Synthetic analysis of chromatin tracing and live-cell imaging indicates pervasive spatial coupling between genes

Christopher H. Bohrer and Daniel R. Larson*

Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

**Correspondence: dan.larson@nih.gov*

Abstract

The role of the spatial organization of chromosomes in directing transcription remains an outstanding question in gene regulation. Here, we analyze two recent single-cell imaging methodologies applied across hundreds of genes to systematically analyze the contribution of chromosome conformation to transcriptional regulation. Those methodologies are: 1) single-cell chromatin tracing with super-resolution imaging in fixed cells; 2) high throughput labeling and imaging of nascent RNA in living cells. Specifically, we determine the contribution of physical distance to the coordination of transcriptional bursts. We find that individual genes adopt a constrained conformation and reposition toward the centroid of the surrounding chromatin upon activation. Leveraging the variability in distance inherent in single-cell imaging, we show that physical distance – but not genomic distance – between genes on individual chromosomes is the major factor driving co-bursting. By combining this analysis with live-cell imaging, we arrive at a corrected transcriptional correlation of $\phi \approx 0.3$ for genes separated by < 400 nm. We propose that this surprisingly large correlation represents a physical property of human chromosomes and establishes a benchmark for future experimental studies.

Introduction

The role of spatial heterogeneity in the nucleus in relationship to gene regulation is an enduring question in cell biology (4). Heterogeneity or compartmentalization is visible at all length and genomic scales, starting from gene loops and proceeding through enhancer-promoter interactions, topologically associated domains, A/B compartments, chromosome territories, up to inter-chromosomal interactions such as the nucleolus, Cajal bodies, and histone locus bodies, and extending to prominent nucleus-wide features such as lamin-associated domains and heterochromatin (35). The synergy between microscopy (mostly light microscopy but also electron microscopy (39)) and chromosome conformation capture approaches has led to fundamental insights of how molecular features drive genome organization, the influence they have on gene regulation, and the extent to which genome organization varies within individual cells.

Yet, the chromatin-transcription relationship at length scales smaller than the wavelength of visible light ($\sim$ 500 nm) remains challenging to dissect. Foundational work from Cook and colleagues introduced the notion of the transcriptional factory. Transcription factories are areas with an enrichment of transcription machinery (10–12, 15, 25), where genes are thought to be transiently bridged to enable efficient transcription (17). Ensemble chromosome conformation capture seems to support this model by revealing that promoter-promoter contacts (smaller than 1 Mb) form as transcription levels increase (26, 32, 47, 51). The model is that actively transcribed genes are positioned to transcription factories. The prediction is that genes which are close in 3d space (nm) will "feel" the same enrichment in transcription machinery and exhibit correlated transcriptional bursts. Indeed, genes on the same chromosome (41, 44, 46, 50) and genes that share the same (ensemble) topologically associated domain are more co-expressed in individual cells (RNA) (45). However, correlations were not seen between nascent transcripts (31) and the genomic distance between genes was found to show a more dominant role in RNA co-expression than Hi-C contact frequency (44). Furthermore, single-cell RNA-seq showed little to no difference in correlation between genes from the same chromosome with an increased contact frequency, given a similar genomic distance between the two, bringing the strength of the hypothesis into question (45).

This static factory view was supplanted by one in which local heterogeneity of the transcription machinery was due to dynamic assembly and disassembly (12, 15, 25). Thus, the "factory" was not a fixed assemblage but rather a transient and movable conglomeration of RNA polymerase II, general transcription factors, and nascent RNA which arose in connection to active transcription units. It is clear that these diffraction-limited spots observed in the fluorescence microscope exchange constituents with the surrounding nucleoplasm. However, the number of terms used to describe these spots – "factories", "foci", "hubs", "clusters", "speckles", "compartments", "condensates", "phases" – emphasises the lack of a consensus model in the field. Further, it should be noted that many of the utilized super-resolution methodologies are prone to artifacts (6). Consequently, the physical interactions between protein, DNA, and RNA and the dynamic changes in chromosome structure which precede RNA synthesis are hotly debated.

Recent advances in single-cell imaging shed light on these questions and motivate the fully theoretical analysis in this paper. First, the development of chromatin tracing of an entire chromosome using super-resolution light microscopy provides a spatial map of the chromatin fiber at $\approx$ 100 nm resolution (27, 43). When coupled with single-molecule fluorescence in situ hybridization (smFISH) to look at nascent RNA, one can then connect chromatin conformation to transcriptional activity with single-cell resolution (43). Specifically, the nascent transcription state of $\sim$80 genes as well as the 3D centroid positions of 651 50kb chromosomal segments was quantified for thousands of individual chromosomes in IMR90 cells (Fig. 1A). Second, the application of single-cell imaging of nascent RNA in living cells provides critical information on temporal heterogeneity to interpret the observations of spatial heterogeneity. For example, transcriptional bursting of human genes expressed in their native genomic context can be monitored with high spatial and temporal precision for hours (42, 49).

Here, we take advantage of two single-cell data sets – chromatin tracing in fixed cells and nascent RNA imaging in living cells to address two questions: 1) do genes reposition upon transcriptional activation? 2) do genes in spatial proximity show correlations in transcriptional activity? Our analysis indicates that with transcription, chromatin adopts a constrained structure and the gene is positioned toward the centroid of the surrounding chromatin. We then probed the distances between genes and found that genes are positioned closer to each other with transcriptional bursts when the genomic distance between them below 5 Mb, and genes were positioned further away from each other with transcription if the genomic distance was above 5 Mb. Importantly, by capitalizing upon the fluctuations of distances between genes on individual chromosomes, we found that the physical distance between genes on individual chromosomes is the major factor driving the transcriptional co-bursting between genes. By incorporating temporal information from live-cell imaging of active genes (duration of active periods and mobility of active genes) we can infer the correlation between transcriptional bursts for proximal genes to be $\phi \approx 0.3$. Overall, our synthetic analysis of these two single-cell data sets indicates that indeed genes do reposition upon activation and show concomitant correlation between individual transcriptional bursts.

Results

Active promoters are positioned to locations defined by chromatin organization

To investigate spatial changes in the chromatin fiber for active and inactive genes, we re-analyzed data from combined super-resolution imaging of DNA and RNA FISH (43). We performed a spatial metagene analysis consisting of "centering" the chromatin around the promoter of the each gene, quantifying the standard deviations (STD) of the distances between the chromosomal loci, and then averaging over all available genes. Note, we utilized the centroid position of the chromosomal segment which contained the transcriptional start site of each gene as the location of the promoter for the gene and only utilized the chromosome tracing by sequential hybridization data (43). This analysis was done for chromosomal segments where genes were 'off' (0) or 'on' (1) (Fig. 1D,E) — we utilize Boolean logic (0 or 1) throughout to describe transcription states based on the absence (0) or presence (1) of nascent RNA. We observed that chromatin centered around the promoter shows less variability while transcribed, again as determined by the presence of nascent RNA. To more clearly visualize distinctions between chromatin configuration +/- nascent RNA, we quantified the difference and found a noticeable red cross (Fig. 1F), indicating that the distances from a promoter to the surrounding chromatin are more restricted with transcription.

The change in confinement could be the result of repositioning active genes to a different nuclear environment. To probe whether gene positioning varies with transcription, we performed a similar analysis but quantified the median physical distance (MPD) between chromosomal loci with and without transcription and quantified the average over all available genes (Fig. 1G,H). Again, we quantified the difference between them and found a similar red cross (Fig. 1I), suggesting that when a gene is active the promoter is on average closer to the surrounding chromatin and the distances between non-promoter chromosomal segments are unperturbed.

Intuitively, a possible reason for the distance to decrease to surrounding chromatin with transcription (on average) is if a gene is located closer to the centroid of the surrounding chromatin for single-chromosomes when active. To test this supposition, we calculated the mean distance of the promoter of the gene to the centroid of the surrounding chromatin with and without transcription (Fig. 1J). The centroid was calculated for windows of various genomic size around each gene — that is, if the amount of chromatin around a promoter site included in the centroid calculation was .5Mb, .25Mb on both sides of the gene were included in the centroid calculation. Tellingly, we found a definitive difference between active promoters (1) and inactive promoters (0): the active promoters were closer to the centroids of the surrounding chromatin (Fig. 1J). Note, that the mean distance from a local centroid to an inactive promoter gives one an idea to natural spread of the chromatin. To understand this phenomenon on a gene by gene basis, we quantified the difference between the active promoter and inactive promoter for each gene (Fig. 1K). We found that even though there are overlaps in the distributions in Fig. 1J, nearly every gene was closer to the centroid with nascent transcription — suggesting a general phenomenon. Overall, these results indicate that transcriptionally active genes are located toward the centroid of surrounding chromatin.

