}$ to 16hr PAC-treated rate $E_{2,b}$) was much higher than that in lower *D* cells (from CBX-treated rate $E_{1,a}$ to 16hr combo rate $E_{2,a}$), resulting in $\delta \sim 4$ (**Fig. 3e**). Likewise, a similar but mitigated effect was observed in initially over-expressed genes (**Fig. 3e**), in strong agreement with the model predictions. Next, we tested whether these trends were independent of cell line and compound. We performed parallel experiments using propranolol as a *D* lowering agent in A2780 cells and celecoxib and propranolol to decrease *D* in more malignant TP53 mutant A2780 (M248) cells. These additional conditions demonstrated a similar effect of *D* on transcriptional malleability in response to paclitaxel stimulation of high *D* compared to low *D* cells (**Fig. S4**). Finally, we tested if observed effect of chromatin packing scaling influences genes specifically involved in functionally relevant stress response pathways. We first identified differentially expressed genes that, on average, increased their expression at least two fold in A2780 cells treated with paclitaxel for 48 hours compared to control cells. Gene ontology analysis of these upregulated genes showed the activation of multiple stress response pathways after stimulation by paclitaxel treatment, including DNA repair, autophagy, cell cycle arrest, and apoptosis (p-value < 0.05, **Fig. 3f**, **Fig. S5**). The effect of *D* on the activation of these established stress response genes was consistent with that observed in all upregulated genes, with δ as high as ~ 4 (**Fig. 3g**).

The scaling behavior of chromatin packing regulates phenotypic plasticity through intercellular transcriptional heterogeneity

Another key aspect of phenotypic plasticity that can be modulated by the disordered packing of chromatin is transcriptional heterogeneity, or the range of expression levels across genes exposed to similar molecular conditions. The CPMC model predicts that transcriptional heterogeneity increases as a function of *D* due to increased variations in both packing density ($\sigma^2_{\phi_{in}}$) and gene accessibility (p_g). To quantify this effect from the CPMC model, the variance in ϵ across any given cell population, Var_ϵ, is (*14*):

$$\mathrm{Var}_\epsilon \approx \frac{1}{2}\left(\left.\frac{\partial \epsilon^2(\vec{m}, \phi_{in})}{\partial \phi_{in}^2}\right|_{\phi_{in}=\phi_{in,0}}\right)^2 {\sigma_{\phi_{in}}}^4 \tag{10}$$

Consequently, intercellular transcriptional heterogeneity, i.e. the standard derivation of steady-state expression rate *E* in **Eq. 3**, becomes:

$$H(D) = p_g \cdot {\mathrm{Var}_\epsilon}^{1/2} \approx \frac{\sqrt{2}}{8} p_g \cdot (\sigma^2_{\phi_{in}})^2 \kappa \left(1 + \sqrt{1 + \frac{16}{(\sigma^2_{\phi_{in}})^2}\frac{\bar{\epsilon}}{\kappa}}\right) \tag{11}$$

and the coefficient of variation (the ratio of the standard deviation to the mean expression) is $COV(D) = \frac{\sqrt{2}}{8}(\sigma_{\phi_{in}}^2)^2 \frac{\kappa}{\bar{\epsilon}}\left(1 + \sqrt{1 + \frac{16}{(\sigma_{\phi_{in}}^2)^2}\frac{\bar{\epsilon}}{\kappa}}\right)$. Both *H* and *COV* increase with *D*, and *COV* also decreases as a function of expression.

To investigate the association between *D* and intercellular transcriptional heterogeneity, we analyzed our scRNA-seq data to quantify the spread in transcriptional states across each treatment condition. Focusing on overall transcriptional differences between cells within the same condition provides better validation to the model than analyzing the spread of all observed genes. Thus, we first used t-Distributed Stochastic Neighbor Embedding (t-SNE) combined with principal component analysis (PCA) to reduce the dimensionality of the system on all cells simultaneously (*48*). The dimensionality reduction mapped each cell onto a three-dimensional projection. Distances between cells in 3D space represented overall differences in transcriptional states, as has been described by van der Maaten and Hinton (*48*). Intercellular transcriptional heterogeneity for each cell population was quantified by the average radius of the cluster of cells, $R_c = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_{mean})^2}$ where r_i is the position of each cell in the reduced spaced, *N* is the total number of cells in each treatment group, and $r_{mean} = \frac{1}{N}\sum_{i=1}^{N} r_i$. Intuitively, R_c can be thought of as the radius of relative genomic space. Consistent with predictions of the CPMC model, we found that transcriptional heterogeneity, as measured by the radius of genomic space, increases with *D* in response to paclitaxel treatment, which preselects for high *D* cells, as shown above. Notably, after 48 hours of paclitaxel treatment, the population of surviving cells had both higher *D* and increased transcriptional heterogeneity compared to control cells (**Fig. 4a-c&f**). In contrast, celecoxib treatment reduces average *D* of a cancer cell population. Accordingly, cells treated with celecoxib for 16 hours had a lower transcriptional heterogeneity compared to control cells. In addition, when these celecoxib-primed cells with initially lower *D* were treated with paclitaxel for 16 hours, they had a decreased transcriptional heterogeneity compared to paclitaxel-treated control cells (**Fig. 4d-f**). Although the resulting projection from t-SNE is non-unique, the trends in the radius of genomic space across conditions are robust to randomly selected choice of seed (**Fig. S6)**. Additional analyses quantifying the Euclidean distance between expression of DNA repair genes upregulated in 48 hour paclitaxel treatment as well as the coefficient of variation of expression between cells in the same treatment condition demonstrate the same effect of chromatin packing scaling on transcriptional heterogeneity as the t-SNE results (**Fig. S7**).

Next, we sought to investigate the effect of chromatin packing scaling on changes in transcriptional heterogeneity in response to stimulation. For higher *D* compared to lower *D* populations, the CPMC model predicts an increase in transcriptional malleability concomitant with an increase in gene expression variability in response to stimulation. As a case in point, consider the upregulation of stress response genes due to a stressor such as chemotherapy. Both transcriptional malleability and heterogeneity may facilitate a response to the stress. An increase in the average expression (malleability) and in the standard deviation of expression levels (heterogeneity) for these genes upon the stimulation would increase the percentage of cells that express these genes above any given level that may facilitate cell survival, regardless of the exact value of this critical level. We used scRNA-seq data on A2780 cells to analyze the distributions of transcriptional responses to paclitaxel treatment, as an example of an exogenous stressor, in cell populations with different initial *D*. We assessed the ratio of the upregulated expression rate due to the stressor versus the

initial expression rate (relative upregulation, E_2/E_1). Focusing on transcriptional responsiveness of genes associated with DNA repair pathways that had been upregulated in response to 48 hour paclitaxel treatment, we found that the higher *D* population had both an increase in the average (malleability) and the variance of relative upregulation compared to those for the lower *D* population (**Fig. 4g**). Next, we examined relative expression levels of genes suppressed in the control condition, specifically those that occupied the bottom 10^{th} percentile of gene expression, and observed a similar behavior (**Fig. 4h**). Importantly, this *D*-dependent increase in intercellular transcriptional heterogeneity was itself a function of expression levels, i.e. genes that were under-expressed in cells prior to PAC treatment had a more significant difference in the heterogeneity of their relative upregulation in the high compared to the low *D* populations than those that were already highly expressed prior to the stimulation (**Fig. 4i**), also in agreement with the CPMC predictions.

Transcriptional divergence is inversely associated with patient survival

As described above, *D* determines a cell's responsiveness to stressors such as chemotherapeutic agents through the effect of chromatin packing scaling on phenotypic plasticity. A logical next step was to establish if these physical regulators play a role in tumor aggressiveness *in vivo*. The effects of chromatin packing scaling on phenotypic plasticity may foster the ability of cancer cells to develop resilience and/or resistance to chemotherapy *in vivo* and may also be involved in other processes fostering increased tumor fitness and aggressiveness. Throughout carcinogenesis, tumors are frequently exposed to a wide range of stressors including attack by a host's immune system, inadequate oxygen supply from nearby blood vessels, or an acidotic microenvironment. To test if such a relationship between phenotypic plasticity and tumor fitness exists, we analyzed publicly available RNA-seq data collected by the TCGA Research Network (*20*) for lung, colorectal, and breast cancers, which are the three most prevalent malignancies in the United States. As the model predicts cellular responsiveness to external stressors, of which chemotherapy is an example, we focus on patients presenting with Stage III and IV tumors at time of diagnosis, as systemic therapy is the standard of care for these patients. Using the *R* package, *TCGAbiolinks* (*49*), we quantified gene expression in units of fragments per kilobase million (FPKM) for each patient. As this data lacks initial control measurements of cancer cells prior to initiation of systemic therapy, transcriptional malleability cannot be measured directly for each patient. Additionally, we do not have information related to chromatin packing scaling and other physical regulators of transcription for these patients. However, the essence of the effect of δ is that elevated *D* amplifies a gene's transcriptional response to stimuli: over-expressed genes are enhanced whereas under-expressed genes are suppressed (**Fig. 2**). Consequently, as the bidirectional behavior of $Se(E)$ curves indicates, an elevated *D* widens the distribution of gene expression resulting in increased transcriptional divergence, which, in turn, is a key determinant of transcriptional malleability (**Fig. 5a**). Thus, quantifying transcriptional divergence within these patient cohorts will, by proxy, measure transcriptional responsiveness, which we have shown above is linked to *D*. Borrowing a method from macroeconomics, transcriptional divergence can be quantified by the ratio of expression of the top 50% of genes and the bottom 50% of genes (P50/P50), for ranked expression, E_k, and total number of detected genes, N :

$$\frac{P50}{P50} = \frac{\sum_{k=\frac{N}{2}+1}^{k=N} E_k}{\sum_{k=1}^{k=\frac{N}{2}} E_k} \quad (12)$$

As age is also a major predictor of cancer mortality, we restricted our analysis to patients under 75 years at time of diagnosis. As CPMC model predicts that higher transcriptional divergence produces more adaptable tumor cells, we would expect that patients with a shorter survival time would have tumor cells with an elevated P50/P50 ratio at time of diagnosis. To test this hypothesis, we compared the P50/P50 ratio calculated at time of diagnosis of patients surviving above or below the median survival time for each cancer type (**Fig. 5b**). We found a statistically significant inverse relationship between P50/P50 ratio and relative patient survival time for lung (**Fig. 5b**, $p<0.021$), breast (**Fig. 5b**, $p<0.0001$), and colon (**Fig. 5b**, $p=0.018$) cancers.

Next, we analyzed the relative contribution of transcriptional divergence to patient survival time compared to effects of other prognostic factors (e.g. demographic factors, tumor molecular subtype, and stage) by performing a multivariate regression on each prognostic factor. We then calculated the relative survival time (RST) for each patient as the observed survival time relative to the expected survival time based on these other prognostic factors. RST < 1 indicates that a patient's survival is shorter than expected (e.g. RST = 0.5 indicates that their survival duration is 50% shorter than expected) whereas RST > 1 indicates the opposite. Patients were then grouped into a high and a low P50/P50 cohort based on if they were in the top or bottom half of P50/P50 values, respectively. Notably, high P50/P50 patients had an RST below 0.8 for all malignancies whereas a low P50/P50 translated into a significantly higher RST > 1 ($p < 0.05$). Next, we analyzed the relationship between patient survival and P50/P50 directly for all malignancies. As survival depends on a multitude of factors, some of which were not available within the TCGA dataset for all patients (e.g. comorbidities), a fixed moving window average was applied to the data (see **Methods** for details). We found a continuous inverse trend between P50/P50 and patient survival for all three malignancies (**Fig. 5d, Fig. S8)**. Finally, Kaplan-Meier survival curves show that patients with high P50/P50 ratios have a median survival of 8 months compared to 28 months for those with a low P50/P50 (**Fig. 5e,** $p=0.01$). In summary, these results support a strong correlation between transcriptional divergence, a facet of phenotypic plasticity that is directly affected by chromatin packing scaling, and patient survival (**Fig. 5**).

DISCUSSION

In this work, we combined multi-scale modeling with high-throughput chromatin conformation capture, single cell RNA sequencing, chromatin electron tomography, and live cell Partial-Wave Spectroscopic microscopy to demonstrate the role of the disordered chromatin polymer on regulating both intercellular transcriptional heterogeneity and transcriptional malleability. Based on predictions from the CPMC model, which were verified experimentally, the spatial arrangement of chromatin packing affects gene expression through a number of physical regulators, including $\phi_{in,0}$, N_d, and D (**Fig. 1&2**). We demonstrate, both computationally and experimentally, that a crucial role of chromatin packing is to determine the level of phenotypic plasticity within a cell

population. In particular, the scaling of chromatin packing, D, modulates both the transcriptional malleability through a chromatin mediated enhancement δ, a "tailwind effect" (**Fig. 3**), and the level of intercellular transcriptional heterogeneity (**Fig. 4**). This effect is further regulated by other physical properties of chromatin. A higher average crowding density within the nucleus suppresses the expression of initially under-expressed genes as D increases (**Fig. 2g&h**). The modulatory effects of N_d are two-fold. Genes localized to domains with a large N_d (Mbp range) are more suppressed than those localized to domains with small N_d (kbp range) owing to the reduced accessibility to transcription factors and polymerases. However, as D increases, expression of genes associated with large N_d is disproportionately enhanced (**Fig. 2i**). Overall, higher D, higher crowding, and lower N_d increase both transcriptional malleability and heterogeneity, with D having a much larger effect compared to the other two chromatin packing properties.