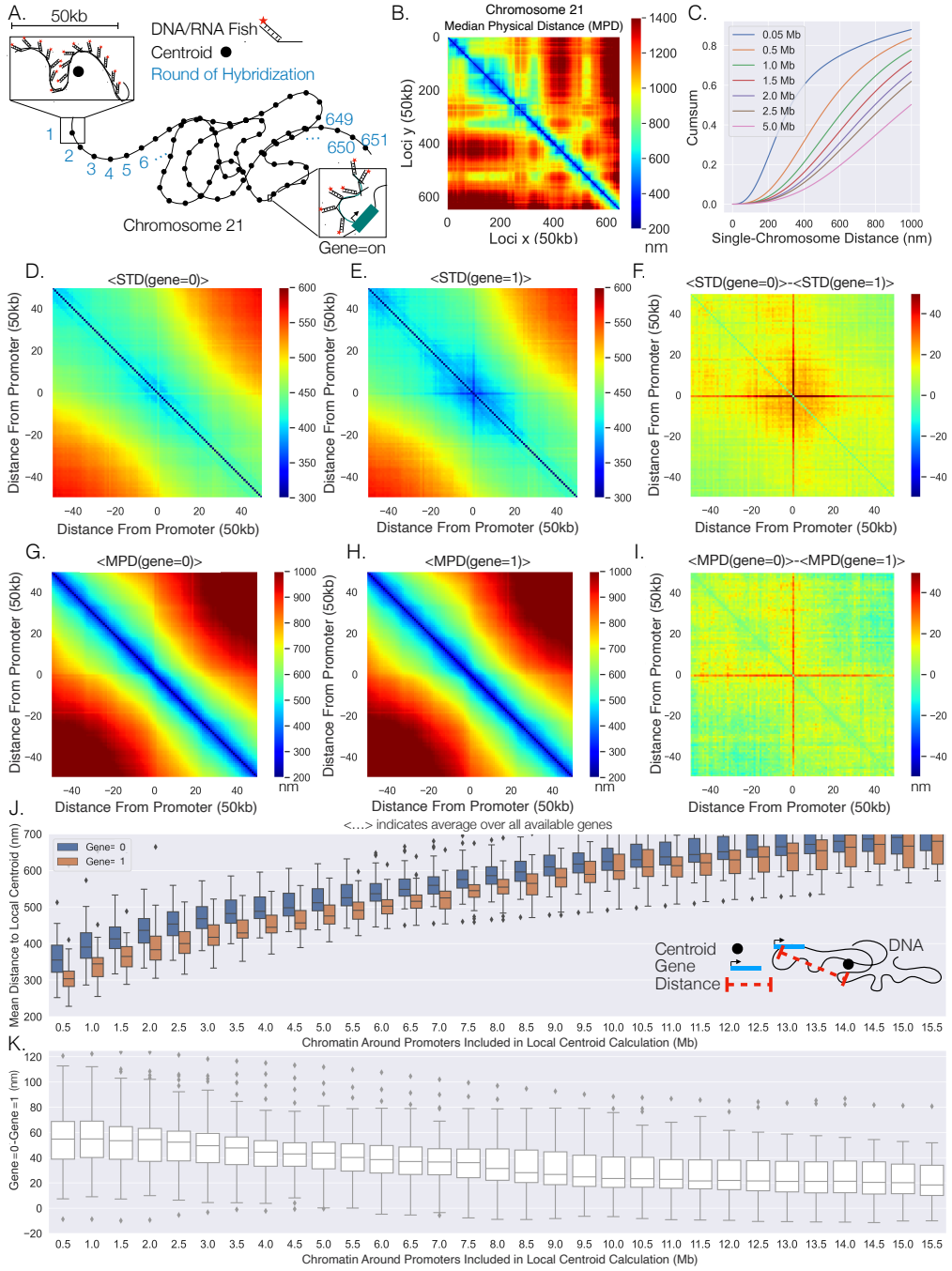

Figure 1: **Transcription confines chromatin and active promoters are located toward the centroid of their surrounding chromatin:** A: An illustration of the chromatin tracing data where each chromosomal loci is imaged through different rounds of hybridization and the centroid of each 50 kb region is determined, and a gene was determined as 'on' (1) or 'off' (0) with RNA Fish. B: The median physical distances (MPD) between all loci determined on chromosome 21. C: The cumulative distribution function of the distance between chromosomal loci separated by various genomic distances — all loci with a given genomic distance were used to generate these distributions. D: An aggregate analysis, calculating the standard deviation (STD) of the distances between chromosomal loci for chromosomes where a gene=0, centered around the loci containing the promoter of an available gene, and then averaging over all genes. E: The same as D but with gene=1. F: The difference in the average centered STD in D and E. G: Similar to D but quantifying the MPD instead of the STD. H: The same as G but for chromosomes where gene=1. I: The difference between the average centered MPD in G and H. J: The mean distances between chromosomal loci containing genes to the centroid of the surrounding chromatin when the genes were either on (1) or off (0) vs the amount of chromatin around the promoter included in the centroid calculation. There is also an illustration of this calculation in the far right corner to aid interpretation. K: The difference between the mean distances to the local centroid when gene=0 and gene=1, showing the results in J on a gene by gene basis. Box plots show quartiles and whiskers expand to 1.5x interquartile range, black diamonds are outliers.

Having considered genes individually based on activity (first order moments), we next sought to quantify higher-order moments such as pairwise interactions in promoter-promoter distances based on transcriptional activity. We first quantified

the average distances between promoters when [both genes were off, (0,0)], [both were on, (1,1)], [one was off and one was on, (0,1)] and quantified them as a function of the genomic separation between them (Fig. 2A). We also quantified the average distances between chromosomal loci that did not contain the investigated genes as a reference control (Fig. 2A). We found that the distances between genes were consistently smaller with transcription for short genomic distances (< 1.5 Mb), as evidenced by the significant decrease in the (0,1) and (1,1) interactions compared to the (0,0) interaction. Importantly, when we compared (0,0) to the no gene control, we saw essentially no difference — suggesting transcriptional bursting (or a consequence of bursting) is correlated with the formation of promoter-promoter contacts.

To probe the distance changes on a gene-pair by gene-pair basis, we first calculated the mean distance between inactive genes on the same chromosome (0,0) and then subtracted the mean distance between the genes when active [(1,1) or (0,1)] — similar to the analysis in Fig. 1K. This analysis is shown as a function of the genomic distance between genes in Fig. 2B. For genomically proximal genes, we observed that when both genes were active the mean distances between the promoters were indeed closer to each other. When we compared the (0,0)-(0,1) to (0,0)-(1,1), the later difference was approximately twice the former difference. Interestingly, we observed that as the genomic distance increased, the difference for both seemed to approach a negative value, suggesting that sufficiently separated genes are positioned to different locations with transcription. However, the spread within the box plots suggests much variability in whether genes are positioned toward the same or different location with transcription. Overall, these analyses provide strong evidence that the spatial separation between genes depends on individual transcriptional bursts.

These analyses suggest a characteristic genomic length scale over which pairwise interactions might occur. However, since genomic distance and physical distance between chromosomal segments are obviously correlated (3, 43, 44), either might define the length scale and drive re-positioning with transcriptional bursting. To probe the general impact of MPD, we characterized the positioning of genes toward the same or different location with transcription based on the 3d distance between the genes. Note, this analysis is only possible with microscopy data sets such as this one (43). We performed the previous analysis as a function of the MPD between the genes (Fig. 2C) and found a strong decay with increasing MPD. The (0,0)-(0,1) resulted in a strong majority of values being negative for MPD above 1300 nm — indicating that the genes move away from each other with bursting above this spatial threshold. The (0,0)-(1,1) had a majority of negative values for MPD above 1300 nm but the proportion with positive values was higher.

To disentangle the effects of the two variables, we quantified how deviations from the expected influenced re-positioning. Given the stronger trend with the MPD, we first quantified the difference as a function of the MPD minus the expected MPD. The expected MPD was calculated utilizing all chromosomal loci and was defined as the average MPD for each genomic distance (Methods). We found that for both scenarios a smaller than expected MPD resulted in genes moving toward each other with transcription and a larger than expected MPD led to the genes moving away from each other (Fig. 2D,E), though the later was less clear for the (0,0)-(1,1). These results suggest that the positioning of genes in physical space influences the outcome of pairwise interactions: genes which are close to each other (MPD $<$ 1100 nm) move closer when bursting, and genes which are far from each other separate when bursting. Similarly, to investigate if the genomic distance plays a role, we performed the analysis but as a function of the genomic distance minus the expected genomic distance — the genomic distance given the MPD (Methods). We found that the analysis did not have a monotonic trend, and instead peaked at zero (Fig. 2F+G). If there were a simple relationship between genomic distance and re-positioning, one would expect a monotonic trend and therefore it seems unlikely that genomic distance drives this phenomenon. Additionally, we found that the zero peak was enriched for gene pairs with low MPDs — as we just demonstrated: low MPDs lead to genes moving toward each other (Fig. 2D+E). In summary, these results suggest that the MPD is predictive of whether genes move toward or away from each other with transcription.

Physical distance – but not genomic distance – correlates with co-expression

Our analysis of the DNA/RNA FISH dataset indicates that spatial gene positioning is correlated with transcriptional activity both in isolation (re-positioning of individual genes with transcription) and in pairwise interactions. One can conceptualize the conclusions of this analysis as understanding spatial position given the transcriptional state. In other words, knowledge of transcription state imparts knowledge of spatial position. We next turned to the inverse question of whether correlations exist between nascent RNA (nRNA, transcriptional state) based on spatial proximity. To do so, we quantified the ϕ correlation coefficient (Methods) between genes on individual chromosomes (Fig. 1A) and plotted it as a function of the genomic distance (Fig. 3A). Note, due to the binary nature of the data (0 or 1), the ϕ correlation coefficient is equivalent to the Pearson and Spearman. With approximately a two-fold increase at smaller genomic distances, the correlation showed a monotonic decay with increasing genomic distance — the .025 plateau persisted with even higher genomic distances (data not shown). The increase in co-expression above the asymptotic baseline persists to $\approx$ 2 Mb. To determine whether, ensemble-chromatin structure is what dictates co-expression, we further quantified the correlation as a function of the contact-frequency (Fig. 3B) and the MPD between their chromosomal segments (Fig. 3C). Here we defined the contact frequency between two genes

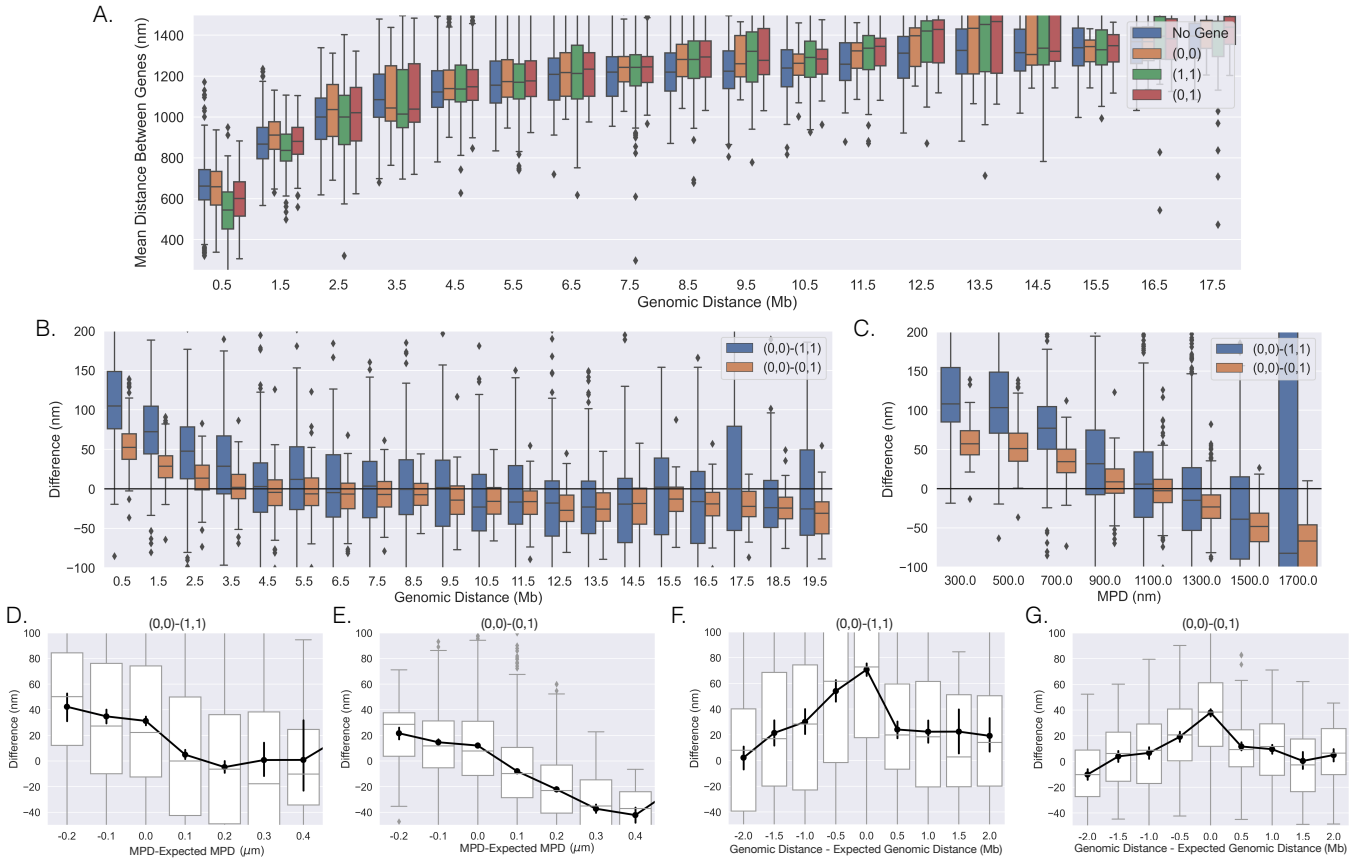

Figure 2: **The distances between genes vary with transcription on individual chromosomes:** A: The mean distances between genes vs. the genomic distance for when both genes were (0,0),(1,1), (0,1) and the mean distances between loci not containing the investigated genes. Box plots show quartiles and whiskers expand to 1.5x interquartile range, black diamonds are outliers. B: The difference between the scenarios shown in A, showing the difference in mean distance on a gene-pair by gene-pair basis, and a black line is shown to aid in visualization of zero. C: The same analysis as in B but vs the MPD between the genes. [D,E,F,G]: The difference shown in B and C but vs either the MPD minus the expected MPD or the genomic distance minus the expected genomic distance (See text). Box plots show quartiles and whiskers expand to 1.5x interquartile range, black diamonds are outliers. Black lines and dots are means and error bars are s.e.m from bootstrapping

as the proportion of chromosomes with distances less than 200 nm between the genes' chromosomal segments using the chromatin tracing data. We observed the predicted monotonic behavior with the average correlation reaching a minimum around .025.

We then attempted to separate the effects of contact-frequency/MPD from genomic distance on the observed correlation, and proceeded to hold one variable constant and quantify the correlation as a function of the other. To do this we calculated the mean correlation given that the contact frequency/MPD and genomic distance between the genes were within a specified range (Fig. 3D,E). Note, we only included averages if more than 40 data-points could be used to calculate the mean. The two showed similar behavior and both had a narrow range for specific genomic distances — making it difficult to uncouple the variables of contact frequency and mean physical distance. For example, we only observed an MPD of 200 nm to 400 nm for genomic distances much less than 1 Mb, therefore we could not determine how the correlation varies with increasing genomic distance for these values. Moreover, most columns and rows did not show significant p-values. In summary, while there is correlation at the nascent RNA level, the limited variability in ensemble-chromatin structure for specific genomic distances obscured the relative contributions of genomic distance, contact-frequency or MPD to co-expression.

A primary advantage of the single-cell dataset (43) is the ability to leverage the large fluctuations of distances between loci across the population (N ≈ 7600 chromosomes) (Fig. 1C). We first quantified the correlation between nascent RNA for genes given that their physical distances were within a specific range, which showed a similar monotonic behavior (Fig. 4A). When calculating these correlation coefficients we only included gene-pairs for specific single-chromosome distance ranges when there was at least 100 chromosomes where the distance between the genes were within that range. We then quantified the mean correlation given that their single-chromosome distance and genomic distance were within specified ranges (Fig. 4B).