The fact that eukaryotic cells have encoded information into the scaling behavior of chromatin packing may have important medical implications. Elevated D is a hallmark of cancer cells and could represent a mechanism by which malignancy gains non-mutational advantages over neighboring healthy cells. As observed *in vitro*, treating cells with a chemotherapeutic agent such as paclitaxel selects for cells with a higher D (**Fig. 3b**), which, as demonstrated within this work, is in part due to the increased phenotypic plasticity compared to cells with a lower D (**Fig. 3&4**). This selects for tumor cell populations with a higher transcriptional adaptive potential, which in turn may facilitate their survival despite future exposure to new stressors. In support of this potential mechanism, our data shows that transcriptional divergence, the cross-sectional measurement of transcriptional malleability, in advanced colorectal, lung, and breast cancers is associated with worse prognosis independent of demographic factors (e.g. age, gender), tumor stage, and molecular transformations (**Fig. 5**).

At present, experimental validation of the CPMC model relies on the measurement of average chromatin packing scaling D and crowding within the entire nucleus. While currently beyond existing experimental capabilities, subsequent studies directly comparing how local (e.g. intra-packing domain) chromatin structure affects transcriptional processes and output would be of considerable importance. Pairing gene-tracking techniques such as CRISPRainbow with imaging modalities that measure chromatin structure, such as live-cell PWS microscopy and ChromEM, as well as super-resolution imaging of molecular factors would help elucidate how intranuclear variations in molecular and physical regulators of transcription contribute to transcriptional heterogeneity and malleability (*12, 27, 50*).

Although not explored in this work, there are several implications of these results on the understanding of multicellular fitness in the context of cell biology. For example, the localization of genes into domains has been demonstrated to be a conserved, albeit heterogeneous, process (*51*). From the predictions of the model, cells would benefit from localizing genes into large domains that are intended to be suppressed at baseline but need rapid amplification if conditions change. Likewise, crowding density could be adjusted by cells either as a preprogrammed response by changing nuclear volume or incidentally from the retention of an extra chromosome during replication. Consequently, as has been hypothesized, this could be a mechanism linking nuclear size and density (e.g. hyperchromasia) with differential gene expression. Interestingly, nuclear size, hyperchromasia, and abnormal nuclear texture are some of the most ubiquitous histological

markers of neoplasia, although their etiology and functional consequences have been poorly understood (*52)*.

In light of the CPMC model conclusions, it should be clear that disordered chromatin packing does not mean that the configurations are random or that observed patterns in gene transcription are the result of configurational noise. While it is beyond the scope of this work, the conformation of a chromatin polymer depends on the balance between chromatin-chromatin and chromatin-nucleoplasm interactions and is further shaped by active chromatin loop formation processes and other constraints such as nuclear lamins (*53*). The shape of the disordered chromatin polymer will ultimately depend on molecular factors such as histone modifications, transcriptional and replication induced supercoiling, DNA motif stiffness, as well as nucleoplasmic factors such as nuclear pH, ionic concentrations, and crowding, which collectively alter chromatin-chromatin and chromatin-nucleoplasm interactions (*26, 54-56*). Therefore, individual cells could utilize a combination of chromatin-chromatin and chromatin-nucleoplasm interactions to appropriately organize the genome while also encoding a pre-determined level of phenotypic plasticity.

In addition, this work may have implications on the open question in chromatin biology regarding the importance of non-coding DNA. Some roles have since been illuminated, including the production of non-coding RNA and the distribution of transcriptional regulatory motifs such as enhancers and insulators (*21, 57*). In light of this work, and in relation to previously suggested hypotheses of the role of macromolecular crowding on gene expression, one of the evolutionary functions of non-coding DNA could be derived from its space-filling role. Consequently, non-coding DNA might be a critical component within the genome to determine phenotypic plasticity as it contains the ability to modulate transcription reactions by influencing the free-energy of these reactions and the diffusion of reactants.

Finally, one could consider how *D* plays a role in the adaptability of cancer cells throughout carcinogenesis. Carcinogenesis depends on cells overcoming aberrations in metabolism, derangements of the microenvironment, inadequate vascular supply, immune surveillance, and acclimation to distal tissue environments during metastasis. As it could take multiple replicative generations to develop a new useful mutation within a population for each of these processes, cancer cells could in addition leverage the physical properties of chromatin packing to increase their transcriptional plasticity in order to acclimate to these conditions over a faster time scale. Thus, it may be worth investigating, for example, whether cancer cells with elevated *D* are better able to survive an immune response and acclimate to distant tissue sites during metastasis. From the therapeutic standpoint, while mutations are difficult to remove from a cell population, this work suggests that limiting cancer cell evolution might be possible pharmacologically by lowering the scaling of disordered chromatin packing.

Materials and Methods:

Gene expression analysis

mRNA Microarray for HT-29 Cells

HT-29 cells were serum deprived for 5 hours 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. 2445 differentially expressed genes were identified for subsequent analysis.

RNA-Seq for A549 and BJ Cells

RNA-seq data for A549 and BJ cells as downloaded from ENCODE and GEO with access codes ENCSR897XFT for A549 cells and GSE81087 for BJ cells (*35*, *58*). 4 replicates are included in both control and 12 hour DXM treated A549 cells. Gene expression was quantified using featureCounts were downloaded from GEO for A549 cells. The length and the counts for each replicate from featureCounts outputs were then changed into Transcripts Per Kilobase Million (TPM) using $\mathrm{TPM}_i = 10^6(\mathrm{cts}_i/\mathrm{L}_i)/\sum_i\left(\frac{\mathrm{cts}_i}{\mathrm{L}_i}\right)$, where TPM_i, cts_i and L_i are the TPM value, the count and the length of gene *i*. The differential expression (DE) analysis for A549 cells was performed using the *DESeq2* packages in *R*. 2292 differentially expressed genes were found after 12 hours 100 nM DXM treatment in A549 cells were 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 $\mathrm{TPM}_i = 10^6\mathrm{FPKM}_i/\sum_i \mathrm{FPKM}_i$, where FPKM_i is the FPKM value of gene *i*. The same differential expression method was used on BJ cells and 7601 genes were identified with p-value <0.01.

RNA-Seq for A2780.M248 Cells

RNA samples from ovarian carcinoma TP53 mutant clone A2780.M248 cells were collected from the cells treated under control, 16 hours celecoxib, 16 hours paclitaxel, 16 hours paclitaxel plus celecoxib and 48 hours paclitaxel conditions with 3 biological replicates per condition. The stranded mRNA-seq was conducted in the Northwestern University NUSeq Core Facility. Briefly, total RNA quantity was determined with Qubit fluorometer, and quality assessed using RINs generated from Agilent Bioanalyzer 2100. To proceed to sequencing library prep, RIN must be at least 7. The Illumina TruSeq Stranded mRNA Library Preparation Kit was used to prepare sequencing libraries from 100 ng RNA. The Kit procedure was performed without modifications. This procedure includes mRNA purification and fragmentation, cDNA synthesis, 3' end adenylation, Illumina adapter ligation, library PCR amplification and validation. lllumina HiSeq 4000 Sequencer was used to sequence the libraries with the production of single-end, 50 bp reads. Single-end FASTQ reads from RNA-seq measurements were aligned and mapped to hg38 using bowtie2. Transcriptions per million (TPM) from mapped reads were estimated using RSEM. Sgnificant genes that are expressed across all conditions and have fold changes larger/smaller than 2-fold/$\frac{1}{2}$-fold of control in cells treated with paclitaxel for 48 hours are selected.

Single cell RNA sequencing for A2780 Cells

The single cell RNA sequencing experiments on A2780 were conducted under Illumina NextSeq 500 platform by University of Illinois at Chicago Research Resources Center Cores using Smart-seq protocol. 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. 46 out of 57 control

cells, 55 out of 58 16hr paclitaxel cells, 53 out of 62 48hr paclitaxel cells, 62 out of 67 16hr celecoxib cells and 59 out of 59 16hr combination (paclitaxel + celecoxib) cells were selected after quality control (excluding cells with less than 4000 genes expressed). Additional quality control was performed using the expression level housekeeping genes (*59*), but no additional cells were excluded. In total, 8415 genes were identified for subsequent analysis for each individual cell after removing 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 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 (*48*). The t-SNE analysis was done using 'Rtsne' package in *R* with initial PCA step performed.

Gene Ontology Analysis

To perform the gene ontology 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. 20 biological processes were shown to be significantly involved by these over-expressed genes (**Fig. S5**). Out of the 20 upregulated biological processes, 11 of them are involved in DNA repair (**Fig. 3d**).

TCGA patients expression analysis

The P50/P50 ratio of each patient's gene expression is calculated by: $\frac{P50}{P50} = \frac{\sum_{k=\frac{N}{2}+1}^{k=N} FPKM_k}{\sum_{k=0}^{k=N/2} FPKM_k}$ where $FPKM_k$ is the sorted FKPM value of the transcribed genes in each patient and N is the total number of measurably transcribed genes. Only transcribed genes (FKPM≠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 the median survival time based on their vital status and survival time after diagnosis. 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) were pooled together to apply a fixed moving window average (MWA) with 15 patients per group to analyze if an overall trend exists between P50/P50 and survival time (days). 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 present in the data set). A linear regression analysis using survival duration, P50/P50, and tumor stage as survival predictors was also conducted using Python showing a significant prediction of patient survival only for P50/P50 (p-value<0.05) with negative coefficient of -33.6day. Notably, regression analysis did not show a strong predictive power of stage at time of diagnosis (p-value>0.05) or an association between tumor stage and P50/P50 level.

Hi-C Topologically Associating Domain (TAD) Analysis

The total mass of chromatin at the upper length scale of self-similarity N_d of genes in the 3D space was estimated using the publicly available high throughput chromatin conformation (Hi-C) data on A549 cells (GEO access code: GSE92819 for control cells and GSE92811 for cells treated with DXM for 12 hours) (*35*). N_d 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 surrounding differentially expressed genes. Genes localized

within the same TAD were assigned with the same N_d. As dissolution of TADs was previously shown to alter access of transcription factors to DNA and we wanted to analyze the effect of N_d size, we selected only TADs that remained intact and of comparable size before and after DXM treatments. Genes within these consistent TADs were divided into two cohorts: a high N_d group and low N_d group. Each group had ~130 genes and the average N_d for each group were approximately 50Kbp for low the N_d group and 2Mbp for high N_d group. Genes with top 5% and bottom 5% N_d were removed from each groups to exclude outliers.

Live cell Partial Wave Spectroscopic (PWS) Microscopy

HT-29 Cell Culture

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℃ and 5% CO_2. All of 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℃ with 5% CO_2.

A2780 Cell Culture

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 Cell Culture

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. 16000044) and 100 µg/mL Penicillin-Streptomycin (ThermoFisher Scientific, Waltham, MA, # 15140122). All cells were maintained and imaged at physiological conditions (5% CO_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 100nm Dexamethasone (Sigma-Aldrich, St. Louis, MO, D6645) for 12 and 32 hours, respectively, in line with published chromatin conformation capture and RNA-seq experiments.

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 (*33, 34*). Briefly, PWS measures the spectral interference resulting from internal light scattering structures within the cell, which captures the mass density distribution. To obtain the interference signal directly related to refractive index fluctuations in the cell, we normalize measurements by the reflectance of the glass-media interface, i.e. to an independent reference measurement acquired in an area without cells. PWS measures a data cube (spatial coordinates of a location within a cell and the light interference spectrum recorded from this location). The data cube then allows to measure spectral standard deviation (Σ), which is related to the spatial variations of refractive index within a given coherence volume. The coherence volume is defined by the spatial coherence in the transverse directions (~200 nm) and the depth of field in the axial direction (~1 μm). In turn, the spatial variations of refractive index depend on the local auto-correlation function (ACF) of the chromatin refractive index. Finite-difference time-domain (FDTD) simulations have shown that PWS is sensitive to ACF within 20 nm to 200 nm range. According to the Gladstone-Dale equation, refractive index is a linear function of local molecular crowding. Therefore, Σ depends on the ACF of the media's macromolecular mass density. Small molecules and other mobile crowders within the nucleus are below the limit of sensitivity of PWS, and PWS is primarily sensitive to chromatin conformation above the level of the nucleosome. To convert Σ for a given location within a nucleus to mass fractal dimension D, we model ACF as a power-law $B_\varphi(r) = \sigma_\varphi^2 \left(\frac{r}{r_{min}}\right)^{D-3}$, where σ_φ^2 is the variance of chromatin volume concentration (*60*). Generally, Σ is a sigmoidal function of D. However, for fractal structures such as a chromatin packing domain where within physiological range $2 < D < 3$, Σ can be approximated as a linear function of D by the relationship $D \approx D_0 + a\Sigma$, where $D_0 = 1.473$ and is comparable to the minimal fractal dimension that an unconstrained polymer can attain and constant $a \sim 7.6$. The measured change in chromatin packing scaling between treatment condition was quantified by first averaging D within each cell's nucleus and then averaging nuclei from over 50 cells per condition.

Chromatin Electron Microscopy (ChromEM)

A549 Cell Culture

Two cells 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 37ºC at 5% CO_2. Confluency of around 60% were reached for all experiments.

EM Sample Preparation and TEM/STEM Data Collection

For EM experiment, all the cells were prepared by the ChromEM staining protocol and embedded in Durcupan resin (EMS) (*27*). 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.

References:

1. P. Rathert, M. Roth, T. Neumann, F. Muerdter, J. S. Roe, M. Muhar, S. Deswal, S. Cerny-Reiterer, B. Peter, J. Jude, T. Hoffmann, L. M. Boryn, E. Axelsson, N. Schweifer, U. Tontsch-Grunt, L. E. Dow, D. Gianni, M. Pearson, P. Valent, A. Stark, N. Kraut, C. R. Vakoc, J. Zuber, Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543-547 (2015).
2. S. Stern, T. Dror, E. Stolovicki, N. Brenner, E. Braun, Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol Syst Biol* 3, 106-106 (2007).
3. L. W. Plasschaert, R. Žilionis, R. Choo-Wing, V. Savova, J. Knehr, G. Roma, A. M. Klein, A. B. Jaffe, A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018).
4. Q. H. Nguyen, N. Pervolarakis, K. Blake, D. Ma, R. T. Davis, N. James, A. T. Phung, E. Willey, R. Kumar, E. Jabart, I. Driver, J. Rock, A. Goga, S. A. Khan, D. A. Lawson, Z. Werb, K. Kessenbrock, Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nature Communications* **9**, 2028 (2018).
5. A. L. Haber, M. Biton, N. Rogel, R. H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T. M. Delorey, M. R. Howitt, Y. Katz, I. Tirosh, S. Beyaz, D. Dionne, M. Zhang, R. Raychowdhury, W. S. Garrett, O. Rozenblatt-Rosen, H. N. Shi, O. Yilmaz, R. J. Xavier, A. Regev, A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333 (2017).
6. D. A. Lawson, K. Kessenbrock, R. T. Davis, N. Pervolarakis, Z. Werb, Tumour heterogeneity and metastasis at single-cell resolution. *Nature Cell Biology* **20**, 1349-1360 (2018).
7. U. Ben-David, B. Siranosian, G. Ha, H. Tang, Y. Oren, K. Hinohara, C. A. Strathdee, J. Dempster, N. J. Lyons, R. Burns, A. Nag, G. Kugener, B. Cimini, P. Tsvetkov, Y. E. Maruvka, R. O'Rourke, A. Garrity, A. A. Tubelli, P. Bandopadhayay, A. Tsherniak, F. Vazquez, B. Wong, C. Birger, M. Ghandi, A. R. Thorner, J. A. Bittker, M. Meyerson, G. Getz, R. Beroukhim, T. R. Golub, Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325-330 (2018).
8. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, 2nd, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).

9. H. Easwaran, H. C. Tsai, S. B. Baylin, Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Molecular cell* **54**, 716-727 (2014).
10. A. Kleppe, F. Albregtsen, L. Vlatkovic, M. Pradhan, B. Nielsen, T. S. Hveem, H. A. Askautrud, G. B. Kristensen, A. Nesbakken, J. Trovik, H. Wæhre, I. Tomlinson, N. A. Shepherd, M. Novelli, D. J. Kerr, H. E. Danielsen, Chromatin organisation and cancer prognosis: a pan-cancer study. *The Lancet Oncology* **19**, 356-369 (2018).
11. V. Bedin, R. L. Adam, B. C. de Sa, G. Landman, K. Metze, Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma. *BMC cancer* **10**, 260 (2010).
12. E. Torre, H. Dueck, S. Shaffer, J. Gospocic, R. Gupte, R. Bonasio, J. Kim, J. Murray, A. Raj, Rare Cell Detection by Single-Cell RNA Sequencing as Guided by Single-Molecule RNA FISH. *Cell systems* **6**, 171-179.e175 (2018).
13. M. J. Aryee, W. Liu, J. C. Engelmann, P. Nuhn, M. Gurel, M. C. Haffner, D. Esopi, R. A. Irizarry, R. H. Getzenberg, W. G. Nelson, J. Luo, J. Xu, W. B. Isaacs, G. S. Bova, S. Yegnasubramanian, DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Science translational medicine* **5**, 169ra110 (2013).
14. L. M. Almassalha, G. M. Bauer, W. Wu, L. Cherkezyan, D. Zhang, A. Kendra, S. Gladstein, J. E. Chandler, D. VanDerway, B.-L. L. Seagle, A. Ugolkov, D. D. Billadeau, T. V. O'Halloran, A. P. Mazar, H. K. Roy, I. Szleifer, S. Shahabi, V. Backman, Macrogenomic engineering via modulation of the scaling of chromatin packing density. *Nature Biomedical Engineering* **1**, 902-913 (2017).
15. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews. Genetics* **2**, 292-301 (2001).
16. L. M. Almassalha, G. M. Bauer, J. E. Chandler, S. Gladstein, I. Szleifer, H. K. Roy, V. Backman, The Greater Genomic Landscape: The Heterogeneous Evolution of Cancer. *Cancer research* **76**, 5605-5609 (2016).
17. L. M. Almassalha, A. Tiwari, P. T. Ruhoff, Y. Stypula-Cyrus, L. Cherkezyan, H. Matsuda, M. A. Dela Cruz, J. E. Chandler, C. White, C. Maneval, H. Subramanian, I. Szleifer, H. K. Roy, V. Backman, The Global Relationship between Chromatin Physical Topology, Fractal Structure, and Gene Expression. *Scientific reports* **7**, 41061 (2017).
18. J. S. Kim, A. Yethiraj, Crowding Effects on Protein Association: Effect of Interactions between Crowding Agents. *The Journal of Physical Chemistry B* **115**, 347-353 (2011).
19. H. Matsuda, G. G. Putzel, V. Backman, I. Szleifer, Macromolecular crowding as a regulator of gene transcription. *Biophysical journal* **106**, 1801-1810 (2014).
20. J. N. Weinstein *et al.*, The cancer genome atlas pan-cancer analysis project. *Nature genetics* **45**, 1113 (2013).
21. S. W. Criscione, M. De Cecco, B. Siranosian, Y. Zhang, J. A. Kreiling, J. M. Sedivy, N. Neretti, Reorganization of chromosome architecture in replicative cellular senescence. *Science Advances* **2**, e1500882 (2016).
22. S. Venkatesh, J. L. Workman, Histone exchange, chromatin structure and the regulation of transcription. Nature reviews. *Molecular cell biology* **16**, 178-189 (2015).
23. K. Chiba, J. Yamamoto, Y. Yamaguchi, H. Handa, Promoter-proximal pausing and its release: molecular mechanisms and physiological functions. *Experimental cell research* **316**, 2723-2730 (2010).

24. R. Jothi, S. Cuddapah, A. Barski, K. Cui, K. Zhao, Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic acids research* **36**, 5221-5231 (2008).
25. M. J. Morelli, R. J. Allen, P. R. Ten Wolde, Effects of macromolecular crowding on genetic networks. *Biophysical journal* **101**, 2882-2891 (2011).
26. J. S. Kim, I. Szleifer, Crowding-induced formation and structural alteration of nuclear compartments: insights from computer simulations. *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals* **307**, 73 (2013).
27. H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, C. C. O'Shea, ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
28. B. B. Mandelbrot, *The fractal geometry of nature*. (Macmillan, 1983), vol. 173.
29. A. Bancaud, C. Lavelle, S. Huet, J. Ellenberg, A fractal model for nuclear organization: current evidence and biological implications. *Nucleic acids research* **40**, 8783-8792 (2012).
30. A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, G. J. Narlikar, Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240 (2017).
31. T. Nozaki, R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, K. S. Wendt, Y. Okada, T. Nagai, K. Maeshima, Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Molecular cell* **67**, 282-293.e287 (2017).
32. C. Francastel, D. Schübeler, D. I. Martin, M. Groudine, Nuclear compartmentalization and gene activity. *Nature reviews Molecular cell biology* **1**, 137-143 (2000).
33. S. Gladstein, L. M. Almassalha, L. Cherkezyan, J. E. Chandler, A. Eshein, A. Eid, D. Zhang, W. Wu, G. M. Bauer, A. D. Stephens, S. Morochnik, H. Subramanian, J. F. Marko, G. A. Ameer, I. Szleifer, V. Backman, Multimodal interference-based imaging of nanoscale structure and macromolecular motion uncovers UV induced cellular paroxysm. *Nature Communications* **10**, 1652 (2019).
34. L. M. Almassalha, G. M. Bauer, J. E. Chandler, S. Gladstein, L. Cherkezyan, Y. Stypula-Cyrus, S. Weinberg, D. Zhang, P. Thusgaard Ruhoff, H. K. Roy, H. Subramanian, N. S. Chandel, I. Szleifer, V. Backman, Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy. *Proceedings of the National Academy of Sciences* **113**, E6372 (2016).
35. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
36. L. Seaman, H. Chen, M. Brown, D. Wangsa, G. Patterson, J. Camps, G. S. Omenn, T. Ried, I. Rajapakse, Nucleome Analysis Reveals Structure-Function Relationships for Colon Cancer. *Molecular cancer research* **15**, 821-830 (2017).
37. A. L. Sanborn, S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences* **112**, E6456-6465 (2015).
38. T. Reya, S. J. Morrison, M. F. Clarke, I. L. Weissman, Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105 (2001).

39. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell systems 3, 95-98 (2016).
40. D. Zink, A. H. Fischer, J. A. Nickerson, Nuclear structure in cancer cells. *Nature reviews cancer* **4**, 677 (2004).
41. L. Cherkezyan, Y. Stypula-Cyrus, H. Subramanian, C. White, M. Dela Cruz, R. K. Wali, M. J. Goldberg, L. K. Bianchi, H. K. Roy, V. Backman, Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: a transmission electron microscopy study. *BMC cancer* **14**, 189 (2014).
42. W. Wu, A. J. Radosevich, A. Eshein, T. Q. Nguyen, J. Yi, L. Cherkezyan, H. K. Roy, I. Szleifer, V. Backman, Using electron microscopy to calculate optical properties of biological samples. *Biomedical optics express* **7**, 4749-4762 (2016).
43. V. J. Konda, L. Cherkezyan, H. Subramanian, K. Wroblewski, D. Damania, V. Becker, M. H. Gonzalez, A. Koons, M. Goldberg, M. K. Ferguson, I. Waxman, H. K. Roy, V. Backman, Nanoscale markers of esophageal field carcinogenesis: potential implications for esophageal cancer screening. *Endoscopy* **45**, 983-988 (2013).
44. S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, A. Raj, Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431-435 (2017).
45. A. L. Paek, J. C. Liu, A. Loewer, W. C. Forrester, G. Lahav, Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing. *Cell* **165**, 631-642 (2016).
46. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357 (2012).
47. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323 (2011).
48. L. v. d. Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of machine learning research* **9**, 2579-2605 (2008).
49. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research* **44**, e71 (2016).
50. H. Ma, L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald, T. Pederson, Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature biotechnology* **34**, 528-530 (2016).
51. E. H. Finn, G. Pegoraro, H. B. Brandao, A. L. Valton, M. E. Oomen, J. Dekker, L. Mirny, T. Misteli, Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e1510 (2019).
52. K. L. Reddy, A. P. Feinberg, Higher order chromatin organization in cancer. *Semin Cancer Biol* **23**, 109-115 (2013).
53. A.-M. Florescu, P. Therizols, A. Rosa, Large scale chromosome folding is stable against local changes in chromatin structure. *PLoS computational biology* **12**, e1004987 (2016).
54. C. Uhler, G. V. Shivashankar, Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nature reviews. Molecular cell biology* **18**, 717-727 (2017).

55. G. Arya, T. Schlick, Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proceedings of the National Academy of Sciences* **103**, 16236-16241 (2006).
56. G. Gerlitz, R. Hock, T. Ueda, M. Bustin, The dynamics of HMG protein-chromatin interactions in living cells. *Biochemistry and cell biology* **87**, 127-137 (2009).
57. S. C. Parker, L. Hansen, H. O. Abaan, T. D. Tullius, E. H. Margulies, Local DNA topography correlates with functional noncoding regions of the human genome. *Science* **324**, 389-392 (2009).
58. H. Chen, J. Chen, L. A. Muir, S. Ronquist, W. Meixner, M. Ljungman, T. Ried, S. Smale, I. Rajapakse, Functional organization of the human 4D Nucleome. *Proceedings of the National Academy of Sciences* **112**, 8002-8007 (2015).
59. N. Nicot, J. F. Hausman, L. Hoffmann, D. Evers, Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of experimental botany* **56**, 2907-2914 (2005).
60. L. Cherkezyan, D. Zhang, H. Subramanian, I. Capoglu, A. Taflove, V. Backman, Review of interferometric spectroscopy of scattered light for the quantification of subdiffractional structure of biomaterials. *Journal of biomedical optics* **22**, 30901 (2017).
61. A. J. Radosevich, J. Yi, J. D. Rogers, V. Backman, Structural length-scale sensitivities of reflectance measurements in continuous random media under the Born approximation. *Optics letters* **37**, 5220-5222 (2012).
62. S. Przibilla, S. Dartmann, A. Vollmer, S. Ketelhut, G. v. Bally, B. Kemper, B. Greve, Sensing dynamic cytoplasm refractive index changes of adherent cells with quantitative phase microscopy using incorporated microspheres as optical probes. *Journal of biomedical optics* **17**, 1-9, 9 (2012).
63. Y. Sung, W. Choi, N. Lue, R. R. Dasari, Z. Yaqoob, Stain-free quantification of chromosomes in live cells using regularized tomographic phase microscopy. *PLoS One* **7**, e49502-e49502 (2012).