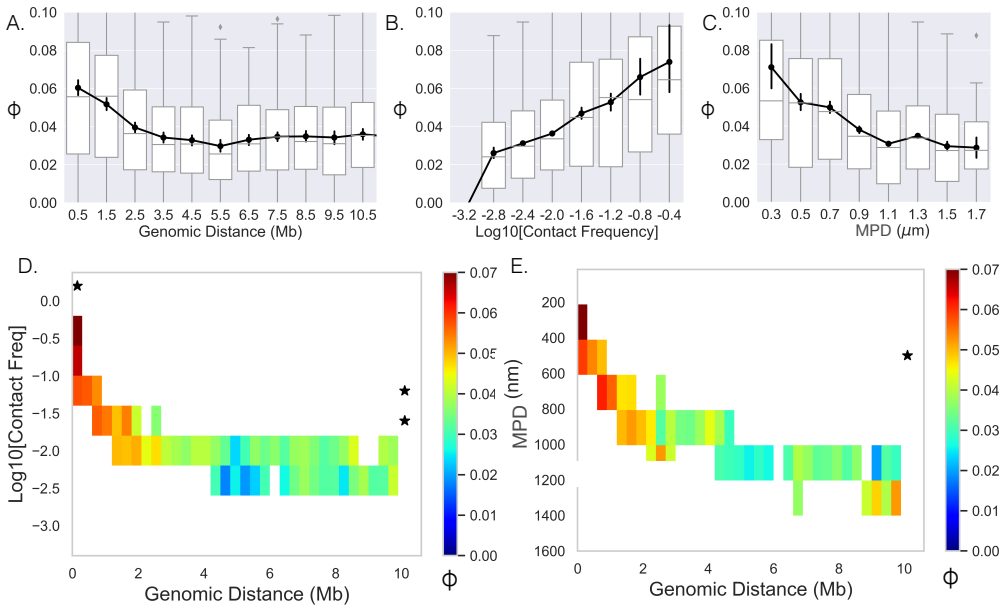

Figure 3: **Limited variability prevents quantification:** [A, B, C]: The Spearman correlation coefficient between genes as a function of genomic distance, contact frequency, and median distance. Black lines and dots are means and error bars are s.e.m from bootstrapping, box plots show the quartiles as above. D: Average correlation coefficients of genes given that their genomic distance and contact frequencies were within a specific range. E: Average correlation coefficient of genes given that their genomic distance and median distance were within the specific range. An * illustrates whether the average correlation coefficients along that dimension are correlated (p-value<.01).

Again, we only included averages if more than 40 data-points (gene-pairs) could be used to calculate the mean. Notably, we observed that co-expression of genes was correlated with the single-chromosome distance between those genes (columns, Fig. 4B). In contrast, we observed no correlation between co-expression and genomic distance (rows). There appeared to be a general decay for the columns with increasing single-chromosome distance, more closely resembling the curve in Fig. 4A, while the rows did not show the behavior. These observations are further solidified by calculations of statistical significance (Fig. 4B).

In summary, these results indicate that co-expression – as quantified through correlations in nascent RNA – is driven by the physical distance between genes on individual chromosomes, uncoupled from genomic distance, which shows no statistical correlation with co-expression.

Chromosome dynamics can obscure the true correlation between physical proximity and gene co-expression.

The single-cell DNA/RNA FISH approach provides exceptional spatial resolution coupled with transcriptional activity, but a potential issue with fixed-cell methodologies is the lack of temporal information. For example, in terms of quantifying the distance dependence on co-expression, the lack of time resolved locus position data could distort the observed distance co-expression relationship. First, the motion of the genes within the on time (defined here as the time it takes for the nascent RNA to dissociate from the DNA) obscures the measurement of the distance at the beginning of a transcriptional co-burst. Second, the stochasticity of the on time would similarly lead to a decrease in the observed co-expression — that is, even if two genes burst at the exact same time, the nascent RNA from one gene will dissociate before the nascent RNA of the other gene, leading to the detection of one and not the other, again decreasing the correlation measured in fixed cells. Third, the error due to the localization precision of the experiment would also distort the distance co-expression curve due to the error in knowing the true distance. Overall, these three sources of noise have the potential to change both the amplitude and distance dependent decay of the co-expression correlation coefficient. Therefore, we utilized a theoretical approach to infer the instantaneous distance co-expression relationship analogous to that shown in Fig. 4A and to thereby understand the contribution of dynamic and temporal fluctuations in gene position and activity. The approach is based on coupling measurements of locus diffusion and activity generated from live-cell imaging of nascent RNA with the fixed cell measurements analyzed thus far. Here, we first discuss our theoretical approach and then our results.

We sought to link the information from live-cell experiments with that of fixed-cell experiments by incorporating the motion

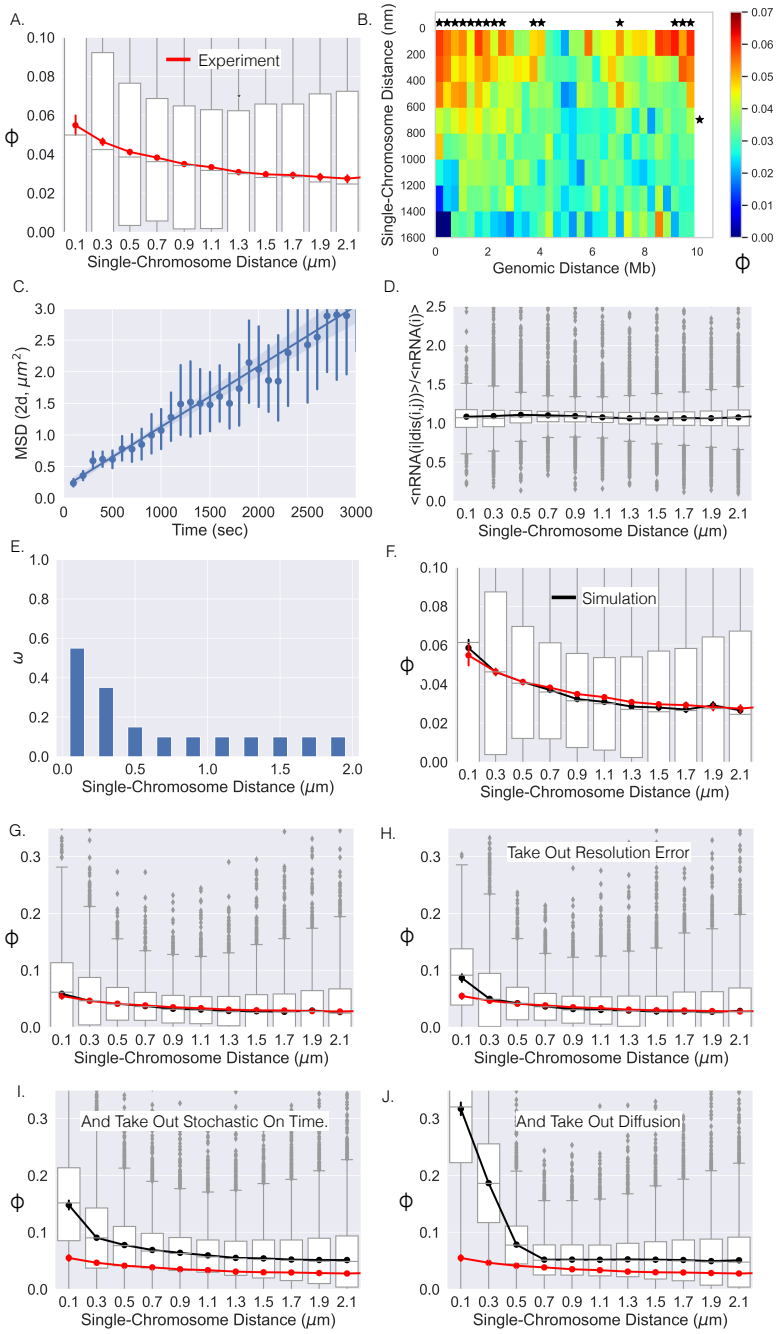

Figure 4: **Single-chromosome distance dictates nRNA correlation:** A: The correlation coefficients between genes as a function of single-chromosome distance. B: Average correlation coefficients of genes given that their genomic distance and single-chromosomal distance were within a specific range. An * illustrates whether the average correlation coefficients along that dimension are correlated (p-value<.01). C: The mean-squared-displacement of active *TFF1*, the fitted line and 95% confidence interval shaded (Error bars are individual 95% confidence intervals). D: The average number chromosomes with nRNA for gene i given the distance between gene j and i divided by the average with all distances. E: The optimal ω function for the model which results in the black curve in F. F: The correlation-distance relationship for all pairs of genes from the simulation utilizing the ω function in E. The box plots here are from simulation, red curve is shown for reference and is the experimental data from A. G: The same as F but on a different scale. H: The results of the simulation without resolution error of the experiment. I: Simulation results without resolution error and with nRNAs having a deterministic on time. J: Simulation results without resolution error, with deterministic on times, and no chromatin diffusion for all pairs of genes.

of chromatin into our model. Chromatin has been suggested to show confined diffusion (7, 8, 13, 34), but this phenomenon is generally quantified over relatively short timescales of < 10 minutes. Considering the on time of a human gene — as measured by the dwell time of nascent RNA — is approximately 10 to 15 minutes (49), we sought to monitor the diffusion of an active gene over a longer timescale. We utilized the live-cell transcriptional bursting data of *TFF1* from Rodriguez et al. (42). This data consists of the spatial coordinates of multiple bursting *TFF1* alleles through time in individual cells,

allowing us to quantify the motion of one allele relative to the other (13). We quantified the mean squared displacement (MSD, Methods) over a timescale of 3,000 seconds and found that the MSD could be fit with a straight line (Fig. 4C) — suggesting Brownian motion of active genes over these timescales (5). We computed a diffusion coefficient of $.25\text{x}10^{-3}\mu m^2/s$, which is comparable to previous results (13).

We chose to utilize the over-dampened Langevin equation to model the temporal dynamics of the distance between genes located on the same polymer. The model describes the time-dependent distance between loci using an arbitrary energy potential of interaction (see Methods) — without the effect of the potential the model exhibits Brownian motion with the determined diffusion coefficient. For each gene pair, we empirically determined a potential that "biases" the distance motion so the steady state distribution matches the empirically determined distance distribution (Methods). We did this using the equivalent Fokker-Planck equation which allowed us to directly convert the empirically defined distance distributions into the potential (Methods). The central advantage of this approach is that it accounts for the unique distance distributions between the various gene pairs on the same chromosome, the diversity of which can be clearly seen with the MPDs in Fig. 1B. The diverse distance distributions result from a multitude of complex context-specific forces which are not considered in the classical polymer models (38, 48). Even with the inclusion of additional factors in polymer models (exp. loop extrusion), reproducing accurate distance distributions is difficult (20) — and would be even more difficult here due to lack of knowledge as to the underlying forces. Also, more simple first order approximations of the Langevin equation have been utilized to model the viscoelastic properties of chromatin (48), which has been shown to adequately determine the potential of the Rouse chain (1). Again, we emphasize that these gene-specific terms were determined empirically (Methods).

The stochastic dwell time of nascent RNA is due to variability in the processes of elongation, termination and splicing. We incorporate this variability in our analysis by setting the nascent RNA decay probability per second (propensity) equal for all genes (P_d) with a characteristic on time equal to ≈13 min. This assumption is motivated by our recent work on high throughput imaging of hundreds of human genes labeled at their endogenous loci using MS2 stem loops (49).

Next we introduce a phenomenological model intended to capture the empirical features of co-expression as observed in the fixed cell data sets. First, we quantified the average fraction of chromosomes with nascent RNA present for gene i as a function of the distance between each pair of genes (genes i and j), normalized by the average fraction of chromosomes with nascent RNA present for gene i over all distances. This metric is a proxy for the burst-frequency and was calculated for each gene for all possible gene pairs. The reasoning is that if this metric is higher at smaller distances, it would suggest that the bursting frequency is dependent upon the distance between genes, hence leading to the higher correlation values at smaller distances. Surprisingly, we found that on a distance binning scale of 200 nm, the metric did not vary, suggesting that the bursting-frequency does not generally change as function of distance between genes at this scale (Fig. 4D). Therefore, we set the probability of nascent RNA production per second equal to a constant for each gene (i), P_i^{tot}, which we determined empirically for each gene (Methods). To account for co-expression, we modeled nascent RNA production as coming either from a co-burst or from an individual burst:

$$P_i^{tot} = P_{ij}(r_{ij}(t)) + P_i(r_{ij}(t)), \tag{1}$$

$$P_j^{tot} = P_{ij}(r_{ij}(t)) + P_j(r_{ij}(t)). \tag{2}$$

Here $P_{ij}(r_{ij})$ is the probability of a transcriptional co-burst per second given the distance between the two genes, $P_i(r_{ij})$ is the probability of an individual burst per second given the distance, and $r_{ij}(t)$ was determined beforehand utilizing the above Langevin equation specific for that gene pair (Methods).

The fact that genes have different expression levels, limits the values of $P_{ij}(r_{ij}(t))$. Arranging the pair of genes so that $P_i^{tot} < P_j^{tot}$, the maximum value that $P_{ij}(r_{ij}(t))$ can be is P_i^{tot} — or else $P_i(r_{ij}(t))$ would have to be negative. With this, we can then rewrite the above as the following:

$$P_{ij}(r_{ij}(t)) = \omega(r_{ij}(t)) \times P_i^{tot}, \tag{3}$$

$$P_i(r_{ij}(t)) = P_i^{tot} - \omega(r_{ij}(t)) \times P_i^{tot}, \tag{4}$$

$$P_j(r_{ij}(t)) = P_j^{tot} - \omega(r_{ij}(t)) \times P_i^{tot}, \tag{5}$$

where $\omega(r_{ij}(t))$ is a function of distance between the genes and ranges between 0 and 1. $\omega(r_{ij}(t))$ is the proportion of gene i's transcriptional bursts that are co-bursts at each distance; if the expression levels of the two genes are approximately equal $\omega(r_{ij})$ is equal to the proportion of bursts that are co-bursts at a given distance for both genes.

Overall, with a single function ($\omega(r_{ij}(t))$), we modeled all pairs of genes with the following stochastic reactions utilizing the Gillespie algorithm (22):

$$0 \xrightarrow{P_{ij}(r_{ij}(t))} nRNA_i + nRNA_j,$$

$$0 \xrightarrow{P_i(r_{ij}(t))} nRNA_i,$$
$$0 \xrightarrow{P_j(r_{ij}(t))} nRNA_j,$$
$$nRNA_i \xrightarrow{P_d} 0,$$
$$nRNA_j \xrightarrow{P_d} 0.$$

More specifically, we simulated thousands of trajectories (15000 seconds each) for each pair of genes for a given $\omega(r_{ij}(t))$ akin to the number of chromosomes within the experimental data. If the amount of nascent RNA for a gene was greater than 0 at the end of the trajectory the gene was considered "on" (Gene=1) — making our simulation data binary like the experimental data. Lastly, we incorporated the error due to the resolution of the experiment (resolution =100 nm, Methods). In total, using this numerical simulation approach, we are able to generate curves like Fig. 4A, for a given coupling coefficient $\omega(r_{ij}(t))$, from the underlying spatiotemporal fluctuations of single genes in living cells. Importantly, the diffusive properties of active genes and the dwell time of nascent RNA are derived empirically from experimental data. Of the parameters described above, the coupling coefficient is the least well-determined and lacks an underlying mechanistic motivation at present.

Is it possible for a single function ($\omega(r_{ij}(t))$) to adequately reproduce the experimental results (Fig. 4A)? To address this question we iterated over many possible monotonically decreasing ($\omega(r_{ij}(t))$) functions. More specifically, we investigated all possible monotonically decreasing functions in .05 increments, with specific values for distances binned at a 200 nm resolution (Methods, Fig. 4E). For each $\omega(r_{ij}(t))$ we quantified the correlation-distance curve for each gene pair and sought to find the one that was closest to Fig. 4A (Methods). The best performing $\omega(r_{ij}(t))$ is shown in Fig. 4E which resulted in the correlation-distance dependence in Fig. 4F — demonstrating that a single general function can adequately describe this phenomenon at the level of the chromatin tracing experiment.

With this dependence in hand, we are able to computationally remove processes that distort the correlation-distance relationship in an effort to uncover the "true" observable degree of correlation for a given distance. The correlation-distance relationship in Fig. 4F is also shown in Fig. 4G with a new y-axis range to aid comparison. We started by simulating all pairs of genes as before but without the resolution error of the experiment with the determined $\omega(r_{ij}(t))$ (Fig. 4H). Removing resolution error associated with light microscopy resulted in a slight increase in the correlation for the first distance bin, resulting in an 66% increase (Fig. 4H). For all other distances, the degree of correlation was basically unchanged. We then simulated the system without resolution error and with a deterministic on time for each nascent RNA — each nascent RNA lasted exactly 800 seconds. We observed a much greater increase across all distances with the first distance bin rising to 250% of its initial value (Fig. 4I). Finally, we simulated the system removing resolution error, with deterministic on times, and without diffusion. Removing these three noise sources resulted in a large increase in correlation for lower distances and a slight decrease for larger distances (Fig. 4J). This latter decrease is due to the correlated bursts at small distances not being able to diffuse to larger distance. For the first distance bin, the removal of all sources of error in fixed cell experiments leads to a $\approx$5 fold increase. The correlation is surprisingly high ($\approx$ 0.3) and extends over a spatial distance of $\approx$ 400 nm. Additionally, this analysis also suggests that if one was able to monitor the distance between genes with high resolution and at time resolution where one could determine the exact start of each transcriptional burst, one should be able to see this true relationship — a clear direction for future pursuit.

Discussion

By capitalizing upon the single-chromosomal nature of chromatin tracing and nascent RNA smFISH data (43), we discovered a variety of phenomena related to the coupling between transcription and higher order chromosome conformation. Specifically, fixed-cell analysis of chromatin conformation and activity coupled with live-cell analysis of transcription dynamics provides two features which are key to the analysis performed here: fluorescence microscopy reveals true physical distances and the variability across single cells. Leveraging these unique features, we find: (1) The chromatin around a gene is "constrained" with transcription; (2) during a transcriptional burst genes are positioned toward the centroid of their surrounding chromatin; (3) transcriptional bursts cause promoters to move toward or away from each other depending on the median physical distance between them; (4) the distance between genes in individual cells is predictive of co-expression; (5) the lack of temporal information and limited imaging resolution greatly reduces the true distance-correlation relationship, with the predicted correlation coefficient of $\sim$ .3 for distance below 400 nm. This last finding relies on theoretical assumptions regarding chromatin mobility and the precise molecular nature of gene co-expression and awaits future experimental validation.