Acknowledgements: We thank Z. Ji and members of his lab for advice on scRNA-seq analysis.

Funding: This work was supported by funding from the National Science Foundation grant EFMA-1830961 and the National Institutes of Health grants R01CA228272, R01CA225002, U54CA193419, and T32GM008152.

Author Contributions: Conceptualization and Project Design: W.W., L.A., H.K.R., I.S., and V.B.; Formal analysis: R.V., L.A., and W.W.; Experiments: L.A., G.B., Y.L., D.V., J.F., and A.E.; Random Media Simulations: D.Z; Writing (original draft): W.W., L.A., I.S. and V.B.; Writing (review and editing): R.V., L.A., and V.B.

Competing Interests: The authors declare they have no competing interests.

Data Availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The results of patient study are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. Additional data related to this paper may be requested from the authors.

List of Supplementary Materials:

Figures:

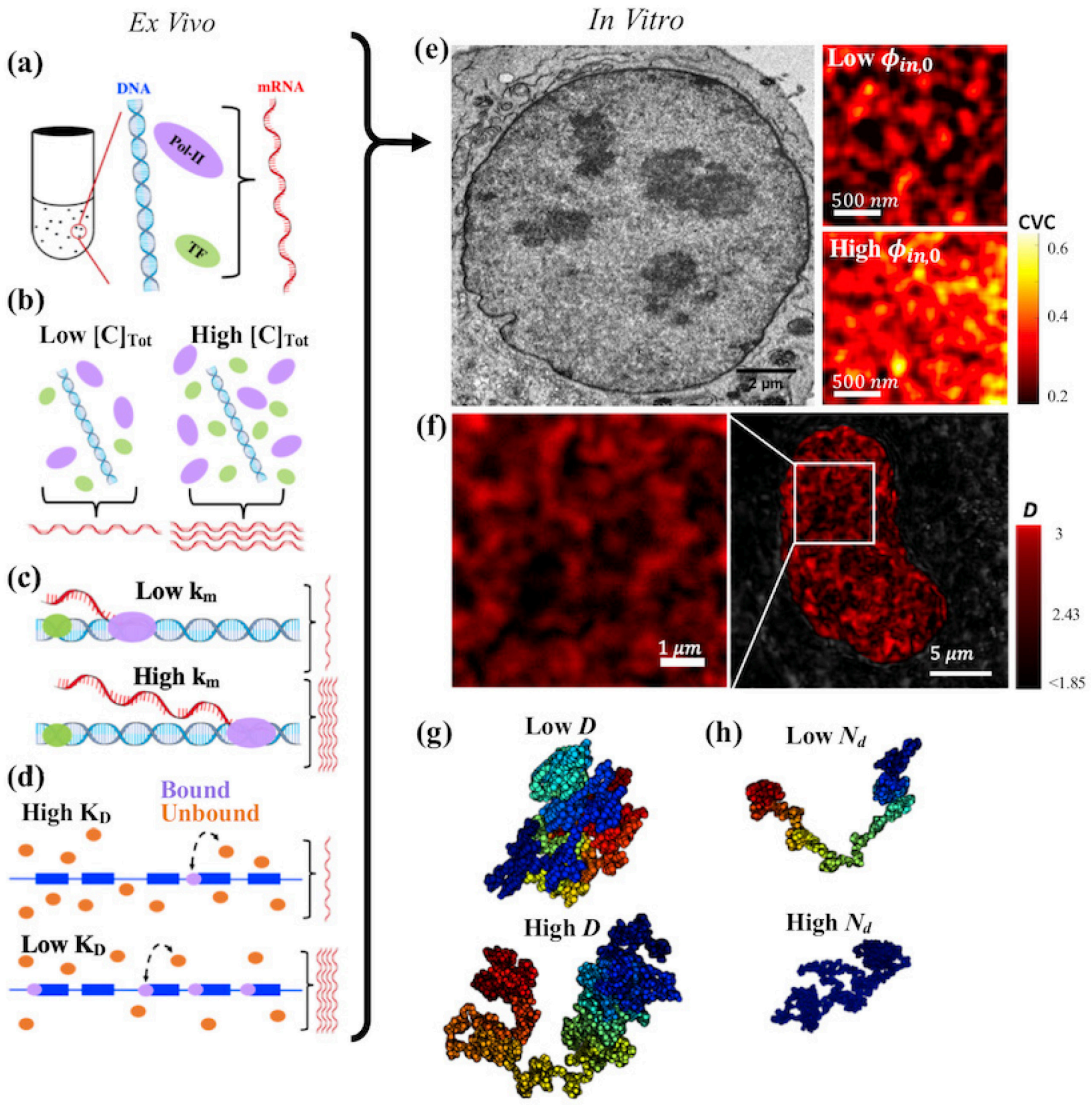

Fig. 1. The Chromatin-Packing Macromolecular-Crowding model integrates molecular and physical regulators of transcription. The regulators influencing transcription reactions can be generally divided into two categories: molecular regulators (k_m, K_D, and $[C]_{tot}$) **(a-d)** and physical regulators (D, $\phi_{in,0}$, and N_d) **(e-h)**. **(a)** The CPMC model describes transcription as a series of diffusion limited chemical reactions. *Ex vivo,* expression depends on (**b**) concentration of

transcriptional reactants $[C]_{tot}$ (TFs (green), Pol-II (yellow)), **(c)** the RNA polymerase elongation rate, k_m, and (d) the disassociation rate of Pol-II from the transcription start site (TSS) K_D. (**e**) (*Left*) In addition to the molecular determinants, transcriptional reactions are influenced by the highly dense and complex nuclear environment. The concentration of the main crowder with the nucleus, chromatin, can be measured by chromatin electron microscopy (ChromEM). As an example, ChromEM of a nucleus of an A549 lung adenocarcinoma cell is shown. (*Right*) ChromEM measurements of chromatin volume concentration (CVC) demonstrates that chromatin density varies throughout the nucleus. Chromatin packing domains can be visualized as areas of higher chromatin packing density. Within each packing domain the average volume fraction of chromatin can range from 15% to 65%. Typical domains are 100 to 200 nm in diameter and may contain, on average, ~400 kb. **(f)** Representative PWS image of an A549 cell demonstrating the existence of chromatin packing domains as regions of elevated chromatin packing scaling (also referred to as fractal dimension) D, which vary throughout the nucleus. **(g)** A polymer with a higher D (right) has a more heterogeneous density distribution and a greater accessible surface area compared to a polymer with a lower D (left). (**h**) N_d is the genomic size (in bp) of a chromatin packing domain and can range from less than 100Kbp to several Mbp. Packing domains are illustrated by color coding with each color representing a separate domain.

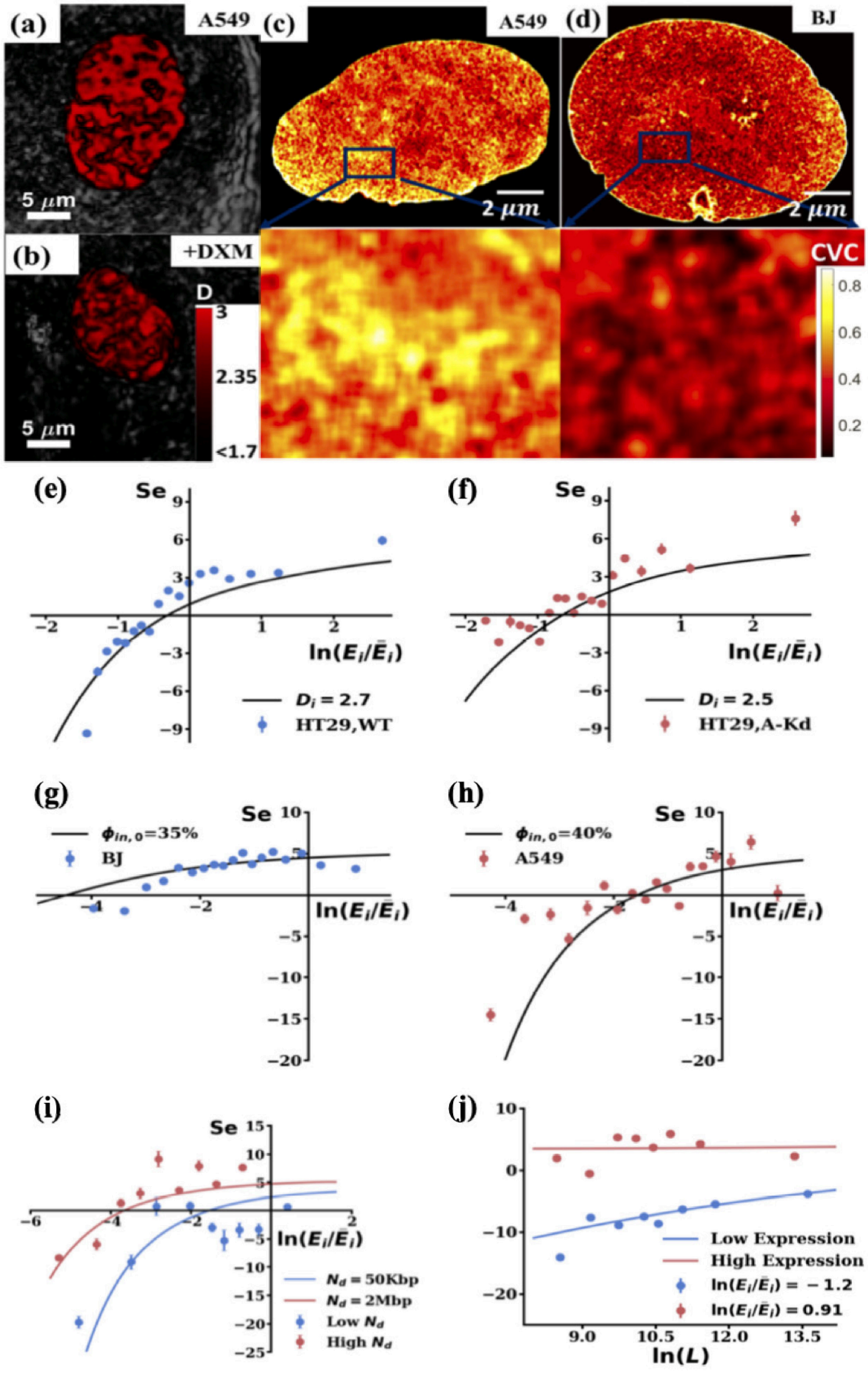
(a)
A549
5 μm
(b)
+DXM
D
3
2.35
<1.7
5 μm
(c)
A549
2 μm
(d)
BJ
2 μm
CVC
0.8
0.6
0.4
0.2
(e)
Se
ln(Ei/Ēi)
Di = 2.7
HT29,WT
(f)
Se
ln(Ei/Ēi)
Di = 2.5
HT29,A-Kd
(g)
Se
ϕin,0=35%
BJ
ln(Ei/Ei)
(h)
Se
ϕin,0=40%
A549
ln(Ei/Ei)
(i)
Se
ln(Ei/Ei)
Nd = 50Kbp
Nd = 2Mbp
Low Nd
High Nd
(j)
Low Expression
High Expression
ln(Ei/Ei) = − 1.2
ln(Ei/Ei) = 0.91
ln(L)

Fig. 2. Comparison of the CPMC model with experimental measurements of gene expression as a function of physical regulators D_i, N_d, and $\phi_{in,0}$ and gene length L. (a,b) Representative live cell PWS microscopy images of nuclear *D* scaled between 2.56 and 2.66 for control (a) and 12 hour dexamethasone treated lung adenocarcinoma A549 cells (b). Brighter red corresponds to higher chromatin packing scaling. (**c,d**) Representative heat maps of CVC values from analysis of ChromEM images of cell nuclei from A549 cells (c) and human fibroblasts BJ (d). Representative magnified regions from each nucleus demonstrates average CVC=0.35 in A549 cell compared to 0.35 in BJ cells, which represents the chromatin contribution to the average crowding volume fraction $\phi_{in,0}$. (**e-j**) Comparison between the CPMC model (solid lines) and experimentally measured (points) sensitivity of gene expression to an incremental change in chromatin *D* (*Se*, y-axis) as a function of the initial gene expression (x-axis). (**e**) Cells with chromatin with a high initial D_i=*2.7* (wild-type HT29 cells) have a bidirectional *Se* curve that becomes attenuated if D_i is lowered to 2.5 (shRNA knockdown Arid-1a HT-29 cells) (**f**). Each point represents the average of 100 genes. Changes in *D* were induced by cell treatment with 10% fetal bovine serum, 100nM epidermal growth factor (EGF), and 100nM phorbol 12-myristate 13-acetate (PMA). The CPMC model was able to explain 86% of the variance of the experimental data for wild-type HT-29 cells and 51% of the variance for Arid-1a HT29 cells. (**g**) *Se* in cells with a lower $\phi_{in,0}$ (BJ cells, $\phi_{in,0} = 35\%$; each point corresponds to 300 genes; explained variance (EV) = 59%) is attenuated in comparison to that of cells with a higher density (**h**) (A549 cells; $\phi_{in,0} = 40\%$; 100 genes per point; EV = 74%). (**i**) Genes located within larger packing domains (N_d ~ 2Mbp, 12 genes/point, EV = 56%) have a lower initial expression but have a positive *Se* to changes in *D* in comparison to genes localized within smaller packing domains (N_d ~ 50Kbp, 12 genes/point, EV = 37%). The change in *D* was induced in A549 cells by treatment with 100 nM of dexamethasone. N_d was approximated based on the corresponding TAD size: 2Mbp TADs for the high N_d group of genes vs. 50Kbp TADs for the low N_d genes. TAD size was measured using the Arrowhead function from the Juicer Tools to analyze Hi-C data. **(j)** Comparison between the CPMC model (solid line) with experimental results (points, 60 genes/point) in HT-29 cells showing the effect of gene length (*L*, x-axis) on *Se* (y-axis). In agreement with the model, shorter, initially under-expressed genes (low expression, blue curve, points, EV = 67%) are disproportionally repressed by an incremental increase in *D* compared to longer genes (high expression, red curve, points). Error bars represent standard error from 4 biological replicates.

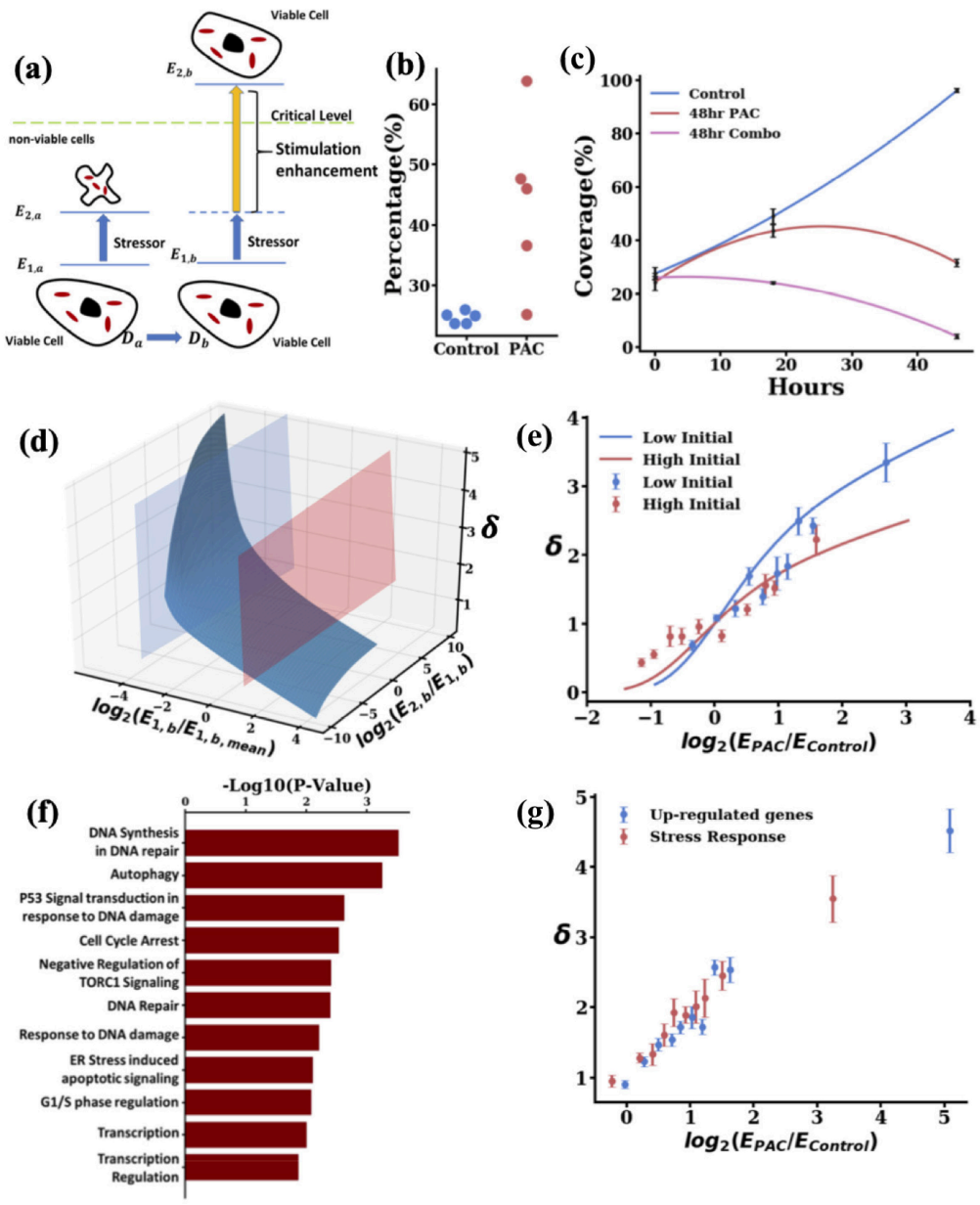

Fig. 3. The scaling of chromatin packing increases the transcriptional malleability of cancer cells. (a) In response to a stressor, such as a chemotherapeutic agent (e.g. paclitaxel), cells with a higher level of transcriptional malleability may have the ability to respond faster, which may lead to an increased survival. Chromatin packing with a higher D (right, D_b) increases a change in the rate of transcription induced by a stimulus/stressor by a factor δ (yellow arrow) relative to a change in the rate of transcription in a cell with a lower $D = D_a < D_b$. If in response to a stressor a cell may increase the probability of retaining viability by reaching a given threshold of expression of pro-acclimation genes, a higher D in cell b would increase the probability of reaching this level of expression compared to cell a. **(b&c)** The fraction of high D cells in a cell culture increases after treatment with paclitaxel for 48 hours (PAC), suggesting that cells with higher D are more likely to survive exposure to a cytotoxic chemotherapeutic agent. **(b)** The percentage of cells having D

above the top quartile of a control cell population (y-axis) increases in cells that survive treatment with paclitaxel for 48 hours. For both conditions, each dot represents percentage of high *D* cells in one replicate for a total number of N=5 replicates per condition. **(c)** Combination treatment with celecoxib, which lowers *D,* and paclitaxel for 48 hours results in increased elimination of cancer cells compared with untreated controls and paclitaxel mono-treated cells. **(d)** CPMC model predictions of the relative transcriptional malleability coefficient δ for initially under-expressed (blue spline) and over-expressed genes (red spline) for D_a=2.3 and D_b=2.5, a difference in *D* relevant to experimentally observed differences in celecoxib-treated versus untreated A2780 cells. **(e)** Single cell RNA sequencing on A2780 cells was performed to compare transcriptional profiles of control A2780 cells (high *D* population) and cells treated with 75 μM of a *D*-lowering agent celecoxib (low *D* population) and their response to treatment with 5nM paclitaxel (stressor) for 16 hours. Initially under-expressed and initially over-expressed genes are defined based on control expression levels. Genes are grouped based on their quantile of $log_2(E_{PAC}/E_{control})$ and the mean and standard error of each quantile for initially under-expressed genes (blue dots, 300 genes/data point) and initially over-expressed genes (red dots, 100 genes/data point) are plotted. **(f)** Gene Ontology (GO) analysis identified biological processes that are most significantly involved in the response to 48 hour paclitaxel treatment. Upregulated genes were defined as those with at least 2 fold increase in expression. (**g**) Chromatin packing scaling-facilitated upregulation (δ) of the stress-response genes identified by the GO analysis (red points, 150 genes/data point) was similar to that for all upregulated genes (blue points, 650 genes/data point).

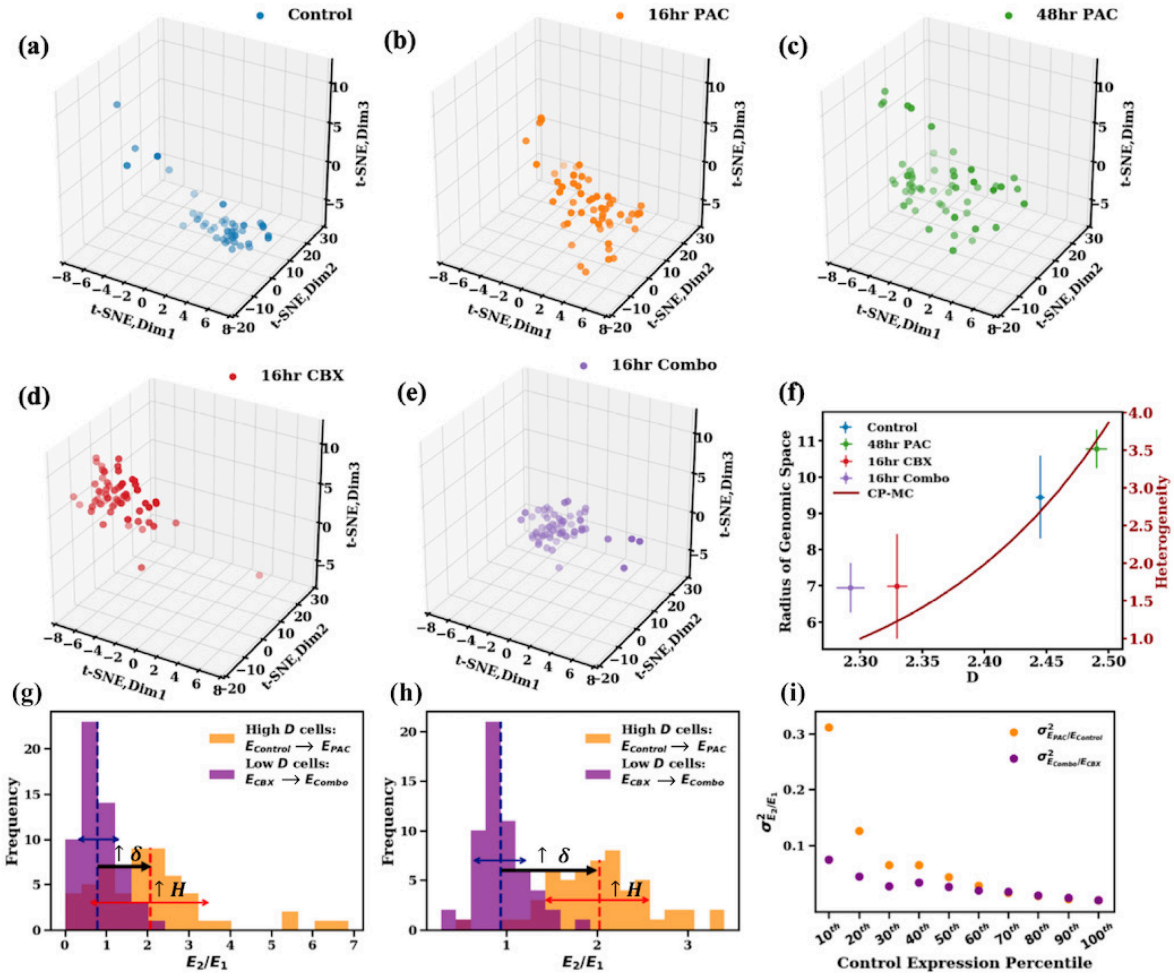

Fig. 4. The scaling of chromatin packing regulates intercellular transcriptional heterogeneity of cancer cells. (a-e) 3D projections of scRNA-seq data (TPM values of 8,275 expressed genes) onto reduced t-SNE space for 5 conditions: (a) control cells (n=46), (b) cells treated with 5nM paclitaxel for 16 hours (16hr PAC, n=55), (c) 5nM paclitaxel for 48 hours (48hr PAC, n=53), (d) 75μm celecoxib for 16 hours (16hr CBX, n=62), (e) and combination of 75μM celecoxib and 5nM paclitaxel for 16 hours (16hr Combo, n=59). The size of the cluster indicates the transcriptional heterogeneity within the population of surviving cells for each condition. **(f)** The radius of genomic space R_c (the radius of clusters through a-e) increases as a function of the chromatin packing scaling D. D was measured by live cell PWS at each time point on cells prior to sequencing. Cells treated with paclitaxel (higher D) have greater transcriptional heterogeneity, especially when compared to cells treated with non-steroidal anti-inflammatory agent, celecoxib, which lowers D. Likewise, the CPMC model (red curve, right side y-axis) shows that intercellular transcriptional heterogeneity increases with D. Error bars represent the standard error of D calculated from PWS measurements (x-axis) and R_c (y-axis) for each condition. **(g)** Relative expression of high D versus low D cells in response to paclitaxel treatment for genes associated with DNA repair pathways which are upregulated in 48 hour paclitaxel treated cells. For each condition (Control, 16hr PAC,

2hr CBX, 16hr Combo), TPM values of these genes (48 in total) was averaged within each cell. Next, expression of paclitaxel-stimulated cells was normalized by the average of the corresponding unstimulated population. The resulting intercellular distribution of relative expression levels is shown. Dashed lines represent mean relative expression. Solid red and blue arrows represent the standard deviation of distributions $E_{PAC}/E_{Control}$ and E_{CBX}/E_{Combo}, respectively. For these stress response genes, cells with a higher initial *D* versus cells with a lower initial *D* had an increase in transcriptional malleability ($\uparrow \delta$) as well as a higher intercellular transcriptional heterogeneity ($\uparrow H$). **(h)** Distribution of relative expression of genes, as described in (g), in the lowest quantile (10^{th} percentile) of control expression levels (839 in total). **(i)** Variance (σ^2) of intercellular distribution of relative expression for each percentile of control expression levels. Initially under-expressed genes show an increased effect of chromatin packing scaling on increasing intercellular transcriptional heterogeneity in response to paclitaxel stimulation compared to that of initially over-expressed genes.