Genes reposition upon transcriptional activation

Our finding that individual transcriptional bursts lead to the repositioning of genes and lower chromatin variability suggests the two phenomena could be linked. The traditional view of transcription influencing the dynamics of chromatin is that transcription leads to more "open" and dynamic chromatin (2). While the traditional view has some empirical support (23), the exact opposite has been observed (21, 36, 37). Accepting the variability of distance distributions as a proxy for the

motion of chromatin puts our observations in agreement with the latter. One possibility is once a gene is positioned toward the centroid of the surrounding chromatin, the confinement could be due to a new micro-environment. Another possibility — which we favor — is that the movement toward the centroid is a steric effect. Active genes recruit large megadalton complexes such as the pre-initiation complex and RNA polymerase II, which 'pushes' and confines the gene to a specific location due to the occluded volume effect. Our analysis thus suggests behavior consistent with the original factory model (genes reposition to a factory upon activation) and also the dynamic self-assembly model (genes assemble their own transcription factory). The order of events is key to distinguishing these alternatives, and these events are not resolved in the fixed-cell data sets analyzed here (10–12, 15, 25). Nevertheless, almost all of the ≈80 genes showed this behavior of repositioning and confinement, suggesting a general phenomenon — illustrating a fundamental aspect of transcription whose mechanistic details await additional study.

On a higher level, promoter-promoter distances (26) are clearly variable with individual transcriptional bursts and are likely important for understanding enhancer-biology and other higher order functional assemblies. Considering the functional similarity between promoters and enhancers (29), we speculate that the rules of promoter-promoter interaction observed here may apply to enhancer-promoter interaction. In most cases the distance change of promoters with transcription is small when compared to the median physical distance (MPD), but for MPD$<$ 400 nm a repositioning of 100 nm could be functionally relevant (Fig. 2C) (6, 9, 19, 24, 32) — putting the distances at the scale of enhancer-promoter communication (9). On the other hand, transcription factories have also been shown to be highly dynamic (12, 15, 25), raising the question of whether these dynamic promoter-promoter distances are linked to the dynamics of the factories (24). The unexpected finding that high MPD promoters tend to move away from each other with transcription suggests the possibility of specific locations for transcription, but this observation might also be used to explain specificity of enhancer-promoter interactions. Intriguingly, whether genes move toward or away from each is dependent upon ensemble chromatin organization, raising the possibility that genes are distributed according to chromatin organization and not genomic distance — given there is an underlying fitness advantage. Finally, it should be noted that for all these results described here there is a lack of temporal information, which obscures the cause and effect of these phenomena (just as we showed for the distance-correlation relationship). It therefore seems likely that these distance changes are likely more significant — a direction for future research.

Genes in spatial proximity show high correlations in transcriptional activity: interpreting $\phi \sim .3$

The hypothesis that genes in close spatial proximity are transcriptionally correlated has long persisted in the field despite conflicting data. Notable studies have taken advantage of single-cell RNA-seq and Hi-C data to disentangle the influence of genomic distance and physical distance on correlation with unclear results (44, 45). For example, while genes from the same (ensemble) topologically associated domain are more co-expressed, intra-chromosomal genes separated by similar genomic distances show essentially no difference in correlation with enrichments in contact frequency (45). The study of Sun et al. even found that the genomic distance is slightly more strongly correlated with co-expression than contact frequency (44) — rightly explained away given the contact frequency was of a lower resolution with high error. Further, nascent RNA FISH found intra-chromosomal genes are not more correlated than when in trans (31). Yet, single-cell imaging experiments coupled with detailed chromosomal perturbations have revealed spatial interactions which dictate a 'hierarchical' organization in multiple genes in response to stimulus (16). Moreover, a recently proposed transcription factor activity 'gradient' model is a diffusion-based model which relies again on the spatial proximity of cis-acting regulatory elements, which might equally well be applied to promoter-promoter interactions (28). Overall, the hypothesis has persisted due to the intuitive mechanism even with the lack of definitive experimental demonstration.

Our results verify the null hypothesis and explain the negative results of previous single-cell studies. We found an enrichment in correlation for nascent RNA given that the genes are separated by a genomic distance of less than 2.5 Mb (Fig. 3A). The fact that the average genomic distance between genes in the previous work was 3 Mb explains why enriched correlations were not seen at the nascent RNA level (31). With our finding that the variability in MPD (or contact frequency) for a given a genomic distance is too low to disentangle these variables (Fig. 3D+E), the defined enrichments in contact frequency for previous studies were likely quite minor in terms of producing a change in correlation (45). Utilizing the large amount of stochasticity in chromatin structure for individual chromosomes (18) definitively shows the physical distance drives co-expression. This result is illustrated with the extremes: we observed an enrichment in correlation for genomic distances up to 10 Mb when the physical distance between genes was less than 200 nm on individual chromosomes, and very low correlations between genes separated by less than .5Mb given that the physical distance was above 1200 nm (Fig. 4B). In summary, our key finding is a correlation gradient with physical distance but not genomic distance.

The lack of temporal data and the spatial resolution limits of the chromatin tracing methodology greatly obscures both the 'true' transcriptional correlation between spatially proximal genes and also the length scale over which transcriptional correlation is measured. The reasons for this reduced correlation are obvious: both the position and the activity status of genes vary randomly. One can imagine for example genes which were far apart at activation and then diffused together and vice versa. Correcting for this behavior requires assumptions about chromatin mobility and also utilization of live-cell nascent RNA

data. We predict that if one were able to measure the distances between genes at the initiation of the transcriptional bursts, one should obtain a correlation of $\sim$.3 if the distance between the promoters of the genes is less than 400 nm. Intriguingly, this level of correlation has been reported between the mRNA levels of adjacent genes in yeast but was attributed to DNA supercoiling (40). Considering the shorter lifetimes of mRNA in yeast, this correlation may be comparable to the nascent RNA in humans. Furthermore, other live-cell studies have seen correlated bursts between spatially proximal genes (in trans and cis), but did not specifically investigate this as a function of the physical distance between the genes or account for the variable on times (19, 24, 32, 33) — finding enrichments in correlation similar to the uncorrected curve (Fig. 4A) (32). The enrichment in co-bursting for genes separated by < 400 nm suggests the working distance of the underlying mechanism is not direct contact. Exactly what mechanism leads to these general correlations is still unknown, however, these results are consistent with the idea of enhancers coordinating transcription with working distances of hundreds of nm (6, 19, 24, 32, 33).

Finally, this analysis suggests co-expression is a general property of the system, that is, unrelated genes show correlated bursts with each other when in spatial proximity. This transcriptional correlation would then be an unavoidable emergent behavior due to the physicality of the system. Hence, the appearance of correlated bursts may not suggest a specific regulatory mechanism. Stated another way: we hypothesize that the physical distance between the vast majority of genes arises from the physical constraints of the nucleus and DNA and is not indicative of a biologically functional relationship requiring coordinated expression conferred by that proximity. Support for this hypothesis comes from the observation that disrupting genomic clusters of metabolic genes such as the *GAL* genes in yeast have no measurable impact on fitness (30). Of course, there are certainly instances where coordinated co-expression conferred by spatial proximity is important, for example in the segmentation clock genes *her1* and *her7* located on the same chromosome and separated by 12 kb (52). The corollary to our hypothesis is that one can look for deviations from the $\phi \sim .3$ to identify bona fide regulatory relationships. Thus, we establish a theoretical benchmark which can be used in future studies.

Methods

Expected MPD and Genomic Distance

To determine the expected MPD for a given genomic distance we simply calculated the average MPD for each specific genomic distance. For example, to determine the expected MPD for a genomic distance of 50 kb, we quantified the average MPD between all loci separated by 50 kb.

To determine the expected genomic distance for a given MPD, we used the same curve and found the genomic distance with the closest average MPD. For example, say the MPD between two loci is 500 nm, using the previously quantified curve, the expected genomic distance is the genomic distance whose average MPD is closest to 500 nm.

Correlations Between Genes

When quantifying the correlations between a pair of genes (aka. whether they were on or off, 1 or 0) we quantified the ϕ coefficient (used for binary data):

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{(n_{11}+n_{10})(n_{11}+n_{01})(n_{00}+n_{10})(n_{00}+n_{01})}},$$

where n_{11} is the number of observations where both genes are active and n_{10} is the number of observations where the first gene is on and the second is off, etc. Here we should state that ϕ is equivalent to the Pearson correlation coefficient and the Spearman correlation coefficient for this data, due to a gene's transcription state being either 1 or 0 — that is, on (1) or off (0).

Determining P_i^{tot}

To determine the bursting propensity for each gene we first conducted many different simulations with P_i^{tot} values ranging from 0 to .05 with our set nRNA decay rate. For each propensity we simulated 2000 trajectories (15000 seconds each). Then, with the last timepoints of each trajectory, we classified the gene as being either "on" or "off", — if the gene's nRNA was greater than zero the gene was classified as "on" (aka 1). We then simply created a lookup table with the average number of "on" states vs. the bursting propensity. To determine a genes specific propensity, we simply calculated the average number of "on" state with the experimental data and found the closest match within the lookup table.

Incorporating Resolution Error

The resolution of the experimental data was previously quantified in the work of Su et al. (43), and the resolution of each chromosomal segment was determined with approximatly 100 nm resolution. The 3d resolution error is not Gaussian due to the Pythagorean theorem and was determined by Churchman et al. (14). Therefore, for our case, the error must be

applied to all three dimensions independently — similar to in Su et al. To do this, with the 'true distance' from the Langivin simulation we randomly decompose the distance into three dimensions — so that the distances along each dimension satisfy the Pythagorean theorem. We then added two random variables of Gaussian noise with standard deviations of 100 nm (one for each loci), generating a new displacement for each dimension with localization error. Lastly, we took the displacements along each dimension with the error and quantified the distance in 3d using the Pythagorean theorem.

Quantifying Best $\omega(r_{ij})$

To determine the ω that captures the behavior of the experimental data, we first generated a large number of unique monotonically decreasing functions. This was first done in .1 iterations and with a distance binning of 200 nm. For example, $\omega^1(r_{ij}) = [.1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$ means genes that are within 200 nm of each other (first number in array) have the value .1, and the rest of the distances have the value 0. We would then iterate and produce the next ω, $\omega^2(r_{ij}) = [.1, .1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$, etc. We then simulated a large number of trajectories for all gene pairs according to the model in the main text with each function. We then quantified the error between each ω's distance-correlation relationship and the experimental data with the following:

$$Error(\omega^k) = \sum_i \sum_j \sum_r |\phi_{ij}^{\omega^k}(r) - \phi_{ij}^{exp}(r)|,$$

where $\phi_{ij}^{\omega^k}(r)$ is the correlation for the gene pair ij given that the observed distances were within the distance bin r (200 nm for each bin) and $\phi_{ij}^{exp}(r)$ is the correlation for the experimental data for that gene pair. Once we found the ω that resulted in the minimum error was found, we then varied the values for distance bins below 1000 nm by plus or minus .05. We then quantified the error again to result in the best fit function shown in the main text.

Mean Squared Displacement (MSD)

We quantified the motion of the *TFF1* gene utilizing the multiple allele data from Rodriguez et al. (42). This live-cell data provided the 2d coordinates of active alleles over extended periods of time, allowing us to monitor the motion of chromatin over a timescale longer than the on time of a gene. To account for the movement of the cell over these long periods, we monitored the motion of one tagged allele relative to another. We then quantified the mean squared displacement (MSD) for a given time (Δt): MSD(Δt)=<[R(t)-R(t-Δt)]^2>. Where R(t) is the position of an allele relative to another, and the arrows are the ensemble average and over all measured trajectories and times.

Modeling distance diffusion

To model the distance between two chromosomal loci we utilized the following Langevin equation:

$$\frac{dr_{ij}}{dt} = -\frac{1}{\gamma_{ij}}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}} + \sqrt{2D} \times g(t).$$

Here r_{ij} is the distance between genes i and j, $V_{ij}(r_{ij})$ is the potential (specific to that gene pair, described below), γ_{ij} is a constant specific for that gene pair, and the last term $\sqrt{2D} \times g(t)$ accounts for the Brownian motion with the determined diffusion coefficient — if the potential is a constant independent of distance, r_{ij} will exhibit Brownian motion. For each gene pair, we empirically determined a $\frac{1}{\gamma_{ij}}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}$ that "biases" the distance's motion so the steady state distribution matches the empirically determined distance distribution (corrected for the resolution of the experiment) — this accounts for the genes being on the same chromosome.

The equivalent Fokker-Planck equation is:

$$\frac{\partial P_{ij}(r_{ij}, t)}{\partial t} = \frac{1}{\gamma_{ij}}\frac{\partial}{\partial r_{ij}}[\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}} P_{ij}(r_{ij}, t)] + D \times \frac{\partial^2 P(r_{ij}, t)}{\partial r_{ij}^2},$$

where the initial condition is dropped for simplicity and $P_{ij}(r_{ij}, t)$ is the probability distribution to have a distance r_{ij} at time t specific to that gene pair. We then set the left hand of the equation equal to zero, defining the steady state distance distribution ($P_{ij}^s(r_{ij})$). The equation then becomes:

$$\frac{1}{\gamma_{ij}}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}} P_{ij}^s(r_{ij}) + D \times \frac{\partial P_{ij}^s(r_{ij})}{\partial r_{ij}} = 0$$

with the solution:

$$P_{ij}^s(r_{ij}) = C_{ij} \times exp(-\frac{V_{ij}(r_{ij})}{\gamma_{ij} D}),$$

where C_{ij} is a normalization constant.

From the experimental data, we can empirically determine $P^s_{ij}(r_{ij})$. To do this we took the naturally observed distance distribution and performed a deconvolution with the resolution distribution. This provided us with $P^s_{ij}(r_{ij})$ minus the resolution error and we can therefore solve for the potential with:

$$\frac{V_{ij}(r_{ij})}{\gamma_{ij}} = D[ln(C_{ij}) - ln(P^s_{ij}(r_{ij}))]$$

With this we can then simulate the Langevin equation with the Euler-Maruyama method, which results in the proper steady state distribution with the approximate diffusion coefficient.

Acknowledgments

This work would not have been possible without the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We would also like to thank the members of the Larson lab for their input especially Dr. Nadezda Fursova. Additionally, we would like to thank Dr. Alexander Englert for the useful discussions.

References

[1] Assaf Amitai, Mathias Toulouze, Karine Dubrana, and David Holcman. Analysis of Single Locus Trajectories for Extracting In Vivo Chromatin Tethering Interactions. *PLoS Computational Biology*, 11(8):1–16, 2015.

[2] Michael Babokhov, Kayo Hibino, Yuji Itoh, and Kazuhiro Maeshima. Local Chromatin Motion and Transcription. *Journal of Molecular Biology*, 432(3):694–700, 2020.

[3] Bogdan Bintu, Leslie J. Mateo, Jun Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413), 2018.

[4] Christopher H. Bohrer and Daniel R. Larson. The Stochastic Genome and Its Role in Gene Expression. *Cold Spring Harbor Perspectives in Biology*, page a040386, 2021.

[5] Christopher H. Bohrer and Jie Xiao. *Complex Diffusion in Bacteria*, volume 1267. 2020.

[6] Christopher H. Bohrer, Xinxing Yang, Shreyasi Thakur, Xiaoli Weng, Brian Tenner, Ryan McQuillen, Brian Ross, Matthew Wooten, Xin Chen, Jin Zhang, Elijah Roberts, Melike Lakadamyali, and Jie Xiao. A pairwise distance distribution correction (DDC) algorithm to eliminate blinking-caused artifacts in SMLM. *Nature Methods*, 18(6):669–677, 2021.

[7] I. Bronshtein, E. Kepten, I. Kanter, S. Berezin, M. Lindner, Abena B. Redwood, S. Mai, S. Gonzalo, R. Foisner, Y. Shav-Tal, and Y. Garini. Loss of lamin A function increases chromatin dynamics in the nuclear interior. *Nature Communications*, 6:1–9, 2015.

[8] Baohui Chen, Luke A. Gilbert, Beth A. Cimini, Joerg Schnitzbauer, Wei Zhang, Gene Wei Li, Jason Park, Elizabeth H. Blackburn, Jonathan S. Weissman, Lei S. Qi, and Bo Huang. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7):1479–1491, 2013.

[9] Hongtao Chen, Michal Levo, Lev Barinov, Miki Fujioka, James B. Jaynes, and Thomas Gregor. Dynamic interplay between enhancer–promoter topology and gene activity. *Nature Genetics*, 50(9):1296–1303, 2018.

[10] Won Ki Cho, Namrata Jayanth, Brian P. English, Takuma Inoue, J. Owen Andrews, William Conway, Jonathan B. Grimm, Jan Hendrik Spille, Luke D. Lavis, Timothée Lionnet, and Ibrahim I. Cisse. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife*, 5(MAY2016):1–31, 2016.

[11] Won Ki Cho, Namrata Jayanth, Susan Mullen, Tzer Han Tan, Yoon J. Jung, and Ibrahim I. Cissé. Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Scientific Reports*, 6(June):1–8, 2016.

[12] Won Ki Cho, Jan Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I. Cisse. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361(6400):412–415, 2018.

[13] Jonathan R. Chubb, Shelagh Boyle, Paul Perry, and Wendy A. Bickmore. Chromatin motion is constrained by association with nuclear compartments in human cells. *Current Biology*, 12(6):439–445, 2002.