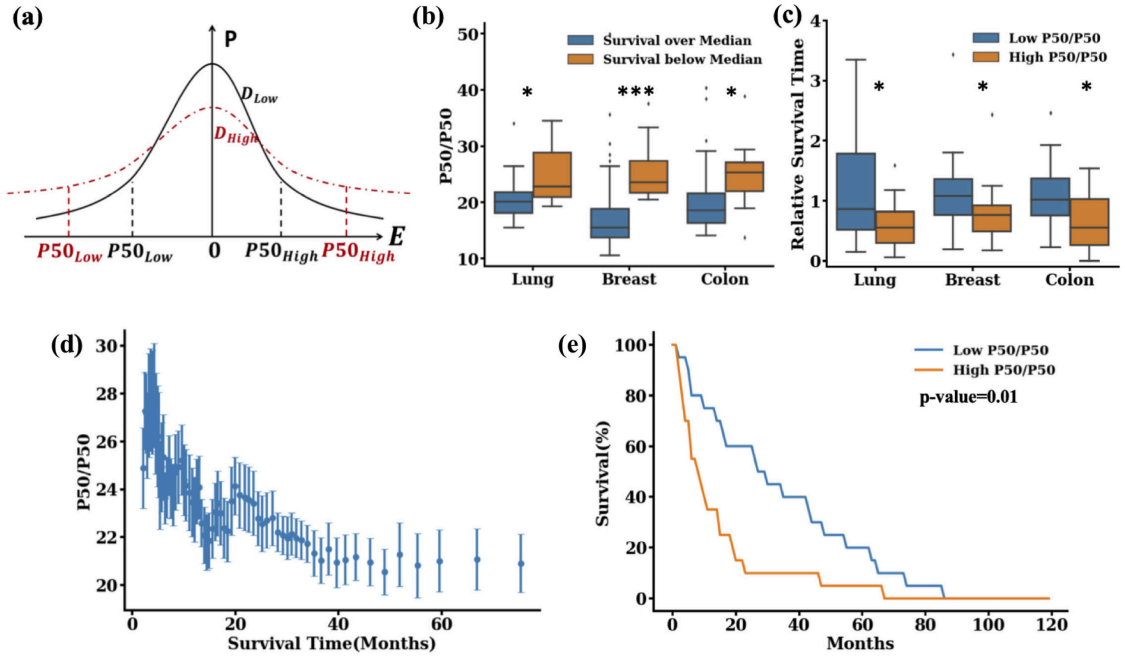

Fig. 5. The relationship between transcriptional divergence (P50/P50) and patient survival in Stage III and IV lung, breast, and colon cancers. (a) From the *Se* curve predicted by the CPMC model, cells with high *D,* such as cancer cells, have a wider distribution of gene expression (transcriptional divergence). Quantitatively, this transcriptional divergence can be calculated by measuring the ratio of the expression of the top 50% of genes to that of the bottom 50% of genes (P50/P50). **(b-e)** Analysis of transcriptional divergence, P50/P50, in the cancer cells of patients with Stage III and IV lung cancer (n=31), breast cancer (n=168), and colon cancer (n=60) verses survival from the time of diagnosis based on The Cancer Genome Atlas dataset for patients <75 years old at the time of diagnosis. **(b)** P50/P50 was elevated in patients with a survival duration below the median for each cancer type ($p = 0.021$, < 0.001 and $= 0.018$ for lung, breast, and colon cancers, respectively). **(c)** The relative survival time (RST; ratio between patient survival time and that predicted by a multidimensional linear regression model based on known prognostic factors such as stage at diagnosis, race, and molecular subtypes of the tumor) is higher for patients with low P50/P50 (P50/P50 below the mean for all patients with a given cancer type). RST < 1 indicates survival shorter than expected based on demographic factors and molecular subtype (all $p < 0.05$). For all three malignancies, RST < 0.8 in high-P50/P50 patients. RST is an independent predictor of survival duration. **(d)** Pooling all patients with these malignancies, we analyzed survival duration (x-axis, in months) vs. P50/P50 at the time of diagnosis. There was an inverse relation between P50/P50 and survival duration. Each point is a moving window average of 10 patients to account for unreported variables (e.g. comorbidities). **(e)** The Kaplan-Meier curve measuring patient survival for the three malignancies. Patients with a high P50/P50 (P50/P50 above the mean) have a shorter survival duration (median survival = 8 months) than patients with low P50/P50 (P50/P50 below the mean, median survival = 28 months, $p = 0.01$).

Science
AAAS

Supplementary Materials for

DISORDERED CHROMATIN PACKING REGULATES PHENOTYPIC PLASTICITY

Ranya K.A. Virk[1], Wenli Wu[1], Luay M. Almassalha[1,2,3], Greta M. Bauer[1], Yue Li[1,4], David VanDerway[1], Jane Frederick[1], Di Zhang[1], Adam Eshein[1], Hemant K. Roy[5], *Igal Szleifer[1,6,7], and *Vadim Backman[1,7]

[1]Department of Biomedical Engineering, Northwestern University, Evanston, Illinois, 60208, USA
[2]Medical Scientist Training Program, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, 60211, USA
[3]Department of Internal Medicine, Northwestern University, Chicago, Illinois, 60211, USA
[4]Applied Physics Program, Northwestern University, Evanston, Illinois, 60208, USA
[5]Section of Gastroenterology, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, 02118, USA
[6]Department of Chemistry, Northwestern University, Evanston, Illinois, 60208, USA
[7]Chemistry of Life Processes Institute, Northwestern University, Evanston, Illinois, 60208, USA

v-backman@northwestern.edu; igalsz@northwestern.edu

This PDF file includes:

Supplementary Text

Section SI. CPMC Model: Variance of Chromatin Packing Density.

The variance of chromatin packing density in each interaction volume within the cell nucleus, $\sigma^2_{\phi_{in}}$, can be expressed as a function of *D*. By definition, the variance of any value, *x,* with probability distribution function *h(x)* can be calculated from the autocorrelation function *H(x)* as *Var(x)=H(x=0)*. The relationship between $\sigma^2_{\phi_{in}}$ and the autocorrelation function of chromatin packing density, ϕ_{in}, is $B_{in}(\vec{r})$, and can be calculated as:

$$\sigma^2_{\phi_{in}} = B_{in}(\vec{r} = 0) \quad (1)$$

According to the definition of an autocorrelation function, $B_{in}(\vec{r})$ is calculated by:

$$B_{in}(\vec{r}) = \int [\phi_{in}(\overrightarrow{r'} + \vec{r}) - \bar{\phi}][\phi_{in}(\overrightarrow{r'}) - \bar{\phi}]d\overrightarrow{r'} = [\phi_{in}(\overrightarrow{r'} + \vec{r}) - \bar{\phi}] * [\phi_{in}(\overrightarrow{r'}) - \bar{\phi}] = [\int [\phi_{in}(\overrightarrow{r'} + \vec{r}) - \bar{\phi}]A_{in}(\overrightarrow{r'})d\overrightarrow{r'}] * [\int [\phi_{in}(\overrightarrow{r'} + \vec{r}) - \bar{\phi}]A_{in}(\overrightarrow{r'})d\overrightarrow{r'}] = \left[[\phi_{in}(\vec{r}) - \bar{\phi}] * A_{in}(\overrightarrow{r'})\right] * \left[[\phi_{in}(\vec{r}) - \bar{\phi}] * A_{in}(\overrightarrow{r'})\right] = \left[[\phi_{in}(\vec{r}) - \bar{\phi}] * [\phi_{in}(\vec{r}) - \bar{\phi}]\right] * [A_{in}(\vec{r}) * A_{in}(\vec{r})] = B(\vec{r}) * ACF_{in}(\vec{r}) = \int B(\overrightarrow{r'} + \vec{r})ACF_{in}(\overrightarrow{r'})d\overrightarrow{r'} \quad (2)$$

where '*' represents the convolution operation, $B(\vec{r})$ is the autocorrelation function of $\phi(\vec{r})$ and $ACF_{in}(\vec{r})$ is the autocorrelation function of $A_{in}(\vec{r})$. As a result, Eq. 1&2 simplifies to:

$$\sigma^2_{\phi_{in}} = \int B(\overrightarrow{r'})ACF_{in}(\overrightarrow{r'})d\overrightarrow{r'} \quad (3)$$

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:

$$\frac{B(r)}{\sigma^2} = \begin{cases} 1, r < r_{min} \\ \left(\frac{r_{min}}{r}\right)^{3-D}, r_{min} \le r \le r_f \\ 0, r > r_f \end{cases} \quad (4)$$

where we can 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 $A_{in}(\vec{r})$ with a radius of r_{in}, the autocorrelation function $ACF_{in}(\vec{r})$ of $A_{in}(\vec{r})$ is also a Gaussian distribution:

$$ACF_{in}(\vec{r}) = \alpha e^{-\frac{r^2}{2r_{in}^2}} \quad (5)$$

where α 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 σ^2, which gives us:

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

In turn, Eq. 6 allows us to solve for alpha:

$$\alpha = \frac{1}{4\pi}\sqrt{\frac{2}{\pi}}\frac{1}{{r_{in}}^3} \quad (7)$$

Therefore, solving Eq. 5 becomes:

$$\sigma_{\phi_{in}}^2 = \sigma^2 \frac{1}{4\pi}\sqrt{\frac{2}{\pi}}\frac{1}{{r_{in}}^3} 4\pi \left[\int_0^{r_{min}} \mathrm{r}^2 e^{-\frac{r^2}{2r_{in}^2}} dr + \int_{r_{min}}^\infty \mathrm{r}^2 e^{-\frac{r^2}{2r_{in}^2}} \left(\frac{r_{min}}{r}\right)^{3-D} dr\right] \approx \sigma^2\sqrt{\frac{2}{\pi}}\frac{1}{{r_{in}}^3}\left[\frac{1}{3} r_{min}^3 + r_{min}^{3-D} 2^{\frac{D}{2}-1} r_{in}^D \int_{\frac{r_{min}^2}{2r_{in}^2}}^\infty x^{\frac{D}{2}-1} e^{-x} dx\right] \approx \sigma^2\sqrt{\frac{2}{\pi}}\left(\frac{r_{min}}{r_{in}}\right)^{3-D} 2^{\frac{D}{2}-1}\left[\Gamma\left(\frac{D}{2}\right) - \frac{2}{D}\frac{1}{2^{\frac{D}{2}}}\left(\frac{r_{min}}{r_{in}}\right)^D\right] \approx \sigma^2\sqrt{\frac{2}{\pi}} 2^{\frac{D}{2}-1}\left(\frac{r_{min}}{r_{in}}\right)^{3-D}\Gamma\left(\frac{D}{2}\right) \approx \sigma^2\left(\frac{r_{min}}{r_{in}}\right)^{3-D} \quad (8)$$

where $\Gamma\left(\frac{D}{2}\right)$ is the Gamma function. Here, we assume that $r_{in} >> r_{min}$ and $\mathrm{v(D)} = \sqrt{\frac{2}{\pi}} 2^{\frac{D}{2}-1}\Gamma\left(\frac{D}{2}\right) \approx 1$ 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, $A_{in}(\vec{r})$. If it has a uniform distribution, $\mathrm{v(D)} = \frac{12}{D(D+1)} \approx 1$, 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:

$$\sigma^2 = \phi_{in,0}(1 - \phi_{in,0}) \quad (9)$$

Using these considerations, $\sigma_{\phi_{in}}^2$ reduces to:

$$\sigma_{\phi_{in}}^2 = \phi_{in,0}(1 - \phi_{in,0})\left(\frac{r_{min}}{r_{in}}\right)^{3-D} \quad (10)$$

The equation shown above indicates that the variance of local crowding increases with an increase of D. To test if this effect of D on $\sigma_{\phi_{in}}^2$ as derived analytically above is conserved, we measured $\sigma_{\phi_{in}}^2$ as a function of D using simulations of random clusters (**Fig. S1**) and random media (**Fig. S2**). As predicted analytically, $\sigma_{\phi_{in}}^2$ increases as a function of D in both sets of simulations. Consequently, this indicates that the effect of chromatin packing on the variance of local crowding is independent of the chosen chromatin model.

Section SII. Random Media Simulations: Calculation of *D* from Mass Density Variations.

To test the relationship between σ^2, D and $\sigma^2_{\phi_{in}}$ 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 was generated from an autocorrelation function (ACF) of random noise (*61*). The mass density ρ in this calculation can be transformed into crowding volume fraction, or crowding density ϕ through $\rho = \phi\rho_c$, where ρ_c is the dry mass density of macromolecular crowders. First, an autocorrelation function was generated with density variance σ^2_ρ=0.2g/cm^3, fractal range (1 to 100 nm) and fractal dimension D (1.2 or 2.5), based on the following equation:

$$ACF = \sigma^2_\rho \exp\left[\frac{-r/r_{max}}{1+(r/r_{min})^{3-D}}\right] \quad (11)$$

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 determining the shape of the ACF. The random media was generated using this ACF for low (D=1.2) and high (D=2.5) fractal dimension (**Fig. S2**). The random media generated through this method had on average the same total mass density variance σ^2_ρ=0.2g/cm^3. For each individual voxel, the mass density was averaged within the interaction volume with radius $r_{in} = 20nm$ to calculate ρ_{in}. Next, the local variance of each interaction volume is calculated to determine $\sigma^2_{\rho_{in}}$ in each random media to compare low and high D cases. The results from these simulations are as shown in **Fig. S1**. As we can see, the random media generated through ACF with higher D (D=2.5) has a larger $\sigma^2_{\rho_{in}}$ compared with the $\sigma^2_{\rho_{in}}$ of the random media generated with lower D (D=1.25) when σ^2_ρ, r_{min} and r_{max} are the same. This confirms a direct relationship between D and σ^2_ρ 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. In the case of the ACF of an alternative arbitrary medium (not necessary a fractal) this can produce a D larger than 3 but would still have the same effective relationship between D and σ^2_ρ.

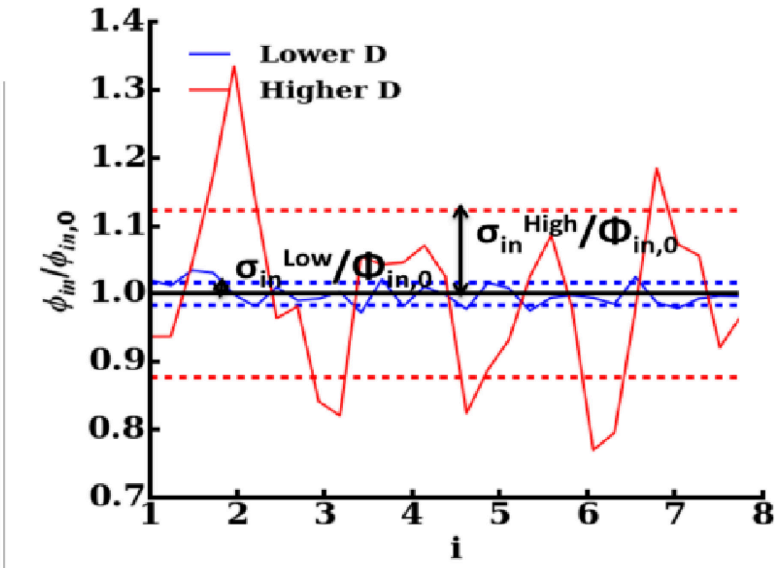

Figure S1. Analysis of the relationship between *D* and $\sigma^2_{\phi_{in}}$. A random media was simulated with clusters of size distribution ~$1/r_s^{4\text{-}D}$ for cluster radius r_s. The mass density for each interaction volume ϕ_{in} was calculated by dividing the 3D random media into 125 separate boxes (5x5x5) and calculating the volume fraction occupied by the randomly distributed clusters in the media. The standard derivation of ϕ_{in}, $\sigma_{\phi_{in}}$, was calculated for media generated from two different *D* (**Figure S2**). The blue dashes in the figure represent $\sigma_{\phi_{in}}$ of the media with lower *D* and the red dashes represent $\sigma_{\phi_{in}}$ of the media with higher *D*, showing a higher $\sigma_{\phi_{in}}$ for the media with higher *D*.

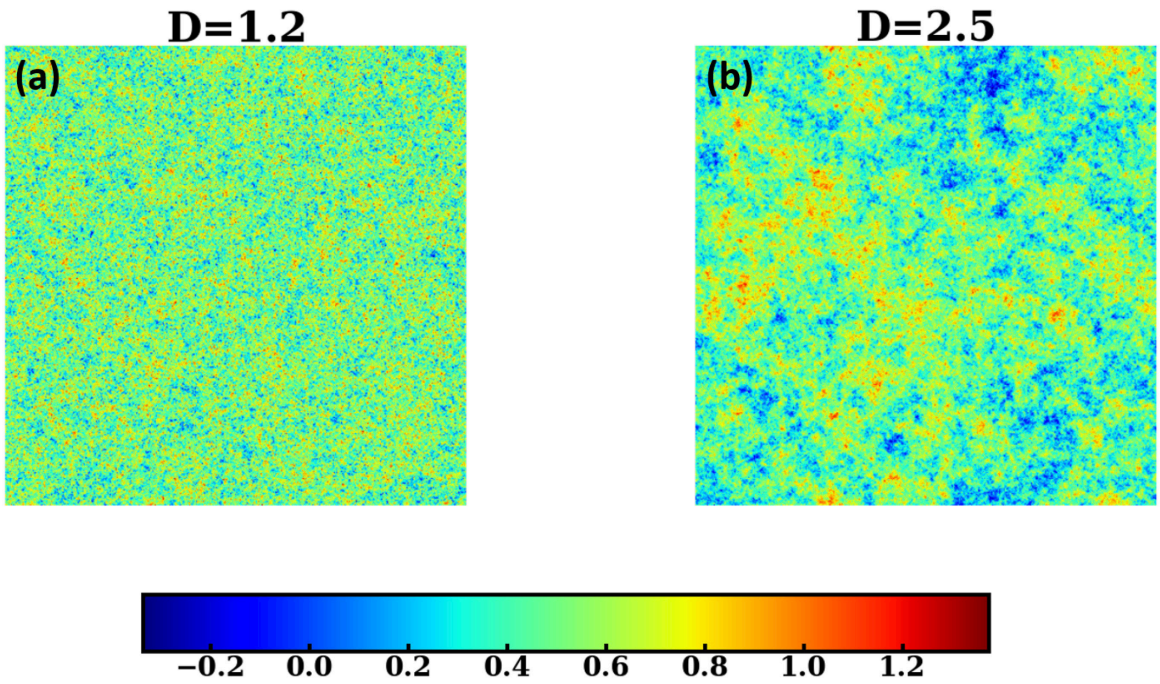

Figure S2. Random media simulations for low and high D.

Section SIII. Calculation of $\phi_{in,0}$ from ChromEM measurements.

Chromatin volume concentration (CVC) was first calculated from ChromEM samples to estimate the contribution of chromatin to average nuclear crowding. 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 nm^3 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, 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. Next, we calculated $\phi_{in,0}$ by adding an additional 5% to average CVC measurements in each nucleus to account for mobile crowders in the nucleus. This proportion of mobile crowders can be obtained from the multiple previously reported measurements of the refractive index (RI) of cell nuclei and other cellular compartments including the cytosol and the Gladstone-Dale relationship between RI and crowding: $n(\vec{r}) = n_{water} + \alpha\rho(\vec{r})$ where $n(\vec{r})$ is the refractive index of the biological material at point $\vec{r}$ in 3D space, n_{water} is the refractive index of water, α is the refractive index increment, and ρ is molecular density (in g/ml) (*42*). Earlier reports indicate that the average value RI of the nucleoplasm and the cytosol are 1.339, which results in the average estimate of 5% volume fraction of mobile crowders (ϕ) (*62*, *63*).

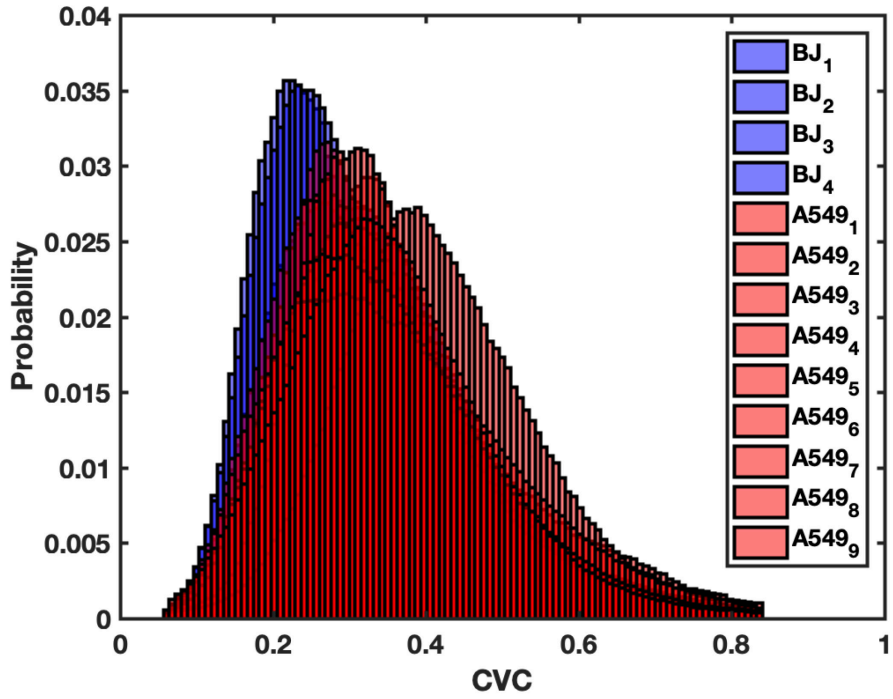

Figure S3. CVC distributions of A549 and BJ cells as measured by ChromEM. Analysis of ChromEM of chromatin volume concentration (CVC) values across N=4 replicates of differentiated BJ fibroblast nuclei and N=9 replicates of A549 lung adenocarcinoma nuclei. A549 nuclei have a pooled CVC average of 0.35 while BJ nuclei have a pooled CVC average of 0.30. These values represent the chromatin contribution to $\phi_{in,0}$.

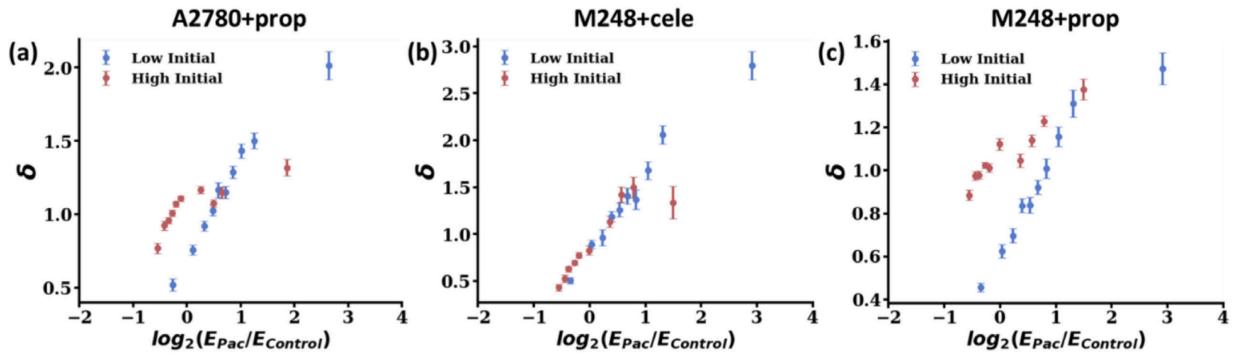

Figure S4. Transcriptional malleability in A2780 and M248 cells. The tailwind factor $\delta = \left(\frac{E_{2,b}}{E_{1,b}}\right)\Big/\left(\frac{E_{2,a}}{E_{1,a}}\right)$ was determined from additional bulk RNA-seq experiments on A2780 cells and TP53 mutated clone A2780.M248 cells along with propranolol, another compound that lowers *D*. PWS measurements showed a 2% decrease in *D* in A2780 cells after propranolol treatment for 16 hours and a ~5% decrease in *D* in M248 cells treated with separately with celecoxib and propranolol for 16 hours. **(a)** δ effect in A2780 cells treated with propranolol to lower *D*. All treatment conditions include: control, 16 hour propranolol, 16 hour paclitaxel, and 16 hours paclitaxel plus celecoxib. **(b&c)** δ tested in M248 cells treated with celecoxib (b) and propranolol (c) as *D* lowering compounds for 16 hours. All treatment conditions include control, 16 hours celecoxib/propranolol, 16 hours paclitaxel, 16 hours paclitaxel plus celecoxib/propranolol. All results are based on the expression data at t=16 hours. Error bars represent the standard error of δ of genes within each quantile. There are three biological replicates for every condition.