[14] L. Stirling Churchman, Henrik Flyvbjerg, and James A. Spudich. A non-gaussian distribution quantifies distances measured with fluorescence localization techniques. *Biophysical Journal*, 90(2):668–671, 2006.

[15] Ibrahim I Cisse, Ignacio Izeddin, Sebastien Z Causse, Lydia Boudarene, Adrien Senecal, Leila Muresan, Claire Dugast-darzacq, and Bassam Hajj. Polymerase II Clustering in Live Human Cells. *Science*, 245(August):664–667, 2013.

[16] Stephanie Fanucchi, Youtaro Shibayama, Shaun Burd, Marc S. Weinberg, and Musa M. Mhlanga. Chromosomal contact permits transcription between coregulated genes. *Cell*, 155(3):606, 2013.

[17] Alexander Feuerborn and Peter R. Cook. Why the activity of a gene depends on its neighbors. *Trends in Genetics*, 31(9):483–490, 2015.

[18] Elizabeth H. Finn and Tom Misteli. Molecular basis and biological function of variability in spatial genome organization. *Science*, 365(6457), 2019.

[19] Takashi Fukaya, Bomyi Lim, and Michael Levine. Enhancer Control of Transcriptional Bursting. *Cell*, 166(2):358–368, 2016.

[20] Michele Gabriele, Hugo B Brandão, Simon Grosse-Holz, Asmita Jha, Gina M Dailey, Claudia Cattoglio, Tsung-Han S Hsieh, Leonid Mirny, Christoph Zechner, and Anders Sejr Hansen. Dynamics of CTCF and cohesin mediated chromatin looping revealed by live-cell imaging. *bioRxiv*, page 2021.12.12.472242, 2021.

[21] Thomas Germier, Silvia Kocanova, Nike Walther, Aurélien Bancaud, Haitham Ahmed Shaban, Hafida Sellou, Antonio Zaccaria Politi, Jan Ellenberg, Franck Gallardo, and Kerstin Bystricky. Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophysical Journal*, 113(7):1383–1394, 2017.

[22] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions with delays. *The Journal of chemical physics*, 81(25), mar 1977.

[23] Bo Gu, Tomek Swigut, Andrew Spencley, Matthew R. Bauer, Mingyu Chung, Tobias Meyer, and Joanna Wysocka. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359(6379):1050–1055, 2018.

[24] Tyler Heist, Takashi Fukaya, and Michael Levine. Large distances separate coregulated genes in living Drosophila embryos. *Proceedings of the National Academy of Sciences of the United States of America*, 116(30):15062–15067, 2019.

[25] Jonathan E. Henninger, Ozgur Oksuz, Krishna Shrinivas, Ido Sagi, Gary LeRoy, Ming M. Zheng, J. Owen Andrews, Alicia V. Zamudio, Charalampos Lazaris, Nancy M. Hannett, Tong Ihn Lee, Phillip A. Sharp, Ibrahim I. Cissé, Arup K. Chakraborty, and Richard A. Young. RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell*, 184(1):207–225.e24, 2021.

[26] Tsung Han S. Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S. Hansen, Oliver J. Rando, Robert Tjian, and Xavier Darzacq. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell*, 78(3):539–553.e8, 2020.

[27] Mengwei Hu and Siyuan Wang. Chromatin Tracing: Imaging 3D Genome and Nucleome. *Trends in Cell Biology*, 31(1):5–8, 2021.

[28] Jonathan P. Karr, John J. Ferrie, Robert Tjian, and Xavier Darzacq. The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer–promoter communication. *Genes and Development*, 36(1-2):7–16, 2022.

[29] Tae Kyung Kim and Ramin Shiekhattar. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*, 162(5):948–959, 2015.

[30] Gregory I. Lang and David Botstein. A test of the coordinated expression hypothesis for the origin and maintenance of the GAL cluster in yeast. *PLoS ONE*, 6(9):1–7, 2011.

[31] Marshall J. Levesque and Arjun Raj. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nature Methods*, 10(3):246–248, 2013.

[32] Michal Levo, João Raimundo, Xin Yang Bing, Zachary Sisco, Philippe J Batut, Sergey Ryabichko, Thomas Gregor, and Michael S Levine. Transcriptional coupling of distant regulatory genes in living embryos. (March 2021), 2022.

[33] Bomyi Lim, Tyler Heist, Michael Levine, and Takashi Fukaya. Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell*, 70(2):287–296.e6, 2018.

[34] W. F. Marshall, A. Straight, J. F. Marko, J. Swedlow, A. Dernburg, A. Belmont, A. W. Murray, D. A. Agard, and J. W. Sedat. Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology*, 7(12):930–939, 1997.

[35] Tom Misteli. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*, 183(1):28–45, 2020.

[36] Ryosuke Nagashima, Kayo Hibino, S. S. Ashwin, Michael Babokhov, Shin Fujishiro, Ryosuke Imai, Tadasu Nozaki, Sachiko Tamura, Tomomi Tani, Hiroshi Kimura, Michael Shribak, Masato T. Kanemaki, Masaki Sasai, and Kazuhiro Maeshima. Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *Journal of Cell Biology*, 218(5):1511–1530, 2019.

[37] Tadasu Nozaki, Ryosuke Imai, Mai Tanbo, Ryosuke Nagashima, Sachiko Tamura, Tomomi Tani, Yasumasa Joti, Masaru Tomita, Kayo Hibino, Masato T. Kanemaki, Kerstin S. Wendt, Yasushi Okada, Takeharu Nagai, and Kazuhiro Maeshima. Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Molecular Cell*, 67(2):282–293.e7, 2017.

[38] Dino Osmanović and Yitzhak Rabin. Dynamics of active Rouse chains. *Soft Matter*, 13(5):963–968, 2017.

[39] Horng D. Ou, Sébastien Phan, Thomas J. Deerinck, Andrea Thor, Mark H. Ellisman, and Clodagh C. O'Shea. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349), 2017.

[40] Heta P. Patel, Stefano Coppola, Wim Pomp, Ineke Brouwer, and Tineke L. Lenstra. DNA supercoiling restricts the transcriptional bursting of neighboring eukaryotic genes. pages 1–31, 2022.

[41] Porfirio Quintero-Cadena and Paul W. Sternberg. Enhancer sharing promotes neighborhoods of transcriptional regulation across eukaryotes. *G3: Genes, Genomes, Genetics*, 6(12):4167–4174, 2016.

[42] Joseph Rodriguez, Gang Ren, Christopher R. Day, Keji Zhao, Carson C. Chow, and Daniel R. Larson. Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. *Cell*, 176(1-2):213–226.e18, jan 2019.

[43] Jun-Han Su, Pu Zheng, Seon Kinrot, Bogdan Bintu, Xiaowei Zhuang, and Seon S Kinrot. Cell Chromosome-and genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell*, pages 1–19, 2020.

[44] Mengyi Sun and Jianzhi Zhang. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLoS Genetics*, 15(9):1–30, 2019.

[45] Marcel Tarbier, Sebastian D. Mackowiak, João Frade, Silvina Catuara-Solarz, Inna Biryukova, Eleni Gelali, Diego Bárcena Menéndez, Luis Zapata, Stephan Ossowski, Magda Bienko, Caroline J. Gallant, and Marc R. Friedländer. Nuclear gene proximity and protein interactions shape transcript covariations in mammalian single cells. *Nature Communications*, 11(1):1–12, 2020.

[46] Hao Tian, Ying Yang, Sirui Liu, Hui Quan, and Yi Qin Gao. Toward an understanding of the relation between gene regulation and 3D genome organization. *Quantitative Biology*, 8(4):295–311, 2020.

[47] Robert Tjian Tsung-Han S. Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S. Hansen, Xavier Darzacq. Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. *bioRxiv*, 2021.

[48] Anat Vivante, Irena Bronshtein, and Yuval Garini. Chromatin Viscoelasticity Measured by Local Dynamic Analysis. *Biophysical Journal*, 118(9):2258–2267, 2020.

[49] Yihan Wan, Dimitrios G. Anastasakis, Joseph Rodriguez, Murali Palangat, Prabhakar Gudla, George Zaki, Mayank Tandon, Gianluca Pegoraro, Carson C. Chow, Markus Hafner, and Daniel R. Larson. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*, 184(11):2878–2895.e20, 2021.

[50] Haiqing Xu, Jing Jing Liu, Zhen Liu, Ying Li, Yong Su Jin, and Jianzhi Zhang. Synchronization of stochastic expressions drives the clustering of functionally related genes. *Science Advances*, 5(10), 2019.

[51] Yizhou Zhu and Yousin Suh. Ultrafine mapping of chromosome conformation at hundred basepair resolution reveals regulatory genome architecture. 2019.

[52] Oriana Q.H. Zinani, Kemal Keseroğlu, Ahmet Ay, and Ertuğrul M. Özbudak. Pairing of segmentation clock genes drives robust pattern formation. *Nature*, 589(7842):431–436, 2021.