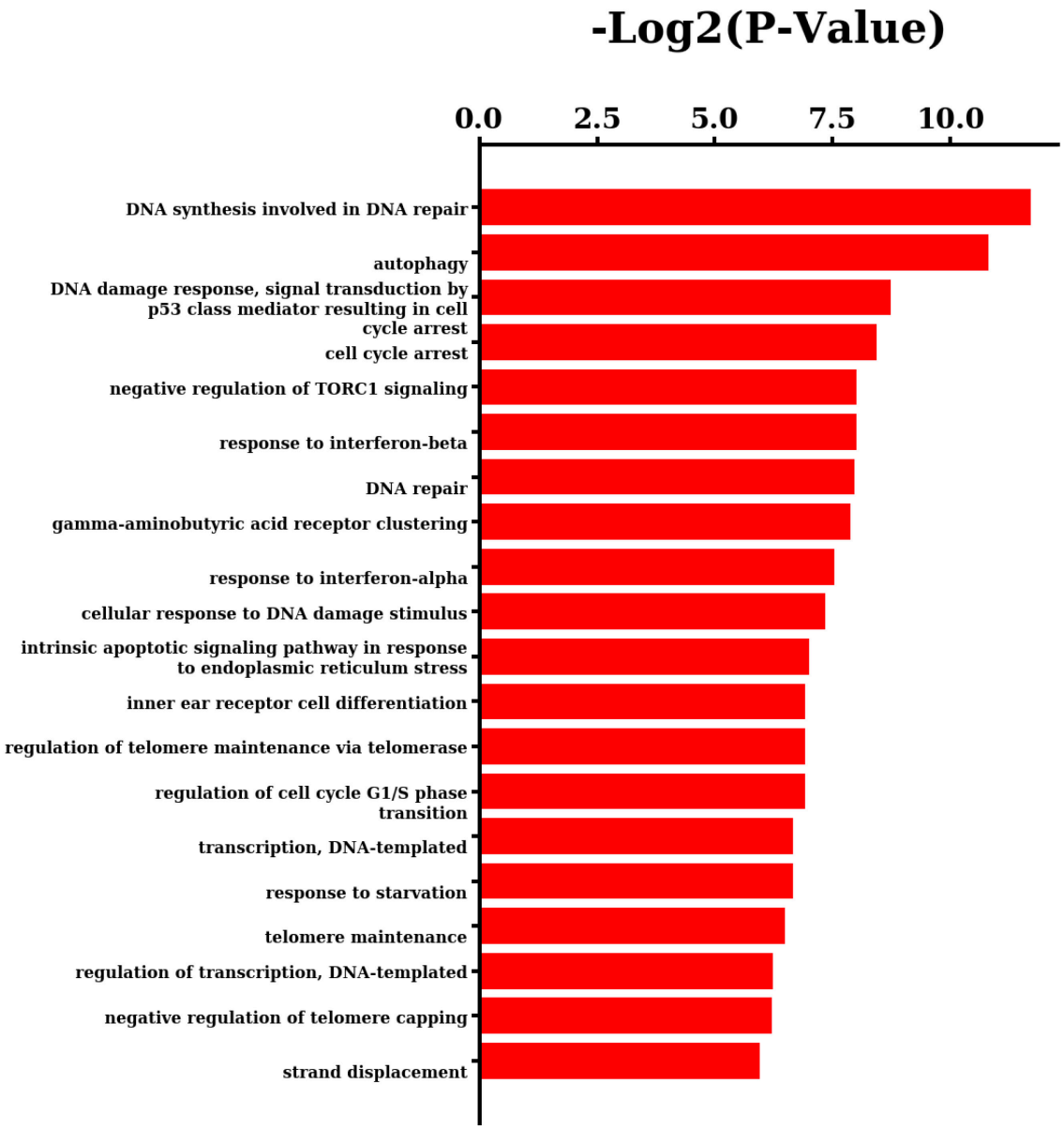

Figure S5. Gene ontology analysis of upregulated genes. Full list of biological processes that contain genes upregulated in paclitaxel-treated A2780 cells compared to control cells.

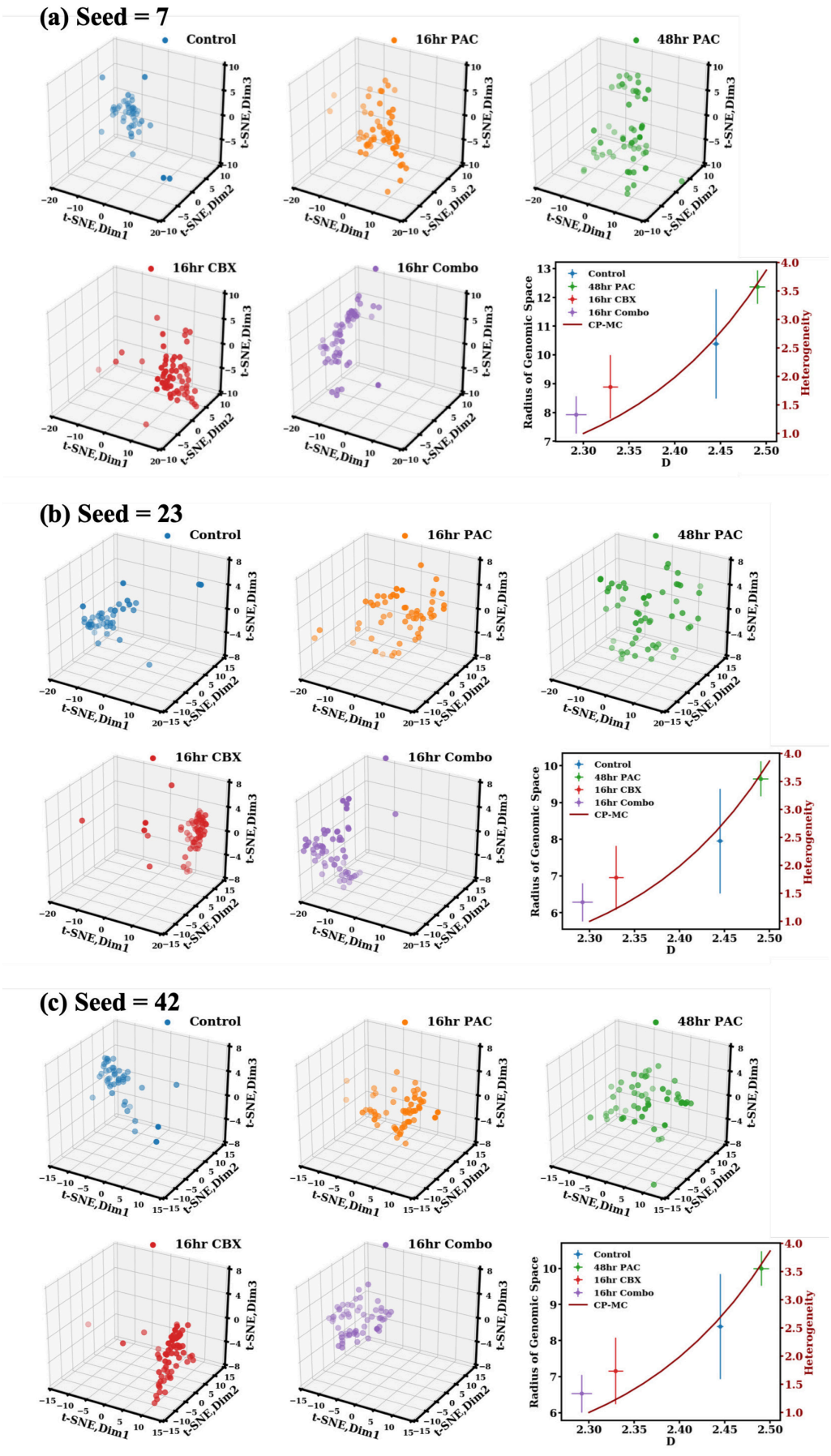
(a) Seed = 7
Control
16hr PAC
48hr PAC
16hr CBX
16hr Combo
t-SNE,Dim1
t-SNE,Dim2
t-SNE,Dim3
Control
48hr PAC
16hr CBX
16hr Combo
CP-MC
Radius of Genomic Space
Heterogeneity
D
(b) Seed = 23
Control
16hr PAC
48hr PAC
16hr CBX
16hr Combo
t-SNE,Dim1
t-SNE,Dim2
t-SNE,Dim3
Control
48hr PAC
16hr CBX
16hr Combo
CP-MC
Radius of Genomic Space
Heterogeneity
D
(c) Seed = 42
Control
16hr PAC
48hr PAC
16hr CBX
16hr Combo
t-SNE,Dim1
t-SNE,Dim2
t-SNE,Dim3
Control
48hr PAC
16hr CBX
16hr Combo
CP-MC
Radius of Genomic Space
Heterogeneity
D

Figure S6. t-SNE transcriptional heterogeneity analysis is independent of seed. (a-c) t-SNE dimensionality reduction from a space of 8,275 genes down to 3D for conditions defined in main text Figure 3. t-SNE is a probabilistic algorithm that attempts to reduce dimensionality while maintaining a similar distribution of Euclidean distance between each cell. Although each iteration results in a different projection in 3D space in terms of coordinates, the overall trend in the spread of transcriptional states between treatment conditions remains the same through all different seeds.

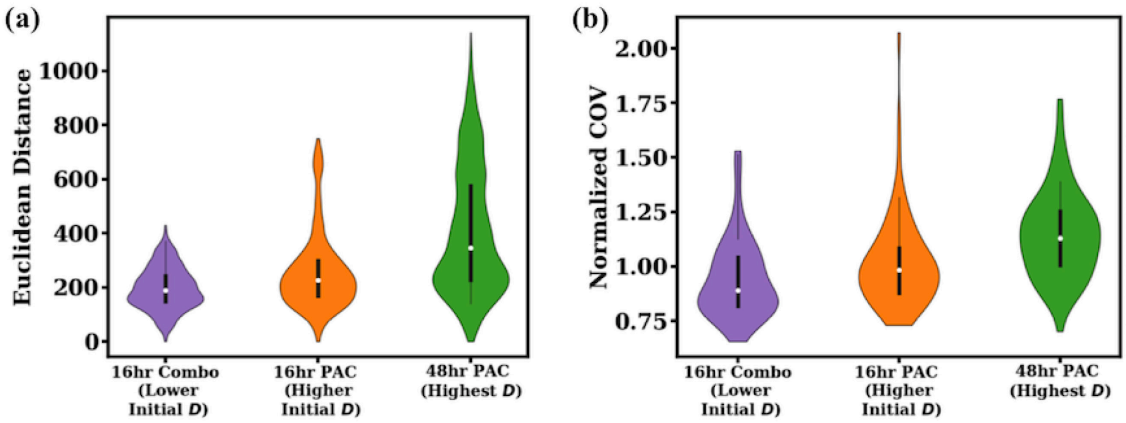

Figure S7. Transcriptional heterogeneity is increased in high *D* cells. (a) Spread of pairwise Euclidean distance was calculated between cells in each condition for genes associated with DNA repair pathways that are upregulated in 48 hour paclitaxel treated cells. **(b)** Coefficient of variation (COV) across treatment populations of genes grouped by control expression levels normalized by control COV. Genes were first binned into groups of ~100 genes (80 quantiles total) each based on relative control expression (exposed to roughly similar molecular regulators of transcription) and expression of these genes was averaged within each cell. $COV_j = \frac{\sigma^2_{E_i}}{\mu_{E_i}}$ was calculated over all average expression levels of cells in treatment condition i for genes in control expression quantile j and each non-control condition was normalized to COV calculated for each bin in the control condition.

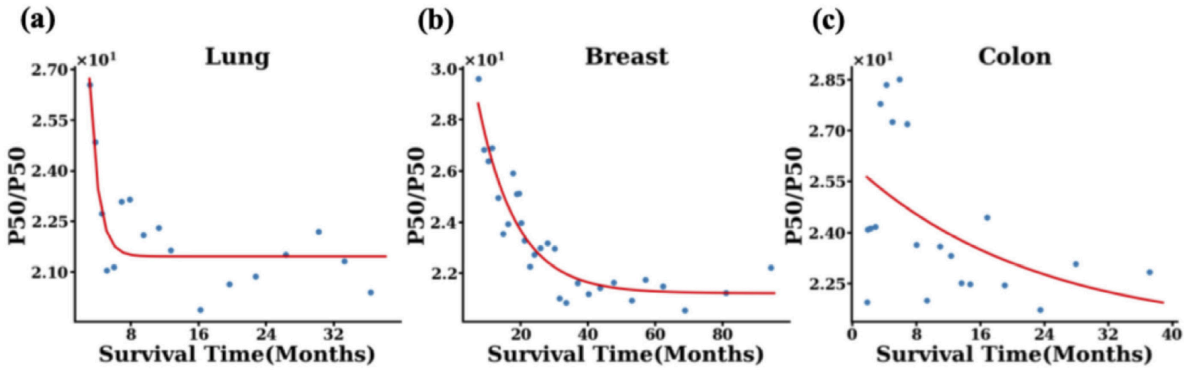

Figure S8. Relationship between survival time and transcriptional divergence. Fixed moving window averaging (MWA) with window size 5 was applied to compare P50/P50 values from sequencing data to patient survival time for **(a)** breast, **(b)** colon, and **(c)** lung cancer patients. Red curves represent the fit to the data points.

FIXED PARAMETERS	DESCRIPTION	VALUE
K_D	Dissociation rate of Pol-II in the absence of crowders	1nM
k_m	Transcription rate of Pol-II in the absence of crowders	$1\mu M^{-1}s^{-1}$
r_{min}	Lower length scale of chromatin self-similarity	1nm
σ^2	Variance of continues crowding density ϕ	$\phi_{in,0}(1-\phi_{in,0})$
L	Average number of base pairs in each gene	6Kbp
r_{in}^0	Radius of interaction volume for single base pair	15nm
N_d	Total mass of upper length scale of chromatin self-similarity	Average for all cell types:~1Mbp Low N_d in A549: 50Kbp High N_d in A549:2Mbp
$\phi_{in,0}$	Average crowding density	HT29 cells: 39% v/v A549 cells:40% v/v BJ cells: 31% v/v A2780: 39% v/v
D_i	Initial chromatin fractal dimension	Wild-type HT-29 cell: 2.7 HT-29 Arid-1a Kd: 2.50 A549 cells: 2.66 BJ cells: 2.66 A2780 cells: 2.50
UNFIXED PARAMETERS	**DESCRPTION**	**VALUE**
$[C]_{tot}$	Total concentration of transcription complexes	[0.035μM, 350μM]

Table S1. Descriptions and values of CPMC model parameters.

Tumor Name	**Tissue**	**Overall Survival**	
		Alive	Dead
Breast Invasive Carcinoma	Breast	152	16
Colon Adenocarcinoma	Colorectal	12	48
Lung Adenocarcinoma	Lung	21	10

Table S2. TGCA Patient Information.